Western University
Scholarship@Western

7-19-2023 3:00 PM

Advances in Copula Estimation and Distribution Theory

Yishan Zang,

Supervisor: Provost, Serge B., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree
in Statistics and Actuarial Sciences
© Yishan Zang 2023

Abstract

This dissertation initially features distributional results related to copulas. Four distinct copula density estimation methodologies, including Bernstein's polynomial approximation, are proposed and criteria for the selection of their tuning parameters are provided. These four approaches were found to produce similar density estimates, which validates their suitability. Moreover, the copula associated with a Wiener process and its running maximum is determined, and an illustrative numerical example is presented. The principal properties of Spearman's, Kendall's, Blomqvist' and Hoeffding's measures of association as well as their representations in terms of copulas are also discussed. Then, a novel method that is based on an arctangent transformation is introduced for classifying the tail behaviour of continuous probability laws. The resulting categories prove consistent with those obtained by applying available criteria. As well, approximations to the distributions of quadratic forms in gamma, inverse Gaussian, binomial and Poisson random variables, which hinge on the determination of their moments via a symbolic approach, are proposed and several applications are pointed out. Additionally, an accurate density approximation that relies of an extension of the generalized gamma distribution is introduced and the case of quadratic forms in Hermitian matrices in complex Gaussian vectors is also addressed. Finally, a methodology involving the use of Fritsch-Carlson monotonic interpolating splines and the Kulback-Leibler measure of divergence is proposed for quantifying the proportion of information that is contained in collections of distributional moments.

Keywords: Copulas, Brownian motion processes, bivariate density approximation, measures of association, data modeling, classification of tail behavior, quadratic forms, information in moments.

Summary for Lay Audience

In the first part of this dissertation, the focus is on copulas which are mathematical tools utilized to describe the relationships between random variables. Four distinct methods for estimating the distribution of copulas are proposed, with the objective consisting of capturing the patterns and characteristics of sets of observations on two variables. These methods were found to produce similar results, which confirms their effectiveness. The study also explores the copula associated with a specific probabilistic model called the Wiener process and its running maximum. An illustrative example is provided. The thesis further examines four measures of association and explains how they can be represented as copulas.

A novel methodology is introduced for classifying the behavior of the tails (extreme values) of continuous probability distributions. This method relies on an arctangent transformation and proves to be consistent with existing criteria.

The dissertation also presents approximations to the distributions of quadratic forms in various types of random variables. These approximations depend on the determination of their moments which are obtained by making use of a symbolic computational approach. The statistical applications of such quadratic forms are discussed. Furthermore, an accurate approximation involving the density of a generalized gamma distribution is introduced. The case of Hermitian quadratic forms in complex Gaussian vectors is addressed as well.

Finally, a methodology is proposed for quantifying the amount of information contained in sets of distributional moments. This approach relies on the use of a non-decreasing curve and a certain measure of divergence.

Overall, this research contributes to an improved understanding of copulas and other distributional concepts. The findings and methods presented in this thesis have the potential to enhance our ability to analyze complex data sets and make informed statistical modeling decisions.

The journey of a thousand miles starts with

a single step.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation comprises several innovative distributional results in connection with the non-parametric estimation of copula density functions, the representation of four principal measures of association in terms of copulas, the characterization of the tail behaviour of continuous random variables, accurate approximations of the density functions of various types of quadratic forms, and the proportion of information that is contained in a set of distributional moments.

Since the thesis format is 'integrated-article', the six main chapters that follow already possess their own introduction. Accordingly, this introductory chapter will essentially consist of describing their contents. It also ought to be noted that given the format of this thesis, some redundancies may occur.

The next three chapters tackle various distributional aspects arising in connection with copulas. Actually, copulas encapsulate all the dependencies between two or more variables. As pointed out in the next chapter, they are currently being utilized in numerous fields of scientific investigation. In this dissertation, we shall focus on the bivariate case.

For a given bivariate data set, the empirical distribution of the resulting copula is discrete. To obtain a continuous estimate of the copula density function, the kernel approach is employed in the second chapter. The resulting estimates manifestly exhibit more flexibility than known functional copula densities. Another technique being discussed is the approximation of

their distribution by means of Bernstein polynomials. An example involving two stock prices is presented. A technique for limiting the support of a bivariate density estimate to a parallelogram is also introduced in that chapter. Finally, a copula density function associated with a Brownian motion process and its running maximum is obtained. The contents of this chapter can be found in part in papers [2] and [3] as listed in the cv.

In the third chapter, we are first advocating with proper justification a repositioning of the pseudo-observations which constitute the distributional support of empirical copulas. Then, four approaches are discussed for securing copula density estimates, including bivariate least-squares approximations, kernel density estimation, the differentiation of linearized empirical copulas and the Bernstein polynomial approximation. Criteria for the determination of their associated tuning parameters such as polynomial orders and kernel bandwidths, will be introduced as well, in order to obtain suitable density estimates. It will also be explained that a joint density function can be constructed with much flexibility from its marginals and a copula density estimate. Illustrative examples involving an actual data set will be presented. The proposed methodologies will as well be applied to a sample generated from a known copula distribution in order to validate their effectiveness. This chapter corresponds to manuscript [5] which is included in the cv as a working paper.

The fourth chapter focuses on four known measures of association, namely, Spearman's, Kendall's, Blomqvist's and Hoeffding's coefficients, and provides representations thereof in terms of copulas, in addition to pointing out certain related distributional results of interest. This chapter contains certain derivations that do not seem to be available in the literature and also includes missing steps that complete some published proofs. The effectiveness of these measures of association in assessing the trends present in data sets that were generated from five distinctive patterns, is assessed in a numerical study. To our knowledge, the four major measures of association that are discussed therein along with their representations in terms of copulas, have not been previously treated altogether in a single source. This chapter's materials are included in manuscript [4] as listed in the cv.

The topics covered in the three subsequent chapters which are all distributional in nature, are unrelated. In the fifth chapter, a methodology based on an arctangent transformation is proposed to characterize the tail behaviour of a distribution. This approach is applied to an array of widely utilized distributions, the resulting characterizations of their tail behaviour being generally consistent with those determined by making use of known criteria. In the case of a sample of observations, one must initially obtain a density estimate to which the proposed approach can then be readily applied. Naturally, the larger the sample, the more reliable are the results. These results can be found in paper [1] which appears in the section *Published Papers* of the cv.

In the sixth chapter, a symbolic approach is employed to evaluate the moments of quadratic forms whose vectors are not necessarily normally distributed. Based on those moments, the density functions the quadratic forms in various types of randoms variables are approximated with great accuracy. Several illustrative examples are presented. For each one, the empirical distribution of a given quadratic form is determined from a large simulated sample. The case of Hermitian quadratic forms in complex Gaussian vectors is also discussed. Finally, some statistical properties of a distribution referred to as the extended generalized gamma distribution are provided. This distribution is then utilized to approximate the density functions of positive definite and indefinite quadratic forms in normal vectors.

The seventh and last chapter aims at assessing the amount of information that is contained in a set of distributional moments. No related results could be found after a thorough search of the statistical literature. Nevertheless, this constitutes a crucial problem in connection with the application of moment-based density estimation methodologies. Indeed, if it is known that beyond moment r, only a minute proportion of the total amount of information is available, then statistical procedures could be reliably based on only the first r moments. Appealing to a theorem that relates sample moments to sample points, the relative information contained in h moments will be quantified via the h most representative points of a sample which, as will be explained, are the set of points having the lowest discrepancy with respect to their

distribution. Given a certain number of points, an empirical cumulative distribution function is approximated by means of a Fritsch-Carlson monotonic piecewise cubic interpolation spline, which is then differentiated to produce a density estimate that is compared to $f(x)$, the target density, by making use of of the Kullback-Leibler divergence measure. The total amount of information that could be gained is taken to be the Kullback-Leibler divergence between a non-informative constant density function and $f(x)$, and the proportion of information that is contained in the first h moments of a distribution can then be determined. The proposed methodology which is applied to two distributions and two data sets, yields results that prove consistent with expectations.

Since much time and effort were devoted to the programming of the original methodologies featured in this dissertation, the *Mathematica* code that was created for their implementation is provided in the Appendix.

Chapter 2

Copula Approximation and an Application to a Brownian Motion Process

2.1 Introduction

Copulas are principally utilized for modeling dependency features in multivariate distributions. As measures of dependence, they have for instance found applications in reliability theory, signal processing, finance, geodesy, hydrology and medicine.

The key idea behind copulas is that the joint distribution of two or more random variables can be represented in terms of their marginal distributions and a specific correlation structure. Copulas enable one to separate the effect that the dependence between the variables is causing from the contribution of each of the marginal variables. We shall address the two-dimensional case in this chapter.

We now review some basic definitions and theorems in connection with copulas. Additional results are discussed for instance in Cherubini *et al.* (2004, 2012), Denuit *et al.* (2005), Joe (1997), Nelsen (2006), and Sklar (1959).

In this framework, a copula function is a bivariate distribution whose support is the unit square $\mathbb{1}^2 = [0\,,1]^2$ and whose marginals are uniformly distributed. A more formal definition

5

is now provided.

A function $C : \mathbb{1}^2 \mapsto \mathbb{1}$ is a bivariate copula if it satisfies the two following properties:

1. For every $u, v \in \mathbb{1}$, $C(u, 1) = u$, $C(1, v) = v$ and

$$C(u, 0) = C(0, v) = 0.$$

2. For every $u_1, u_2, v_1, v_2 \in \mathbb{1}$ such that $u_1 \leq u_2$ and $v_1 \leq w_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

This last inequality implies that $C(u, v)$ is increasing in both variables.

The following theorem which was introduced by Sklar (1959), constitutes a seminal result in the theory of copulas and its application.

Result 1   *A Bivariate Formulation of Sklar's Theorem*

Let $F_{\mathbf{X}}(x_1, x_2)$ be the joint cumulative distribution function (cdf) of the random variables X_1 and X_2 whose continuous marginal distribution functions are denoted by $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$. Then, there exists a unique bivariate copula $C(\cdot, \cdot) : \mathbb{1}^2 \mapsto \mathbb{1}$ such that

$$F_{\mathbf{X}}(x_1, x_2) = C\left(F_{X_1}(x_1), F_{X_2}(x_2)\right) \tag{2.1}$$

where $C(\cdot, \cdot)$ is a joint cumulative distribution function having uniform marginals. Conversely, for any continuous cumulative distribution function $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ and any copula $C(\cdot, \cdot)$, the function $F_{\mathbf{X}}(\cdot, \cdot)$ as given in (2.1) is a joint distribution function with marginal distributions $F_{X_1}(\cdot)$ and $F_{X_2}(\cdot)$.

Sklar's theorem provides a technique for constructing copulas. Indeed, the function

$$C(u, v) = F_{\mathbf{X}}\left(F_{X_1}^{-1}(u), F_{X_2}^{-1}(v)\right) \tag{2.2}$$

is a bivariate copula, where the quasi-inverses $F_{X_1}^{-1}(\cdot)$ and $F_{X_2}^{-1}(\cdot)$ are given by

$$F_{X_1}^{-1}(u) = \inf\{x_1 | F_{X_1}(x_1) \geq u\}, \quad \forall u \in (0, 1), \tag{2.3}$$

and

$$F_{X_2}^{-1}(v) = \inf\{x_2 | F_{X_2}(x_2) \geq v\}, \quad \forall\, v \in (0,\, 1). \tag{2.4}$$

Copulas are invariant with respect to strictly increasing transformations. More specifically, letting X_1 and X_2 be two continuous random variables whose associated copula is $C(\cdot\,,\cdot)$, if $\alpha(\cdot)$ and $\beta(\cdot)$ are two strictly increasing functions and $C_{\alpha,\beta}(\cdot\,,\cdot)$ is the copula obtained from $\alpha(X_1)$ and $\beta(X_2)$, then for all $(u, v) \in \mathbb{1}^2$, $C_{\alpha,\beta}(u, v) = C(u, v)$.

We shall denote the *probability density function* or pdf corresponding to the copula $C(u, v)$ by

$$c(u,\, v) = \frac{\partial^2}{\partial u \partial v} C(u, v). \tag{2.5}$$

The following relationship between the joint density function of X_1 and X_2, denoted by $f_{\mathbf{X}}(\cdot\,,\cdot)$, and the associated pdf $c(\cdot\,,\cdot)$ can readily be obtained from (2.1) and (2.5):

$$f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2)\, c(F_{X_1}(x_1), F_{X_2}(x_2)) \tag{2.6}$$

where $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ respectively denote the marginal density functions of X_1 and X_2. Accordingly, the copula density function can be expressed as follows:

$$c(u,\, v) = \frac{f_{\mathbf{X}}(F_{X_1}^{-1}(u),\, F_{X_2}^{-1}(v))}{f_{X_1}(F_{X_1}^{-1}(u))\, f_{X_2}(F_{X_2}^{-1}(v))}. \tag{2.7}$$

Given a random sample $(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})$ from the continuous random variables X_1 and X_2, let

$$(u_i, v_i) = (F_{X_1}(x_{1i}), F_{X_2}(x_{2i})), \quad i = 1, \ldots, n$$

where $F_{X_1}(\cdot)$ and $F_{X_2}(\cdot)$ are the usually unknown marginal *cumulative distribution functions* or cdf's of X_1 and X_2. Throughout this chapter, n shall denote the sample size. Now, since the underlying distributions of the variables are assumed to be continuous, the x_{1i}'s are, in theory, all distinct, and so are the x_{2i}'s. Should a data set happen to contain replicates due to rounding,

the data could be minimally randomly perturbed to ensure that the ranks on each variable be distinct.

In Section 2.2, we discuss the use of kernel density estimates (kde's) in connection with the determination of copula densities. A moment-based approximation technique that applies to bivariate continuous functions is presented in Section 2.3. As a result, flexible copula densities– as opposed to the functional forms that are available—are obtained in a convenient form. An illustrative example involving the closing prices of two stocks is featured in Section 2.4. A technique for delimiting the support of a bivariate density estimate is introduced in Section 2.5. A copula associated with a Brownian motion process and its running maximum is considered in Section 2.6 and an application is presented. Bernstein polynomial approximations to empirical copulas and their density functions are discussed in Section 2.7 which also includes illustrative examples. It should be noted that certain aspects of this chapter could be regarded as preparatory material in connection with some of the results to be developed in the next chapter.

2.2   Copula density based on kernel density estimation

Suppose that a bivariate data set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is at hand. Then, on applying the kernel density estimation (kde) method, one can obtain an estimate of the bivariate probability density of $\mathbf{X}$. The kernel density estimator is given by

$$\hat{f}_H(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_H(\mathbf{x} - \mathbf{x_i}) = \frac{1}{n} |H|^{-1/2} \sum_{i=1}^{n} K(H^{-1/2}(\mathbf{x} - \mathbf{x_i})), \tag{2.8}$$

where K denotes the kernel – a continuous bivariate density function – and H is a 2×2 bandwidth (or smoothing) matrix which is symmetric and positive definite.

In light of equation (2.7), the copula density can also be represented as follows:

$$\hat{c}(u, v) = \frac{\hat{f}_{\mathbf{X}}(Q_{X_1}(u), Q_{X_2}(v))}{\hat{f}_{X_1}(Q_{X_1}(u)) \hat{f}_{X_2}(Q_{X_2}(v))}, \tag{2.9}$$

where $\hat{f}_{\mathbf{X}}$ is a bivariate kde, $\hat{f}_{X_1}$ and $\hat{f}_{X_2}$ are obtained by integration as explained below, and $Q(\cdot)$ denotes the estimated quantile function. The resulting copula density estimate can be used to obtain a joint density estimate in light the decomposition given in (2.6). This approach provides more flexibility than the direct joint density estimation methods that are usually employed.

Thus, one has to determine the marginal density functions $\hat{f}_{X_1}$ and $\hat{f}_{X_2}$ and the inverse cumulative distribution functions (cdf's), Q_{X_1} and Q_{X_2}.

The marginal density function $\hat{f}_{X_1}$ can be obtained by integrating out X_2 from the bivariate kde $\hat{f}_{\mathbf{X}}$:

$$\hat{f}_{X_1}(x_1) = \int_A \hat{f}_{\mathbf{X}}(x_1, x_2) \; dx_2, \tag{2.10}$$

where A is the support of X_2. Similarly, the marginal density function $\hat{f}_{X_2}$ can be obtained as follows:

$$\hat{f}_{X_2}(x_2) = \int_{A'} \hat{f}_{\mathbf{X}}(x_1, x_2) \; dx_1, \tag{2.11}$$

where A' is the support of X_1.

For the inverse cdf's Q_{X_1} and Q_{X_2}, we can use a moment-based method or the method of least squares to determine an approximation in polynomial form.

We are then ready to evaluate the copula density $\hat{c}$. This copula density is based on the bivariate kernel density estimate since it involves $\hat{f}_{\mathbf{X}}$ and the marginal densities $\hat{f}_{X_1}$ and $\hat{f}_{X_2}$ along with their inverse cdfs, Q_{X_1} and Q_{X_2}, all of which are obtained from the kde $\hat{f}_{\mathbf{X}}$.

A simple representation of this copula density that is suitable for reporting purposes or carrying out further algebraic manipulation can also be determined by making use of Result 2 which is stated in Section 2.3.

2.3   A moment-based bivariate polynomial approximation of a copula density

Once a copula density is determined from (2.9), it can be approximated by the product of a base density and a bivariate polynomial whose coefficients are obtained from the joint moments associated with the copula density. The proposed procedure for achieving this is described in the next result which is a bivariate extension of a proposition stated in Provost (2005).

Result 2   *Moment-Based Bivariate Polynomial Approximation*

Let $f_{\mathbf{Y}}(y_1, y_2)$ be the density function of a bivariate continuous random variable $\mathbf{Y}$ defined in the rectangle $(l_1, u_1) \times (l_2, u_2)$. The joint moments of orders i and j of $f_{\mathbf{Y}}$ are denoted as

$$\mu_{\mathbf{Y}}(i, j) \equiv \int_{l_1}^{u_1} \int_{l_2}^{u_2} y_1^i \, y_2^j \, f_{\mathbf{Y}}(y_1, y_2) \, \mathrm{d}y_2 \, \mathrm{d}y_1. \tag{2.12}$$

Let $\psi_{\mathbf{Y}}(y_1, y_2)$ be a nontrivial base density function of (y_1, y_2) on the same support $(l_1, u_1) \times (l_2, u_2)$ whose distributional features are reminiscent of those of $f_{\mathbf{Y}}(y_1, y_2)$. In the case of a copula, a uniformly distributed based density is generally suitable. We denote the joint moments of $\psi_{\mathbf{Y}}$ as

$$m_{\mathbf{Y}}(i, j) \equiv \int_{l_1}^{u_1} \int_{l_2}^{u_2} y_1^i \, y_2^j \, \psi_{\mathbf{Y}}(y_1, y_2) \, \mathrm{d}y_2 \, \mathrm{d}y_1. \tag{2.13}$$

Assuming that the sequence $\mu_{\mathbf{Y}}(i, j)$, $i = 0, 1, 2, \ldots$, $j = 0, 1, 2, \ldots$ uniquely defines the distribution of $\mathbf{Y}$, the density function of $\mathbf{Y}$ can be approximated by

$$f_n(y_1, y_2) = \psi(y_1, y_2) \sum_{i=0}^{n} \sum_{j=0}^{n} \xi_{i,j} \, y_1^i \, y_2^j, \tag{2.14}$$

where $\xi_{i,j}$ can be determined by letting

$$\int_{l_1}^{u_1} \int_{l_2}^{u_2} y_1^h \, y_2^g \, f_{\mathbf{Y}}(y_1, y_2) \, \mathrm{d}y_2 \, \mathrm{d}y_1 = \int_{l_1}^{u_1} \int_{l_2}^{u_2} \psi(y_1, y_2) \sum_{i=0}^{n} \sum_{j=0}^{n} \xi_{i,j} \, y_1^{i+h} \, y_2^{j+g} \, \mathrm{d}y_2 \, \mathrm{d}y_1, \tag{2.15}$$

$h = 0, 1, \ldots, n;\ g = 0, 1, \ldots, n.$

The above equation can be re-expressed as

$$\mu_{\mathbf{Y}}(h, g) = \sum_{i=0}^{n} \sum_{j=0}^{n} \xi_{i,j}\, m_{\mathbf{Y}}(i + h, y + g), \tag{2.16}$$

$h = 0, 1, \ldots, n;\ g = 0, 1, \ldots, n$ *where $m_{\mathbf{Y}}(\cdot, \cdot)$ denotes the joint moments associated with $\psi(\cdot)$.*

Thus, we can determine the polynomial coefficients $\xi_{i,j}$ of $f_n(y_1, y_2)$ from the moments of $f_{\mathbf{Y}}(\cdot)$ and $\psi_{\mathbf{Y}}(\cdot)$ by solving the linear system (2.16).

A technique for estimating a copula density is described in Section 2.2. Then, the result presented in this section can be utilized to express it in a convenient form. The base function $\psi_{\mathbf{Y}}(y_1, y_2)$ can be the density function of the continuous uniform distribution on the support or some other density function selected on the basis of the distributional features of the copula density. The degree n used in the polynomial adjustment should be selected so that f_n provides an accurate approximation to the estimate of the copula density, which can be determined for instance by evaluating their integrated squared differences.

2.4   An application: Two stocks' closing prices

The two stocks selected are GOOG (Alphabet Inc.) and AAPL (Apple Inc.). The bivariate data are the daily closing prices of (GOOG, AAPL) in the entire year 2019. Each component of the data has been standardized.

2.4.1   Moment-based approach to approximating the copula density

The bivariate kernel density estimate and the marginals are plotted in Figures 2.1-2.3.
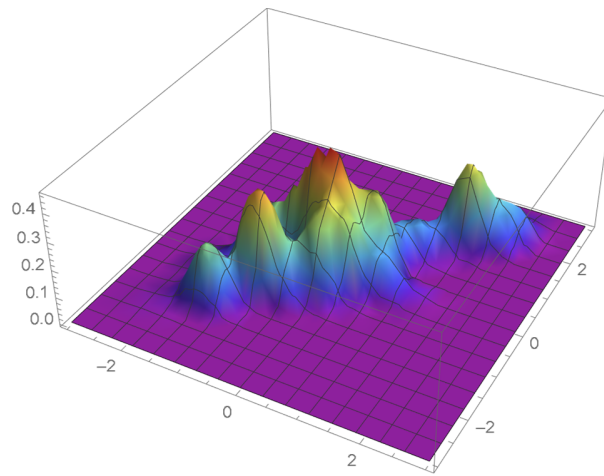
Figure 2.1: Bivariate kde
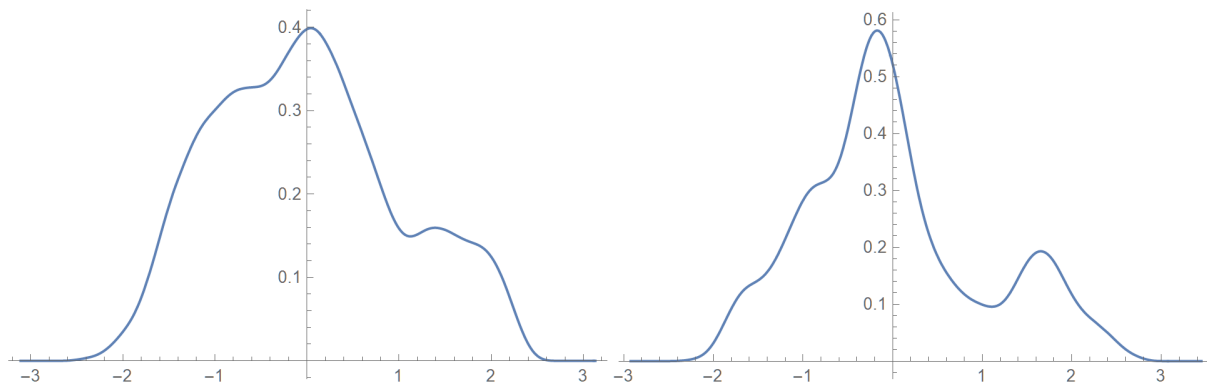
Figure 2.2: The marginal pdf of X_1 obtained by integration

Figure 2.3: The marginal pdf of X_2 obtained by integration

The marginal densities can be approximated by polynomials upon proceeding as explained in the remainder of this section. Consider for example,

$$f_{X_1}(x_1) = \int_C f_{\mathbf{X}}(x_1, x_2) \, dx_2.$$

We can generate a list of points $(z_i, f_{X_1}(z_i))$ from $f_{X_1}(x_1)$ where the z_i's are equidistant and then, fit a polynomial based on the method of least squares. We proceed similarly for f_{X_2}.

We denote the marginal cdf's as $F_{X_1}(x_1) = \int_{y=l_1}^{x_1} f_{X_1}(y) \, dy$ and $F_{X_2}(x_2) = \int_{y=l_2}^{x_2} f_{X_2}(y) \, dy$. Then, for instance, one can generate a list of sample points $(F_{X_1}(x_{1i}), x_{1i})$ from $F_{X_1}(x_1)$, using a interpolation technique to obtain a continuous function. The approximation to the inverse cdf of X_1 can be determined by applying a moment-based approximation method which is actually the univariate counterpart to the approximation method described in Section 2.3, to that continuous function. One would then proceed similarly for F_{X_2}. The resulting estimates of the inverse cdf's, Q_{X_1} and Q_{X_2}, are plotted in Figs 2.4 and 2.5.
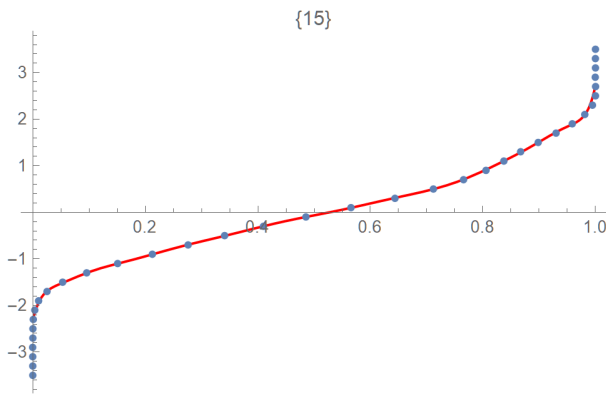
Figure 2.4: Polynomially approximated inverse cdf Q_{X_1}
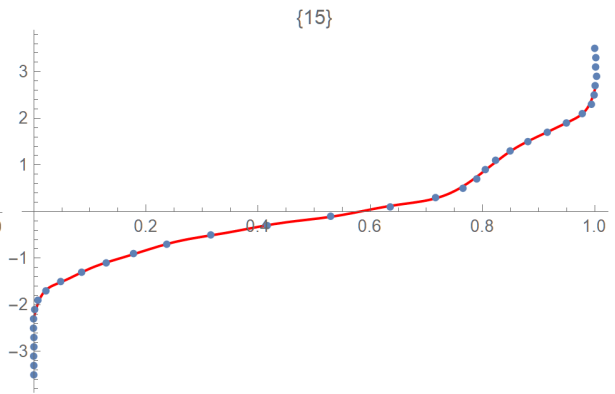
Figure 2.5: Polynomially approximated inverse cdf Q_{X_2}

Given the bivariate kde and the estimates of f_{X_1}, f_{X_2}, Q_{X_1} and Q_{X_2}, the copula density is determined from the representation specified in (2.9). Then, one can apply Result 2 as given in Section 2.3 to obtain an approximation to the copula density. The copula density and its bivariate polynomial estimate of degree 5 in each variable are plotted in Figures 2.6

and 2.7. The base function ψ being utilized is the bivariate continuous uniform distribution. This approach which is based on the joint moments of the copula density estimate provides a conveniently simple representation of the estimated copula density.
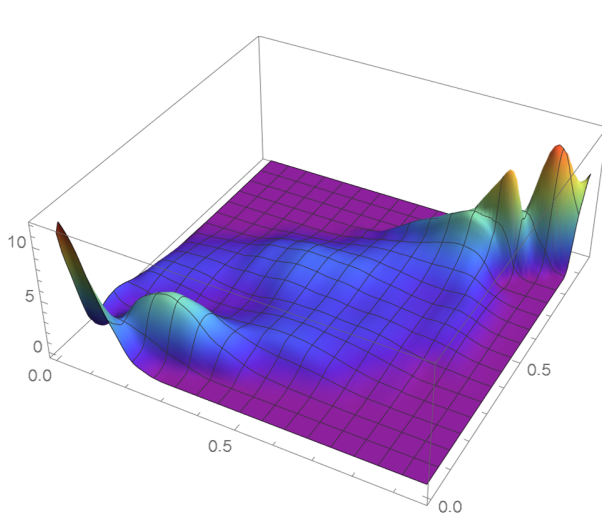
Figure 2.6: The copula density $\hat{c}$
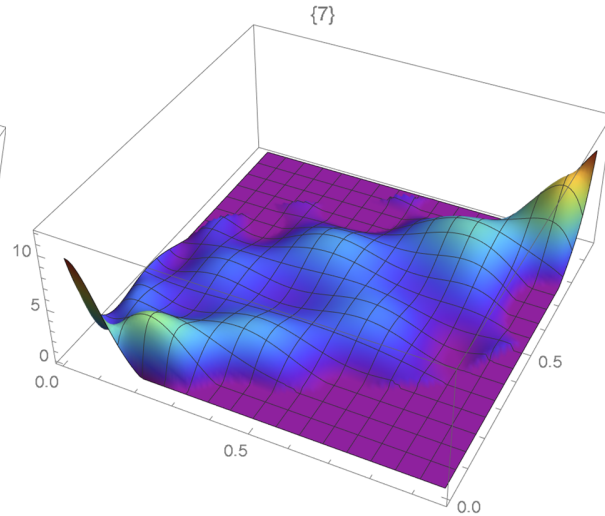
Figure 2.7: The bivariate polynomial estimate of the copula density using a uniform base density

2.4.2   Determining the marginals separately and using 'InverseFunction'

An alternative approach is described in this subsection. In this instance, the marginal densities are estimated with kde's from the data available on each variable rather than by integrating out the other variable from the joint density. As well, we make use of the built-in *Mathematica* function *InverseFunction* to obtain the inverse of the distribution functions. We observed that proceeding this way yields a more accurate copula density. Accordingly, it is advocated that this approach be utilized in future work. However, when the built-in function *InverseFunction* cannot be applied, one can always employ the methodology described in the previous subsection.

The scatter plot and marginal kde's are plotted in Figures 2.8-2.10. Then, we estimate the inverse cdf's from the cdf's determined from the marginal kde's by resorting to the function

InverseFunction. Finally, the copula density is determined by making use of the kde's and the approximations to the inverse cdf's in conjunction with formula (2.9). The estimate of the inverse cdf's and the resulting copula density f_c are plotted in Figures 2.11-2.13.
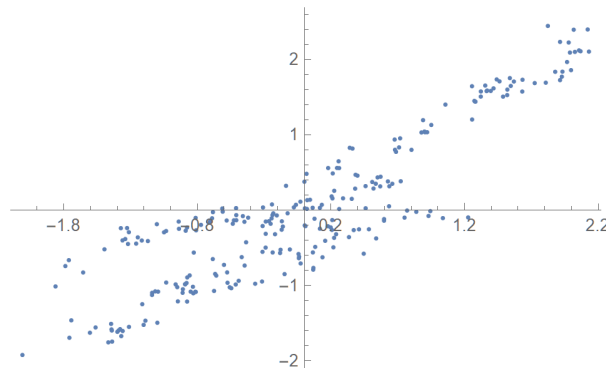
Figure 2.8: The scatter plot
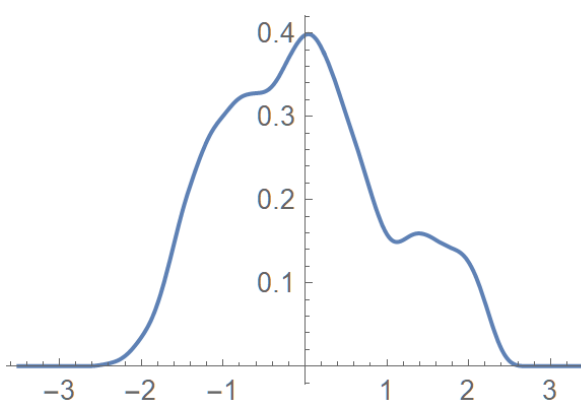
Figure 2.9: The marginal kde of X_1
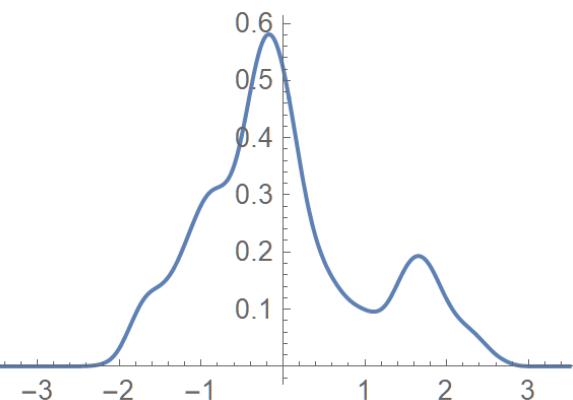
Figure 2.10: The marginal kde of X_2
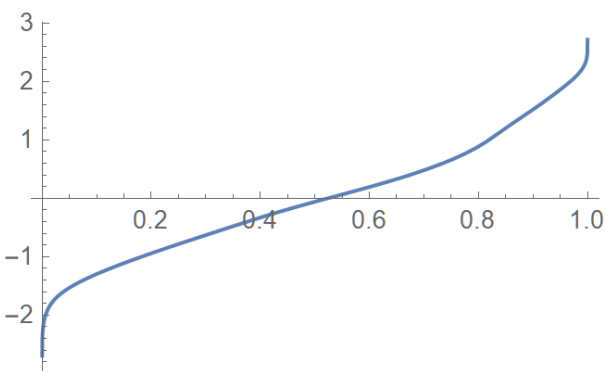
Figure 2.11: Approximated inverse cdf Q_{X_1} obtained with 'InverseFunction'
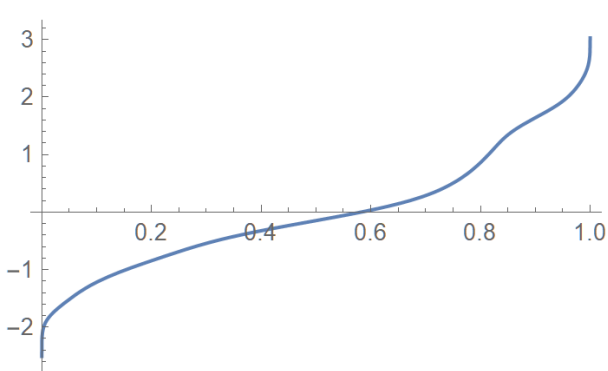
Figure 2.12: Approximated inverse cdf Q_{X_2} obtained with 'InverseFunction'
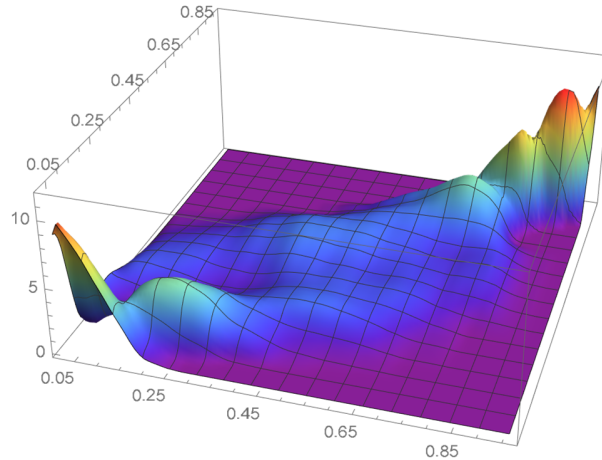
Figure 2.13: The copula density $\hat{c}$

2.5   Parallelograms as domains for joint density functions

We now propose a technique for confining the domain of a bivariate distribution to a parallelogram. For instance, this approach ought to enable one to readily identify outliers and properly exclude them from further statistical analyses.

Let Y_1 and Y_2 be two dependent random variables. On applying the transformation

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \Sigma^{-1/2} \begin{pmatrix} Y_1 - \mu_{Y_1} \\ Y_2 - \mu_{Y_2} \end{pmatrix}, \tag{2.17}$$

where Σ is the covariance matrix associated with $(Y_1, Y_2)'$ and the μ_{Y_i}'s are the means of Y_i's, $i = 1, 2$, one obtains the standardized random variables X_1 and X_2. The marginal densities of X_1 and X_2 are then estimated for instance by making use of kde's or certain techniques based on moments. The product of the two marginal density estimates whose support is taken to be a rectangle outside of which the density function is negligible, is used as the domain of a base density function that can be adjusted with a bivariate polynomial as explained in Section 2.3. This yields an estimate of the joint density of X_1 and X_2. We note that even though X_1 and X_2 are uncorrelated, they may still be dependent to some extent. The last step consists of applying

the inverse transformation

$$
\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \Sigma^{1/2} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{pmatrix} \tag{2.18}
$$

in order to obtain a joint density estimate for Y_1 and Y_2 whose support can be bounded by a parallelogram as explained below. For instance, such a joint density can then be utilized to determine a copula density.

We now show that when applying the inverse transformation, the support of the joint density function of Y_1 and Y_2 turns out to be a parallelogram. Let the domain of the joint density of (X_1, X_2) be $(a_1, a_2) \times (b_1, b_2)$. Denote the line specified by $x_1 = a_1$ as ℓ_1. We have the following equations resulting from the transformation (2.18):

$$
y_1 = \alpha_{11} a_1 + \alpha_{12} x_2 + \mu_{Y_1}, \tag{2.19}
$$

$$
y_2 = \alpha_{12} a_1 + \alpha_{22} x_2 + \mu_{Y_2}, \tag{2.20}
$$

which are equivalent to

$$
\frac{\alpha_{22}}{\alpha_{12}} x_1 = \frac{\alpha_{22}}{\alpha_{12}} \alpha_{11} a_1 + \alpha_{22} x_2 + \frac{\alpha_{22}}{\alpha_{12}} \mu_{Y_1}, \tag{2.21}
$$

$$
y_2 = \alpha_{12} a_1 + \alpha_{22} x_2 + \mu_{Y_2}, \tag{2.22}
$$

the α_{ij}'s being the entries of the matrix

$$
\Sigma^{1/2} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}. \tag{2.23}
$$

On subtracting (2.21) from (2.22), we obtain the transformed line

$$
\tilde{\ell}_1 : y_2 = \frac{\alpha_{22}}{\alpha_{12}} y_1 - (\alpha_{11}\alpha_{22} - \alpha_{12}^2) \frac{a_1}{\alpha_{12}} + \mu_{Y_2} - \frac{\alpha_{22}}{\alpha_{12}} \mu_{Y_1}. \tag{2.24}
$$

On proceeding similarly, one can obtain the other three transformed lines. The domain of Y_1 and Y_2 will then be a parallelogram bounded by the following four lines:

$$\tilde{\ell}_1 : y_2 = \frac{\alpha_{22}}{\alpha_{12}} y_1 - (\alpha_{11}\alpha_{22} - \alpha_{12}^2)\frac{a_1}{\alpha_{12}} + \mu_{Y_2} - \frac{\alpha_{22}}{\alpha_{12}}\mu_{Y_1}, \text{ when } x_1 = a_1; \qquad (2.25)$$

$$\tilde{\ell}_2 : y_2 = \frac{\alpha_{12}}{\alpha_{11}} y_1 - (\alpha_{11}\alpha_{22} - \alpha_{12}^2)\frac{b_1}{\alpha_{11}} + \mu_{Y_2} - \frac{\alpha_{12}}{\alpha_{11}}\mu_{Y_1}, \text{ when } x_2 = b_1; \qquad (2.26)$$

$$\tilde{\ell}_3 : y_2 = \frac{\alpha_{22}}{\alpha_{12}} y_1 - (\alpha_{11}\alpha_{22} - \alpha_{12}^2)\frac{a_2}{\alpha_{12}} + \mu_{Y_2} - \frac{\alpha_{22}}{\alpha_{12}}\mu_{Y_1}, \text{ when } x_1 = a_2; \qquad (2.27)$$

$$\tilde{\ell}_4 : y_2 = \frac{\alpha_{12}}{\alpha_{11}} y_1 - (\alpha_{11}\alpha_{22} - \alpha_{12}^2)\frac{b_2}{\alpha_{11}} + \mu_{Y_2} - \frac{\alpha_{12}}{\alpha_{11}}\mu_{Y_1}, \text{ when } x_2 = b_2. \qquad (2.28)$$

The joint density of Y_1 and Y_2 is then set equal to zero for any point (y_1, y_2) lying outside the parallelogram.

2.5.1   The two stock's closing prices revisited

Consider the two stocks' data as described in Section 2.4. The bivariate data is then standardized using Equation (2.17) and the support of the standardized random vector is a rectangle, which is $(-3.26, 3.04) \times (-2.46, 3.09)$. The joint density function of (Y_1, Y_2) can be determined as before except that the domain of (Y_1, Y_2) being delimited by a parallelogram as previously explained. The four lines forming its boundary are plotted in Figure 2.14. We now define the indicator function

$$I(y_1, y_2) = \begin{cases} 1, & \text{if } (y_1, y_2) \text{ lies in this domain.} \\ 0, & \text{otherwise.} \end{cases} \qquad (2.29)$$

The estimated joint density function of (Y_1, Y_2) is multiplied by this indicator function so that the density is zero when (y_1, y_2) lies outside the parallelogram. The indicator function I is shown in Figure 2.15. The joint density function of (Y_1, Y_2) and the copula density function are plotted in Figures 2.16 and 2.17.

In this case, delimiting the domain of the joint density function will not affect the resulting

estimate of copula density as obtained from the technique described in Subsection 2.4.2.
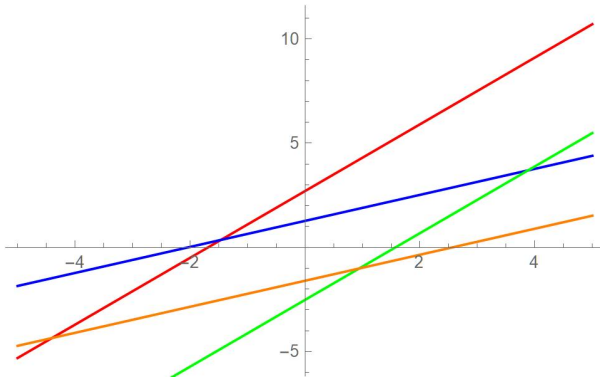
Figure 2.14: The lines delimiting the domain
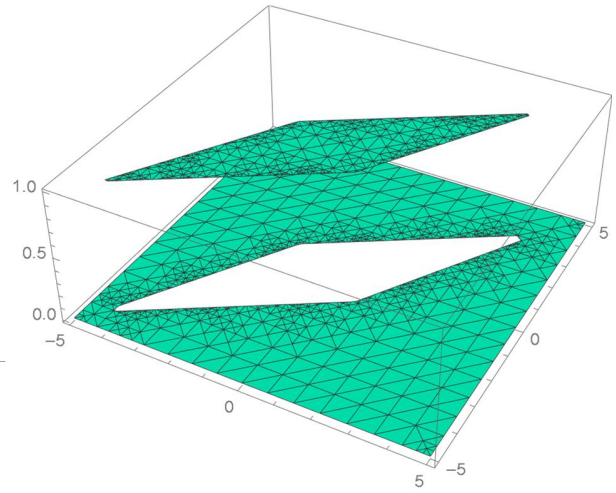
Figure 2.15: Plot of $I(y_1, y_2)$
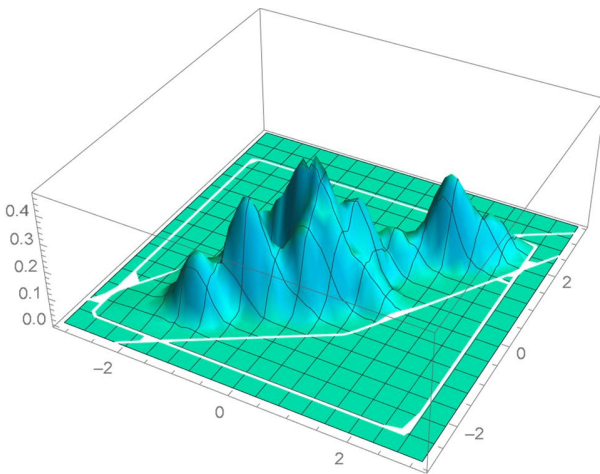
Figure 2.16: The estimated joint density of (Y_1, Y_2)
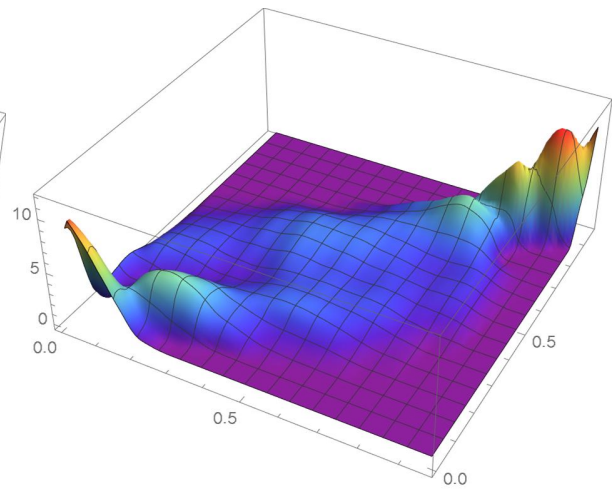
Figure 2.17: The resulting copula density

2.6   Copula associated with a Brownian motion process and its running maximum

2.6.1   Introduction

This section initially presents relevant background information on Brownian motion ($\mathcal{BM}$). Then, the connection to copulas will be discussed and appropriate references will be provided.

The mathematician Norbert Wiener provided a rigorous formulation of Brownian motion in 1918; as a result, the alternative name, Wiener process, is also utilized in the literature. $\mathcal{BM}$ constitutes a useful modelling tool in many areas such as Economics, Biology, Communications Theory, Business Administration and Quantitative Finance.

The standard $\mathcal{BM}$ will be denoted by $\{W_t\}_{t \geq 0}$. Letting $M_t = \max_{0 \leq s \leq t} W_s$, we shall consider the joint distributions of W_t and its maximum M_t, which has previously been studied by Harrison (1985) and Lee (2003), among others.

Darsow *et al.* (1992, Example 4.3) seem to have been the first to introduce a copula in connection with the Brownian motion, which is given by

$$C(u, v) = \int_0^u \Phi\left(\frac{\sqrt{t}\,\Phi^{-1}(v) - \sqrt{s}\,\Phi^{-1}(x)}{\sqrt{t - s}} \right) \mathrm{d}x, \tag{2.30}$$

for $s < t$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable. Later, Schmitz (2003) developed a copula for a Brownian motion and its supremum.

As explained for instance in Etheridge (2002), Harrison (1990), Karlin and Taylor (1975), Revuz and Yor (2005) and Rogers and Williams (2000a), the joint distribution of (W_t, M_t) and the marginal distributions of W_t and M_t are respectively given by

$$\mathbb{P}\{W_t \leq x, M_t \leq a\} \equiv F_{W_t, M_t}(x, a) = \begin{cases} \Phi\left(\frac{x}{\sqrt{t}}\right) - \Phi\left(\frac{x-2a}{\sqrt{t}}\right) & \text{if } x \leq a \\ 2\Phi\left(\frac{a}{\sqrt{t}}\right) - 1 & \text{if } x > a \end{cases}, \tag{2.31}$$

$$\mathbb{P}\{W_t \le x\} \equiv F_{W_t}(x) = \Phi\left(\frac{x}{\sqrt{t}}\right), \tag{2.32}$$

and

$$\mathbb{P}\{M_t \le a\} \equiv F_{M_t}(a) = 2\Phi\left(\frac{a}{\sqrt{t}}\right) - 1, \tag{2.33}$$

$\forall\, t \in \mathbb{R}_+$ (the set of positive real numbers).

As explained in Adès *et al.* (2022), the copula $C_{W_t, M_t}(u, v)$ generated by a standard $\mathcal{BM}$ and its maximum is

$$C_{W_t, M_t}(u, v) = \begin{cases} u - \Phi\left(\Phi^{-1}(u) - 2\Phi^{-1}\left(\frac{v+1}{2}\right)\right) & \text{if } u \le \frac{v+1}{2} \\ v & \text{if } u > \frac{v+1}{2}, \end{cases} \tag{2.34}$$

its associated density function $c_{M_t}(u, v)$ then being given by

$$\begin{aligned} c_{W_t, M_t}(u, v) &= \frac{\partial^2}{\partial u \partial v} C_{W_t, M_t}(u, v) \\ &= \frac{\left[2\Phi^{-1}\left(\frac{v+1}{2}\right) - \Phi^{-1}(u)\right]\phi\left(2\Phi^{-1}\left(\frac{v+1}{2}\right) - \Phi^{-1}(u)\right)}{\phi\left(\Phi^{-1}\left(\frac{v+1}{2}\right)\right)\phi\left(\Phi^{-1}(u)\right)} \end{aligned} \tag{2.35}$$

whenever $u \le \frac{v+1}{2}$, and zero otherwise.

Some further related results are available in the statistical literature. For example, representations of the joint density function of a $\mathcal{BM}$ process and its minimum and maximum, which are given for instance in Borodin and Salminen (2002), were shown to be convergent by Choi and Roh (2013). Upper and lower bounds for the distribution of the maximum of a two-parameter $\mathcal{BM}$ process were obtained by Cabaña and Wschebor (1982). Vardar-Acara *et al.* (2013) provided explicit expressions for the correlation between the supremum and the infimum of a $\mathcal{BM}$ with drift. Kou and Zhong (2016) studied the first-passage times of two-dimensional $\mathcal{BM}$ processes. Haugh (2004) explained how to generate correlated Brownian motions and pointed out some applications involving security pricing and portfolio evaluation.

Bibbona *et al.* (2016) obtained a copula for the Ornstein-Uhlenbeck process. Cherubini

and Romagnoli (2010) expressed (2.30) in the following form:

$$C(u, v) = \int_0^u \Phi\left(\frac{\Phi^{-1}(v) - \rho(s, t)\Phi^{-1}(x)}{\sqrt{1 - \rho(s, t)^2}}\right) dx, \tag{2.36}$$

where $\rho(s, t) = \sqrt{\frac{s}{t}}$.

Nadarajah *et al.* (2017) pointed out that it follows from equation (2.30) that independence corresponds to $t - s \to \infty$, while full dependence corresponds to $t - s \to 0$. Cherubini and Romagnoli (2010) emphasized the importance of Brownian copulas by pointing out that several non-Gaussian processes can be converted into a Brownian motion by means of the time change technique which is discussed in Dambis (1965), Dubins and Schwarz (1965) and Monroe (1978), among others. Jaworski and Krzywda (2013) and Bosc (2012) obtained the copulas corresponding to certain correlated Brownian motions. Lagerås (2010) provided an explicit representation of the copulas associated with Brownian motion processes that are reflected at 0 and 1. Chen *et al.* (2019) explain that correlated Brownian motions and their associated copulas can be utilized in the case of correlated assets occurring in risk management, pairs trading and derivative's pricing. Deschatre (2016a,b) proposed to make use of asymmetric copulas generated from a Brownian motion and its reflection to model and control the distribution of their difference with applications to the energy market and the pricing of spread options.

2.6.2   A stock's closing price and its running maximum

The stock selected is AC.TO (Air Canada). The data consists of the daily closing prices of AC.TO during the entire year 2019. To relate the data to a standard Wiener process, the first data point should be 0, the differences between successive observations should ideally often change signs and have a variance of one and there should be one unit of time between successive observations. Hence the following transformation is applied. Let $U_1, U_2, \ldots, U_n$ denote the closing prices and $V_1, V_2, \ldots, V_{n-1}$ be the differences between successive closing prices, that is, $V_i = U_{i+1} - U_i$; denoting by σ_D the standard deviation of the differences $V_1, V_2, \ldots, V_{n-1}$,

the following transformation is applied

$$W_i = \frac{U_i - U_1}{\sigma_D}$$

and the resulting data is denoted by $W_1, W_2, \ldots, W_n$. Let Z_i be the ith running maximum, that is, $Z_i = \text{Max}\{W_1, W_2, \ldots, W_i\}$, $i = 1, 2, \ldots, n$. Then, the resulting bivariate data, (W_i, Z_i), $i = 1, 2, \ldots, n$, has the features of a Brownian motion process and its running maximum. The distribution of the transformed data and its running maximum along with the joint distribution are depicted in Figs 2.18-2.23.

The inverse cdf's which were obtained from the marginal kernel density estimates are plotted in Figs 2.24-2.25. Finally, the resulting copula density f_c as defined in (2.9), is shown in Fig 2.26.
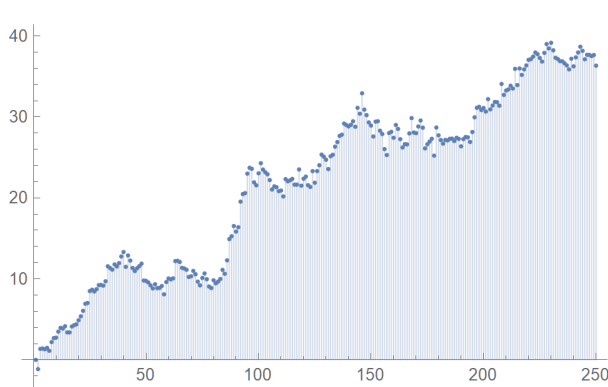
Figure 2.18: Plot of the first component
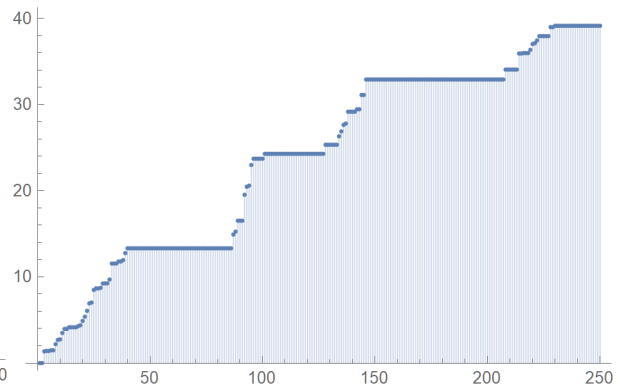
Figure 2.19: Plot of the second component (running maximum)
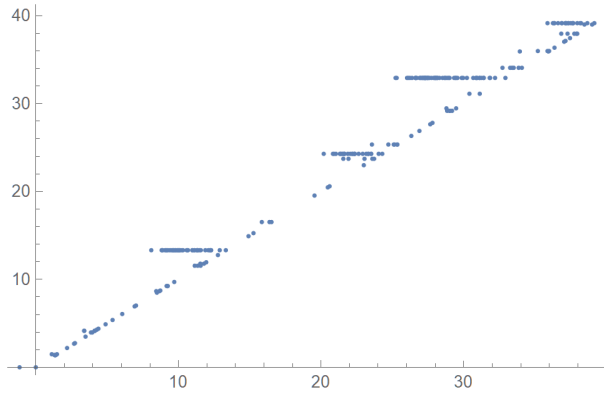
Figure 2.20: The scatter plot of the transformed data
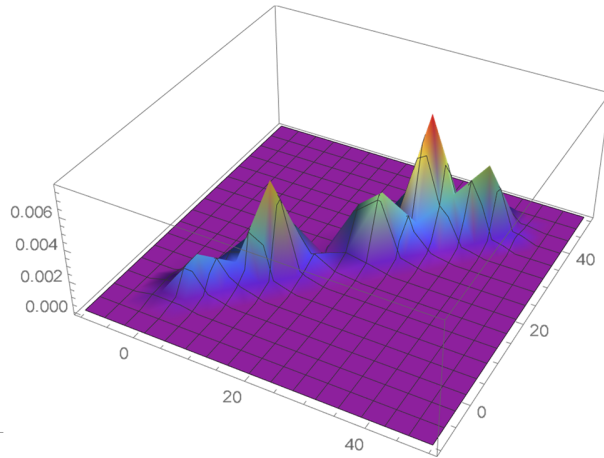
Figure 2.21: kde of the transformed data
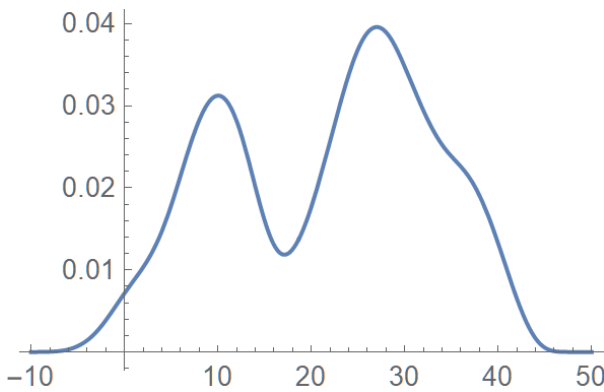
Figure 2.22: The marginal kde of the first variable
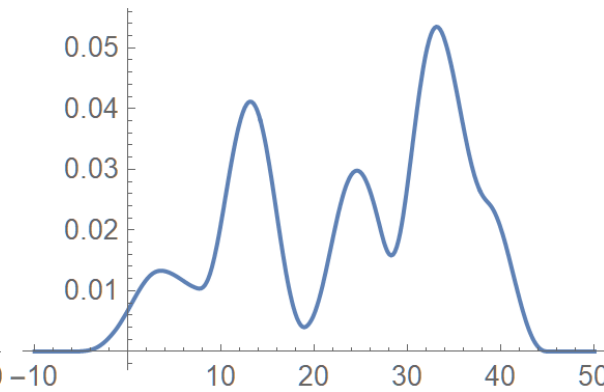
Figure 2.23: The marginal kde of the second variable
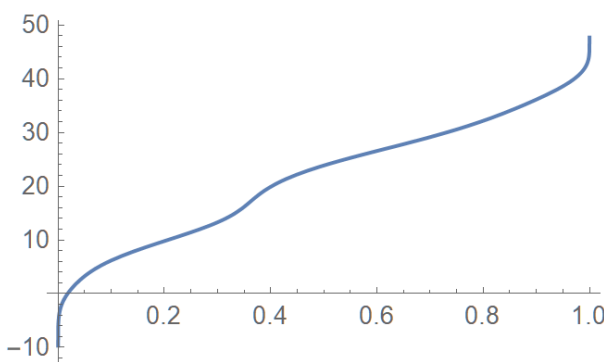
Figure 2.24: The estimate of inverse cdf of the first variable
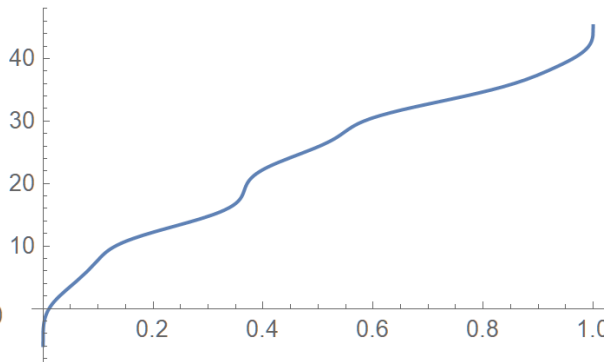
Figure 2.25: The estimate of inverse cdf of the second variable
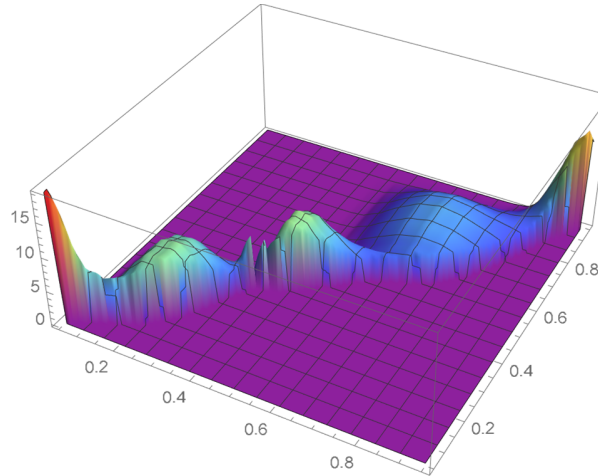
Figure 2.26: The resulting copula density $\hat{c}$

2.7 Bernstein's approximation to copulas

2.7.1 Introduction

This section initially presents with appropriate references relevant background information on Bernstein's empirical copulas and copula densities. A novel methodology for obtaining a copula density by differentiating an approximation of Deheuvels' empirical copula is introduced, and a technique for determining a suitable degree for the polynomials defining a Bernstein empirical copula density function will be proposed.

First we define Bernstein polynomials, explain how it is utilized to approximate continuous functions and present some of its properties. A Bernstein polynomial of degree n is defined as follows:

$$B_n(x) = \sum_{v=0}^{n} \beta_v \, b_{v,n}(x), \tag{2.37}$$

where the β_v's are called the Bernstein coefficients and $b_{v,n}(x) = \binom{n}{v} x^v (1 - x)^{n-v}$ is the Bernstein basis polynomial of degree n, which is also a binomial probability mass function when $x \in [0, 1]$. Let f be a continuous function on the interval $[0, 1]$. The Bernstein basis polynomials have the following properties:

- $b_{v,n}(x) = 0,$ if $v < 0$ or $v > n$.

- $b_{v,n}(x) \geq 0$ for $x \in [0, 1]$.

- $b_{v,n}(1 - x) = b_{n-v,n}(x)$.

- The derivative can be written as a combination of two polynomials of lower degree :

$$b'_{v,n}(x) = n(b_{v-1,n-1}(x) - b_{v,n-1}(x)).$$

Consider the Bernstein polynomial

$$B_n(f)(x) = \sum_{v=0}^{n} f\left(\frac{v}{n}\right) b_{v,n}(x). \tag{2.38}$$

It can be shown that $\lim_{n\to\infty} B_n(f) = f$ uniformly on the interval $[0, 1]$. This approximation approach can be generalized to d dimensions: Letting $g(x_1, \ldots, x_d)$ be a continuous function on $[0, 1]^d$, $g(x_1, \ldots, x_d)$ can be approximated by

$$\sum_{v_1=0}^{n_1} \cdots \sum_{v_d=0}^{n_d} g\left(\frac{v_1}{n_1}, \ldots, \frac{v_d}{n_d}\right) \prod_{j=1}^{d} b_{v_j,n_j}(x_j). \tag{2.39}$$

Bernstein's empirical copula was first proposed and investigated by Sancetta & Satchell (2004) for identically and independently distributed (i.i.d.) data. Bernstein's approximation of order k, $k > 0$, of a copula function C, the so-called Bernstein copula function, can be defined as follows:

$$B_k(\mathbf{u}) = \sum_{v_1=0}^{k} \cdots \sum_{v_d=0}^{k} C\left(\frac{v_1}{k}, \ldots, \frac{v_d}{k}\right) \prod_{j=1}^{d} P_{v_j,k}(u_j), \quad \text{for } \mathbf{u} = (u_1, \ldots, u_d) \in [0, 1]^d, \tag{2.40}$$

where k plays the role of bandwidth parameter and $P_{v_j,k}(u_j)$ is the binomial probability mass function:

$$P_{v_j,k}(u_j) = \binom{k}{v_j} u_j^{v_j} (1 - u_j)^{k-v_j}. \tag{2.41}$$

It has been shown that

$$\lim_{k\to\infty} B_k(\mathbf{u}) = C(\mathbf{u}), \quad \text{uniformly in } \mathbf{u} \in [0, 1]^d.$$

In addition, under the conditions specified in their Theorem 1, Sancetta & Satchell (2004) established that B_k in (2.40) is itself a copula. Thus, to estimate the copula function C, they proposed the following estimator referred to as Bernstein's empirical copula (cdf):

$$C_{k,n}(\mathbf{u}) = \sum_{v_1=0}^{k} \cdots \sum_{v_d=0}^{k} C_n\Big(\frac{v_1}{k}, \ldots, \frac{v_d}{k}\Big) \prod_{j=1}^{d} P_{v_j,k}(u_j), \quad \text{for } \mathbf{u} = (u_1, \ldots, u_d) \in [0, 1]^d, \qquad (2.42)$$

where C_n is the standard empirical copula estimator that was introduced by Deheuvels (1979) and is defined as

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \mathbb{I}\{F_{j;n}(X_{i,j}) \leq u_j\}, \quad \text{for } \mathbf{u} = (u_1, \ldots, u_d) \in [0, 1]^d, \qquad (2.43)$$

where $\mathbb{I}$ is the indicator function and $F_{j;n}$ is the empirical cumulative distribution function of the component X_j with n being the sample size. Recently Jansen *et al.* (2012) have shown that Bernstein's empirical copula estimator outperforms the classical empirical copula estimator.

If it exists, the copula density, denoted by c, which corresponds to the copula function C is given by

$$c(\mathbf{u}) = \partial^d C(\mathbf{u})/\partial u_1 \cdots \partial u_d. \qquad (2.44)$$

Since the Bernstein copula function in (2.40) is absolutely continuous, it follows that Bernstein's copula density is defined as:

$$c_k(\mathbf{u}) = \sum_{v_1=0}^{k} \cdots \sum_{v_d=0}^{k} C\Big(\frac{v_1}{k}, \ldots, \frac{v_d}{k}\Big) \prod_{j=1}^{d} P'_{v_j,k}(u_j), \qquad (2.45)$$

where $P'_{v_j,k}(u_j)$ is the derivative of $P_{v_j,k}$ with respect to u_j. Accordingly, Sancetta & Satchell

(2004) proposed and investigated the following estimator of Bernstein's copula density:

$$\hat{c}_n(\mathbf{u}) = \sum_{v_1=0}^{k} \cdots \sum_{v_d=0}^{k} C_n\left(\frac{v_1}{k}, \ldots, \frac{v_d}{k}\right) \prod_{j=1}^{d} P'_{v_j,k}(u_j). \tag{2.46}$$

Later, Bouezmarni *et al.* (2010) have used Bernstein's empirical copula density to estimate the copula density in the presence of dependent data. Recently, Jansen *et al.* (2014) have reinvestigated this estimator by establishing its asymptotic normality under i.i.d. data. Bouezmarni *et al.* (2010) also proposed another form of Bernstein's copula density estimator given by

$$\hat{c}_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} K_k(\mathbf{u}, S_i), \quad \text{for } \mathbf{u} \in [0, 1]^d, \tag{2.47}$$

where

$$K_k(\mathbf{u}, S_i) = k^d \sum_{i=1}^{k-1} \cdots \sum_{i=1}^{k-1} A_{S_i,v} \prod_{j=1}^{d} P_{v_j,k-1}(u_j), \tag{2.48}$$

with $S_i = (F_{i;n}(X_{i1}), \ldots, F_{d;n}(X_{id}))$, where $F_{j;n}(\cdot)$, for $j = 1, \ldots, d$, is the empirical distribution of the random component X_j, $A_{S_i,v} = \mathbf{1}_{\{S_i \in B_v\}}$, for $B_v = [\frac{v_1}{k}, \frac{v_1+1}{k}] \times \cdots \times [\frac{v_d}{k}, \frac{v_d+1}{k}]$.

For a given data set, we will first determine Deheuvels' empirical copula as given in (2.43) and approximate it by means of a bivariate least-squares polynomial. Then, on differentiating this polynomial with respect to both variables, an estimate of the copula density, say, $Dc(x, y)$, is obtained. Next, Bernstein's empirical copula density, say, $Bc_k(x, y)$ as defined in (2.40) is then evaluated and plotted for various values of the degree k. We will finally select the value of k that will minimize the integrated squared difference between $Bc_k(x, y)$ and $Dc(x, y)$ or such that it is observed that $Bc_k(x, y)$ and $Dc(x.y)$ share similarly features. Alternatively, we can select k such that there are no significant differences between the density estimate obtained with k and those determined with higher values of k.

For example, using the data set considered in Subsection 2.6.2, we can generate Bernstein copula densities of different degrees and compare them with an estimate of the copula density $D_c(x, y)$. The copula densities are plotted in Figures 2.27-2.31.

