

Electronic Thesis and Dissertation Repository

---

12-12-2013 12:00 AM

## Credibility theory for phase-type distributions

Yanyan Zang, *The University of Western Ontario*

Supervisor: David Stanford, *The University of Western Ontario*

Joint Supervisor: Amin Hassan-Zadeh, *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

© Yanyan Zang 2013

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

---

### Recommended Citation

Zang, Yanyan, "Credibility theory for phase-type distributions" (2013). *Electronic Thesis and Dissertation Repository*. 1833.

<https://ir.lib.uwo.ca/etd/1833>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

CREDIBILITY THEORY FOR PHASE-TYPE DISTRIBUTIONS

(Thesis format: Monograph)

by

Yanyan Zang

Graduate Program in Statistical & Actuarial Sciences

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies  
The University of Western Ontario  
London, Ontario, Canada

© Yanyan Zang 2013

# Abstract

Credibility theory provides important guidelines for insurers in the practice of experience rating. It recognizes multiple sources of risk and proposes potential premium adjustments by considering individual experiences along with the class experiences. Two popular tools in credibility theory are Bayesian and Bühlmann premium estimators. This thesis develops both models assuming a phase-type distribution of losses, following a Bayesian inference approach. A family of conjugate priors is first established accordingly. The solutions for both Bayesian and Bühlmann estimators are then obtained in explicit forms. Simulation studies are performed to evaluate each estimator individually as well as to conduct comparisons where appropriate. Mean squared errors for each estimator are computed based on different prior choices and outcomes are compared against theoretical results.

**Keywords:** credibility theory, Bühlmann premium, Bayesian premium, phase-type distribution

## Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. David Stanford and co-supervisor, Dr. Amin Hassan-Zadeh for their considerable help, excellent advice and great support, which guided me through this work. Their insights and knowledge had a profound effect on my thesis and have helped me overcome various challenges.

I would also like to thank Dr. Gordon Willmot, Dr. Geoff Wild, Dr. Xiaoming Liu and Dr. Jiandong Ren for many stimulating discussions and providing valuable advice during my thesis defense.

My special thanks to my office mate Cheng Qian, Huan Gao, Zhe Sheng for providing a nice and comfortable office environment. I am grateful that all people from the department are extremely supportive and kind.

Finally, I am deeply grateful to my family and friends, especially to my parents, my uncle Zhongning, my aunt Yan and my boyfriend Harry for their unlimited support and encouragement along the way, which made the completion of this thesis possible.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Preliminaries</b>	<b>9</b>
2.1 Credibility Theory . . . . .	9
2.1.1 Limited Fluctuation Credibility Theory . . . . .	9
2.1.2 Greatest Accuracy Credibility Theory . . . . .	13
2.2 Phase-type Distributions . . . . .	16
2.2.1 Continuous Phase-type Distributions . . . . .	16

2.2.2	Discrete Phase-type Distributions . . . . .	19
2.3	Proofs . . . . .	21
2.3.1	Proof of Proposition 2.2 . . . . .	21
<b>3</b>	<b>Bühlmann Premium for Phase-type Distributed Losses</b>	<b>24</b>
3.1	The Bühlmann Premium . . . . .	24
3.2	Expressing Phase-type Distributions as Infinite Mixtures of Erlang Distributions . . . . .	29
3.3	Conjugate Prior for Phase-type Distributions . . . . .	33
3.4	Bühlmann Premium for Phase-type Distributed Losses . . . . .	36
3.5	Examples . . . . .	38
3.6	Proofs . . . . .	50
3.6.1	Proof of Proposition 3.1 . . . . .	50
3.6.2	Proof of Corollary 3.2 . . . . .	52
3.6.3	Proof of Theorem 3.4 . . . . .	53
3.6.4	Proof of Theorem 3.5 . . . . .	59
<b>4</b>	<b>Bayesian Premium for Phase-type Distributed Losses</b>	<b>63</b>
4.1	The Bayesian Premium . . . . .	64
4.2	Bayesian Premium for Phase-type Distributed Losses . . . . .	66
4.3	Marginal and Predictive Distributions for the Losses . . . . .	69

4.4	Examples . . . . .	72
4.5	Proofs . . . . .	97
4.5.1	Proof of Lemma 4.1 . . . . .	97
4.5.2	Proofs of Theorem 4.2 . . . . .	103
4.5.3	Proof of Corollary 4.3 . . . . .	106
4.5.4	Proofs of Corollary 4.4 . . . . .	109
<b>5</b>	<b>Conclusions and Future Work</b>	<b>111</b>
	<b>Bibliography</b>	<b>115</b>
	<b>Curriculum Vitae</b>	<b>122</b>

# List of Figures

3.1	MSE comparison Bühlmann (solid line) vs. $\bar{X}$ (dashed line) . . .	45
3.2	Bühlmann premium (solid line) vs. $\mu(\theta)$ (dashed line) with $\theta = 0.2$	48
3.3	Bühlmann premium (solid line) vs. $\mu(\theta)$ (dashed line) with $\theta = 0.65$	49
4.1	Density plot of prior distribution: finite mixtures of 3 Gammas . .	79
4.2	Density estimation based on 200 samples of $\theta$ . . . . .	80
4.3	$MSE(P_{Cr})$ minus $MSE(P_{Ba})$ : finite mixtures of 3 Gammas . . .	81
4.4	Bayesian premiums at different $\theta$ : finite mixtures of 3 Gammas . .	83
4.5	Difference in premiums, $P_{Cr} - P_{Ba}$ : finite mixtures of 3 Gammas .	84
4.6	Density plot of prior distribution: finite mixtures of 6 Gammas . .	86
4.7	$MSE(P_{Cr})$ minus $MSE(P_{Ba})$ : finite mixtures of 6 Gammas . . .	87
4.8	Bühlmann premiums (left) and Bayesian (right) premiums at dif- ferent $\theta$ : finite mixtures of 6 Gammas . . . . .	88
4.9	Difference in premiums, $P_{Cr} - P_{Ba}$ : finite mixtures of 6 Gammas .	89