It is seen from those figures that when $k = 75$, Bernstein's empirical copula density and the preliminary density estimate $Dc(x, y)$ exhibit similar features. The copula densities obtained with $k = 25$ and 50 are clearly not sufficiently accurate. Since the plot resulting from $k = 90$ does not differ significantly from that obtained for $k = 75$, we deem that $k = 75$ is adequate for modeling the copula density. It should also be noted that this density estimate is similar to that shown in Figure 2.26 which was obtained independently using a distinct approach.
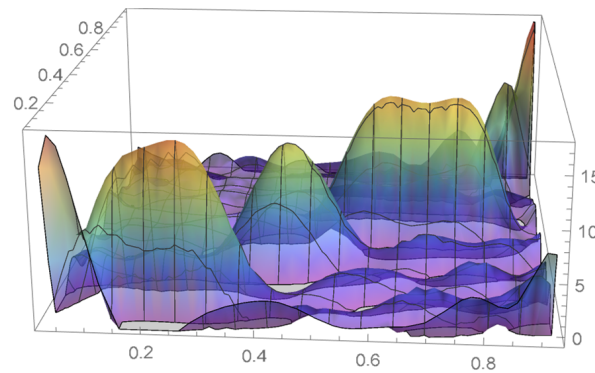
Figure 2.27: The preliminary copula density Dc
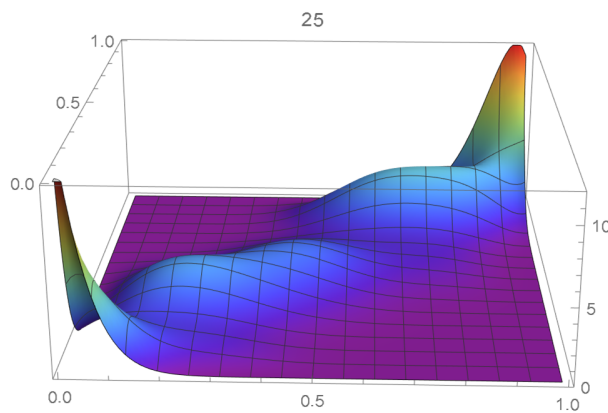
Figure 2.28: Bernstein copula density, degree 25
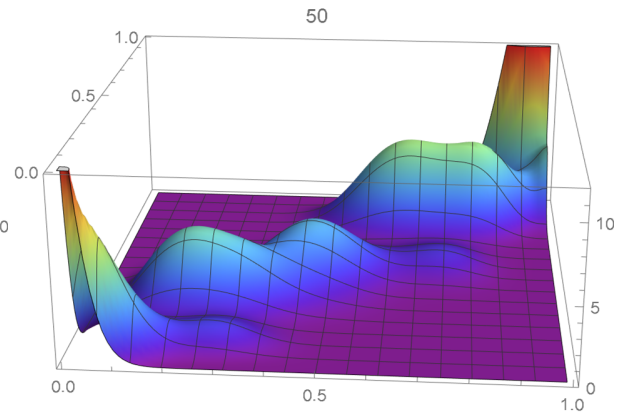
Figure 2.29: Bernstein copula density, degree 50
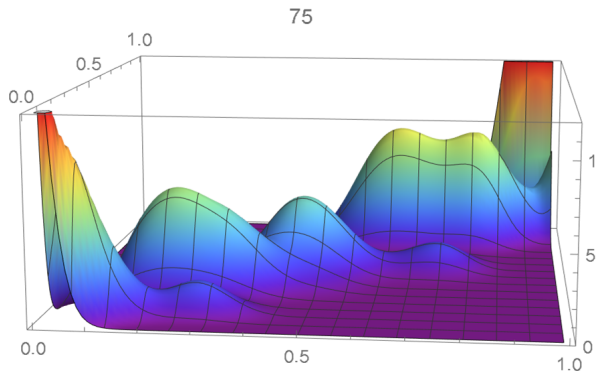
Figure 2.30: Bernstein copula density, degree 75
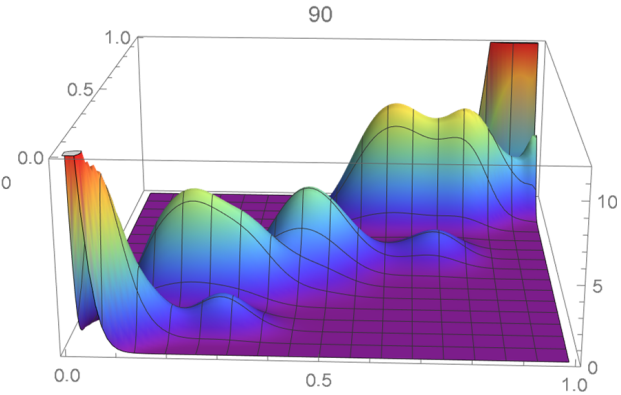
Figure 2.31: Bernstein copula density, degree 90

2.7.2   The two stocks' closing prices revisited

Using the same data as in Section 2.4, we determined a Bernstein estimate $\hat{c}_n$ of the copula density which is already a bivariate polynomial. The resulting density is plotted in Fig. 2.33. It shares several features with the preliminary copula density estimate shown in Fig. 2.32, which was obtained by differentiating a least-square approximation of Deheuvels' empirical copula. As expected, the copula density estimate appearing in Figure 2.33 is similar to that shown in Figure 2.13 which was determined by making use of a completely different approach.
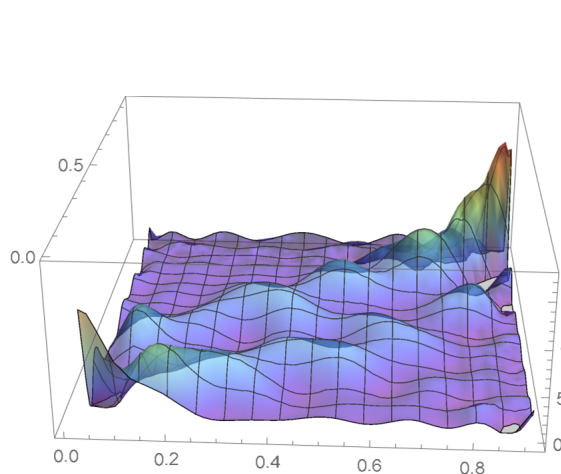
Figure 2.32: The preliminary copula density *Dc*
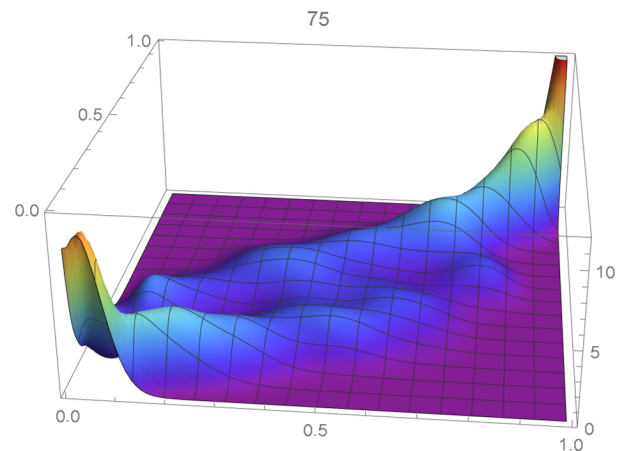
Figure 2.33: The copula density $\hat{c}_n$

Bibliography

[1] Adès, M., Dufour, M., Provost, S., Vachon, M. and Zang, Y. (2022). A class of copulas associated with Brownian motion processes and their maxima. *Journal of Applied Mathematics and Computation*, 6(1), 96–120.

[2] Bibbona, E., Sacerdote, L. and Torre, E. (2016). A copula-based method to build diffusion models with prescribed marginal and serial dependence. *Methodology and Computing in Applied Probability*, 18(3), 765–783.

[3] Borodin, A. N. and Salminen, P. (2002). *Handbook of Brownian Motion–Facts and Formulae*, Second Edition. Birkhäuser, Basel.

[4] Bosc, D. (2012). Three essays on modeling the dependence between financial assets. Thèse, Ecole Polytechnique X, https://pastel.archives-ouvertes.fr/pastel-00721674

[5] Bouezmarni, T., Rombouts, J. and Taamouti, A. (2010). Asymptotic properties of the Bernstein density copula estimator for α-mixing data. *Journal of Multivariate Analysis*, 101, 1–10.

[6] Cabaña, E. M. and Wschebor, M. (1982). The two-parameter Brownian bridge: Kolmogorov inequalities and upper and lower bounds for the distribution of the maximum. *The Annals of Probability*, 10, 289–302.

[7] Chen, T., Cheng, X. and Yang, J. (2019). Common decomposition of correlated Brownian motions and its financial applications. *Quantitative Finance - Financial Mathematics*. arXiv:1907.03295 [q-fin.MF], 47 pages.

[8] Cherubini, U., Fabio, G., Mulinacci, S. and Romagno, S. (2012). *Dynamic Copula Methods in Finance*. John Wiley & Sons, New York.

[9] Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley & Sons, New York.

[10] Cherubini, U. and Romagnoli, S. (2010). The dependence structure of running maxima and minima: results and option pricing applications.*Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 20(1), 35–58.

[11] Choi, B.-S. and Roh, J.-H. (2013). On the trivariate joint distribution of Brownian motion and its maximum and minimum. *Statistics & Probability Letters*, 83, 1046–1053.

[12] Dambis, K. E. (1965). On the decomposition of continuous submartingales. *Theory of Probability & Its Applications*, 10(3), 401–410.

[13] Darsow, W. F., Nguyen, B. and Olsen, E. T. (1992). Copulas and Markov Processes. *Illinois journal of mathematics*, 36(4), 600–642.

[14] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'independance. *Bullettin de l'Académie Royale de Belgique, Classe des Sciences*, 65, 274–292.

[15] Denuit, M., Daehe, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons, New York.

[16] Deschatre, T. (2016a). On the control of the difference between two Brownian motions: a dynamic copula approach. *Dependence Modeling*, 4, 141–160.

[17] Deschatre, T. (2016b). On the control of the difference between two Brownian motions: an application to energy market modeling. *Dependence Modeling*, 4, 161–183.

[18] Dubins, L. E. and Schwarz, G. (1965). On continuous martingales.*Proceedings of the National Academy of Sciences of the United States of America*, 53(5), 913.

[19] Etheridge, A. (2002). *A Course in Financial Calculus.* Cambridge University Press.

[20] Harrison, M. (1990). *Brownian Motion and Stochastic Flow Systems.* Robert E. Krieger, Malabar, Florida.

[21] Haugh, M. (2004). The Monte Carlo framework, examples from finance and generating correlated random variables. *Monte Carlo Simulation: IEOR E4703*, 1–10.

[22] Jansen, P., Swanepoel, J. and Veraverbeke, N. (2012). Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planing and Inference*, 142, 1189–1197.

[23] Jansen, P., Swanpoel, J. and Veraverbek, N. (2014). A note on the asymptotic bevahior of the Bernstein estimator of the copula density. *Journal of Multivariate Analysis*, 124, 480–487.

[24] Jaworski, P. and Krzywda, M. (2013). Coupling of Wiener processes by using copulas. *Statistics & Probability Letters*, 83(9), 2027–2033.

[25] Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall/CRC, Boca Raton, Florida.

[26] Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, Second Edition. Academic Press, New York.

[27] Kou, S. and Zhong, H. (2016). First passage times of two-dimensional Brownian motion. *Advances in Applied Probability*, 48, 1045–1060.

[28] Lagerås, A. N. (2010). Copulas for Markovian dependence. *Bernoulli*, 16, 331–342.

[29] Lee, H. (2003). Pricing equity-indexed annuities with path-dependent options. *Insurance: Mathematics and Economics*, 33, 677–690.

[30] Monroe, I. (1978). Processes that can be embedded in Brownian motion. *The Annals of Probability*, 6(1), 42–56.

[31] Nadarajah, S., Afuecheta, E., and Chan, S. (2017). A compendium of copulas. *Statistica*, 77(4), 279–328.

[32] Nelsen, R. B. (2006). *An Introduction to Copulas*, Second Edition. Springer, New York.

[33] Provost, S.B. (2005). Moment-based density approximants, *The Mathematica Journal*, 9, 727–756.

[34] Revuz, D. and Yor, M. (2005). *Continuous Martingales and Brownian Motion*, Third Edition. Springer-Verlag, New York.

[35] Rogers, L. C. G. and Williams, D. (2000a). *Diffusions, Markov Processes and Martingales.* Vol. 1, *Foundations*, Second Edition. Cambridge University Press.

[36] Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20, 535–562.

[37] Schmitz, V. (2003). *Copulas and Stochastic Processes* (Doctoral dissertation), Bibliothek der RWTH Aachen).

[38] Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

[39] Vardar-Acara, C., Zirbel, C. L. and Székelyc, G. (2013). On the correlation of the supremum and the infimum and of maximum gain and maximum loss of Brownian motion with drift. *Journal of Computational and Applied Mathematics*, 248, 61–75.

Chapter 3

Nonparametric Copula Density Estimation Methodologies

3.1 Introduction

This chapter aims at providing objective criteria for the selection of optimal copula density estimates in conjunction with four main nonparametric approaches. It is assumed that the reader is already familiar with the notation and results introduced in Section 2.1.

The *pseudo-observations* $(\hat{u}_i, \hat{v}_i)$, $i = 1, \ldots, n$, are then defined in terms of the empirical marginal cdf's denoted by $\hat{F}(\cdot)$ and $\hat{G}(\cdot)$, that is,

$$(\hat{u}_i, \hat{v}_i) = (\hat{F}(x_i), \hat{G}(y_i)), \quad i = 1, \ldots, n, \tag{3.1}$$

where the empirical cdf's (ecdf's) are given by

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(x_j \le x) \text{ and } \hat{G}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(y_j \le y), \tag{3.2}$$

with $\mathcal{I}(\mathcal{A})$ denoting the indicator function which is equal to 1 if condition $\mathcal{A}$ is verified and 0,

otherwise. Equivalently, one has

$$(\hat{u}_i, \hat{v}_i) = (r_i/n, \rho_i/n), \tag{3.3}$$

where r_i is the rank of x_i among $\{x_1, \ldots, x_n\}$ and ρ_i, the rank of y_i among $\{y_1, \ldots, y_n\}$. Actually, $(\hat{u}_1, \hat{v}_1), \ldots, (\hat{u}_n, \hat{v}_n)$ are *copula frequency points* with $1/n$ being the common frequency associated with these n points which form the support of the empirical copula probability mass function (pmf).

The the *probability mass function* or pmf of an empirical copula, which is either equal to 0 or $1/n$ can be expressed as follows:

$$
\begin{aligned}
\hat{c}(u, v) &= \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(\hat{F}(x_i) = u)\, \mathcal{I}(\hat{G}(y_i) = v) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(r_i/n = u)\, \mathcal{I}(\rho_i/n = v),
\end{aligned}
\tag{3.4}
$$

and the corresponding empirical copula (distribution function), which is a consistent estimate of $C(u, v)$, is then given by

$$
\begin{aligned}
\hat{C}(u, v) &= \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(\hat{F}(x_i) \le u)\, \mathcal{I}(\hat{G}(y_i) \le v) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(r_i/n \le u)\, \mathcal{I}(s_i/n \le v).
\end{aligned}
\tag{3.5}
$$

Additional results related to copulas are discussed in Cherubini *et al.* (2004, 2012), Denuit *et al.* (2005), Joe (1997), Nelsen (2006), and Sklar (1959), among others.

3.1.1   Repositioning of the pseudo-observations

Given a random sample of n bivariate observations, we advance that the favored placement of the pseudo-observations is at the center of the cells of an $n \times n$ grid of the unit square.

For illustrative purposes, consider the sample $\{(2, 4), (3, 2), (7, 2), (8, 3)\}$ of size $n = 4$, which is plotted in Figure 3.1. In this instance, the ranks of the first component observations are $(1, 2, 3, 4)$, whereas those of the second component observations are $(3, 4, 1, 2)$. Accordingly, the pseudo-observations (pso's), $(\hat{u}_i, \hat{v}_i)$, $i = 1, \ldots, 4$, are $\{(1/4, 3/4), (1/2, 1), (3/4, 1/4), (1, 1/2)\}$. When one makes use of pseudo-observations in the context of kernel density estimation, undesirable boundary effects are observed and the resulting copula density functions will be less concentrated near the origin than they would be when centered as explained below. Moreover, they will not integrate to one.

For a sample of size n, The corresponding centered pseudo-observations (cpso's) also referred to as centered copula points (ccp's), denoted by $(\hat{u}_i^*, \hat{v}_i^*)$, are obtained by subtracting $1/(2 \times n)$ from each coordinate of $(\hat{u}_i, \hat{v}_i)$, $i = 1, \ldots, n$. These pseudo-observations and centered pseudo-observations are respectively shown in Figures (3.2) and (3.4) for the example at hand.
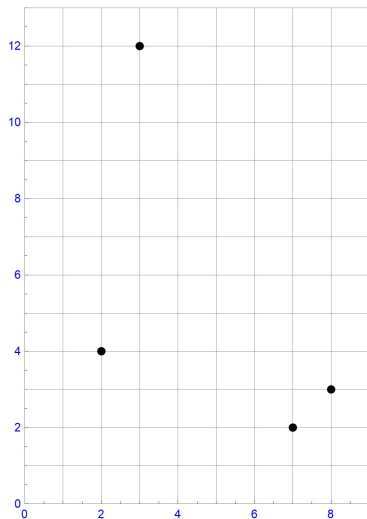
Figure 3.1: The four data points
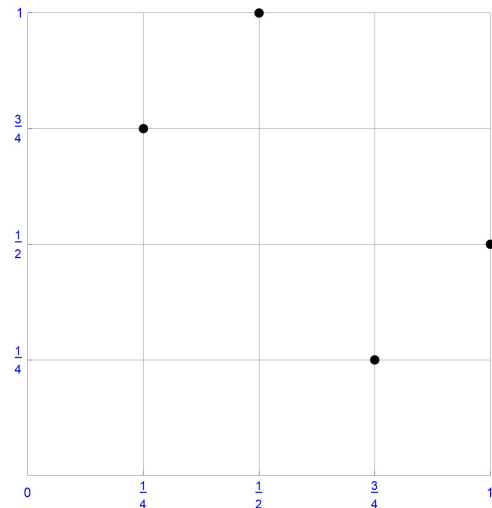
Figure 3.2: The pseudo-observations

An approach that is suggested in the literature for mitigating the edge effects consists of multiplying the pseudo-observations by $n/(n + 1)$. In this case, the resulting points are $(1/5, 3/5)$, $(2/5, 4/5)$, $(3/5, 1/5)$ and $(4/5, 2/5)$, which are plotted in Figure 3.3. As can be

seen from this graph, these points occupy irregular positions within the corresponding grid cells and their uneven distribution will result in a copula density that will be less concentrated near the ends of the unit intervals in each direction. Discrete uniform marginal distributions can be achieved by making use of the centered pseudo-observations. The marginal probabilities will then be $1/n$ at the points $(2i - 1)/(2n)$, $i = 1, \ldots, n$.

Wang and Fang (1990) and Pérez *et al.* (2005, p.100) discuss the following measure of representativeness of a sample $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ with respect to the distribution function $F(x)$, which is referred to as F–discrepancy:

$$D_F(\mathcal{S}) = \sup_{x \in \Re} |F_n(x) - F(x)|$$

where $F_n(x)$ denotes the empirical distribution function of $\mathcal{S}$. We observe that $D_F(\mathcal{S})$ is nothing but the Kolmogorov-Smirnov statistic for assessing the goodness-of-fit with respect to F(x). In one dimension,

$$\{F^{-1}(\tfrac{2i-1}{2n}), \ i = 1, 2, \ldots, n\}$$

is the set of points having the lowest F–discrepancy. In that sense, this set of n points form the most representative sample with respect to the distribution specified by $F(x)$. Thus, when $F(\cdot)$ is the distribution function of a uniform distribution on the unit interval, the sample having the lowest F–discrepancy is $\{\frac{2-1}{2n}, \ldots, \frac{2n-1}{2n}\}$—precisely the support of the marginals of the distribution of empirical copula pmf when the centered pseudo-observations are utilized.

As well, note that by constructing cuboidal kernels over the $1/4 \times 1/4$ grid cells containing the cpso's, one actually obtains uniform marginals on the unit intervals as can be seen from Figure 3.7, the corresponding copula being plotted in Figure 3.8. This clearly would not the case with any other repositioning of the pseudo-observations. Accordingly, we shall make use of centered pseudo-observations when determining copula kde's, which also have the advantage of not lying on the boundary of the support of the copula. We observe that since there are n distinct ranks with respect to each coordinate, the n pairs of points will fill exactly one grid

cell in each row and each column of grid cells. The empirical copulas as determined from the
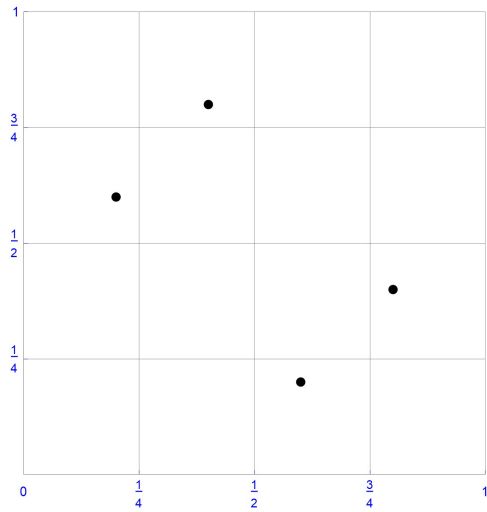
Figure 3.3: Pseudo-observations $\times \frac{n}{n+1}$
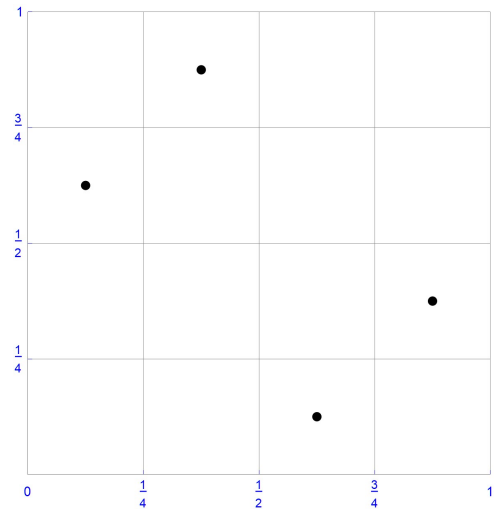
Figure 3.4: Centered pseudo-observations

pseudo-observations and their centered counterparts are respectively shown in Figures 3.5 and 3.6.
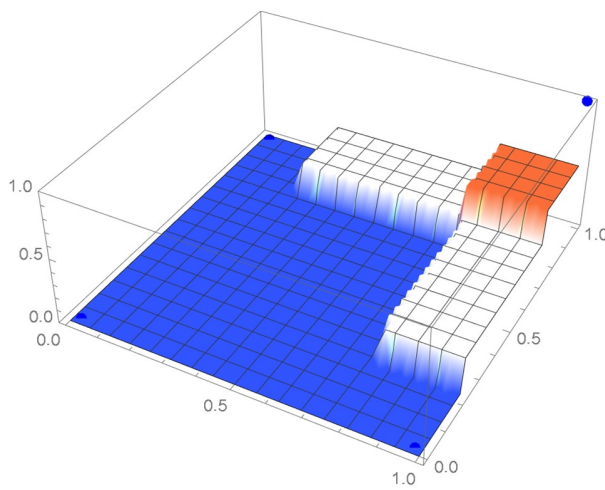
Figure 3.5: Empirical copula as evaluated from the pso's
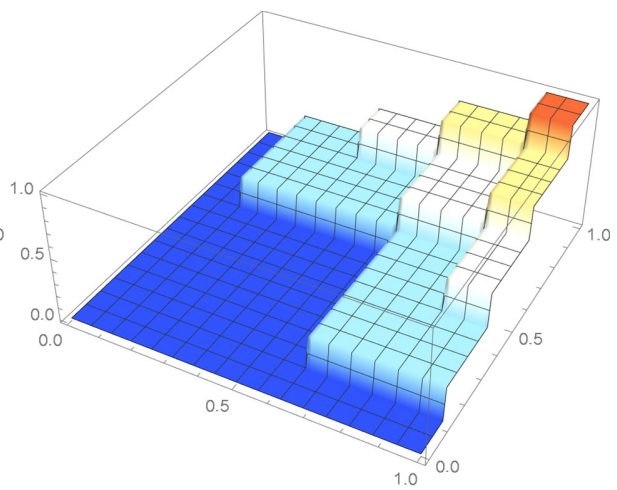
Figure 3.6: Empirical copula as evaluated from the cpso's

The centered pseudo-observations form a discrete distribution, their respective pmf being $1/n = 1/4$. As can be seen from Figure 3.7, continuous discrete uniform marginal pdf's on

[0,1] will be achieved by making use of cuboidal kernels density estimates whose bases are the grid cells at the center of which lie the centered pseudo-observations and whose height is $n = 4$. As well, centering the pseudo-observation lessens the boundary issues in the context of kernel density estimation. A *bona fide* copula which is plotted in 3.8, can be secured by integrating this copula density. The kde obtained with cuboidal kernels and the copula determined by integrating the kde are shown in Figures 3.7 and 3.8.
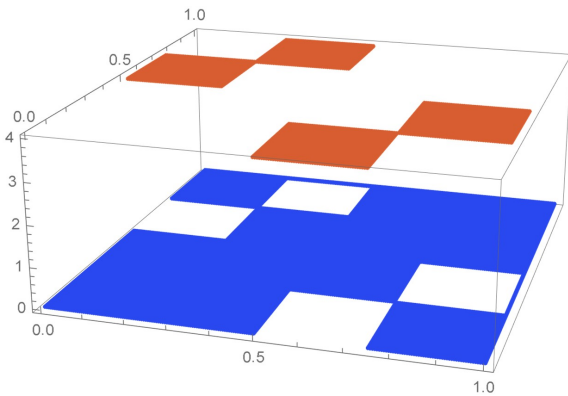
Figure 3.7: Cuboidal kernel kde
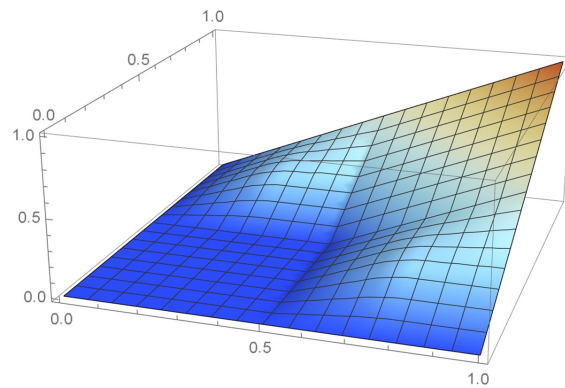
Figure 3.8: Copula obtained from the cuboidal kernel kde

3.1.2   Additional considerations and structure of the chapter

The reader may refer to Section 2.3 for details on implementing a moment-based bivariate polynomial approximation of continuous functions such as density functions. When the distributional features of a preliminary estimate of the copula density warrant it, one may select a nontrivial base density function. However, when the empirical copula appears to exhibit an irregular pattern, as is often the case, a uniform density function whose support area slightly exceeds that of the copula, ought to be taken as base density. Accordingly, unless specified otherwise, we will utilize such a base density for the purpose of approximating or smoothing

copula densities.

We will make use of integrated squared differences (ISD) or average squared differences (ASD) to compare the copula density estimates to reference copula density functions over the unit square. The integrated squared difference between two functions is equal the integral of the square of their differences over the domain of interest, the average squared difference being the sum of the square of the differences between the two functions at s selected points divided by s. When the density estimates fluctuate highly near the boundary, a subset of the unit square, namely, $[0.1, 0.9] \times [0.1, 0.9]$ referred to as the reduced unit square, will be utilized for comparison purposes. Moreover, in order to ensure that the resulting density function be *bona fide* within the unit square, the final approximation will be taken to be $c(\hat{f}(y_1, y_2) + |\hat{f}(y_1, y_2)|)/2$ where c denotes the normalizing constant.

The remainder of this chapter is organized as follows. Section 3.2 proposes four approaches for securing copula density estimates and specify criteria for determining those that are nearly optimal. It is explained in Section 3.3 that such density estimates can be utilized to obtain joint density functions. The proposed copula density estimation techniques are applied to a sample generated from a bivariate t distribution in Section 3.4. Some concluding remarks are offered in the last section.

3.2 Methodologies for estimating copula densities

3.2.1 Differentiated least-squares copula estimates as initial density approximations

Let $(x_1, y_1), \ldots, (x_n, y_n)$ denote the data set at hand and $\hat{C}(u, v)$ be the associated empirical copula as specified in (3.5). A least-squares approximating polynomial of degree $t + 1$ in each variable which is denoted by $P_{t+1}^{\mathcal{LS}}(u, v)$, is fitted to the n^2 points $(j/n, k/n, \hat{C}(j/n, k/n))$ where $j, k = 1, 2, \ldots, n$. The resulting polynomial is then differentiated with respect to u and v to obtain a copula density estimate denoted by $\hat{c}_t^{\mathcal{LS}}(u, v)$ whose domain is delimited by the unit

square. For a derivation of bivariate least-squares regression polynomials, the reader is referred to Fox (2015, Section 5.2.1).

By comparing the graphs of density estimates $\hat{c}_t^{\mathcal{LS}}(u, v)$ for $t = 5, 10, 15, \ldots$, one will notice that at a certain point, say t_o, some of the successive plots will fluctuate more. For instance, one could then select $\hat{c}_{t_o+r}^{\mathcal{S}}(u, v)$, $r = 0, 5$ or 10 as an initial copula density approximation and utilize this bivariate polynomial as a yardstick to choose a suitable density estimate as determined by making use of a different approach. The optimal degree of $\hat{c}_t^{\mathcal{LS}}(u, v)$ could also be obtained more formally by evaluating the integrated squared difference between density estimates that are five degrees apart in the truncated domain $[0.9, 0.9]^2$ for a range of degrees within which the density functions remain more or less stable.

An illustrative example

The set of observations to which all the proposed copula density estimation techniques will be applied is the Old Faithful geyser data which consists of 272 bivariate observations on two variables, the first component being the the duration of an eruption in seconds and the second one, the waiting time to the next eruption in minutes.

A scatter plot and a kernel density estimate (kde) of the data are shown in Figures 3.9 and 3.10. The linearized copula C_L resulting from applying linear interpolation to the points $\hat{C}(j/n, k/n)$, for $j, k = 1, \ldots, 272$, is plotted in Figure 3.11, while Figure 3.12 displays the probability mass function (pmf) of the empirical copula which is equal to $1/272$ at the copula frequency points $(\hat{u}_i, \hat{v}_i)$, $i = 1, \ldots, n$, as defined in (3.1) and (3.3), and zero, elsewhere.

Given the 272 copula frequency points, a bivariate least-squares approximating polynomial of degree $t + 1$ in each variable is determined, and a copula density estimate is then obtained by differentiation. The resulting copula density estimates are plotted in Figures 3.13-3.20 for $t = 5, 10, \ldots, 40$, respectively.
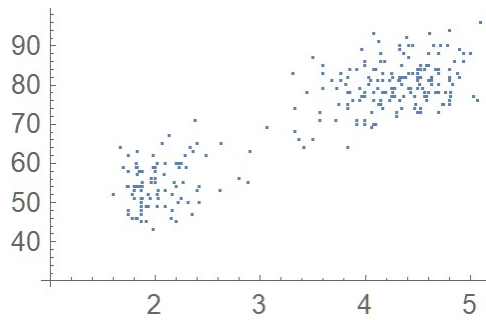
Figure 3.9: Scatter plot of the data
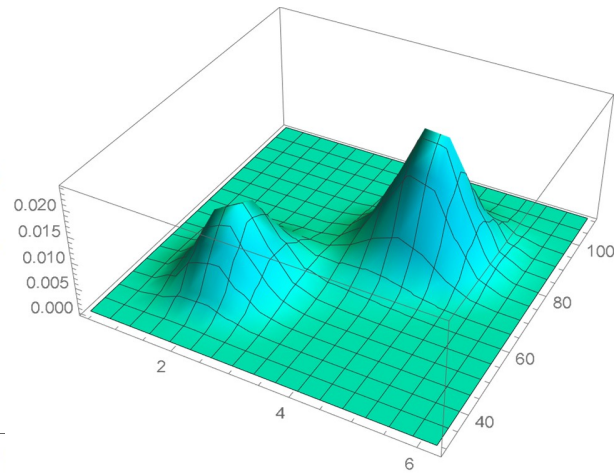
Figure 3.10: A bivariate kde
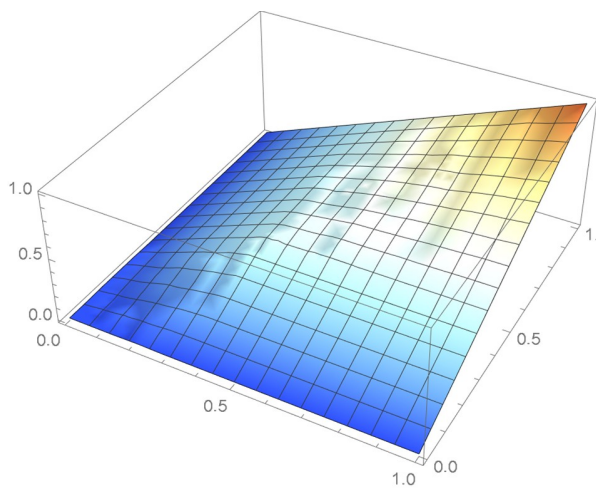
Figure 3.11: The linearized empirical copula
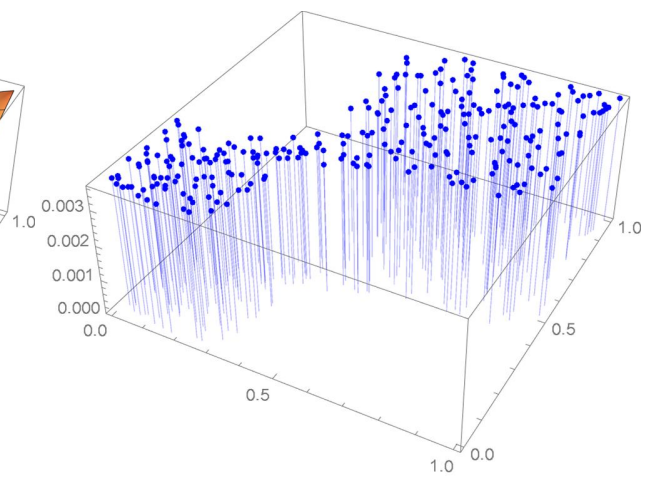
Figure 3.12: The empirical copula pmf
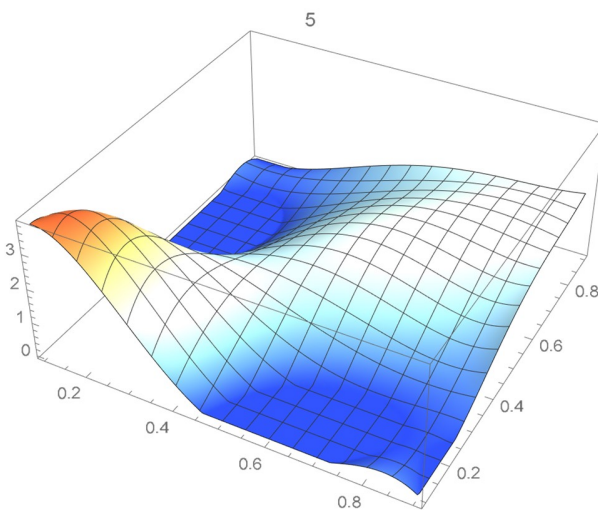
Figure 3.13: Estimated copula pdf, $t = 5$
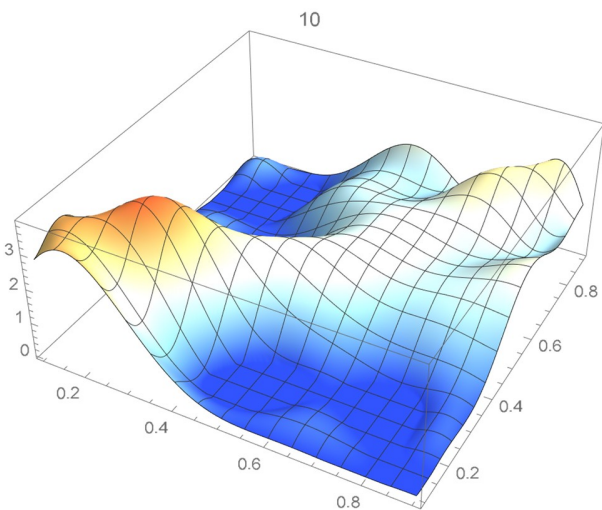
Figure 3.14: Estimated copula pdf, $t = 10$
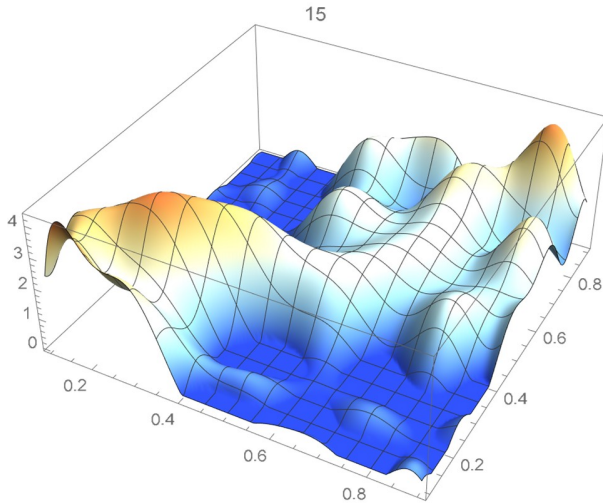
Figure 3.15: Estimated copula pdf, $t = 15$
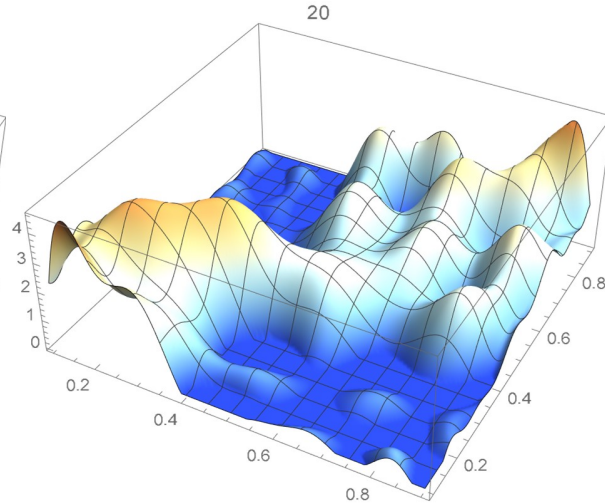
Figure 3.16: Estimated copula pdf, $t = 20$
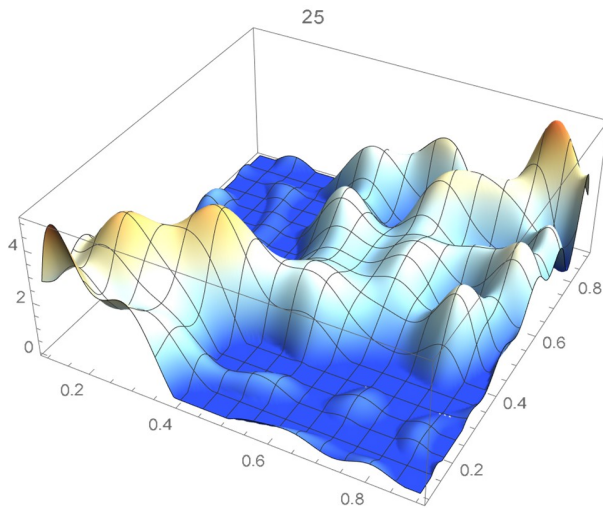
Figure 3.17: Estimated copula pdf, $t = 25$
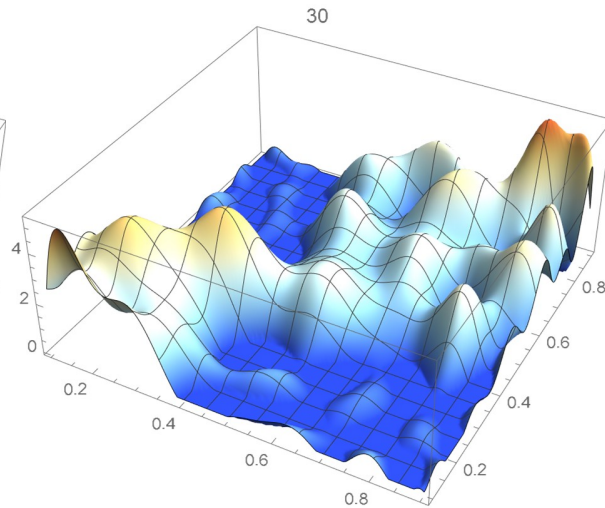
Figure 3.18: Estimated copula pdf, $t = 30$

By mere visual inspection, one can observe that the copula density estimates of degrees 20, 25 and 30 are quite similar. Appealing to the principle of parsimony, one could select the copula density of degree 20 in each variables as a yardstick for the distribution of the copula. This can be mathematically corroborated by noting that, within the truncated domain, the integrated squared differences between successive copula estimates indicate that there is little to be gained
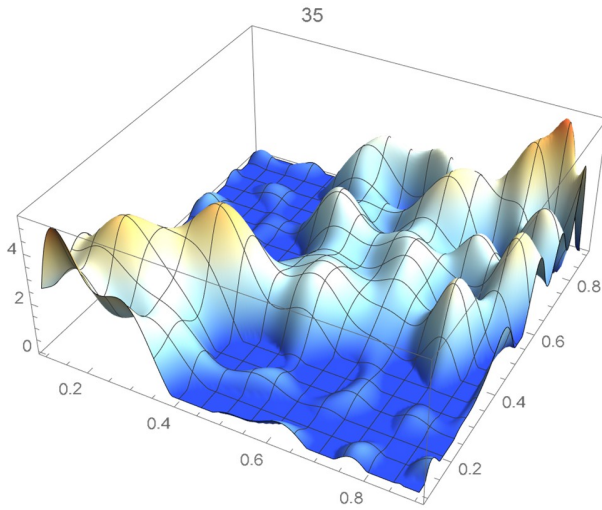
Figure 3.19: Estimated copula pdf, $t = 35$ Figure 3.20: Estimated copula pdf, $t = 40$
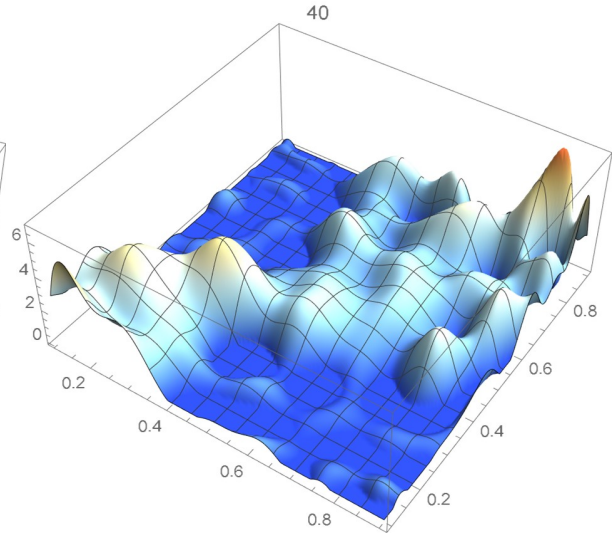
by choosing copulas of greater than 20, as can be seen from Table 3.1 and Fig 3.21. As well, the stability of the density estimates for such a wide array of degrees beyond 20 is indicative of their reliability.

t	ISD's between estimates of degrees $t + 5$ and t
5	1.75743×10^{-5}
10	3.55799×10^{-6}
15	1.06553×10^{-6}
20	2.65551×10^{-7}
25	3.15104×10^{-7}
30	1.01601×10^{-7}
35	1.11607×10^{-7}

Table 3.1: Successive ISD's

Least-squares polynomial approximations are underfitting when $t < 20$ as they do not adequately capture the distinctive features of the copula whereas approximations of degrees at least 20 in each variable turn out to be quite similar up to degree 40 in which case the value of the density function at the mode becomes higher, which is indicative of overfitting. Thus the least-squares copula density estimate of degree 20 is selected as initial reference density.

Although the selected least-squares density estimate can be regarded as preliminary, as it turns out, it already provides a reasonably accurate representation of the copula density.
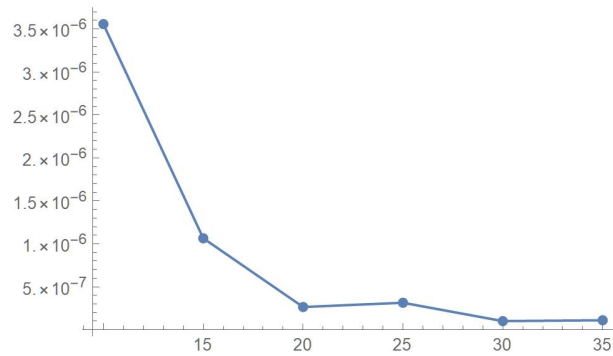
Figure 3.21: ISD's between successive least-squares copula estimates

3.2.2 Bernstein's polynomial approximation and degree selection

The reader is referred to Section 2.3 for an introduction to Bernstein's polynomial approximation as applied to copulas. We shall focus on the choice of an appropriate degree when applying this approximation. This will be achieved by comparing them to the selected least-squares copula density. In the case of the Old Faithful geyser eruption data set which is being considered throughout Section 3.2, it was determined that the least-squares polynomial approximation of degree 20 in each variable would be used as the yardstick copula.

The reference copula density and the Bernstein's copula densities of degree 25, 50, 75, 100, 125 and 150 are plotted in Figures 3.22-3.28. The integrated squared differences (ISD's) between Bernstein's copulas that are twenty-five degrees apart and the reference least-squares copula were calculated.
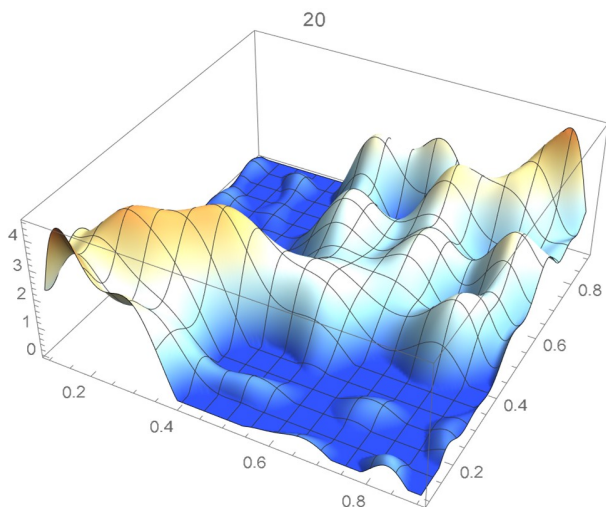
Figure 3.22: Reference least-squares copula density
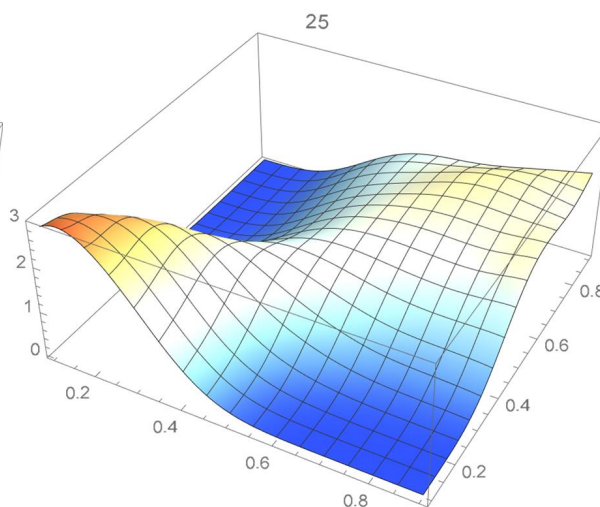
Figure 3.23: Bernstein's copula density of degree 25
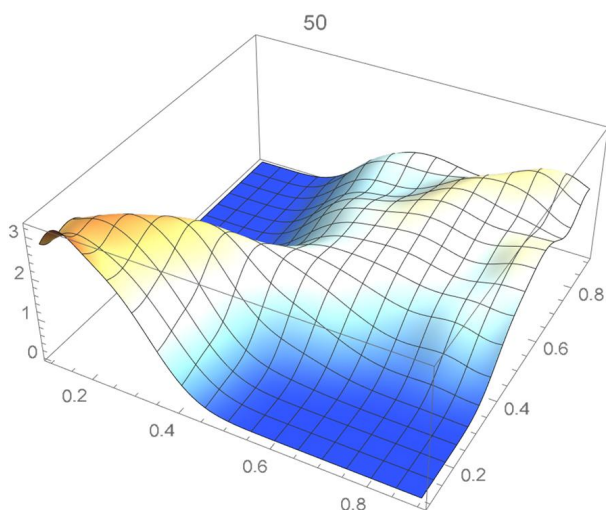
Figure 3.24: Bernstein's copula density of degree 50
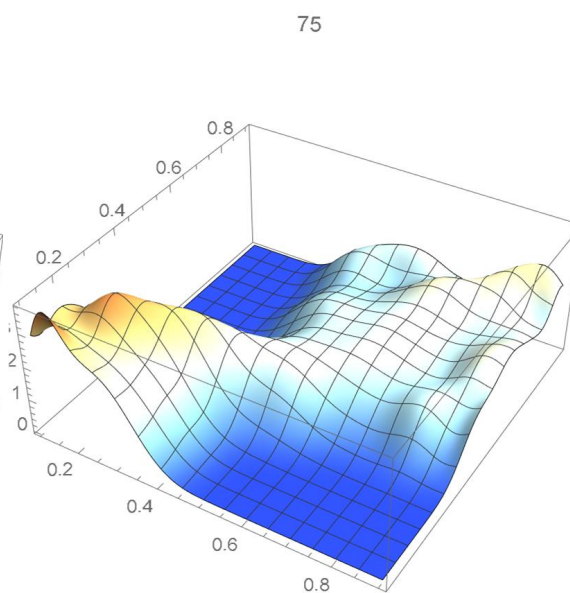
Figure 3.25: Bernstein's copula density of degree 75
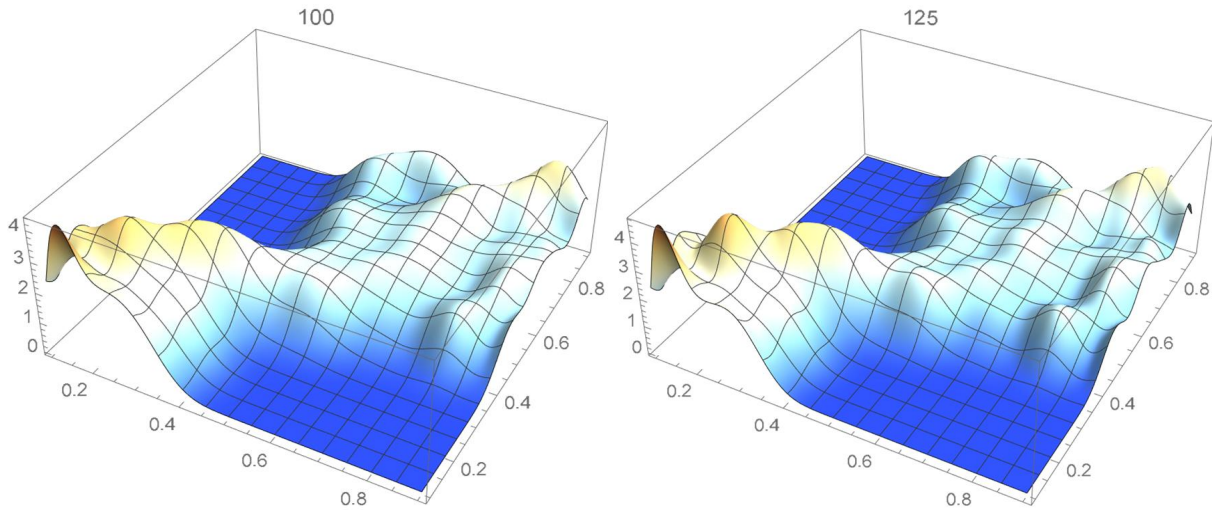
Figure 3.26: Bernstein's copula density of of degree 100

Figure 3.27: Bernstein's copula density of degree 125
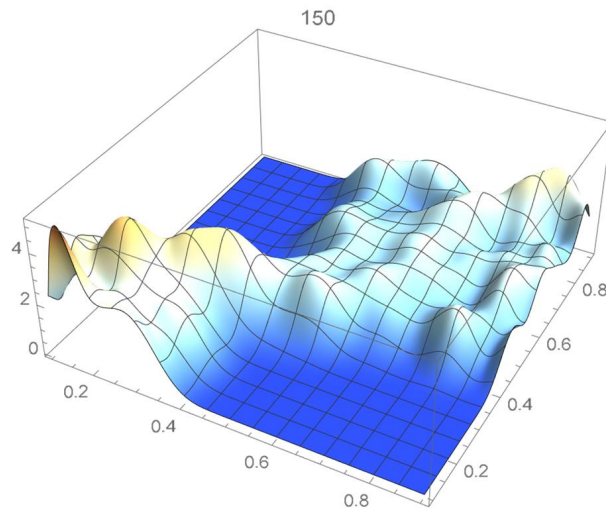
Figure 3.28: Bernstein's copula density of degree 150

It is seen from Table 3.2 that the ISD's keep decreasing as the degrees of the Bernstein's copula keep increasing from 25 to 200. However, beyond the degree 125, the ISD's with respect to the least-squares copula turn out to be of the same order. This can also observed by taking the relative differences between successive ISD's, i.e. |ISD(t)-ISD(t+25)|/ISD(t), referring to Table 3.3 and Figure 3.29.

The minimum relative difference is achieved between degrees 125 and 150. Accordingly,

t	ISD's of degrees t
25	1.02525×10^{-4}
50	3.47305×10^{-5}
75	1.57839×10^{-5}
100	1.12940×10^{-5}
125	8.42562×10^{-6}
150	7.08102×10^{-6}
175	5.42979×10^{-6}
200	5.96616×10^{-6}

Table 3.2: ISD's between Bernstein's copula and the reference copula

t	Relative differences between ISD's of degrees $t + 25$ and t
25	0.661248
50	0.545531
75	0.284463
100	0.253973
125	0.159585
150	0.233190

Table 3.3: Successive relative differences between ISD's
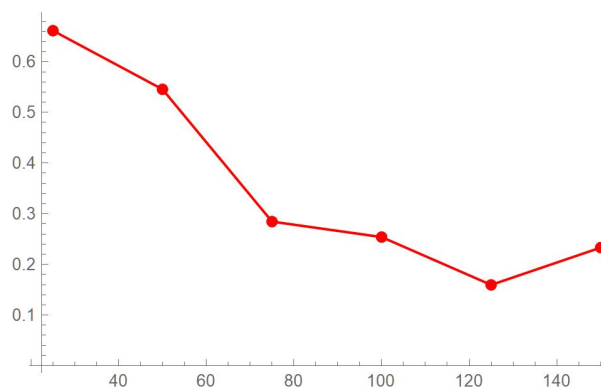
Figure 3.29: Successive relative differences between ISD's

we may conclude that a Bernstein's copula density of degree 125 in each variable is a suitable density estimate, which is in agreement with the assessment resulting from a visual comparison with the selected least-squares density estimate. Since the Bernstein polynomial approximation technique produces *bona fide* copula density estimates, it will be utilized as reference copula estimate with respect to the two subsequent approaches to copula density estimation.

3.2.3   Kernel-based copula density estimates

Among others, Li and Silvapulle (2015), Geenens et al. (2017), and Wen and Wu (2020) have utilized kernel density estimates (kde's) to estimate the density functions of copulas. Since the support of copulas is finite, kde's can produce what is referred to as 'boundary bias'. This drawback has previously been circumvented by making use of Epanechnikov kernels or a transformation method as discussed in Charpentier *et al.* (2007), or by applying a mirror reflexion technique as suggested by Gijbels and Mielniczuk (1990). It will be explained that it can as well be mitigated by repositioning the usual pseudo-observations and making use of biweight kernels.

Consider the four pseudo-observations discussed in Section 3.1.1 for the purpose of illustrating that utilizing centered pseudo-observations significantly alleviates the boundary issues resulting from the use of the original pseudo-observations when it comes to kernel density estimation; the kde's of the copula density obtained from the latter and the former (centered case) are plotted in Figures 3.30 and 3.31. The corresponding copulas, which were obtained by integrating the kde's, are shown in Figures 3.32 and 3.33.

It is seen that two of the kernels centered at the original pseudo-observations are truncated and as the graph of the cdf indicates, the resulting kde does not integrate to one. This does not occur when the centered pseudo-observation are utilized.
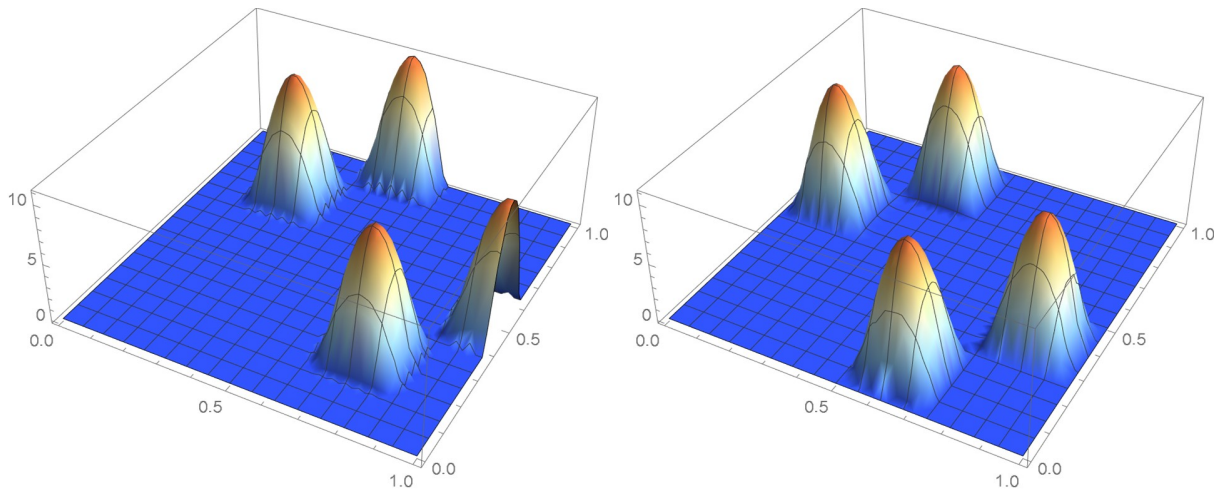
Figure 3.30: kde obtained from the original pseudo-observations

Figure 3.31: kde obtained from the centered pseudo-observations
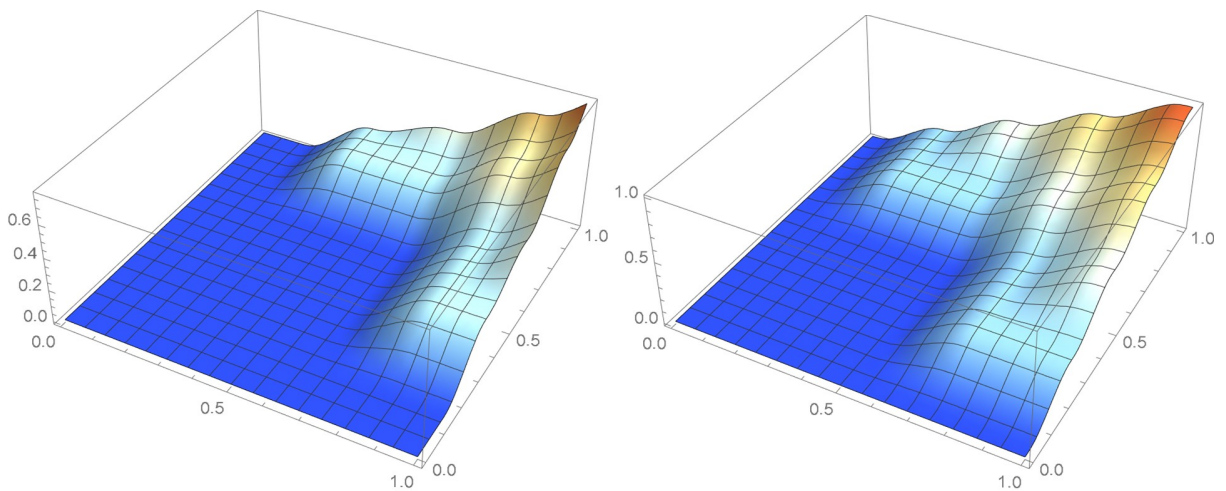
Figure 3.32: Copula obtained from the kde based on the original pseudo-observations

Figure 3.33: Copula obtained from the kde based on centered pseudo-observations

Optimal kde bandwidth selection

In the case of kde based copula densities, the narrower the kernels, the higher the peaks and the lower the troughs. A mathematical criterion for selecting the optimal bandwidth is proposed in this section. It has been shown that the centered copula points or centered pseudo-observations can be utilized to obtain improved copula density functions. Then, kde's of various bandwidths centered at those points are utilized for comparison purposes with respect to a reliable reference copula which in this instance is taken to be the selected Bernstein's copula density estimate of degree 125.

The selection criterion will rely on the integrated squared difference (on the reduced square) between the reference copula density and kde's of bandwidths 0.045, 0.04, 0.035, 0.030 and 0.025 in conjunction with Epanechnikov kernels, which are plotted in Figures 3.34-3.39.
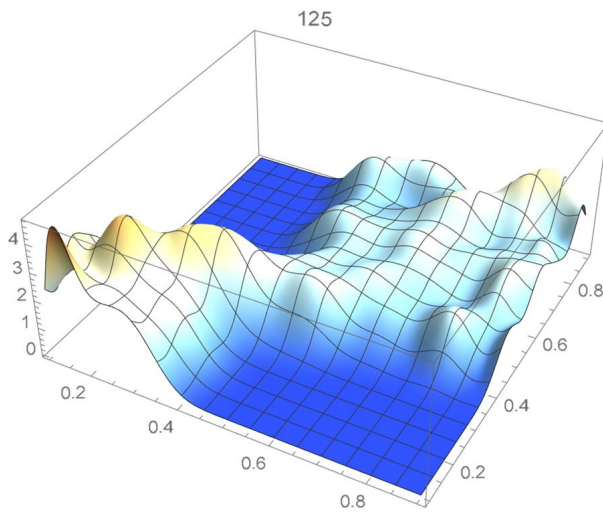
Figure 3.34: Reference copula density
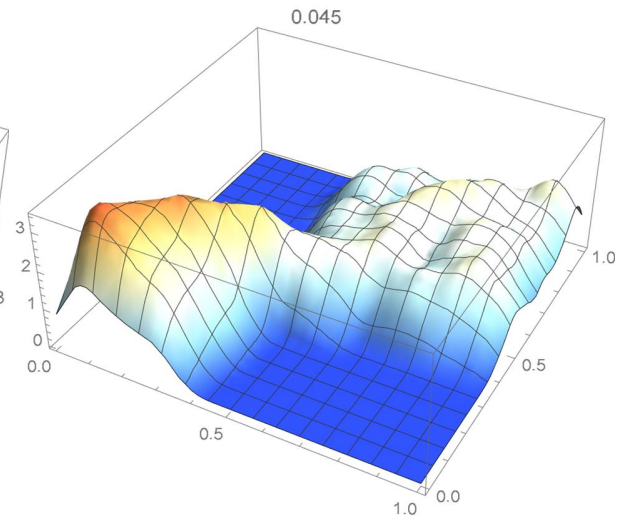
Figure 3.35: kde using the ccp's with bandwidth 0.045

It is seen from the resulting ISD's which are tabulated in Table 3.4 and displayed Figure 3.40, that the smallest one corresponds to a bandwidth of 0.035. Accordingly, the kde having this bandwidth is selected as being the most suitable one, a conclusion that could also have been reached by visual inspection.
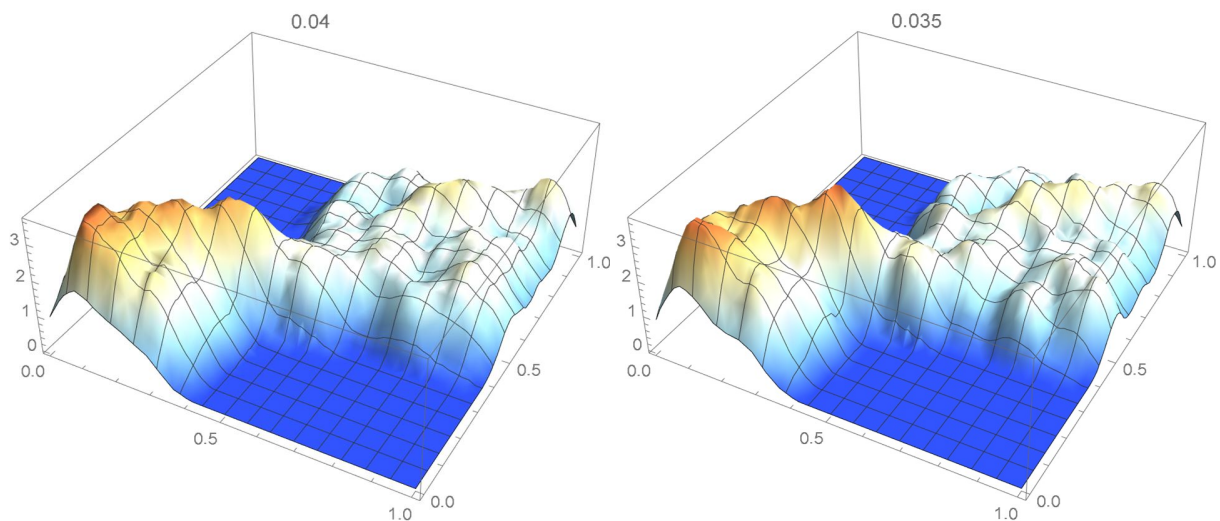
Figure 3.36: kde using the ccp's with bandwidth 0.040

Figure 3.37: kde using the ccp's with bandwidth 0.035
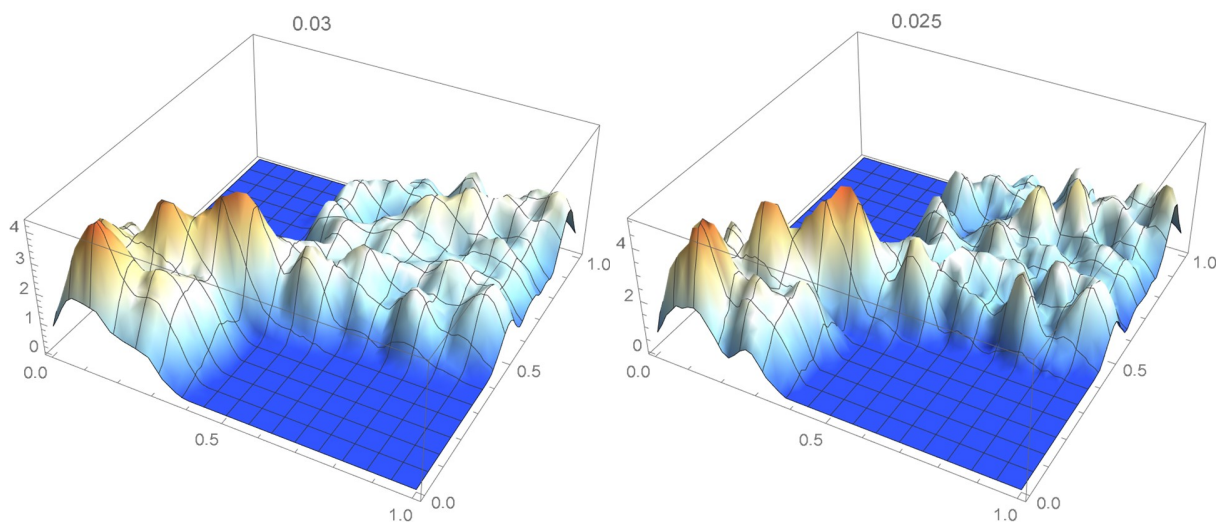
Figure 3.38: kde using the ccp's with bandwidth 0.030

Figure 3.39: kde using the ccp's with bandwidth 0.025

Bandwidth	ISD
0.045	0.0310480
0.040	0.0249804
0.035	0.0241592
0.030	0.0407768
0.025	0.0796507

Table 3.4: ISD's between the reference copula density and kde's having various bandwidths
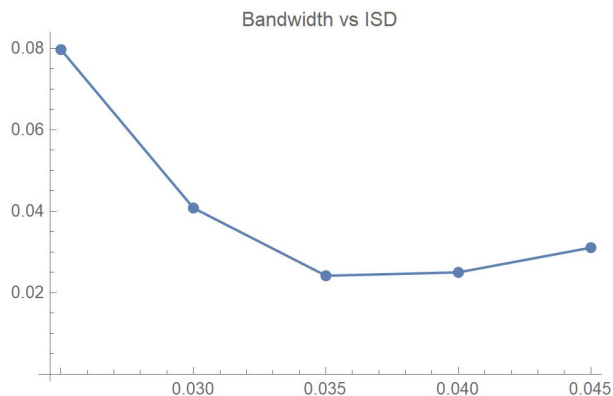
Figure 3.40: ISD's between the reference copula density and kde's having various bandwidths

3.2.4   Differentiated linearized empirical copulas

Another novel approach to copula density estimation is described in this subsection. A Deheuvels' empirical copula is first determined for the data set at hand. The empirical copula (cdf) is initially evaluated at grid points of the unit square whose associated spacing along both directions is denoted by c. Then, linear interpolation is applied to those points and the resulting surface is then differentiated, which yields an approximate density function. The resulting copula density which, as explained, is obtained from differentiating the linearized copula, will be referred to as a DL copula density. The spacing parameter c is selected in such a way that a reference copula density (for instance, a chosen Bernstein approximation) and the differentiated linearized copula share similar distributional features. Mathematically, the spacing parameter c is taken to be the minimizer of the integrated squared difference between the selected differentiated linearized copula and the chosen reference copula density.

Grid points spaced 1/12 apart of the empirical copula are plotted in Figure 3.41 for the Old Faithful geyser eruption data. The linear interpolation between these points and the resulting DL copula density are plotted in Figure 3.42-3.43. In this case, the optimal spacing parameter c based on ISD's (as shown in Table 3.5) is 1/12. The DL copula density functions are also plotted for $c = 1/11$ and $c = 1/13$ in Figures 3.44-3.45.
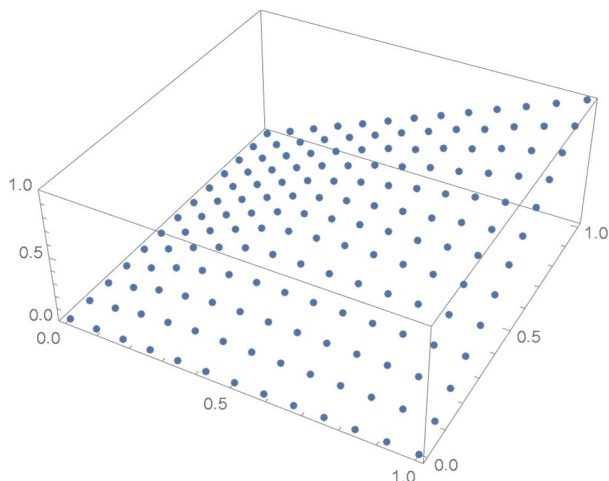
Figure 3.41: Empirical copula at grid points Figure 3.42: Linearly interpolated ecdf of the empirical copula at grid points spaced $1/12$ apart
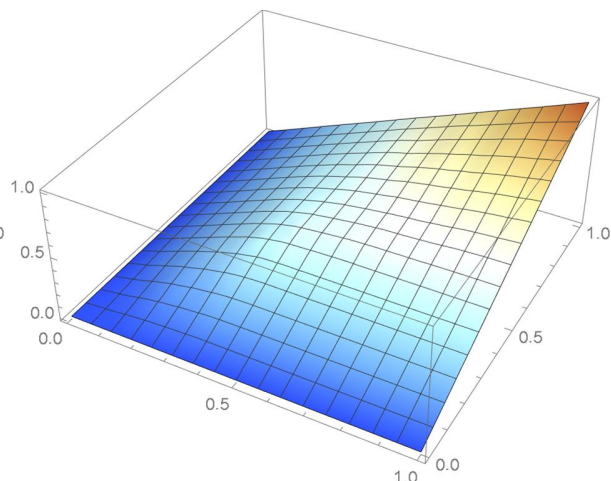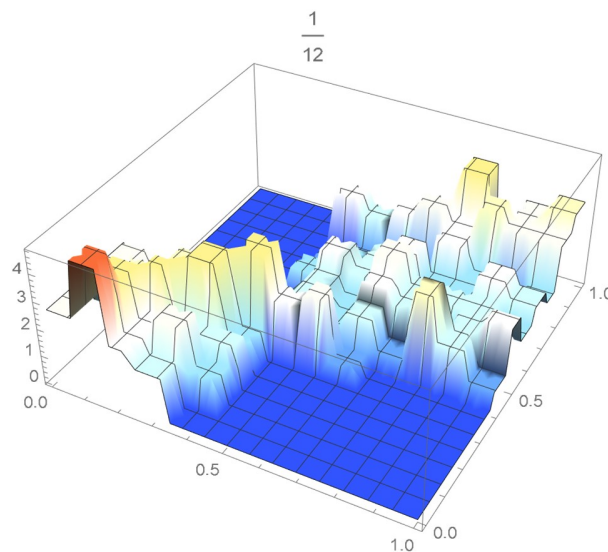
Figure 3.43: DL copula density with spacing parameter $c = 1/12$

Spacing parameter c	ISD
1/11	0.396789
1/12	0.360477
1/13	0.488780

Table 3.5: ISD's between the reference copula density and certain DL copula densities
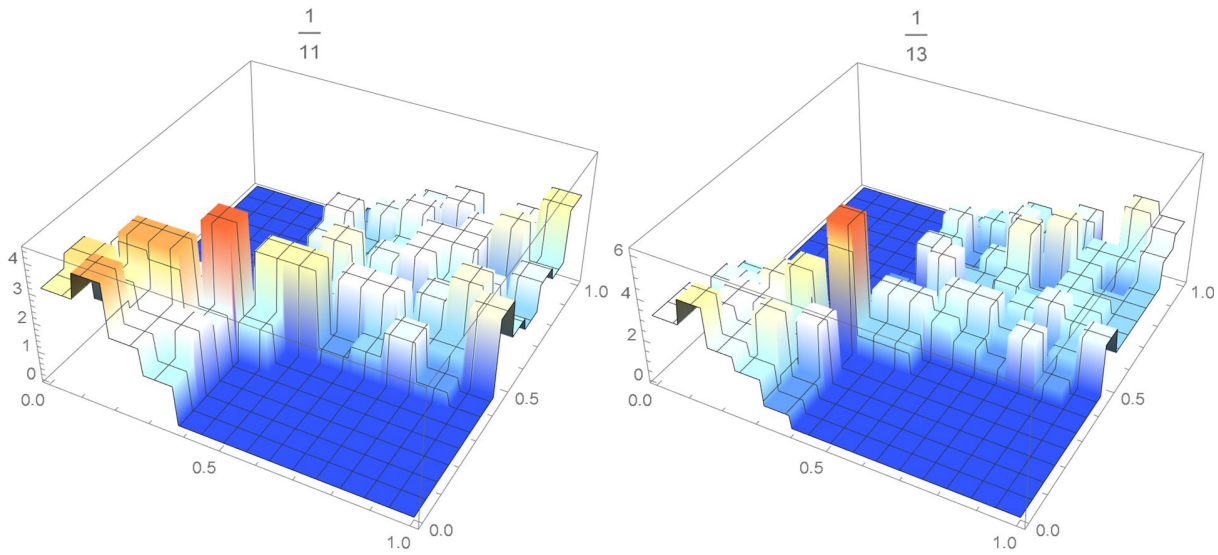
Figure 3.44: DL copula density with spacing parameter $c = 1/11$

Figure 3.45: DL copula density with spacing parameter $c = 1/13$

Smoothing a DL copula density by means of a bivariate polynomial

A smooth eleventh-degree bivariate polynomial approximation of the DL copula density (see Section 2.3) is shown in Figure 3.46. In this case, the base density was taken to be the uniform distribution.
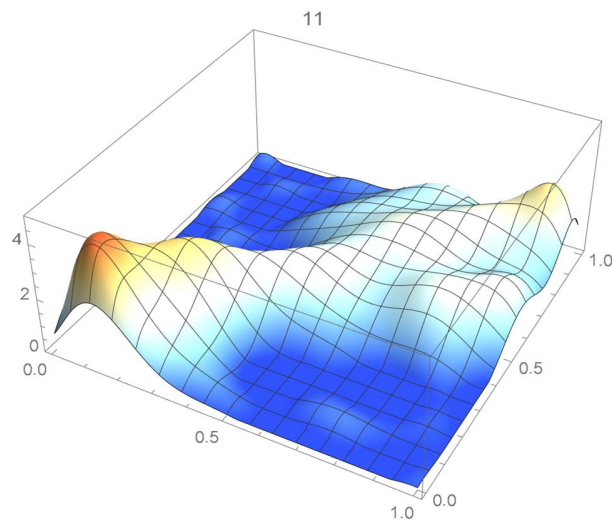
Figure 3.46: Smooth bivariate polynomial estimate of the DL copula density

3.3 On estimating joint density functions via copula density estimates

3.3.1 Introduction

The following formula which is consequence of Sklar's theorem, expresses a joint density function estimate in terms of estimates of the marginal density and distribution functions and a copula density estimate:

$$\hat{f}(x_1, x_2) \approx \hat{f}_1(x_1)\hat{f}_2(x_2)\,\hat{c}(\hat{F}_1(x_1), \hat{F}_2(x_2)). \tag{3.6}$$

Thus, once a copula density estimate has been secured, it is a rather simple matter to obtain a joint density estimate. More specifically, we proceed as follows: First, the marginal densities $f_1(x_1)$ and $f_2(x_2)$ associated with the random variables X_1 and X_2 are estimated and their respective distribution functions are obtained by integration; then, a copula density estimate is determined by making use of one of the proposed approaches such as those based on smoothing a differentiated linearized copula or evaluating a Bernstein polynomial approximation. This alternative approach to determining joint density function estimates allows for more flexibility than the direct approach, as one could for instance rely on some prior knowledge or information in order to choose distinct bandwidths for the marginal kde's *and* select an appropriate degree of smoothness for the copula density estimate.

3.3.2 An illustrative example

Consider once again the Old Faithful geyser eruption data as described in Section 3.2.1. A bivariate histogram of these observations is shown in Figure 3.47. A kde of the copula density whose optimal bandwidth was found to be 0.03, is plotted in Figure 3.48. Kernel density estimates of the marginal densities have then been obtained. They are superimposed on histograms of the observations on each variables in Figures 3.49 and 3.50. It is seen that the bivariate kde

shown in Figure 3.51, which was secured directly from the data, and the estimated joint density

obtained from equation (3.6), which appears in Figure 3.52, share similar features.
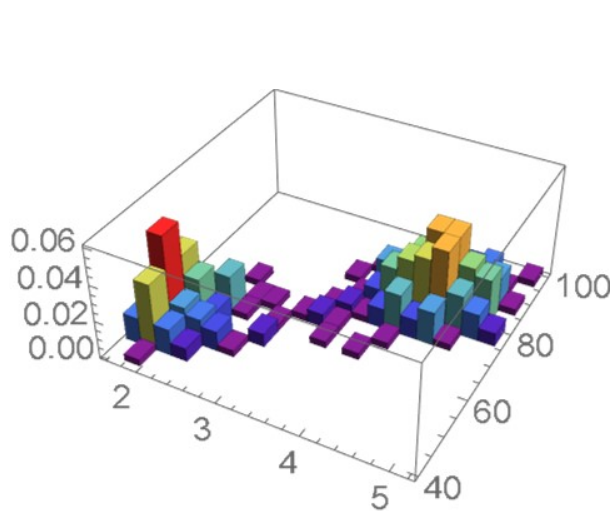
Figure 3.47: Bivariate histogram
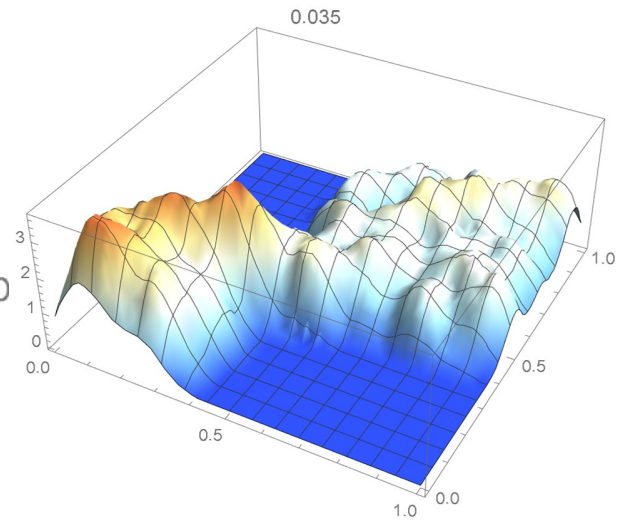
Figure 3.48: Copula kde with a bandwidth of 0.035
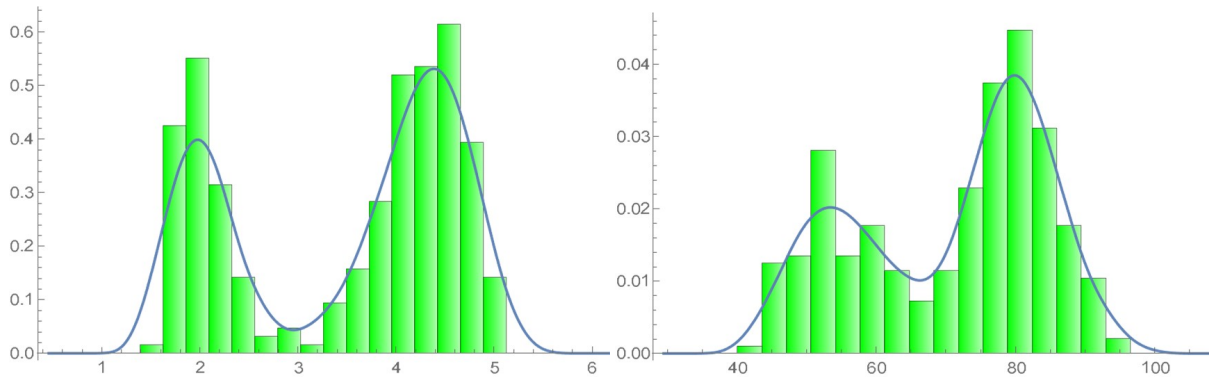
Figure 3.49: The estimated marginal density of the first variable and histogram

Figure 3.50: The estimated marginal density of the second variable and histogram
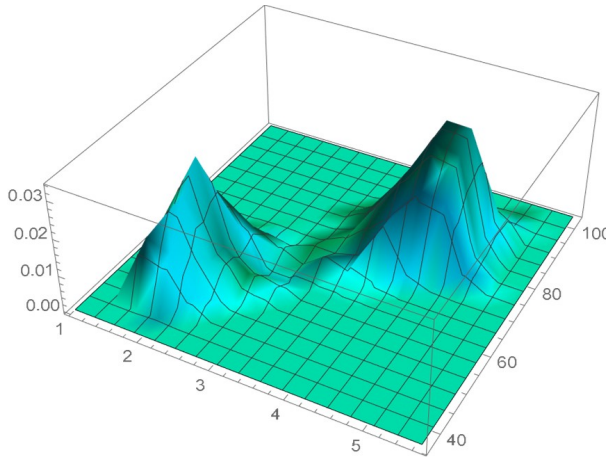
Figure 3.51: Bivariate kde of the copula
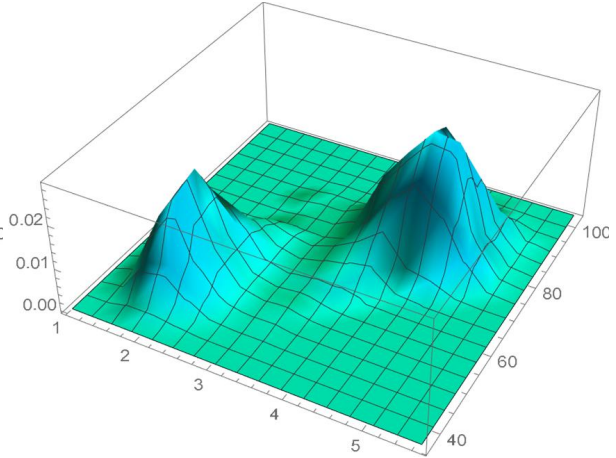
Figure 3.52: Joint density estimate resulting from applying Sklar's theorem

3.4 Estimating a *t*-distributed copula density

3.4.1 Introduction

In this section, we apply the proposed density estimation approaches to a random sample of 2000 points generated from a know copula distribution, namely a bivariate T on one degree of freedom, the marginal distributions being assumed to be standard normal and uniform on the interval $[0, 2]$. Note that the selected copula is challenging to model as its density function tends to plus infinity at each of the four vertices of the unit square. The joint density function and the copula density are plotted in Figures 3.53 and 3.54.

3.4.2 Application of the proposed approaches

Proceeding as explained in Section 3.2.1, it was determined that it is appropriate to use the differentiated least-squares bivariate polynomial approximation of degree 30 that is plotted in Figure 3.55, as initial density estimate and reference density.

On following the methodology advocated in Section 3.2.3, it was found that the kde-based estimate having 0.025 as its bandwidth which is shown in Figure 3.56, is suitable.

Referring to Section 3.2.2, its was determined that an appropriate degree for Bernstein's
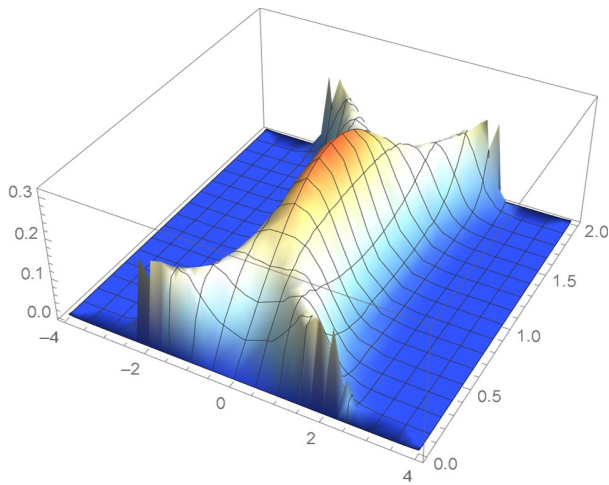
Figure 3.53: The joint density function
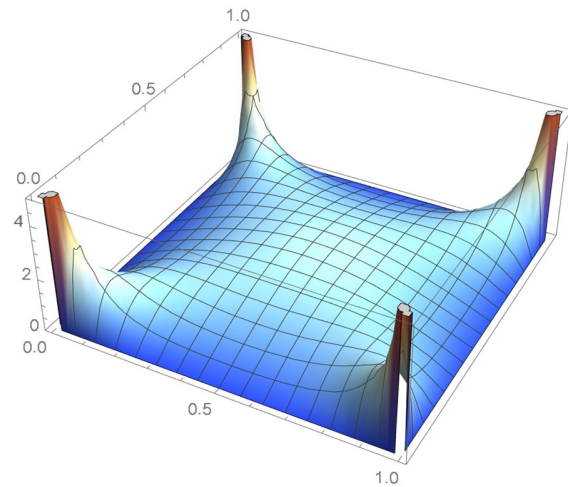
Figure 3.54: The bivariate T copula density on one degree of freedom
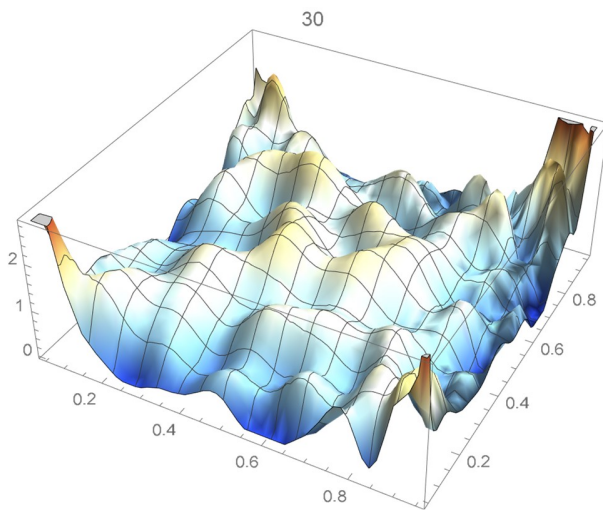
Figure 3.55: Differentiated least-squares estimate of degree 30
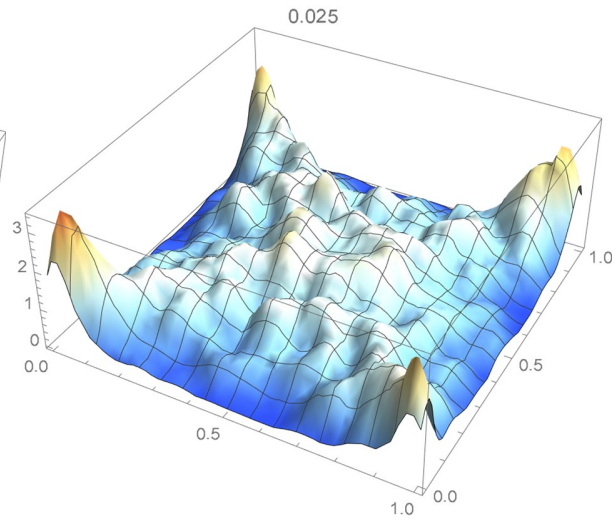
Figure 3.56: kde-based copula density whose bandwidth is 0.025

copula density estimate is 100. This copula density is plotted in Figure 3.57.

Proceeding as explained in Section 3.2.4, the optimal spacing for the DL copula density is $c = 1/12$. This copula density is shown in Figure 3.58. We note that the Bernstein polynomial approximation has the advantage of already having a smooth representation.
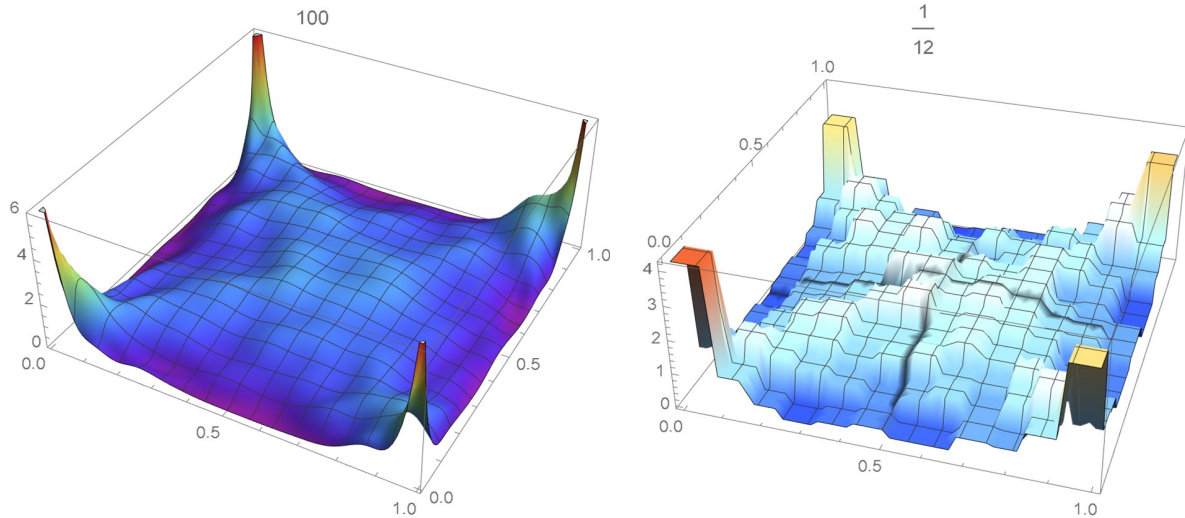
Figure 3.57: Bernstein's copula density of degree 100

Figure 3.58: Differentiated linearized empirical copula

A copula density estimate can also be obtained from Equation (2.7) which results from Sklar's theorem. In this instance, kde's were first obtained for the joint and marginal density functions, and the quantile functions were then determined from the latter. The copula density so obtained is plotted in Figure 3.59.

All these density estimates turn out to be similar to one another and consistent with the underlying distribution, which supports the validity of the various methodologies being advocated in this chapter. As a further step, bivariate splines could be utilized to approximate them.

3.4.3   Identification of the underlying distribution

Given a previously obtained copula density estimate of the underlying bivariate *t* copula distribution, we now verify whether it can be correctly identified when compared to several parametric copula density functions by means of the Hellinger distance.
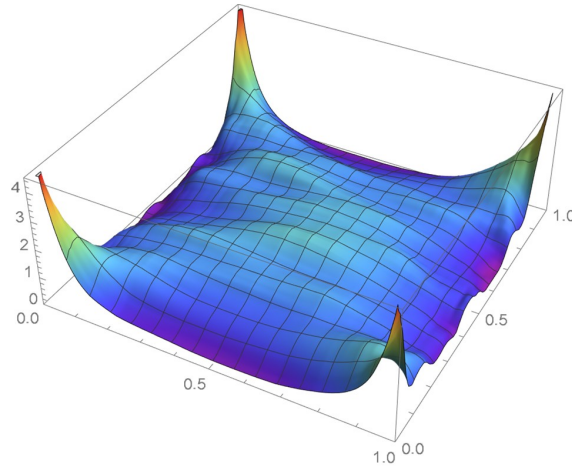
Figure 3.59: Copula density obtained by applying Sklar's theorem

If we denote the probability density functions of two bivariate distributions as $f(\cdot)$ and $g(\cdot)$, the square of the Hellinger distance is given by

$$H^2(f, g) = \frac{1}{2} \iint (\sqrt{f(x, y)} - \sqrt{g(x, y)})^2 dxdy. \tag{3.7}$$

The Hellinger distances between the Bernstein copula density approximation of degree 100 that is plotted in Figure 3.57 and the following copula density functions were evaluated: bivariate Student's t on 1, 3 and 10 degrees of freedom, bivariate Gaussian, Farlie-Gumbel-Morgenstern, Ali-Mikhail-Haq, Gumbel-Hougaard, Frank, and Clayton-Pareto. In this instance, the bounds of integration are zero and one.

As anticipated, the Hellinger distance between the estimated copula density and the bivariate t copula density on one degree of freedom turned out to be the smallest. The Hellinger distance approach constitutes an alternative to the Kullback-Leibler divergence and could be employed as a goodness-of-fit measure between a given copula density estimate and parametric copula density functions sharing similar features in order to determine the most appropriate model.

3.5   Concluding remarks

Although quite distinct in nature, the proposed approaches to estimating copula density functions were found to produce similar density estimates, which to a certain extent, validates their suitability. It was explained that such copula density estimates can be utilized to secure joint density functions. It was also verified that, on the basis of a sample drawn from a known copula distribution, the density estimation methodologies that were introduced in this chapter could yield quite accurate density estimates. It should be finally be mentioned that the various approaches advocated in this chapter can readily be extended to the estimation of multivariate copula density functions.

Bibliography

[1] Charpentier, A. , Fermanian, J. D. and Scaillet, O. (2007). The estimation of copulas: theory and practice, 35 – 60 Posted: 2007

[2] Cherubini, U., Fabio, G., Mulinacci, S. and Romagno, S. (2012). *Dynamic Copula Methods in Finance*. John Wiley & Sons, New York.

[3] Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley & Sons, New York.

[4] Denuit, M., Daehe, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons, New York.

[5] Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. Sage Publications, Los Angeles.

[6] Geenens, G., Charpentier, A. and Paindaveine, D. (2017). Probit transformation for non-parametric kernel estimation of the copula density. *Bernoulli*, 23(3), 1848–1873.

[7] Gijbels. I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics – Theory and Methods*, 19(2), 445 – 464.

[8] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, Boca Raton, Florida.

[9] Li, S. and Silvapulle, P. (2015). Kernel estimation of copula densities and applications. SSRN: https://ssrn.com/abstract=2620511 or http://dx.doi.org/10.2139/ssrn.2620511

[10] Nelsen, R. B. (2006). *An Introduction to Copulas.* Second Edition, Springer, New York.

[11] Pérez, C. J., Martín, J., Rufo, M. J. and Rojano, C. (2005). Quasi-random sampling importance resampling. *Communications in Statistics - Simulation and computation*, 34 (1), 97-112.

[12] Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

[13] Wang, Y. and Fang, K. T. (1990). Number theoretic methods in applied statistics. *Chinese Annals of Mathematics, Series B*, 11, 41–55.

[14] Wen, K. and Wu, X. (2020). Transformation-kernel estimation of copula densities. *Journal of Business & Economic Statistics*, 38 (1), 148–164

Chapter 4

Representations of Certain Measures of Association in terms of Copulas

4.1 Introduction

In this chapter, four correlation coefficients, namely, Spearman's ρ, Kendall's τ, Blomqvist's β and Hoeffding's Φ^2, are expressed in terms of copulas. Representations of their empirical counterparts are provided as well. As this chapter addresses mainly the bivariate case, such coefficients measure the strength of association and type of relationship between two variables.

The readers are assumed to be familiar with the notation and results that were introduced in Section 2.1. Now, given a random sample $(x_1, y_1), \ldots, (x_n, y_n)$ generated from the continuous random vector (X, Y), let

$$(u_i, v_i) = (F(x_i), G(y_i)), \quad i = 1, \ldots, n \tag{4.1}$$

where $F(\cdot)$ and $G(\cdot)$ are the usually unknown marginal cumulative distribution functions (cdf's) of X and Y. The empirical marginal cdf's, $\hat{F}(\cdot)$ and $\hat{G}(\cdot)$ are then utilized to determine the *pseudo-observations*:

$$(\hat{u}_i, \hat{v}_i) = (\hat{F}(x_i), \hat{G}(y_i)), \quad i = 1, \ldots, n, \tag{4.2}$$

where the empirical cdf's (ecdf's) are given by $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(x_j \leq x)$ and $\hat{G}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(y_j \leq y)$, $\mathrm{I}(\aleph)$ denoting the indicator which is equal to 1 if the condition $\aleph$ is verified and, 0, otherwise. Equivalently, one has

$$(\hat{u}_i, \hat{v}_i) = (r_i/n, s_i/n), \tag{4.3}$$

where r_i is the rank of x_i among $\{x_1, \ldots, x_n\}$ and s_i, the rank of y_i among $\{y_1, \ldots, y_n\}$.

The frequencies or probability mass function of an empirical copula can be expressed as

$$\begin{aligned}
\hat{c}(u, v) &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(\hat{F}(x_i) = u) \, \mathrm{I}(\hat{G}(y_i) = v) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(r_i/n = u) \, \mathrm{I}(s_i/n = v),
\end{aligned} \tag{4.4}$$

and the corresponding empirical copula is then given by

$$\begin{aligned}
\hat{C}(u, v) &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(\hat{F}(x_i) \leq u) \, \mathrm{I}(\hat{G}(y_i) \leq v) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(r_i/n \leq u) \, \mathrm{I}(s_i/n \leq v),
\end{aligned} \tag{4.5}$$

which is a consistent estimate of $C(u, v)$.

Note that in practice, the ranks are often divided by $n + 1$ instead of n so as to avoid certain boundary effects, and that other adjustments that are discussed in Section 4.2 may also be applied.

Additional properties of copulas that are not directly relevant to the results presented in this chapter are discussed for instance in Joe (1997), Cherubini *et al.* (2004, 2012), Denuit *et al.* (2005) and Nelsen (2006).

Finally, it should be pointed out that Pearson's correlation coefficient cannot be expressed in terms of copulas since it is a function of the observations themselves rather than their ranks.

This chapter contains certain derivations that do not seem to be available in the literature and also provides missing steps that complete the published proofs. It is structured as follows:

Sections 4.2, 4.3, 4.4 and 4.5 respectively focus on Spearman's, Kendall's, Blomqvist's and Hoeffding's measures of dependence and provides their representation in terms of copulas, in addition to pointing out certain related distributional results of interest. The effectiveness of these measures of association in assessing the trends present in data generated from five distinctive patterns is compared in a numerical study that is presented in Section 4.6.

To our knowledge, the four major measures of association are herein discussed along with their representations in terms of copulas, have not been altogether previously treated in a single source.

4.2   Spearman's Correlation Coefficient

Spearman's correlation statistic, also referred to as Spearman's ρ, measures the extent to which the relationship between two variables is monotonic—either increasing or decreasing.

First, a theoretical representation of Spearman's ρ is expressed in terms of a copula denoted by $C(U, V)$. Then, some equivalent representations of Spearman's rank correlation nonparametric statistic are provided, and it is explained that one of them can be obtained by replacing $C(U, V)$ by its empirical counterpart.

Let (X, Y) be a bivariate continuous random vector with joint density function $h(x, y)$ and $F(X)$ and $G(Y)$ denote the respective marginal distribution functions of X and Y.

Theoretically, Spearman's correlation is given by

$$\rho_S = \frac{Cov[F(X), G(Y)]}{\sqrt{Var[F(X)]Var[G(Y)]}} \tag{4.6}$$

$$= \frac{\iint_{\mathbb{R}^2} F(x)G(y)h(x, y)\mathrm{d}x\mathrm{d}y - (\int_{\mathbb{R}} F(x)\,\mathrm{d}F(x))(\int_{\mathbb{R}} G(y)\,\mathrm{d}G(y))}{\sqrt{[\int_{\mathbb{R}} F(x)^2\,\mathrm{d}F(x) - (\int_{\mathbb{R}}(F(x)\,\mathrm{d}F(x))^2][\int_{\mathbb{R}} G(y)^2\,\mathrm{d}G(y) - (\int_{\mathbb{R}}(G(y)\,\mathrm{d}G(y))^2]}} \tag{4.7}$$

$$= \frac{\int_0^1 \int_0^1 u\,v\,c(u,v)\,\mathrm{d}u\mathrm{d}v - (1/2)(1/2)}{\sqrt{(1/12)(1/12)}} \quad \text{in light of (2.7),} \tag{4.8}$$

with the transformation $\{x = F^{-1}(u) \text{ and } y = G^{-1}(v)\}$ whose Jacobian is the

inverse of the Jacobian associated with the following transformation :

$\{u = F(x) \text{ and } v = G(y)\}$, that is, $1/[f(F^{-1}(u))g(G^{-1}(v))]$,

$$= 12 \int_0^1 \int_0^1 C(u,v)\,\mathrm{d}u\mathrm{d}v - 3, \tag{4.9}$$

where $C(\cdot,\cdot)$ and $c(\cdot,\cdot)$ are respectively the copula and copula density associated with X and Y.

It is stated without proof in Kojanovik and Yan (2010) and Schmid and Schmidt (2007) that

the double integral in (4.8) can be expressed as that appearing in (4.9). We now prove that this

is indeed the case. First, recall that the copula density $c(u,v)$ is given by $\frac{\partial^2 C(u,v)}{\partial u \partial v}$. Integrating by

parts twice, one has

$$
\begin{aligned}
\int_0^1 \int_0^1 u\,v\,\mathrm{d}C(u,v) &= \int_0^1 \int_0^1 uv \frac{\partial^2 C(u,v)}{\partial u \partial v}\,\mathrm{d}v\mathrm{d}u \\
&= \int_0^1 u \left[\int_0^1 v \frac{\partial}{\partial v}\left(\frac{\partial C(u,v)}{\partial u} \right)\mathrm{d}v \right]\mathrm{d}u \\
&= \int_0^1 u \left[v\left.\frac{\partial C(u,v)}{\partial u}\right|_0^1 - \int_0^1 \frac{\partial C(u,v)}{\partial u}\mathrm{d}v \right]\mathrm{d}u \\
&= \int_0^1 u \left[1 - \int_0^1 \frac{\partial C(u,v)}{\partial u}\mathrm{d}v \right]\mathrm{d}u, \quad \text{as } C(u,1) = u \\
&= \int_0^1 u\mathrm{d}u - \int_0^1 \int_0^1 u\frac{\partial C(u,v)}{\partial u}\mathrm{d}v\mathrm{d}u \\
&= \frac{1}{2} - \int_0^1 \left[\left. uC(u,v)\right|_0^1 - \int_0^1 C(u,v)\mathrm{d}u \right]\mathrm{d}v \\
&= \frac{1}{2} - \frac{1}{2} + \int_0^1 \int_0^1 C(u,v)\mathrm{d}u\mathrm{d}v, \quad \text{as } C(1,v) = v \\
&= \int_0^1 \int_0^1 C(u,v)\mathrm{d}u\mathrm{d}v.
\end{aligned}
\tag{4.10}
$$

Now, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample generated from the random vector (X, Y),

and denote by $\hat{F}(X)$ and $\hat{G}(Y)$ the empirical distribution functions of X and Y, respectively.

Throughout this chapter, the sample size will be assumed to be n. On respectively denoting by

R_i and S_j, the rank of X_i among $\{X_1, \ldots, X_n\}$ and the rank of Y_j among $\{Y_1, \ldots, Y_n\}$, one has $\hat{F}(X_i) = R_i/n \equiv U_i$ and $\hat{G}(Y_j) = S_j/n \equiv V_j$, where U_i and V_j denote the canonical pseudo-observations on each component. Note that $\bar{R} = \bar{S} = (n+1)/2$. Then, as a random quantity, Spearman's rank correlation admits the following equivalent representations:

$$\hat{\rho}_S = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}} \tag{4.11}$$

$$= \frac{(\sum_{i=1}^{n} R_i S_i) - n\bar{R}\bar{S}}{\sqrt{[(\sum_{i=1}^{n} R_i^2) - n\bar{R}^2][(\sum_{i=1}^{n} S_i^2) - n\bar{S}^2)]}} \tag{4.12}$$

$$= \frac{(\sum_{i=1}^{n} \hat{F}(x_i) \hat{G}(y_i)) - n(n+1)^2/4}{\sqrt{[(\sum_{i=1}^{n} \hat{F}(x_i)^2) - n(n+1)^2/4][(\sum_{i=1}^{n} \hat{G}(y_i)^2) - n(n+1)^2/4]}} \tag{4.13}$$

$$= \frac{(\sum_{i=1}^{n} U_i V_i) - n(n+1)^2/4}{\sqrt{[(\sum_{i=1}^{n} U_i^2) - n(n+1)^2/4][(\sum_{i=1}^{n} V_i^2) - n(n+1)^2/4]}} \tag{4.14}$$

$$= \frac{\sum_{i=1}^{n}(U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^{n}(U_i - \bar{U})^2 \sum_{i=1}^{n}(V_i - \bar{V})^2}}, \tag{4.15}$$

where $\bar{U} = \sum_{i=1}^{n} U_i/n$ and $\bar{V} = \sum_{i=1}^{n} V_i/n$.

Of course (4.15) follows directly from (4.11), and it is seen from either one of these expressions that Spearman's correlation will not be affected by any monotonic affine transformation, whether applied to the ranks or the canonical pseudo-observations. Actually, the pseudo-observations are often taken to be

$$\hat{U}_i = \frac{R_i}{n+1} = \frac{n}{n+1}\hat{F}(x_i) = \frac{1}{n+1}\sum_{k=1}^{n} \mathrm{I}(x_k \le x_i) \tag{4.16}$$

and

$$\hat{V}_j = \frac{S_j}{n+1} = \frac{n}{n+1}\hat{G}(y_j) = \frac{1}{n+1}\sum_{k=1}^{n} \mathrm{I}(y_k \le y_j), \tag{4.17}$$

see for instance, Genest *et al.* (1995).

Another way to keep the pseudo-observations away from the upper bound of each unit

interval consists of defining them as follows:

$$\tilde{U}_i = \frac{R_i}{n} - \frac{1}{2n} = \hat{F}(x_i) - \frac{1}{2n} \tag{4.18}$$

and

$$\tilde{V}_j = \frac{S_j}{n} - \frac{1}{2n} = \hat{G}(y_j) - \frac{1}{2n}. \tag{4.19}$$

In a simulation study, Dias (2022) observed that such pseudo-observations have a lower bias than those obtained by dividing the ranks by $n + 1$. What is more, it should be observed that if we extend the pseudo-observations $\tilde{U}_i$, $i = 1, \ldots, n$, and $\tilde{V}_j$, $j = 1, \ldots, n$, by $\frac{1}{2n}$ on each side and assign their respective probability, namely $\frac{1}{n}$, to each of the n resulting intervals, the resulting marginal distributions will be uniformly distributed within the interval $[0, 1]$, which happens to be a requirement for a copula density. However, this will not be the case for any other affine transformation of the ranks or the empirical distribution functions. The alternative transformations $\frac{\text{rank}-1/3}{n+1/3}$ and $\frac{\text{rank}-1}{n-1}$ were also respectively considered by Kerman (2011) and Dias (2022). As established in Dias (2022), the pseudo-observation estimators resulting from any of the above-mentioned transformations as well as the canonical pseudo-observations are consistent estimators of the underlying distribution functions.

Kojanovik and Yan (2010) pointed out that $\hat{\rho}_S$ as specified in (4.12) can also be expressed as

$$\hat{\rho}_S = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^{n} R_i S_i - 3\frac{n+1}{n-1}, \tag{4.20}$$

where $\hat{\rho}_S$ is a consistent estimator of ρ_S.

As well, it can be algebraically shown that, alternatively,

$$\hat{\rho}_S = 1 - 6 \sum_{i=1}^{n} \frac{(R_i - S_i)^2}{n(n^2 - 1)} \tag{4.21}$$

when the ranks are distinct integers.

On writing (4.9) as

$$12 \int_0^1 \int_0^1 u\,v\,\mathrm{d}C(u,v) - 3,\tag{4.22}$$

and replacing $C(u,v)$ by $\hat{C}(u,v)$ as defined in (4.5), the double integral becomes

$$\frac{1}{n} \sum_{i=1}^n \int_0^1 u\,\mathrm{d}(\mathrm{I}(r_i/n \le u)) \int_0^1 v\,\mathrm{d}(\mathrm{I}(s_i/n \le v)).$$

For instance, on integrating the first integral by parts, one has

$$u\,\mathrm{I}(r_i/n \le u)\big|_0^1 - \int_0^1 \mathrm{I}(r_i/n \le u)\,\mathrm{d}u = 1 - (1 - r_i/n) = r_i/n.$$

Thus, the resulting estimator of Spearman's rank correlation is given by

$$\frac{12}{n^3} \sum_{i=1}^n R_i S_i - 3,$$

which is approximately equal to the representation specified in (4.20).

Now, letting $C_\theta(u,v)$ be a copula whose functional representation is known, and assuming that it is a one-to-one function of the dependence parameter θ, it follows from (4.9) that

$$\rho_S(\theta) = 12 \int_0^1 \int_0^1 C_\theta(u,v)\mathrm{d}u\mathrm{d}v - 3,\tag{4.23}$$

which provides an indication of the extent to which the variables are monotonically related. Moreover, since $\hat{\rho}_S$ as defined in (4.12), (4.20) or (4.21) tends to $\rho_S(\theta)$, an estimate of θ can be obtained as $\hat{\theta} = \rho_S^{-1}(\hat{\rho}_S)$.

It follows from (4.9) that Spearman's ρ can be expressed as

$$\rho_{X,Y} = \rho_C = 12 \iint_{\mathbb{I}^2} [C(u,v) - uv]\mathrm{d}u\mathrm{d}v.\tag{4.24}$$

If we replace $[C(u,v) - uv]$ in (4.24) by $|C(u,v) - uv|$, we have a measure based on the

L_1 distance between the copula C and the product copula $\Pi = uv$ (Nelsen 2006). This is the so-called Schweizer-Wolff's sigma that was defined in Schweizer and Wolff (1981), and which is given by

$$\sigma_{X,Y} = \sigma_C = 12 \iint_{\mathbb{1}^2} |C(u, v) - uv| \mathrm{d}u\mathrm{d}v. \tag{4.25}$$

The expression (4.25) is a measure of dependence which satisfies the properties of Rényi's axioms (Rényi, 1959) for measures of dependence (Schweizer and Wolff, 1981; Lai and Balakrishnan, 2009, p. 145).

Note that the Pearson's correlation coefficient

$$\hat{r} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.26}$$

only measures the strength of a *linear* relationship between X and Y whereas Spearman's rank correlation ρ_S assesses the strength of any monotonic relationship between X and Y. The latter is always well-defined, which is not the case for the former. Both vary between -1 and 1 and $\rho_S = \pm 1$ indicates that Y is either an increasing or a decreasing function of X.

4.3 Kendall's Correlation Coefficient

Kendall's τ is a non-parametric measure of association between two variables, which is based on the number of concordant pairs minus the number of discordant pairs. Also referred to as Kendall's rank correlation coefficient, it is named after Maurice Kendall who introduced this measure of dependence in an article published in 1938. He also proposed its estimate $\hat{\tau}$ and published several papers and a monograph in connection with certain ordinal measures of correlation. Further historical details about Kendall's τ are provided in Kruskal (1958).

Consider two observations (x_i, y_i) and (x_j, y_j), with $(i, j) \in \{1, \ldots, n\}$ such that $i \neq j$, that are drawn from a vector (X, Y) of continuous random variables. Then, for any such assignment of pairs, define each pair as being concordant, discordant, or equal, as follows:

- (x_i, y_i) and (x_j, y_j) are concordant if:

  $\{x_i < x_j$ and $y_i < y_j$ or if $x_i > x_j$ and $y_i > y_j\}$, or equivalently

  $(x_i - x_j)(y_i - y_j) > 0$, i.e., the slope of the line going through the two points is positive.

- In the same way, (x_i, y_i) and (x_j, y_j) are discordant if:

  $\{x_i < x_j$ and $y_i > y_j$ or if $x_i > x_j$ and $y_i < y_j\}$, or equivalently

  $(x_i - x_j)(y_i - y_j) < 0$, i.e., the slope of the line going through the two points is negative.

- (x_i, y_i) and (x_j, y_j) are equal if $x_i = x_j$ or $y_i = y_j$. Pair equality can be disregarded since the random variables X and Y are assumed to be continuous.

The empirical Kendall's τ

Let $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ be a random sample of n pairs drawn from the vector (X, Y) of continuous random variables. There are $C_2^n = \binom{n}{2}$ possible ways of selecting distinct pairs (x_i, y_i) and (x_j, y_j) of observations in the sample, with each pair being either concordant or discordant.

Let S_{ij} be defined as follows:

$$S_{ij} = \text{sign}(X_i - X_j)\,\text{sign}(Y_i - Y_j), \tag{4.27}$$

where

$$\text{sign}(u) = \begin{cases} -1 & \text{if } u < 0 \\ 0 & \text{if } u = 0 \\ 1 & \text{if } u > 0. \end{cases}$$

Then, the values taken by S_{ij} are:

$$
s_{ij} = \begin{cases} -1 & \text{when the pairs are discordant} \\ 0 & \text{when the pairs are neither concordant nor discordant} \\ 1 & \text{when the pairs are concordant.} \end{cases}
$$

Since we have $s_{ij} = s_{ji}$ and $s_{ii} = 0$, then there are C_2^n sets of pairs to consider. Kendall's sample $\hat{\tau}$ is defined as follows:

$$
\hat{\tau} = \sum_{1 \leq i < j \leq n} \frac{s_{ij}}{C_2^n} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} s_{ij}. \tag{4.28}
$$

Using an alternative approach, let c denote the number of concordant pairs and d, the number of discordant pairs, then Kendall's τ for that sample is given by

$$
\hat{\tau} = \frac{c-d}{c+d} = \frac{c-d}{C_2^n} = \frac{2(c-d)}{n(n-1)}. \tag{4.29}
$$

As it is assumed that there can be no equality between pairs, $C_2^n = c + d$ so that

$$
\hat{\tau} = \frac{4c}{n(n-1)} - 1 \ \text{ or } \ \hat{\tau} = 1 - \frac{4d}{n(n-1)}. \tag{4.30}
$$

Note that $\hat{\tau}$ is an unbiased estimator of τ. Additionally, Kendall and Gibbons (1990, Chapter 5) and Valz and McLeod (1990) established that $\text{Var}(\hat{\tau}) = \frac{2(2n+5)}{9n(n-1)}$.

The population Kendall's τ

Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed random vectors, with the joint distribution function of (X_i, Y_i) being $H(x, y)$, and let $F(x)$ and $G(y)$ denote the distribution functions of X_i and Y_j, $i, j = 1, 2$, and the associated copula be $C(u, v) = H(F^{-1}(u), G^{-1}(v))$.

The population Kendall's τ is defined as:

$$\tau = \tau_{X,Y} = \Pr[\text{concordant pairs}] - \Pr[\text{discordant pairs}]$$

$$\equiv p_c - p_d$$

$$= \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_2) < 0] \tag{4.31}$$

$$= 2\Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 \tag{4.32}$$

$$= 4\Pr[(X_1 < X_2, Y_1 < Y_2)] - 1 \tag{4.33}$$

$$= 4 \iint_{\mathbb{R}^2} \Pr(X_2 \le x, Y_2 \le y)\mathrm{d}H(x,y) - 1, \text{ where } H(x,y) = C(F(x), G(y)) \tag{4.34}$$

$$= 4 \iint_{\mathbb{R}^2} H(x,y)c(F(x), G(y))f(x)g(y)\mathrm{d}x\mathrm{d}y - 1$$

$$= 4 \iint_{\mathbb{1}^2} \frac{H(F^{-1}(u), G^{-1}(v))c(u,v)f(F^{-1}(u))g(G^{-1}(v))}{f(F^{-1}(u))g(G^{-1}(v))} \mathrm{d}u\mathrm{d}v - 1$$

$$= 4 \int_0^1 \int_0^1 C(u,v)\mathrm{d}C(u,v) - 1 \tag{4.35}$$

$$= 4\,\mathrm{E}[C(U,V)] - 1, \tag{4.36}$$

where

- U and V have a Uniform $(0,1)$ distribution with joint cdf C.

- $u = F_X(x)$ and $v = F_Y(y)$.

- $\mathbb{R}^2 \equiv \{(x,y) \mid x \text{ and } y \text{ are real numbers}\}$.

- $\mathrm{d}C(u,v) = \frac{\partial^2 C(u,v)}{\partial u_1 \partial v}\mathrm{d}u\mathrm{d}v = c(u,v)\mathrm{d}u\mathrm{d}v$.

Note that (4.32) follows from (4.31) since

$$\Pr[(X_1 - X_2)(Y_1 - Y_2) < 0] = 1 - \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0].$$

Marginal probability of S_{ij}

The marginal probability of S_{ij} is

$$p_{S_{ij}}(s_{ij}) = \begin{cases} p_c, & s_{ij} = 1 \\ p_d, & s_{ij} = -1 \\ 1 - p_c - p_d, & s_{ij} = 0. \end{cases}$$

Gibbons and Chakraborti (2003) proved that

$$E(S_{ij}) = 1 p_c + (-1)p_d = \tau.$$

Some properties of τ

- The coefficient correlation τ is invariant with respect to strictly increasing transformations.

- This correlation coefficient takes on values in the interval of $[-1, 1]$.

- If X_1 and Y_1 are independent, then the value of τ is zero:

$$\tau(X_1, Y_1) = 2 \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1$$
$$= 2\{\Pr[X_1 - X_2 > 0, Y_1 - Y_2 > 0] + \Pr[X_1 - X_2 < 0, Y_1 - Y_2 < 0]\} - 1$$
$$= 2\left(\frac{1}{4} + \frac{1}{4}\right) - 1 = 0.$$

- When the number of discordant pairs is 0, then the value of τ is maximum and equals 1, which means a perfect relationship; the variables are then comonotonic, i.e., one variable is an increasing transform of the other. If the variables are countermonotonic, i.e., one variable is a decreasing transform of the other, the correlation coefficient τ equals -1, see Joe (1997).

Note that the last two properties do not hold for Pearson's correlation coefficient. Moreover, Kendall's τ is more appropriate when the joint distribution is not Gaussian.

4.4   Blomqvist's Correlation Coefficient

Blomqvist (1950) proposed a measure of dependence similar in its structure to Kendall's correlation coefficient, except that medians are being used. Blomqvist's correlation coefficient can be defined as follows:

$$
\begin{aligned}
\beta = \beta_{X,Y} &= P[(X - F_X^{-1}(1/2))(Y - G_Y^{-1}(1/2)) > 0] \\
&\quad - P[(X - F_X^{-1}(1/2))(Y - G_Y^{-1}(1/2)) < 0],
\end{aligned}
\tag{4.37}
$$

where $F_X^{-1}(1/2) \equiv \tilde{x}$ and $G_Y^{-1}(1/2) \equiv \tilde{y}$ are the medians of X and Y respectively, which explain why this coefficient is also known as the median correlation coefficient.

Now, let X and Y be continuous random variables having $H(\cdot, \cdot)$ as joint distribution function, $F(\cdot)$ and $G(\cdot)$ as respective marginals, and $C(\cdot, \cdot)$ as their copula; then,

$$
F(\tilde{x}) = F(F_X^{-1}(1/2)) = G(\tilde{y}) = G(G_Y^{-1}(1/2)) = 1/2,
$$

and

$$
\beta = \beta_{X,Y} = 2\Pr[(X - F_X^{-1}(1/2))(Y - G_Y^{-1}(1/2)) > 0] - 1 \tag{4.38}
$$

$$
= 2\{\Pr[X < F_X^{-1}(1/2), Y < G_Y^{-1}(1/2)] + \Pr[X > F_X^{-1}(1/2), Y > G_Y^{-1}(1/2)]\} - 1 \tag{4.39}
$$

$$
= 4H(F_X^{-1}(1/2), G_Y^{-1}(1/2)) - 1 \tag{4.40}
$$

$$
= 4C(1/2, 1/2) - 1. \tag{4.41}
$$

In the development of these equations, the following relationships were utilized in addition to $H(x, y) = C(F(x), G(y))$:

$$
\begin{aligned}
P[(X - F_X^{-1}(1/2))(Y - G_Y^{-1}(1/2)) > 0] &= P[X - F_X^{-1}(1/2) > 0, Y - G_Y^{-1}(1/2) > 0] \\
&\quad + P[X - F_X^{-1}(1/2) < 0, Y - G_Y^{-1}(1/2) < 0];
\end{aligned}
\tag{4.42}
$$

$$P[X > F_X^{-1}(1/2), Y > G_Y^{-1}(1/2)] = P[X < F_X^{-1}(1/2), Y < G_Y^{-1}(1/2)]. \tag{4.43}$$

Estimation of β

Let $\tilde{x}_n$ and $\tilde{y}_n$ be the medians of the samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. The computation of Blomqvist's correlation coefficient is based on a 2×2 contingency table constructed from these two samples.

According to Blomqvist's suggestion, the $x-y$ plane is divided into four regions by drawing the lines $x = \tilde{x}_n$ and $y = \tilde{y}_n$.

Let n_1 and n_2 be the number of points belonging respectively to the first or third quadrants, and to the second or fourth quadrants.

Blomqvist's sample β_n or the median correlation coefficient is defined by

$$\beta_n = \frac{n_1 - n_2}{n_1 + n_2} = 2\frac{n_1}{n_1 + n_2} - 1. \tag{4.44}$$

If the sample size n is even, then, it is obvious that no sample points will fall on the lines $x = \tilde{x}_n$ and $y = \tilde{y}_n$. Moreover, it follows that n_1 and n_2 will be even. However, if n is odd, then one or two sample points must fall on the lines $x = \tilde{x}_n$ and $y = \tilde{y}_n$. In the first case (a single point lying on a median), Blomqvist proposed that this point shall not be counted. For the second case, one point has to fall on each line. Then, one of the points is assigned to the quadrant touched by both points, while the other is not counted, so that, n_1 and n_2 remain even.

Genest *et al.* (2013) provide an accurate interpretation of β_n as "the difference between the proportion of sample points having both components either smaller or greater than their respective medians, and the proportion of the other sample points". Finally, as pointed out by Blomqvist (1950), the definition of β_n was not new (Mosteller, 1946); however, its statistical properties had never been fully investigated.

Some properties of Blomqvist's correlation coefficient:

- The coefficient β is invariant under strictly increasing transformations of X and Y.

- The correlation coefficient β takes on values in the interval $[-1, 1]$.

- If X and Y are independent, then $C(1/2, 1/2) = F(1/2)G(1/2) = 1/4$, and $\beta = 0$.

4.5   Hoeffding's Dependence Coefficient

To measure the strength of relationships that are not necessarily monotonic, one may make use of Hoeffding's dependence coefficient. Letting $H(X, Y)$ denote the joint distribution function of X and Y, and $F(X)$ and $G(Y)$ stand for the marginal distribution functions of X and Y, Hoeffding's nonparametric rank test for bivariate independence, which is consistent against all bivariate dependence alternatives, is based on

$$D(x, y) = H(x, y) - F(x)G(y), \tag{4.45}$$

which is equal to zero if and only if X and Y are independently distributed. The nonparametric estimator of the quantity $\mathcal{D}^2 = 30 \int D^2(x, y) \mathrm{d}H(x, y)$ results in the statistic,

$$\hat{\mathcal{D}}^2 = 30 \frac{Q - 2(n-2)R + (n-2)(n-3)S}{n(n-1)(n-2)(n-3)(n-4)}, \tag{4.46}$$

where

$$Q = \sum_{i=1}^{n} (R_i - 1)(R_i - 2)(S_i - 1)(S_i - 2), \tag{4.47}$$

$$R = \sum_{i=1}^{n} (R_i - 2)(S_i - 2)C_i, \tag{4.48}$$

and

$$S = \sum_{i=1}^{n}(C_i - 1)C_i, \tag{4.49}$$

C_i being the number of bivariate observations (X_j, Y_j) for which $X_j \leq X_i$ and $Y_j \leq Y_i$.

We now state Hoeffding's lemma (1940): *Let X and Y be random variables with joint distribution function $H(x, y)$ and marginal distributions $F(x)$ and $G(y)$. If $E(XY)$ and $E(X)E(Y)$ are finite, then*

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[H(x, y) - F(x)G(y)]\mathrm{d}x\mathrm{d}y. \tag{4.50}$$

This result became known when it was cited by Lehmann (1966). Jogdeo (1968) proposed an extension to several variables. Two decades later, Block and Fang (1988) discussed a multivariate version of this lemma.

The correlation coefficient is thus given by

$$\text{Cor}(X, Y) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[H(x, y) - F(x)G(y)]\mathrm{d}x\mathrm{d}y}{\sqrt{Var(X)}\sqrt{Var(Y)}} \tag{4.51}$$

or

$$\text{Cor}(X, Y) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[C(F(x), G(y)) - F(x)G(y)]\mathrm{d}x\mathrm{d}y}{\sqrt{Var(X)}\sqrt{Var(Y)}}, \tag{4.52}$$

with (4.52) resulting from Sklar's theorem.

Using Hoeffding's lemma, Hofert *et al.* (2019, p. 47) pointed out two fallacies about the uniqueness and independence of random variables. Hoeffding made use of his lemma to identify the bivariate distributions with given marginal distribution functions $F(x)$ and $G(y)$, which minimize or maximize the correlation between X and Y.

Hoeffding's Φ^2

Hoeffding (1940) defined the stochastic dependence index of the random variables X and Y as

$$\Phi_{X,Y}^2 = 90 \int_0^1 \int_0^1 (C(u, v) - uv)^2 \mathrm{d}u\mathrm{d}v, \tag{4.53}$$

where

$$\Phi^2_{X,Y} = \begin{cases} 0 & \text{in the case of independence since then } C(u,v) = uv \\ 1 & \text{in the case of monotone dependence} \\ \Phi^2 \in (0,1) & \text{otherwise,} \end{cases}$$

90 being a normalization factor.

Hoeffding (1940) showed that $\Phi^2_{X,Y}$ takes the value 1 in the cases of monotonically increasing and monotonically decreasing continuous functional dependence; it is otherwise less than 1 and greater than zero.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an (i.i.d.) random sample generated from the 2-dimensional random vector $\mathbf{X}$ with distribution function H and copula C. We assume that both F and C as well as the marginal distribution functions F_i, $i = 1, 2$, are completely unknown. The copula C is estimated by the empirical copula $\hat{C}_n$, which is defined as

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{2} I(\hat{U}_{ij,n} \leq u_i) \quad \text{for } \mathbf{u} \in [0,1]^2, \tag{4.54}$$

with pseudo-observations $\hat{U}_{ij,n} = \hat{F}_{i,n}(X_{ij})$ for $i = 1, 2$ and $j = 1, \ldots, n$, and $\hat{F}_{i,n}(x) = \frac{1}{n} \sum_{j=1}^{n} I(X_{ij} \leq x)$ for $x \in \mathbb{R}$. Since $\hat{U}_{ij,n} = \frac{1}{n}(\text{rank of } X_{ij} \text{ in } X_{i1}, \ldots, X_{in})$, statistical inference is based on the ranks of the observations. When it is clear from the context, the index n is suppressed, and the pseudo-observations are denoted by $\hat{U}_{ij}$.

A nonparametric estimator of Φ^2 is then obtained by replacing the copula C in

$$\Phi^2 := \Phi^2(C) = 90 \iint_{\mathbb{I}^2} \{C(\mathbf{u}) - \Pi(\mathbf{u})\}^2 d\mathbf{u}, \tag{4.55}$$

by the empirical copula $\hat{C}_n$, i.e.,

$$\hat{\Phi}^2_n := \Phi^2(\hat{C}_n) = 90 \iint_{\mathbb{I}^2} \{\hat{C}_n(\mathbf{u}) - \Pi(\mathbf{u})\}^2 d\mathbf{u}, \tag{4.56}$$

where $\Pi(\mathbf{u}) = u_1 u_2$ denotes the independence copula.

As explained in Gaißer *et al.* (2010), this estimator can be evaluated as follows:

$$\hat{\Phi}_n^2 = 90 \left\{ \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \prod_{i=1}^{2} (1 - \max\{\hat{U}_{ij}, \hat{U}_{ik}\}) - \frac{1}{2n} \sum_{j=1}^{n} \prod_{i=1}^{2} (1 - \hat{U}_{ij}^2) + \left(\frac{1}{3}\right)^2 \right\}. \qquad (4.57)$$

The asymptotic distribution of $\hat{\Phi}_n^2$ can be deduced from the asymptotic behavior of the empirical copula process which, for instance, has been discussed by Rüschendorf (1976), Gänßler and Stute (1987), Van der Vaart and Wellner (1996), and Tsukahara (2005).

The quantity $\Phi_{X,Y}^2$ was introduced by Blum *et al.* (1961) without the normalizing factor 90, as a distribution-free statistic to test for the independence of X and Y.

Referring to Schweizer and Wolff (1981), Nelsen (2006, p. 210) states that "... any L_p distance should yield a symmetric nonparametric measure of dependence". For any $p, 1 < p < \infty$, the L_p distance between the copula C and the product copula Π is given by the following expression:

$$\left(k_p \iint_{\mathbb{I}^2} |C(u,v) - uv|^p \mathrm{d}u\mathrm{d}v \right)^{\frac{1}{p}}, \qquad (4.58)$$

where k_p is the normalizing factor. On letting $p = 2$, one obtains $\Phi_{X,Y}$.

4.6 Illustrative Examples

In order to compare the measures of association discussed in the previous sections, five two-dimensional data sets exhibiting different patterns are considered.

First, 500 random values of x, denoted by $\mathcal{S}$, were generated within the interval $(-3, 3)$. Now, letting

- $f_A(x) = -x/5 + 1 + \epsilon$,
- $f_B(x) = -x^5 + \epsilon$,
- $f_C(x) = \sin(x) + \epsilon$,
- $f_D(x) = -\sqrt{|x^{3/2}|} + \epsilon$,

• $f_E = \tan(x)^3 + \epsilon$,

where ϵ represents a slight perturbation consisting of a multiple of random values gener-

ated from a uniform distribution on the interval $[-1, 1]$, the five resulting data sets are $A = \{(x, f_A(x))|x \in S\}$, $B = \{(x, f_B(x))|x \in S\}$, $C = \{(\cos(x), f_C(x))|x \in S\}$, $D = \{(x, f_D(x))|x \in S\}$, and $E = \{(x, f_E(x))|x \in S\}$. These data sets are plotted in Figures 4.1-4.5.
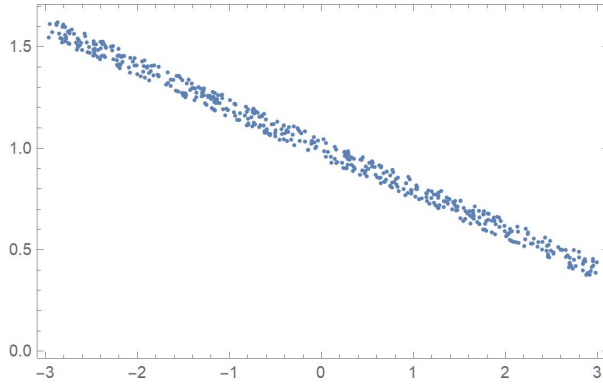
Figure 4.1: Plot of data set A
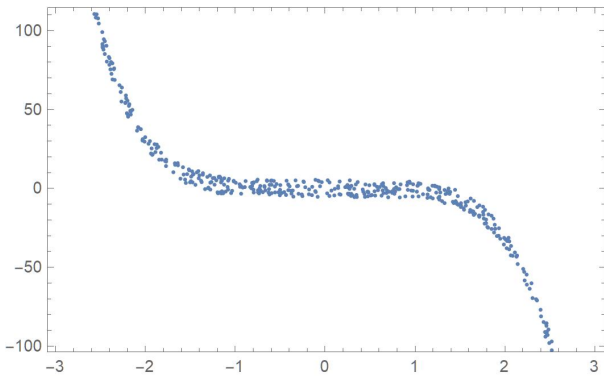
Figure 4.2: Plot of data set B
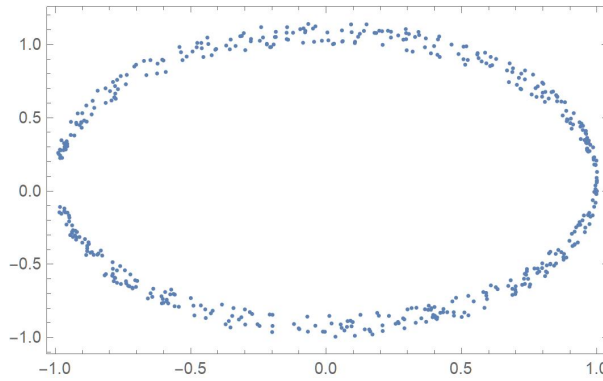
Figure 4.3: Plot of data set C
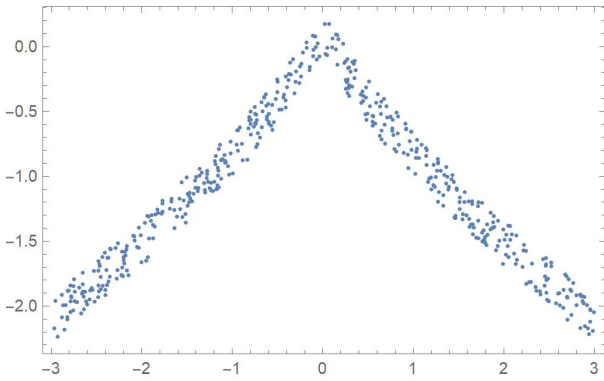
Figure 4.4: Plot of data set D

We then evaluated Spearman's, Kendall's, Blomqvist's and Hoeffding's statistics, as well as Pearson's correlation coefficient for each data set, and reported their numerical values as well as their associated p-values in Table 4.1.

Hoeffding's statistic strongly rejects the null hypothesis of independence since the p-values are all virtually equal to zero, correctly indicating that in all five cases there exist some functional relationship between the variables.
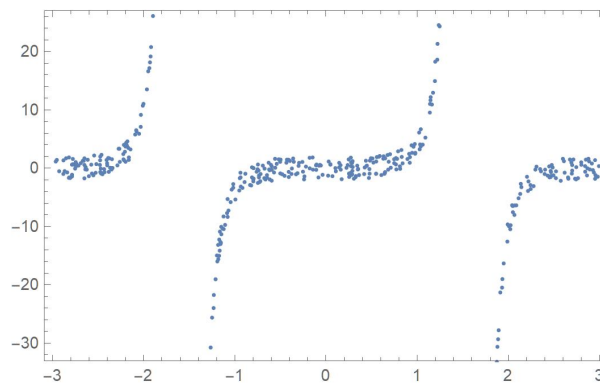
Figure 4.5: Plot of data set *E*

Statistics and *p*-values	A	B	C	D	E
Spearman	{−0.9963, 0}	{−0.9218, 0}	{0.0022, 0.9602}	{0.1028, 0.0215}	{−0.0745, 0.0961}
Kendall	{−0.9456, 0}	{−0.8072, 0}	{0.0071, 0.8136}	{0.0919, 0.0021}	{−0.0350, 0.2419}
Blomqvist	{−0.9600, 0}	{−0.6160, 0}	{0.0320, 0.4209}	{0.0720, 0.0891}	{0.0640, 0.1283}
Hoeffding	{0.8679, 0}	{0.5302, 0}	{0.0472, 0}	{0.1902, 0}	{0.0104, 0}
Pearson	{−0.9964, 0}	{−0.8207, 0}	{0.0202, 0.6529}	{0.0555, 0.2152}	{−0.0092, 0.8377}

Table 4.1: Five statistics and associated *p*-values

As expected, Pearson's correlation coefficient is higher in absolute value in the case of linear relationship (data set *A*) with a value of −0.9964 than in the case of a monotonic relationship (data set *B*) with a value of −0.8207.

Spearman's, Kendall's and Blomqvist's statistics readily detect the monotonic relationships that data sets *A* and *B* exhibit. Interestingly, at the 5% significance level, Spearman's and Kendall's statistics can reject the independence assumption in favor of the relationship occurring between the variables in data set *D*, which happens to be monotonically increasing and then, decreasing.

Bibliography

[1] Block, H. W. and Fang, Z. (1988). A multivariate extension of Hoeffding's lemma. *The Annals of Probability*, 16(4), 1803–1820.

[2] Blomqvist, N. (1950). On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, 21, 593–600.

[3] Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, 32(2), 485–498.

[4] Cherubini, U., Fabio, G., Mulinacci, S. and Romagno, S. (2012). *Dynamic Copula Methods in Finance*. John Wiley & Sons, New York.

[5] Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley & Sons, New York.

[6] Denuit, M., Daehe, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons, New York.

[7] Dias, A. (2022). Maximum pseudo-likelihood estimation in copula models for small weakly dependent samples. arXiv:2208.01322 [stats.ME].

[8] Gaißer, S., Ruppert, M. and Schmid, F. (2010). A multivariate version of Hoeffding's phi-square. *Journal of Multivariate Analysis*, 101(10), 2571–2586.

[9] Gänssler, P. and Stute, W. (1987). *Seminar on Empirical Processes.* DMV Seminar 9.

[10] Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distribution. *Biometrika*, 82(3), 543–552.

[11] Genest, C., Carabarín-Aguirre, A. and Harvey, F. (2013). Copula parameter estimation using Blomqvist's beta. *Journal de la Société Française de Statistique*, 154(1), 5–24.

[12] Gibbons, J. D. and Chakraborti, S. (2003). *Nonparametric Statistical Inference*, Fourth Edition, revised and expanded. Statistics Textbooks and Monographs, 168. Marcel Dekker, Inc., New York.

[13] Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5, 181--233. (Reprinted as Scale-Invariant correlation theory, in *The Collected Works of Wassily Hoeffding*, Editors Fisher, N. I., Sen, P. K., Springer, New York, 57–107, 2012.)

[14] Hofert, M., Kojadinovic, I., Mächler, M. and Yan, J. (2019). *Elements of Copula Modeling with R.* Springer, New York.

[15] Joe H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, Florida.

[16] Jogdeo, K. (1968). Characterizations of independence in certain families of bivariate and multivariate distributions. *The Annals of Mathematical Statistics*, 39(2), 433–441.

[17] Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81–93.

[18] Kendall, M. G. and Gibbons, J. D. (1990). *Rank Correlation Methods: Fifth Edition.* Griffin, London, UK.

[19] Kerman, J. (2011). A closed-form approximation for the median of the beta distribution. arXiv:1111.0433 [math.ST].

[20] Kojadinovic, I. and Yan, J. (2010). Comparison of three semiparametric methods for esti-
     mating dependence parameters in copula models. *Mathematics and Economics*, 47(1), 52–
     63.

[21] Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statis-
     tical Association*, 53(284), 814–861.

[22] Lai, C. D. and Balakrishnan, N. (2009). *Continuous Bivariate Distributions: Second Edi-
     tion.* Springer, New York.

[23] Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical
     Statistics*, 37(5), 1137–1153.

[24] Mosteller, F. (1946). On some useful "inefficient" statistics. *The Annals of Mathematical
     Statistics*, 17(4),377–408.

[25] Nelsen, R. B. (2006). *An Introduction to Copulas, Second Edition.* Springer, New York.

[26] Rényi, A. (1959). On measures of dependence. *Acta Mathematica Hungarica*, 10(3-4),
     441–451.

[27] Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics.
     The Annals of Statistics, 4(5), 912—923.

[28] Schmid, F. and Schmidt, R. (2007). Multivariate conditional versions of Spearman's rho
     and related measures of tail dependence. *Journal of Multivariate Analysis*, 98(6), 1123–
     1140.

[29] Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for
     random variables. *The Annals of Statistics*, 9(4), 879–885.

[30] Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of
     Statistics*, 33(3), 357–375.

[31] Valz, P. D. and McLeod, A. I. (1990). A simplified derivation of the variance of Kendall's rank correlation coefficient. *The American Statistician*, 44(1), 39–40.

[32] Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Chapter 5

A Criterion for Characterizing the Tail Behavior of a Distribution

5.1 Introduction

This section provides an overview of certain criteria that have been previously introduced in the statistical literature for the purpose of labeling distributional tail behavior. When selecting a base density function for determining a density estimate or approximant that is based on sample moments, ideally, this function should exhibit a tail behavior that is similar to that of the underlying distribution. Hence, the need to objectively and efficiently identify this behaviour, which is the objective of this chapter.

Fairly recently, Klugman *et al.* (2012) provided several classification categories for identifying light to heavy-tailed distributions, these being based on moments, hazard rate functions and mean excess loss functions. Previously, Parzen (1979) examined the limiting behavior of density quantile functions which can be expressed as

$$f(Q(u)) \sim \begin{cases} (1-u)^{\alpha} & \text{for } \alpha > 0 \text{ and } \alpha \neq 1 \\ (1-u)(\log \frac{1}{1-u})^{1-\beta} & \text{for } \alpha = 1 \text{ and } 0 \leq \beta \leq 1 \end{cases} \tag{5.1}$$

where f and Q represent the density and quantile function, respectively, and $f_1(u) \sim f_2(u)$ denotes that the ratio $f_1(u)/f_2(u)$ converges to a positive finite constant as $u \to 1$. The parameter α determines three types of tail behavior: short tails, medium tails and long tails which correspond to $\alpha < 1$, $\alpha = 1$ and $\alpha > 1$, respectively.

In order to refine the tail classification advocated by Parzen (1979), Schuster (1984) relied on two quantities, namely,

$$\alpha = \lim_{u \to 1^-} -\frac{1-u}{f(Q(u))} \frac{\partial \log[f(Q(u))]}{\partial u} \tag{5.2}$$

and

$$c = \lim_{u \to 1^-} (1-u)/f(Q(u)) = \lim_{u \to 1^-} 1/h(Q(u)), \tag{5.3}$$

where f, Q, and h represent the density, quantile and hazard function, respectively, to obtain five categories of tail behavior:

$$
\begin{array}{lll}
\text{Short} & 0 < \alpha < 1 & \\
\text{Medium-Short} & \alpha = 1 & c = 0 \\
\text{Medium-Medium} & \alpha = 1 & 0 < c < \infty \\
\text{Medium-Long} & \alpha = 1 & c = \infty \\
\text{Long} & \alpha > 1. &
\end{array}
$$

Actually, this criterion has a theoretical connection with the limiting size of extreme spacings.