4.10	Bühlmann (solid line) and Bayesian premiums (dashed line) vs. $\theta$ :	
	higher dimension PH . . . . .	91
4.11	Bühlmann (solid line) and Bayesian premiums (dashed line) vs. $\theta$ :	
	higher dimension PH . . . . .	94
4.12	Histogram for the difference $P_{Cr} - P_{Ba}$ : higher dimension PH . .	95

# List of Tables

3.1	Descriptive statistics for prior distribution . . . . .	44
3.2	Squared error losses (S.E.L): Bühlmann vs. $\bar{X}$ . . . . .	47
4.1	Bühlmann and Bayesian Premiums . . . . .	92
4.2	Bühlmann and Bayesian Premiums for Figure 4.11 . . . . .	96

# Chapter 1

## Introduction

Experience rating is one of the most important practices for insurers. Insurance companies set up experience rating systems to determine pricing of premiums for different groups or individuals based on their past experience. In a competitive market nowadays, insurers want to determine individual premiums as precisely as possible. It is crucial for them to know the optimal pricing: if it were set too low, the insurer would face solvency issues when large claims occur; if it were set too high then the company would compromise its competitiveness.

Individual policyholders are usually divided into different groups according to their deemed “risk levels”, which are often assessed during the underwriting process based on a variety of relevant factors. A manual rate is then introduced for each group to represent the expected experience arising from the unique risk

characteristics of the class.

In the insurance industry, manual rates are typically calculated based on large volumes of data obtained from respective blocks of business. One implicit assumption embedded in the manual rate is that the underlying risk level is uniformly the same for each member of the class, which is sometimes referred to as “homogeneity” by actuaries. However, as Bühlmann and Gisler (2005) pointed out, there are actually no homogeneous risk classes in insurance. Empirical evidence has suggested that individual experiences may vary considerably even within the same risk group, because no risk is exactly the same as another. Norberg (1979) also point out that such heterogeneity may only appear to the insurer through the individual claims records. Therefore, insurance premiums cannot be solely determined by manual rates. Unique individual experiences also need to be taken into account.

Credibility theory provides additional tools to adjust the risk premiums by combining individual experiences with class experiences. It quantifies the difference between a unique “individual risk” and a more general “collective risk”. The premium of a policyholder is then adjusted according to his or her history of claims. For instance, the policyholder with a favorable record of claims may demand a discount while the ones with larger claim sizes may be subjected to certain levels of premium increase. Credibility models also consider how “credible” the experience data is. It helps insurers to understand whether a favorable claim is just essentially some random fluctuation in the individual experiences or

indicating a genuine better risk.

There are two main approaches in credibility theory. The first and earlier approach is called limited fluctuation credibility, also sometimes referred to as American credibility. This approach was first introduced by Mowbray (1914) in the study of workers compensation insurance. He intended to find the minimal number of employees needed so that the risk experience of the employer would be statistically stable enough to be considered as fully credible. According to this approach, an adjustment of the premium is appropriate only when the experience of the insured is significant enough to produce a reliable estimate.

The approach of limited fluctuation credibility is relatively simple in principle and straightforward to apply. However, it has several drawbacks. First of all, full credibility is usually difficult to achieve for insurers in practice as it usually requires a large amount of experience data. Secondly, this framework lacks a fine probabilistic structure. While it may be hard to weigh its pros and cons, this does limit our options of applying certain statistical inference tools. Finally, the calculation of either full or partial credibility involves certain parameters that are selected without mathematical guidance. In the absence of an underlying statistical basis, it may make it difficult to evaluate the effectiveness of the estimator and make comparison with other results. We will discuss this in more detail in the next chapter.

The second and more modern approach is known as greatest accuracy credibility theory. Under this approach, a hidden risk parameter denoted by  $\Theta$  is

assumed for each risk class. To reflect the heterogeneity of individual experiences within a class, this risk parameter is presumed to follow a postulated probability law instead of being a constant. Different values of this random variable are then assigned to policyholders as their risk level. This approach accounts for heterogeneities both within and outside of a risk class: the variation of individual risk within a class is represented by different values of the risk parameter while the difference between classes is indicated by choosing different distributions for this parameter.

For a particular policyholder, his or her risk characteristic is featured by the realized value of  $\Theta$ , which we denote by  $\theta$ . Each claim is then viewed as a random number drawn from a conditional distribution depending on  $\theta$ . Therefore, the expectation of the claim can be calculated based on this conditional distribution, which is also known as the risk premium. This quantity would be a natural candidate for the purpose of pricing but unfortunately, since  $\theta$  is never observed the risk premium is only a hypothetical true value.

In practice, different models have been proposed to find the best estimator of this true premium. Formulations could be accomplished through either frequentist or Bayesian inference. The latter one is usually preferred as it introduces multiple sources of variations through different distributions. The process is Bayesian in nature and starts by imposing a prior distribution for  $\Theta$  and another distribution describing the risk experience given a particular  $\theta$ . Conditional on that information the posterior distribution which establishes the dependence of  $\theta$  upon

the experience history can be obtained. The Bayesian premium is defined as the expectation of the future claim given past experience. For more general accounts of Bayesian analysis refer to Berger (1985), Klugman et al. (2008), Gelman et al. (2003).

It is well known that the Bayesian premium enjoys the advantage of being the estimator with the least squared error loss, cf. Klugman et al. (2008). However, the explicit form of this estimator could be quite difficult to obtain as it involves a number of integrations, whose computational complexity depends upon our choice of the distributions. An alternative approach that address this problem is to restrict the consideration within the class of linear estimators. This idea was first explored by several authors including Whitney (1918), Bailey (1945, 1950) and Mayerson (1964). The most important contribution is due to Bühlmann (1967, 1969) where a general linear estimator was obtained without imposing any specific distributional assumption. He also showed that the obtained estimator has the lowest squared error loss with respect to hypothetical risk premium amongst all linear estimators. This result immediately attracted a great amount of attention and has been widely applied by practitioners. The resulting estimate is known nowadays as the Bühlmann premium or credibility premium.

Both Bayesian and Bühlmann premiums are important tools in actuarial science. For comprehensive accounts of credibility models, see Goulet (2008) and the monograph by Klugman (1992). There are also a number of other models available proposed by different authors, for examples Heilmann (1989), Landsman

and Makov (1998, 2000), Bühlmann and Gisler (2005). In recent years, credibility models have been extensively applied not only around property and casualty insurance but also increasingly in group life insurance, for example see Tschupp (2011). Klugman et al. (2009) also pointed out that life and annuity actuaries now face issues of using company experience data for both the organizational overall risk assessment and the preparation of principle-based reserving. These issues tend to be resolved based on the statistical credibility methods along with actuarial judgement.