The reader may also refer to Rojo (1996), whose classification is based on the residual lifetime distributions. The aforementioned classification techniques are reviewed in Rojo and Ott (2010).

As explained in Heyde and Kou (2004), there exists a variety of methodologies for determining whether a distribution has an exponential or a power tail, including QQ-plots, likelihood methods, and plots of the mean residual life functions.

As well, Loh (1984) and Doksum (1969), among others, proposed criteria that take into account the behaviour of a sizeable portion of the distribution at hand. An efficiency-based tail

ordering technique was introduced by Lehmann (1988) in the context of location experiments. Another approach consists in relating the limiting distribution of the standardized maximum of a given distribution (when it exists) to that of the Weibull, Gumbel or Fréchet distributions (the three extremal distributions), which leads to categorizing the distribution as having a short, medium or long tail, respectively. It is manifestly desirable to develop procedures that will yield additional categories with a view to identifying more precisely the tail behavior of a wide array of distributions.

The conceptually simple technique being advocated in this chapter results in eight categories. It is applied to numerous theoretical distributions, the resulting tail behaviours being generally consistent with those determined by making use of other criteria. In the case of a sample of observations, one must initially obtain a density estimate to which the proposed approach can then be readily applied. Of course, the larger the sample, the more reliable the results, which is corroborated by the small-scale simulation study presented in Section 5.3. The proposed methodology, which is described in Section 5.2, is applied to an array of widely used distributions in Section 5.4. The proposed methodology is compared with three other criteria with respect to various distributions in Section 5.5.

5.2   A methodology based on the arctangent transformation

We are proposing to make use of the percentiles of a transformed distribution to define a criterion for characterising the tail behaviour of a given distribution. More specifically, on letting X represent a distribution having finite mean μ, finite variance σ^2 and probability density function (PDF) $f(x)$, the standardized random variable $Y = (X - \mu)/\sigma$ is mapped onto $(-1, 1)$ or a subset thereof via the transformation $Z = (2/\pi) \arctan(Y)$. The density functions of certain distributions that have been so transformed are plotted in Figures 2.1 to 2.14. These distributions include the normal, Weibull with shape parameters 0.5 and 2, logistic, extreme value, exponential, Student t on 3, 5 and 20 degrees of freedom (df), lognormal, Uniform(0,1), Beta(5,2),

Type-II Beta(5,3) and (50,30), and Gamma(50,1). The dots indicate the 90th and 99.9999^{th} percentiles of Z.

Let $q(\alpha)$ represent the $(100 \times \alpha)^{th}$ quantile of the distribution of Z. We propose to employ the difference between the 90th and 99.9999^{th} percentiles of Z as a criterion for classifying the right-tail behavior of X. We denote this *tail index* by

$$p = q(0.999999) - q(0.90). \tag{5.4}$$
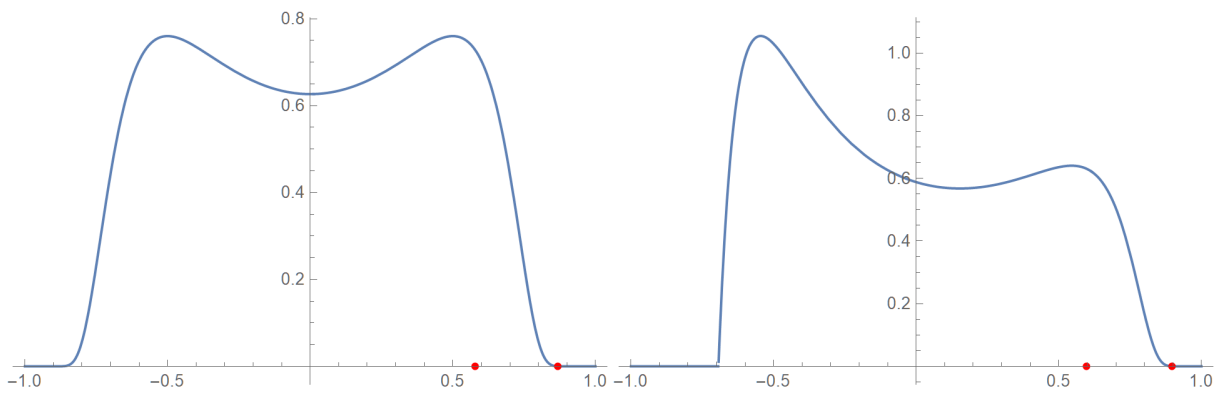
Figure 5.1:
PDF of Z for the normal distribution

Figure 5.2:
PDF of Z for the Weibull distribution (k=2)
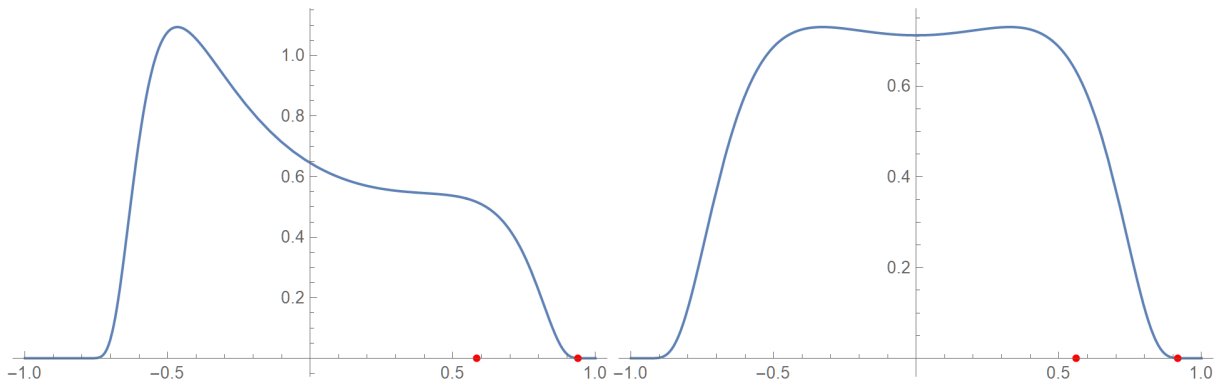
Figure 5.3:
PDF of Z for the extreme value distribution

Figure 5.4:
PDF of Z for the logistic distribution
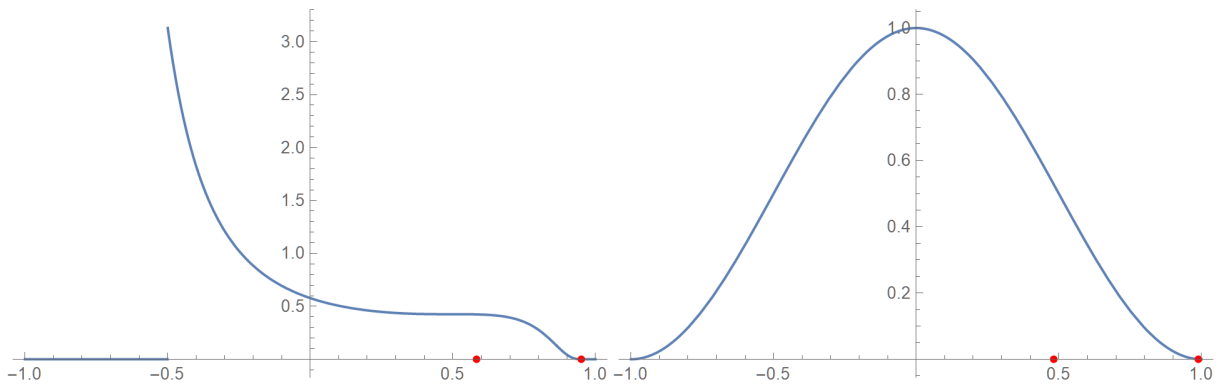
Figure 5.5:
PDF of Z for the exponential distribution

Figure 5.6:
PDF of Z for the t distribution on 3 df
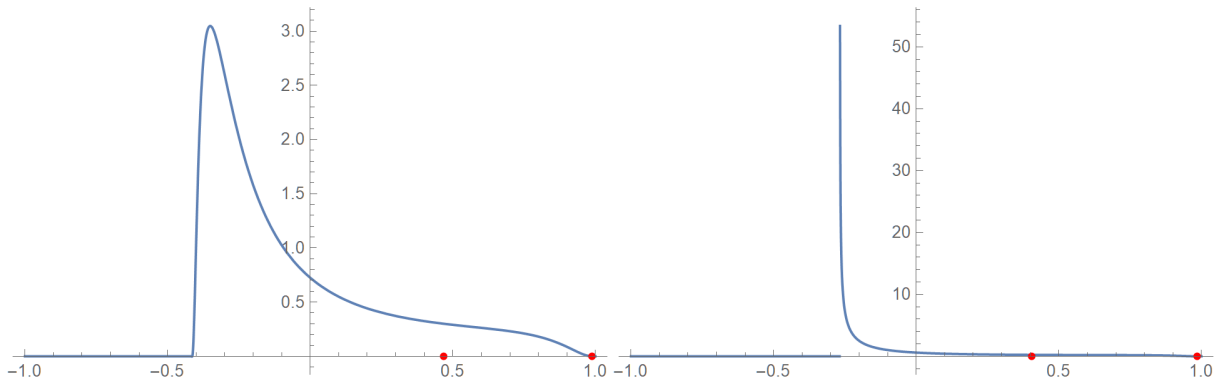
Figure 5.7:
PDF of Z for the lognormal distribution

Figure 5.8:
PDF of Z for the Weibull distribution (k=0.5)
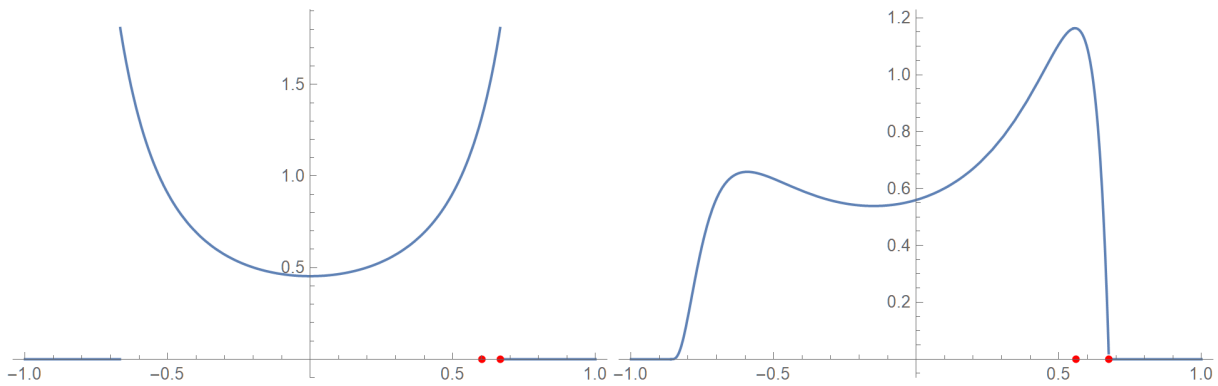
Figure 5.9:
PDF of Z for the Uniform(0,1) distribution

Figure 5.10:
PDF of Z for the Beta(5,2) distribution
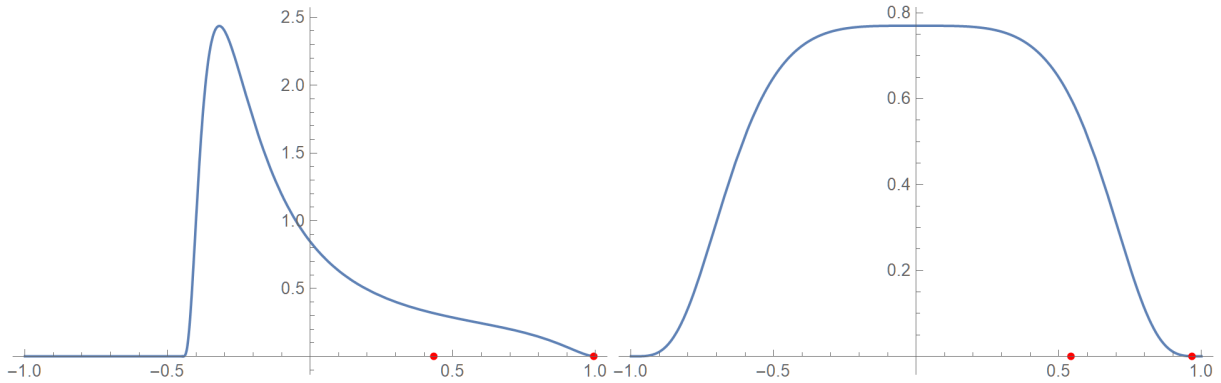
Figure 5.11:
PDF of Z for the type-II Beta(5,3) distribution

Figure 5.12:
PDF of Z for the t distribution on 5 df
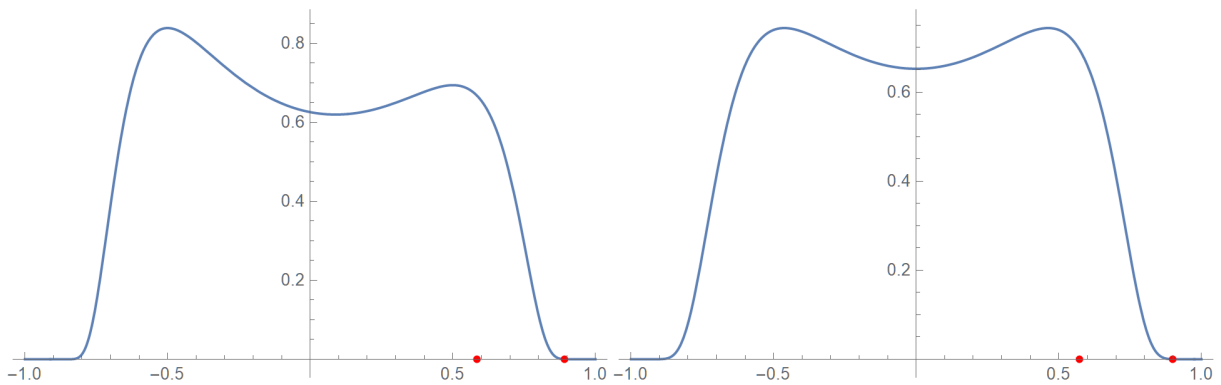
Figure 5.13:
PDF of Z for the Gamma(50,1) distribution

Figure 5.14:
PDF of Z for the t distribution on 20 df
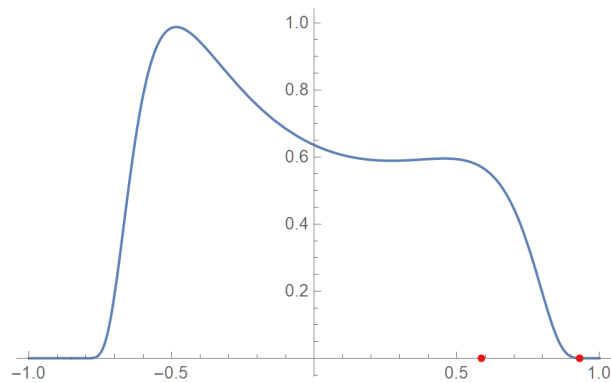
Figure 5.15:
PDF of Z for the type-II Beta(50,30) distribution

We define distributions whose mean is finite but whose variance is infinite as having a long tail and distributions whose mean is undefined as having an extremely long tail. Generally, the fewer the number of finite moments a distribution possesses, the heavier its tail is. The proposed tail behavior categories and their associated tail index ranges are:

- Distinctly Short : $p < 0.1$

- Short : $0.1 \leq p < 0.2$

- Nearly Medium : $0.2 \leq p < 0.3$

- Medium : $0.3 \leq p < 0.4$

- Extended Medium : $0.4 \leq p < 0.5$

- Relatively long : $0.5 \leq p$

- Long :   Indefinite second moment

- Extremely Long :   Indefinite first moment.

The left-tail behavior of a distribution is similarly characterized by defining the correspond-

ing tail index as

$$p^* = q(0.10) - q(0.000001). \tag{5.5}$$

The specified ranges for p also apply to p^*.

When this methodology is implemented, a distribution is deemed to be heavy-tailed if it belongs to one of the following categories: Extended Medium ($0.4 \le p < 0.5$), Relatively Long ($p \ge 0.5$), Long or Extremely Long. For instance, the lognormal, Weibull(k) with $k < 1$, Pareto, Student t with fewer than 6 degrees of freedom and Cauchy are such distributions. Heavy-Tailed distributions belong to the medium-long and long tail categories in Schuster's classification.

5.3   An illustration of the convergence of the tail index

We generated samples of sizes 100, 1000, 50000 and 1000000 from the standard exponential distribution and obtained the following tail indices from their kde's: 0.300144, 0.309942, 0.342787, 0.361315. We observe that these values indeed tend to the theoretical p-index for the standard exponential distribution, which is 0.3672. The corresponding density functions of Z are plotted in Figs. 5.16-5.19.
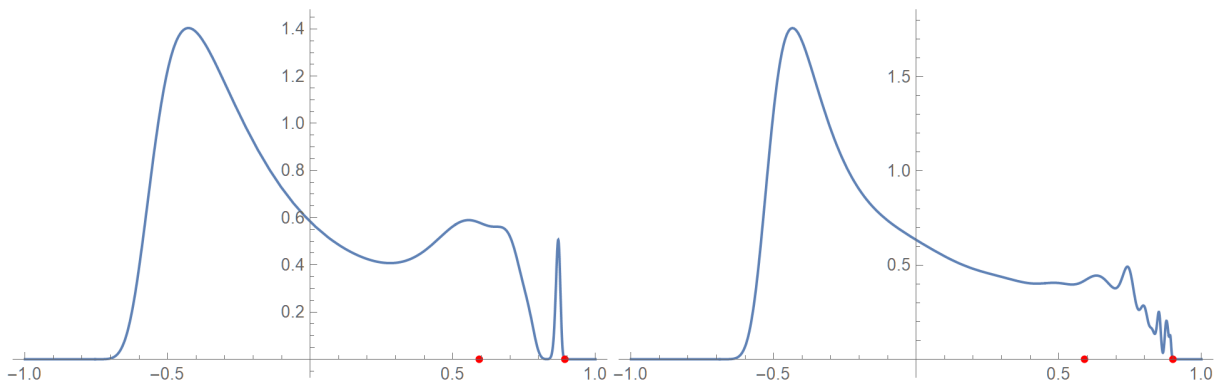
Figure 5.16:
PDF of Z, 100 Exp(1) sample points

Figure 5.17:
PDF of Z, 1,000 Exp(1) sample points
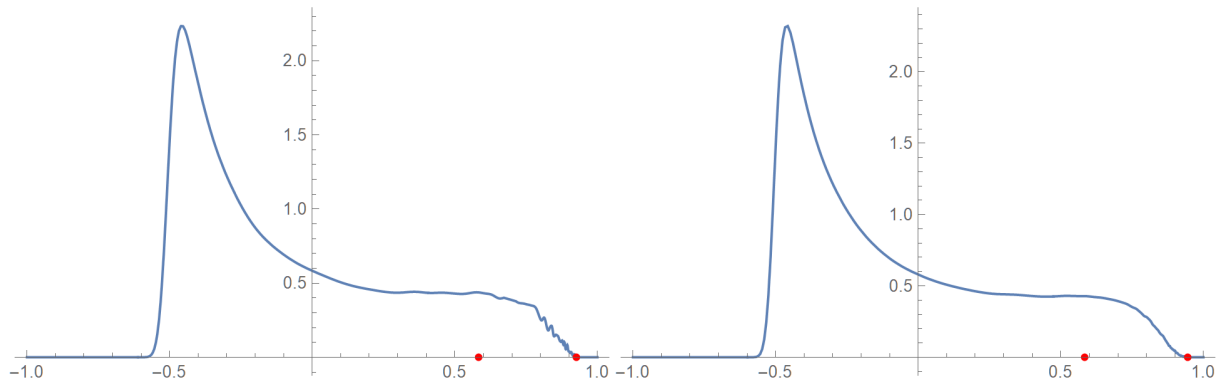
Figure 5.18:
PDF of Z, 50,000 Exp(1) sample points

Figure 5.19:
PDF of Z, 1,000,000 Exp(1) sample points

5.4 Application of the tail index criterion

Some illustrative classification results obtained by applying of the proposed tail index criterion
are presented in Table 5.1 for certain commonly utilized distributions.

Table 5.1: The tail behavior of certain distributions

Distribution	Tail Behavior	Tail Index
Uniform	Distinctly Short	0.0646
Beta(5,2)	Short	0.1152
Normal	Nearly Medium	0.2898
Rayleigh	Nearly Medium	0.2998
Type-II Beta(50,30)	Medium	0.3444
Extreme value	Medium	0.3549
Logistic	Medium	0.3562
Exponential	Medium	0.3672
Student t on 5 df	Extended Medium	0.4244
Type-II Beta(2,5)	Extended Medium	0.4591
Student t on 3 df	Relatively Long	0.5071
Lognormal	Relatively Long	0.5201
Type-II Beta(5,3)	Relatively Long	0.5609
Weibull(0.5)	Relatively Long	0.5800
Student t on 2 df	Long	**
Cauchy	Extremely Long	**

** Moment-based.

It should be noted that for the beta, Weibull, Type-II Beta and lognormal distributions, the
shapes of the standardized densities and thus the tail behavior and the associated value of the

tail index vary with the parameters.

5.5 Comparison with other criteria

In this section, we compare the tail classification criterion introduced in Section 5.2 with three other criteria with respect to various distributions. The results are summarized in Table 5.2. It is observed that, overall, the proposed criterion and three other criteria yield similar classification results.

Table 5.2: Comparative classification of tail behavior for certain distributions

Distribution	Parzen	Schuster	Rojo	Our category	p
Uniform	Short	Short	Super-Short	Distinctly Short	0.0646
Beta(5, 2)	Short	Short	Super-Short	Short	0.1152
Normal	Medium	Medium-Short	Weakly-Short	Nearly Medium	0.2898
Weibull ($k = 2$)	Medium	Medium-Short	Weakly-Short	Nearly Medium	0.2998
Extreme value	Medium	Medium-Short	Moderately-Short	Medium	0.3549
Logistic	Medium	Medium-Medium	Medium	Medium	0.3562
Exponential	Medium	Medium-Medium	Medium	Medium	0.3672
Weibull ($k = 1$)	Medium	Medium-Medium	Medium	Medium	0.3672
Lognormal(0, 1)	Medium	Medium-Long	Weakly-Long	Relatively Long	0.5201
Weibull ($k = 0.5$)	Medium	Medium-Long	Weakly-Long	Relatively Long	0.5800
Pareto ($k = 3$)	Long	Long	Weakly-Long	Relatively Long	0.5824
Pareto ($k = 2$)	Long	Long	Weakly-Long	Long	**
Pareto ($k = 1$)	Long	Long	Moderately-Long	Super-Long	**
Pareto ($k = 0.5$)	Long	Long	Super-Long	Super-Long	**
Cauchy	Long	Long	Weakly-Long	Super-Long	**

** Moment-based.

5.6 Conclusion

While being readily implementable, the proposed approach to characterising distributional tail behaviour produces easily identifiable categories that, generally, prove consistent with those obtained by making use of other criteria.

Bibliography

[1] Doksum, K. (1969). Starshaped transformations and the power of rank tests. *The Annals of Mathematical Statistics*, 40, 1167–1176.

[2] Heyde, C. C. and Kou, S. G. (2004). On the controversy over tailweight of distributions. *Operations Research Letters*, 32, 399–408.

[3] Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2012). *Loss Models: from Data to Decisions*, 4th Edition. John Wiley & Sons, New York.

[4] Lehmann, E. (1988). Comparing location experiments. *The Annals of Statistics*, 16, 521–533.

[5] Loh, W.-Y. (1984). Bounds on AREs for restricted classes of distributions defined via tail-orderings. *Annals of Statistics*, 12, 685–701.

[6] Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, 74, 105–121.

[7] Rojo, J. (1996). On tail categorization of probability laws. *Journal of the American Statistical Association*, 91, 378–384.

[8] Rojo, J. and Ott, R. C. (2010). Testing for tail behavior using extreme spacings. *Mathematics - Statistics Theory*. arXiv:1011.6458 [math.ST], 25 pages.

[9] Schuster, E. F. (1984). Classification of probability laws by tail behavior. *Journal of the American Statistical Association*, 79, 936–939.

Chapter 6

Distribution of Quadratic Forms in Various Types of Random Variables

6.1   Introduction

Numerous distributional results are available in connection with quadratic forms in normal random variables. Several representations of the density function of a quadratic form have been derived and a number procedures have been introduced for computing percentage points. Among others, Gurland (1953), Box (1954), Pachares (1955), Ruben (1960, 1962), Shah and Khatri (1961), and Kotz *et al*. (1967a,b), obtained various series representations. Gurland (1956) and Shah (1963) respectively considered central and noncentral indefinite quadratic forms. Exact distributional results were obtained by Imhof (1961), Davis (1973) and Rice (1980).

As mentioned in Mathai and Provost (1992), a wide array of test statistics can be expressed in terms of quadratic forms in normal random vectors. For example, one may consider the lagged regression residuals developed by De Gooijer and MacNeill (1999) and discussed in Provost (2005), or certain change point test statistics derived by MacNeill (1978).

Since the case of quadratic forms in real Gaussian variables has been discussed at length

in Mathai and Provost (1992), we will initially focus on approximating the density functions of quadratic forms in other types of random variables. It will be explained in the next section that moments of quadratic forms can be obtained symbolically, which enables one to consider non-Gaussian types of distributions. Additionally, the variables comprising the vector of a given quadratic form need not be identically distributed when making use of this innovative approach.

More specifically, we will approximate the distribution of quadratic forms in gamma, inverse Gaussian, binomial and Poisson random variables that are not necessarily identically distributed. Some distributional limit theorems such as those that are discussed in Barrio *et al.* (2005) in connection with a certain empirical quantile process, involve quadratic forms in exponential random variables which are a particular case of gamma random variables. As well, three test statistics that can be expressed as quadratic forms in exponential random variables, are described in Donald and Paarsch (2002). Whittle (1960) considered the case of random variables in Poisson variables in connection with contingency tables. In generalized linear models, the conditional distribution of the response variable is often taken to be a member of an exponential family, such as the Gaussian, binomial, Poisson, gamma, or inverse Gaussian families of distributions, and the determination of the distribution of quadratic forms in such random variables could contribute to further developments in data modeling.

The accuracy of initial approximations will be improved by applying a technique introduced in Provost (2005), which consists of adjusting an appropriate base density by a polynomial whose coefficients are determined from the moments of the quadratic forms and those of the base density. As explained in Provost *et al.* (2009), this approach which is formally stated in the following result, can also be utilized to approximate discrete distributions.

Result 3 *Let $f_Y(y)$ be the density function of a continuous random variable Y defined in the interval (a, b), $E(Y^j) \equiv \mu_Y(j)$, $X = (Y - u)/s$ be an affine transformation (oftentimes, $u = E(Y)$ and $s = \sqrt{Var(Y)}$), where $u \in \mathbb{R}$ and $s \in \mathbb{R}^+$, $a_0 = (a-u)/s$, $b_0 = (b-u)/s$, $f_X(x) = s f_Y(u+s x)$ denote the density function of X whose support is the interval (a_0, b_0), $E(X^j) = E[((Y-u)/s)^j] \equiv$*

$\mu_X(j)$, and let the base density function $\psi_X(x) \equiv c_T w(x)$, where c_T is a positive normalizing constant, be an initial density approximation to $f_X(x)$ with $\int_{a_0}^{b_0} x^j \psi_X(x) \, dx \equiv m_X(j)$. Assuming that the sequence $\mu_X(i), i = 0, 1, 2, \ldots$, uniquely defines the distribution of X, that $m_X(j)$ exists for $j = 0, 1, \ldots, 2n$, and that whenever $\psi_X(x)$ is nontrivial function of x, its tail behavior is congruent to that of $f_X(x)$, the density function of X can be approximated by

$$f_{X_n}(x) = \psi_X(x) \sum_{\ell=0}^{n} \xi_\ell \, x^\ell \qquad (6.1)$$

with $(\xi_0, \ldots, \xi_n)' = \mathcal{M}^{-1}(\mu_X(0), \ldots, \mu_X(n))'$, where $\mathcal{M}$ is an $(n + 1) \times (n + 1)$ matrix whose $(h + 1)^{th}$ row is $m_X(h), \ldots, m_X(h + n), h = 0, 1, \ldots, n$. The corresponding density approximant for Y is then

$$f_{Y_n}(y) = \psi_X\left(\frac{y - u}{s}\right) \sum_{\ell=0}^{n} \frac{\xi_\ell}{s} \left(\frac{y - u}{s}\right)^\ell. \qquad (6.2)$$

Let $X_1, X_2, \ldots, X_n$ be random variables and A be a $n \times n$ symmetric matrix, where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}.$$

The quadratic form,

$$Y = \mathbf{X}'A\mathbf{X},$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ can be expanded as follows:

$$Y = a_{11}X_1^2 + a_{22}X_2^2 + \cdots + a_{nn}X_n^2 + 2a_{12}X_1X_2 + 2a_{13}X_1X_3 + \cdots + 2a_{n-1,n}X_{n-1}X_n$$

$$= \sum_{i=1}^{n} a_{ii}X_{ii}^2 + \sum_{1 \leq i < j \leq n} 2a_{ij}X_iX_j.$$

The symbolic approach to the evaluation of the moments of quadratic forms is described in Section 6.2. Several illustrative examples are provided in Section 6.3. In each of the examples, the empirical distribution of a given quadratic form is determined from a simulated sample of size 10,000. The case of Hermitian quadratic forms in complex normal vectors is discussed in Section 6.4 where several applications are pointed out. In the last section, an extended version of the generalized gamma density function is shown to provide accurate approximations to the density functions of indefinite quadratic forms in Gaussian vectors without having to resort to polynomial adjustments.

6.2 Evaluation of the exact moments of quadratic forms via the symbolic approach

Suppose that the random vector $\mathbf{X}$ follows a distribution with density function $f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$. We can make use of the moments associated with $f_{\mathbf{X}}(\cdot)$ to determine the hth moment of the random variable Y. Let $\mu_{\mathbf{X}}(h_1, h_2, \ldots, h_n)$ be the joint moment of order $(h_1, h_2, \ldots, h_n)$ of $X_1, X_2, \ldots, X_n$. The hth moment of Y is then

$$
\begin{aligned}
E[Y^h] &= E\left[\left(\sum_{i=1}^{n} a_{ii}X_{ii}^2 + \sum_{1\le i<j\le n} 2a_{ij}X_iX_j\right)^h\right] \\
&= E\left(\sum_i c_i X_1^{h_{i1}} X_2^{h_{i2}} \cdots X_n^{h_{in}}\right),
\end{aligned}
\tag{6.3}
$$

where the c_i's and h_{ij}'s can be determined by expanding symbolically the expression within the square brackets. Moreover, in light of the linearity property of mathematical expectations, we have

$$
\begin{aligned}
E[Y^h] &= \sum_i E[c_i X_1^{h_{i1}} X_2^{h_{i2}} \cdots X_n^{h_{in}}] \\
&= \sum_i c_i \, \mu_X(h_{i1}, h_{i2}, \ldots, h_{in}).
\end{aligned}
\tag{6.4}
$$

Thus, we can evaluate the moments of Y from the joint moments of $\mathbf{X}$ without having to determine the exact distribution of Y. The proposed symbolic approach for determining the moments of quadratic forms applies to any type of quadratic forms and can be readily implemented. Besides, when moment expressions are available, they are usually quite complicated.

For example, letting $\mathbf{X} = (X_1, X_2)$ follow a bivariate normal distribution with mean $(-1, 2)'$ and covariance matrix

$$\begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$$

and

$$Y = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},$$

the second moment of Y is obtained as follows:

$$\begin{aligned} E[Y^2] &= E\left[\left(\begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right)^2\right] \\ &= E[X_1^4 + 16X_1^3 X_2 + 60X_1^2 X_2^2 - 32X_1 X_2^3 + 4X_2^4] \\ &= E[X_1^4] + 16E[X_1^3 X_2] + 60E[X_1^2 X_2^2] - 32E[X_1 X_2^3] + 4E[X_2^4] \\ &= \mu_{\mathbf{X}}(4, 0) + 16\mu_{\mathbf{X}}(3, 1) + 60\mu_{\mathbf{X}}(2, 2) - 32\mu_{\mathbf{X}}(1, 3) + 4\mu_{\mathbf{X}}(0, 2) \\ &= 3481, \end{aligned}$$

where $\mu_{\mathbf{X}}(i, j)$ denotes the joint moment of orders i and j of the distribution of $\mathbf{X}$.

One could also evaluate the second moment of Y by integration:

$$E[Y^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)^2 f_{\mathbf{X}}(x_1, x_2) \, dx_1 \, dx_2. \tag{6.5}$$

However, the proposed approach proves more efficient, especially when n is large.

The moments of a quadratic forms in Gaussian variables can also be evaluated from math-

ematical representations. As shown in Mathai and Provost (1992), the sth cumulant of $\mathbf{X}'A\mathbf{X}$ where $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, which denotes a p-variate normal distribution whose mean is $\boldsymbol{\mu}$ and nonsingular covariance matrix is Σ, is given by

$$
\begin{aligned}
k(s) &= 2^{s-1} s! \left(\text{tr}(A\,\Sigma)^s / s + \boldsymbol{\mu}'(A\Sigma)^{s-1} A\boldsymbol{\mu} \right) \\
&= 2^{s-1}(s-1)!\, \theta_s \,,
\end{aligned}
\tag{6.6}
$$

where $\text{tr}(\cdot)$ denotes the trace of $(\cdot)$ and $\theta_s = \sum_{j=1}^{p} \lambda_j^s (1 + s\, b_j^2)$, $s = 1, 2, \ldots$, the λ_j's, $j = 1, \ldots, p$, denoting the eigenvalues of $A\Sigma$, and $\text{tr}(A\Sigma)^s = \sum_{j=1}^{p} \lambda_j^s$. The hth moment of $\mathbf{X}'A\mathbf{X}$ can be obtained from its cumulants by means of the following recursive relationship, derived for instance in Smith (1995):

$$
\mu(h) = \sum_{i=0}^{h-1} \frac{(h-1)!}{(h-1-i)!\, i!}\, k(h-i)\, \mu(i),
\tag{6.7}
$$

where $k(s)$ is as given in Equation (6.6). Note that the expressions for the cumulants become more complex in the case of quadratic forms in singular Gaussian vectors.

6.3   Illustrative Examples

In this section, nearly exact density functions are obtained for quadratic forms in four types of random variables.

6.3.1   Quadratic forms in gamma random variables

First, consider the case where the matrix of the quadratic form is

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

and the component of the vector $\mathbf{X}$, i.e., X_1, X_2, X_3, and X_4 are independently distributed gamma random variables with respective parameters (2,2), (9,1), (2,1), (12,1), the density function of a gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ being given by

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0.$$

The pdf's of X_1, X_2, X_3, and X_4 are plotted in Figure 6.1.
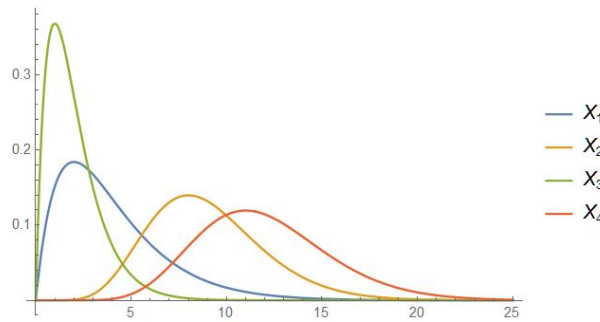
Figure 6.1: pdf's of the X_i's

An approximant to the the distribution of $Y = (X_1, X_2, X_3, X_4)A(X_1, X_2, X_3, X_4)'$ is being sought. Since the exact probability density or cumulative distribution functions of such a quadratic form are not known, we resort to a histogram of a simulated sample (with 10,000 replications) and the resulting empirical cumulative distribution for assessing the accuracy of the approximants. We employ the integrated squared difference (ISD) between the empirical cumulative distribution function and the approximated cumulative distribution function to determine the degree of the polynomial adjustment. More specifically, we select the degree beyond which the ISD does not decrease significantly or starts to increase. This criterion will be utilized for all the quadratic forms in continuous r.v.'s considered in this section.

We make use of a gamma pdf as base density and determine its parameters by setting the first two moments of the gamma distribution equal to those of the quadratic form which are obtained from the symbolic approach. This base pdf and the corresponding cdf are respectively shown in Figure 6.2 and 6.3. A polynomial adjustment of degree 8 (obtained by applying Result 3 as stated in the Introduction of this chapter) is made in this case. The resulting pdf and cdf approximants are respectively plotted in Figures 6.4 and 6.5. In this case as well as in all the other cases presented in this chapter, it is seen that the empirical and approximated cdf's are nearly identical.
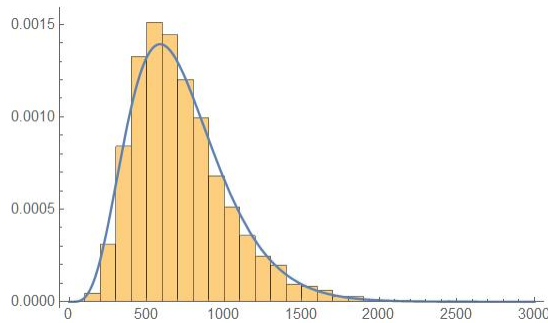
Figure 6.2: Histogram and base pdf
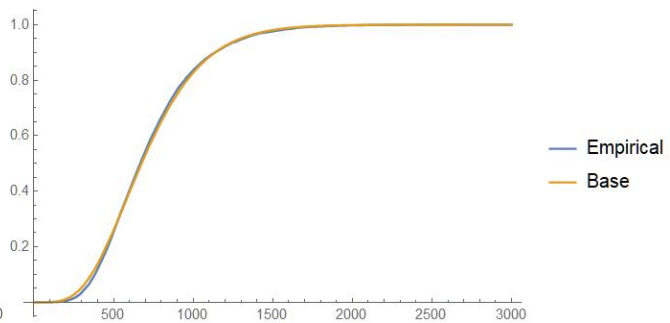
Figure 6.3: Empirical and base cdf's
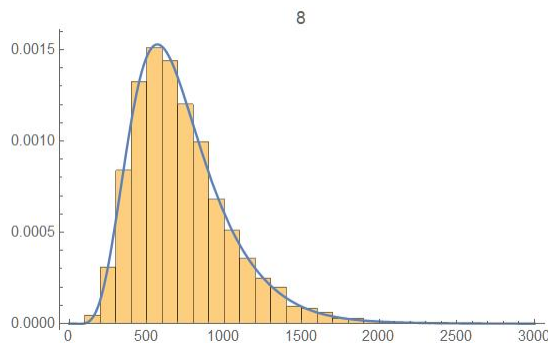
Figure 6.4: Histogram and pdf approximant
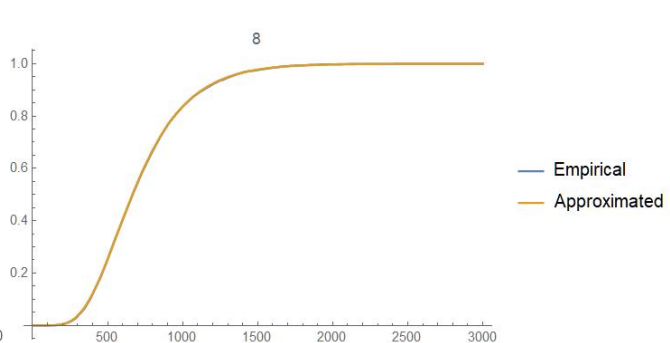
Figure 6.5: Empirical and approximated cdf's

6.3.2   Quadratic forms in inverse Gaussian random variables

Consider the case where the matrix of the quadratic form is

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

and the component of the vector $\mathbf{X}$, i.e., X_1, X_2, X_3 and X_4 are independently distributed inverse Gaussian random variables with respective parameters (2,5), (3,6), (2,2), (3,4), the density function of an inverse Gaussian distribution with parameters (λ, μ) being

$$f(x) = \left(\frac{\lambda}{2\pi x^3} \right)^{1/2} e^{-\lambda(x-\mu)^2/(2\mu^2 x)}, \quad x > 0, \ \lambda > 0, \ \mu \in \mathbb{R}.$$

The pdf's of X_1, X_2, X_3 and X_4 are plotted in Figure 6.6.
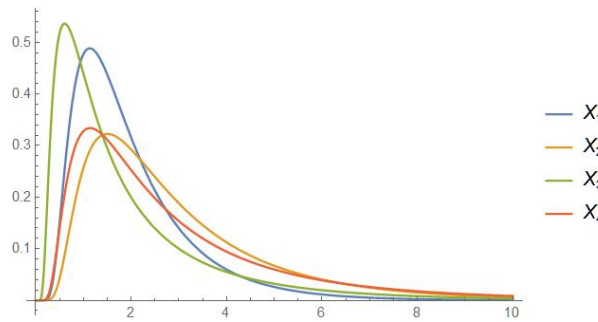
Figure 6.6: pdf's of the X_i's

An approximant to the distribution of $Y = (X_1, X_2, X_3, X_4)A(X_1, X_2, X_3, X_4)'$ is being sought. Proceeding as explained in Subsection 6.3.1, a random sample was initially generated and the resulting empirical cumulative distribution function was determined.

In this case, we make use of a inverse Gaussian density as base density and estimate its parameters by setting the first two moments of the inverse Gaussian distribution equal to those of the quadratic form which are obtained from the symbolic approach. This base pdf and the corresponding cdf are respectively shown in Figure 6.7 and 6.8. A polynomial adjustment of degree 5 (determined by applying Result 3 and calculating the ISD) was deemed appropriate in this case. The resulting pdf and cdf approximants are respectively plotted in Figures 6.9 and

6.10. It is seen that the empirical and approximated cdf's are in close agreement.
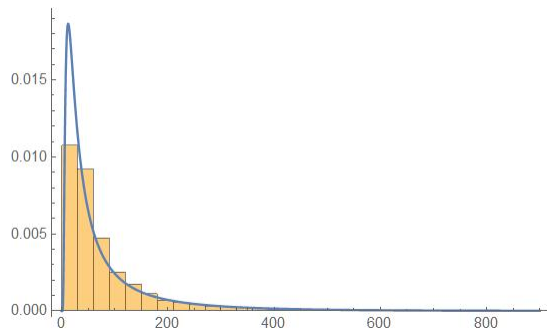
Figure 6.7: Histogram and base pdf
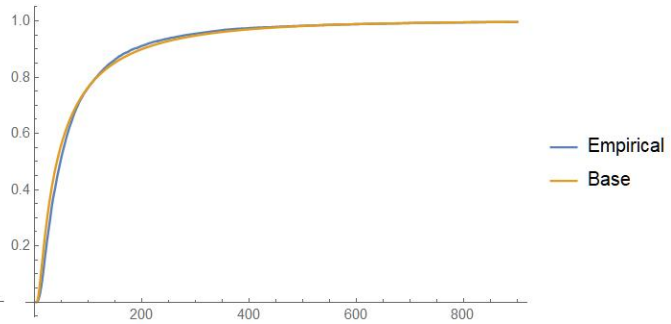
Figure 6.8: Empirical and base cdf's
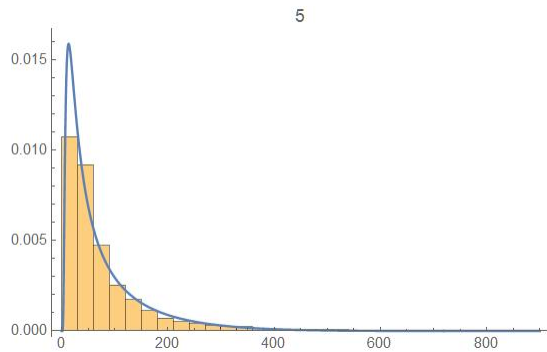
Figure 6.9: Histogram and pdf approximant
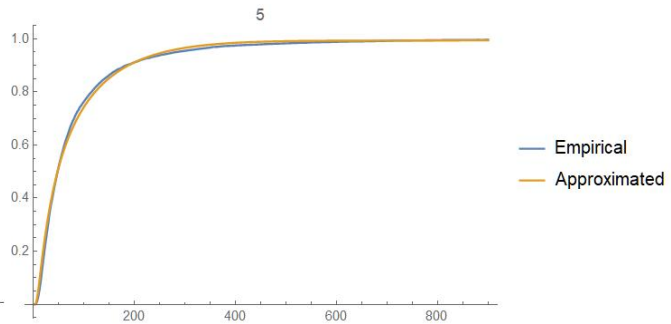
Figure 6.10: Empirical and approximated cdf's

As the following subsections illustrate, the methodology also applies in the case of discrete random variables.

6.3.3   Quadratic forms in binomial random variables

Consider the case where the matrix of the quadratic form is

$$
A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}
$$

and the component of the vector $\mathbf{X}$, i.e., X_1, X_2, X_3 and X_4 are independently distributed binomial random variables with respective parameters (2,2), (9,1), (2,1), (12,1), the probability mass function of a binomial distribution with parameters (n, p) being

$$p(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

The cdf's of X_1, X_2, X_3 and X_4 are plotted in Figure 6.11.
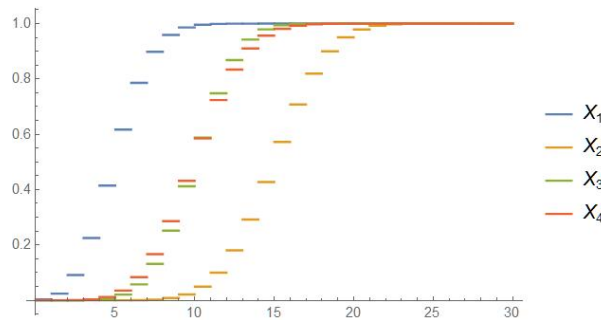
Figure 6.11: cdf's of the X_i's

An approximant to the distribution of $Y = (X_1, X_2, X_3, X_4) A (X_1, X_2, X_3, X_4)'$ is being sought. We proceed as explained in Subsection 6.3.1 in order to obtain an approximation of the pdf of Y. However, in the case of quadratic forms in discrete r.v.'s, the sum of squared differences (SSD's) between the empirical and the approximated cdf's is utilized instead of the ISD in order to determine the degree of the polynomial adjustment.

We make use of a gamma density as base density and estimate its parameters by setting its first two moments equal to those of the quadratic form which were obtained by implementing the symbolic approach. This base pdf and the corresponding cdf are respectively shown in Figure 6.12 and 6.13. A polynomial adjustment of degree 9 (determined by applying Result 3) was made in this case. The resulting pdf and cdf approximants are respectively plotted in Figures 6.14 and 6.15. It is seen that the empirical and approximated cdf's nearly coincide.
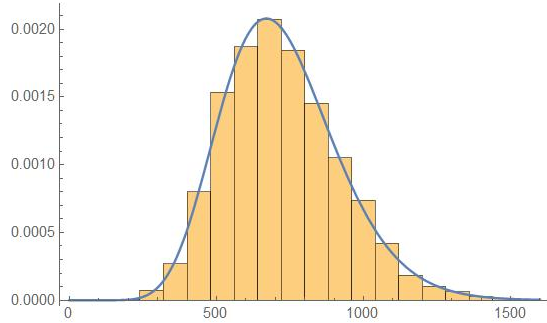
Figure 6.12: Histogram and base pdf
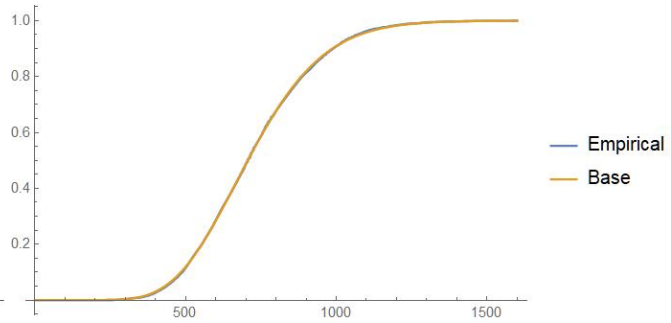
Figure 6.13: Empirical and base cdf's
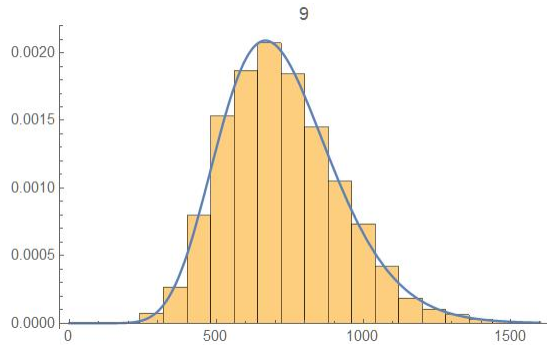
Figure 6.14: Histogram and pdf approximant
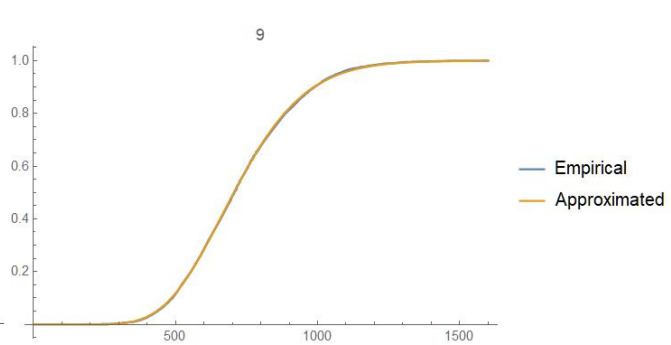
Figure 6.15: Empirical and approximated cdf's

6.3.4   Quadratic forms in Poisson random variables

Consider the case where the matrix of the quadratic form is

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

and the component of the vector $\mathbf{X}$, i.e., X_1, X_2, X_3 and X_4 are independently distributed Poisson random variables with respective parameters 3, 4, 5 and 6, the probability mass function of a Poisson distribution with parameter $\lambda > 0$ being

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots.$$

The cdf's of X_1, X_2, X_3 and X_4 are plotted in Figure 6.16.
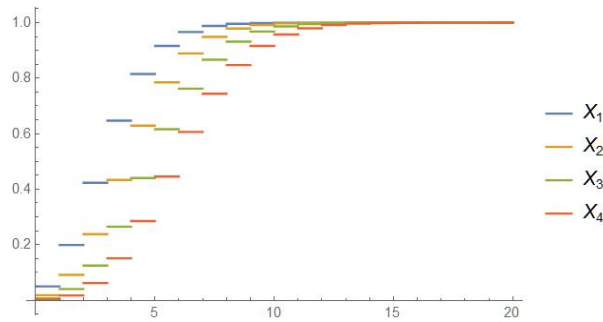
Figure 6.16: cdf's of the X_i's

An approximant to the distribution of $Y = (X_1, X_2, X_3, X_4)A(X_1, X_2, X_3, X_4)'$ is being sought. Since the random variables are discrete, we proceed as explained in Subsection 6.3.3. We also make use of a gamma pdf as base density. This base pdf and the corresponding cdf are respectively shown in Figure 6.17 and 6.18. A polynomial adjustment of degree 10 was deemed appropriate in this case. The resulting pdf and cdf approximants are respectively plotted in Figures 6.19 and 6.20. It is seen that the empirical and approximated cdf's are almost identical.
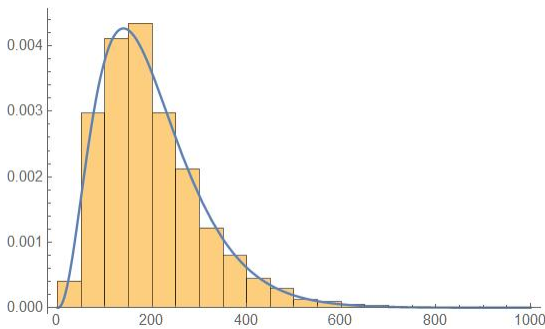
Figure 6.17: Histogram and base pdf
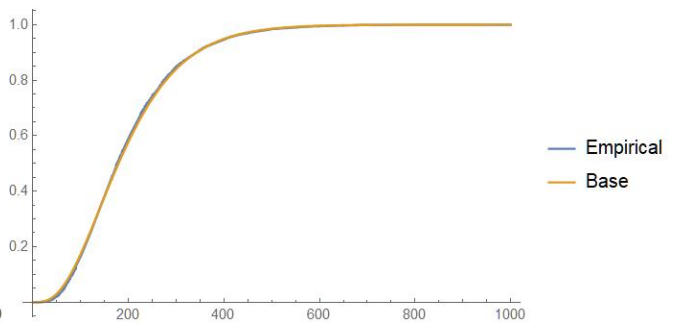
Figure 6.18: Empirical and base cdf's
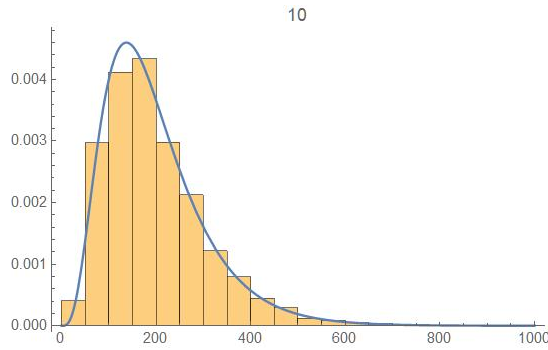
Figure 6.19: Histogram and pdf approximant
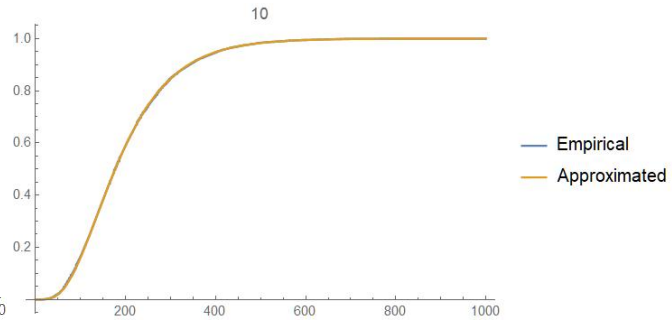
Figure 6.20: Empirical and approximated cdf's

6.4   Hermitian quadratic forms

A Hermitian quadratic form in a complex Gaussian vector $\mathbf{W}$ is an expression of the type $\mathbf{W}^* H \mathbf{W}$ where $\mathbf{W}^*$ denotes the conjugate transpose of $\mathbf{W}$ and H is a Hermitian matrix, that is, $H = H^*$. The distributional properties of Hermitian quadratic forms in complex Gaussian random vectors have been discussed in Bello and Nelin (1962), Khatri (1970), Goodman (1963), Fang *et al.* (1990), Sultan (1999) and Mathai (1997), among others. A representation of the moment generating function is obtained in Shah and Li (2005) by contour integration. A representation of the characteristic function of Hermitian quadratic forms in complex normal variables is provided in Turin (1960). The expected values of certain Hermitian quadratic forms are given in closed forms in Soong (1984).

Hermitian quadratic forms in complex normal variables are frequently encountered in binary hypothesis testing problems, especially in the performance analysis of systems whose inputs are affected by random noise such as radars, sonars, communications receivers and signal acquisition devices. This is explained for instance in Kac and Sieger (1947), Divsalar *et al.* (1990) and Kailath (1960). As pointed out by Biyari and Lindsey (1993), the decision variables in many systems can also be characterized by means of Hermitian quadratic forms in complex Gaussian vectors. Moreover, as explained in Provost and Rudiuk (1995), several statistics utilized for testing hypotheses on the parameters of complex random vectors involve Hermitian quadratic forms. Hermitian quadratic forms were also utilized as cost functions in

Kwon *et al.* (1994), and as characteristic functions in correlated Rician fading environments in Annamalai *et al.* (2005). Kac and Sieger (1947), Turin (1958, 1959), Kailath (1960), Bello and Nelin (1962), Simon and Divsalar (1988), Divsalar *et al.* (1990), Cavers and Ho (1992) and Biyari and Lindsey (1993) expressed pairwise error probabilities of system output decision variables in terms of Hermitian forms. As well, Shah and Li (2005) mentioned an application involving bit error rate calculation in a certain wireless relay network. When studying a full-duplex decode-and-forward relay system in a Rician fading environment, Zhu *et al.* (2008) expressed the highest achievable information rate of the system as a Hermitian quadratic form. As explained in Kay (1989) and Monzigo and Miller (1980), complex random vectors are used in several areas of signal processing such as array processing and spectral analysis. An informative account of various applications involving complex normal vectors and useful related distributional results are included in Picinbono (1996).

We shall focus on Hermitian quadratic forms in Gaussian vectors since applications of Hermitian quadratic forms in other types of random vectors do not seem to be discussed in the statistical literature. It was established in Provost and Chong (2002) that upon diagonalizing a Hermitian quadratic form in a complex Gaussian vectors via a unitary transformation, it can be expressed as a real quadratic form. Accordingly, the technique outlined in Section 6.2 for evaluating the distribution of real quadratic forms will also apply to Hermitian quadratic forms. As will be explained next, the real and complex components of the complex Gaussian vectors will then become the components of a real vector that is twice as long, and the quadratic form will be expressible as a sum of products of some of the components of that real vector. It is thus important to know the distribution of that vector so that its joint moments could be determined by making use of their joint moments.

A complex random vector $\mathbf{W}$ belonging to C^n, the set of n-dimensional complex vectors, can be written as $\mathbf{W} = \mathbf{U} + i\mathbf{V}$ where $\mathbf{U}$ and $\mathbf{V}$ are real random vectors in $\mathfrak{R}^n$. Accordingly, certain problems involving a complex random vector $\mathbf{W}$, can be re-expressed in terms of the real random vector $(\mathbf{U}', \mathbf{V}')'$ in $\mathfrak{R}^{2n}$ where, for instance, $\mathbf{U}'$ denotes the transpose of $\mathbf{U}$. When

$\mathbf{U}$ and $\mathbf{V}$ are correlated n-dimensional real normal vectors with means $\boldsymbol{\mu}_{\mathbf{U}}$ and $\boldsymbol{\mu}_{\mathbf{V}}$, respectively, the random vector $\mathbf{W} = \mathbf{U} + i\mathbf{V}$ has the complex normal distribution $C\mathcal{N}_n(\boldsymbol{\mu}_{\mathbf{W}}, \Gamma, C)$ where $\boldsymbol{\mu}_{\mathbf{W}} = \boldsymbol{\mu}_{\mathbf{U}} + i\boldsymbol{\mu}_{\mathbf{V}} = E(\mathbf{W})$,

$$\Gamma = E[(\mathbf{W} - \boldsymbol{\mu}_{\mathbf{W}})(\overline{\mathbf{W}} - \bar{\boldsymbol{\mu}}_{\mathbf{W}})'] \text{ and } C = E[(\mathbf{W} - \boldsymbol{\mu}_{\mathbf{W}})(\mathbf{W} - \boldsymbol{\mu}_{\mathbf{W}})'], \tag{6.8}$$

$\overline{\mathbf{W}}$ denoting the complex conjugate of $\mathbf{W}$. The covariance matrix Γ is Hermitian and non-negative definite and the relation matrix C is symmetric and non-negative definite. Moreover, as pointed out in Picinbono (1996), the matrices Γ and C must be such that the matrix $\bar{\Gamma} - \bar{C}'\Gamma^{-1/2}C$ be also non-negative definite (which will be assumed throughout), $\Gamma^{-1/2}$ denoting the inverse of the symmetric square root of Γ. We note that in many applications, C is taken to be the null matrix. For instance, that assumption was made in Mathai (1997) when defining the multivariate normal density in the complex case.

It follows from (6.8) that the matrices Γ and C are related to the covariance matrices associated with $\mathbf{U}$ and $\mathbf{V}$ as follows:

$$\text{Cov}(\mathbf{U}) = E[(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})'] = \frac{1}{2}\text{Re}[\Gamma + C],$$

$$\text{Cov}(\mathbf{U}, \mathbf{V}) = E[(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})'] = \frac{1}{2}\text{Im}[-\Gamma + C],$$

$$\text{Cov}(\mathbf{U}, \mathbf{V})' = E[(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})'] = \frac{1}{2}\text{Im}[\Gamma + C],$$

and

$$\text{Cov}(\mathbf{V}) = E[(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})'] = \frac{1}{2}\text{Re}[\Gamma - C],$$

where $\text{Re}[\cdot]$ and $\text{Im}[\cdot]$ respectively denote the real and imaginary parts of $[\cdot]$.

Thus, the real random vector $(\mathbf{U}', \mathbf{V}')'$ corresponding to the complex normal random vector $(\mathbf{U}' + i\mathbf{V}') \sim C\mathcal{N}_n(\boldsymbol{\mu}_{\mathbf{U}} + i\boldsymbol{\mu}_{\mathbf{V}}, \Gamma, C)$ has the following distribution:

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim \mathcal{N}_{2n}\left(\begin{pmatrix} \mu_{\mathbf{U}} \\ \mu_{\mathbf{V}} \end{pmatrix}, \Sigma \right) \tag{6.9}$$

where

$$\Sigma_{2n\times 2n} = \frac{1}{2}\begin{pmatrix} Re[\Gamma + C] & Im[-\Gamma + C] \\ Im[\Gamma + C] & Re[\Gamma - C] \end{pmatrix} \tag{6.10}$$

and $\mathcal{N}_\nu(\mu, \Sigma)$ denotes a real ν-dimensional normal vector whose mean and covariance matrix are respectively μ and Σ.

6.4.1 A numerical example

Let

$$A = \begin{pmatrix} 6 & -2i & 1 \\ 2i & 2 & 1-i \\ 1 & 1+i & 6 \end{pmatrix},$$

and $Y = (X_1, X_2, X_3)A(X_1, X_2, X_3)^*$, where $(\cdot)^*$ denotes the conjugate transpose and X_1, X_2 and X_3 are independent complex normal variables, with the respective parameters $\mu_1 = 3$; $\sigma_1 = 2$; $\mu_2 = -5$; $\sigma_2 = 1$; $\mu_3 = 4$; $\sigma_3 = 3$ for both their real and complex parts. The pdf's of the real parts of X_1, X_2, and X_3 are shown in the Figure 6.21.

In this case, we also use a gamma pdf as the base density. The parameters are determined by setting the first two moments of the gamma distribution equal to those of the quadratic form, which are obtained by expanding Y and determining the expected value of the resulting expression and its square.

Since the base pdf is already very accurate, which can be seen from Figure 6.22 and in particular Figure 6.23, there is no need for an adjustment in this case and this base density function can serve as an approximation to the density of Y.
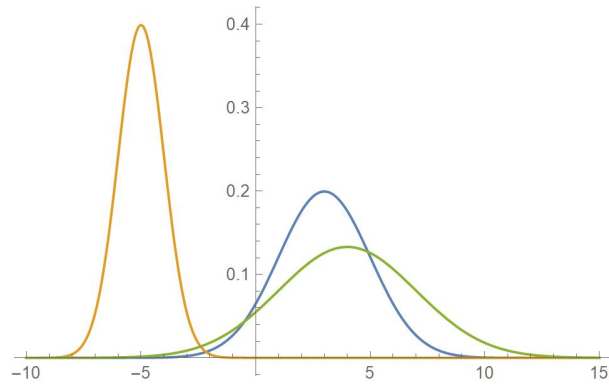
Figure 6.21: pdf's of the real parts of X_1, X_2 and X_3
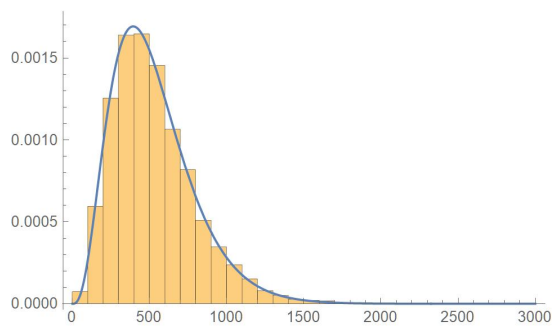
Figure 6.22: Histogram and the base pdf
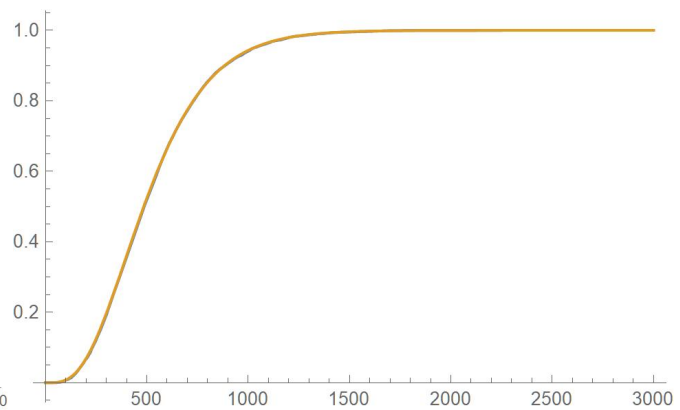
Figure 6.23: Empirical and base cdf's

6.5 A parametric approach for quadratic forms in Gaussian random vectors

6.5.1 An extended generalized gamma distribution

The currently available parametric families of distributions may prove inadequate for modeling purposes. A very flexible model that comprises numerous standard distributions will be utilised to approximate the distribution of quadratic forms. It is referred to a the q-extended generalised gamma distribution.

We initially define the *q-exponential distribution* as an example of a q-analogue distribution. This distribution generalises the exponential distribution whose pdf is given by

$$f(x) = \begin{cases} (2 - q)\lambda\, e_q(-\lambda\, x)\, \mathbb{1}_{\{x>0\}}, & 1 \le q < 2,\ \lambda > 0, \\ (2 - q)\lambda\, e_q(-\lambda\, x)\, \mathbb{1}_{\{0<x<[(1-q)\lambda]^{-1}\}}, & q < 1,\ \lambda > 0, \end{cases} \tag{6.11}$$

where $\mathbb{1}_{\{\,.\,\}}$ is the indicator function and

$$e_q(y) = \begin{cases} e^y, & \text{as } q \to 1, \\ [1 + (1 - q)\, y]^{1/(1-q)}, & \text{for } q \ne 1,\ 1 + (1 - q)\, x > 0, \\ 0, & \text{otherwise}, \end{cases}$$

is the *q-exponential* function. Note that he exponential distribution is retrieved when q tends to 1.

We will use this *q*-exponential function to secure the *q-extended generalized gamma (q-EGG) distribution*. The functional parts of the pdf's associated with the three types of q-EGG random variables are given in 6.5.2 where the moments and normalizing constants are given. This pdf is used in Section 6.5.3 to approximate the distribution of a positive definite quadratic form as well as an indefinite quadratic form in Gaussian random variables.

6.5.2   The density function and moments

The generalised gamma density can be expressed as

$$f(x; a, d, p) = \frac{p/a}{\Gamma(d/p)} (x/a)^{d-1} e^{-(x/a)^p}, \quad x > 0, \, a > 0, \, d > 0, \, p > 0. \tag{6.12}$$

On substituting $e_q(-(x/a)^p)$ to $e^{-(x/a)^p}$ in 6.12, on obtains the pdf

$$f(x; a, d, p, q) \propto \left(\frac{x}{a}\right)^{d-1} \left[1 - (1-q)\left(\frac{x}{a}\right)^p\right]^{\frac{1}{1-q}}, \quad x > 0, \, a > 0, \, d > 0, \, p > 0. \tag{6.13}$$

Now, adding a location parameter ν and reparameterising a, d, and p as $a = \tau$, $d = \lambda\delta$, and $p = \delta$, one has

$$f(x; \nu, \tau, \lambda, \delta, q) \propto \left(\frac{x-\nu}{\tau}\right)^{\lambda\delta-1} \left[1 - (1-q)\left(\frac{x-\nu}{\tau}\right)^\delta\right]^{\frac{1}{1-q}}, \tag{6.14}$$

where $1 - (1-q)[(x-\nu)/\tau]^\delta > 0$ and $\lambda > 0$. We observe that letting $\tau < 0$ and/or $\delta < 0$ still produces bona fide pdf's. Accordingly, the parameter space is $\tau, \delta \in \mathbb{R} \setminus \{0\}$, $\lambda > 0$, $\nu \in \mathbb{R}$ (with $x > \nu$ if $\tau > 0$ and $x < \nu$ if $\tau < 0$). We refer to the model whose pdf is specified in 6.14 as the *q-analogue of the extended generalized gamma (q-EGG) distribution*.

We consider three types of q-EGG distributions.

- When $q < 1$, the q-EGG density can be expressed as

$$f_1(x; \nu, \tau, \lambda, \delta, q) = c_{1,\lambda} \left(\frac{x-\nu}{\tau}\right)^{\lambda\delta-1} \left[1 - (1-q)\left(\frac{x-\nu}{\tau}\right)^\delta\right]^{\frac{1}{1-q}}, \tag{6.15}$$

where $c_{1,\lambda}$, the normalizing constant which is given in 6.18a, also depends on δ, λ and ν, with $\nu \in \mathbb{R}$, $\tau \neq 0$, $\lambda > 0$, $\delta \neq 0$, and it support is as follows:

1. $\nu < x < \nu + \tau(1-q)^{-1/\delta}$ if $\tau > 0$, $\delta > 0$ and $x > \nu + \tau(1-q)^{-1/\delta}$ if $\tau > 0$, $\delta < 0$;

2. $\nu + \tau(1-q)^{-1/\delta} < x < \nu$ if $\tau < 0$, $\delta > 0$ and $x < \nu + \tau(1-q)^{-1/\delta}$ if $\tau < 0$, $\delta < 0$.

The model 6.15 will be called a *type-1 q-EGG distribution*.

- When $q > 1$, we may express $(1 - q)$ as $-(q - 1)$, so that the q-EGG density becomes

$$f_2(x; \nu, \tau, \lambda, \delta, q) = c_{2,\lambda} \left(\frac{x - \nu}{\tau}\right)^{\lambda\delta - 1} \left[1 + (q - 1) \left(\frac{x - \nu}{\tau}\right)^\delta\right]^{-\frac{1}{q-1}}, \qquad (6.16)$$

where $c_{2,\lambda}$ is the normalizing constant, $\nu \in \mathbb{R}$, $\tau \neq 0$, $0 < \lambda < (q - 1)^{-1}$, $\delta \neq 0$; in this instance, the support is given by $x > \nu$ if $\tau > 0$ and $x < \nu$ if $\tau < 0$. The distribution specified by 6.16 will be referred to as a *type-2 q-EGG distribution*.

- When q tends to 1, 6.15 boils down to the *Amoroso distribution* whose density function is

$$f_3(x; \nu, \tau, \lambda, \delta) = c_{3,\lambda} \left(\frac{x - \nu}{\tau}\right)^{\lambda\delta - 1} \exp\left\{-\left(\frac{x - \nu}{\tau}\right)^\delta\right\}, \qquad (6.17)$$

where $c_{3,\lambda}$ is the normalizing constant, $\nu \in \mathbb{R}$, $\tau \neq 0$, $\lambda > 0$, $\delta \neq 0$. Its support is $x > \nu$ if $\tau > 0$ and $x < \nu$ if $\tau < 0$. This model is referred to as a *type-3 q-EGG distribution*.

As the location parameter ν does not change the normalizing constants $c_{1,\lambda}$, $c_{2,\lambda}$ and $c_{3,\lambda}$, we assume without any loss of generality that $\nu = 0$ in the derivations.

The normalizing constants

It suffices to integrate the probability density functions 6.15, 6.16, 6.17 and to equate the results to 1, to secure the normalizing constants which are:

$$c_{1,\lambda} = \frac{|\delta/\tau| (1 - q)^\lambda \, \Gamma(1 + \frac{1}{1-q} + \lambda)}{\Gamma(1 + \frac{1}{1-q}) \Gamma(\lambda)}, \qquad (6.18a)$$

$$c_{2,\lambda} = \frac{|\delta/\tau| (q - 1)^\lambda \, \Gamma(\frac{1}{q-1})}{\Gamma(\frac{1}{q-1} - \lambda) \Gamma(\lambda)}, \qquad (6.18b)$$

$$c_{3,\lambda} = \frac{|\delta/\tau|}{\Gamma(\lambda)}. \qquad (6.18c)$$

For example, $c_{1,\lambda}$ can be obtained as follows.

Case 1. Assuming that $\tau > 0$ and $\delta > 0$, letting $y = x/\tau$ and integrating 6.15, one has

$$1 = c_{1,\lambda}\,\tau \int_0^{(1-q)^{-1/\delta}} y^{\lambda\delta-1}\left[1 - (1-q)\,y^\delta\right]^{\frac{1}{1-q}}\,dy. \tag{6.19}$$

With the substitution $u = (1-q)\,y^\delta$, 6.19 can be expressed as follows:

$$1 = c_{1,\lambda}\,\frac{\tau}{\delta\,(1-q)^\lambda}\int_0^1 u^{\lambda-1}\,(1-u)^{\frac{1}{1-q}}\,du$$

$$= c_{1,\lambda}\,\frac{\tau}{\delta\,(1-q)^\lambda}\,B(\lambda,\ 1 + \tfrac{1}{1-q}),$$

where $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}\,(1-t)^{\beta-1}\,dt = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ denotes the *beta function*. Then,

$$c_{1,\lambda} = \frac{(\delta/\tau)\,(1-q)^\lambda}{B(\lambda,\ 1 + \tfrac{1}{1-q})} = \frac{(\delta/\tau)\,(1-q)^\lambda\,\Gamma(1 + \tfrac{1}{1-q} + \lambda)}{\Gamma(1 + \tfrac{1}{1-q})\,\Gamma(\lambda)}. \tag{6.20}$$

Case 2. If $\tau > 0$ and $\delta < 0$, on setting $y = x/\tau$ and integrating 6.15, one has

$$1 = c_{1,\lambda}\,\tau \int_{(1-q)^{-1/\delta}}^{\infty} y^{\lambda\delta-1}\left[1 - (1-q)\,y^\delta\right]^{\frac{1}{1-q}}\,dy \tag{6.21}$$

$$= c_{1,\lambda}\,\tau \int_0^{(1-q)^{1/\delta}} z^{-\lambda\delta-1}\left[1 - (1-q)\,z^{-\delta}\right]^{\frac{1}{1-q}}\,dz,$$

with $z = 1/y$. On letting $u = (1-q)\,z^{-\delta}$, one obtains

$$1 = -c_{1,\lambda}\,\frac{\tau}{\delta\,(1-q)^\lambda}\int_0^1 u^{\lambda-1}\,(1-u)^{\frac{1}{1-q}}\,du$$

$$= -c_{1,\lambda}\,\frac{\tau}{\delta\,(1-q)^\lambda}\,B(\lambda,\ 1 + \tfrac{1}{1-q}).$$

Consequently,

$$c_{1,\lambda} = \frac{(-\delta/\tau)\,(1-q)^\lambda}{B(\lambda,\ 1 + \tfrac{1}{1-q})} = \frac{(-\delta/\tau)\,(1-q)^\lambda\,\Gamma(1 + \tfrac{1}{1-q} + \lambda)}{\Gamma(1 + \tfrac{1}{1-q})\,\Gamma(\lambda)}. \tag{6.22}$$

Cases 3 and 4. Whenever $\tau < 0$, we set $y = x/\tau$. For the cases $\delta > 0$ and $\delta < 0$, we end

up with 6.19 and 6.21 wherein τ is replaced by $-\tau$. Accordingly, the normalizing constants corresponding to the cases $\tau < 0, \delta > 0$ and $\tau < 0, \delta < 0$ are in fact specified by 6.22 and 6.20.

Thus, the normalizing constants $c_{1,\lambda}$ can altogether be determined from the expression given in 6.18a.

The normalizing constants specified in 6.18b and 6.18c can be similarly derived.

For all three types, the hth moment is given by

$$\frac{c_{i,\lambda}\tau^h}{c_{i,\lambda+\frac{h}{\delta}}}, \quad i = 1, 2, 3.$$

These results were derived in Chen (2022).

6.5.3   A parametric approximation to the distribution of quadratic forms

We are initially determining the distribution of a positive definite quadratic form by making use of a q-EGG distribution. We first specify the matrix of the quadratic form which is

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

The three random variables, denoted as T_1, T_2 and T_3, comprising the vector of this quadratic form are independent standard normal variables. The quadratic form is denoted as $Q = (T_1, T_2, T_3)A(T_1, T_2, T_3)'$. The kernel density estimate which is based on 10,000 simulated values, is plotted in Figure 6.24.
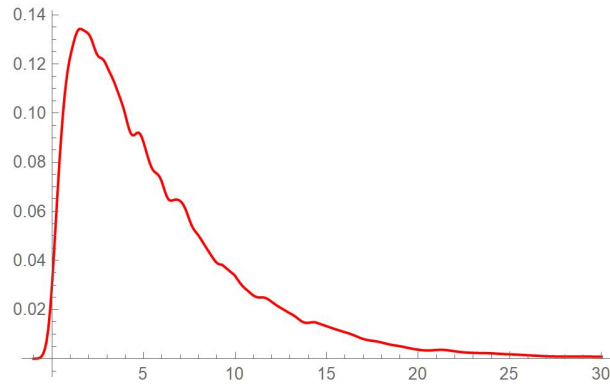
Figure 6.24: The kernel density estimate of Q

Making use of the method of moments, we first secured a gamma approximation whose pdf and cdf are shown in Figures 6.25 and 6.26. We employed the integrated squared difference between cdf's as goodness-of-fit measure, which in this case is equal to 0.00163148. On the basis of the results of the gamma approximation, we were able to determine initial intervals for two of the parameters of a 4-parameter q-EGG approximation. We could then secure more accurate pdf and cdf approximations which are respectively shown in Figures 6.27 and 6.28, the associated integrated squared difference, 0.000676887, being smaller than that of the gamma approximation. The details of the calculations are provided in the Appendix.
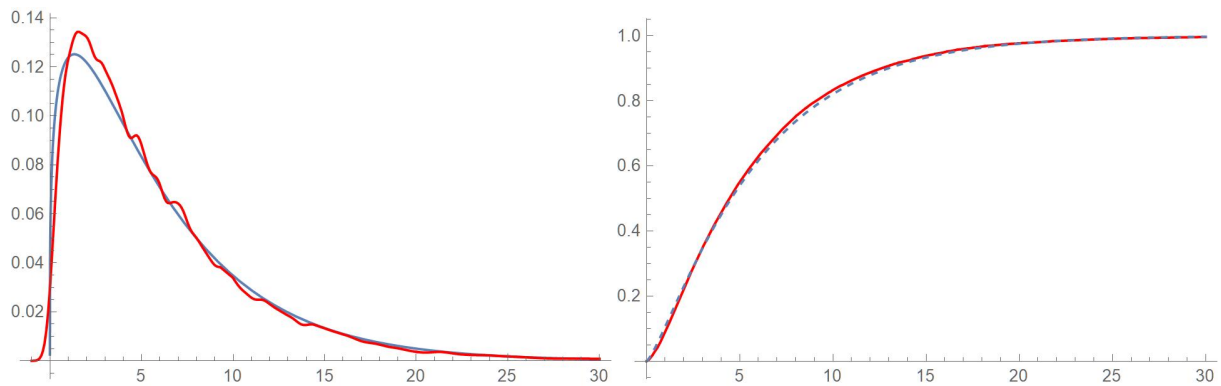
Figure 6.25: gamma pdf approximation (blue) and the kde (red)

Figure 6.26: gamma cdf approximation (blue, dashed) and the empirical cdf (red)
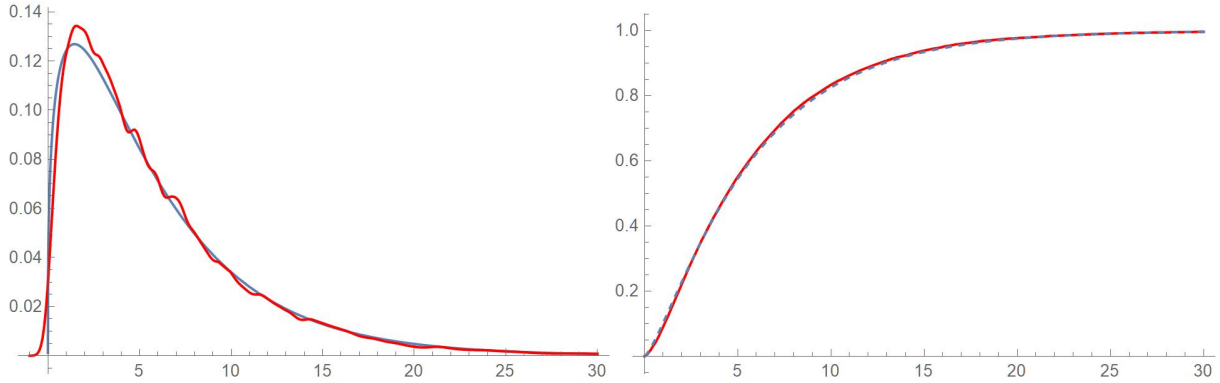
Figure 6.27: q-EGG pdf approximation (blue) and the kde (red)

Figure 6.28: q-EGG cdf approximation (blue, dashed) and the empirical cdf (red)

We also secured an approximation to the distribution of an indefinite quadratic form which is based on two q-EGG density approximants. The matrix of this quadratic form is

$$
A = \begin{pmatrix}
-12 & 0 & 0 & 0 & 0 \\
0 & -8 & 0 & 0 & 0 \\
0 & 0 & -4 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 5
\end{pmatrix}.
$$

Its random vector is $(T_1, T_2, T_3, T_4, T_5)$ where the independent T_i's are assumed to have a standard normal distribution. The kernel density estimate obtained from 10,000 simulated values is shown in Figure (6.29). Note that only diagonal matrices need be considered since any symmetric matrix can be diagonalized in light of the spectral decomposition theorem.
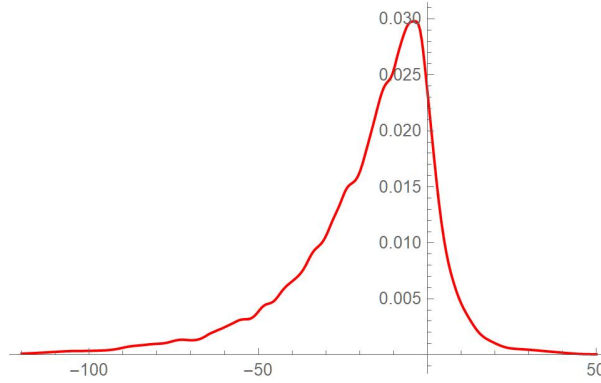
Figure 6.29: The kernel density estimate of the indefinite quadratic form Z

This quadratic form also has the following representation:

$$Z = (T_1, T_2, T_3, T_4, T_5)A(T_1, T_2, T_3, T_4, T_5)' = -(12T_1^2 + 8T_2^2 + 4T_3^2) + (T_4^2 + 5T_5^2).$$

Thus, it can be expressed as the difference of two positive definite quadratic forms: $X = (T_4, T_5)A_1(T_4, T_5)'$ and $Y = (T_1, T_2, T_3)A_2(T_1, T_2, T_3)'$, where

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}.$$

and

$$A_2 = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

The distributions of each of the quadratic forms X and Y is approximated with a q-EGG model. As shown in the Appendix, the parameters of these distributions are obtained by minimizing the sum of the squares of the differences between the first four theoretical moments of the selected model and the corresponding moments of each quadratic form. On letting $Z = X - Y$ and $W = Y$ and applying the transformation of variables technique, on can secure a pdf approximation to the target distribution from the marginal distribution of Z. We first obtained the approximated

pdf's of X and Y, their joint pdf being the product of their density functions. Upon securing the joint distribution of Z and W, that of Z was obtained by integrating out W. The pdf and cdf estimates of X and Y and that of the indefinite quadratic form are plotted in Figures 6.30-6.34.
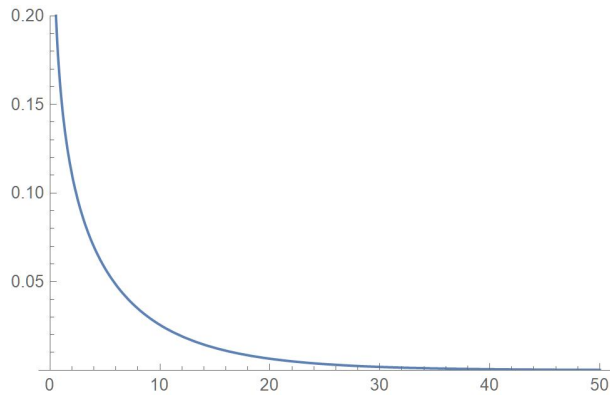
Figure 6.30: The estimated density of X
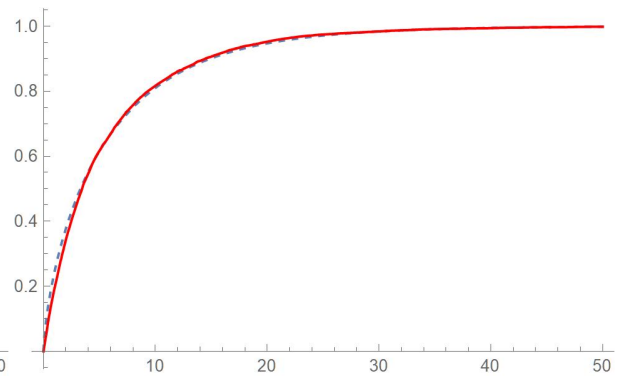
Figure 6.31: The estimated cdf (blue, dashed) and the empirical cdf (red) of X
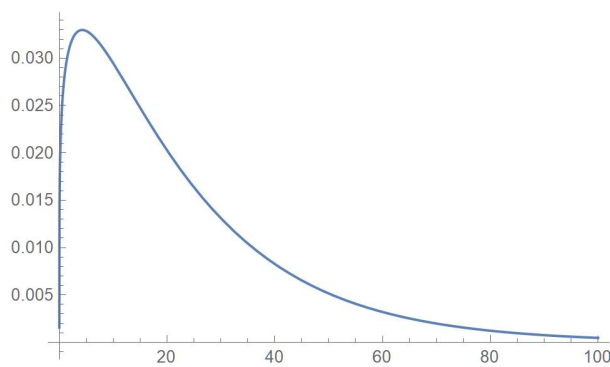
Figure 6.32: The estimated density of Y
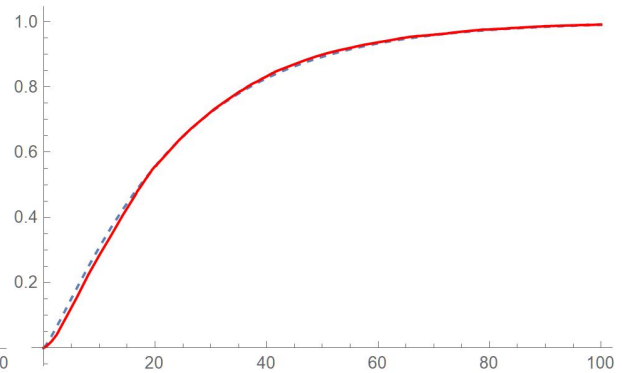
Figure 6.33: The estimated cdf (blue, dashed) and the empirical cdf (red) of Y
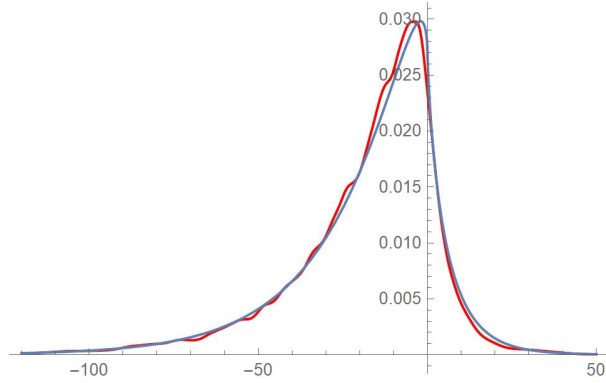
Figure 6.34: The kernel density estimate (red) and the approximated density (blue) of the indefinite quadratic form Z

Bibliography

[1] Annamalai, A., Tellambura, C. and Bhargava, V. K. A. (2005). A general method for calculating error probabilities over fading channels. *IEEE Transactions on Communications*, 53, 841–852.

[2] Barrio, E. D., Gine, E. and Utzet, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11, 131–189.

[3] Bello, P. and Nelin, B. D. (1962). Predetection diversity combining with selectively fading channels. *IRE Transactions on Communications Systems*, 10, 32–42.

[4] Biyari, K. H. and Lindsey, W. C. (1993). Statistical distribution of Hermitian quadratic forms in complex Gaussian variables. *IEEE Transactions on Information Theory*, 39, 1076–1082.

[5] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Ann. Math. Statist.*, 25, 290–302.

[6] Cavers, J. K. and Ho, P. (1992). Analysis of the error performance of Trellis-coded modulations in Rayleigh fading channels. *IEEE Transactions on Communications*, 40, 74–83.

[7] Chen, W. (2022). The q-Analogue of the extended generalized gamma distribution. https://ir.lib.uwo.ca/usri/usri2022/ReOS/205/

[8] Davis, R. B. (1973). Numerical inversion of a characteristic function, *Biometrika*, 60, 415–417.

[9] Divsalar, D., Simon M. K. and Shahshahani, M. (1990). The performance of Trellis-coded MDPSK with multiple symbol detection. *IEEE Transactions on Communications*, 38, 1391–1403.

[10] Donald, S. G. and Paarsch, H. J. (2002). Superconsistent estimation and inference in structural econometric models using extreme order statistics. *Journal of Econometrics*, 109, 305–340.

[11] Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London.

[12] Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34, 152–177.

[13] Gurland, J. (1953). Distribution of quadratic forms and ratios of quadratic forms. *Ann. Math. Statist.*, 24, 416–427.

[14] Gurland, J. (1956). Quadratic forms in normally distributed random variables. *Sankhya*, 17, 37–50.

[15] Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419–426.

[16] Kac, M. and Sieger, A. J. F. (1947). On theory of noise in radio receivers with square-law detectors. *Journal of Applied Physics*, 18, 383–397.

[17] Kailath, T. (1960). Optimum receivers for randomly varying channels. In Cherry, C. editor, *Information Theory*. Butterworth and Co., London, 109–122.

[18] Kay, S. (1989). *Modern Spectral Analysis, Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ.

[19] Khatri, C. J. (1970). On the moments of traces of two matrices in three situations for complex multivariate normal populations,*Sankhyā Series A* , 32, 65–80.

[20] Kotz, S., Johnson, N. L. and Boyd, D. W. (1967a). Series representation of distribution of quadratic forms in normal variables I. Central case. *Ann. Math. Statist.*, 38, 823–837.

[21] Kotz, S., Johnson, N. L. and Boyd, D. W. (1967b). Series representation of distribution of quadratic forms in normal variables II. Non-central case. *Ann. Math. Statist.*, 38, 838–848.

[22] Kwon, O., Kim, B. and Ih, J. (1994). On the positioning of control sources in active noise control of three-dimensional interior space. *Journal of Mechanical Science and Technology*, 8, 283–292.

[23] MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *The Annals of Statistics*, 6, 422–433.

[24] Mathai, A. M. (1997). *Jacobians of Matrix Transformations and Functions of Matrix Arguments*. World Scientific Publishing, New York.

[25] Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms in Random Variables, Theory and Application*. Marcel Dekker Inc., New York.

[26] Monzigo, R. A. and Miller, T. W. (1980). *Introduction to Adaptive Arrays*. Wiley, New York.

[27] Pachares, J. (1955). Note on the distribution of a definite quadratic form. *Ann. Math. Statist.*, 26, 128–131.

[28] Picinbono, B. (1993). *Random Signals and Systems*. Prentice-Hall, Englewood Cliffs, NJ.

[29] Picinbono, B. (1996). Second-order complex random vectors and normal distributions. *IEEE Transactions on Signal Processing*, 44, 2637–2640.

[30] Provost, S. B. (1989). The distribution function of a statistic for testing the equality of scale parameters in two gamma populations. *Metrika*, 36, 337–345.

[31] Provost, S. B. (2005). Moment-based density approximants. *The Mathematica Journal*, 9, 727–756.

[32] Provost, S. B. and Cheong, Y. (2002). The distribution of Hermitian quadratic forms in elliptically contoured random vectors. *Journal of Statistical Planning and Inference*, 102, 303–316.

[33] Provost, S. B., Jiang, M., and Ha, H.-T. (2009). Moment-based approximations of probability mass functions with applications involving rank-sum test statistics. *Communications in Statistics - Theory and Methods*, 38, 1969–1981.

[34] Provost, S. B. and Rudiuk, E. (1995). Moments of test statistics for covariance structures and their exact densities as inverse Mellin transforms. *International Journal of Mathematical and Statistical Sciences*, 4, 71–90.

[35] Rice, S. O. (1980). Distribution of quadratic forms in normal variables. Evaluation by numerical integration. *SIAM J. Scient. Statist. Comput.*, 1, 438–448.

[36] Ruben, H. (1960). Probability content of regions under spherical normal distribution. I. *Ann. Math. Statist.*, 31, 598–619.

[37] Ruben, H. (1962). Probability content of regions under spherical normal distribution. IV: The distribution of homogeneous and nonhomogeneous quadratic functions in normal variables. *Ann. Math. Statist.*, 33, 542–570.

[38] Shah, B. K. (1963). Distribution of definite and of indefinite quadratic forms from a non-central normal distribution. *Ann. Math. Statist.*, 34, 186–190.

[39] Shah, B. K. and Khatri, C. G. (1961). Distribution of a definite quadratic form for non-central normal variates.*Ann. Math. Statist.*, 32, 883–887.

[40] Shah, N. and Li, H. (2005). Distribution of quadratic form in Gaussian mixture variables and an application in relay networks. *Signal Processing Advances in Wireless Communications IEEE 6th Workshop*, 490–494.

[41] Simon, M. and Divsalar, D. (1988). The performance of Trellis-coded multilevel DPSK on a fading mobile satellite channel. *IEEE Transactions on Vehicular Technology*, 37, 78–91.

[42] Smith, P. J. (1995). A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. *The American Statistician*, 49, 217–219.

[43] Soong, T. T. (1984). A note on expectations of a random quadratic forms. *Stochastic Analysis and Applications*, 2, 295–298.

[44] Sultan, S. A. (1999). The distribution of Hermitian indefinite quadratic forms. *Stochastic Analysis and Applications*, 17, 275–293.

[45] Turin, G. L. (1958). Error probabilities for binary symmetric ideal reception through non-selective slow fading and noise. *Proceedings of the Institute of Radio Engineers*, 46, 1603–1619.

[46] Turin, G. L. (1959). Some computations of error rates for selectively fading multipath channels. *Proceedings of the National Electronics Conference*, Chicago, Illinois.

[47] Turin, G. L. (1960). The characteristic function of Hermitian quadratic forms in complex normal variables. *Biometrika*, 47, 199–201.

[48] Whittle, P. (1960). Quadratic forms in Poisson and multi-nomial variables. *Journal of the Australian Mathematical Society*, 1, 233–240.

[49] Zhu, Y., Xin, Y. and Kam, P. (2008). Outage probability of Rician fading relay channels. *IEEE Transactions on Vehicular Technology*, 57, 2648–2652.

Chapter 7

Estimating the Proportion of Information Contained in Sets of Moments

7.1 Introduction

An approach for quantifying the amount of information that is contained in a set of moments is proposed in this chapter. This information turns out to be useful in assessing how many moments ought to be utilized when applying moment-based density estimation methodologies such as those developed in Provost (2005) and Provost *et al.* (2022). To our knowledge, this very problem has not yet been tackled in the statistical literature.

Rather than attempting to solve it directly, we will make use of a set of points that are representative of the distribution under consideration, which is equivalent to utilizing the corresponding number of moments in light of the following result which appeared in Provost *et al.* (2022).

Result 4 *A set of n observations is uniquely determined by the first n associated sample moments.*

Since the proof is rather short, it is included herein for the sake of completeness.

135

Proof Let $S = \{x_1, x_2, \ldots, x_n\}$, $\mathcal{M} = \{m_1, m_2, \ldots, m_n\}$ and $m_h = \sum_{i=1}^n x_i^h / n$. According to the fundamental theorem of algebra, $p(z) = a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n$ is uniquely specified by its n roots denoted by x_i, $i = 1, 2, \ldots, n$.

Moreover, given S, the coefficients of $p(x)$ can be expressed in terms of the sequence of moments $\mathcal{M}$ via the Newton-Girard identity. Accordingly, a given polynomial of degree n, say $p(x)$, can be represented as follows:

$$\prod_{i=1}^n (x - x_i) = \sum_{k=0}^n (-1)^{n-k} e_{n-k} \, x^k, \tag{7.1}$$

where $e_0 = 1$ and

$$e_\ell = \frac{n}{\ell} \sum_{j=1}^\ell (-1)^{j-1} e_{\ell-j} m_j, \quad \ell = 1, \ldots, n. \tag{7.2}$$

Thus, given the first n sample moments associated with a sample of size n, one can determine the right hand side of Equation (7.1), whose roots are precisely $x_1, x_2, ..., x_n$. This establishes that S is uniquely specified by $\mathcal{M}$.

Although it follows from Result 4 that a set of n data points and its first n associated sample moments contain exactly the same amount information, only a subset of the latter usually suffices to elicit the relevant characteristics of the underlying distribution. This is indeed the case since low-order moments are much more informative than those of higher order whereas individual data point holds the same proportion of the total information, namely $1/n$. Accordingly, moment-based density estimation techniques ought to prove for instance more efficient than kernel-based approaches which rely on every sample points—especially in the case of massive data sets. As the relevant information can be conveyed by a moderate number of moments, it is in general pertinent to make use of sample moments for modeling, classification or inference purposes.

In this chapter, our objective consists of quantifying the relative amount of information that a set of moments up to a certain order holds with respect to a given distribution. The proposed methodology which is described in Section 7.3 relies on the most representative points of a

distribution as specified in the next section. As shown is Section 7.4, the approach that is herein advocated produces results that are consistent with expectations.

7.2 The most representative sample points of a distribution

Since there exists an equivalency between the number of observations in a random samples and the number of sample moments available, we shall determine the information contained in a certain number of moments via the information provided by the same number of representative sample points with respect to a given distribution.

A representative sample of size h is determined as follows. Fang and Wang (1993) proposed the following measure of representativeness of a sample $S = \{x_1, x_2, \ldots, x_h\}$ with respect to the distribution function $F(x)$, which is referred to as F-discrepancy:

$$D_F(S) = \sup_{x \in \mathcal{R}} |F_h(x) - F(x)|, \tag{7.3}$$

where $F_h(x)$ denotes the empirical distribution function of S. We observe that $D_F(S)$ is in fact the Kolmogorov-Smirnov statistic for assessing goodness-of-fit with respect to $F(x)$. In one dimension,

$$\left\{ F^{-1}\left(\frac{2i-1}{2h} \right), \ i = 1, 2, \ldots, h \right\}$$

constitutes the set of points having lowest F-discrepancy. In that sense, this set of points form the most representative sample with respect to a distribution whose cdf is $F(x)$. It is referred to as a pseudo-random sample.

Remark. Since the information that is solely imparted by the median is not sufficient to provide a utilizable pdf estimate, the proposed methodology is initially applied to at least two representative points of the distribution at hand. As well, we note that by letting the base density function $\psi_X(x)$ be constant in Equation (6.1), the resulting estimate is a polynomial of degree n whose coefficients require n moments. Thus, a single point or equivalently a single

moment would not suffice to produce a satisfactory pdf estimate as a linear function would manifestly be inadequate. Furthermore, if a distribution happens to have δ *distinct* modes, one could only proceed at first with 2δ points/moments since on implementing the moment-based estimation technique described in Result 3 of Section 6.1, one would need a polynomial of degree at least 2δ in order to secure a reasonably accurate density estimate. Polynomials of lesser degrees would inevitably miss some distinctive features of the distribution at hand. For instance, in the case of a bimodal distribution, a polynomial of degree three could not provide an adequate pdf estimate. Although it might be of some interest to know how informative individual moments are, only the number of lower moments corresponding to a large cumulative proportion of the information—which in most instances, is markedly larger than 2δ—would customarily be required in applications, as higher moments could be then disregarded.

7.3   Quantifying the proportion of information contained in sets of moments

In order to quantify the amount of information that moments are holding through the most representative points, we make use of Fritsch-Carlson (FC) monotonic piecewise cubic interpolants and the Kullback-Leibler (KL) measure of divergence. Fritsch and Carlson (1980) proposed a method for constructing a monotone piecewise cubic interpolant for a monotone set of points, which is suitable for approximating cdf's. Kullback and Leibler (1951) introduced the concept of the KL divergence, which is a measure of the mean information for discriminating between two distributions, which can be formally defined as follows: Given continuous distributions P and Q, the KL divergence from Q to P is the following expectation with respect to P:

$$KL(p \,\|\, q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx, \qquad (7.4)$$

where $p(\cdot)$ and $q(\cdot)$ denote the probability density functions corresponding to P and Q. It is also the relative entropy from Q to P or relative entropy of P with respect to Q. Most relevant in the context of this chapter, it is also known as the information gain from P over Q. Alternatively, as pointed out in Burnham and Anderson (2002, p.51), $KL(P\|Q)$ represents the amount of information lost when Q is used to approximate P.

Upper and lower bounds for the support of X, say lb and ub, are specified and added to the n most representative points. They could be determined for instance from a preliminary kernel density estimate. We then apply the FC interpolation technique to the following monotonic set of cdf points:

$$\left\{ \{lb, 0\}, \left\{ F^{-1}\left(\frac{1}{2h}\right), \frac{1}{2h} \right\}, \left\{ F^{-1}\left(\frac{3}{2h}\right), \frac{3}{2h} \right\}, \ldots, \left\{ F^{-1}\left(\frac{2h-1}{2h}\right), \frac{2h-1}{2h} \right\}, \{ub, 1\} \right\} \quad (7.5)$$

and secure a density estimate denoted by $f_h(x)$ by taking the derivative of the FC spline which incidentally is not a smooth function. The target density denoted by $f(x)$ is either known or estimated from a sample.

The KL divergence between the target density $f(x)$ and $f_h(x)$, i.e., $KL(f \| f_h)$ is used as a criterion for determining how much more information remains to be gained with more than h points/moments. This quantity is denoted by $KLf(h)$.

The uniform distribution on the interval (lb, ub) denoted by $f_0(x)$ is viewed as being non informative. Accordingly, the KL divergence between $f(x)$ and the constant density function $f_0(x)$, i.e., $KL(f \| f_0)$ is regarded as the total information held by $f(x)$ and is denoted by $KLfc$.

Then, $KLfc - KLf(h)$ is taken to be the amount of information one can gain by making use of the first h points/moments. Thus, the proportion of information included in the first h moments relative to $KLfc$ is evaluated as follows: $\frac{KLfc - KLf(h)}{KLfc}$. This quantity will increase from 0 to nearly one as h increases.

Additionally, $\frac{KLf(h) - KLf(h+s)}{KLfc}$ will be taken as the proportion of the total information that the set of moments of orders $h + 1$ to $h + s$ are holding.

7.4   Illustrative examples

Two exact distributions and two data sets possessing distinct distributional features are considered.

7.4.1   Two exact distributions involving beta random variables

The first distribution being considered in this section is a single skewed beta distribution with parameters (3,9), the second one being an equal mixture of two beta density functions with parameters (7,15) and (12,8), which is a bimodal distribution. Their pdf's are plotted in Figures 7.1 and 7.2. In both instances, the end points are 0 and 1. The Fritsch-Carlson monotonic piecewise cubic interpolation technique is applied to the points specified in (7.5) and their derivatives are taken to be the corresponding density estimates of the underlying density. To visually illustrate the process, plots of the FC cdf interpolants and their derivatives are shown for selected values of h in Figures 7.3-7.14 for the case of the single skewed beta density function.
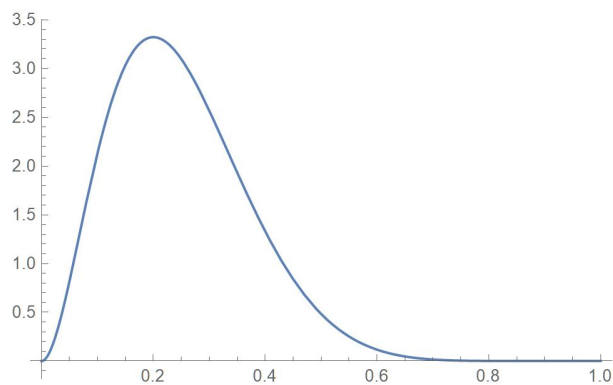
Figure 7.1: A single skewed beta density   Figure 7.2: A mixture of two beta densities
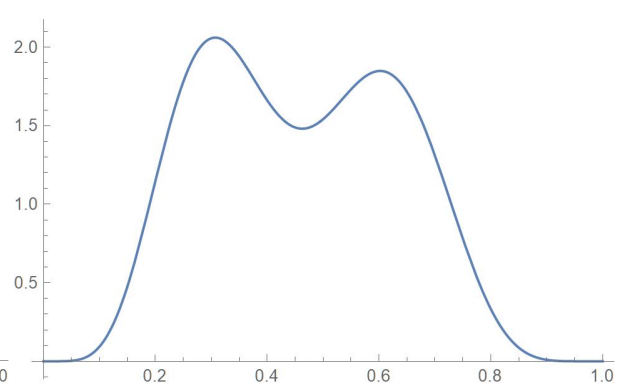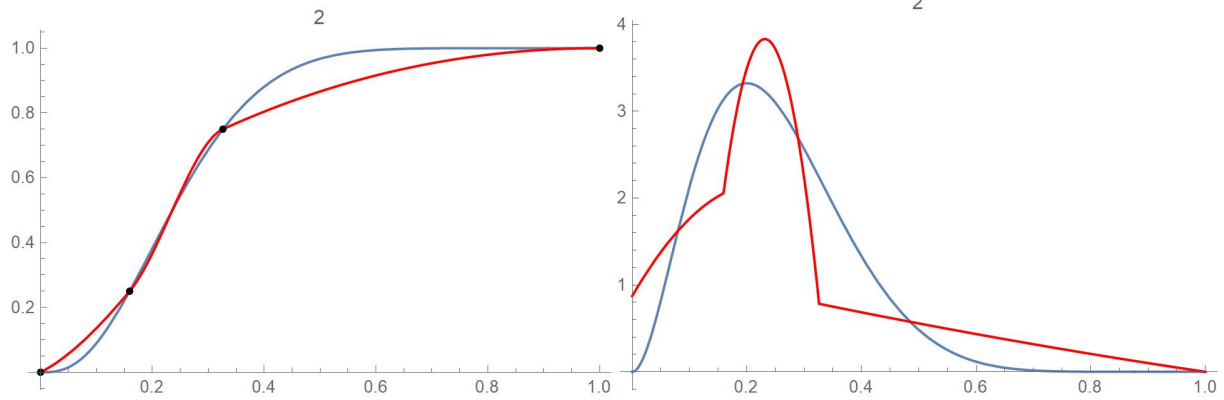
Figure 7.3: Estimate of the skewed beta cdf with 2 representative points

Figure 7.4: Estimate of the skewed beta density with 2 representative points
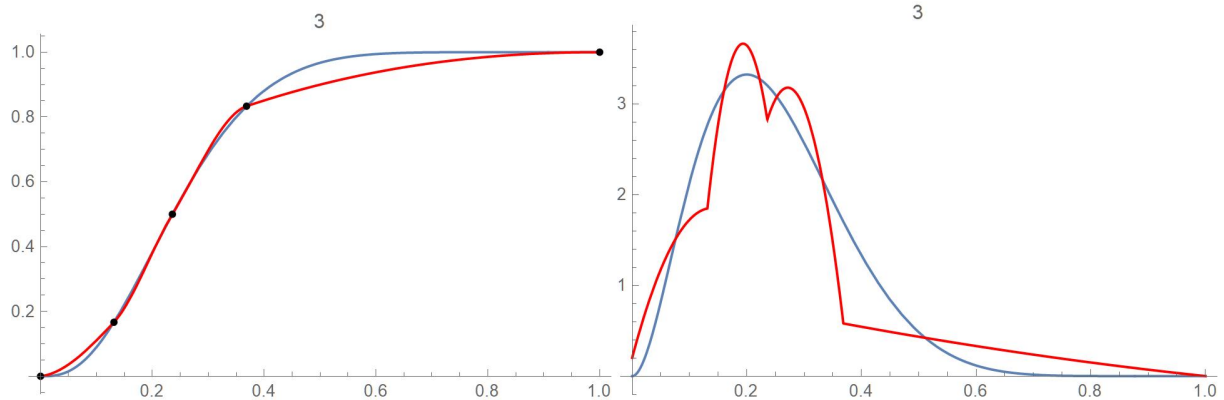
Figure 7.5: Estimate of the skewed beta cdf with 3 representative points

Figure 7.6: Estimate of the skewed beta density with 3 representative points
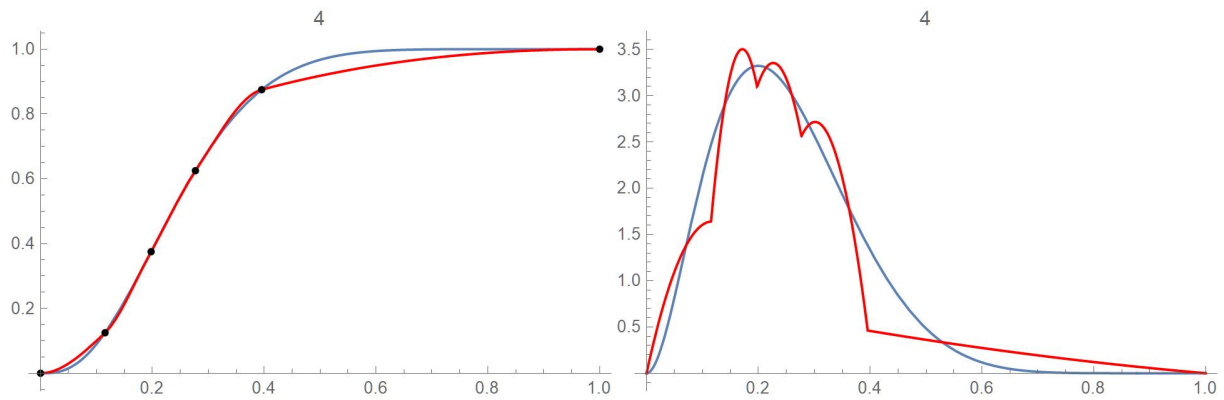
Figure 7.7: Estimate of the skewed beta cdf with 4 representative points

Figure 7.8: Estimate of the skewed beta density with 4 representative points
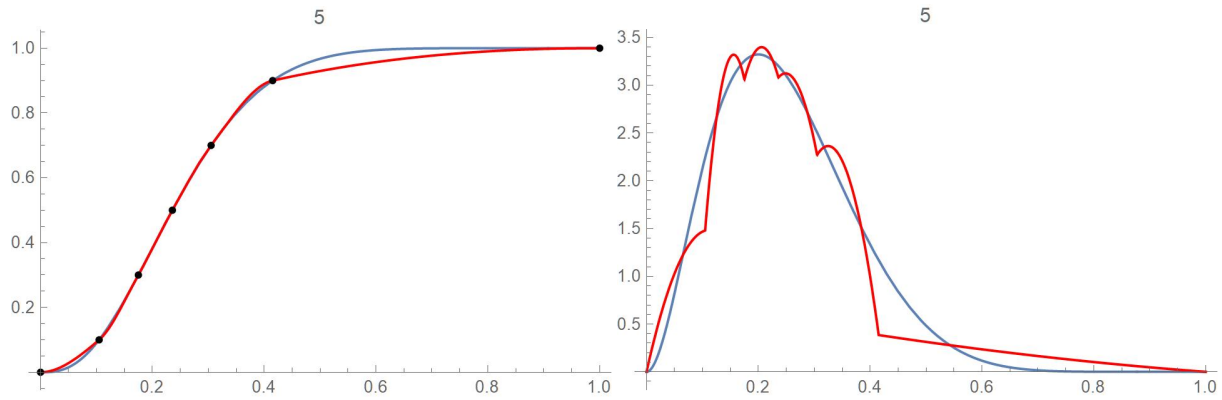
Figure 7.9: Estimate of the skewed beta cdf with 5 representative points

Figure 7.10: Estimate of the skewed beta density with 5 representative points
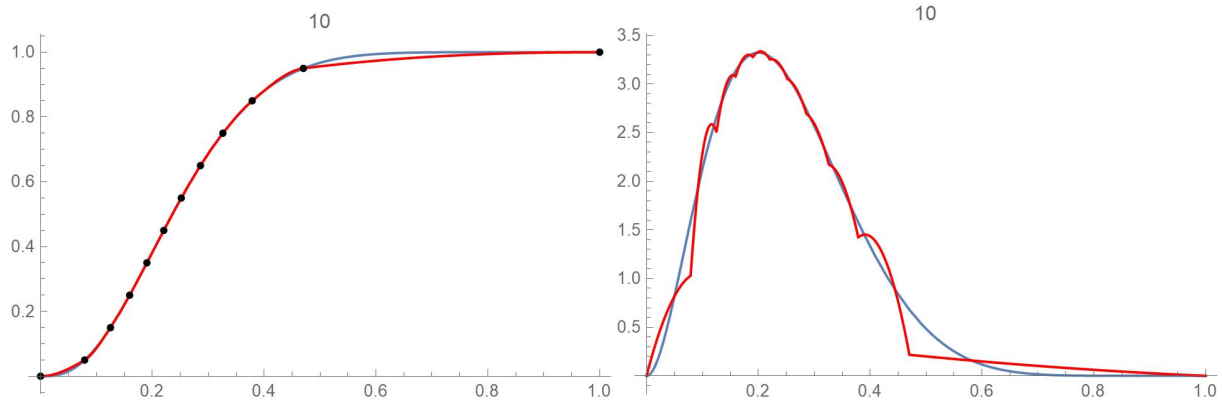
Figure 7.11: Estimate of the skewed beta cdf with 10 representative points

Figure 7.12: Estimate of the skewed beta density with 10 representative points
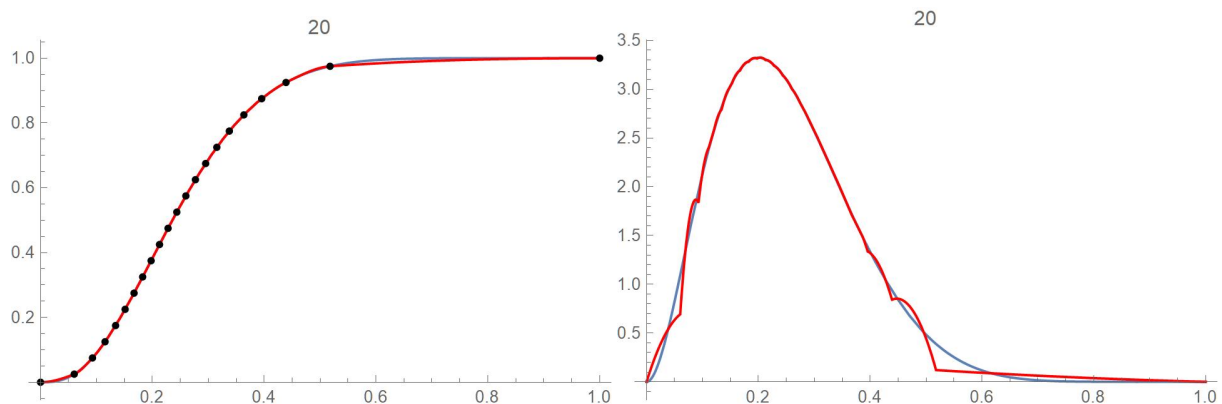
Figure 7.13: Estimate of the skewed beta cdf with 20 representative points

Figure 7.14: Estimate of the skewed beta density with 20 representative points

The proportion of information present in the first h^{th} moments which is $1 - \frac{KLf(h)}{KLfc}$, $h = 2, 3, \ldots$, and that contained in each subsequent moment i.e. $\frac{KLf(h-1) - KLf(h)}{KLfc}$ are respectively plotted in Figures 7.15 and 7.16 and Figures 7.17 and 7.18. The numerical values of the proportion of information present in the first h^{th} moments, $h = 2, 3, \ldots, 20$, i.e., $(1 - \frac{KLf(h)}{KLfc})$, are included in Table 7.1 for both distributions.

For the single skewed data case, the results indicate that most of the information is contained in the first two moments, which makes sense since this a simple unimodal density function. The third moment would have contained relatively more information if the distribution had been more skewed. For the case of a mixture of two beta densities, the results indicate that most of the information is contained in the first four moments, which is also sensible. We note that 95% of the total information is reached with 9 points/moments in the case of the single beta pdf whereas it takes 19 points/moments to attain this proportion in the case of the mixture of beta pdf's.
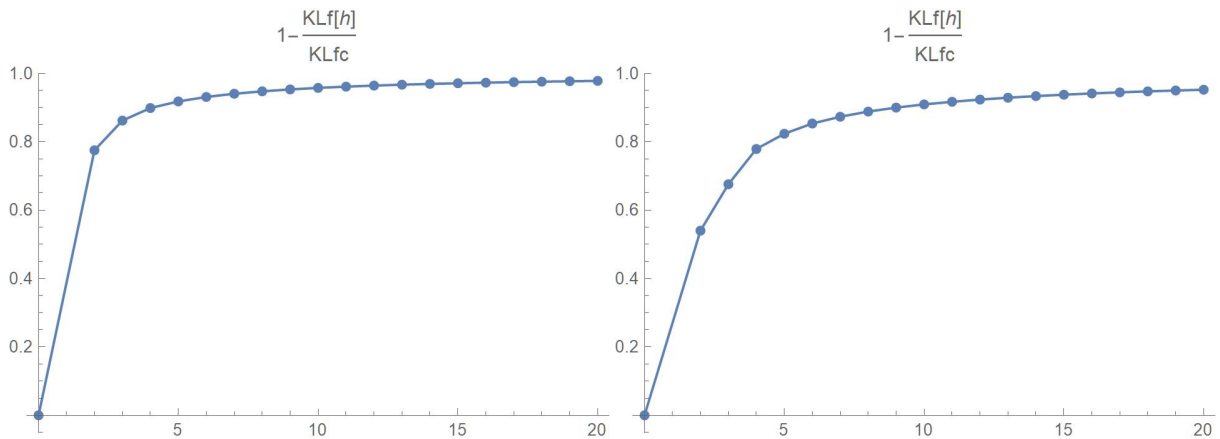
Figure 7.15: Proportion of information in the first h^{th} moments of the skewed beta density

Figure 7.16: Proportion of information in the first h^{th} moments of the mixture of two beta densities
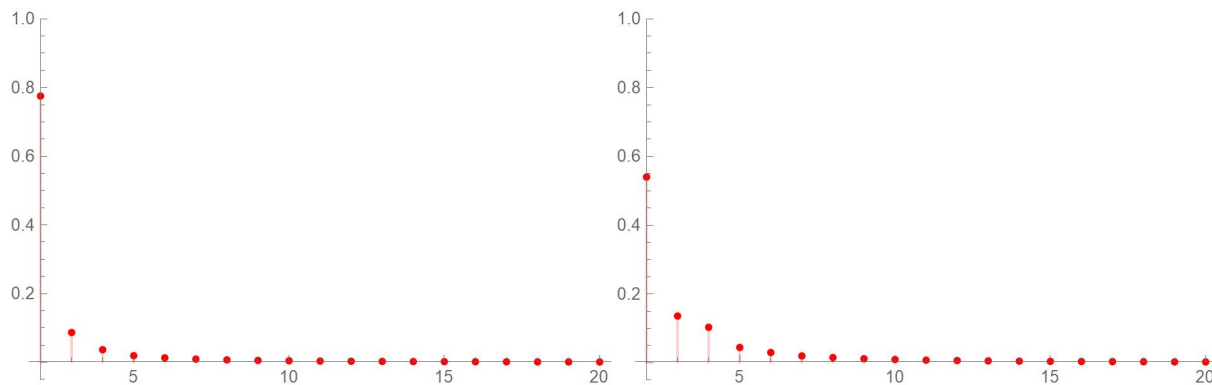
Figure 7.17: Proportion of information held in the first two moments and each subsequent moment of the single skewed beta density

Figure 7.18: Proportion of information held in the first two moments and each subsequent moment of the mixture of two beta densities

h	$1 - \frac{KLf(h)}{KLfc}$ for the skewed beta density	$1 - \frac{KLf(h)}{KLfc}$ for the mixture of two beta densities
2	0.775253	0.540069
3	0.862104	0.675933
4	0.898833	0.779093
5	0.918207	0.823617
6	0.931244	0.853528
7	0.940637	0.873409
8	0.947738	0.888493
9	0.953299	0.900107
10	0.957775	0.909463
11	0.961457	0.917139
12	0.964541	0.923574
13	0.967162	0.929046
14	0.969417	0.933764
15	0.971379	0.937873
16	0.973102	0.941488
17	0.974626	0.944693
18	0.975985	0.947556
19	0.977205	0.950129
20	0.978305	0.952454

Table 7.1: The proportion of information present in the first h^{th} moments for the skewed beta density with $KLfc$ equal to 0.748641 and for the mixture of two beta densities with $KLfc$ equal to 0.400459

7.4.2 Two actual data sets

In this subsection, two real world data sets are considered. The first one is the Buffalo snowfall data, which consists of the snowfall accumulations in inches in Buffalo, New York, from 1910 to 1972. The second one is the life expectancies in 183 countries/regions for the year 2015 as recorded by the World Health Organisation.

First, an estimate of the target distribution, namely a kernel density estimate (KDE) in this instance, is secured for each data sets. These KDE's which were directly obtained from default settings, are shown in Figures 7.19 and 7.20. Then, the end points were chosen to be the values at which the KDE cdf's equal 10^{-4} and $1 - 10^{-4}$. They are 1.70395 and 151.651 for the first data set, and 45.1232 and 93.5663 for the second. For each set comprising the representative points and the end points, an FC interpolating spline was constructed. To visually illustrate the process, plots of the FC cdf interpolants and their derivatives are shown for selected values of h in Figures 7.21-7.32 for the case of the life expectancy data.
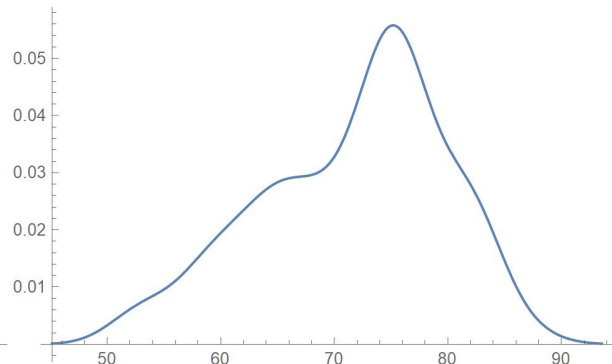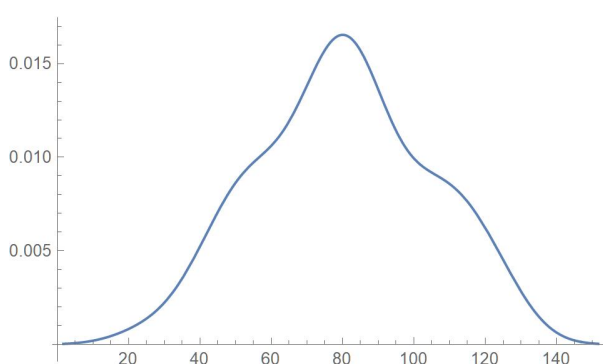
Figure 7.19: KDE of the Buffalo snowfall data   Figure 7.20: KDE of the life expectancy data
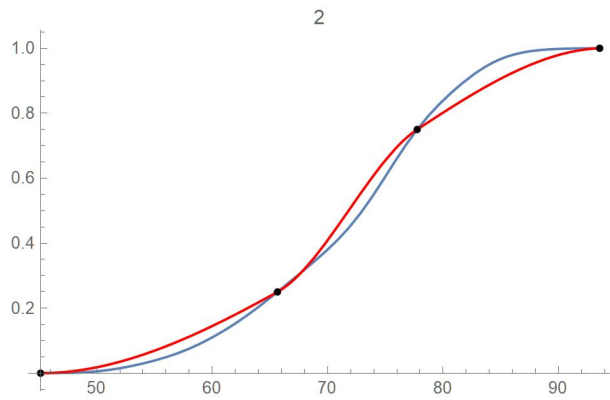
Figure 7.21: CDF estimate of the life ex-
pectancy data with 2 representative points
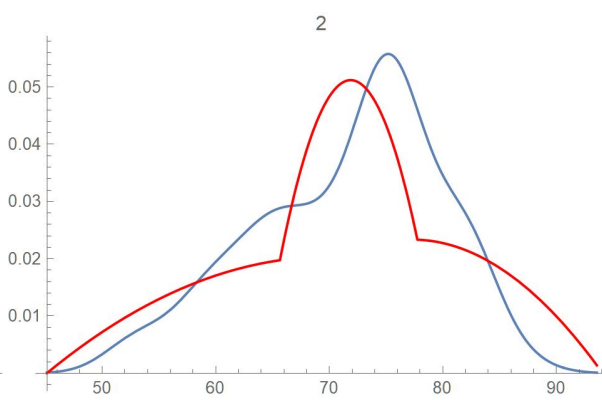
Figure 7.22: Density estimate of the life ex-
pectancy data with 2 representative points
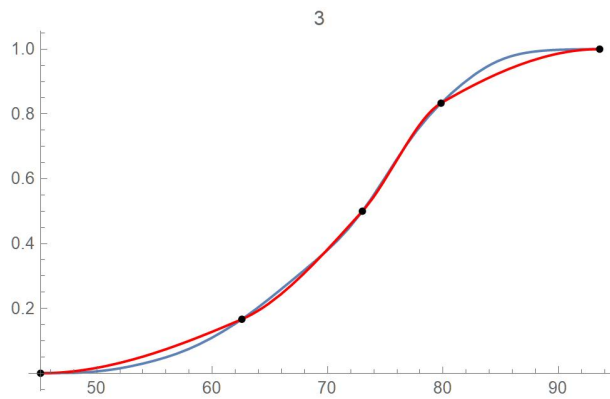
Figure 7.23: CDF estimate of the life ex-
pectancy data with 3 representative points
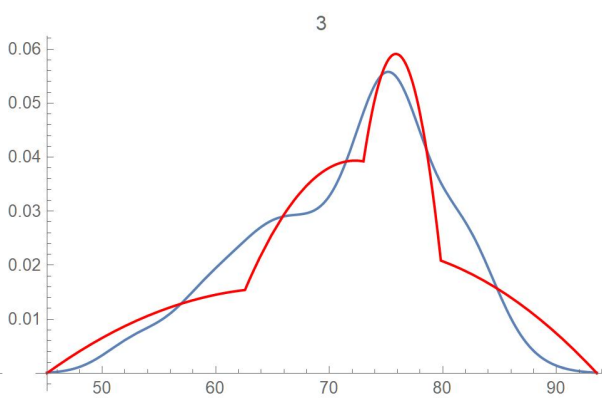
Figure 7.24: Density estimate of the life ex-
pectancy data with 3 representative points
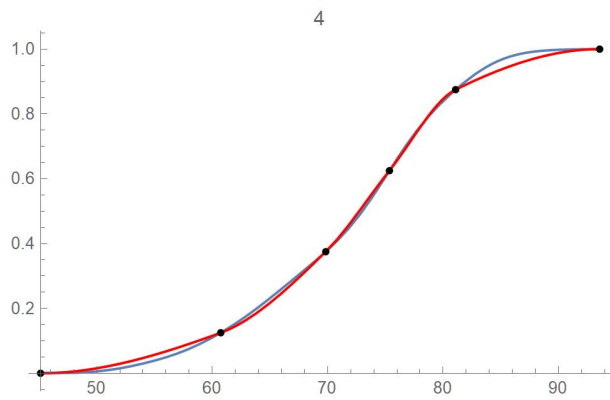
Figure 7.25: CDF estimate of the life ex-
pectancy data with 4 representative points
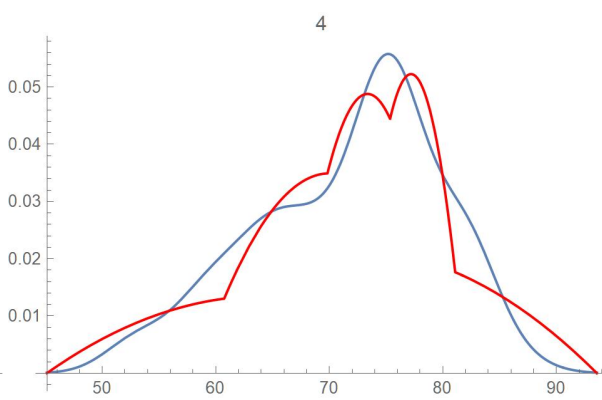
Figure 7.26: Density estimate of the life ex-
pectancy data with 4 representative points
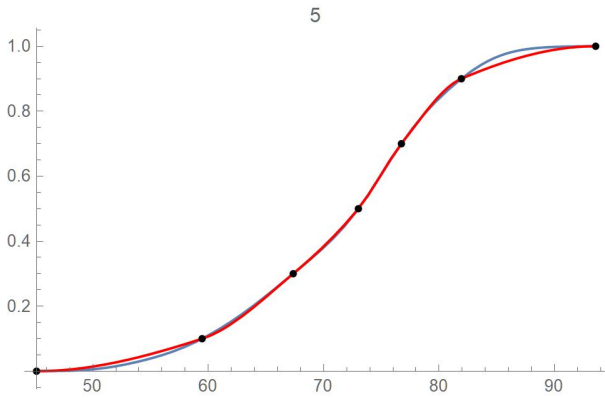
Figure 7.27: CDF estimate of the life expectancy data with 5 representative points
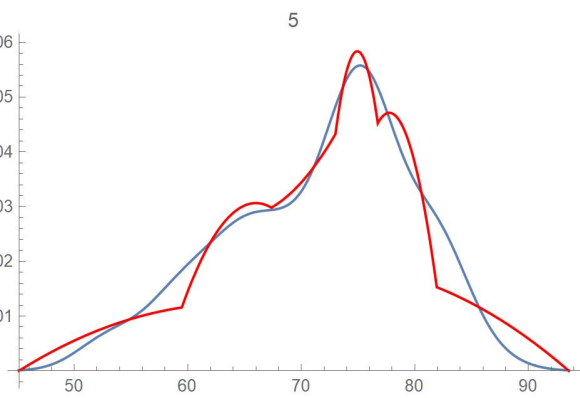
Figure 7.28: Density estimate of the life expectancy data with 5 representative points
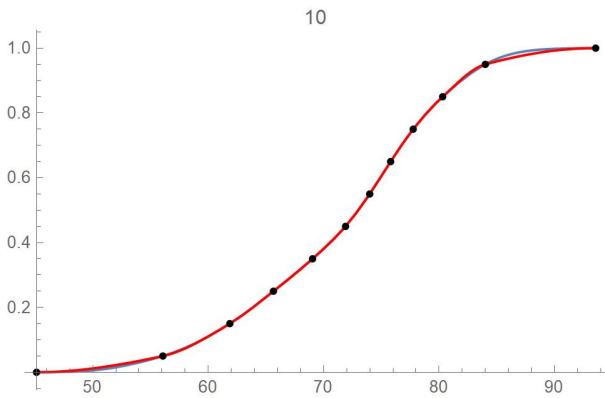
Figure 7.29: CDF estimate of the life expectancy data with 10 representative points
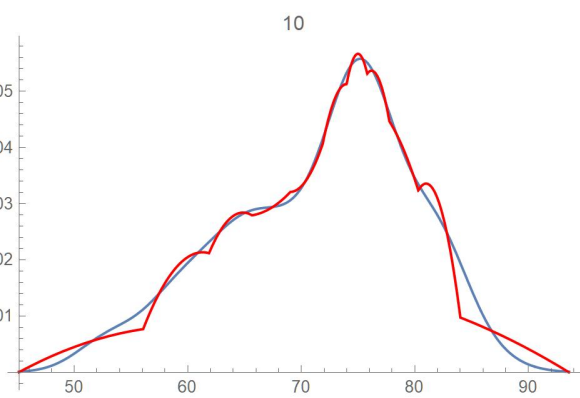
Figure 7.30: Density estimate of the life expectancy data with 10 representative points
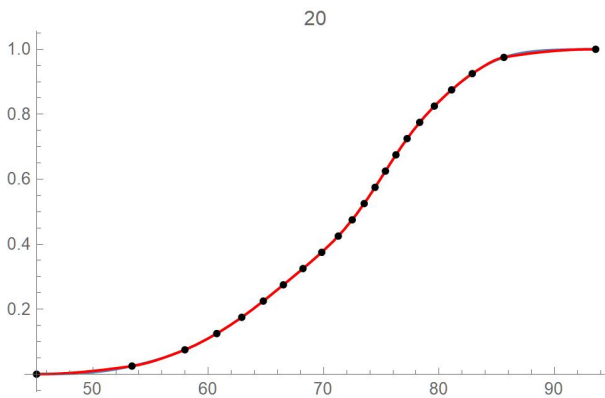
Figure 7.31: CDF estimate of the life expectancy data with 20 representative points
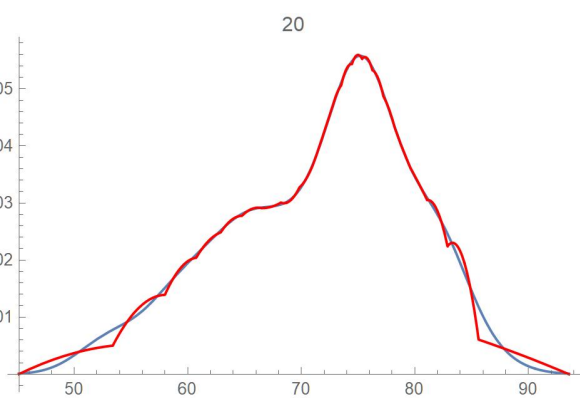
Figure 7.32: Density estimate of the life expectancy data with 20 representative points

The proportion of information present in the first h^{th} moments which is $1 - \frac{KLf(h)}{KLfc}$, $h =$

$2, 3, \ldots$, and that contained in each subsequent moment i.e. $\frac{KLf(h-1) - KLf(h)}{KLfc}$ are respectively plotted in Figures 7.33 and 7.34 and Figures 7.35 and 7.36. The numerical values of the proportion of information present in the first h^{th} moments, $h = 2, 3, \ldots, 20$, i.e., $(1 - \frac{KLf(h)}{KLfc})$, are included in Table 7.2 for both distributions.

As expected the third moment of the life expectancy data which is skewed contains a larger proportion of information than the third moment of the Buffalo snowfall data whose first two moments contain relatively more information.

In all the examples, the proportion of information determined by applying the proposed methodology is consistent with that being expected.
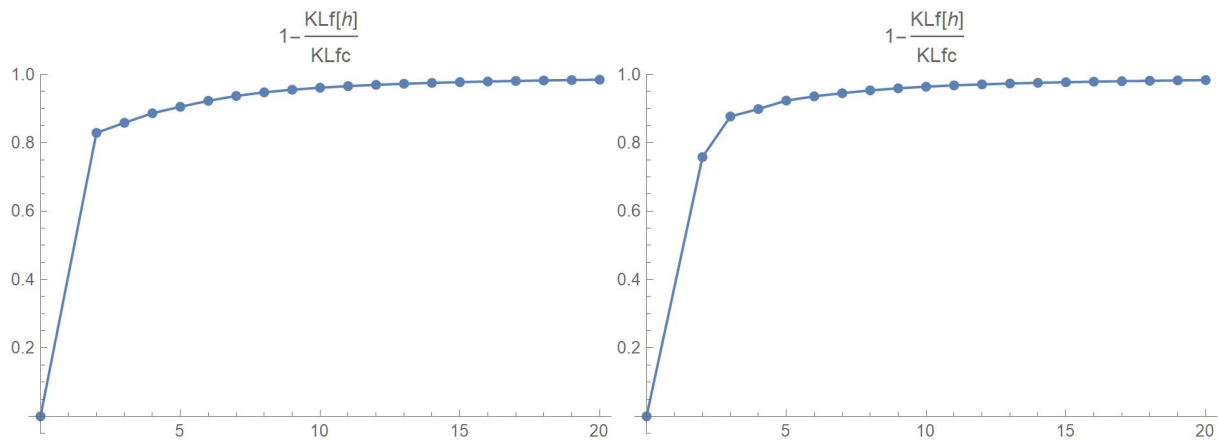
Figure 7.33: Proportion of information in the first h^{th} moments of the Buffalo snowfall data

Figure 7.34: Proportion of information in the first h^{th} moments of the life expectancy data
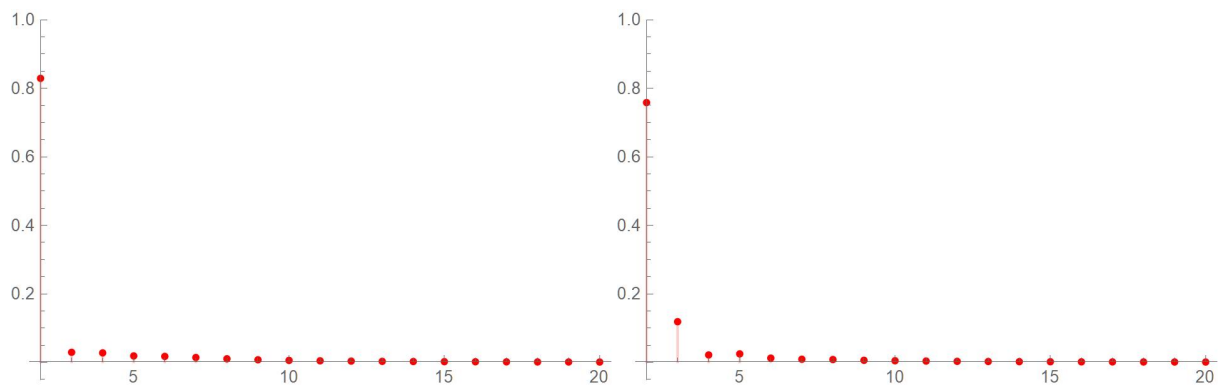
Figure 7.35: Proportion of information held in the first two moments and each subsequent moment of the Buffalo snowfall data

Figure 7.36: Proportion of information held in the first two moments and each subsequent moment of the life expectancy data

h	$1 - \frac{KLf(h)}{KLfc}$ for the Buffalo snowfall data	$1 - \frac{KLf(h)}{KLfc}$ for the life expectancy data
2	0.829169	0.758417
3	0.858646	0.877124
4	0.886463	0.898778
5	0.905378	0.923591
6	0.922889	0.935992
7	0.937178	0.945120
8	0.947716	0.953389
9	0.955368	0.959459
10	0.961226	0.964186
11	0.965893	0.968010
12	0.969669	0.971011
13	0.972754	0.973502
14	0.975338	0.975560
15	0.977521	0.977341
16	0.979390	0.978890
17	0.980992	0.980247
18	0.982397	0.981446
19	0.983639	0.982533
20	0.984737	0.983516

Table 7.2: The proportion of information present in the first h^{th} moments for the Buffalo snowfall data with $KLfc$ equal to 0.373586 and for the life expectancy data with $KLfc$ equal to 0.364324

Bibliography

[1] Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multi-Model Inference*, Second Edition. Springer, New York.

[2] Fang, K.T. and Wang, Y. (1993). *Number Theoretic Methods in Applied Statistics*. Chapman and Hall, London.

[3] Fritsch, F. N. and Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal of Numerical Analysis*, 17(2), 238–246.

[4] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.

[5] Provost, S.B. (2005). Moment-based density approximants, *The Mathematica Journal*, 9, 727–756.

[6] Provost, S. B., Zareamoghaddam, H., Ahmed, S. E., Ha, H.-T. (2022). The generalized Pearson family of distributions and explicit representation of the associated density functions. *Communications in Statistics – Theory and Methods*, 51(16), 5590–5606.

Appendix A

Mathematica Code

The *Mathematica* code utilized for implementing the main numerical examples presented in this dissertation is included in this appendix.