The focus of this thesis is on the determination of Bayesian and Bühlmann premiums within greatest accuracy credibility theory. One important aspect for both models is the choice of the distribution of losses, which depicts the pattern of the experience of a policyholder conditional upon his risk parameter value. Empirical evidence often suggests that the individual's claim data can sometimes be volatile and hard to predict. For this reason, we intend to adopt a more general framework that leads to better versatility and robustness of the estimator. Our choice is the phase-type family of distributions popularized by Neuts (1981).

The phase-type distribution is used to describe the time until absorption with a finite number of transient states and one absorbing state. It is known to have the ability to approximate any distribution with positive support, cf. Cox (1955), Bolch et al. (2001). Another motivation for using this distribution in statistical modeling is pointed out by Asmussen et al. (1996): “very often, problems which have an explicit solution assuming exponential distributions are algorithmically



tractable when one replaces the exponential distribution with a phase-type distribution". These features make phase-type distributions very popular in many areas: see Herbertsson (2011), Meester and Sander (2007), Fackrell (2009) for applications of phase-type distributions in health care, finance and transportation infrastructure. In recent years, there have been a large number of applications available concerning risk theory, where the claim sizes were frequently assumed to be phase-type distributed, cf. Asmussen (2000), Bladt (2005) for thorough reviews of relevant literature.

Despite its popularity in other areas, virtually no attention has been given to the phase-type distribution in the research of credibility theory. Therefore, it is the goal of this thesis to exploit this deficit and develop new actuarial tools. To be more specific, the main task of this research is to obtain explicit solutions for Bayesian and Bühlmann premiums under the assumption that experience data follows a phase-type distribution. On the other hand, we also want to investigate the exact credibility property that are well known for linear exponential family, cf. Jewell (1974a). We are interested to see how widely it may extend within the family of phase-type distributions.

This thesis does not address the questions of parameters estimations. Nevertheless, we are able to obtain the numerical values of both Bühlmann and Bayesian premiums when the parameters of losses distributions and prior distributions are appropriately presumed. Thus, a practical question of how our premium estimators accurately reflect the business of an insurance portfolio has

risen. In order to solve this question, Klugman et al. (2008) pointed out that when there is limited prior information available, we may need to use the data provided by the insures to estimate the prior parameters. This approach is also referred to empirical Bayes estimation. Therefore, parameters estimations is a more practical aspect to be investigated when there is real insurance data available.

This thesis is organized as follows. The mathematical formulations underlying credibility theory including Bayesian and Bühlmann premiums are carefully reviewed in Chapter 2.

In Chapter 3, we start by reviewing an alternative representation of the density function of phase-type distributions based on the uniformization technique. The conjugate prior for our phase-type sampling distribution is then established. The explicit solution of the Bühlmann premium estimator is then obtained. Simulation study will be performed under different model settings and comparisons will be conducted accordingly.

The Bayesian premium estimator is treated in Chapter 4. Using the same conjugate prior obtained previously, we derive the algebraic form of the estimator. Similar numeric experiment is going to be performed as in Chapter 3. Additionally, we will also compare the result with Bühlmann premiums algebraically as well as through a number of examples.

Conclusions and future work are presented in Chapter 5.

# Chapter 2

## Preliminaries

### 2.1 Credibility Theory

In the first chapter, we introduced the credibility theory as a set of quantitative tools used by insurers for performing experience ratings. Two branches of credibility theory, known as limited fluctuation credibility theory and greatest accuracy credibility theory have also been briefly discussed. In this section, we will provide more detailed discussions around mathematical assumptions and formulations of these two approaches.

#### 2.1.1 Limited Fluctuation Credibility Theory

Suppose that the history of a policyholder's experience consisting  $n$  losses which are denoted by  $x_1, \dots, x_n$ . They are viewed as realized values of i.i.d. random variables  $X_1, \dots, X_n$  with  $E(X_j) = \xi$  and  $Var(X_j) = \sigma^2$ . Then the average loss

amount of this policyholder is determined by

$$\bar{X} = (X_1 + \dots + X_n)/n.$$

It is not difficult for one to verify that  $E(\bar{X}) = \xi$  and  $Var(\bar{X}) = \sigma^2/n$ . Under the limited fluctuation credibility theory,  $\bar{X}$  could be the premium estimator for the next period if the policyholder's past experience is stable enough, which is known as the case of "full credibility". Otherwise if the experience appears to be more volatile, a manual rate  $m$  reflecting the risk experience of the entire risk class would be a more appropriate choice, which case is also referred to as "no credibility".

From a statistical point of view, the stability of the policyholder's past experience can be investigated by studying the distance between the sample average  $\bar{X}$  and the expected value  $\xi$ , i.e.  $|\bar{X} - \xi|$ , which is sometimes also known as the bias of the estimate. The past experience is then deemed stable if this bias is bounded above by a small fraction of  $\xi$  with relatively high probability. To be more specific, assuming two real numbers  $r > 0$  and  $0 < p < 1$  with  $r$  close to 0 and  $p$  close to 1. The past experience is stable if the following inequality holds:

$$Pr(|\bar{X} - \xi| \leq r\xi) \geq p.$$

The above inequality can also be expressed equivalently as

$$Pr\left(\left|\frac{\bar{X} - \xi}{\sigma/\sqrt{n}}\right| \leq \frac{r\xi\sqrt{n}}{\sigma}\right) \geq p. \quad (2.1)$$

Now define  $y_p$  by

$$y_p = \inf_{y \in \mathbb{R}} \left\{ Pr\left(\left|\frac{\bar{X} - \xi}{\sigma/\sqrt{n}}\right| \leq y\right) \geq p \right\}.$$

Then comparing with (2.1), a sufficient condition for full credibility can be found to be  $r\xi\sqrt{n}/\sigma \geq y_p$ . In other words, to ensure full credibility the number of exposure units needs to satisfy

$$n \geq \left(\frac{y_p\sigma}{r\xi}\right)^2. \quad (2.2)$$

The experience is considered zero credibility when the above inequality is not satisfied.