Chapter 2
Copula Approximation and an Application to a Brownian Motion Process

---

Section 2.4
An application : Two stocks' closing prices

Subsection 2.4.1
Moment - based approach to approximating the copula density

1.data

```
GOOG = {1045.85`, 1016.06`, 1070.71`, 1068.39`, 1076.28`, 1074.66`, 1070.33`, 1057.19`,
    1044.69`, 1077.15`, 1080.97`, 1089.9`, 1098.26`, 1070.52`, 1075.57`, 1073.9`,
    1090.99`, 1070.08`, 1060.62`, 1089.06`, 1116.37`, 1110.75`, 1132.8`, 1145.99`,
    1115.23`, 1098.71`, 1095.06`, 1095.01`, 1121.37`, 1120.16`, 1121.67`, 1113.65`,
    1118.56`, 1113.8`, 1096.97`, 1110.37`, 1109.4`, 1115.13`, 1116.05`, 1119.92`,
    1140.99`, 1147.8`, 1162.03`, 1157.86`, 1143.3`, 1142.32`, 1175.76`, 1193.2`,
    1193.32`, 1185.55`, 1184.46`, 1184.26`, 1198.85`, 1223.97`, 1231.54`, 1205.5`,
    1193.`, 1184.62`, 1173.02`, 1168.49`, 1173.31`, 1194.43`, 1200.49`, 1205.92`, 1215.`,
    1207.15`, 1203.84`, 1197.25`, 1202.16`, 1204.62`, 1217.87`, 1221.1`, 1227.13`,
    1236.34`, 1236.37`, 1248.84`, 1264.55`, 1256.`, 1263.45`, 1272.18`, 1287.58`,
    1188.48`, 1168.08`, 1162.61`, 1185.4`, 1189.39`, 1174.1`, 1166.27`, 1162.38`,
    1164.27`, 1132.03`, 1120.44`, 1164.21`, 1178.98`, 1162.3`, 1138.85`, 1149.63`,
    1151.42`, 1140.77`, 1133.47`, 1134.15`, 1116.46`, 1117.95`, 1103.63`, 1036.23`,
    1053.05`, 1042.22`, 1044.34`, 1066.04`, 1080.38`, 1078.72`, 1077.03`, 1088.77`,
    1085.35`, 1092.5`, 1103.6`, 1102.33`, 1111.42`, 1121.88`, 1115.52`, 1086.35`,
    1079.8`, 1076.01`, 1080.91`, 1097.95`, 1111.25`, 1121.58`, 1131.59`, 1116.35`,
    1124.83`, 1140.48`, 1144.21`, 1144.9`, 1150.34`, 1153.58`, 1146.35`, 1146.33`,
    1130.1`, 1138.07`, 1146.21`, 1137.81`, 1132.12`, 1250.41`, 1239.41`, 1225.14`,
    1216.68`, 1209.01`, 1193.99`, 1152.32`, 1169.95`, 1173.99`, 1204.8`, 1188.01`,
    1174.71`, 1197.27`, 1164.29`, 1167.26`, 1177.6`, 1198.45`, 1182.69`, 1191.25`,
    1189.53`, 1151.29`, 1168.89`, 1167.84`, 1171.02`, 1192.85`, 1188.1`, 1168.39`,
    1181.41`, 1211.38`, 1204.93`, 1204.41`, 1206.`, 1220.17`, 1234.25`, 1239.56`,
    1231.3`, 1229.15`, 1232.41`, 1238.71`, 1229.93`, 1234.03`, 1218.76`, 1246.52`,
    1241.39`, 1225.09`, 1219.`, 1205.1`, 1176.63`, 1187.83`, 1209.`, 1207.68`, 1189.13`,
    1202.31`, 1208.67`, 1215.45`, 1217.14`, 1243.01`, 1243.64`, 1253.07`, 1245.49`,
    1246.15`, 1242.8`, 1259.13`, 1260.99`, 1265.13`, 1290.`, 1262.62`, 1261.29`,
    1260.11`, 1273.74`, 1291.37`, 1292.03`, 1291.8`, 1308.86`, 1311.37`, 1299.19`,
    1298.8`, 1298.`, 1311.46`, 1334.87`, 1320.7`, 1315.46`, 1303.05`, 1301.35`, 1295.34`,
    1306.69`, 1313.55`, 1312.99`, 1304.96`, 1289.92`, 1295.28`, 1320.54`, 1328.13`,
    1340.62`, 1343.56`, 1344.66`, 1345.02`, 1350.27`, 1347.83`, 1361.17`, 1355.12`,
    1352.62`, 1356.04`, 1349.59`, 1348.84`, 1343.56`, 1360.4`, 1351.89`, 1336.14`};
```

```
AAPL = {157.92`, 142.19`, 148.26`, 147.93`, 150.75`, 153.31`, 153.8`, 152.29`, 150.`,
    153.07`, 154.94`, 155.86`, 156.82`, 153.3`, 153.92`, 152.7`, 157.76`, 156.3`,
    154.68`, 165.25`, 166.44`, 166.52`, 171.25`, 174.18`, 174.24`, 170.94`, 170.41`,
    169.43`, 170.89`, 170.18`, 170.8`, 170.42`, 170.93`, 172.03`, 171.06`, 172.97`,
    174.23`, 174.33`, 174.87`, 173.15`, 174.97`, 175.85`, 175.53`, 174.52`, 172.5`,
    172.91`, 178.9`, 180.91`, 181.71`, 183.73`, 186.12`, 188.02`, 186.53`, 188.16`,
    195.09`, 191.05`, 188.74`, 186.79`, 188.47`, 188.72`, 189.95`, 191.24`, 194.02`,
    195.35`, 195.69`, 197.`, 200.1`, 199.5`, 200.62`, 198.95`, 198.87`, 199.23`,
    199.25`, 203.13`, 203.86`, 204.53`, 207.48`, 207.16`, 205.28`, 204.3`, 204.61`,
    200.67`, 210.52`, 209.15`, 211.75`, 208.48`, 202.86`, 202.9`, 200.72`, 197.18`,
    185.72`, 188.66`, 190.92`, 190.08`, 189.`, 183.09`, 186.6`, 182.78`, 179.66`,
    178.97`, 178.23`, 177.38`, 178.3`, 175.07`, 173.3`, 179.64`, 182.54`, 185.22`,
    190.15`, 192.58`, 194.81`, 194.19`, 194.15`, 192.74`, 193.89`, 198.45`, 197.87`,
    199.46`, 198.78`, 198.58`, 195.57`, 199.8`, 199.74`, 197.92`, 201.55`, 202.73`,
    204.41`, 204.23`, 200.02`, 201.24`, 203.23`, 201.75`, 203.3`, 205.21`, 204.5`,
    203.35`, 205.66`, 202.59`, 207.22`, 208.84`, 208.67`, 207.02`, 207.74`, 209.68`,
    208.78`, 213.04`, 208.43`, 204.02`, 193.34`, 197.`, 199.04`, 203.43`, 200.99`,
    200.48`, 208.97`, 202.75`, 201.74`, 206.5`, 210.35`, 210.36`, 212.64`, 212.46`,
    202.64`, 206.49`, 204.16`, 205.53`, 209.01`, 208.74`, 205.7`, 209.19`, 213.28`,
    213.26`, 214.17`, 216.7`, 223.59`, 223.09`, 218.75`, 219.9`, 220.7`, 222.77`,
    220.96`, 217.73`, 218.72`, 217.68`, 221.03`, 219.89`, 218.82`, 223.97`, 224.59`,
    218.96`, 220.82`, 227.01`, 227.06`, 224.4`, 227.03`, 230.09`, 236.21`, 235.87`,
    235.32`, 234.37`, 235.28`, 236.41`, 240.51`, 239.96`, 243.18`, 243.58`, 246.58`,
    249.05`, 243.29`, 243.26`, 248.76`, 255.82`, 257.5`, 257.13`, 257.24`, 259.43`,
    260.14`, 262.2`, 261.96`, 264.47`, 262.64`, 265.76`, 267.1`, 266.29`, 263.19`,
    262.01`, 261.78`, 266.37`, 264.29`, 267.84`, 267.25`, 264.16`, 259.45`, 261.74`,
    265.58`, 270.71`, 266.92`, 268.48`, 270.77`, 271.46`, 275.15`, 279.86`,
    280.41`, 279.74`, 280.02`, 279.44`, 284.`, 284.27`, 289.91`, 289.8`, 291.52`};

Print["{Min[GOOG],Max[GOOG]}=", {Min[GOOG], Max[GOOG]}]
Print["{Min[AAPL],Max[AAPL]}=", {Min[AAPL], Max[AAPL]}]
GraphicsRow[{ListPlot[GOOG, Filling → Axis],
  ListPlot[AAPL, AxesOrigin → {0, 120}, Filling → Axis]}, ImageSize → Large]

Y1 = GOOG;  Y2 = AAPL;
Print["Length[Y1]=", n = Length[Y1]]
Print["{Mean[Y1],SD[Y1]}=", {m1 = Mean[Y1], sd1 = StandardDeviation[Y1]}]
Print["{Mean[Y2],SD[Y2]}=", {m2 = Mean[Y2], sd2 = StandardDeviation[Y2]}]

Y = Table[{Y1[[i]], Y2[[i]]}, {i, n}];
{HY = Histogram3D[Y, 15, "Probability", ColorFunction → "Rainbow", ImageSize → Medium],
 ListPlot[Y, PlotStyle → {Purple}, AxesOrigin → {1000, 120}, ImageSize → Medium]}

(*Standardization*)
```

```
X1 = (Y1 - m1)/sd1;

X2 = (Y2 - m2)/sd2;

X = Table[{X1[[i]], X2[[i]]}, {i, n}];
Print["{Min[X1],Max[X1]=", {Min[X1], Max[X1]} // N]
Print["{Min[X2],Max[X2]}=", {Min[X2], Max[X2]} // N]
edis = EmpiricalDistribution[Y];
EmpCDF = Table[CDF[edis, Y[[j]]], {j, 1, n}];
ListPlot[X, PlotStyle → {Blue}]
```

2. The joint density of the two variables

```
𝒟 = SmoothKernelDistribution[X, {"Adaptive", Automatic, .2},
    "Epanechnikov", PerformanceGoal → "Quality"];
{Plot3D[Evaluate@PDF[𝒟, {x, y}], {x, -3, 3}, {y, -3, 3},
  PlotRange → All, ColorFunction → "Rainbow", ImageSize → Medium],
 ListPlot[X, ImageSize → Medium]}

ClearAll[ff]
ff[x_, y_] := Evaluate@PDF[𝒟, {x, y}]

l1 = -3.5; u1 = 3.5; l2 = -3.5; u2 = 3.5;
```

3.a The density function and the inverse cdf of the first random variable, f11 and InvF1a

```
ClearAll[ff1h]
ff1h[x_] :=
 ff1h[x] = NIntegrate[Evaluate@PDF[𝒟, {x, y}], {y, -3.5, 3.5}, AccuracyGoal → 5]

ta1 = Table[{u, ff1h[u]}, {u, -3.5, 3.5, 0.05}];
Table[ta1[[i]], {i, 1, Length[ta1], 5}]

ClearAll[methodfx]
methodfx[t_] := methodfx[t] = (
   ClearAll[f11];
   f11[x_] := Fit[ta1, Table[w^i, {i, 0, t}], w] /. w → x;
   Show[
    {ListPlot[ta1], Plot[f11[u], {u, l1, u1}, PlotStyle → Red]}, PlotLabel → {t}]
  )
Table[methodfx[5 i + j], {i, 1, 4}, {j, 1, 5}] // TableForm
```

```mathematica
(*choose degree=20*)
degree1 = 20;
c11 = NIntegrate[Fit[ta1, Table[w^i, {i, 0, degree1}], w] /. w → x, {x, l1, u1}]
ClearAll[f11, F1a]

f11[x_] := f11[x] = (Fit[ta1, Table[w^i, {i, 0, degree1}], w] /. w → x) / c11

F1a[x_] := F1a[x] = Integrate[f11[u], {u, l1, x}]

f11[x] // N
{Show[{ListPlot[ta1] , Plot[f11[u], {u, l1, u1}, PlotStyle → Red]},
  PlotLabel → {degree1}],
 Plot[F1a[x], {x, l1, u1}, PlotLabel → {"F1a"}]}

ta11 = Table[{F1a[x], x}, {x, l1, u1, 0.2}];
Table[ta11[[i]], {i, 1, Length[ta11], 5}]

it1 = Interpolation[ta11, InterpolationOrder → 1];
Show[Plot[it1[x], {x, 0, 1}, PlotStyle → Red], ListPlot[ta11]]

ClearAll[methodfx1]

methodfx1[t_] := methodfx1[t] = (

   ClearAll[mo, InvF1a];
   mo[h_] := mo[h] = Rationalize[NIntegrate[x^h * it1[x], {x, 0, 1}], 10^-300];
   M2 = Table[Moment[UniformDistribution[{0, 1}], i + j], {i, 0, t}, {j, 0, t}];
   μ2 = Table[mo[i], {i, 0, t}];
   coe2 = Inverse[M2].μ2;

   InvF1a[y_] := InvF1a[y] = Sum[coe2[[i]] y^(i-1), {i, 1, t+1}];

   Show[{Plot[InvF1a[w], {w, 0, 1}, PlotStyle → Red, PlotRange → All],
     ListPlot[ta11]}, PlotLabel → {t}]

   )

Table[methodfx1[5 i + j], {i, 1, 3}, {j, 1, 5}] // TableForm
```

156

```
(*choose degree=15*)
t = 15;
ClearAll[mo, InvF1a];
mo[h_] := mo[h] = Rationalize[NIntegrate[x^h * it1[x], {x, 0, 1}], 10^-300];
M2 = Table[Moment[UniformDistribution[{0, 1}], i + j], {i, 0, t}, {j, 0, t}];
μ2 = Table[mo[i], {i, 0, t}];
coe2 = Inverse[M2].μ2;

InvF1a[y_] := InvF1a[y] = ∑_{i=1}^{t+1} coe2[[i]] y^{i-1};


InvF1a[x] // N
Show[{Plot[InvF1a[w], {w, 0, 1}, PlotStyle → Red, PlotRange → All], ListPlot[ta11]},
  PlotLabel → {t}]
```

3.bThe density function and the inverse cdf of the second random variable, f22 and InvF2a

```
ClearAll[ff2h]
ff2h[y_] :=
  ff2h[y] = NIntegrate[Evaluate@PDF[𝒟, {x, y}], {x, -3.5, 3.5}, AccuracyGoal → 5]

ta2 = Table[{u, ff2h[u]}, {u, -3.5, 3.5, 0.1}];
Table[ta2[[i]], {i, 1, Length[ta2], 5}]

{{-3.5, 0.}, {-3., 0.}, {-2.5, 0.}, {-2., 0.018962}, {-1.5, 0.144438},
 {-1., 0.321815}, {-0.5, 0.376445}, {0., 0.566495}, {0.5, 0.186357}, {1., 0.103897},
 {1.5, 0.189698}, {2., 0.102735}, {2.5, 0.0284651}, {3., 0.}, {3.5, 0.}}

ClearAll[methodfy]
methodfy[t_] := methodfy[t] = (
   ClearAll[f22];
   f22[x_] := Fit[ta2, Table[w^i, {i, 0, t}], w] /. w → x;
   Show[
     {ListPlot[ta2], Plot[f22[u], {u, l1, u1}, PlotStyle → Red]}, PlotLabel → {t}]
  )
Table[methodfy[5 i + j], {i, 1, 4}, {j, 1, 5}] // TableForm
```

```
(*choose degree=20*)
degree2 = 20;
c22 = NIntegrate[Fit[ta2, Table[w^i, {i, 0, degree2}], w] /. w → x, {x, l1, u1}]
ClearAll[f22, F2a]

f22[x_] := f22[x] = (Fit[ta2, Table[w^i, {i, 0, degree2}], w] /. w → x) / c22

F2a[x_] := F2a[x] = Integrate[f22[u], {u, l1, x}]


f22[x] // N
{Show[{ListPlot[ta2] , Plot[f22[u], {u, l1, u1}, PlotStyle → Red]},
   PlotLabel → {degree2}],
 Plot[F2a[x], {x, l1, u1}, PlotLabel → {"F2a"}]]}

ta22 = Table[{F2a[x], x}, {x, l1, u1, 0.2}];
Table[ta22[[i]], {i, 1, Length[ta22], 5}]

it2 = Interpolation[ta22, InterpolationOrder → 1];
Show[Plot[it2[x], {x, 0, 1}, PlotStyle → Red], ListPlot[ta22]]

ClearAll[methodfy2]

methodfy2[t_] := methodfy2[t] = (

   ClearAll[mo, InvF2a];
   mo[h_] := mo[h] = Rationalize[NIntegrate[x^h * it2[x], {x, 0, 1}], 10^-300];
   M2 = Table[Moment[UniformDistribution[{0, 1}], i + j], {i, 0, t}, {j, 0, t}];
   μ2 = Table[mo[i], {i, 0, t}];
   coe2 = Inverse[M2].μ2;

   InvF2a[y_] := InvF2a[y] = Sum_{i=1}^{t+1} coe2[[i]] y^{i-1};

   Show[{Plot[InvF2a[w], {w, 0, 1}, PlotStyle → Red, PlotRange → All],
     ListPlot[ta22]}, PlotLabel → {t}]

   )

Table[methodfy2[5 i + j], {i, 1, 3}, {j, 1, 5}] // TableForm
```

```
(*choose degree=15*)
t = 15;
ClearAll[mo, InvF2a];
mo[h_] := mo[h] = Rationalize[NIntegrate[x^h * it2[x], {x, 0, 1}], 10^-300];
M2 = Table[Moment[UniformDistribution[{0, 1}], i + j], {i, 0, t}, {j, 0, t}];
μ2 = Table[mo[i], {i, 0, t}];
coe2 = Inverse[M2].μ2;

InvF2a[y_] := InvF2a[y] = ∑_{i=1}^{t+1} coe2[[i]] y^{i-1};


InvF2a[x] // N
Show[{Plot[InvF2a[w], {w, 0, 1}, PlotStyle → Red, PlotRange → All], ListPlot[ta22]},
 PlotLabel → {t}]
```

4. The copula density

```
ClearAll[cpdfa]
cpdfa[r_, s_] :=
  cpdfa[r, s] = (ff[InvF1a[r], InvF2a[s]]) / (f11[InvF1a[r]] * f22[InvF2a[s]]);

pcpdfa = ListPlot3D[Flatten[Table[{r, s, cpdfa[r, s]},
    {r, 2 / 100, 96 / 100, 1 / 100}, {s, 2 / 100, 96 / 100, 1 / 100}], 1],
  ColorFunction → "Rainbow", PlotRange → {-.3, 12}, ImageSize → Medium]
```

Bernstein' s empirical copula has very similar features

5. Approximating the copula density with a moment - based bivariate polynomial

```
ClearAll[jm, f3, mm2]
jm[m_, n_] := jm[m, n] = Rationalize[NIntegrate[r^m * s^n * cpdfa[r, s],
    {r, 0, 1}, {s, 0, 1}, Method → "QuasiMonteCarlo"], 10^-10];
Needs["MultivariateStatistics`"];
Off[MLE::shdw]; Off[Inner::"normal"];
Off[NIntegrate::izero]; Off[NIntegrate::"slwcon"];
f3[L1_List, L2_List] := Inner[Plus, L1, L2, List];

mm2[r_, s_] := mm2[r, s] = 1 / ((r + 1) (s + 1));

(*mm2[r_,s_]:=mm2[r,s]=Moment[UniformDistribution[{{0,1},{0,1}}],{r,s}]*)
mm2[0, 0] = 1;

(*Show some degrees*)
```

```
ClearAll[method]
method[t_] := method[t] = (
    L3 = Flatten[Table[{j, i}, {i, 0, t}, {j, 0, t}], 1];
    P3 = Table[f3[L3[[i]], L3[[j]]], {i, 1, (t + 1)^2}, {j, 1, (t + 1)^2}];
    M4 = Rationalize[Table[mm2[P3[[i, j]][[1]], P3[[i, j]][[2]]],
        {i, Length[L3]}, {j, Length[L3]}], 10^-25];


    ClearAll[Zv1, Gms, t50, t5];
    Zv1[x_, y_] := Zv[x, y] = Flatten[Table[x^j y^i, {i, 0, t}, {j, 0, t}], 1];
    Gms[i_] := Gms[i] = jm[L3[[i, 1]], L3[[i, 2]]];
    μ = Table[Gms[i], {i, Dimensions[L3][[1]]}];
    c4 = Inverse[M4].μ;
    t50[x_, y_] := t50[x, y] = c4.Zv1[x, y];
    (*remove negative parts and renormalize*)
    Print["tc=",
     tc = NIntegrate[If[t50[x, y] > 0, t50[x, y], 0], {x, 0, 1}, {y, 0, 1}]];
    t5[x_, y_] := t5[x, y] = If[t50[x, y] > 0, t50[x, y], 0] / tc;
    {ListPlot3D[Flatten[Table[{r, s, t5[r, s]}, {r, 1 / 100, 99 / 100, 1 / 100},
        {s, 1 / 100, 99 / 100, 1 / 100}], 1], ColorFunction → "Rainbow",
      PlotLabel → {t}, PlotRange → {All, 12}, ImageSize → Medium], pcpdfa}
  )

result = AbsoluteTiming[Quiet[method[2]]];
Print["AbsoluteTiming=", result[[1]]]
result[[2]]
```

Repeat with other degrees, degree 7 is selected as the optimal

```
result = AbsoluteTiming[Quiet[method[7]]];
Print["AbsoluteTiming=", result[[1]]]
result[[2]]
```

Eliminate the unwanted fluctuations

Obtain a parallelogram support, based on the centered copula points

```
SeedRandom[0];
X1r = X1 + Flatten[Table[RandomReal[-1, 1], {i, 1, Length[X1]}] * 10^-8];
X2r = X2 + Flatten[Table[RandomReal[-1, 1], {i, 1, Length[X2]}] * 10^-8];
DuplicateFreeQ[X1r]
DuplicateFreeQ[X2r]

p1 = Position[Sort[X1r], #] & /@ X1r // Flatten;
p2 = Position[Sort[X2r], #] & /@ X2r // Flatten;

n = Length[X1];
ccp = Table[{{1/n p1[[i]] - 1/(2 n), 1/n p2[[i]] - 1/(2 n)}, {i, n}];
ListPlot[ccp]
```

2. Obtain the parallelogram support of the ccps

160

```mathematica
me1 = ∑_{j=1}^{n} ccp[[j, 1]] / n;

me2 = ∑_{j=1}^{n} ccp[[j, 2]] / n;

Var = 1/(n - 1) {{Sum[(ccp[[j, 1]] - me1)^2, {j, n}],
    Sum[(ccp[[j, 1]] - me1) (ccp[[j, 2]] - me2), {j, n}]},
   {Sum[(ccp[[j, 1]] - me1) (ccp[[j, 2]] - me2), {j, n}],
    Sum[(ccp[[j, 2]] - me2)^2, {j, n}]}}};

Vhi = MatrixPower[Var, -1 / 2];
Print["Vhi=", Vhi // MatrixForm]
Vh = MatrixPower[Var, 1 / 2];
Print["Vh=", Vh // MatrixForm]
detVhi = Det[Vhi] // N;
V1 = Var[[1, 1]];
V2 = Var[[2, 2]];

sccp1 = Vhi[[1, 1]] (ccp[[All, 1]] - ConstantArray[me1, n]) +
    Vhi[[1, 2]] (ccp[[All, 2]] - ConstantArray[me2, n]) // N;
sccp2 = Vhi[[2, 1]] (ccp[[All, 1]] - ConstantArray[me1, n]) +
    Vhi[[2, 2]] (ccp[[All, 2]] - ConstantArray[me2, n]) // N;
sccp = Table[{sccp1[[i]], sccp2[[i]]}, {i, n}];
ksccp1 = SmoothKernelDistribution[sccp1, "Silverman"];
ksccp2 = SmoothKernelDistribution[sccp2, "Silverman"];
pk1 = Plot[PDF[ksccp1, x], {x, -4, 4}, PlotRange → All];
pk2 = Plot[PDF[ksccp2, x], {x, -4, 4}, PlotRange → All];
{pk1, pk2}
{a1, a2} = {Round[FindRoot[PDF[ksccp1, x] == 0.001, {x, -2.5}][[1, 2]], 0.01],
   Round[FindRoot[PDF[ksccp1, x] == 0.001, {x, 2}][[1, 2]], 0.01]}

{b1, b2} = {Round[FindRoot[PDF[ksccp2, x] == 0.001, {x, -2}][[1, 2]], 0.01],
   Round[FindRoot[PDF[ksccp2, x] == 0.001, {x, 1.5}][[1, 2]], 0.01]}

s11 = Vh[[1, 1]];
s12 = Vh[[1, 2]];
s21 = Vh[[2, 1]];
s22 = Vh[[2, 2]];
```

161

```
Clear[s1, s2, s3, s4]
s1[x_] := s22/s12 x - (s11 * s22 - s12^2) a1/s12 + me2 - s22/s12 me1

s2[x_] := s12/s11 x - (s11 * s22 - s12^2) b1/s11 + me2 - s12/s11 me1

s3[x_] := s22/s12 x - (s11 * s22 - s12^2) a2/s12 + me2 - s22/s12 me1

s4[x_] := s12/s11 x - (s11 * s22 - s12^2) b2/s11 + me2 - s12/s11 me1

ps1 = Plot[s1[x], {x, -2.5, 2.5}, PlotStyle → Red];
ps2 = Plot[s2[x], {x, -2.5, 2.5}, PlotStyle → Blue];
ps3 = Plot[s3[x], {x, -2.5, 2.5}, PlotStyle → Green];
ps4 = Plot[s4[x], {x, -2.5, 2.5}, PlotStyle → Orange];
Show[ps1, ps2, ps3, ps4]

Clear[DI]
DI[x_, y_] := If[y ≤ s1[x] && y ≤ s2[x] && y ≥ s3[x] && y ≥ s4[x], 1, 0]
Plot3D[DI[x, y], {x, -1, 2}, {y, -1, 2}, Mesh → All,
 PlotStyle → Cyan(*,PlotPoints→50,MaxRecursion→10*)]

t = 7;
L3 = Flatten[Table[{j, i}, {i, 0, t}, {j, 0, t}], 1];
P3 = Table[f3[L3[[i]], L3[[j]]], {i, 1, (t + 1)^2}, {j, 1, (t + 1)^2}];
M4 = Rationalize[Table[mm2[P3[[i, j]][[1]], P3[[i, j]][[2]]],
    {i, Length[L3]}, {j, Length[L3]}], 10^-25];


ClearAll[Zv1, Gms, t50, t5];
Zv1[x_, y_] := Zv[x, y] = Flatten[Table[x^j y^i, {i, 0, t}, {j, 0, t}], 1];
Gms[i_] := Gms[i] = jm[L3[[i, 1]], L3[[i, 2]]];
μ = Table[Gms[i], {i, Dimensions[L3][[1]]}];
c4 = Inverse[M4].μ;
t50[x_, y_] := t50[x, y] = c4.Zv1[x, y];
(*remove negative parts and renormalize*)
(*Print["tc=",tc=NIntegrate[If[t50[x,y]>0,t50[x,y],0],{x,0,1},{y,0,1}]];*)
t5[x_, y_] := t5[x, y] = If[t50[x, y] > 0, t50[x, y], 0](*/tc*);
{ListPlot3D[Flatten[Table[{r, s, t5[r, s]}, {r, 2 / 100, 96 / 100, 1 / 100},
    {s, 2 / 100, 96 / 100, 1 / 100}], 1], ColorFunction → "Rainbow",
  PlotLabel → {t}, PlotRange → All, ImageSize → Medium], pcpdfa}

ListPlot3D[Flatten[Table[{r, s, t5[r, s] * DI[r, s]},
   {r, 2 / 100, 96 / 100, 1 / 100}, {s, 2 / 100, 96 / 100, 1 / 100}], 1],
 ColorFunction → "Rainbow", PlotLabel → {t}, PlotRange → {0, 12}, ImageSize → Medium]
```

```
l1 = -3.5; u1 = 3.5; l2 = -3.5; u2 = 3.5;
k1 = SmoothKernelDistribution[X1, .6, "Triweight"];
k2 = SmoothKernelDistribution[X2, .6, "Triweight"];

ClearAll[f11, f22, InvF1a, InvF2a]
f11[x_] := PDF[k1, x]
f22[x_] := PDF[k2, x]
InvF1a[r_] := InverseFunction[CDF[k1, #] &][r]
InvF2a[s_] := InverseFunction[CDF[k2, #] &][s]

{Plot[PDF[k1, x], {x, l1, u1}],
 Plot[PDF[k2, y], {y, l2, u2}]}

{Plot[InvF1a[r], {r, 0, 1}],
 Plot[InvF2a[r], {r, 0, 1}]}

ClearAll[cpdfa]
cpdfa[r_, s_] :=
  cpdfa[r, s] = (ff[InvF1a[r], InvF2a[s]]) / (f11[InvF1a[r]] * f22[InvF2a[s]]);

ListPlot3D[Flatten[Table[{r, s, cpdfa[r, s]},
    {r, 2 / 100, 96 / 100, 1 / 100}, {s, 2 / 100, 96 / 100, 1 / 100}], 1],
 PlotRange → All, ImageSize → Medium, ColorFunction → "Rainbow"]
```

Section 2.5

Parallelograms as domains for joint density functions

1. data

GOOG = {1045.85`, 1016.06`, 1070.71`, 1068.39`, 1076.28`, 1074.66`, 1070.33`, 1057.19`,
    1044.69`, 1077.15`, 1080.97`, 1089.9`, 1098.26`, 1070.52`, 1075.57`, 1073.9`,
    1090.99`, 1070.08`, 1060.62`, 1089.06`, 1116.37`, 1110.75`, 1132.8`, 1145.99`,
    1115.23`, 1098.71`, 1095.06`, 1095.01`, 1121.37`, 1120.16`, 1121.67`, 1113.65`,
    1118.56`, 1113.8`, 1096.97`, 1110.37`, 1109.4`, 1115.13`, 1116.05`, 1119.92`,
    1140.99`, 1147.8`, 1162.03`, 1157.86`, 1143.3`, 1142.32`, 1175.76`, 1193.2`,
    1193.32`, 1185.55`, 1184.46`, 1184.26`, 1198.85`, 1223.97`, 1231.54`, 1205.5`,
    1193.`, 1184.62`, 1173.02`, 1168.49`, 1173.31`, 1194.43`, 1200.49`, 1205.92`, 1215.`,
    1207.15`, 1203.84`, 1197.25`, 1202.16`, 1204.62`, 1217.87`, 1221.1`, 1227.13`,
    1236.34`, 1236.37`, 1248.84`, 1264.55`, 1256.`, 1263.45`, 1272.18`, 1287.58`,
    1188.48`, 1168.08`, 1162.61`, 1185.4`, 1189.39`, 1174.1`, 1166.27`, 1162.38`,
    1164.27`, 1132.03`, 1120.44`, 1164.21`, 1178.98`, 1162.3`, 1138.85`, 1149.63`,
    1151.42`, 1140.77`, 1133.47`, 1134.15`, 1116.46`, 1117.95`, 1103.63`, 1036.23`,
    1053.05`, 1042.22`, 1044.34`, 1066.04`, 1080.38`, 1078.72`, 1077.03`, 1088.77`,
    1085.35`, 1092.5`, 1103.6`, 1102.33`, 1111.42`, 1121.88`, 1115.52`, 1086.35`,
    1079.8`, 1076.01`, 1080.91`, 1097.95`, 1111.25`, 1121.58`, 1131.59`, 1116.35`,
    1124.83`, 1140.48`, 1144.21`, 1144.9`, 1150.34`, 1153.58`, 1146.35`, 1146.33`,
    1130.1`, 1138.07`, 1146.21`, 1137.81`, 1132.12`, 1250.41`, 1239.41`, 1225.14`,
    1216.68`, 1209.01`, 1193.99`, 1152.32`, 1169.95`, 1173.99`, 1204.8`, 1188.01`,
    1174.71`, 1197.27`, 1164.29`, 1167.26`, 1177.6`, 1198.45`, 1182.69`, 1191.25`,
    1189.53`, 1151.29`, 1168.89`, 1167.84`, 1171.02`, 1192.85`, 1188.1`, 1168.39`,
    1181.41`, 1211.38`, 1204.93`, 1204.41`, 1206.`, 1220.17`, 1234.25`, 1239.56`,
    1231.3`, 1229.15`, 1232.41`, 1238.71`, 1229.93`, 1234.03`, 1218.76`, 1246.52`,
    1241.39`, 1225.09`, 1219.`, 1205.1`, 1176.63`, 1187.83`, 1209.`, 1207.68`, 1189.13`,
    1202.31`, 1208.67`, 1215.45`, 1217.14`, 1243.01`, 1243.64`, 1253.07`, 1245.49`,
    1246.15`, 1242.8`, 1259.13`, 1260.99`, 1265.13`, 1290.`, 1262.62`, 1261.29`,
    1260.11`, 1273.74`, 1291.37`, 1292.03`, 1291.8`, 1308.86`, 1311.37`, 1299.19`,
    1298.8`, 1298.`, 1311.46`, 1334.87`, 1320.7`, 1315.46`, 1303.05`, 1301.35`, 1295.34`,
    1306.69`, 1313.55`, 1312.99`, 1304.96`, 1289.92`, 1295.28`, 1320.54`, 1328.13`,
    1340.62`, 1343.56`, 1344.66`, 1345.02`, 1350.27`, 1347.83`, 1361.17`, 1355.12`,
    1352.62`, 1356.04`, 1349.59`, 1348.84`, 1343.56`, 1360.4`, 1351.89`, 1336.14`};

```
AAPL = {157.92`, 142.19`, 148.26`, 147.93`, 150.75`, 153.31`, 153.8`, 152.29`, 150.`,
    153.07`, 154.94`, 155.86`, 156.82`, 153.3`, 153.92`, 152.7`, 157.76`, 156.3`,
    154.68`, 165.25`, 166.44`, 166.52`, 171.25`, 174.18`, 174.24`, 170.94`, 170.41`,
    169.43`, 170.89`, 170.18`, 170.8`, 170.42`, 170.93`, 172.03`, 171.06`, 172.97`,
    174.23`, 174.33`, 174.87`, 173.15`, 174.97`, 175.85`, 175.53`, 174.52`, 172.5`,
    172.91`, 178.9`, 180.91`, 181.71`, 183.73`, 186.12`, 188.02`, 186.53`, 188.16`,
    195.09`, 191.05`, 188.74`, 186.79`, 188.47`, 188.72`, 189.95`, 191.24`, 194.02`,
    195.35`, 195.69`, 197.`, 200.1`, 199.5`, 200.62`, 198.95`, 198.87`, 199.23`,
    199.25`, 203.13`, 203.86`, 204.53`, 207.48`, 207.16`, 205.28`, 204.3`, 204.61`,
    200.67`, 210.52`, 209.15`, 211.75`, 208.48`, 202.86`, 202.9`, 200.72`, 197.18`,
    185.72`, 188.66`, 190.92`, 190.08`, 189.`, 183.09`, 186.6`, 182.78`, 179.66`,
    178.97`, 178.23`, 177.38`, 178.3`, 175.07`, 173.3`, 179.64`, 182.54`, 185.22`,
    190.15`, 192.58`, 194.81`, 194.19`, 194.15`, 192.74`, 193.89`, 198.45`, 197.87`,
    199.46`, 198.78`, 198.58`, 195.57`, 199.8`, 199.74`, 197.92`, 201.55`, 202.73`,
    204.41`, 204.23`, 200.02`, 201.24`, 203.23`, 201.75`, 203.3`, 205.21`, 204.5`,
    203.35`, 205.66`, 202.59`, 207.22`, 208.84`, 208.67`, 207.02`, 207.74`, 209.68`,
    208.78`, 213.04`, 208.43`, 204.02`, 193.34`, 197.`, 199.04`, 203.43`, 200.99`,
    200.48`, 208.97`, 202.75`, 201.74`, 206.5`, 210.35`, 210.36`, 212.64`, 212.46`,
    202.64`, 206.49`, 204.16`, 205.53`, 209.01`, 208.74`, 205.7`, 209.19`, 213.28`,
    213.26`, 214.17`, 216.7`, 223.59`, 223.09`, 218.75`, 219.9`, 220.7`, 222.77`,
    220.96`, 217.73`, 218.72`, 217.68`, 221.03`, 219.89`, 218.82`, 223.97`, 224.59`,
    218.96`, 220.82`, 227.01`, 227.06`, 224.4`, 227.03`, 230.09`, 236.21`, 235.87`,
    235.32`, 234.37`, 235.28`, 236.41`, 240.51`, 239.96`, 243.18`, 243.58`, 246.58`,
    249.05`, 243.29`, 243.26`, 248.76`, 255.82`, 257.5`, 257.13`, 257.24`, 259.43`,
    260.14`, 262.2`, 261.96`, 264.47`, 262.64`, 265.76`, 267.1`, 266.29`, 263.19`,
    262.01`, 261.78`, 266.37`, 264.29`, 267.84`, 267.25`, 264.16`, 259.45`, 261.74`,
    265.58`, 270.71`, 266.92`, 268.48`, 270.77`, 271.46`, 275.15`, 279.86`,
    280.41`, 279.74`, 280.02`, 279.44`, 284.`, 284.27`, 289.91`, 289.8`, 291.52`};

Print["{Min[GOOG],Max[GOOG]}=", {Min[GOOG], Max[GOOG]}]
Print["{Min[AAPL],Max[AAPL]}=", {Min[AAPL], Max[AAPL]}]
GraphicsRow[{ListPlot[GOOG, Filling → Axis],
  ListPlot[AAPL, AxesOrigin → {0, 120}, Filling → Axis]}, ImageSize → Large]
```

$$Y1 = \frac{GOOG - Mean[GOOG]}{StandardDeviation[GOOG]}; \quad Y2 = \frac{AAPL - Mean[AAPL]}{StandardDeviation[AAPL]};$$

```
Print["Length[Y1]=", n = Length[Y1]]
Print["{Mean[Y1],SD[Y1]}=", {m1 = Mean[Y1], sd1 = StandardDeviation[Y1]}]
Print["{Mean[Y2],SD[Y2]}=", {m2 = Mean[Y2], sd2 = StandardDeviation[Y2]}]
{Min[Y1] - StandardDeviation[Y1], Max[Y1] + StandardDeviation[Y1]}
{Min[Y2] - StandardDeviation[Y2], Max[Y2] + StandardDeviation[Y2]}

Y = Table[{Y1[[i]], Y2[[i]]}, {i, n}];
{HY = Histogram3D[Y, 15, "Probability", ColorFunction → "Rainbow"],
 ListPlot[Y, PlotStyle → {Purple}]}
```

2. Obtain the parallelogram support of the joint density of (Y_1, Y_2)

```
me1 = ∑_{j=1}^{n} Y1[[j]] / n;

me2 = ∑_{j=1}^{n} Y2[[j]] / n;

Var =
    1
  ──────── {{Sum[(Y1[[j]] - me1)^2, {j, n}], Sum[(Y1[[j]] - me1) (Y2[[j]] - me2), {j, n}]},
   (n - 1)

    {Sum[(Y1[[j]] - me1) (Y2[[j]] - me2), {j, n}], Sum[(Y2[[j]] - me2)^2, {j, n}]}}};


Vhi = MatrixPower[Var, -1 / 2];
Print["Vhi=", Vhi // MatrixForm]
Vh = MatrixPower[Var, 1 / 2];
Print["Vh=", Vh // MatrixForm]
detVhi = Det[Vhi] // N;
V1 = Var[[1, 1]];
V2 = Var[[2, 2]];

X1 = Vhi[[1, 1]] (Y1 - ConstantArray[me1, n]) +
     Vhi[[1, 2]] (Y2 - ConstantArray[me2, n]) // N;
X2 = Vhi[[2, 1]] (Y1 - ConstantArray[me1, n]) +
     Vhi[[2, 2]] (Y2 - ConstantArray[me2, n]) // N;
X = Table[{X1[[i]], X2[[i]]}, {i, n}];
kX1 = SmoothKernelDistribution[X1, "Silverman"];
kX2 = SmoothKernelDistribution[X2, "Silverman"];
pk1 = Plot[PDF[kX1, x], {x, -4, 4}, PlotRange → All];
pk2 = Plot[PDF[kX2, x], {x, -4, 4}, PlotRange → All];
{pk1, pk2}
{a1, a2} = {Round[FindRoot[PDF[kX1, x] == 0.001, {x, -3}][[1, 2]], 0.01],
   Round[FindRoot[PDF[kX1, x] == 0.001, {x, 2}][[1, 2]], 0.01]}
(*Let the domain for X1 be {-3.26,3.04}*)
{b1, b2} = {Round[FindRoot[PDF[kX2, x] == 0.001, {x, -2}][[1, 2]], 0.01],
   Round[FindRoot[PDF[kX2, x] == 0.001, {x, 2}][[1, 2]], 0.01]}
(*Let the domain for X2 be {-2.46,3.09}*)

s11 = Vh[[1, 1]];
s12 = Vh[[1, 2]];
s21 = Vh[[2, 1]];
s22 = Vh[[2, 2]];
```

```
Clear[s1, s2, s3, s4]
```

$$s1[x\_] := \frac{s22}{s12} x - \left(s11 * s22 - s12^2\right) \frac{a1}{s12} + me2 - \frac{s22}{s12} me1$$

$$s2[x\_] := \frac{s12}{s11} x - \left(s11 * s22 - s12^2\right) \frac{b1}{s11} + me2 - \frac{s12}{s11} me1$$

$$s3[x\_] := \frac{s22}{s12} x - \left(s11 * s22 - s12^2\right) \frac{a2}{s12} + me2 - \frac{s22}{s12} me1$$

$$s4[x\_] := \frac{s12}{s11} x - \left(s11 * s22 - s12^2\right) \frac{b2}{s11} + me2 - \frac{s12}{s11} me1$$

```
ps1 = Plot[s1[x], {x, -5, 5}, PlotStyle → Red];
ps2 = Plot[s2[x], {x, -5, 5}, PlotStyle → Blue];
ps3 = Plot[s3[x], {x, -5, 5}, PlotStyle → Green];
ps4 = Plot[s4[x], {x, -5, 5}, PlotStyle → Orange];
Show[ps1, ps2, ps3, ps4]

Clear[DI]
DI[x_, y_] := If[y ≤ s1[x] && y ≤ s2[x] && y ≥ s3[x] && y ≥ s4[x], 1, 0]
Plot3D[DI[x, y], {x, -5, 5}, {y, -5, 5}, Mesh → All,
 PlotStyle → Cyan(*,PlotPoints→50,MaxRecursion→10*)]
```

3. Determine the joint density of (Y_1, Y_2) on the parallelogram support

```
𝒟 = SmoothKernelDistribution[Y, {"Adaptive", Automatic, .2},
    "Epanechnikov", PerformanceGoal → "Quality"];

(*set domain for Y1,Y2 and renormalize*)
cff = Quiet[NIntegrate[PDF[𝒟, {x, y}] * DI[x, y],
    {x, -5, 5}, {y, -5, 5}, Method → "QuasiMonteCarlo"]]
ClearAll[ff]
ff[x_, y_] := Evaluate@PDF[𝒟, {x, y}] * DI[x, y] / cff

{Plot3D[Evaluate@PDF[𝒟, {x, y}],
  {x, Min[Y1] - StandardDeviation[Y1], Max[Y1] + StandardDeviation[Y1]},
  {y, Min[Y2] - StandardDeviation[Y2], Max[Y2] + StandardDeviation[Y2]},
  PlotRange → All, PlotStyle → Cyan, ImageSize → Medium],
 Plot3D[ff[x, y], {x, Min[Y1] - StandardDeviation[Y1],
   Max[Y1] + StandardDeviation[Y1]},
  {y, Min[Y2] - StandardDeviation[Y2], Max[Y2] + StandardDeviation[Y2]},
  PlotRange → All, PlotStyle → Cyan, ImageSize → Medium]}
```

4. Marginal Densities and their Inverse CDFs

```
k1 = SmoothKernelDistribution[Y1, 0.6, "Triweight"];
k2 = SmoothKernelDistribution[Y2, .6, "Triweight"];
{Pk1 = Plot[PDF[k1, x], {x, Min[Y1] - StandardDeviation[Y1],
    Max[Y1] + StandardDeviation[Y1]}, PlotRange → Full],
Pk2 = Plot[PDF[k2, y], {y, Min[Y2] - StandardDeviation[Y2],
    Max[Y2] + StandardDeviation[Y2]}, PlotRange → Full]}
```

```
ClearAll[f11, f22, InvF1a, InvF2a]
f11[x_] := PDF[k1, x]
f22[x_] := PDF[k2, x]
InvF1a[r_] := InverseFunction[CDF[k1, #] &][r]
InvF2a[s_] := InverseFunction[CDF[k2, #] &][s]

{Plot[InvF1a[x], {x, 0, 1}, PlotRange → All],
 Plot[InvF2a[x], {x, 0, 1}, PlotRange → All]}
```

5. The copula density

```
ClearAll[cpdfa]
cpdfa[r_, s_] :=
   cpdfa[r, s] = (ff[InvF1a[r], InvF2a[s]]) / (f11[InvF1a[r]] * f22[InvF2a[s]]);

cp0 = ListPlot3D[Flatten[Table[{r, s, cpdfa[r, s]},
     {r, 2 / 100, 96 / 100, 1 / 100}, {s, 2 / 100, 96 / 100, 1 / 100}], 1],
   ColorFunction → "Rainbow", PlotRange → {- .3, 12}, ImageSize → Medium]
```

Section 2.6
Copula associated with a Brownian motion process and its running maximum

```
(*1. Data generated using the Air Canada stock prices*)
```

```
X = {25.83`, 25.11`, 26.69`, 26.72`, 26.66`, 26.77`, 26.54`, 27.22`, 27.53`, 27.57`,
    28.04`, 28.34`, 28.28`, 28.46`, 27.98`, 27.98`, 28.44`, 28.54`, 28.61`, 28.93`,
    29.24`, 29.67`, 30.22`, 30.28`, 31.21`, 31.32`, 31.18`, 31.36`, 31.68`, 31.69`,
    31.64`, 31.98`, 33.15`, 33.02`, 32.89`, 33.3`, 33.14`, 33.4`, 33.92`, 34.27`,
    33.11`, 34, 33.6`, 33.02`, 32.78`, 33, 33.16`, 33.36`, 32.04`, 32.01`, 31.89`,
    31.66`, 31.42`, 31.75`, 31.44`, 31.46`, 31.61`, 30.96`, 31.92`, 32.2`, 32.13`,
    32.21`, 33.56`, 33.58`, 33.49`, 33.03`, 32.98`, 32.88`, 32.32`, 32.37`, 32.79`,
    32.53`, 31.95`, 31.66`, 32.24`, 32.59`, 32.14`, 31.57`, 31.45`, 32.06`,
    31.82`, 31.95`, 32.16`, 32.87`, 32.57`, 33.62`, 35.28`, 35.5`, 36.3`, 35.87`,
    36.21`, 38.21`, 38.8`, 38.88`, 40.4`, 40.86`, 40.78`, 39.72`, 39.49`, 40.43`,
    41.22`, 40.72`, 40.51`, 40.35`, 39.9`, 39.16`, 39.41`, 39.34`, 39.03`, 39.08`,
    38.62`, 39.97`, 39.8`, 39.88`, 39.99`, 39.54`, 39.54`, 40.73`, 39.46`, 40,
    40.17`, 39.5`, 39.36`, 40.6`, 39.69`, 40.6`, 41.06`, 41.89`, 41.73`, 41.49`,
    40.76`, 41.76`, 41.88`, 42.51`, 42.87`, 43.35`, 43.45`, 44.32`, 44.21`, 44.1`,
    44.22`, 44.5`, 44.07`, 45.55`, 45.09`, 46.69`, 45.41`, 44.98`, 44.41`, 44.16`,
    43.31`, 44.47`, 44.5`, 43.76`, 43.5`, 42.32`, 41.86`, 43.58`, 43.67`, 43.21`,
    44.21`, 43.9`, 43.1`, 42.45`, 42.71`, 42.69`, 43.55`, 44.75`, 43.62`, 43.57`,
    44.09`, 44.56`, 44, 42.38`, 42.7`, 42.91`, 43.13`, 41.81`, 44.01`, 43.4`,
    43.03`, 42.75`, 43.04`, 43, 43.13`, 43.14`, 42.95`, 43.21`, 43.12`, 42.54`,
    43.1`, 43.27`, 43.24`, 42.88`, 43.66`, 44.81`, 45.55`, 45.63`, 45.38`, 45.54`,
    45.26`, 46.23`, 45.43`, 45.73`, 46.01`, 45.99`, 45.72`, 47.42`, 46.56`, 46.9`,
    46.99`, 47.28`, 47.06`, 48.6`, 47.33`, 48.63`, 48.13`, 48.56`, 48.87`, 49.3`,
    49.36`, 49.56`, 49.88`, 49.75`, 49.45`, 49.17`, 49.85`, 50.54`, 50.2`, 50.64`,
    50.05`, 49.47`, 49.37`, 49.2`, 49.19`, 49.03`, 48.87`, 48.56`, 49.39`, 48.8`,
    49.5`, 49.88`, 50.33`, 50.01`, 49.35`, 49.7`, 49.69`, 49.6`, 49.68`, 48.85`};
Print["length=", t = Length[X]]
```

To relate the data to a standard Wiener process, the first data point should be 0, the differences between successive observations should ideally often change signs and have a variance of one and there should be one unit of time between successive observations (as was assumed). Hence the following transformation.

```
data = (X – X[[1]]) / StandardDeviation[Table[X[[i + 1]] – X[[i]], {i, t – 1}]];
ListPlot[data, Filling → Axis]
```

```
dd = Table[data[[i + 1]] – data[[i]], {i, t – 1}];
Print["max[dd]=", Max[dd]]
Print["var[dd]=", Variance[dd]]
n = Length[data];
```

```
(*1.1 The first variable is the data after the transformation, a1*)
```

```
a1 = data;
Print["{mean[a1],var[a1]}=", {mn = Mean[a1], va = Variance[a1]}]
Print["{min[a1],max[a1]}=", {Min[a1], Max[a1]}]
```

{mean[a1],var[a1]}={21.7377, 120.204}

{min[a1],max[a1]}={–1.13638, 39.1579}

```
(*1.2 The second variable is the running maximum, mc*)
```

```
ClearAll[mc]
mc[j_] := Max[Table[a1[[i]], {i, 1, j}]]
tmc = Table[mc[j], {j, n}];
Print["{mean[tmc],var[tmc]}=", {Mean[tmc], Variance[tmc]}]
Print["{min[tmc],max[tmc]}=", {Min[tmc], Max[tmc]}]
ListPlot[tmc, Filling → Axis]

(*1.3 The bivariate data, pa*)

pa = Table[{a1[[j]], mc[j]}, {j, n}];

(*2. The joint density of the two variables*)

ListPlot[pa]

𝒟 = SmoothKernelDistribution[pa, Automatic, "Triweight", PerformanceGoal → "Quality"];
{Plot3D[PDF[𝒟, {x, y}], {x, -10, 50}, {y, -10, 50},
  PlotRange → All, ColorFunction → "Rainbow", ImageSize → Medium],
 ListPlot[pa, ImageSize → Medium]}

ClearAll[ff]
cff = NIntegrate[If[x ≤ y, PDF[𝒟, {x, y}], 0],
  {x, -10, 50}, {y, -10, 50}, Method → "AdaptiveQuasiMonteCarlo"]
ff[x_, y_] := ff[x, y] = If[x ≤ y, PDF[𝒟, {x, y}], 0] / cff
{Plot3D[ff[x, y], {x, -10, 50}, {y, -10, 50},
  PlotRange → All, ColorFunction → "Rainbow", ImageSize → Medium],
 ListPlot[pa, ImageSize → Medium]}

(*3.Marginal Densities and their Inverse CDFs*)

l1 = -10; u1 = 50; l2 = -10; u2 = 50;

k1 = SmoothKernelDistribution[a1, 8.8, "Triweight"];
Pk1 = Plot[PDF[k1, x], {x, l1, u1}, PlotRange → Full]
k2 = SmoothKernelDistribution[tmc, 6, "Triweight"];
Pk2 = Plot[PDF[k2, y], {y, l2, u2}, PlotRange → Full]

ClearAll[f11, f22, InvF1a, InvF2a]

f11[x_] := PDF[k1, x]
f22[x_] := PDF[k2, x]
InvF1a[r_] := InverseFunction[CDF[k1, #] &][r]
InvF2a[s_] := InverseFunction[CDF[k2, #] &][s]

{Plot[InverseFunction[CDF[k1, #] &][x], {x, 0, 1}],
 Plot[InverseFunction[CDF[k2, #] &][x], {x, 0, 1}, PlotRange → All]}

(*4. The copula density*)

ClearAll[cpdfa]
cpdfa[r_, s_] :=
  cpdfa[r, s] = (ff[InvF1a[r], InvF2a[s]]) / (f11[InvF1a[r]] * f22[InvF2a[s]]);
```

```
ListPlot3D[Flatten[Table[{r, s, cpdfa[r, s]},
    {r, 5 / 100, 95 / 100, 1 / 100}, {s, 5 / 100, 95 / 100, 1 / 100}], 1],
  PlotRange → All, ColorFunction → "Rainbow", ImageSize → Medium]
```

Section 2.7
Bernstein's approximation to copulas

Subsection 2.7.1

```
(*1. Data generated using the Air Canada stock prices*)
```

```
X = {25.83`, 25.11`, 26.69`, 26.72`, 26.66`, 26.77`, 26.54`, 27.22`, 27.53`, 27.57`,
    28.04`, 28.34`, 28.28`, 28.46`, 27.98`, 27.98`, 28.44`, 28.54`, 28.61`, 28.93`,
    29.24`, 29.67`, 30.22`, 30.28`, 31.21`, 31.32`, 31.18`, 31.36`, 31.68`, 31.69`,
    31.64`, 31.98`, 33.15`, 33.02`, 32.89`, 33.3`, 33.14`, 33.4`, 33.92`, 34.27`,
    33.11`, 34, 33.6`, 33.02`, 32.78`, 33, 33.16`, 33.36`, 32.04`, 32.01`, 31.89`,
    31.66`, 31.42`, 31.75`, 31.44`, 31.46`, 31.61`, 30.96`, 31.92`, 32.2`, 32.13`,
    32.21`, 33.56`, 33.58`, 33.49`, 33.03`, 32.98`, 32.88`, 32.32`, 32.37`, 32.79`,
    32.53`, 31.95`, 31.66`, 32.24`, 32.59`, 32.14`, 31.57`, 31.45`, 32.06`,
    31.82`, 31.95`, 32.16`, 32.87`, 32.57`, 33.62`, 35.28`, 35.5`, 36.3`, 35.87`,
    36.21`, 38.21`, 38.8`, 38.88`, 40.4`, 40.86`, 40.78`, 39.72`, 39.49`, 40.43`,
    41.22`, 40.72`, 40.51`, 40.35`, 39.9`, 39.16`, 39.41`, 39.34`, 39.03`, 39.08`,
    38.62`, 39.97`, 39.8`, 39.88`, 39.99`, 39.54`, 39.54`, 40.73`, 39.46`, 40,
    40.17`, 39.5`, 39.36`, 40.6`, 39.69`, 40.6`, 41.06`, 41.89`, 41.73`, 41.49`,
    40.76`, 41.76`, 41.88`, 42.51`, 42.87`, 43.35`, 43.45`, 44.32`, 44.21`, 44.1`,
    44.22`, 44.5`, 44.07`, 45.55`, 45.09`, 46.69`, 45.41`, 44.98`, 44.41`, 44.16`,
    43.31`, 44.47`, 44.5`, 43.76`, 43.5`, 42.32`, 41.86`, 43.58`, 43.67`, 43.21`,
    44.21`, 43.9`, 43.1`, 42.45`, 42.71`, 42.69`, 43.55`, 44.75`, 43.62`, 43.57`,
    44.09`, 44.56`, 44, 42.38`, 42.7`, 42.91`, 43.13`, 41.81`, 44.01`, 43.4`,
    43.03`, 42.75`, 43.04`, 43, 43.13`, 43.14`, 42.95`, 43.21`, 43.12`, 42.54`,
    43.1`, 43.27`, 43.24`, 42.88`, 43.66`, 44.81`, 45.55`, 45.63`, 45.38`, 45.54`,
    45.26`, 46.23`, 45.43`, 45.73`, 46.01`, 45.99`, 45.72`, 47.42`, 46.56`, 46.9`,
    46.99`, 47.28`, 47.06`, 48.6`, 47.33`, 48.63`, 48.13`, 48.56`, 48.87`, 49.3`,
    49.36`, 49.56`, 49.88`, 49.75`, 49.45`, 49.17`, 49.85`, 50.54`, 50.2`, 50.64`,
    50.05`, 49.47`, 49.37`, 49.2`, 49.19`, 49.03`, 48.87`, 48.56`, 49.39`, 48.8`,
    49.5`, 49.88`, 50.33`, 50.01`, 49.35`, 49.7`, 49.69`, 49.6`, 49.68`, 48.85`};
Print["length=", t = Length[X]]
```

To relate the data to a standard Wiener process, the first data point should be 0, the differences between successive observations should ideally often change signs and have a variance of one and there should be one unit of time between successive observations (as was assumed). Hence the following transformation.

```
data = (X – X[[1]]) / StandardDeviation[Table[X[[i + 1]] – X[[i]], {i, t – 1}]];
ListPlot[data, Filling → Axis]
```

```
dd = Table[data[[i + 1]] - data[[i]], {i, t - 1}];
Print["max[dd]=", Max[dd]]
Print["var[dd]=", Variance[dd]]
n = Length[data];

(*1.1 The first variable is the data after the transformation, a1*)

a1 = data;
Print["{mean[a1],var[a1]}=", {mn = Mean[a1], va = Variance[a1]}]
Print["{min[a1],max[a1]}=", {Min[a1], Max[a1]}]

(*1.2 The second variable is the running maximum, mc*)

ClearAll[mc]
mc[j_] := Max[Table[a1[[i]], {i, 1, j}]]
tmc = Table[mc[j], {j, n}];
Print["{mean[tmc],var[tmc]}=", {Mean[tmc], Variance[tmc]}]
Print["{min[tmc],max[tmc]}=", {Min[tmc], Max[tmc]}]
ListPlot[tmc, Filling → Axis]

(*1.3 The bivariate data, pa*)

pa = Table[{a1[[j]], mc[j]}, {j, n}];

(*Joint Density and Scatterplot*)

T = EmpiricalDistribution[pa];
𝒟 = SmoothKernelDistribution[pa, {"Adaptive", 3, .2}, PerformanceGoal → "Quality"];
GraphicsRow[{Plot3D[Evaluate@PDF[𝒟, {x, y}],
    {x, -10, 50}, {y, -10, 50}, PlotRange → All, PlotStyle → Cyan],
  ListPlot[pa]}, ImageSize → Large]

(*Approximating the empirical copula by a least-squares approximating polynomial
    and estimating its associated copula density function by differentiating the
    polynomial. This provides an initial representation of the copula density,
  which can be utilized to calibrate the parameter k of Bernstein'
 s empirical copula density.*)

data1 = Sort[a1];
data2 = tmc;
tbcop = Rationalize[
  Flatten[Table[{j / Length[a1], k / Length[a1], CDF[T, {data1[[j]], data2[[k]]}]},
    {j, Length[a1]} , {k, Length[a1]}], 1], 10^-10]

ecp1 = ListPlot3D[tbcop, ColorFunction → "Rainbow"];
```

A least-squares approximating polynomial P2(x,y) is fitted to the points tbcop

```
ClearAll[P2]
b2 = Flatten[Table[x^i y^j, {i, 0, 25}, {j, 0, 25}]];
P2 = Fit[tbcop, b2, {x, y}];
```

```
Show[Plot3D[P2, {x, .01, .99}, {y, .01, .99},
   PlotStyle -> Opacity[.2], PlotStyle -> GrayLevel, PlotRange → {0, 1}],
  Graphics3D[{Blue, PointSize[0.0002], Map[Point, tbcop]}]]

(*P2 is differentiated to obtain a preliminary copula density
   which will inform the choice of k in Bernstein's empirical copula:*)

ClearAll[f2]

f2[x_, y_] := D[P2, x, y]

Plot3D[f2[x, y] /. {x → u, y → v}, {u, .03, .92}, {v, .05, .97},
  ColorFunction → "Rainbow", PlotRange → {-1, 18}, PlotStyle -> Opacity[.6]]

(*Determining Bernstein's empirical copula density*)

ED1 = EmpiricalDistribution[a1];
ED2 = EmpiricalDistribution[tmc];
ClearAll[Cn]
```

$$Cn[x\_,\ y\_]\ :=\ Cn[x,\ y]\ =\ \frac{1}{Length[a1]}$$

$$\sum_{i=1}^{Length[a1]}\ Which[CDF[ED1,\ a1[[i]]]\ >\ x,\ 0,\ \ CDF[ED2,\ tmc[[i]]]\ >\ y,\ 0,\ \ True,\ 1]$$

```
k = 25;
ClearAll[dbin]
dbin[u_, v_] := dbin[u, v] = D[PDF[BinomialDistribution[k, x], v], x] /. x → u
dbin[u, v]
dbin[0.5, 2]

ClearAll[bcop]
```

$$bcop[x\_,\ y\_]\ :=\ bcop[x,\ y]\ =\ \sum_{i=0}^{k}\sum_{j=0}^{k}\left(Cn\left[\frac{i}{k},\ \frac{j}{k}\right]\ *\ dbin[x,\ i]\ *\ dbin[y,\ j]\right)$$

```
pbcop = ListPlot3D[Flatten[Table[{r, s, bcop[r, s]},
     {r, 1 / 100, 99 / 100, 1 / 100}, {s, 1 / 100, 99 / 100, 1 / 100}], 1],
   ColorFunction → "Rainbow", ImageSize → Medium, PlotRange → {0, 12}, PlotLabel → k]

(*repeat the above code with k=50,75 and 90*)
```

Subsection 2.7.2

```
(*input data*)
```

GOOG = {1045.85`, 1016.06`, 1070.71`, 1068.39`, 1076.28`, 1074.66`, 1070.33`, 1057.19`,
    1044.69`, 1077.15`, 1080.97`, 1089.9`, 1098.26`, 1070.52`, 1075.57`, 1073.9`,
    1090.99`, 1070.08`, 1060.62`, 1089.06`, 1116.37`, 1110.75`, 1132.8`, 1145.99`,
    1115.23`, 1098.71`, 1095.06`, 1095.01`, 1121.37`, 1120.16`, 1121.67`, 1113.65`,
    1118.56`, 1113.8`, 1096.97`, 1110.37`, 1109.4`, 1115.13`, 1116.05`, 1119.92`,
    1140.99`, 1147.8`, 1162.03`, 1157.86`, 1143.3`, 1142.32`, 1175.76`, 1193.2`,
    1193.32`, 1185.55`, 1184.46`, 1184.26`, 1198.85`, 1223.97`, 1231.54`, 1205.5`,
    1193.`, 1184.62`, 1173.02`, 1168.49`, 1173.31`, 1194.43`, 1200.49`, 1205.92`, 1215.`,
    1207.15`, 1203.84`, 1197.25`, 1202.16`, 1204.62`, 1217.87`, 1221.1`, 1227.13`,
    1236.34`, 1236.37`, 1248.84`, 1264.55`, 1256.`, 1263.45`, 1272.18`, 1287.58`,
    1188.48`, 1168.08`, 1162.61`, 1185.4`, 1189.39`, 1174.1`, 1166.27`, 1162.38`,
    1164.27`, 1132.03`, 1120.44`, 1164.21`, 1178.98`, 1162.3`, 1138.85`, 1149.63`,
    1151.42`, 1140.77`, 1133.47`, 1134.15`, 1116.46`, 1117.95`, 1103.63`, 1036.23`,
    1053.05`, 1042.22`, 1044.34`, 1066.04`, 1080.38`, 1078.72`, 1077.03`, 1088.77`,
    1085.35`, 1092.5`, 1103.6`, 1102.33`, 1111.42`, 1121.88`, 1115.52`, 1086.35`,
    1079.8`, 1076.01`, 1080.91`, 1097.95`, 1111.25`, 1121.58`, 1131.59`, 1116.35`,
    1124.83`, 1140.48`, 1144.21`, 1144.9`, 1150.34`, 1153.58`, 1146.35`, 1146.33`,
    1130.1`, 1138.07`, 1146.21`, 1137.81`, 1132.12`, 1250.41`, 1239.41`, 1225.14`,
    1216.68`, 1209.01`, 1193.99`, 1152.32`, 1169.95`, 1173.99`, 1204.8`, 1188.01`,
    1174.71`, 1197.27`, 1164.29`, 1167.26`, 1177.6`, 1198.45`, 1182.69`, 1191.25`,
    1189.53`, 1151.29`, 1168.89`, 1167.84`, 1171.02`, 1192.85`, 1188.1`, 1168.39`,
    1181.41`, 1211.38`, 1204.93`, 1204.41`, 1206.`, 1220.17`, 1234.25`, 1239.56`,
    1231.3`, 1229.15`, 1232.41`, 1238.71`, 1229.93`, 1234.03`, 1218.76`, 1246.52`,
    1241.39`, 1225.09`, 1219.`, 1205.1`, 1176.63`, 1187.83`, 1209.`, 1207.68`, 1189.13`,
    1202.31`, 1208.67`, 1215.45`, 1217.14`, 1243.01`, 1243.64`, 1253.07`, 1245.49`,
    1246.15`, 1242.8`, 1259.13`, 1260.99`, 1265.13`, 1290.`, 1262.62`, 1261.29`,
    1260.11`, 1273.74`, 1291.37`, 1292.03`, 1291.8`, 1308.86`, 1311.37`, 1299.19`,
    1298.8`, 1298.`, 1311.46`, 1334.87`, 1320.7`, 1315.46`, 1303.05`, 1301.35`, 1295.34`,
    1306.69`, 1313.55`, 1312.99`, 1304.96`, 1289.92`, 1295.28`, 1320.54`, 1328.13`,
    1340.62`, 1343.56`, 1344.66`, 1345.02`, 1350.27`, 1347.83`, 1361.17`, 1355.12`,
    1352.62`, 1356.04`, 1349.59`, 1348.84`, 1343.56`, 1360.4`, 1351.89`, 1336.14`};

```
AAPL = {157.92`, 142.19`, 148.26`, 147.93`, 150.75`, 153.31`, 153.8`, 152.29`, 150.`,
    153.07`, 154.94`, 155.86`, 156.82`, 153.3`, 153.92`, 152.7`, 157.76`, 156.3`,
    154.68`, 165.25`, 166.44`, 166.52`, 171.25`, 174.18`, 174.24`, 170.94`, 170.41`,
    169.43`, 170.89`, 170.18`, 170.8`, 170.42`, 170.93`, 172.03`, 171.06`, 172.97`,
    174.23`, 174.33`, 174.87`, 173.15`, 174.97`, 175.85`, 175.53`, 174.52`, 172.5`,
    172.91`, 178.9`, 180.91`, 181.71`, 183.73`, 186.12`, 188.02`, 186.53`, 188.16`,
    195.09`, 191.05`, 188.74`, 186.79`, 188.47`, 188.72`, 189.95`, 191.24`, 194.02`,
    195.35`, 195.69`, 197.`, 200.1`, 199.5`, 200.62`, 198.95`, 198.87`, 199.23`,
    199.25`, 203.13`, 203.86`, 204.53`, 207.48`, 207.16`, 205.28`, 204.3`, 204.61`,
    200.67`, 210.52`, 209.15`, 211.75`, 208.48`, 202.86`, 202.9`, 200.72`, 197.18`,
    185.72`, 188.66`, 190.92`, 190.08`, 189.`, 183.09`, 186.6`, 182.78`, 179.66`,
    178.97`, 178.23`, 177.38`, 178.3`, 175.07`, 173.3`, 179.64`, 182.54`, 185.22`,
    190.15`, 192.58`, 194.81`, 194.19`, 194.15`, 192.74`, 193.89`, 198.45`, 197.87`,
    199.46`, 198.78`, 198.58`, 195.57`, 199.8`, 199.74`, 197.92`, 201.55`, 202.73`,
    204.41`, 204.23`, 200.02`, 201.24`, 203.23`, 201.75`, 203.3`, 205.21`, 204.5`,
    203.35`, 205.66`, 202.59`, 207.22`, 208.84`, 208.67`, 207.02`, 207.74`, 209.68`,
    208.78`, 213.04`, 208.43`, 204.02`, 193.34`, 197.`, 199.04`, 203.43`, 200.99`,
    200.48`, 208.97`, 202.75`, 201.74`, 206.5`, 210.35`, 210.36`, 212.64`, 212.46`,
    202.64`, 206.49`, 204.16`, 205.53`, 209.01`, 208.74`, 205.7`, 209.19`, 213.28`,
    213.26`, 214.17`, 216.7`, 223.59`, 223.09`, 218.75`, 219.9`, 220.7`, 222.77`,
    220.96`, 217.73`, 218.72`, 217.68`, 221.03`, 219.89`, 218.82`, 223.97`, 224.59`,
    218.96`, 220.82`, 227.01`, 227.06`, 224.4`, 227.03`, 230.09`, 236.21`, 235.87`,
    235.32`, 234.37`, 235.28`, 236.41`, 240.51`, 239.96`, 243.18`, 243.58`, 246.58`,
    249.05`, 243.29`, 243.26`, 248.76`, 255.82`, 257.5`, 257.13`, 257.24`, 259.43`,
    260.14`, 262.2`, 261.96`, 264.47`, 262.64`, 265.76`, 267.1`, 266.29`, 263.19`,
    262.01`, 261.78`, 266.37`, 264.29`, 267.84`, 267.25`, 264.16`, 259.45`, 261.74`,
    265.58`, 270.71`, 266.92`, 268.48`, 270.77`, 271.46`, 275.15`, 279.86`,
    280.41`, 279.74`, 280.02`, 279.44`, 284.`, 284.27`, 289.91`, 289.8`, 291.52`};

Print["{Min[GOOG],Max[GOOG]}=", {Min[GOOG], Max[GOOG]}]
Print["{Min[AAPL],Max[AAPL]}=", {Min[AAPL], Max[AAPL]}]
GraphicsRow[{ListPlot[GOOG, Filling → Axis],
  ListPlot[AAPL, AxesOrigin → {0, 120}, Filling → Axis]}, ImageSize → Large]

Y1 = GOOG; Y2 = AAPL;
Print["Length[Y1]=", n = Length[Y1]]
Print["{Mean[Y1],SD[Y1]}=", {m1 = Mean[Y1], sd1 = StandardDeviation[Y1]}]
Print["{Mean[Y2],SD[Y2]}=", {m2 = Mean[Y2], sd2 = StandardDeviation[Y2]}]

Y = Table[{Y1[[i]], Y2[[i]]}, {i, n}];
HY = Histogram3D[Y, 15, "Probability", ColorFunction → "Rainbow"]

ListPlot[Y, PlotStyle → {Purple}, AxesOrigin → {1000, 120}]

(*Standardization of the components*)
```

```
      Y1 – m1
X1 = ─────────;
         sd1
      Y2 – m2
X2 = ─────────;
         sd2
X = Table[{X1[[i]], X2[[i]]}, {i, n}];
Print["{Min[X1],Max[X1]=", {Min[X1], Max[X1]} // N]
Print["{Min[X2],Max[X2]}=", {Min[X2], Max[X2]} // N]
edis = EmpiricalDistribution[Y];
EmpCDF = Table[CDF[edis, Y[[j]]], {j, 1, n}];
ListPlot[X, PlotStyle → {Blue}]

l1 = -3.5; u1 = 3.5; l2 = -3.5; u2 = 3.5;
H1 = Histogram[X1, {l1, u1, 0.2}, "PDF",
   ChartElementFunction → "FadingRectangle", ChartStyle → Green, PlotLabel → "X1"];
H2 = Histogram[X2, {l2, u2, 0.2}, "PDF", ChartElementFunction → "FadingRectangle",
   ChartStyle → Green, PlotLabel → "X2"];
HX = Histogram3D[X, 15, "Probability", ColorFunction → "Rainbow"];
GraphicsRow[{H1, H2, HX}, ImageSize → Large]

kp = SmoothKernelDistribution[X, "Silverman"];
P2 =
  Plot3D[PDF[kp, {x, y}], {x, l1, u1}, {y, l2, u2}, PlotRange → Full, PlotStyle → Cyan];
H1 = Histogram3D[X, 15, "PDF", ChartStyle → Red];
GraphicsRow[{P2, H1}, ImageSize → Large]

(*Approximating the empirical copula by a least-
   squares approximating polynomial and estimating its associated
    copula density function by differentiating the polynomial.
     This provides an initial representation of the copula density,
  which can be utilized to calibrate the parameter k of Bernstein'
 s empirical copula density.*)

T = EmpiricalDistribution[X];

data1 = Sort[X1];
data2 = Sort[X2];
tbcop = Rationalize[
  Flatten[Table[{j / Length[X1], k / Length[X1], CDF[T, {data1[[j]], data2[[k]]}]},
    {j, Length[X1]} , {k, Length[X1]}], 1], 10^-10]

ecp1 = ListPlot3D[tbcop, ColorFunction → "Rainbow"]
```

A least-squares approximating polynomial P2(x,y) is fitted to the points tbcop

```
ClearAll[P2]
b2 = Flatten[Table[x^i y^j, {i, 0, 25}, {j, 0, 25}]];
P2 = Fit[tbcop, b2, {x, y}];

Show[Plot3D[P2, {x, .01, .99}, {y, .01, .99},
  PlotStyle -> Opacity[.2], PlotStyle -> GrayLevel, PlotRange → {0, 1}],
 Graphics3D[{Blue, PointSize[0.0002], Map[Point, tbcop]}]]
```

```
(*P2 is differentiated to obtain a preliminary copula density
    which will inform the choice of k in Bernstein's empirical copula:*)

ClearAll[f2]
f2[x_, y_] := D[P2, x, y]

Plot3D[f2[x, y] /. {x → u, y → v}, {u, .03, .92}, {v, .05, .97},
 PlotRange → {-1, 18}, ColorFunction → "Rainbow", PlotStyle → Opacity[0.6]]

(*Determining Bernstein's empirical copula density*)

ED1 = EmpiricalDistribution[X1];
ED2 = EmpiricalDistribution[X2];
ClearAll[Cn]
```

$$Cn[x\_, y\_] := Cn[x, y] = \frac{1}{Length[X1]}$$

$$\sum_{i=1}^{Length[X1]} Which[CDF[ED1, X1[[i]]] > x, 0, CDF[ED2, X2[[i]]] > y, 0, True, 1]$$

```
k = 75;
ClearAll[dbin]
dbin[u_, v_] := dbin[u, v] = D[PDF[BinomialDistribution[k, x], v], x] /. x → u
dbin[u, v]

ClearAll[bcop]
```

$$bcop[x\_, y\_] := bcop[x, y] = \sum_{i=0}^{k} \sum_{j=0}^{k} \left(Cn\left[\frac{i}{k}, \frac{j}{k}\right] * dbin[x, i] * dbin[y, j] \right)$$

```
AbsoluteTiming[
  pbcop = ListPlot3D[Flatten[Table[{r, s, bcop[r, s]}, {r, 1 / 100, 99 / 100, 1 / 100},
      {s, 1 / 100, 99 / 100, 1 / 100}], 1], ColorFunction → "Rainbow",
    ImageSize → Medium, PlotRange → {0, 12}, PlotLabel → k]] // TableForm
```

For k = 75, Bernstein' s empirical copula density and the preliminary density estimate exhibit similar features. The copula densities obtained with k=25 and 50 are not sufficiently accurate.