One concern with the above method is that it only considers “binary” results of either full credibility or zero credibility based on a simple criteria (2.2). To achieve a more balanced approach, an extension of this model was proposed by Whitney (1918) which is called partial credibility. In his work, the premium estimator was expressed as a weighted average of individual risk experience and class experience in the form of

$$P = z\bar{X} + (1 - z)m, \quad (2.3)$$

where  $z$  is called the credibility factor and it takes a value between 0 and 1. The larger  $z$  becomes, more consideration is given toward the individual experience and vice versa.

In practice, there are many formulae available for calculating the credibility factor. Popular choices as mentioned by Goulet (2008) include

$$z = \min \left\{ \frac{r\xi}{y_p\sigma} \sqrt{n}, 1 \right\},$$

or

$$z = \min \left\{ \left( \frac{nr\xi}{y_p\sigma} \right)^{2/3}, 1 \right\}.$$

Although limited fluctuation credibility is relatively simple to implement, it suffers a few drawbacks as commented by Klugman et al. (2008). First of all, it provides no systematic guidance for the selection of  $r$  and  $p$  thus their choices are somewhat arbitrary. Secondly, there is no underlying theoretical model for the distribution of losses, which makes it difficult to prove why a premium estimator in the format of (2.3) is preferable to the manual rate  $m$ . Finally, this approach does not examine the difference between  $\xi$  and  $m$  therefore the reliability of  $m$  as an estimator for the collective risk level is unclear. For more discussion of limited fluctuation credibility theory refer to Longley-Cook (1962), Norberg (2004) and Goulet (2008).

### 2.1.2 Greatest Accuracy Credibility Theory

Greatest accuracy credibility theory treats the experience rating problem following a model based approach. Certain distributional assumptions are implemented under this approach.

The approach starts with assuming a hypothetical risk parameter  $\theta_i$  for policyholder  $i$ ,  $i = 1, \dots, k$  where  $k$  is the total number of policyholders within a same risk class. The value  $\theta_i$  represents the unobservable risk characteristics of policyholder  $i$  and as Norberg (2004) stated, it could be viewed as a random selection from a portfolio of similar but not identical risks whose variation is described by some probability distribution. In other words, we assume a random variable  $\Theta$  with a prior distribution  $\pi(\theta)$  which describes the risk structure within a class. Then an individual risk parameter  $\theta_i$  can be viewed as a realized value of  $\Theta$ .

The second assumption is that individual experiences are random selections from some postulated distribution depending on his risk parameter. Suppose that  $n$  losses have been observed for individual  $i$  denoted by  $x_{i1}, \dots, x_{in}$ , they are treated as observations from i.i.d. random variables  $X_{i1}, \dots, X_{in}$  with a density function  $f(x_{ij}|\theta_i)$ , which is also known as the likelihood in Bayesian inference. Together with the prior distribution, the posterior distribution could then be calculated as

$$\pi(\theta|x_{i1}, \dots, x_{in}) \propto f(x_{i1}, \dots, x_{in}|\theta)\pi(\theta). \quad (2.4)$$

Ideally we would like to use the risk premium defined by  $\mu_{n+1}(\theta_i) = E(X_{i,(n+1)}|\theta_i)$

to assess the premium for policyholder  $i$  in the future. But since  $\theta_i$  is not observable this expectation cannot be calculated. Fortunately the posterior distribution (2.4) supplies us a valuable tool to infer the value of  $\theta_i$  given the experience data. Thus, instead of calculating the above expectation  $\mu_{n+1}(\theta_i)$  we can alternatively compute  $E(X_{i,(n+1)}|X_{i1}, \dots, X_{in})$  and use it as an estimate of the risk premium. This quantity is often referred to as the Bayesian premium.

As already mentioned in Chapter 1, Bayesian premiums can sometimes be very difficult to obtain. In this light, Bühlmann (1967, 1969) proposed the credibility premium by imposing certain linear structure in the estimator. The Bühlmann premium is in the form of  $\alpha_0 + \sum \alpha_j X_{ij}$  where the weights  $\alpha_0, \dots, \alpha_n$  are chosen such that the squared error loss with respect to  $\mu_{n+1}(\Theta)$  is minimized. Bühlmann has shown that this solution can also be expressed in a similar fashion as (2.3). We will discuss this issue in more details in Chapter 3.

By definition, Bühlmann credibility is the best linear estimator of the risk premium  $\mu_{n+1}(\theta_i)$ . Moreover, research has also shown that it is the best linear approximation of the Bayesian premium  $E(X_{i,(n+1)}|X_{i1}, \dots, X_{in})$  as well as  $X_{i,(n+1)}$ . See Herzog (1990) for a thorough comparison of Bayesian and Bühlmann models. Also refer to Hewitt(Jr.) (1970) for some historical notes. In some instances, both models can yield the exactly same estimator. This phenomenon is termed as exact credibility and was first examined by Jewell (1974a,b) where he showed that the exact credibility occurs when the likelihood is of exponential form and



the prior is conjugate <sup>1</sup>. This problem was also rigorously studied by Diaconis and Ylvisaker (1979) and they showed that the above condition was in fact sufficient and necessary. For a review of exact credibility see Goel (1982). Schmidt (1980) conducted another comparison focusing on the large sample property of the two estimators and it was shown that when  $n$  goes to infinity, both estimators converge to the same quantity which was termed as “individual premium” by the author.

Besides Bayesian and Bühlmann premiums, a number of other credibility models have also been introduced by various researchers. For example, the Bühlmann-Straub model of Bühlmann and Straub (1990) extended the Bühlmann model by considering  $X_{i1}, \dots, X_{in}$  as independent but not identically distributed random variables: experiences were still assumed to have the same conditional mean but different conditional variances. Other popular models include the random coefficients regression credibility model introduced by Hachemeister (1975), the hierarchical credibility model and crossed classification credibility model. For details and formulas see surveys by Goulet (1998), Makov et al. (1996), Bühlmann and Gisler (2005).