Chapter 3

Nonparametric copula density estimation methodologies

Section 3.1 Introduction

An illustrative example involving four points

The data:

```
dt = {{2, 4}, {3, 12}, {7, 2}, {8, 3}}
n = Length[dt]
```

The empirical cdf of the data:

```
T = EmpiricalDistribution[dt];
```

```
cdc =
 Flatten[Table[{x, y, CDF[T, {x, y}]}, {x, 0, 8, 1 / 10}, {y, 0, 12, 1 / 10}], 1]
ListPointPlot3D[cdc, ColorFunction -> "Rainbow", PlotRange -> All]
Graphics[{PointSize[Large], Point[{{2, 4}, {3, 12}, {7, 2}, {8, 3}}]},
 GridLines → {Range[0, 9], Range[0, 13]}, Ticks → Automatic,
 TicksStyle → Blue, PlotRange → {{0, 9}, {0, 13}}, Axes → True]
```

The original Deheuvels' copula points

The components of each variable:

```
Y1 = Table[dt[[i, 1]], {i, n}]
Y2 = Table[dt[[j, 2]], {j, n}]
```

The ranks of the components:

```
p1 = Position[Sort[Y1], #] & /@ Y1 // Flatten
p2 = Position[Sort[Y2], #] & /@ Y2 // Flatten
```

Deheuvels' copula frequency points whose pmf is 1/n :

```
Table[{p1[[i]], p2[[i]]}, {i, n}] / n
```

$$\text{Graphics}\left[\left\{\text{PointSize[Large]}, \text{Point}\left[4\left\{\left\{\frac{1}{4}, \frac{3}{4}\right\}, \left\{\frac{1}{2}, 1\right\}, \left\{\frac{3}{4}, \frac{1}{4}\right\}, \left\{1, \frac{1}{2}\right\}\right\}\right]\right\},\right.$$

```
 GridLines → {Range[0, 4], Range[0, 4]},
 Ticks → {{{0, 0}, {1, 1 / 4}, {2, 1 / 2}, {3, 3 / 4}, {4, 1}},
   {{1, 1 / 4}, {2, 1 / 2}, {3, 3 / 4}, {4, 1}}},
 TicksStyle → Blue, PlotRange → {{0, 4}, {0, 4}}, Axes → True]
```

178

When utilized for example in conjunction with kernel density estimation, these copula points produce undesirable boundary effects and the resulting empirical copula pdf's will noticeably be less concentrated near zero within the interval [0,1].

Ranks divided by n + 1

The usual approach that is advocated in the literature to attenuate the edge effects consists of multiplying Deheuvels' copula points by n/(n+1), which in this case results in the following points: $\left\{\left\{\frac{1}{5}, \frac{3}{5}\right\}, \left\{\frac{2}{5}, \frac{4}{5}\right\}, \left\{\frac{3}{5}, \frac{1}{5}\right\}, \left\{\frac{4}{5}, \frac{2}{5}\right\}\right\}$.

```
Graphics[{PointSize[Large], Point[4 {{1/5, 3/5}, {2/5, 4/5}, {3/5, 1/5}, {4/5, 2/5}}]},
 GridLines → {Range[0, 4], Range[0, 4]},
 Ticks → {{{0, 0}, {1, 1 / 4}, {2, 1 / 2}, {3, 3 / 4}, {4, 1}},
   {{1, 1 / 4}, {2, 1 / 2}, {3, 3 / 4}, {4, 1}}},
 TicksStyle → Blue, PlotRange → {{0, 4}, {0, 4}}, Axes → True]
```

These points occupy various positions within the corresponding grid cells. Their uneven distribution will result in an empirical copula pdf that will be less concentrated near the ends of the unit interval and thus be inadequate as regards to achieving a uniform distribution on [0,1] in each variable.

Centered Copula Points

Conceptually, the copula points which merely depend on the ranks of the components of each variable, should preferably be uniformly spread out within the unit intervals. It is thus sensible to place each one of them at the center of the grid cell to which they belong. This is achieved by subtracting 1/(2n) from each of Deheuvel's copula points coordinates.
Note that there are only 4 pairs of points that are allocated to 16 grid cells in such a way that there will be only one point in each column and one point in each row, as there are 4 distinct ranks with respect to each coordinate.

The repositioned copula points which are centered in the corresponding grid cell and referred to as Centered Copula Points or CCPs:

```
ccp = Table[{1/n p1[[i]] - 1/2n, 1/n p2[[i]] - 1/2n}, {i, n}]
```

```
Graphics[{PointSize[Large],
  Point[{{4 × 1/8, 4 × 5/8}, {4 × 3/8, 4 × 7/8}, {4 × 5/8, 4 × 1/8}, {4 × 7/8, 4 × 3/8}}]},
 GridLines → {Range[0, 4], Range[0, 4]}, TicksStyle → Blue,
 Ticks → {{{0, 0}, {1, 1 / 4}, {2, 1 / 2}, {3, 3 / 4}, {4, 1}},
   {{1, 1 / 4}, {2, 1 / 2}, {3, 3 / 4}, {4, 1}}}, PlotRange → {{0, 4}, {0, 4}}, Axes → True]
```

With these points, the support of the marginal distributions constitutes the lowest discrepancy deterministic sequence with respect ot the uniform distribution in the unit interval. They form a

discrete distribution, their respective pmf being 1/n. Continuous discrete uniform marginal pdf's on [0,1] will be achieved by making use of a kde with **cuboidal kernels** whose height is n and whose bases are the cells containing the CCPs. The marginals are then uniformly distributed as Sklar's theorem stipulates. As well, the CCPs mitigate the boundary issues when used in conjunction with the kernel density estimation.

Constant cuboidal kernels kn[i,x,y] whose height is *n* over the *n* cells (of the *nxn* grid) containing the copula points

```
ClearAll[kn, kde, Ck]
kn[i_, x_, y_] := kn[i, x, y] = If[ccp[[i, 1]] – 1 / (2 n) ≤ x < ccp[[i, 1]] + 1 / (2 n) &&
    ccp[[i, 2]] – 1 / (2 n) ≤ y < ccp[[i, 2]] + 1 / (2 n), n, 0]
```

```
ccp[[2, 1]]
```

```
kn[2, 3 / 8, 7 / 8]
```

The copula density whose marginal distributions are uniformly distributed over the interval [0,1]:

$$kde[x\_, y\_] := kde[x, y] = \sum_{j=1}^{n} kn[j, x, y]$$

```
ListPointPlot3D[Table[{x, y, kde[x, y]}, {x, 0, 1, 1 / 200}, {y, 0, 1, 1 / 200}],
 ColorFunction → "LightTemperatureMap", PlotRange → All]
```

The copula cdf obtained by integration of this *copula density* :

$$Ck[u\_, v\_] := Ck[u, v] =$$
$$\sum_{j=1}^{n} Integrate[kn[j, x, y], \{x, ccp[[j, 1]] – 1 / (2 n), u\}, \{y, ccp[[j, 2]] – 1 / (2 n), v\}]$$

```
Ck[1 / 2, 2 / 3]
```

```
Ck[3 / 4, 5 / 6]
```

```
Ck[1, 1]
```

```
Plot3D[Evaluate[Ck[u, v]], {u, 0, 1}, {v, 0, 1}, ColorFunction → "LightTemperatureMap"]
```

The resulting copula is smoother than Deheuvels' copula which is obtained below.

Deheuvels' copula

```
Y1 = {2, 3, 7, 8};
Y2 = {4, 12, 2, 3};
```

```
ED1 = EmpiricalDistribution[Y1];
ED2 = EmpiricalDistribution[Y2];
ClearAll[Cnn]
```

```
Cnn[x_, y_] := Cnn[x, y] =  1
                           ───────────
                           Length[Y1]

        Length[Y1]
          ___
          ╲      Which[CDF[ED1, Y1[[i]]] > x, 0,  CDF[ED2, Y2[[i]]] > y, 0,  True, 1]
          ╱
          ‾‾‾
          i=1

ftp = Flatten[Table[{i / n, j / n, Cnn[i / n, j / n]}, {i, 1, n}, {j, 1, n}], 1]

Cnn[1, 1]

p0 = ListPointPlot3D[{{0.02, 0.02, 0}, {0.02, .98, 0}, {.98, 0.02, 0}, {1, 1, 1}},
      PlotStyle → {Blue, PointSize[0.02]}];

p3 = Plot3D[Cnn[u, v], {u, 0, 1}, {v, 0, 1},
      PlotRange → {0, 1}, ColorFunction → "LightTemperatureMap"];

Show[p0, p3, PlotRange → {0, 1}]
```

The copula obtained from the CCPs

```
Graphics[{PointSize[Large],

    Point[{{4 × 1/8, 4 × 5/8}, {4 × 3/8, 4 × 7/8}, {4 × 5/8, 4 × 1/8}, {4 × 7/8, 4 × 3/8}}]},

  GridLines → {Range[0, 4], Range[0, 4]}, TicksStyle → Blue,
  Ticks → {{{0, 0}, {1, 1 / 4}, {2, 1 / 2}, {3, 3 / 4}, {4, 1}},
      {{1, 1 / 4}, {2, 1 / 2}, {3, 3 / 4}, {4, 1}}}, PlotRange → {{0, 4}, {0, 4}}, Axes → True]

ccp = Table[{1/n p1[[i]] - 1/(2 n), 1/n p2[[i]] - 1/(2 n)}, {i, n}]

Dc = EmpiricalDistribution[ccp];

Cnt[x_, y_] :=

  Cnt[x, y] =  1        Length[Y1]
             ───────────    ___
             Length[Y1]     ╲    Which[CDF[ED1, Y1[[i]]] - 1 / (2 n) > x, 0,
                            ╱
                            ‾‾‾
                            i=1
        CDF[ED2, Y2[[i]]] - 1 / (2 n) > y, 0,  True, 1]

ftr = Flatten[Table[{i / n, j / n, Cnt[i / n, j / n]}, {i, 1, n}, {j, 1, n}], 1]

ListPlot3D[Flatten[Table[{u, v, Cnt[u, v]}, {u, 0, 1, 1 / 100}, {v, 0, 1, 1 / 100}], 1],
  PlotRange → {0, 1}, ColorFunction → "LightTemperatureMap"]

Plot3D[CDF[Dc, {x, y}], {x, 0, 1}, {y, 0, 1},
  PlotRange → {0, 1}, ColorFunction → "Rainbow"]
```

Section 3.2 Methodologies for estimating copula densities

Subsection 3.2.1
Differentiated least-squares copula estimates as initial density approximations

Data

Note that whether or not the data is standardized, the same copula will be obtained since it is only based on ranks and the standardization process does not affect the ranks.

```
faithful = ExampleData[{"Statistics", "OldFaithful"}];
Y = faithful;

Y1 = faithful[[All, 1]];
Y2 = faithful[[All, 2]];
n = Length[Y1];
m1 = Mean[Y1]; sd1 = StandardDeviation[Y1];
m2 = Mean[Y2]; sd2 = StandardDeviation[Y2];

Y = Table[{Y1[[i]], Y2[[i]]}, {i, n}];
𝒟Y1 = SmoothKernelDistribution[Y1];
𝒟Y2 = SmoothKernelDistribution[Y2];
𝒟Y = SmoothKernelDistribution[Y];
{Plot[PDF[𝒟Y1, x], {x, Min[Y1] – sd1, Max[Y1] + sd1}],
 Plot[PDF[𝒟Y2, x], {x, Min[Y2] – sd2, Max[Y2] + sd2}],
 Plot3D[PDF[𝒟Y, {x, y}], {x, Min[Y1] – sd1, Max[Y1] + sd1},
   {y, Min[Y2] – sd2, Max[Y2] + sd2}, PlotStyle → Cyan, PlotRange → All],
 HY = Histogram3D[Y, 15, "Probability", ColorFunction → "Rainbow", ImageSize → Small],
 ListPlot[Y, AxesOrigin → {1, 30}, ImageSize → Small]}
```

Express the copula frequency points in terms of the ranks

```
p1 = Position[Sort[Y1], #] & /@ Y1 // Flatten;
p2 = Position[Sort[Y2], #] & /@ Y2 // Flatten;
```

The centered copula frequency points and their common frequency (= 1/n)

```
ccfp = Table[{p1[[i]] / n, p2[[i]] / n} – 1 / (2 n), {i, n}];
ccfpf = Table[{p1[[i]] / n – 1 / (2 n), p2[[i]] / n – 1 / (2 n), 1 / n}, {i, n}];
ListPointPlot3D[ccfpf, Filling → Bottom, PlotStyle → Blue]
```

182

```
(*Standardization*)
X1 = Y1 - m1 / sd1 ;

X2 = Y2 - m2 / sd2 ;

X = Table[{X1[[i]], X2[[i]]}, {i, n}];
Print["{Min[X1],Max[X1]=", {Min[X1], Max[X1]} // N]
Print["{Min[X2],Max[X2]}=", {Min[X2], Max[X2]} // N]
edis = EmpiricalDistribution[Y];
EmpCDF = Table[CDF[edis, Y[[j]]], {j, 1, n}];

Xs = Table[{Rationalize[X1[[i]], 10^-12], Rationalize[X2[[i]], 10^-12]}, {i, 1, n}];

{l1, u1} = {-3, 3};
{l2, u2} = {-3.5, 3.5};

H1 = Histogram[X1, {l1, u1, 0.2}, "PDF", ChartElementFunction → "FadingRectangle",
    ChartStyle → Green, PlotLabel → "X1", ImageSize → Small];
H2 = Histogram[X2, {l2, u2, 0.2}, "PDF", ChartElementFunction → "FadingRectangle",
    ChartStyle → Green, PlotLabel → "X2", ImageSize → Small];
HX = Histogram3D[Xs, 15, "Probability", ColorFunction → "Rainbow", ImageSize → Small];
LX = ListPlot[X, PlotStyle → {Blue}, ImageSize → Small];


kp = SmoothKernelDistribution[X, "Silverman"];
PD = Plot3D[PDF[kp, {x, y}], {x, l1, u1},
    {y, l2, u2}, PlotRange → Full, PlotStyle → Cyan, ImageSize → Small];


{H1, H2, LX, HX, PD}
```

Linearized copula FI

```
X1r = Flatten[Table[X[[i, 1]] + RandomReal[-1, 1] 10^-8, {i, 1, Length[X1]}], 1];
X2r = Flatten[Table[X[[i, 2]] + RandomReal[-1, 1] 10^-8, {i, 1, Length[X1]}], 1];

D1 = EmpiricalDistribution[X1r];
D2 = EmpiricalDistribution[X2r];
```

The empirical copula to which linear interpolation is applied:

```
ClearAll[Cn]
Cn[x_, y_] :=
 Cn[x, y] = 1/n Σ_{i=1}^{n} Which[CDF[D1, X1r[[i]]] > x, 0,  CDF[D2, X2r[[i]]] > y, 0,  True, 1]

ftp = Flatten[Table[{i/n, j/n, Cn[i/n, j/n]}, {i, 1, n}, {j, 1, n}], 1];
```

```
ClearAll[FI]
FI = Interpolation[ftp, InterpolationOrder → 1];

FI[.55, .79]

Plot3D[Evaluate[FI[x, y]] /. {x → w, y → z}, {w, 0, 1},
  {z, 0, 1}, ColorFunction → "LightTemperatureMap"] // Quiet
```

Least-squares approximation of the copula (cdf) and its derivative

A least-squares approximating polynomial $F2_d[u, v]$ is fitted to the points tbcop.

```
T = EmpiricalDistribution[X];

data1o = Sort[Table[X[[i, 1]] + RandomReal[-1, 1] 10^-8, {i, 1, Length[X1]}]];
data2o = Sort[Table[X[[i, 2]] + RandomReal[-1, 1] 10^-8, {i, 1, Length[X1]}]];

tbcop = Rationalize[Flatten[
    Table[{j / Length[X1], k / Length[X1], CDF[T, {data1o[[j, 1]], data2o[[k, 1]]}]},
     {j, Length[X1]} , {k, Length[X1]}], 1], 10^-10];

d = 5;
b2_d = Flatten[Table[x^i y^j, {i, 0, d}, {j, 0, d}]];
P2_d = Fit[tbcop, b2_d, {x, y}];
F2_d[u_, v_] := F2_d[u, v] = P2_d /. {x → u, y → v};
```

Empirical copula pdf obtained by differentiation of the least - squares approximation of the copula cdf

```
f2_d[x_, y_] := f2_d[x, y] = D[P2_d, x, y];
ListPlot3D[
 Flatten[Table[{u, v, Evaluate[(Abs[f2_d[x, y]] + f2_d[x, y] /. {x → u, y → v}) / 2 ]},
   {u, .1, .9, .005}, {v, .1, .9, .005}], 1],
 ColorFunction → "LightTemperatureMap", PlotLabel → d, ImageSize → Medium]
```

Repeat the above codes for d=10, 15, 20, ...

Although the resulting density estimates can be regarded as preliminary, they turn out to be quite representative of the distributional features of the copula.

Since the least-squares polynomial approximations become unstable in the neighborhood of the boundary of the domain, we shall consider a subset thereof that is truncated by 10% on each side.

Finding the optimal degree

By mere visual inspection, one can observe that the copula density estimates of degrees 20, 25.30, and 35 are quite similar. Appealing to the principle of parsimony, one could select the copula density of degree 20 in each variables as a yardstick for the distribution of the copula.

184

This can be mathematically corroborated by noting that, within the truncated domain, the integrated squared difference between successive copula estimates are quite small when the degree is greater than or equal to 20.

Least-squares polynomial approximations for d < 20 are under-fitting as they do not adequately capture the structure of the data whereas approximations of degrees at least 20 in each variable turn out to be adequate up to degree 40 in which case the value of the pdf at the mode (around 7) is noticeably higher, which is indicative of overfitting.

```
ClearAll[ISDcdf]
ISDcdf[t_] :=
  ISDcdf[t] = NIntegrate[(F2_{t+5}[x, y] - F2_t[x, y])^2, {x, 0.1, 0.9}, {y, 0.1, 0.9}] // Quiet

{Table[{5 * t, ISDcdf[5 t]}, {t, 1, 7}] // Insert[{"t", "ISDcdf[t]"}, 1] // MatrixForm,
 ListLinePlot[Table[{5 t, ISDcdf[5 t]}, {t, 2, 7}],
   Mesh → All, ImageSize → Medium, PlotRange -> All]}
```

Subsection 3.2.2
Bernstein's polynomial approximation and degree selection

1. Data

Note that whether or not the data is standardized, the same copula will be obtained since it is only based on ranks and the standardization process does not affect the ranks.

```
faithful = ExampleData[{"Statistics", "OldFaithful"}];
Y = faithful;

Y1 = faithful[[All, 1]];
Y2 = faithful[[All, 2]];
n = Length[Y1];
m1 = Mean[Y1]; sd1 = StandardDeviation[Y1];
m2 = Mean[Y2]; sd2 = StandardDeviation[Y2];

Y = Table[{Y1[[i]], Y2[[i]]}, {i, n}];
𝒟Y1 = SmoothKernelDistribution[Y1];
𝒟Y2 = SmoothKernelDistribution[Y2];
𝒟Y = SmoothKernelDistribution[Y];
{Plot[PDF[𝒟Y1, x], {x, Min[Y1] - sd1, Max[Y1] + sd1}],
 Plot[PDF[𝒟Y2, x], {x, Min[Y2] - sd2, Max[Y2] + sd2}],
 Plot3D[PDF[𝒟Y, {x, y}], {x, Min[Y1] - sd1, Max[Y1] + sd1},
   {y, Min[Y2] - sd2, Max[Y2] + sd2}, PlotStyle → Cyan, PlotRange → All],
 HY = Histogram3D[Y, 15, "Probability", ColorFunction → "Rainbow", ImageSize → Small],
 ListPlot[Y, AxesOrigin → {1, 30}, ImageSize → Small]}
```

```
X1 = (Y1 - m1) / sd1;
X2 = (Y2 - m2) / sd2;
X = Table[{X1[[i]], X2[[i]]}, {i, n}];
Print["{Min[X1],Max[X1]=", {Min[X1], Max[X1]} // N]
Print["{Min[X2],Max[X2]}=", {Min[X2], Max[X2]} // N]
edis = EmpiricalDistribution[Y];
EmpCDF = Table[CDF[edis, Y[[j]]], {j, 1, n}];

Xs = Table[{Rationalize[X1[[i]], 10^-12], Rationalize[X2[[i]], 10^-12]}, {i, 1, n}];

{l1, u1} = {-3, 3};
{l2, u2} = {-3.5, 3.5};

H1 = Histogram[X1, {l1, u1, 0.2}, "PDF", ChartElementFunction → "FadingRectangle",
    ChartStyle → Green, PlotLabel → "X1", ImageSize → Small];
H2 = Histogram[X2, {l2, u2, 0.2}, "PDF", ChartElementFunction → "FadingRectangle",
    ChartStyle → Green, PlotLabel → "X2", ImageSize → Small];
HX = Histogram3D[Xs, 15, "Probability", ColorFunction → "Rainbow", ImageSize → Small];
LX = ListPlot[X, PlotStyle → {Blue}, ImageSize → Small];

kp = SmoothKernelDistribution[X, "Silverman"];
PD = Plot3D[PDF[kp, {x, y}], {x, l1, u1},
    {y, l2, u2}, PlotRange → Full, PlotStyle → Cyan, ImageSize → Small];


{H1, H2, LX, HX, PD}
```

2. Bernstein copula density

```
ED1 = EmpiricalDistribution[X1];
ED2 = EmpiricalDistribution[X2];
ClearAll[Cn]
```

$$Cn[x\_, y\_] := Cn[x, y] = \frac{1}{Length[X1]}$$

$$\sum_{i=1}^{Length[X1]} Which[CDF[ED1, X1[[i]]] > x, 0, \ CDF[ED2, X2[[i]]] > y, 0, \ True, 1]$$

```
Plot3D[Cn[x, y], {x, 0, 1}, {y, 0, 1}, ColorFunction → "LightTemperatureMap"]

ClearAll[bin, bcdf]
bin[k_, u_, v_] := bin[k, u, v] = PDF[BinomialDistribution[k, u], v];
```

$$bcdf[k\_, u\_, v\_] := \sum_{i=0}^{k} \sum_{j=0}^{k} \left(Cn\left[\frac{i}{k}, \frac{j}{k}\right] * bin[k, u, i] * bin[k, v, j] \right)$$

```
bcdf[25, 0.3, 0.1]
bcdf[25, 0.2, 0.1]
bcdf[50, 0.1, 0.1]

Plot3D[bcdf[25, x, y], {x, 0, 1}, {y, 0, 1}, ColorFunction → "LightTemperatureMap"]
```

186

```
ClearAll[dbin, bcop]
dbin[k_, u_, v_] := dbin[k, u, v] = D[PDF[BinomialDistribution[k, x], v], x] /. x → u;
```

$$\text{bcop}[k\_, x\_, y\_] := \text{bcop}[k, x, y] = \sum_{i=0}^{k} \sum_{j=0}^{k} \left(\text{Cn}\left[\frac{i}{k}, \frac{j}{k}\right] * \text{dbin}[k, x, i] * \text{dbin}[k, y, j] \right);$$

```
bcop[25, 0.1, 0.1]
```

Plots of Bernstein' s copula densities

```
d = 25;
ListPlot3D[
 Flatten[Table[{u, v, bcop[d, u, v]}, {u, .1, .9, .005}, {v, .1, .9, .005}], 1],
 ColorFunction → "LightTemperatureMap", PlotLabel → d, ImageSize → Medium]
```

Repeat the above code with d = 50, 75, 100, 125, 150

3. Optimal degree selection

Using the least-squares copula cdf with degree 20 as the reference copula.

Calculate the ISD between the reference copula and Bernstein copula cdf's of different degrees.

The reference (or yardstick) copula is the LS approximating polynomial denoted by F_y that was selected in the previous section.

```
T = EmpiricalDistribution[X];
```

$$\text{data1o} = \text{Sort}\left[\text{Table}\left[X[[i, 1]] + \text{RandomReal}[-1, 1] \, 10^{-8}, \{i, 1, \text{Length}[X1]\}\right]\right];$$
$$\text{data2o} = \text{Sort}\left[\text{Table}\left[X[[i, 2]] + \text{RandomReal}[-1, 1] \, 10^{-8}, \{i, 1, \text{Length}[X1]\}\right]\right];$$

$$\text{tbcop} = \text{Rationalize}\left[\text{Flatten}\left[\right.\right.$$
$$\text{Table}[\{j / \text{Length}[X1], k / \text{Length}[X1], \text{CDF}[T, \{\text{data1o}[[j, 1]], \text{data2o}[[k, 1]]\}]\},$$
$$\{j, \text{Length}[X1]\}, \{k, \text{Length}[X1]\}], 1\right], 10^{-10}\right];$$

```
d = 20;
```
$$\text{b2} = \text{Flatten}\left[\text{Table}\left[x^i \, y^j, \{i, 0, d\}, \{j, 0, d\}\right]\right];$$
```
P2 = Fit[tbcop, b2, {x, y}];
ClearAll[Fy]
Fy[u_, v_] := Fy[u, v] = P2 /. {x → u, y → v};
```

Plot Fy

```
ListPlot3D[
 Flatten[Table[{u, v, Evaluate[(Abs[Fy[x, y]] + Fy[x, y] /. {x → u, y → v}) / 2]},
   {u, .1, .9, .005}, {v, .1, .9, .005}], 1],
 ColorFunction → "LightTemperatureMap", PlotLabel → 20, ImageSize → Medium]
```

ISDs and relative differences based on ISD between t+25 and t

```
ClearAll[ISD]
```
$$\text{Table}\left[\text{Print}\left["\{t,ISD\}=", \{t, \text{Quiet}\left[\text{ISD}[25 * t] = \right.\right.\right.$$
$$\text{NIntegrate}\left[(\text{bcdf}[25 * t, x, y] - \text{Fy}[x, y])^2, \{x, 0, 1\}, \{y, 0, 1\}\right]\right]\}\right], \{t, 8\}\right]$$

```
{Table[{25 * n, ISD[25 * n]}, {n, 8}] // Insert[{"n", "ISD"}, 1] // MatrixForm,
 ListLinePlot[Table[{25 * n, ISD[25 * n]}, {n, 8}],
   Mesh → Full, PlotLabel → "ISD", ImageSize → Medium, PlotRange → All]}
```

Add relative differences between 'successive' ISD's 25 degrees apart

$$
A1 = \begin{pmatrix}
\text{"n"} & \text{"ISD"} \\
25 & 0.00010252481955737453` \\
50 & 0.00003473051580290645` \\
75 & 0.00001578393420385734` \\
100 & 0.0000112939946143552`5 \\
125 & 8.425623652017525`*\text{\textasciicircum}{-6} \\
150 & 7.081017244973576`*\text{\textasciicircum}{-6} \\
175 & 5.429791820783687`*\text{\textasciicircum}{-6}
\end{pmatrix};
$$

```
{v1 =
  Table[{(i - 2) * 25, Abs[(A1[[i, 2]] - A1[[i - 1, 2]])] / A1[[i - 1, 2]]}, {i, 3, 8}] //
    MatrixForm,
 ListLinePlot[Table[{(i - 2) * 25, Abs[(A1[[i, 2]] - A1[[i - 1, 2]])] / A1[[i - 1, 2]]},
    {i, 3, 8}], Mesh → Full,
   (*PlotLabel→"successive relative difference based on ISD",*)
   ImageSize → Medium, PlotRange → All, PlotStyle → Red]}
```

Subsection 3.2.3
Kernel-based copula density estimates

CCP example

With Deheuvels' copula points

$$
\text{dcp} = \left\{ \left\{ \frac{1}{4}, \frac{3}{4} \right\}, \left\{ \frac{1}{2}, 1 \right\}, \left\{ \frac{3}{4}, \frac{1}{4} \right\}, \left\{ 1, \frac{1}{2} \right\} \right\};
$$

```
ddc = SmoothKernelDistribution[dcp, .05, "Epanechnikov"];
```

```
Plot3D[PDF[ddc, {x, y}], {x, 0, 1}, {y, 0, 1},
 ColorFunction → "LightTemperatureMap", PlotRange → All]
```

```
Plot3D[CDF[ddc, {x, y}], {x, 0, 1}, {y, 0, 1},
 ColorFunction → "LightTemperatureMap", PlotRange → All]
```

The density function is truncated on parts of the boundary and does not integrate to one. This does not occur with the CCPs. Moreover, the marginal distributions are more evenly distributed over the unit intervals when making use of the CCPs.

With the CCPs

```
n = Length[dcp];
```

$$
\text{ccp} = \left\{ \left\{ \frac{1}{4}, \frac{3}{4} \right\}, \left\{ \frac{1}{2}, 1 \right\}, \left\{ \frac{3}{4}, \frac{1}{4} \right\}, \left\{ 1, \frac{1}{2} \right\} \right\} - \frac{1}{2\,n};
$$

```
cdc = SmoothKernelDistribution[ccp, .05, "Epanechnikov"];
```

```
Plot3D[PDF[cdc, {x, y}], {x, 0, 1}, {y, 0, 1},
 ColorFunction → "LightTemperatureMap", PlotRange → All]

Plot3D[CDF[cdc, {x, y}], {x, 0, 1}, {y, 0, 1},
 ColorFunction → "LightTemperatureMap", PlotRange → All]
```

Optimal kde bandwidth selection

1. Data

Note that whether or not the data is standardized, the same copula will be obtained since it is only based on ranks and the standardization process does not affect the ranks.

```
faithful = ExampleData[{"Statistics", "OldFaithful"}];
Y = faithful;

Y1 = faithful[[All, 1]];
Y2 = faithful[[All, 2]];
n = Length[Y1];
m1 = Mean[Y1]; sd1 = StandardDeviation[Y1];
m2 = Mean[Y2]; sd2 = StandardDeviation[Y2];

Y = Table[{Y1[[i]], Y2[[i]]}, {i, n}];
{ListPlot[Y1, AxesOrigin → {0, 3}, ImageSize → Small],
 ListPlot[Y2, AxesOrigin → {0, 60}, ImageSize → Small],
 HY = Histogram3D[Y, 15, "Probability", ColorFunction → "Rainbow", ImageSize → Small],
 ListPlot[Y, AxesOrigin → {1, 30}, PlotStyle → {Purple}, ImageSize → Small]}
```

$$X1 = \frac{Y1 - m1}{sd1};$$
$$X2 = \frac{Y2 - m2}{sd2};$$

```
X = Table[{X1[[i]], X2[[i]]}, {i, n}];
Print["{Min[X1],Max[X1]=", {Min[X1], Max[X1]} // N]
Print["{Min[X2],Max[X2]}=", {Min[X2], Max[X2]} // N]
edis = EmpiricalDistribution[Y];
EmpCDF = Table[CDF[edis, Y[[j]]], {j, 1, n}];

Xs = Table[{Rationalize[X1[[i]], 10⁻¹²], Rationalize[X2[[i]], 10⁻¹²]}, {i, 1, n}];

{l1, u1} = {-3, 3};
{l2, u2} = {-3.5, 3.5};
```

```
H1 = Histogram[X1, {l1, u1, 0.2}, "PDF", ChartElementFunction → "FadingRectangle",
    ChartStyle → Green, PlotLabel → "X1", ImageSize → Small];
H2 = Histogram[X2, {l2, u2, 0.2}, "PDF", ChartElementFunction → "FadingRectangle",
    ChartStyle → Green, PlotLabel → "X2", ImageSize → Small];
HX = Histogram3D[Xs, 15, "Probability", ColorFunction → "Rainbow", ImageSize → Small];
LX = ListPlot[X, PlotStyle → {Blue}, ImageSize → Small];

kp = SmoothKernelDistribution[X, "Silverman"];
PD = Plot3D[PDF[kp, {x, y}], {x, l1, u1},
    {y, l2, u2}, PlotRange → Full, PlotStyle → Cyan, ImageSize → Small];


{H1, H2, LX, HX, PD}
```

2. Use the Bernstein polynomial approximation of degree 125 as the yardstick

```
ED1 = EmpiricalDistribution[X1];
ED2 = EmpiricalDistribution[X2];
ClearAll[Cn]
```

$$Cn[x\_, y\_] := Cn[x, y] = \frac{1}{Length[X1]}$$

$$\sum_{i=1}^{Length[X1]} Which[CDF[ED1, X1[[i]]] > x, 0, \ CDF[ED2, X2[[i]]] > y, 0, \ True, 1]$$

```
k = 125;
ClearAll[dbin, bcop];
dbin[u_, v_] := dbin[u, v] = D[PDF[BinomialDistribution[k, x], v], x] /. x → u;
```

$$bcop[x\_, y\_] := bcop[x, y] = \sum_{i=0}^{k} \sum_{j=0}^{k} \left(Cn\left[\frac{i}{k}, \frac{j}{k}\right] * dbin[x, i] * dbin[y, j]\right);$$

```
ListPlot3D[Flatten[Table[{u, v, bcop[u, v]}, {u, .1, .9, .005}, {v, .1, .9, .005}], 1],
  ColorFunction → "LightTemperatureMap", PlotLabel → 125, ImageSize → Medium]
```

3. KDE based copula densities and bandwidth selection criterion

The narrower the kernels, the higher the peaks and the lower the troughs .

Thus, it seems reasonable to have a bandwidth selection criterion that is a function of the distance between the reference LS - copula pdf and a given KDE .

```
SeedRandom[0];
```
$$X1r = Table\left[X[[i, 1]] + RandomReal[-1, 1] \, 10^{-8}, \{i, 1, Length[X1]\}\right];$$
$$X2r = Table\left[X[[i, 2]] + RandomReal[-1, 1] \, 10^{-8}, \{i, 1, Length[X1]\}\right];$$

```
p1 = Position[Sort[X1r], #] & /@ X1r // Flatten;
p2 = Position[Sort[X2r], #] & /@ X2r // Flatten;
```

The repositioned copula points (which form the lowest discrepancy *n* point set whose discrepancy is 1/(2n)).

These points are located at the center of each cell of an *nxn* grid of the unit square. There will be

referred to as the *Centered Copula Points or ccp's*. The pmf at each point is 1/n

```
ccp = Table[{1/n p1[[i]] - 1/(2 n), 1/n p2[[i]] - 1/(2 n)}, {i, n}];
```

Try bandwidth 0.045, 0.040, 0.035, 0.030, 0.025

```
cd45 = SmoothKernelDistribution[ccp, .045, "Epanechnikov"];
cd40 = SmoothKernelDistribution[ccp, .040, "Epanechnikov"];
cd35 = SmoothKernelDistribution[ccp, .035, "Epanechnikov"];
cd30 = SmoothKernelDistribution[ccp, .030, "Epanechnikov"];
cd25 = SmoothKernelDistribution[ccp, .025, "Epanechnikov"];
```

Show plots

```
Plot3D[PDF[cd45, {x, y}], {x, 0, 1},
 {y, 0, 1}, ColorFunction -> "LightTemperatureMap",
 PlotRange -> All, PlotLabel → .045, ImageSize → Medium]

Plot3D[PDF[cd40, {x, y}], {x, 0, 1},
 {y, 0, 1}, ColorFunction -> "LightTemperatureMap",
 PlotRange -> All, PlotLabel → .040, ImageSize → Medium]

Plot3D[PDF[cd35, {x, y}], {x, 0, 1},
 {y, 0, 1}, ColorFunction -> "LightTemperatureMap",
 PlotRange -> All, PlotLabel → .035, ImageSize → Medium]

Plot3D[PDF[cd30, {x, y}], {x, 0, 1},
 {y, 0, 1}, ColorFunction -> "LightTemperatureMap",
 PlotRange -> All, PlotLabel → .030, ImageSize → Medium]

Plot3D[PDF[cd25, {x, y}], {x, 0, 1},
 {y, 0, 1}, ColorFunction -> "LightTemperatureMap",
 PlotRange -> All, PlotLabel → .025, ImageSize → Medium]
```

Comparison

ISD on a subset of the unit square, (0.1, 0.9)^2

```
(ISD[45] = NIntegrate[(kde45[x, y] - bcop[x, y])², {x, 0.1, 0.9},
    {y, 0.1, 0.9}, Method → "QuasiMonteCarlo"]) // AbsoluteTiming

(ISD[40] = NIntegrate[(kde40[x, y] - bcop[x, y])², {x, 0.1, 0.9},
    {y, 0.1, 0.9}, Method → "QuasiMonteCarlo"]) // AbsoluteTiming

(ISD[35] = NIntegrate[(kde35[x, y] - bcop[x, y])², {x, 0.1, 0.9},
    {y, 0.1, 0.9}, Method → "QuasiMonteCarlo"]) // AbsoluteTiming

(ISD[30] = NIntegrate[(kde30[x, y] - bcop[x, y])², {x, 0.1, 0.9},
    {y, 0.1, 0.9}, Method → "QuasiMonteCarlo"]) // AbsoluteTiming

(ISD[25] = NIntegrate[(kde25[x, y] - bcop[x, y])², {x, 0.1, 0.9},
    {y, 0.1, 0.9}, Method → "QuasiMonteCarlo"]) // AbsoluteTiming
```

Show the ISDs in table and figure

```
{(A2 = Table[{i * 0.001, ISD[i]}, {i, {45, 40, 35, 30, 25}}]) //
    Insert[{"Bandwidth", "ISD"}, 1] // MatrixForm,
  ListLinePlot[A2, Mesh → Full, PlotLabel → "Bandwidth vs ISD",
    ImageSize → Medium, PlotRange → All]}
```

Subsection 3.2.4
Differentiated linearized empirical copulas

data

```
OF = ExampleData[{"Statistics", "OldFaithful"}];
n = Length[OF];
mx1 = Max[Table[OF[[i, 1]], {i, n}]];
mx2 = Max[Table[OF[[i, 2]], {i, n}]];
```

The data is slightly perturbed to avoid replicates .

This is how to proceed to obtain the *same * vectors Y1 and Y2 every time the notebook is run .

```
ClearAll[Y1, Y2]
Y1 = Flatten[N[Table[OF[[j, 1]] + {SeedRandom[j + 5];
        RandomReal[{-1, 1}]} mx1 10⁻⁸, {j, n}], 8], 1];
DuplicateFreeQ[Y1]
Y2 = Flatten[N[Table[OF[[j, 2]] + {SeedRandom[j];
        RandomReal[{-1, 1}]} RandomReal[{-1, 1}] mx2 10⁻⁸, {j, n}], 8], 1];
DuplicateFreeQ[Y2]
Y = Table[{Y1[[i]], Y2[[i]]}, {i, n}];
```

```
m1 = Mean[Y1];
sd1 = StandardDeviation[Y1];
m2 = Mean[Y2];
sd2 = StandardDeviation[Y2];

MinY1 = Min[Y1];
MaxY1 = Max[Y1];
MinY2 = Min[Y2];
MaxY2 = Max[Y2];

k1 = SmoothKernelDistribution[Y1, "LeastSquaresCrossValidation"];
P1 = Plot[PDF[k1, y1], {y1, MinY1 - sd1, MaxY1 + sd1},
    PlotRange → Full, PlotStyle → Cyan, ImageSize → Small];
k2 = SmoothKernelDistribution[Y2, "LeastSquaresCrossValidation"];
P2 = Plot[PDF[k2, y2], {y2, MinY2 - sd2, MaxY2 + sd2},
    PlotRange → Full, PlotStyle → Cyan, ImageSize → Small];
LY = ListPlot[Y, PlotStyle → {Blue}, ImageSize → Small];
kp = SmoothKernelDistribution[Y];
Pk =
  Plot3D[PDF[kp, {y1, y2}], {y1, MinY1 - sd1, MaxY1 + sd1}, {y2, MinY2 - sd2, MaxY2 + sd2},
    PlotRange → Full, PlotStyle → Cyan, ImageSize → Small];

{P1, P2, LY, Pk}
```

Express the copula frequency points in terms of the ranks

```
p1 = Position[Sort[Y1], #] & /@ Y1 // Flatten;
p2 = Position[Sort[Y2], #] & /@ Y2 // Flatten;
```

The centered copula frequency points and their common frequency (= 1/n)

```
ccfp = Table[{p1[[i]] / n, p2[[i]] / n} - 1 / (2 n), {i, n}];
ccfpf = Table[{p1[[i]] / n - 1 / (2 n), p2[[i]] / n - 1 / (2 n), 1 / n}, {i, n}];
ListPointPlot3D[ccfpf, Filling → Bottom, PlotStyle → Blue]
```

Differentiated linearized empirical copulas

The empirical copula (cdf) is initially evaluated at all the intersection points of a $1/c$ x $1/c$ grid of the unit square which comprises $1/c$ x $1/c$ squares of dimension c^2.

Then, linear interpolation is applied to those copula points and the resulting surface is differentiated, which yields an approximate density function referred to as a differentiated linearized copula.

The selected spacing parameter c is such that the chosen Bernstein approximation and the differentiated linearized copula share similar distributional features, the neighborhoods wherein the reference density approximation increases sharply, if any, being discarted.

The spacing parameter c is taken to be the minimizer of the integrated squared difference, denoted by ISD(c) between the selected differentiated linearized copula and the chosen Bernstein density approximant denoted by bcop[ ]. Since ISD[1/12] is smaller than ISD[1/11] and ISD[1/13],

we let c equal 1/12 = 0.08333. The resulting density estimate shares similar distributional features with the other density approximants and could as well be utilized to identify suitable copula models.

The centered empirical copula

```
D1 = EmpiricalDistribution[Y1];
D2 = EmpiricalDistribution[Y2];
Ds = EmpiricalDistribution[Y];
```

Empirical copula (cdf)

```
ClearAll[Cn];
Cn[x_, y_] :=
```

$$Cn[x, y] = \frac{1}{n} \sum_{i=1}^{n} \text{Which}[\text{CDF}[D1, Y1[[i]]] > x, 0, \ \text{CDF}[D2, Y2[[i]]] > y, 0, \ \text{True}, 1]$$

The copula is evaluated at points from 0 to 1 by increments of c in each directions where c can be calibrated so that the resulting copula and the reference copula share similar distributional features; for instance, its value at the mode should be approximately the same as that of the chosen LS copula density or that of the selected Bernstein copula density- about 4 in this case.

$c = 1/11$

```
ftp11 = Flatten[Table[{x, y, Cn[x, y]}, {x, 0, 1, 1 / 11}, {y, 0, 1, 1 / 11}], 1];
fI11 = Interpolation[ftp11, InterpolationOrder → 1];
ListPlot3D[Flatten[Table[{w, z, Evaluate[D[fI11[x, y], x, y] /. {x → w, y → z}]},
    {w, 0, 1, 1 / 121}, {z, 0, 1, 1 / 121}], 1],
 ColorFunction → "LightTemperatureMap", PlotLabel → 1 / 11]
```

$c = 1/12$

```
ftp12 = Flatten[Table[{x, y, Cn[x, y]}, {x, 0, 1, 1 / 12}, {y, 0, 1, 1 / 12}], 1];
fI12 = Interpolation[ftp12, InterpolationOrder → 1];
ListPlot3D[Flatten[Table[{w, z, Evaluate[D[fI12[x, y], x, y] /. {x → w, y → z}]},
    {w, 0, 1, 1 / 144}, {z, 0, 1, 1 / 144}], 1],
 ColorFunction → "LightTemperatureMap", PlotLabel → 1 / 12]
```

$c = 1/13$

```
ftp13 = Flatten[Table[{x, y, Cn[x, y]}, {x, 0, 1, 1 / 13}, {y, 0, 1, 1 / 13}], 1];
fI13 = Interpolation[ftp13, InterpolationOrder → 1];
ListPlot3D[Flatten[Table[{w, z, Evaluate[D[fI13[x, y], x, y] /. {x → w, y → z}]},
    {w, 0, 1, 1 / 169}, {z, 0, 1, 1 / 169}], 1],
 ColorFunction → "LightTemperatureMap", PlotLabel → 1 / 13]
```

```
ClearAll[dbin, bcop]
dbin[k_, u_, v_] := dbin[k, u, v] = D[PDF[BinomialDistribution[k, x], v], x] /. x → u;
```

$$bcop[k\_, x\_, y\_] := bcop[k, x, y] = \sum_{i=0}^{k} \sum_{j=0}^{k} \left(Cn\left[\frac{i}{k}, \frac{j}{k}\right] * dbin[k, x, i] * dbin[k, y, j] \right);$$

```
(*Choose Bernstein's copula density with degree 125 as reference copula density*)

NIntegrate[((Evaluate[D[fI11[u, v], u, v]] - bcop[125, u, v]) /. {u → x, v → y})^2,
 {x, 0, 1}, {y, 0, 1}, Method → "AdaptiveQuasiMonteCarlo", MaxRecursion → 200]

NIntegrate[((Evaluate[D[fI12[u, v], u, v]] - bcop[125, u, v]) /. {u → x, v → y})^2,
 {x, 0, 1}, {y, 0, 1}, Method → "AdaptiveQuasiMonteCarlo", MaxRecursion → 200]

NIntegrate[((Evaluate[D[fI13[u, v], u, v]] - bcop[125, u, v]) /. {u → x, v → y})^2,
 {x, 0, 1}, {y, 0, 1}, Method → "AdaptiveQuasiMonteCarlo", MaxRecursion → 200]
```

This suggests that letting the spacings equal 1/12 is appropriate. This agrees with a visual assessment of the distributional features of the differentiated linearized empirical copulas as compared with the reference distributions.

Polynomial copula density estimate (h[x,y]) obtained from the joint moments (jm[r,s]) of the derivative of the linearized copula.
The marginals are approximately uniformly distributed:

```
ClearAll[df, df1, df2]
df[w_, z_] := D[fI12[x, y], x, y] /. {x → w, y → z}
df1[w_] := NIntegrate[df[w, z], {z, 0, 1}]
Plot[df1[w], {w, 0, 1}, PlotRange → {0, 1.1}]
df2[z_] := NIntegrate[df[w, z], {w, 0, 1}]
Plot[df2[z], {z, 0, 1}, PlotRange → {0, 1.1}]
```

Smoothing a DL copula by means of a bivariate polynomial

```
ClearAll[jm]
```

```
jm[r_, s_] := jm[r, s] = Rationalize[NIntegrate[x^r y^s df[x, y], {x, 0, 1}, {y, 0, 1},
      Method → "CartesianRule", MaxRecursion → 20, PrecisionGoal → 50], 10^-200] // Quiet
```

Various options to NIntegrate could be tried so as to obtain more precision.

The support of the uniform distribution is extended by 10 % in each direction.

```
ClearAll[f3, base, mm2]
Needs["MultivariateStatistics`"];
Off[MLE::shdw]; Off[Inner::"normal"];
Off[NIntegrate::izero]; Off[NIntegrate::"slwcon"];
f3[L1_List, L2_List] := Inner[Plus, L1, L2, List];
base[x_, y_] := base[x, y] = 100 / 144;
```

$$mm2[k\_, h\_] := mm2[k, h] = \frac{100}{144}\left(\frac{(11/10)^{k+1} - (-1/10)^{k+1}}{k+1}\right)\left(\frac{(11/10)^{h+1} - (-1/10)^{h+1}}{h+1}\right)$$

```
t = 11;
ClearAll[Zv1, Gms, t5, h];
L3 = Flatten[Table[{j, i}, {i, 0, t}, {j, 0, t}], 1];
P3 = Table[f3[L3[[i]], L3[[j]]], {i, 1, (t + 1)^2}, {j, 1, (t + 1)^2}];
M4 =
   Table[mm2[P3[[i, j]]][[1]], P3[[i, j]][[2]]], {i, Length[L3]}, {j, Length[L3]}];
Zv1[x_, y_] := Zv1[x, y] = Flatten[Table[x^j y^i, {i, 0, t}, {j, 0, t}], 1]
Gms[i_] := Gms[i] = jm[L3[[i, 1]], L3[[i, 2]]]
(*The vector of the joint moments of the selected linearized empirical copula*)
μ = Table[Gms[i], {i, Dimensions[L3][[1]]}];
(*The polynomial coefficients:*)
c4 = Inverse[M4].μ;
t5[x_, y_] := t5[x, y] = c4.Zv1[x, y];
h[x_, y_] := h[x, y] = base[x, y] * t5[x, y];

Unprotect[Power]; 0^0 = 1; Protect[Power];
ListPlot3D[Flatten[
   Table[{x, y, Abs[h[x, y]] / 2 + h[x, y] / 2}, {x, 0, 1, 1 / 100}, {y, 0, 1, 1 / 100}], 1],
  PlotRange → {0, 5}, ColorFunction -> "LightTemperatureMap", PlotLabel → t]
```

Section 3.3
On estimating joint density functions via copula density estimates

1. Data

```
faithful = ExampleData[{"Statistics", "OldFaithful"}];
Y = faithful;
```

```
Y1 = faithful[[All, 1]];
Y2 = faithful[[All, 2]];
n = Length[Y1];
m1 = Mean[Y1]; sd1 = StandardDeviation[Y1];
m2 = Mean[Y2]; sd2 = StandardDeviation[Y2];

Y = Table[{Y1[[i]], Y2[[i]]}, {i, n}];
𝒟Y1 = SmoothKernelDistribution[Y1];
𝒟Y2 = SmoothKernelDistribution[Y2];
𝒟Y = SmoothKernelDistribution[Y];
{Plot[PDF[𝒟Y1, x], {x, Min[Y1] - sd1, Max[Y1] + sd1}],
 Plot[PDF[𝒟Y2, x], {x, Min[Y2] - sd2, Max[Y2] + sd2}],
 Plot3D[PDF[𝒟Y, {x, y}],
```
$$\left\{x, \text{Min[Y1]} - \frac{sd1}{2}, \text{Max[Y1]} + \frac{sd1}{2}\right\}, \left\{y, \text{Min[Y2]} - \frac{sd2}{2}, \text{Max[Y2]} + \frac{sd2}{2}\right\},$$
```
  ColorFunction -> "LightTemperatureMap", PlotRange → All],
 HY = Histogram3D[Y, 15, "Probability", ColorFunction → "Rainbow", ImageSize → Small],
 ListPlot[Y, AxesOrigin → {1, 30}, ImageSize → Small]}
```

2. Applying Sklar' s result

```
SeedRandom[0];
Y1r = Table[Y[[i, 1]] + RandomReal[-1, 1] 10⁻⁸, {i, 1, Length[Y1]}];
Y2r = Table[Y[[i, 2]] + RandomReal[-1, 1] 10⁻⁸, {i, 1, Length[Y1]}];
p1 = Position[Sort[Y1r], #] & /@ Y1r // Flatten;
p2 = Position[Sort[Y2r], #] & /@ Y2r // Flatten;
```

The repositioned copula points have the lowest discrepancy for an *n* point set - whose discrepancy is 1/(2n).

These points are located at the center of each cell of an *nxn* grid of the unit square. There will be referred to as the *Centered Copula Points or CCPs*. The pmf at each point is 1/n. The undesirable edge effects encountered with the original Deheuvels' copula points when utilized in conjunction with the kde approach to density estimation, are mitigated when the CCPs are used.

$$\text{ccp} = \text{Table}\left[\left\{\frac{1}{n}\, \text{p1[[i]]} - \frac{1}{2\,n}, \frac{1}{n}\, \text{p2[[i]]} - \frac{1}{2\,n}\right\}, \{i, n\}\right];$$

The selected copula kde with bandwidth 0.035

```
cd35 = SmoothKernelDistribution[ccp, .035, "Epanechnikov"];

Plot3D[PDF[cd35, {x, y}], {x, 0, 1}, {y, 0, 1},
 ColorFunction -> "LightTemperatureMap", PlotRange -> All, PlotLabel → .035]
```

Marginal Distributions

```
kde1 = SmoothKernelDistribution[Y1, 0.75, "Triweight"];
ClearAll[g1, G1]
g1[x_] := PDF[kde1, x]
G1[x_] := CDF[kde1, x]
pd1 = Plot[g1[x], {x, Min[Y1] - sd1, Max[Y1] + sd1}, ImageSize → Medium];
```

$$\text{Show}\left[\text{Histogram}\left[\text{Y1}, \left\{\text{Min}[\text{Y1}] - \text{sd1}, \text{Max}[\text{Y1}] + \text{sd1}, \frac{\text{Max}[\text{Y1}] - \text{Min}[\text{Y1}]}{15}\right\},\right.\right.$$

```
   "PDF", ChartElementFunction → "FadingRectangle", ChartStyle → Green], pd1]
```

```
kde2 = SmoothKernelDistribution[Y2, 10, "Triweight"];
ClearAll[g2, G2]
g2[x_] := PDF[kde2, x]
G2[x_] := CDF[kde2, x]
pd2 = Plot[g2[x], {x, Min[Y2] - sd2, Max[Y2] + sd2}, ImageSize → Medium];
```

$$\text{Show}\left[\text{Histogram}\left[\text{Y2}, \left\{\text{Min}[\text{Y2}] - \text{sd2}, \text{Max}[\text{Y2}] + \text{sd2}, \frac{\text{Max}[\text{Y2}] - \text{Min}[\text{Y2}]}{15}\right\},\right.\right.$$

```
   "PDF", ChartElementFunction → "FadingRectangle", ChartStyle → Green], pd2]
```

Product of the marginal pdfs

```
Plot3D[g1[w] * g2[z], {w, Min[Y1] - sd1, Max[Y1] + sd1}, {z, Min[Y2] - sd2, Max[Y2] + sd2},
 PlotRange → All, ColorFunction → "LightTemperatureMap"]
```

Application of Sklar' s result

$$\text{Psk} = \text{Plot3D}\left[\text{g1}[w] * \text{g2}[z] * \text{Evaluate}[\text{PDF}[\text{cd35}, \{\text{G1}[w], \text{G2}[z]\}]],\right.$$

$$\left\{w, \text{Min}[\text{Y1}] - \frac{\text{sd1}}{2}, \text{Max}[\text{Y1}] + \frac{\text{sd1}}{2}\right\}, \left\{z, \text{Min}[\text{Y2}] - \frac{\text{sd2}}{2}, \text{Max}[\text{Y2}] + \frac{\text{sd2}}{2}\right\},$$

$$\left.\text{PlotRange} \rightarrow \text{All}, \text{PlotStyle} \rightarrow \text{Cyan}, \text{ImageSize} \rightarrow \text{Medium}\right]$$

Bivariate KDE obtained directly from the data

```
bandwidthy1 = 0.27;
kdeY = SmoothKernelDistribution[Y, {bandwidthy1, 10 * bandwidthy1}, "Cosine"];
Plot3D[PDF[kdeY, {x, y}],
```

$$\left\{x, \text{Min}[\text{Y1}] - \frac{\text{sd1}}{2}, \text{Max}[\text{Y1}] + \frac{\text{sd1}}{2}\right\}, \left\{y, \text{Min}[\text{Y2}] - \frac{\text{sd2}}{2}, \text{Max}[\text{Y2}] + \frac{\text{sd2}}{2}\right\},$$

$$\left.\text{ImageSize} \rightarrow \text{Medium}, \text{PlotRange} \rightarrow \text{All}, \text{PlotStyle} \rightarrow \text{Cyan}\right]$$

Section 3.4
Estimating a *t*-distributed copula density

1. Data

2000 points are generated from a bivariate T copula on one degree of freedom, the marginal distributions being standard normal and uniform on the interval $[0, 2]$.

```
jointpdf = CopulaDistribution[{"MultivariateT", {{1, 0}, {0, 1}}, 1},
   {NormalDistribution[], UniformDistribution[{0, 2}]}];
Plot3D[PDF[jointpdf, {x, y}], {x, -4, 4}, {y, 0, 2}, Evaluated -> True,
  ColorFunction -> "LightTemperatureMap"] // Quiet
```

The copula density :

```
pdfcop = CopulaDistribution[{"MultivariateT", {{1, 0}, {0, 1}}, 1},
   {UniformDistribution[{0, 1}], UniformDistribution[{0, 1}]}];
Plot3D[PDF[pdfcop, {x, y}], {x, 0, 1}, {y, 0, 1}, Evaluated -> True,
  ColorFunction -> "LightTemperatureMap", PlotRange -> {0, 5}] // Quiet
```

```
SeedRandom[13];
n = 2000;
X = RandomVariate[jointpdf, n];
ListPlot[X, PlotStyle -> Red]
```

```
X1 = Table[X[[i, 1]], {i, n}];
X2 = Table[X[[i, 2]], {i, n}];
```

```
HX = Histogram3D[X, 10, "PDF", ColorFunction -> "LightTemperatureMap"]
```

```
kdeX = SmoothKernelDistribution[X];
Plot3D[PDF[kdeX, {x, y}], {x, -4, 4}, {y, 0, 2},
 ImageSize -> Medium, ColorFunction -> "LightTemperatureMap"]
```

```
kdeX1 = SmoothKernelDistribution[X1];
Plot[PDF[kdeX1, x], {x, -4, 4}, ImageSize -> Medium, ColorFunction -> Blue]
```

```
kdeX2 = SmoothKernelDistribution[X2];
Plot[PDF[kdeX2, x], {x, 0, 2}, ImageSize -> Medium, PlotRange -> {0, .6}]
```

2. Differentiated least - squares approximations of the empirical copula as initial density estimates

```
T = EmpiricalDistribution[X];
```

```
data1o = Sort[Table[X[[i, 1]] + RandomReal[-1, 1] 10^-8, {i, 1, Length[X1]}]];
data2o = Sort[Table[X[[i, 2]] + RandomReal[-1, 1] 10^-8, {i, 1, Length[X1]}]];
```

The empirical copula cdf evaluated at 40,000 points lying on a 200 x 200 grid. This subsample is

used to make the calculations more manageable in this section.

```
tbcop = Rationalize[Flatten[Table[{10 j / Length[X1],
      10 k / Length[X1], CDF[T, {data1o[[10 j, 1]], data2o[[10 k, 1]]}]},
     {j, 1, Length[X1] / 10} , {k, 1, Length[X1] / 10}], 1], 10⁻¹⁰];
```

A least-squares approximating polynomial P2(x,y) is fitted to the points tbcop and differentiated. The copula density approximation is taken to be half the sum of the resulting polynomial and its absolute value to ensure nonnegativity on the unit square.

```
d = 5;
b2_d = Flatten[Table[xⁱ yʲ, {i, 0, d}, {j, 0, d}]];
P2_d = Fit[tbcop, b2_d, {x, y}];
F2_d_[u_, v_] := F2_d[u, v] = P2_d /. {x → u,  y → v};
f2_d_[x_, y_] := f2_d[x, y] = D[P2_d, x, y];
Pf2 = Plot3D[(Abs[f2_d[x, y]] + f2_d[x, y] /. {x → u, y → v}) / 2 , {u, 5 / 100, 95 / 100},
   {v, 5 / 100, 95 / 100}, ColorFunction → "LightTemperatureMap",
   PlotLabel → d, ImageSize → Medium, PlotLabel → d]
```

Repeat the above procedure similarly with d = 10, 15, 20, 25, 30, 35, 40

The degree selection criterion is based on the integrated squared difference between cdf approximations that are 5 degrees apart of over $[0.1, 0.9]^2$ so as to avoid the less stable behavior of the approximations close to the boundary of the unit square.

Visually, the graphs of the approximate copula densities for degrees 20 to 40 are similar. Since the difference between d30 and d35 is relatively small, the approximation of degree 30 in each variable can be chosen as reference distribution.

```
ClearAll[ISDcdf]
ISDcdf[t_] :=
 ISDcdf[t] = NIntegrate[(F2_{t+5}[x, y] - F2_t[x, y])², {x, 0.1, 0.9}, {y, 0.1, 0.9}] // Quiet

{Table[{5 * t, ISDcdf[5 t]}, {t, 1, 7}] // Insert[{"t", "ISDcdf[t]"}, 1] // MatrixForm,
 ListLinePlot[Table[{5 t, ISDcdf[5 t]}, {t, 2, 7}],
  Mesh → All, ImageSize → Medium, PlotRange -> All]}
```

3. KDE based copula densities and bandwidth selection criterion

The narrower the kernels, the higher the peaks and the lower the troughs .
Thus, it seems reasonable to have a bandwidth selection criterion that is a function of the distance between the reference Least-Squares copula pdf and a given KDE .

```
n = Length[X];

X1r = Table[X[[i, 1]] + RandomReal[-1, 1] 10⁻⁸, {i, 1, Length[X1]}];
X2r = Table[X[[i, 2]] + RandomReal[-1, 1] 10⁻⁸, {i, 1, Length[X1]}];
```

```
p1 = Position[Sort[X1r], #] & /@ X1r // Flatten;
p2 = Position[Sort[X2r], #] & /@ X2r // Flatten;
```

The repositioned copula points (which form the lowest discrepancy *n* point set whose discrepancy is 1/(2n)).