In this thesis, our focus will be restricted to Bayesian and Bühlmann premiums. One of the most important steps for implementing both models is the choice of the likelihood density  $f(x_{ij}|\theta_i)$  as in (2.4). This is where the phase-type

---

<sup>1</sup>The prior distribution  $\pi(\theta)$  is said to be conjugate for likelihood if the posterior distribution  $\pi(\theta|\mathbf{x})$  is in the same family as the prior distribution.

distribution comes to play.

## 2.2 Phase-type Distributions

Choosing a proper distribution for  $X_{ij}|\theta_i$  can be difficult: individual experiences may exhibit different types of behaviors which are difficult to be summarized by one particular distribution. For this reason, we decide to adopt a more versatile framework with the ability to model various types of structures. The phase-type distribution excels in this regard due to its capacity to approximating other distributions.

The earliest work regarding phase-type distributions can be tracked back to 1900s and was due to Erlang (1909). However, it was not until the late 70s that Neuts (1981) established the modern theory of phase-type distributions which has then been widely applied in different areas. See also Neuts (1989, 1995). A phase-type distribution can be either continuous or discrete. In this section, we review both forms as well as their important properties.

### 2.2.1 Continuous Phase-type Distributions

As Latouche and Ramaswami (1999) pointed out, the construction of phase-type distributions is based on the method of stages. The key idea is to model random time intervals as being made up a number of exponentially distributed segments and to exploit the resulting Markovian structure to simplify the analysis.

Consider a continuous-time Markov process  $\{J(x), x \geq 0\}$  with  $m+1$  states such that states  $1, \dots, m$  are transient and state 0 is an absorbing state. In addition, we suppose the process starts with an initial probability measure defined on the  $m+1$  states specified as  $(\alpha_0, \boldsymbol{\alpha})$ , where  $\alpha_0 = 1 - \boldsymbol{\alpha}'\mathbf{1}$  and  $\mathbf{1}$  is a  $m$ -dimensional column vector of ones. We further denote the infinitesimal generator<sup>1</sup>  $Q$  as

$$Q = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{t}_0 & \mathbf{T} \end{pmatrix},$$

where  $\mathbf{t}_0$  is an  $m \times 1$  column vector and  $\mathbf{T}$  is an  $m \times m$  matrix. Since  $Q$  is the generator of a Markov process we know that

$$T_{ii} < 0, T_{ij} \geq 0, \mathbf{t}_0 \geq 0, \text{ for } 1 \leq i \neq j \leq m$$

and

$$\mathbf{T}\mathbf{1} + \mathbf{t}_0 = \mathbf{0}.$$

Now we are in the position to formally define the phase-type distribution.

**Definition 2.1.** *The distribution of the time  $X$  till absorption into the absorbing state 0 is called the PH distribution with representation  $(\boldsymbol{\alpha}, \mathbf{T})$ , denoted by  $X \sim PH(\boldsymbol{\alpha}, \mathbf{T})$ .*

Based on this design one may derive the distribution function and density

---

<sup>1</sup>For a continuous-time Markov process it is also known as the transition rate matrix.

function of phase-type distribution by exploiting the Markov structure. The following result is given by Neuts (1981). However, his derivation was quite sketchy. To make the result more self-explanatory we have included a more detailed proof.

**Proposition 2.2.** *Assume that  $X \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{T})$ . The cumulative distribution function of  $X$  is given by*

$$F(x) = 1 - \boldsymbol{\alpha}' \exp \mathbf{T}x \mathbf{1}, \quad x \geq 0 \quad (2.5)$$

and its probability density function is given by

$$f(x) = \boldsymbol{\alpha}' \exp \mathbf{T}x \mathbf{t}_0, \quad x \geq 0, \quad (2.6)$$

where  $\mathbf{t}_0 = -\mathbf{T}\mathbf{1}$  and the matrix exponential for some matrix  $A$  is defined by

$$\exp(A) = \sum_{n=0}^{\infty} \frac{1}{n!} A^n.$$

*Proof.* See section 2.4.1. □

The moment structures of the continuous Phase-type distribution can then be studied and the results are outlined in the following corollary.

**Corollary 2.3.** *Suppose  $X \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{T})$  and  $t\mathbf{I} + \mathbf{T}$  is not singular. The moment*

generating function of  $X$  is given by

$$M(t) = E(e^{tX}) = -\boldsymbol{\alpha}'(t\mathbf{I} + \mathbf{T})^{-1}\mathbf{t}_0 + \alpha_0.$$

The  $k$ -th moments of  $X$  is

$$E(X^k) = k!\boldsymbol{\alpha}'(-\mathbf{T}^{-1})^k\mathbf{1}.$$

*Proof.* The moment generating function can be easily derived from the Laplace-Stieltjes transform of  $X$  obtained by Neuts (1981). The calculation of  $k^{\text{th}}$  moments then follows by  $E(X^k) = \frac{d^k}{dt^k}M(t)|_{t=0}$ .  $\square$

### 2.2.2 Discrete Phase-type Distributions

The discrete phase-type distribution is constructed in a similar fashion to the continuous case. Assuming a discrete-time Markov chain with initial probability measure  $(\alpha_0, \boldsymbol{\alpha})$  with transient states  $1, 2, \dots, m$  and an absorbing state 0. The transition probability matrix is given by

$$P = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{t}_0 & \mathbf{T} \end{pmatrix},$$

where  $\mathbf{t}_0 \geq 0$ ,  $T_{ij} \geq 0$  for  $1 \leq i, j \leq m$  and  $\mathbf{t}_0 + \mathbf{T}\mathbf{1} = \mathbf{1}$ . Suppose  $X$  is the absorption time into state 0 in this discrete Markov chain, then  $X$  is said to have

a discrete phase-type distribution denoted by  $PH_d(\boldsymbol{\alpha}, \mathbf{T})$ . Notice that it is the exactly same as Definition 2.1 except that the Markov context has been slightly altered.

The cumulative distribution function and probability mass function are given in the proposition below.

**Proposition 2.4.** *Assume that  $X$  is  $PH_d(\boldsymbol{\alpha}, \mathbf{T})$ . We have that*

$$P(X = 0) = \alpha_0, \quad (2.7)$$

$$P(X = k) = \boldsymbol{\alpha} \mathbf{T}^{k-1} \mathbf{t}_0, \quad k \geq 1, \quad (2.8)$$

$$P(X \leq k) = 1 - \boldsymbol{\alpha} \mathbf{T}^k \mathbf{t}_0, \quad k \geq 0. \quad (2.9)$$

*Proof.* The  $k^{th}$  step transition matrix can be obtained by block matrix multiplication:

$$P^k = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{1} - \mathbf{T}^k \mathbf{1} & \mathbf{T}^k \end{pmatrix}.$$

The probability mass function can then be derived based on conditional probability arguments, cf. Neuts (1981).  $\square$

Based on the above result, the probability generating function and factorial moments are obtained.

**Corollary 2.5.** *Suppose  $X \sim PH_d(\boldsymbol{\alpha}, \mathbf{T})$  and  $I - z\mathbf{T}$  is nonsingular. The prob-*

ability generating function of  $X$  is given by

$$G(z) = E(z^X) = \alpha_0 + z\alpha(I - zT)^{-1}\mathbf{t}_0, \text{ for } |z| \leq 1.$$

The factorial moment of  $X$  is

$$E[X(X-1)\cdots(X-k+1)] = k!\alpha(I - T)^{-k}T^{k-1}\mathbf{1}, \text{ for } k \geq 1.$$

*Proof.* One can easily calculate the probability generating function by  $\sum_{k=0}^{\infty} z^k P(X = k)$ . The factorial moments are followed by differentiating the p.g.f. successively, cf. Latouche and Ramaswami (1999).  $\square$

## 2.3 Proofs

### 2.3.1 Proof of Proposition 2.2

Consider a Markov process with the infinitesimal generator  $Q$  is in the state  $J(x)$  at time  $x$  for all  $x \geq 0$ . The transition function  $P(x)$  with elements  $P_{ij}(x) = [J(x) = j | J(0) = i]$  is given by  $P(x) = \exp(Qx)$ . We also know that by definition,

$$\exp(Qx) = \sum_{n=0}^{\infty} \frac{(Qx)^n}{n!}$$

It is easy to verify that the infinitesimal generator  $Q$  can be rewritten as

$$Q = \begin{pmatrix} 0 & \mathbf{0} \\ -\mathbb{T}\mathbf{1} & \mathbb{T} \end{pmatrix},$$

and its  $n$ th power is

$$Q^n = \begin{pmatrix} 0 & \mathbf{0} \\ -\mathbb{T}^n\mathbf{1} & \mathbb{T}^n \end{pmatrix}.$$

Thus,

$$\begin{aligned} \exp(Qx) &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{0} \\ -\mathbb{T}x\mathbf{1} & \mathbb{T}x \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{0} \\ \frac{-\mathbb{T}^2x^2\mathbf{1}}{2} & \frac{\mathbb{T}^2x^2}{2} \end{pmatrix} + \dots \\ &= \begin{pmatrix} 1 & \mathbf{0} \\ -(\mathbb{T}x + \frac{-\mathbb{T}^2x^2}{2} + \dots)\mathbf{1} & I + \mathbb{T}x + \frac{-\mathbb{T}^2x^2}{2} + \dots \end{pmatrix} \\ &= \begin{pmatrix} 1 & \mathbf{0} \\ (I - \exp(\mathbb{T}x))\mathbf{1} & \exp(\mathbb{T}x) \end{pmatrix} \\ &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{1} - \exp(\mathbb{T}x)\mathbf{1} & \exp(\mathbb{T}x) \end{pmatrix} \end{aligned}$$

Then we could obtain that

$$\begin{aligned} F(x) &= P[J(x) = 0] \\ &= \sum_{0 \leq i \leq m} P[J(0) = i]P[J(x) = 0 | J(0) = i] \end{aligned}$$



$$\begin{aligned} &= \alpha_0 + \sum_{0 \leq i \leq m} \alpha_i P_{i0}(x) \\ &= \alpha_0 + \boldsymbol{\alpha} \mathbf{1} - \boldsymbol{\alpha} \exp(\mathbb{T}x) \mathbf{1} \\ &= 1 - \boldsymbol{\alpha} \exp(\mathbb{T}x) \mathbf{1}. \end{aligned}$$

After differentiation, the density function is given by

$$f(x) = \boldsymbol{\alpha}' \exp(\mathbb{T}x) \mathbf{t}_0,$$

which completes the proof.

# Chapter 3

## Bühlmann Premium for

## Phase-type Distributed Losses

The focus of this chapter is the Bühlmann model. The assumption of phase-type distributed losses will be applied to obtain the corresponding Bühlmann estimator. The issue of conjugate prior will also be discussed.

### 3.1 The Bühlmann Premium

We will first start by outlaying the mathematical grounds for a general Bühlmann type estimator without any distribution assumptions. This section is heavily based on the original work due to Bühlmann (1967, 1969). In the interests of presenting a self-contained development, in what follows below we present several pages which follow closely from Klugman et al. (2008).

Assume that some policyholder has a history of losses  $x_i, i = 1, \dots, n$  and his hypothetical risk parameter is  $\theta$ . As already been discussed in the last chapter, we treat those losses as observations from i.i.d. random variables  $X_i$ , whose distribution is specified by  $f_{X|\Theta}(x|\theta)$ .

To estimate the future premium one natural choice would be the risk premium  $\mu_{n+1}(\theta) = E(X_{n+1}|\Theta = \theta)$ . Since this object is not directly workable, the Bühlmann premium was proposed to provide a linear approximation. The estimator is defined by

$$P_{Cr} = \alpha_0 + \sum_{j=1}^n \alpha_j X_j, \quad (3.1)$$

and the weights  $\alpha_0, \dots, \alpha_n$  are chosen so that the squared error loss with regard to  $\mu_{n+1}(\Theta)$

$$Q = E \left\{ \left[ \mu_{n+1}(\Theta) - \alpha_0 - \sum_{j=1}^n \alpha_j X_j \right]^2 \right\} \quad (3.2)$$

is minimized, where expectation is over the joint distribution of  $X_1, \dots, X_n$  and  $\Theta$ . We would like to make a remark here to emphasize the difference between the so-called “squared error loss” and the commonly known mean squared error. By the law of total expectation, we can rewrite (3.2) as

$$Q = E \left\{ E \left[ \mu_{n+1}(\Theta) - \alpha_0 - \sum_{j=1}^n \alpha_j X_j \middle| \Theta = \theta \right]^2 \right\}.$$

This equation shows that the calculation of  $Q$  actually involves two integrals.