These points are located at the center of the corresponding cells of an *nxn* grid of the unit square. They will be referred to as the *Centered Copula Points or CCPs*. The pmf at each point is 1/n. The kernels of the density estimates are located at these points.

```
ccp = Table[{1/n p1[[i]] - 1/(2 n), 1/n p2[[i]] - 1/(2 n)}, {i, n}];
```

Bandwidth .025

```
cd25 = SmoothKernelDistribution[ccp, .025, "Epanechnikov"];
```

```
Plot3D[PDF[cd25, {x, y}], {x, 0, 1}, {y, 0, 1},
 ColorFunction -> "LightTemperatureMap", PlotRange -> All, PlotLabel → .025]
```

4. Bernstein polynomial copula densities and model selection

```
ED1 = EmpiricalDistribution[X1];
ED2 = EmpiricalDistribution[X2];
ClearAll[Cn]
Cn[x_, y_] :=
  Cn[x, y] = 1/n Sum[Which[CDF[ED1, X1[[i]]] > x, 0, CDF[ED2, X2[[i]]] > y, 0, True, 1], {i, 1, n}]
```

degree 100

```
ED1 = EmpiricalDistribution[X1];
ED2 = EmpiricalDistribution[X2];
ClearAll[Cn]
Cn[x_, y_] :=
  Cn[x, y] = 1/n Sum[Which[CDF[ED1, X1[[i]]] > x, 0, CDF[ED2, X2[[i]]] > y, 0, True, 1], {i, 1, n}]
```

```
ClearAll[cdfbcop, bcop]
k = 100;
ClearAll[dbin, bin]
dbin[u_, v_] := dbin[u, v] = D[PDF[BinomialDistribution[k, x], v], x] /. x → u
dbin[u, v]
dbin[0.5, 4]
```

```
bin[u_, v_] := bin[u, v] = PDF[BinomialDistribution[k, x], v] /. x → u
bin[u, v]
bin[0.5, 3]
```

```
ClearAll[cdfbcop]
```

$$\text{cdfbcop}[x\_, y\_] := \text{cdfbcop}[x, y] = \sum_{i=0}^{k} \sum_{j=0}^{k} \left(\text{Cn}\left[\frac{i}{k}, \frac{j}{k}\right] * \text{bin}[x, i] * \text{bin}[y, j] \right)$$

The cdf ISD within the unit square

```
NIntegrate[(cdfbcop[x, y] - F2₃₀[x, y])², {x, 0, 1},
 {y, 0, 1}, Method → "AdaptiveQuasiMonteCarlo", MaxRecursion → 200]
```

This is the smallest cdf ISD with respect to the reference LS copula and bcop can also be utilized as a reference copula density function.

$$\text{bcop}[x\_, y\_] := \text{bcop}[x, y] = \sum_{i=0}^{k} \sum_{j=0}^{k} \left(\text{Cn}\left[\frac{i}{k}, \frac{j}{k}\right] * \text{dbin}[x, i] * \text{dbin}[y, j] \right)$$

```
pbcop =
 ListPlot3D[Flatten[Table[{r, s, bcop[r, s]}, {r, 0, 1, 1 / 100}, {s, 0, 1, 1 / 100}], 1],
  ColorFunction → "Rainbow", ImageSize → Medium, PlotRange → {0, 6}, PlotLabel → k]
```

5. Differentiated linearized empirical copulas

```
ftp12 = Flatten[Table[{x, y, Cn[x, y]}, {x, 0, 1, 1 / 12}, {y, 0, 1, 1 / 12}], 1];
```

```
fI12 = Interpolation[ftp12, InterpolationOrder → 1];
```

```
Plot3D[Evaluate[D[fI12[x, y], x, y] /. {x → w, y → z}], {w, 0, 1}, {z, 0, 1},
 ColorFunction → "LightTemperatureMap", PlotRange → All, PlotLabel → 1 / 12]
```

```
NIntegrate[((Evaluate[D[fI12[u, v], u, v]] - bcop[u, v]) /. {u → x, v → y})²,
 {x, .1, .9}, {y, .1, .9}, Method → "AdaptiveQuasiMonteCarlo", MaxRecursion → 200]
```

Chapter 4
Representations of Certain Measures of Association in terms of Copulas

Section 4.6
Illustrative Examples

```
SeedRandom[1];
uni = RandomReal[{-3, 3}, 500];
f[x_] := {
   {x, -x / 5 + 1 + RandomReal[{-.05, .05}]},
   {x, -x⁵ + RandomReal[{-5.5, 5.5}]},
   {Cos[x], Sin[x] + RandomReal[.15]},
   {x, -Sqrt[Abs[x^(3/2)]] + RandomReal[.35]},
   {x, Tan[x]³ + RandomReal[{-1.9, 1.9}]}}
data = f /@ uni;

Table[ListPlot[data[[All, i]], Frame → True, Axes → None,
   PlotStyle → {Directive[PointSize[.007]], Purple}], {i, 5}]

NumberForm[
   Table[HoeffdingDTest[Sequence @@ Transpose[data[[All, i]]], "TestDataTable"], {i, 5}],
   {4, 4}] // Chop
NumberForm[Table[SpearmanRankTest[Sequence @@ Transpose[data[[All, i]]],
      "TestDataTable"], {i, 5}], {4, 4}] // Chop
NumberForm[Table[KendallTauTest[Sequence @@ Transpose[data[[All, i]]], "TestDataTable"],
      {i, 5}], {4, 4}] // Chop
NumberForm[Table[BlomqvistBetaTest[Sequence @@ Transpose[data[[All, i]]],
      "TestDataTable"], {i, 5}], {4, 4}] // Chop
NumberForm[Table[PearsonCorrelationTest[Sequence @@ Transpose[data[[All, i]]],
      "TestDataTable"], {i, 5}] // Quiet, {4, 4}] // Chop
```

203

Chapter 5

A Criterion for Characterizing the Tail Behavior of a Distribution

Section 5.2
A methodology based on the arctangent transformation

```
ClearAll[tail];

tail[𝒟_] := (

  μ = Mean[𝒟];
  σ = StandardDeviation[𝒟];
       2
  Qt9 = ─ ArcTan[(Quantile[𝒟, 0.9] - μ) / σ];
       π

            2
  Qt999999 = ─ ArcTan[(Quantile[𝒟, 0.999999] - μ) / σ];
            π

  lpt = ListPlot[{{Qt9, 0}, {Qt999999, 0}}, PlotStyle → Red];
  Print["p=", Qt999999 - Qt9];
  Print[Show[Plot[PDF[𝒟, σ Tan[(π / 2) y] + μ] σ (π / 2) Sec[(π / 2) y]²,
    {y, -1, 1}, PlotRange → All, PlotLegends → {𝒟}], lpt]];

)

dis = {
    NormalDistribution[0, 1],
    WeibullDistribution[0.5, 1], WeibullDistribution[2, 1],
    ExtremeValueDistribution[0, 1],
    ExponentialDistribution[1],
    StudentTDistribution[3], StudentTDistribution[5], StudentTDistribution[20],
    LogNormalDistribution[0, 1],
    UniformDistribution[{0, 1}],
    BetaDistribution[5, 2],
    BetaPrimeDistribution[5, 3], BetaPrimeDistribution[50, 30],
    GammaDistribution[50, 1],
    LogisticDistribution[0, 1]};
Table[tail[dis[[i]]], {i, 1, Length[dis]}]
```

Section 5.3
An illustration of the convergence of the tail index

```
ClearAll[simulation]
simulation[n_] := (
   𝒟 = ExponentialDistribution[1];
   SeedRandom[5];
   X0 = Sort[RandomVariate[𝒟, n]];
   P0 = SmoothKernelDistribution[X0];
   tail[P0]
  )
Table[simulation[n], {n, {100, 1000, 50 000, 1 000 000}}]
```

Section 5.4
Application of the tail index criterion

```
dis1 = {RayleighDistribution[1],
    BetaPrimeDistribution[2, 5]};
Table[tail[dis1[[i]]], {i, 1, Length[dis1]}]

(*Comparison with other criteria*)

dis = {
   WeibullDistribution[1, 1],
   ParetoDistribution[1, 3]
  };
Table[tail[dis[[i]]], {i, 1, Length[dis]}]
```

Chapter 6

Distribution of Quadratic Forms in Various Types of Random Variables

Section 6.3
Illustrative examples

Subsection 6.3.1
Quadratic forms in gamma random variables

```
(*A={{1,0,0,0},{0,2,1,0},{0,1,2,0},{0,0,0,3}}*)

A = {{1, 0, 0, 0}, {0, 2, 1, 0}, {0, 1, 2, 0}, {0, 0, 0, 3}};
% // TableForm
Eigenvalues[A] // N
α1 = 2; β1 = 2; α2 = 9; β2 = 1;
α3 = 2; β3 = 1; α4 = 12; β4 = 1;
Plot[{PDF[GammaDistribution[α1, β1], y], PDF[GammaDistribution[α2, β2], y],
  PDF[GammaDistribution[α3, β3], y], PDF[GammaDistribution[α4, β4], y]},
 {y, 0, 25}, PlotRange → All, PlotLegends → {"X₁", "X₂", "X₃", "X₄"}]
```

```
𝒟1 = GammaDistribution[α1, β1];
𝒟2 = GammaDistribution[α2, β2];
𝒟3 = GammaDistribution[α3, β3];
𝒟4 = GammaDistribution[α4, β4];
size = 10 000;
SeedRandom[1];
data1 = RandomVariate[𝒟1, size];
SeedRandom[2];
data2 = RandomVariate[𝒟2, size];
SeedRandom[3];
data3 = RandomVariate[𝒟3, size];
SeedRandom[4];
data4 = RandomVariate[𝒟4, size];
data = Sort[Table[{data1[[i]], data2[[i]], data3[[i]], data4[[i]]}.
     A.{data1[[i]], data2[[i]], data3[[i]], data4[[i]]}, {i, 1, size}]];
ED = EmpiricalDistribution[data];
ClearAll[F]
F[x_] := F[x] = CDF[ED, x];
ub = 3000;
F[ub]
H1 = Histogram[data, {0, ub, ub / 30}, "PDF"]
Plot[CDF[ED, x], {x, 0, ub}]

ClearAll[μ];
μ[h_] := μ[h] = Expand[({X₁, X₂, X₃, X₄}.A. {X₁, X₂, X₃, X₄})^h] /.
   {X₁^(j-·) :→ Moment[GammaDistribution[α1, β1], j],
    X₂^(j-·) :→ Moment[GammaDistribution[α2, β2], j],
    X₃^(j-·) :→ Moment[GammaDistribution[α3, β3], j],
    X₄^(j-·) :→ Moment[GammaDistribution[α4, β4], j]}
```

$$\left\{ \alpha = \frac{\mu[1]^2}{\mu[2] - \mu[1]^2}, \ \beta = \frac{\mu[2] - \mu[1]^2}{\mu[1]} \right\}$$

```
ClearAll[fb, Fb];
fb[y_] := fb[y] = PDF[GammaDistribution[α, β], y]
Fb[y_] := Fb[y] = CDF[GammaDistribution[α, β], y]
pf = Plot[fb[y], {y, 0, ub}];
Show[H1, pf]
Plot[{F[y], Fb[y]}, {y, 0, ub}, PlotRange → All, PlotLegends → {"Empirical", "Base"}]
ISD = Quiet[NumberForm[NIntegrate[(F[y] - Fb[y])², {y, 0, ub}], 4]]
```

```
n = 8;
M1 = Table[Moment[GammaDistribution[α, β], i + j], {i, 0, n}, {j, 0, n}];
μ1 = Table[μ[i], {i, 0, n}];
coe1 = Inverse[M1].μ1;
ClearAll[p, f1, F1]
```

$$p[y\_] := p[y] = \sum_{i=1}^{n+1} coe1[[i]] \; y^{i-1}$$

```
f1[y_] := f1[y] = fb[y] × p[y]
pf1 = Plot[f1[y], {y, 0, ub}, PlotRange → All];
Show[H1, pf1, PlotLabel → n]
F1[y_] := F1[y] = NIntegrate[f1[x], {x, 0, y}]
Plot[{F[y], F1[y]}, {y, 0, ub}, PlotRange → All,
  PlotLegends → {"Empirical", "Approximated"}, PlotLabel → n]
ISD1 = Quiet[NumberForm[NIntegrate[(F[y] - F1[y])², {y, 0, ub}], 4]]
```

Subsection 6.3.2
Quadratic forms in inverse Gaussian random variables

```
(*A={{1,0,0,0},{0,2,1,0},{0,1,2,0},{0,0,0,3}}*)

A = {{1, 0, 0, 0}, {0, 2, 1, 0}, {0, 1, 2, 0}, {0, 0, 0, 3}};
% // TableForm
Eigenvalues[A] // N
α1 = 2; β1 = 5; α2 = 3; β2 = 6;
α3 = 2; β3 = 2; α4 = 3; β4 = 4;
Plot[{PDF[InverseGaussianDistribution[α1, β1], y],
  PDF[InverseGaussianDistribution[α2, β2], y],
  PDF[InverseGaussianDistribution[α3, β3], y],
  PDF[InverseGaussianDistribution[α4, β4], y]}, {y, 0, 10},
 PlotRange → All, PlotLegends → {"X₁", "X₂", "X₃", "X₄"}]
```

```mathematica
𝒟1 = InverseGaussianDistribution[α1, β1];
𝒟2 = InverseGaussianDistribution[α2, β2];
𝒟3 = InverseGaussianDistribution[α3, β3];
𝒟4 = InverseGaussianDistribution[α4, β4];
size = 10000;
SeedRandom[1];
data1 = RandomVariate[𝒟1, size];
SeedRandom[2];
data2 = RandomVariate[𝒟2, size];
SeedRandom[3];
data3 = RandomVariate[𝒟3, size];
SeedRandom[4];
data4 = RandomVariate[𝒟4, size];
data = Sort[Table[{data1[[i]], data2[[i]], data3[[i]], data4[[i]]}.
      A.{data1[[i]], data2[[i]], data3[[i]], data4[[i]]}, {i, 1, size}]];
ED = EmpiricalDistribution[data];
ClearAll[F]
F[x_] := F[x] = CDF[ED, x];
ub = 900;
F[ub]
H1 = Histogram[data, {0, ub, ub / 30}, "PDF"]
Plot[CDF[ED, x], {x, 0, ub}, PlotRange → All]

ClearAll[μ];
μ[h_] := μ[h] = Expand[ ({X₁, X₂, X₃, X₄}.A. {X₁, X₂, X₃, X₄})^h] /.
    {X₁^j-· :→ Moment[InverseGaussianDistribution[α1, β1], j],
     X₂^j-· :→ Moment[InverseGaussianDistribution[α2, β2], j],
     X₃^j-· :→ Moment[InverseGaussianDistribution[α3, β3], j],
     X₄^j-· :→ Moment[InverseGaussianDistribution[α4, β4], j]}
```

$$\left\{\alpha = \mu[1], \; \beta = \frac{\mu[1]^3}{\mu[2] - \mu[1]^2}\right\}$$

```mathematica
ClearAll[fb, Fb];
fb[y_] := fb[y] = PDF[InverseGaussianDistribution[α, β], y]
Fb[y_] := Fb[y] = CDF[InverseGaussianDistribution[α, β], y]
pf = Plot[fb[y], {y, 0, ub}, PlotRange → All];
Show[H1, pf]
Plot[{F[y], Fb[y]}, {y, 0, ub}, PlotRange → All, PlotLegends → {"Empirical", "Base"}]
ISD = Quiet[NumberForm[NIntegrate[ (F[y] - Fb[y])², {y, 0, ub}], 4]]
```

```
n = 5;
M1 = Table[Moment[InverseGaussianDistribution[α, β], i + j], {i, 0, n}, {j, 0, n}];
μ1 = Table[μ[i], {i, 0, n}];
coe1 = Inverse[M1].μ1;
ClearAll[p, f1, F1]
```

$$p[y\_] := p[y] = \sum_{i=1}^{n+1} coe1[[i]] \, y^{i-1}$$

```
f1[y_] := f1[y] = fb[y] × p[y]
pf1 = Plot[f1[y], {y, 0, ub}, PlotRange → All];
Show[H1, pf1, PlotLabel → n]
F1[y_] := F1[y] = NIntegrate[f1[x], {x, 0, y}]
Plot[{F[y], F1[y]}, {y, 0, ub}, PlotRange → All,
  PlotLegends → {"Empirical", "Approximated"}, PlotLabel → n]
ISD1 = Quiet[NumberForm[NIntegrate[(F[y] - F1[y])², {y, 0, ub}], 4]]
```

Subsection 6.3.3
Quadratic forms in binomial random variables

```
(*A={{1,0,0,0},{0,2,-1,0},{0,-1,2,0},{0,0,0,3}}*)

A = {{1, 0, 0, 0}, {0, 2, -1, 0}, {0, -1, 2, 0}, {0, 0, 0, 3}};
% // TableForm
Eigenvalues[A] // N
α1 = 20; β1 = 1 / 4; α2 = 30; β2 = 1 / 2;
α3 = 20; β3 = 1 / 2; α4 = 30; β4 = 1 / 3;
Plot[{CDF[BinomialDistribution[α1, β1], y], CDF[BinomialDistribution[α2, β2], y],
  CDF[BinomialDistribution[α3, β3], y], CDF[BinomialDistribution[α4, β4], y]},
  {y, 0, 30}, PlotRange → All, PlotLegends → {"X₁", "X₂", "X₃", "X₄"}]
```

```
𝒟1 = BinomialDistribution[α1, β1];
𝒟2 = BinomialDistribution[α2, β2];
𝒟3 = BinomialDistribution[α3, β3];
𝒟4 = BinomialDistribution[α4, β4];
size = 10 000;
SeedRandom[1];
data1 = RandomVariate[𝒟1, size];
SeedRandom[2];
data2 = RandomVariate[𝒟2, size];
SeedRandom[3];
data3 = RandomVariate[𝒟3, size];
SeedRandom[4];
data4 = RandomVariate[𝒟4, size];
data = Sort[Table[{data1[[i]], data2[[i]], data3[[i]], data4[[i]]}.
     A.{data1[[i]], data2[[i]], data3[[i]], data4[[i]]}, {i, 1, size}]];

ED = EmpiricalDistribution[data];
ClearAll[F]
F[x_] := F[x] = CDF[ED, x];

ub = 1600;
F[ub] // N

H1 = Histogram[data, {0, ub, ub / 20}, "PDF"]
Plot[CDF[ED, x], {x, 0, ub}, PlotRange → All]

ClearAll[μ];
μ[h_] := μ[h] = Expand[({X₁, X₂, X₃, X₄}.A. {X₁, X₂, X₃, X₄})^h] /.
   {X₁^j-· ⧴ Moment[BinomialDistribution[α1, β1], j],
    X₂^j-· ⧴ Moment[BinomialDistribution[α2, β2], j],
    X₃^j-· ⧴ Moment[BinomialDistribution[α3, β3], j],
    X₄^j-· ⧴ Moment[BinomialDistribution[α4, β4], j]}
```

$$\left\{\alpha = \frac{\mu[1]^2}{\mu[2] - \mu[1]^2}, \; \beta = \frac{\mu[2] - \mu[1]^2}{\mu[1]}\right\}$$

```
ClearAll[fb, Fb];
fb[y_] := fb[y] = PDF[GammaDistribution[α, β], y]
Fb[y_] := Fb[y] = CDF[GammaDistribution[α, β], y]
SDD = NumberForm[Sum[(F[y] - Fb[y])², {y, 0, ub}], 4] // N
pf = Plot[fb[y], {y, 0, ub}];
Show[H1, pf]
Plot[{F[y], Fb[y]}, {y, 0, ub}, PlotRange → All, PlotLegends → {"Empirical", "Base"}]
```

```
n = 9;
M1 = Table[Moment[GammaDistribution[α, β], i + j], {i, 0, n}, {j, 0, n}];
μ1 = Table[μ[i], {i, 0, n}];
coe1 = Inverse[M1].μ1;
ClearAll[p, f1, F1]
```

$$p[y\_] := p[y] = \sum_{i=1}^{n+1} coe1[[i]] \, y^{i-1}$$

```
f1[y_] := f1[y] = N[fb[y] × p[y]]
pf1 = Plot[f1[y], {y, 0, ub}, PlotRange → All];
Show[H1, pf1, PlotLabel → n]
F1[y_] := F1[y] = NIntegrate[f1[x], {x, 0, y}]
Plot[{F[y], F1[y]}, {y, 0, ub}, PlotRange → All,
  PlotLegends → {"Empirical", "Approximated"}, PlotLabel → n]
```

$$SDD = NumberForm\left[Sum\left[(F[y] - F1[y])^2, \{y, 0, ub\}\right], 4\right]$$

Subsection 6.3.4
Quadratic forms in Poisson random variables

```
(*A={{1,0,0,0},{0,2,-1,0},{0,-1,2,0},{0,0,0,3}}*)

A = {{1, 0, 0, 0}, {0, 2, -1, 0}, {0, -1, 2, 0}, {0, 0, 0, 3}};
% // TableForm
Eigenvalues[A] // N
α1 = 3; α2 = 4;
α3 = 5; α4 = 6;
Plot[{CDF[PoissonDistribution[α1], y], CDF[PoissonDistribution[α2], y],
  CDF[PoissonDistribution[α3], y], CDF[PoissonDistribution[α4], y]},
  {y, 0, 20}, PlotRange → All, PlotLegends → {"X₁", "X₂", "X₃", "X₄"}]
```

```
𝒟1 = PoissonDistribution[α1];
𝒟2 = PoissonDistribution[α2];
𝒟3 = PoissonDistribution[α3];
𝒟4 = PoissonDistribution[α4];
size = 10000;
SeedRandom[1];
data1 = RandomVariate[𝒟1, size];
SeedRandom[2];
data2 = RandomVariate[𝒟2, size];
SeedRandom[3];
data3 = RandomVariate[𝒟3, size];
SeedRandom[4];
data4 = RandomVariate[𝒟4, size];
data = Sort[Table[{data1[[i]], data2[[i]], data3[[i]], data4[[i]]}.
     A.{data1[[i]], data2[[i]], data3[[i]], data4[[i]]}, {i, 1, size}]];

ED = EmpiricalDistribution[data];
ClearAll[F]
F[x_] := F[x] = CDF[ED, x];
ub = 1000;
F[ub]

H1 = Histogram[data, {0, ub, ub / 20}, "PDF"]
Plot[CDF[ED, x], {x, 0, ub}, PlotRange → All]

ClearAll[μ];
μ[h_] := μ[h] = Expand[ ({X₁, X₂, X₃, X₄}.A. {X₁, X₂, X₃, X₄})ʰ] /.
   {X₁ʲ⁻· :→ Moment[PoissonDistribution[α1], j],
    X₂ʲ⁻· :→ Moment[PoissonDistribution[α2], j],
    X₃ʲ⁻· :→ Moment[PoissonDistribution[α3], j],
    X₄ʲ⁻· :→ Moment[PoissonDistribution[α4], j]}
```

$$\left\{\alpha = \frac{\mu[1]^2}{\mu[2] - \mu[1]^2}, \; \beta = \frac{\mu[2] - \mu[1]^2}{\mu[1]}\right\}$$

```
ClearAll[fb, Fb];
fb[y_] := fb[y] = PDF[GammaDistribution[α, β], y]
Fb[y_] := Fb[y] = CDF[GammaDistribution[α, β], y]
SDD = NumberForm[Sum[ (F[y] - Fb[y])², {y, 0, ub}], 4] // N
pf = Plot[fb[y], {y, 0, ub}];
Show[H1, pf]
Plot[{F[y], Fb[y]}, {y, 0, ub}, PlotRange → All, PlotLegends → {"Empirical", "Base"}]
```

```
n = 10;
M1 = Table[Moment[GammaDistribution[α, β], i + j], {i, 0, n}, {j, 0, n}];
μ1 = Table[μ[i], {i, 0, n}];
coe1 = Inverse[M1].μ1;
ClearAll[p, f1, F1]

p[y_] := p[y] = Sum_{i=1}^{n+1} coe1[[i]] y^{i-1}

f1[y_] := f1[y] = N[fb[y] × p[y]]
pf1 = Plot[f1[y], {y, 0, ub}, PlotRange → All];
Show[H1, pf1, PlotLabel → n]
F1[y_] := F1[y] = NIntegrate[f1[x], {x, 0, y}]
Plot[{F[y], F1[y]}, {y, 0, ub}, PlotRange → All,
  PlotLegends → {"Empirical", "Approximated"}, PlotLabel → n]


SDD = NumberForm[Sum[(F[y] - F1[y])^2, {y, 0, ub}], 4]
```

Section 6.4
Hermitian quadratic forms

Subsection 6.4.1
A numerical example

```
A = {{6, -2 ί, 1}, {2 ί, 2, 1 - ί}, {1, 1 + ί, 6}};
MatrixForm[A]
es = Eigensystem[A] // N

α1 = 3; β1 = 2; α2 = -5; β2 = 1; α3 = 4; β3 = 3;
Plot[{PDF[NormalDistribution[α1, β1], y],
  PDF[NormalDistribution[α2, β2], y], PDF[NormalDistribution[α3, β3], y]},
 {y, -10, 15}, PlotRange → All, PlotLegends → {"X_1", "X_2", "X_3"}]
𝒟1 = NormalDistribution[α1, β1];
𝒟2 = NormalDistribution[α2, β2];
𝒟3 = NormalDistribution[α3, β3];
𝒟1c = NormalDistribution[α1, β1];
𝒟2c = NormalDistribution[α2, β2];
𝒟3c = NormalDistribution[α3, β3];
size = 10 000;
SeedRandom[1];
data1 = RandomVariate[𝒟1, size];
SeedRandom[2];
data2 = RandomVariate[𝒟2, size];
SeedRandom[3];
data3 = RandomVariate[𝒟3, size];
```

```
SeedRandom[4];
data1c = RandomVariate[𝒟1c, size];
SeedRandom[5];
data2c = RandomVariate[𝒟2c, size];
SeedRandom[6];
data3c = RandomVariate[𝒟3c, size];

data = Chop[Table[{data1[[i]] - 𝕚 data1c[[i]], data2[[i]] - 𝕚 data2c[[i]],
        data3[[i]] - 𝕚 data3c[[i]]}.A.{data1[[i]] + 𝕚 data1c[[i]],
        data2[[i]] + 𝕚 data2c[[i]], data3[[i]] + 𝕚 data3c[[i]]}, {i, 1, 10 000}]];
ub = 3000;
H1 = Histogram[data, {0, ub, ub / 30}, "PDF"]
ED = EmpiricalDistribution[data];
ClearAll[F]
F[x_] := F[x] = CDF[ED, x];
F[150];
Plot[CDF[ED, x], {x, 0, ub}, PlotRange → All]

ClearAll[μ];
μ[h_] :=
 μ[h] = Expand[({X₁ - 𝕚 X₁c, X₂ - 𝕚 X₂c, X₃ - 𝕚 X₃c}.A. {X₁ + 𝕚 X₁c, X₂ + 𝕚 X₂c, X₃ + 𝕚 X₃c})ʰ] /.
    {X₁^(j-.) :> Moment[NormalDistribution[α1, β1], j],
     X₂^(j-.) :> Moment[NormalDistribution[α2, β2], j],
     X₃^(j-.) :> Moment[NormalDistribution[α3, β3], j],
     X₁c^(j-.) :> Moment[NormalDistribution[α1, β1], j],
     X₂c^(j-.) :> Moment[NormalDistribution[α2, β2], j],
     X₃c^(j-.) :> Moment[NormalDistribution[α3, β3], j]}
μ[1];
μ[2];
```

$$\left\{\alpha = \frac{\mu[1]^2}{\mu[2] - \mu[1]^2}, \ \beta = \frac{\mu[2] - \mu[1]^2}{\mu[1]}\right\};$$

```
ClearAll[fb, Fb];
fb[y_] := fb[y] = PDF[GammaDistribution[α, β], y]
Fb[y_] := Fb[y] = CDF[GammaDistribution[α, β], y]
Show[H1, Plot[fb[y], {y, 2, ub}, PlotRange → All]]

Plot[{F[y], Fb[y]}, {y, 0, ub}, PlotLegends → {"Empirical", "Base"}]
ISD = Quiet[NIntegrate[(F[y] - Fb[y])², {y, 0, 1500}]]
```

Section 6.5
A parametric approach for quadratic forms in Gaussian random vectors

Subsection 6.5.3

A parametric approximation to the distribution of quadratic forms

Example 1

```
A = {{1, 0, 0}, {0, 2, 0},
{0, 0, 3}};
MatrixForm[A]
es = Eigensystem[A] // N
```

Simulated distribution

```
α1 = 0; β1 = 1; α2 = 0; β2 = 1; α3 = 0; β3 = 1;
Plot[{PDF[NormalDistribution[α1, β1], y]}, {y, -4, 4}, PlotRange → All];
𝒟1 = NormalDistribution[α1, β1];
𝒟2 = NormalDistribution[α2, β2];
𝒟3 = NormalDistribution[α3, β3];

size = 10000;
SeedRandom[1];
data1 = RandomVariate[𝒟1, size];
SeedRandom[2];
data2 = RandomVariate[𝒟2, size];
SeedRandom[3];
data3 = RandomVariate[𝒟3, size];

data = Chop[Table[{data1[[i]], data2[[i]], data3[[i]]}.
      A.{data1[[i]], data2[[i]], data3[[i]]}, {i, 1, 10000}]];
ED = EmpiricalDistribution[data];
ClearAll[F]
F[x_] := F[x] = CDF[ED, x];
sP = Plot[CDF[ED, x], {x, 0, 40}, PlotRange → All]

𝒟 = SmoothKernelDistribution[data, {"Adaptive", Automatic, .45}];
pkde = Plot[PDF[𝒟, x], {x, -1, 30}, PlotStyle → Red]
```

The moments

```
ClearAll[μ];
μ[h_] := μ[h] = Expand[({X₁, X₂, X₃}.A.{X₁, X₂, X₃})^h] /.
    {X₁^(j-.) :> Moment[NormalDistribution[α1, β1], j],
      X₂^(j-.) :> Moment[NormalDistribution[α2, β2], j],
      X₃^(j-.) :> Moment[NormalDistribution[α3, β3], j]}
μ[1]
μ[2]
μ[3]
μ[4]
μ[5]
```

The Gamma approximation

Using the method of moments, we have

$$\left\{\alpha = \frac{\mu[1]^2}{\mu[2] - \mu[1]^2}, \ \beta = \frac{\mu[2] - \mu[1]^2}{\mu[1]}\right\}$$

```
ClearAll[fb, Fb];
fb[y_] := fb[y] = PDF[GammaDistribution[α, β], y]
Fb[y_] := Fb[y] = CDF[GammaDistribution[α, β], y]
pg = Plot[fb[y], {y, 0, 30}, PlotRange → All];
Show[pg, pkde]
ep = Plot[F[y], {y, 0, 30}, PlotRange → All,
    PlotStyle → Red, PlotLegends → {"The empirical cdf"}];
bp = Plot[Fb[y], {y, 0, 30}, PlotRange → All, PlotStyle → Dashed,
    PlotLegends → {"The estimated cdf of the gamma approximation"}];
Show[
 ep,
 bp]

ISD = Quiet[NIntegrate[(F[y] - Fb[y])², {y, 0, 100}]]
```

The 4 - parameter q - EGG estimate:

Note that since the quadratic form is distributed as a linear combination of chi - square r.v.' s, the upper bound of the support is + infinity. Thus, the appropriate type of q - EGG distribution is **type-2** ($q > 1$) in which case the condition $\lambda < (q - 1)^{-1}$ must be satisfied.

Since the gamma approximation produces an excellent approximation, we will seek a solution **in the neighborhood** of $\left\{\alpha = \frac{9}{7}, \ \beta = \frac{14}{3}\right\}$. Now referring to the particular cases, it is seen that the gamma distribution is obtained by letting $\delta = 1$ and $q \to 1$ with τ being β and λ being α.

This suggests as a useful modelling approach to begin with a distribution having fewer parameters and then to proceed with q-EGG distributions having more parameters.

Type-2 q-EGG pdf:

```
ClearAll[g]
g[x_, q_, τ_, λ_, δ_] :=
```

$$g[x, q, \tau, \lambda, \delta] = \frac{\text{Abs}\left[\frac{\delta}{\tau}\right] (q - 1)^\lambda \, \text{Gamma}\left[\frac{1}{q-1}\right]}{\text{Gamma}\left[\frac{1}{q-1} - \lambda\right] \text{Gamma}[\lambda]} \left(\frac{x}{\tau}\right)^{\lambda * \delta - 1} \left(1 + (q - 1) \left(\frac{x}{\tau}\right)^\delta\right)^{-\frac{1}{q-1}}$$

First attempt: $1 < q < 1.05$, $4.5 < \tau < 4.8$, $1.1 < \lambda < 1.4$, $.98 < \delta < 1.02$.

$$\text{NMinimize}\Big[\Big\{\Big(\frac{(\tau)^1\,(q-1)^{\frac{-1}{\delta}}\,\text{Gamma}\big[\frac{1}{q-1}-\frac{1}{\delta}-\lambda\big]\,\text{Gamma}\big[\frac{1}{\delta}+\lambda\big]}{\text{Gamma}\big[\frac{1}{q-1}-\lambda\big]\,\text{Gamma}[\lambda]}-6\Big)^2+$$

$$\Big(\frac{(\tau)^2\,(q-1)^{\frac{-2}{\delta}}\,\text{Gamma}\big[\frac{1}{q-1}-\frac{2}{\delta}-\lambda\big]\,\text{Gamma}\big[\frac{2}{\delta}+\lambda\big]}{\text{Gamma}\big[\frac{1}{q-1}-\lambda\big]\,\text{Gamma}[\lambda]}-64\Big)^2+$$

$$\Big(\frac{(\tau)^3\,(q-1)^{\frac{-3}{\delta}}\,\text{Gamma}\big[\frac{1}{q-1}-\frac{3}{\delta}-\lambda\big]\,\text{Gamma}\big[\frac{3}{\delta}+\lambda\big]}{\text{Gamma}\big[\frac{1}{q-1}-\lambda\big]\,\text{Gamma}[\lambda]}-1008\Big)^2+$$

$$\Big(\frac{(\tau)^4\,(q-1)^{\frac{-4}{\delta}}\,\text{Gamma}\big[\frac{1}{q-1}-\frac{4}{\delta}-\lambda\big]\,\text{Gamma}\big[\frac{4}{\delta}+\lambda\big]}{\text{Gamma}\big[\frac{1}{q-1}-\lambda\big]\,\text{Gamma}[\lambda]}-21312\Big)^2,$$

$$1<q<1.05,\ 4.5<\tau<4.8,\ 1.1<\lambda<1.4,\ \lambda<(q-1)^{-1},\ .98<\delta<1.02\Big\},$$

$$\{q,\ \tau,\ \lambda,\ \delta\},\ \text{MaxIterations}\to 10\,000\Big]$$

Since the set lower limit of 4.5 is reached for the estimate of τ, that interval was extended: $4.2<\tau<4.8$.

$$\text{NMinimize}\Big[\Big\{\Big(\frac{(\tau)^1\,(q-1)^{\frac{-1}{\delta}}\,\text{Gamma}\big[\frac{1}{q-1}-\frac{1}{\delta}-\lambda\big]\,\text{Gamma}\big[\frac{1}{\delta}+\lambda\big]}{\text{Gamma}\big[\frac{1}{q-1}-\lambda\big]\,\text{Gamma}[\lambda]}-6\Big)^2+$$

$$\Big(\frac{(\tau)^2\,(q-1)^{\frac{-2}{\delta}}\,\text{Gamma}\big[\frac{1}{q-1}-\frac{2}{\delta}-\lambda\big]\,\text{Gamma}\big[\frac{2}{\delta}+\lambda\big]}{\text{Gamma}\big[\frac{1}{q-1}-\lambda\big]\,\text{Gamma}[\lambda]}-64\Big)^2+$$

$$\Big(\frac{(\tau)^3\,(q-1)^{\frac{-3}{\delta}}\,\text{Gamma}\big[\frac{1}{q-1}-\frac{3}{\delta}-\lambda\big]\,\text{Gamma}\big[\frac{3}{\delta}+\lambda\big]}{\text{Gamma}\big[\frac{1}{q-1}-\lambda\big]\,\text{Gamma}[\lambda]}-1008\Big)^2+$$

$$\Big(\frac{(\tau)^4\,(q-1)^{\frac{-4}{\delta}}\,\text{Gamma}\big[\frac{1}{q-1}-\frac{4}{\delta}-\lambda\big]\,\text{Gamma}\big[\frac{4}{\delta}+\lambda\big]}{\text{Gamma}\big[\frac{1}{q-1}-\lambda\big]\,\text{Gamma}[\lambda]}-21312\Big)^2,$$

$$1<q<1.05,\ 4.2<\tau<4.8,\ 1.1<\lambda<1.4,\ \lambda<(q-1)^{-1},\ .98<\delta<1.02\Big\},$$

$$\{q,\ \tau,\ \lambda,\ \delta\},\ \text{MaxIterations}\to 10\,000\Big]$$

```
qo = %[[2, 1, 2]];
τo = %%[[2, 2, 2]];
λo = %%%[[2, 3, 2]];
δo = %%%%[[2, 4, 2]];

ClearAll[f]
f[x_] := Chop[g[x, qo, τo, λo, δo]]
f[2]

px = Plot[Evaluate[f[x]], {x, 0, 30}, PlotRange → All,
    PlotLegends → {"PDF Estimate (Method of Moments)"},
    PlotLegends → {"The estimated density of the q-EGG approximation"}];
Show[
 px,
 pkde]

ClearAll[F1];
F1[y_] := F1[y] = NIntegrate[f[x], {x, 0, y}]
F1[5]

p2 = Plot[F1[x], {x, 0, 30}, PlotStyle → Dashed, PlotRange → All,
    PlotLegends → {"The estimated CDF of the q-EGG approximation"}];
Show[
 ep,
 p2]

ISD = Quiet[Chop[NIntegrate[(F[y] - F1[y])², {y, 0, 100}]]]
```

As expected, the ISD is smaller (by approximately a factor of 2) than in the case of the gamma approximation.

Example 2
Approximating the distribution of an indefinite quadratic form

```
A = {{-12, 0, 0, 0, 0}, {0, -8, 0, 0, 0},
    {0, 0, -4, 0, 0}, {0, 0, 0, 1, 0}, {0, 0, 0, 0, 5}};
MatrixForm[A]
es = Eigensystem[A] // N
Q = {X₁, X₂, X₃, X₄, X₅}.A.{X₁, X₂, X₃, X₄, X₅}
```

$Q = X - Y;$

$X = X_4^2 + 5\, X_5^2;$

$Y = 12\, X_1^2 + 8\, X_2^2 + 4\, X_3^2;$

```
α1 = 0; β1 = 1; α2 = 0; β2 = 1; α3 = 0; β3 = 1; α4 = 0; β4 = 1; α5 = 0; β5 = 1;
Plot[{PDF[NormalDistribution[α1, β1], y]}, {y, -4, 4}, PlotRange → All]
𝒟1 = NormalDistribution[α1, β1];
𝒟2 = NormalDistribution[α2, β2];
𝒟3 = NormalDistribution[α3, β3];
𝒟4 = NormalDistribution[α4, β4];
𝒟5 = NormalDistribution[α5, β5];

size = 10 000;
SeedRandom[1];
data1 = RandomVariate[𝒟1, size];
SeedRandom[2];
data2 = RandomVariate[𝒟2, size];
SeedRandom[3];
data3 = RandomVariate[𝒟3, size];
SeedRandom[4];
data4 = RandomVariate[𝒟4, size];
SeedRandom[5];
data5 = RandomVariate[𝒟5, size];

data = Chop[Table[{data1[[i]], data2[[i]], data3[[i]], data4[[i]], data5[[i]]}.A.
        {data1[[i]], data2[[i]], data3[[i]], data4[[i]], data5[[i]]}, {i, 1, 10 000}]];
ED = EmpiricalDistribution[data];
ClearAll[F]
F[x_] := F[x] = CDF[ED, x];
sP = Plot[CDF[ED, x], {x, -120, 50},
    PlotRange → All, PlotLegends → {"Exact CDF(the whole density)"}]

𝒟 = SmoothKernelDistribution[data, {"Adaptive", Automatic, .45}];
pkde = Plot[PDF[𝒟, x], {x, -120, 50}, PlotStyle → Red, PlotRange → All]

Table[Quantile[data, i / 25], {i, 25}]

ClearAll[μ];
μ[h_] := μ[h] = Expand[({X₁, X₂, X₃, X₄, X₅}.A. {X₁, X₂, X₃, X₄, X₅})^h] /.
    {X₁^(j-·) :→ Moment[NormalDistribution[α1, β1], j],
     X₂^(j-·) :→ Moment[NormalDistribution[α2, β2], j],
     X₃^(j-·) :→ Moment[NormalDistribution[α3, β3], j],
     X₄^(j-·) :→ Moment[NormalDistribution[α4, β4], j],
     X₅^(j-·) :→ Moment[NormalDistribution[α5, β5], j]}
μ[1]
μ[2]
μ[3]
μ[4]
μ[5]
```

Distribution of Y:

```mathematica
datan =
  Chop[Table[{data1[[i]], data2[[i]], data3[[i]]}.{{12, 0, 0}, {0, 8, 0}, {0, 0, 4}}.
      {data1[[i]], data2[[i]], data3[[i]]}, {i, 1, 10 000}]];
EDn = EmpiricalDistribution[datan];
ClearAll[Fn]
Fn[x_] := Fn[x] = CDF[EDn, x];
sPn = Plot[CDF[EDn, x], {x, 0, 100}, PlotRange → All, PlotStyle → Red]

𝒟n = SmoothKernelDistribution[datan, {"Adaptive", Automatic, .45}];
pn = Plot[PDF[𝒟n, x], {x, 0, 100}, PlotStyle → Red, PlotRange → All]

A1 = {{12, 0, 0}, {0, 8, 0}, {0, 0, 4}};

ClearAll[μ];
μ[h_] := μ[h] = Expand[({X₁, X₂, X₃}.A1. {X₁, X₂, X₃})^h] /.
    {X₁^(j-·) :→ Moment[NormalDistribution[α1, β1], j],
     X₂^(j-·) :→ Moment[NormalDistribution[α2, β2], j],
     X₃^(j-·) :→ Moment[NormalDistribution[α3, β3], j]}
μ[1]
μ[2]
μ[3]
μ[4]
```
$$\left\{\alpha = \frac{\mu[1]^2}{\mu[2] - \mu[1]^2}, \ \beta = \frac{\mu[2] - \mu[1]^2}{\mu[1]}\right\}$$

```mathematica
ClearAll[g]
g[x_, q_, τ_, λ_, δ_] :=
```
$$g[x, q, \tau, \lambda, \delta] \ = \ \frac{\text{Abs}\left[\frac{\delta}{\tau}\right] (q-1)^\lambda \, \text{Gamma}\left[\frac{1}{q-1}\right]}{\text{Gamma}\left[\frac{1}{q-1} - \lambda\right] \text{Gamma}[\lambda]} \left(\frac{x}{\tau}\right)^{\lambda*\delta-1} \left(1 + (q-1)\left(\frac{x}{\tau}\right)^\delta\right)^{-\frac{1}{q-1}}$$

$$\text{NMinimize}\Bigg[\Bigg\{\left(\frac{(\tau)^1\,(q-1)^{\frac{-1}{\delta}}\,\text{Gamma}\left[\frac{1}{q-1}-\frac{1}{\delta}-\lambda\right]\,\text{Gamma}\left[\frac{1}{\delta}+\lambda\right]}{\text{Gamma}\left[\frac{1}{q-1}-\lambda\right]\,\text{Gamma}[\lambda]}-24\right)^2+$$

$$\left(\frac{(\tau)^2\,(q-1)^{\frac{-2}{\delta}}\,\text{Gamma}\left[\frac{1}{q-1}-\frac{2}{\delta}-\lambda\right]\,\text{Gamma}\left[\frac{2}{\delta}+\lambda\right]}{\text{Gamma}\left[\frac{1}{q-1}-\lambda\right]\,\text{Gamma}[\lambda]}-1024\right)^2+$$

$$\left(\frac{(\tau)^3\,(q-1)^{\frac{-3}{\delta}}\,\text{Gamma}\left[\frac{1}{q-1}-\frac{3}{\delta}-\lambda\right]\,\text{Gamma}\left[\frac{3}{\delta}+\lambda\right]}{\text{Gamma}\left[\frac{1}{q-1}-\lambda\right]\,\text{Gamma}[\lambda]}-64\,512\right)^2+$$

$$\left(\frac{(\tau)^4\,(q-1)^{\frac{-4}{\delta}}\,\text{Gamma}\left[\frac{1}{q-1}-\frac{4}{\delta}-\lambda\right]\,\text{Gamma}\left[\frac{4}{\delta}+\lambda\right]}{\text{Gamma}\left[\frac{1}{q-1}-\lambda\right]\,\text{Gamma}[\lambda]}-5\,455\,872\right)^2,$$

$$1<q<1.05,\ 18.2<\tau<18.8,\ 1.1<\lambda<1.5,\ \lambda<(q-1)^{-1},\ 0.99<\delta<1.01\Big\},$$

$$\{q,\ \tau,\ \lambda,\ \delta\},\ \text{MaxIterations}\to 10\,000\Bigg]$$

```
qo = %[[2, 1, 2]];
τo = %%[[2, 2, 2]];
λo = %%%[[2, 3, 2]];
δo = %%%%[[2, 4, 2]];

ClearAll[f, F]
f[x_] := g[x, qo, τo, λo, δo]
F[y_] := NIntegrate[f[x], {x, 0, y}]
f[x]

ep = Plot[f[x], {x, 0, 100}, PlotRange → All]
Show[pn, ep]
cdf = Plot[F[y], {y, 0, 100}, PlotRange → All,
    PlotStyle → Dashed, PlotLegends → {"CDF estimation(Negative part)"}];
Show[
 cdf,
 sPn]
```

Distribution of X:

```
A2 = {{1, 0}, {0, 5}};

datap =
  Chop[Table[{data4[[i]], data5[[i]]}.A2.{data4[[i]], data5[[i]]}, {i, 1, 10000}]];
EDp = EmpiricalDistribution[datap];
ClearAll[Fp]
Fp[x_] := Fp[x] = CDF[EDp, x];
sPp = Plot[CDF[EDp, x], {x, 0, 50}, PlotRange → All, PlotStyle → Red]
```

```
𝒟1 = SmoothKernelDistribution[datan, {"Adaptive", Automatic, .45}];
p1 = Plot[PDF[𝒟1, x], {x, 0, 50}, PlotStyle → Red, PlotRange → All]

ClearAll[μ];
μ[h_] := μ[h] =
   Expand[({X₄, X₅}.A2. {X₄, X₅})^h] /. {X₄^j_. :> Moment[NormalDistribution[α4, β4], j],
     X₅^j_. :> Moment[NormalDistribution[α5, β5], j]}
μ[1]
μ[2]
μ[3]
μ[4]
```

$$\left\{\alpha = \frac{\mu[1]^2}{\mu[2] - \mu[1]^2}, \ \beta = \frac{\mu[2] - \mu[1]^2}{\mu[1]}\right\}$$

```
ClearAll[g]
g[x_, q_, τ_, λ_, δ_] :=
```

$$g[x, q, τ, λ, δ] \ = \ \frac{\text{Abs}\left[\frac{\delta}{\tau}\right] \ (q-1)^\lambda \ \text{Gamma}\left[\frac{1}{q-1}\right]}{\text{Gamma}\left[\frac{1}{q-1} - \lambda\right] \ \text{Gamma}[\lambda]} \left(\frac{x}{\tau}\right)^{\lambda*\delta - 1} \left(1 + (q-1)\left(\frac{x}{\tau}\right)^\delta\right)^{-\frac{1}{q-1}}$$

$$\text{NMinimize}\Bigg[\Bigg\{\left(\frac{(\tau)^1 \ (q-1)^{\frac{-1}{\delta}} \ \text{Gamma}\left[\frac{1}{q-1} - \frac{1}{\delta} - \lambda\right] \ \text{Gamma}\left[\frac{1}{\delta} + \lambda\right]}{\text{Gamma}\left[\frac{1}{q-1} - \lambda\right] \ \text{Gamma}[\lambda]} - 6\right)^2 +$$

$$\left(\frac{(\tau)^2 \ (q-1)^{\frac{-2}{\delta}} \ \text{Gamma}\left[\frac{1}{q-1} - \frac{2}{\delta} - \lambda\right] \ \text{Gamma}\left[\frac{2}{\delta} + \lambda\right]}{\text{Gamma}\left[\frac{1}{q-1} - \lambda\right] \ \text{Gamma}[\lambda]} - 88\right)^2 +$$

$$\left(\frac{(\tau)^3 \ (q-1)^{\frac{-3}{\delta}} \ \text{Gamma}\left[\frac{1}{q-1} - \frac{3}{\delta} - \lambda\right] \ \text{Gamma}\left[\frac{3}{\delta} + \lambda\right]}{\text{Gamma}\left[\frac{1}{q-1} - \lambda\right] \ \text{Gamma}[\lambda]} - 2160\right)^2 +$$

$$\left(\frac{(\tau)^4 \ (q-1)^{\frac{-4}{\delta}} \ \text{Gamma}\left[\frac{1}{q-1} - \frac{4}{\delta} - \lambda\right] \ \text{Gamma}\left[\frac{4}{\delta} + \lambda\right]}{\text{Gamma}\left[\frac{1}{q-1} - \lambda\right] \ \text{Gamma}[\lambda]} - 74880\right)^2,$$

$$1 < q < 1.05, \ 8.2 < \tau < 8.8, \ 0.3 < \lambda < 0.8, \ \lambda < (q-1)^{-1}, \ 0.99 < \delta < 1.01\Bigg\},$$

$$\{q, \ \tau, \ \lambda, \ \delta\}, \text{MaxIterations} \to 10000\Bigg]$$

```
qo = %[[2, 1, 2]];
τo = %%[[2, 2, 2]];
λo = %%%[[2, 3, 2]];
δo = %%%%[[2, 4, 2]];
```

```
ClearAll[f2, F2]
f2[x_] := g[x, qo, τo, λo, δo]
f2[x]
Plot[f2[x], {x, 0, 50}, PlotRange → {0, 0.2}]
F2[y_] := NIntegrate[f2[x], {x, 0, y}]
cdf2 = Plot[F2[y], {y, 0, 50}, PlotRange → All,
    PlotStyle → Dashed, PlotLegends → {"CDF estimation(Negative part)"}];
Show[
 cdf2,
 sPp]

ClearAll[gj, g1]
gj[z_, w_] :=
```

$$\frac{0.0297755 \, w^{0.224522241}}{\left(1 + 0.000869 \, w^{1.01`}\right)^{59.549088}} * \frac{0.17570}{(z+w)^{0.32280} \left(1 + 0.003029 \, (z+w)^{1.00045`}\right)^{39.61737}}$$

```
gj[z, w]
g1[z_] := NIntegrate[gj[z, w], {w, 0, ∞}]
L1 = Plot[g1[z], {z, 0, 50}, PlotRange → All]

ClearAll[gt]
gt[z_] := NIntegrate[gj[z, w], {w, -z, Infinity}, MaxRecursion → 500] // Quiet
L2 = Plot[gt[z], {z, -120, 0}, PlotRange → All]

Show[pkde, L2, L1]
```

Chapter 7
Estimating the Proportion of Information Contained in Sets of Moments

---

Section 7.4
Illustrative examples

Subsection 7.4.1
Two exact distributions involving beta random variables

Single skewed beta

Generate points from a single skewed beta density .

The ideal sample with size n consists of the (i/n-1/2n)*100th percentiles (i=1,2,...,n) of the distribution.

The percentiles can be determined using the function Quantile[dist,q] in Mathematica.

The underlying distribution

```
α1 = 3; β1 = 9;
𝒟 = BetaDistribution[α1, β1];
ClearAll[f]
f[x_] := f[x] = PDF[𝒟, x]
lb = 0; ub = 1;
{Plot[PDF[𝒟, x], {x, 0, 1}, ImageSize → Medium],
 po = Plot[CDF[𝒟, x], {x, 0, 1}, ImageSize → Medium]}
```

Determining the $\left(\frac{i}{n} - \frac{1}{2n}\right)$th percentiles and calculating the Kullback-Leibler (KL) divergence (also called relative entropy)
https://blogs.sas.com/content/iml/2020/06/01/the-kullback-leibler-divergence-between-continuous-probability-distributions.html

```
ClearAll[c, point, sample, KLf]
```

```mathematica
ClearAll[method1]

method1[n_] := (

  point[n] = Table[Quantile[𝒟, i/n - 1/(2 n)], {i, n}] // N;

  sample[n] = Table[{point[n][[i]], i/n - 1/(2 n)}, {i, 1, n}] // Insert[{lb, 0}, 1] //
    Insert[{ub, 1}, -1];
  c[n] := ResourceFunction["CubicMonotonicInterpolation"][
    sample[n], Method → "FritschCarlson"];
  KLf[n] = Quiet[NIntegrate[f[x] * (Log[f[x]] - Log[c[n]'[x]]), {x, lb, ub}]];
  Print["sample[", n, "]=", sample[n]];
  Print["KLf[", n, "]", "=", NumberForm[KLf[n], {4, 4}]];
  Plot[{f[x], c[n]'[x]}, {x, lb, ub},
   AxesOrigin → {lb, 0}, PlotStyle → {, Red}, PlotLabel → n, PlotRange → All]

 )

Table[method1[n], {n, 2, 20}]
```

Also show some cdf's with the representative sample points

```mathematica
ClearAll[method2]

method2[n_] := (

  point[n] = Table[Quantile[𝒟, i/n - 1/(2 n)], {i, n}] // N;

  sample[n] = Table[{point[n][[i]], i/n - 1/(2 n)}, {i, 1, n}] // Insert[{lb, 0}, 1] //
    Insert[{ub, 1}, -1];
  c[n] := ResourceFunction["CubicMonotonicInterpolation"][
    sample[n], Method → "FritschCarlson"];
  {Show[Plot[CDF[𝒟, x], {x, lb, ub}], Plot[c[n][x], {x, lb, ub}, PlotStyle → Red],
    ListPlot[sample[n], PlotStyle → Black], PlotLabel → n],
   Plot[{f[x], c[n]'[x]}, {x, lb, ub}, AxesOrigin → {lb, 0},
    PlotStyle → {, Red}, PlotRange → All, PlotLabel → n]}

 )

Table[method2[i], {i, {2, 3, 4, 5, 10, 20}}]
```

Results

Note that the uniform distribution whose pdf is a constant and is therefore non-informative.

Proportion of information in the first h moments
KLf[h] means KL[f[x], c[h]'[x]]

```mathematica
Table[{h, NumberForm[KLf[h], {6, 6}]}, {h, 2, 20}] //
  Insert[{"h", "KLf[h]"}, 1] // MatrixForm
```

Results based on $1 - \dfrac{\text{KLf}[h]}{\text{KLfc}}$

1. Proportion info in the first h^{th} moment

2. Successive info in h^{th} moment (in first two and in the third ~ 20 th)

```
KLfc = NIntegrate[f[x] * (Log[f[x]] - Log[ 1/(ub - lb) ]), {x, lb, ub}]

{Table[{h, NumberForm[1 - KLf[h]/KLfc, {6, 6}]}, {h, 2, 20}] //

   Insert[{"h", "1- KLf[h]/KLfc "}, 1] // MatrixForm,

 ListLinePlot[Table[{h, 1 - KLf[h]/KLfc}, {h, 2, 20}] // Insert[{0, 0}, 1],

   PlotLabel → "1- KLf[h]/KLfc ", ImageSize → Medium,

   PlotMarkers → Automatic, PlotRange → {All, 1}]}

info = Table[{h, (KLf[h - 1] - KLf[h])/KLfc}, {h, 3, 20}] // Insert[{2, 1 - KLf[2]/KLfc}, 1];

{info // Insert[{"h", " (KLf[h - 1] - KLf[h])/KLfc ( (KLfc - KLf[2])/KLfc  when h=2)"}, 1] // MatrixForm,

 DiscretePlot[info[[h - 1, 2]], {h, 2, 20},

   ImageSize → Medium, PlotStyle → Red, PlotRange → {All, {All, 1}}]}
```

Mixture of beta pdf's

Generate points from an equal mixture of two beta densities .

The ideal sample with size n consists of the (i/n-1/2n)*100th percentiles (i=1,2,…,n) of the distribution.

The percentiles can be determined using the function Quantile[dist,q] in Mathematica.

The underlying distribution

```
α1 = 7; β1 = 15; α2 = 12; β2 = 8;
𝒟1 = BetaDistribution[α1, β1];
𝒟2 = BetaDistribution[α2, β2];
ClearAll[f]

f[x_] := f[x] = 1/2 (PDF[𝒟1, x] + PDF[𝒟2, x])

lb = 0; ub = 1;

𝒟 = ProbabilityDistribution[f[x], {x, 0, 1}];
{Plot[f[x], {x, 0, 1}, PlotRange → All, ImageSize → Medium],
 po = Plot[CDF[𝒟, x], {x, 0, 1}, ImageSize → Medium]}
```

Determining the $\left(\dfrac{i}{n} - \dfrac{1}{2n}\right)$th percentiles and calculating the Kullback-Leibler (KL) divergence (also called relative entropy)

https://blogs.sas.com/content/iml/2020/06/01/the-kullback-leibler-divergence-between-continu-ous-probability-distributions.html

```
ClearAll[c, point, sample, KLf]
```

n=2~20

```
ClearAll[method1]

method1[n_] := (

  point[n] = Table[Quantile[𝒟, i/n - 1/(2 n)], {i, n}] // N;

  sample[n] = Table[{point[n][[i]], i/n - 1/(2 n)}, {i, 1, n}] // Insert[{lb, 0}, 1] //
    Insert[{ub, 1}, -1];
  c[n] := ResourceFunction["CubicMonotonicInterpolation"][
    sample[n], Method → "FritschCarlson"];

  KLf[n] = Quiet[NIntegrate[f[x] * (Log[f[x]] - Log[c[n]'[x]]), {x, lb, ub}]];

  Print["sample[", n, "]=", sample[n]];
  Print["KLf[", n, "]", "=", NumberForm[KLf[n], {4, 4}]];
  Plot[{f[x], c[n]'[x]}, {x, lb, ub},
    AxesOrigin → {lb, 0}, PlotStyle → {, Red}, PlotLabel → n, PlotRange → All]

)

Table[method1[n], {n, 2, 20}]
```

Also show some cdf's with the representative sample points

```
ClearAll[method2]

method2[n_] := (

  point[n] = Table[Quantile[𝒟, i/n - 1/(2 n)], {i, n}] // N;

  sample[n] = Table[{point[n][[i]], i/n - 1/(2 n)}, {i, 1, n}] // Insert[{lb, 0}, 1] //
    Insert[{ub, 1}, -1];
  c[n] := ResourceFunction["CubicMonotonicInterpolation"][
    sample[n], Method → "FritschCarlson"];
  {Show[Plot[CDF[𝒟, x], {x, lb, ub}], Plot[c[n][x], {x, lb, ub}, PlotStyle → Red],
    ListPlot[sample[n], PlotStyle → Black], PlotLabel → n],
   Plot[{f[x], c[n]'[x]}, {x, lb, ub}, AxesOrigin → {lb, 0},
    PlotStyle → {, Red}, PlotRange → All, PlotLabel → n]}

)

Table[method2[i], {i, {2, 3, 4, 5, 10, 20}}]
```

Results

Note that the uniform distribution whose pdf is a constant and is therefore non-informative.

Proportion of information in the first h moments

KLf[h] means KL[f[x], c[h]'[x]]

```
Table[{h, NumberForm[KLf[h], {6, 6}]}, {h, 2, 20}] //
  Insert[{"h", "KLf[h]"}, 1] // MatrixForm
```

Results based on $1 - \dfrac{\text{KLf}[h]}{\text{KLfc}}$

1. Proportion info in the first h^{th} moment

2. Successive info in h^{th} moment (in first two and in the third ~ 20 th)

```
KLfc = NIntegrate[f[x] * (Log[f[x]] - Log[ 1/(ub - lb) ]), {x, lb, ub}]
```

$$\{ \text{Table}[\{h, \text{NumberForm}[1 - \dfrac{\text{KLf}[h]}{\text{KLfc}}, \{6, 6\}]\}, \{h, 2, 20\}] //$$

$$\text{Insert}[\{"h", "1-\dfrac{\text{KLf}[h]}{\text{KLfc}}"\}, 1] // \text{MatrixForm},$$

$$\text{ListLinePlot}[\text{Table}[\{h, 1 - \dfrac{\text{KLf}[h]}{\text{KLfc}}\}, \{h, 2, 20\}] // \text{Insert}[\{0, 0\}, 1],$$

$$\text{PlotLabel} \to "1-\dfrac{\text{KLf}[h]}{\text{KLfc}}", \text{ImageSize} \to \text{Medium},$$

$$\text{PlotMarkers} \to \text{Automatic}, \text{PlotRange} \to \{\text{All}, 1\}]\}$$

$$\text{info = Table}[\{h, \dfrac{\text{KLf}[h-1] - \text{KLf}[h]}{\text{KLfc}}\}, \{h, 3, 20\}] // \text{Insert}[\{2, 1 - \dfrac{\text{KLf}[2]}{\text{KLfc}}\}, 1];$$

$$\{\text{info} // \text{Insert}[\{"h", "\dfrac{\text{KLf}[h-1] - \text{KLf}[h]}{\text{KLfc}} (\dfrac{\text{KLfc} - \text{KLf}[2]}{\text{KLfc}} \text{ when h=2)}"\}, 1] // \text{MatrixForm},$$

$$\text{DiscretePlot}[\text{info}[[h-1, 2]], \{h, 2, 20\},$$

$$\text{ImageSize} \to \text{Medium}, \text{PlotStyle} \to \text{Red}, \text{PlotRange} \to \{\text{All}, \{\text{All}, 1\}\}]\}$$

Subsection 7.4.2
Two actual data sets

Buffalo snowfall data

An actual data set, Buffalo snowfall data.

The ideal sample with size n consists of the (i/n-1/2n)*100th percentiles (i=1,2,...,n) of the distribution.

The percentiles can be determined using the function Quantile[dist,q] in Mathematica.

The underlying distribution is a density estimate based on the data set

```
Y = ExampleData[{"Statistics", "BuffaloSnow"}]
Length[Y]
{Mean[Y], StandardDeviation[Y]}
{Min[Y], Max[Y]}
```

Use a KDE as the underlying distribution

229

```
𝒟 = SmoothKernelDistribution[Y];
(*Plot[PDF[𝒟,x],{x,Min[Y]-StandardDeviation[Y],Max[Y]+StandardDeviation[Y]}]*)
{lb = FindRoot[CDF[𝒟, x] == 10^-4, {x, 20}][[1, 2]],
  ub = FindRoot[CDF[𝒟, x] == 1 - 10^-4, {x, 130}][[1, 2]]};
Print["{lb,ub}=", {lb, ub}]
{po = Plot[CDF[𝒟, x], {x, lb, ub}, ImageSize → Medium],
 Plot[PDF[𝒟, x], {x, lb, ub}, ImageSize → Medium]}
```

Determining the $\left(\frac{i}{n} - \frac{1}{2n}\right)$th percentiles and calculating the Kullback-Leibler (KL) divergence (also called relative entropy)

https://blogs.sas.com/content/iml/2020/06/01/the-kullback-leibler-divergence-between-continuous-probability-distributions.html

```
ClearAll[c, point, sample, KLf, f]
f[x_] := f[x] = PDF[𝒟, x]
```

n=2~20

```
ClearAll[method1]

method1[n_] := (

  point[n] = Table[Quantile[𝒟, i/n - 1/(2 n)], {i, n}] // N;

  sample[n] = Table[{point[n][[i]], i/n - 1/(2 n)}, {i, 1, n}] // Insert[{lb, 0}, 1] //
    Insert[{ub, 1}, -1];
  c[n] := ResourceFunction["CubicMonotonicInterpolation"][
    sample[n], Method → "FritschCarlson"];

  KLf[n] = Quiet[NIntegrate[
      f[x] * (Log[f[x]] - Log[c[n]'[x]]), {x, lb, ub}, Method → "QuasiMonteCarlo"]];

  Print["sample[", n, "]=", sample[n]];
  Print["KLf[", n, "]", "=", NumberForm[KLf[n], {4, 4}]];
  Plot[{f[x], c[n]'[x]}, {x, lb, ub},
   AxesOrigin → {lb, 0}, PlotStyle → {, Red}, PlotLabel → n, PlotRange → All]

  )

Table[method1[n], {n, 2, 20}]
```

Also show some cdf's with the representative sample points

```mathematica
ClearAll[method2]
method2[n_] := (

  point[n] = Table[Quantile[𝒟, i/n - 1/(2 n)], {i, n}] // N;

  sample[n] = Table[{point[n][[i]], i/n - 1/(2 n)}, {i, 1, n}] // Insert[{lb, 0}, 1] //
    Insert[{ub, 1}, -1];
  c[n] := ResourceFunction["CubicMonotonicInterpolation"][
    sample[n], Method → "FritschCarlson"];
  {Show[Plot[CDF[𝒟, x], {x, lb, ub}], Plot[c[n][x], {x, lb, ub}, PlotStyle → Red],
    ListPlot[sample[n], PlotStyle → Black], PlotLabel → n],
   Plot[{f[x], c[n]'[x]}, {x, lb, ub}, AxesOrigin → {lb, 0},
    PlotStyle → {, Red}, PlotRange → All, PlotLabel → n]}

)

Table[method2[i], {i, {2, 3, 4, 5, 10, 20}}]
```

Results

Note that the uniform distribution whose pdf is a constant and is therefore non-informative.

Proportion of information in the first h moments
KLf[h] means KL[f[x], c[h]'[x]]

```mathematica
Table[{h, NumberForm[KLf[h], {6, 6}]}, {h, 2, 20}] //
  Insert[{"h", "KLf[h]"}, 1] // MatrixForm
```

Results based on $1 - \dfrac{KLf[h]}{KLfc}$

1. Proportion info in the first h^{th} moment

2. Successive info in h^{th} moment (in first two and in the third ∼ 20 th)

```mathematica
KLfc =
  NIntegrate[f[x] * (Log[f[x]] - Log[1/(ub - lb)]), {x, lb, ub}, Method → "QuasiMonteCarlo"]
{Table[{h, NumberForm[1 - KLf[h]/KLfc, {6, 6}]}, {h, 2, 20}] //

   Insert[{"h", "1-KLf[h]/KLfc"}, 1] // MatrixForm,

 ListLinePlot[Table[{h, 1 - KLf[h]/KLfc}, {h, 2, 20}] // Insert[{0, 0}, 1],

  PlotLabel → "1-KLf[h]/KLfc", ImageSize → Medium,
  PlotMarkers → Automatic, PlotRange → {All, 1}]}
```

```
info = Table[{h, (KLf[h - 1] - KLf[h])/KLfc}, {h, 3, 20}] // Insert[{2, 1 - KLf[2]/KLfc}, 1];
{info // Insert[{"h", "(KLf[h - 1] - KLf[h])/KLfc (KLfc - KLf[2])/KLfc when h=2)"}, 1] // MatrixForm,
  DiscretePlot[info[[h - 1, 2]], {h, 2, 20},
    ImageSize → Medium, PlotStyle → Red, PlotRange → {All, {All, 1}}]}
```

Life expectancy data

An actual data set

Life expectancy data in 183 countries/regions for the year 2015 recorded by WHO

This data set is a left skewed

The ideal sample with size n consists of the (i/n-1/2n)*100th percentiles (i=1,2,...,n) of the distribution.

The percentiles can be determined using the function Quantile[dist,q] in Mathematica.

The underlying distribution is a density estimate based on the data set

```
Y = {65, 77.8, 75.6, 52.4, 76.4, 76.3, 74.8, 82.8, 81.5, 72.7, 76.1, 76.9, 71.8, 75.5, 72.3,
  81.1, 71, 60, 69.8, 77, 77.4, 65.7, 75, 77.7, 74.5, 59.9, 59.6, 53.3, 73.3, 68.7,
  57.3, 82.2, 52.5, 53.1, 85, 76.1, 74.8, 63.5, 64.7, 79.6, 78, 79.1, 85, 78.8,
  76, 59.8, 86, 63.5, 73.9, 76.2, 79, 73.5, 58.2, 64.7, 77.6, 64.8, 69.9, 81.1,
  82.4, 66, 61.1, 74.4, 81, 62.4, 81, 73.6, 71.9, 59, 58.9, 66.2, 63.5, 74.6, 75.8,
  82.7, 68.3, 69.1, 75.5, 68.9, 81.4, 82.5, 82.7, 76.2, 83.7, 74.1, 72, 63.4, 66.3,
  74.7, 71.1, 65.7, 74.6, 74.9, 53.7, 61.4, 72.7, 73.6, 82, 65.5, 58.3, 75, 78.5,
  58.2, 81.7, 63.1, 74.6, 76.7, 69.4, 68.8, 76.1, 74.3, 57.6, 66.6, 65.8, 69.2,
  81.9, 81.6, 74.8, 61.8, 54.5, 81.8, 76.6, 66.4, 77.8, 62.9, 74, 75.5, 68.5, 77.5,
  81.1, 78.2, 82.3, 72.1, 75, 75, 66.1, 75.2, 73.2, 74, 67.5, 74.5, 66.7, 75.6,
  73.2, 51, 83.1, 76.7, 88, 69.2, 55, 62.9, 57.3, 82.8, 74.9, 64.1, 71.6, 58.9,
  82.4, 83.4, 64.5, 69.7, 74.9, 75.7, 68.3, 59.9, 73.5, 71.2, 75.3, 75.8, 66.3,
  62.3, 71.3, 77.1, 81.2, 61.8, 79.3, 77, 69.4, 72, 74.1, 76, 65.7, 61.8, 67};
Length[Y]
{Mean[Y], StandardDeviation[Y]}
{Min[Y], Max[Y]}
```

Use a KDE as the underlying distribution

```
𝒟 = SmoothKernelDistribution[Y];
(*Plot[PDF[𝒟,x],{x,Min[Y]-StandardDeviation[Y],Max[Y]+StandardDeviation[Y]}]*)
{lb = FindRoot[CDF[𝒟, x] == 10^-4, {x, 55}][[1, 2]],
  ub = FindRoot[CDF[𝒟, x] == 1 - 10^-4, {x, 85}][[1, 2]]};
Print["{lb,ub}=", {lb, ub}]
{po = Plot[CDF[𝒟, x], {x, lb, ub}, ImageSize → Medium],
 Plot[PDF[𝒟, x], {x, lb, ub}, ImageSize → Medium]}
```

Determining the $\left(\frac{i}{n} - \frac{1}{2n}\right)$th percentiles and calculate the Kullback-Leibler (KL) divergence (also called relative entropy)

https://blogs.sas.com/content/iml/2020/06/01/the-kullback-leibler-divergence-between-continu-

ous-probability-distributions.html

```
ClearAll[c, point, sample, KLf, f]
f[x_] := f[x] = PDF[𝒟, x]
```

n=2~20

```
ClearAll[method1]

method1[n_] := (

  point[n] = Table[Quantile[𝒟, i/n - 1/(2 n)], {i, n}] // N;

  sample[n] = Table[{point[n][[i]], i/n - 1/(2 n)}, {i, 1, n}] // Insert[{lb, 0}, 1] //
     Insert[{ub, 1}, -1];
  c[n] := ResourceFunction["CubicMonotonicInterpolation"][
     sample[n], Method → "FritschCarlson"];

  KLf[n] = Quiet[NIntegrate[
      f[x] * (Log[f[x]] - Log[c[n]'[x]]), {x, lb, ub}, Method → "QuasiMonteCarlo"]];

  Print["sample[", n, "]=", sample[n]];
  Print["KLf[", n, "]", "=", NumberForm[KLf[n], {4, 4}]];
  Plot[{f[x], c[n]'[x]}, {x, lb, ub},
   AxesOrigin → {lb, 0}, PlotStyle → {, Red}, PlotLabel → n, PlotRange → All]

)

Table[method1[n], {n, 2, 20}]
```

Also show some cdf's with the representative sample points

```
ClearAll[method2]

method2[n_] := (

  point[n] = Table[Quantile[𝒟, i/n - 1/(2 n)], {i, n}] // N;

  sample[n] = Table[{point[n][[i]], i/n - 1/(2 n)}, {i, 1, n}] // Insert[{lb, 0}, 1] //
     Insert[{ub, 1}, -1];
  c[n] := ResourceFunction["CubicMonotonicInterpolation"][
     sample[n], Method → "FritschCarlson"];
  {Show[Plot[CDF[𝒟, x], {x, lb, ub}], Plot[c[n][x], {x, lb, ub}, PlotStyle → Red],
    ListPlot[sample[n], PlotStyle → Black], PlotLabel → n],
   Plot[{f[x], c[n]'[x]}, {x, lb, ub}, AxesOrigin → {lb, 0},
    PlotStyle → {, Red}, PlotRange → All, PlotLabel → n]}

)

Table[method2[i], {i, {2, 3, 4, 5, 10, 20}}]
```

Results

Note that the uniform distribution whose pdf is a constant and is therefore non-informative.

Proportion of information in the first h moments
KLf[h] means KL[f[x], c[h]'[x]]

```
Table[{h, NumberForm[KLf[h], {6, 6}]}, {h, 2, 20}] //
   Insert[{"h", "KLf[h]"}, 1] // MatrixForm
```

Results based on $1 - \dfrac{KLf[h]}{KLfc}$

1. Proportion info in the first h^{th} moment

2. Successive info in h^{th} moment (in first two and in the third ~ 20 th)

```
KLfc =
  NIntegrate[f[x] * (Log[f[x]] - Log[ 1/(ub - lb) ]), {x, lb, ub}, Method → "QuasiMonteCarlo"]
{Table[{h, NumberForm[1 - KLf[h]/KLfc, {6, 6}]}, {h, 2, 20}] //
    Insert[{"h", "1- KLf[h]/KLfc "}, 1] // MatrixForm,
  ListLinePlot[Table[{h, 1 - KLf[h]/KLfc}, {h, 2, 20}] // Insert[{0, 0}, 1],
    PlotLabel → "1- KLf[h]/KLfc ", ImageSize → Medium,
   PlotMarkers → Automatic, PlotRange → {All, 1}]}

info = Table[{h, (KLf[h - 1] - KLf[h])/KLfc}, {h, 3, 20}] // Insert[{2, 1 - KLf[2]/KLfc}, 1];
{info // Insert[{"h", " (KLf[h - 1] - KLf[h])/KLfc ( (KLfc - KLf[2])/KLfc  when h=2)"}, 1] // MatrixForm,
 DiscretePlot[info[[h - 1, 2]], {h, 2, 20},
   ImageSize → Medium, PlotStyle → Red, PlotRange → {All, {All, 1}}]}
```

234

Curriculum Vitae

Education

Ph.D. candidate in Statistics   2019 - Present

The University of Western Ontario, London, ON

- Supervisor: Dr. Serge B. Provost

- Thesis title: Advances in Copula Estimation and Distribution Theory

Master of Science in Statistics   2017 - 2019

The University of Western Ontario, London, ON

- Supervisor: Dr. Serge B. Provost

- Thesis title: Advances in Moment-Based Distributional Methodologies

Bachelor of Science in Statistics   2013 - 2017

Shandong University, Jinan, China

Work Experience

Teaching Assistant   2018 - Present

The University of Western Ontario, London, ON

Research Assistant   2018 - Present

The University of Western Ontario, London, ON

Conference

The 4th International Conference on Statistics: Theory and Applications, Prague, Czech Republic, 2022. Presenter

Publications

[1] A methodology for characterising the tail behaviour of a distribution. *International Journal of Mathematical and Computational Sciences*, 14(6), 57–60.

[2] A class of copulas associated with Brownian motion processes and their maxima. *Journal of Applied Mathematics and Computation*, 6(1), 96–120.

[3] Practical representations of copula and joint density estimates. *Journal of Machine Intelligence and Data Science*, 4, 1–5.

Working Papers

[4] Certain measures of association and their representations in terms of copulas. (To be submitted to Statistical Reviews).

[5] On various copula density estimates and the determination of their tuning parameters. (To be submitted to Communications in Statistics - Theory and Methods).