The inner integral is in fact equivalent to

$$MSE(P_{Cr}) = E \left[ \mu_{n+1}(\theta) - \alpha_0 - \sum_{j=1}^n \alpha_j X_j \right]^2, \quad (3.3)$$

which is the mean squared error of the estimator  $P_{Cr}$  with regard to  $\mu_{n+1}(\theta)$  while fixing  $\theta$ . Then  $Q$  can be obtained by taking another expectation with respect to the distribution of  $\Theta$ .

To obtain solutions of  $\alpha_0, \dots, \alpha_n$  we first minimize  $Q$ . Denote  $\tilde{\alpha}_0, \dots, \tilde{\alpha}_n$  to be the appropriate values that minimize  $Q$ . Then we know the respective partial derivatives of  $Q$  evaluated at  $\tilde{\alpha}_0, \dots, \tilde{\alpha}_n$  should equal to zero. This gives us the following relations:

$$E(X_{n+1}) = \tilde{\alpha}_0 + \sum_{j=1}^n \tilde{\alpha}_j E(X_j), \quad (3.4)$$

$$E(X_i X_{n+1}) = \tilde{\alpha}_0 E(X_i) + \sum_{j=1}^n \tilde{\alpha}_j E(X_i X_j). \quad (3.5)$$

Equations (3.4) and (3.5) are often referred to as the “unbiased equation” and the “normal equation” respectively.

Based on equations (3.4) - (3.5)  $\tilde{\alpha}_0, \dots, \tilde{\alpha}_n$  can be solved. Introduce the following notations:

$$\mu(\theta) = E(X_i | \Theta = \theta), \quad v(\theta) = Var(X_i | \Theta = \theta).$$

Also define

$$\mu = E[\mu(\Theta)], \quad v = E[v(\Theta)], \quad a = \text{Var}[\mu(\Theta)].$$

It can be shown that the mean, variance, covariance and correlation coefficient of the experiences are

$$E(X_i) = E(\mu(\theta)) = \mu,$$

$$\text{Var}(X_i) = E(v(\theta)) + \text{Var}(\mu(\theta)) = v + a,$$

$$\text{Cov}(X_i, X_j) = \text{Var}(\mu(\theta)) = a,$$

$$\rho(X_i, X_j) = \frac{a}{v + a}.$$

Then the solutions of  $\tilde{\alpha}_0, \dots, \tilde{\alpha}_n$  are given by

$$\tilde{\alpha}_0 = \frac{(1 - \rho)\mu}{1 - \rho + n\rho},$$

$$\tilde{\alpha}_i = \frac{\rho}{1 - \rho + n\rho},$$

where  $\rho = \rho(X_i, X_j)$ .

Substituting  $\tilde{\alpha}_0, \dots, \tilde{\alpha}_n$  back into (3.1) we have

$$P_{Cr} = Z\bar{X} + (1 - Z)\mu, \tag{3.6}$$

where  $Z = n\rho/(1 - \rho + n\rho)$  and  $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ . By replacing  $\rho$  with  $a/(v + a)$

we can also calculate  $Z$  by

$$Z = \frac{n}{n + k}, \quad (3.7)$$

$$k = \frac{v}{a} = \frac{E[\text{Var}(X_j|\Theta)]}{\text{Var}[E(X_j|\Theta)]}. \quad (3.8)$$

Moving forward we will use (3.6) as the definition for Bühlmann premium.

We would like to point out the resemblance between equations (3.6) and (2.3).  $Z$  from (3.6) is also known as the credibility factor despite of the difference in their definitions. For a homogeneous portfolio, there is no need to charge different premiums to the insureds since  $v$  would be minimum leading to a small value of  $Z$  close to 0. Conversely, the more heterogeneous the portfolio, the greater the consideration of the individual experience, hence the higher the credibility factor.

On the other hand, when there is no prior information available, one can always refer to the approach of nonparametric estimation to find the unbiased estimation of the parameters  $\mu$ ,  $v$  and  $a$  involved in the Bühlmann premium. The materials presented below will follow Klugman et al. (2008) closely.

Suppose that, for policyholder  $i$ , we have the loss vector

$$\mathbf{X}_i = (X_{i1}, \dots, X_{in})^T, \quad i = 1, \dots, r.$$

Furthermore, conditional on  $\Theta_i = \theta_i$ ,  $X_{ij}$  has mean and variance such as

$$\mu(\theta_i) = E(X_{ij}|\Theta_i = \theta_i),$$

$$v(\theta_i) = Var(X_{ij}|\Theta_i = \theta_i).$$

and  $X_{i1}|\Theta_i = \theta_i, \dots, X_{in}|\Theta_i = \theta_i$  are independent. Therefore, the unbiased estimators of the Bühlmann quantities are given by

$$\begin{aligned} \hat{\mu} &= \bar{X} = (rn)^{-1} \sum_{i=1}^r \sum_{j=1}^n X_{ij}, \\ \hat{v} &= \frac{1}{r(n-1)} \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2, \\ \hat{a} &= \frac{1}{r-1} \sum_{i=1}^r r(\bar{X}_i - \bar{X})^2 - \frac{1}{rn(n-1)} \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2. \end{aligned}$$

Klugman et al. (2008) also stated that if  $f_{X_j|\Theta}(x_j|\theta)$  is assumed to be of parametric form but not  $\pi(\theta)$ , then we refer to the problem as being of a semiparametric nature. In another case, the fully parametric approach can be investigated when both  $f_{X_j|\Theta}(x_j|\theta)$  and  $\pi(\theta)$  are assumed to be of parametric forms.

## 3.2 Expressing Phase-type Distributions as Infinite Mixtures of Erlang Distributions

Phase-type distributions entail rich mathematical structure which sometimes leads to considerable amount of complexity in modeling and computation. Luckily,

research has shown that an alternative representation is available for phase-type distributions, which is in the form of an infinite mixture of Erlang distribution<sup>1</sup>. This representation was obtained using the technique of uniformization, which was first discussed by Jensen (1953). Johnson and Taaffe (1988) showed the denseness of both phase-type distributions and this infinite mixtures representation. For related discussions see Shanthikumar (1985), Stanford (2011). The major advantage of using this representation is to reduce the complexity associated with deriving Bayesian and Bühlmann estimators later.

In this section we review the formulation of this representation. Suppose that  $X|\theta \sim PH(\boldsymbol{\alpha}, T)$  where matrix  $T$  depends on  $\theta$ . Without loss of generality, we can presume that  $T$  can be written in the form of

$$T = \begin{pmatrix} -\theta_1 & \theta_1 p_{12} & \theta_1 p_{13} & \dots & \theta_1 p_{1m} \\ \theta_2 p_{21} & -\theta_2 & \theta_2 p_{23} & \dots & \theta_2 p_{2m} \\ \theta_3 p_{31} & \theta_3 p_{32} & -\theta_3 & \dots & \theta_3 p_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_m p_{m1} & \theta_m p_{m2} & \theta_m p_{m3} & \dots & -\theta_m \end{pmatrix}, \quad (3.9)$$

where  $\theta_i, i = 1, \dots, m$  represents the rate of leaving state  $i$  and  $p_{ij}, 1 \leq i \neq j \leq m$  represents the rate of leaving state  $i$  then immediately entering state  $j$ . To satisfy

---

<sup>1</sup>Essentially an Erlang distribution is a special case of Gamma distribution with the shape parameter being an integer.



the Markov process context we assume that

$$p_{ij} \geq 0, \quad \theta_i \geq 0 \quad (1 \leq i \neq j \leq m), \quad \sum_{j \neq i} p_{ij} < 1 \text{ for some } 1 \leq i \leq m. \quad (3.10)$$

Setting  $\theta = \max_i \{\theta_i\}$ , we then define

$$P = I + (1/\theta)T, \quad (3.11)$$

where the matrix  $I$  is an identity matrix with a proper dimension. Therefore we have:

$$P = \begin{pmatrix} 1 - \theta_1/\theta & \theta_1 p_{12}/\theta & \theta_1 p_{13}/\theta & \dots & \theta_1 p_{1m}/\theta \\ \theta_2 p_{21}/\theta & 1 - \theta_2/\theta & \theta_2 p_{23}/\theta & \dots & \theta_2 p_{2m}/\theta \\ \theta_3 p_{31}/\theta & \theta_3 p_{32}/\theta & 1 - \theta_3/\theta & \dots & \theta_3 p_{3m}/\theta \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_m p_{m1}/\theta & \theta_m p_{m2}/\theta & \theta_m p_{m3}/\theta & \dots & 1 - \theta_m/\theta \end{pmatrix}. \quad (3.12)$$

It is not difficult to see that  $P$  is also sub-stochastic under condition (3.10).

By using the property  $e^{X+Y} = e^X e^Y$  for commutative matrices  $X$  and  $Y$ , we can calculate  $\exp(Tx)$  as:

$$\begin{aligned} \exp(Tx) &= \exp(\theta(P - I)x) \\ &= \exp(\theta x P) \exp(-\theta x I) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{k=0}^{\infty} \frac{(\theta x)^k}{k!} P^k \sum_{l=0}^{\infty} \frac{(-\theta x)^l}{l!} I^l \\
 &= \sum_{k=0}^{\infty} \frac{(\theta x)^k}{k!} P^k (e^{-\theta x} I) \\
 &= \sum_{k=0}^{\infty} \frac{e^{-\theta x} (\theta x)^k}{k!} \cdot P^k.
 \end{aligned} \tag{3.13}$$

The following two results are presented by Hassan-Zadeh and Stanford (2013) which give the exact representation.

**Proposition 3.1.** *The density function of  $X|\theta \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{T})$  can be rewritten as an infinite mixture of Erlang densities as follows:*

$$f(x|\Theta = \theta) = \sum_{n=0}^{\infty} q_{n+1} \theta e^{-\theta x} (\theta x)^n / n! \tag{3.14}$$

where  $\theta = \max_i \{\theta_i\}$  with  $\theta_i$  specified in (3.9) and

$$q_{n+1} = \boldsymbol{\alpha}' P^n (I - P) \mathbf{1} \geq 0; \quad n = 0, 1, \dots, \tag{3.15}$$

with  $P$  specified by (3.12). The coefficients  $q_{n+1}$  satisfy  $\sum_{n=0}^{\infty} q_{n+1} = 1$ .

*Proof.* See section 3.6.1. □

Based on the above result, the joint probability density function for a collection of phase-type distributed losses can be easily derived which is shown below.

**Corollary 3.2.** *Assume that given  $\Theta = \theta$ ,  $X_1, \dots, X_n$  are random samples from*

(3.26). Then their joint probability density is

$$f(x_1, \dots, x_n | \Theta = \theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{l=0}^{\infty} \frac{(\theta \sum_{i=1}^n x_i)^l}{l!} C(l, x_1, \dots, x_n), \quad (3.16)$$

where

$$C(l, x_1, \dots, x_n) = \sum_{i=0}^l \binom{l}{i} q_{i+1} \left( \frac{x_n}{\sum_{j=1}^n x_j} \right)^i \left( 1 - \frac{x_n}{\sum_{j=1}^n x_j} \right)^{l-i} C(l-i, x_1, \dots, x_{n-1}) \quad (3.17)$$

for all  $l \in \mathbb{Z}^+ \cup \{0\}$  and  $n \in \mathbb{N}$ . The above iteration is initiated by

$$C(l, x_1) = q_{l+1}, \quad l = 0, 1, 2, \dots$$

*Proof.* See section 3.6.2. □

This corollary identifies the likelihood function we are going to use for the construction of Bayesian and Bühlmann premiums. It will help us to derive the conjugate prior in the next chapter.

### 3.3 Conjugate Prior for Phase-type Distributions

The first step to construct premium estimators is to specify a prior distribution for the risk parameter  $\Theta$ . While the choice of prior distributions are generally subjective, we intend to use the natural conjugate prior for our phase-type likelihood.

The following result from Bühlmann and Gisler (2005) is helpful for constructing a class  $\mathcal{U}$  of conjugate priors to a distribution family  $\mathcal{F}$ .

**Proposition 3.3.** *Assumptions:  $\mathcal{F}$  and  $\mathcal{U}$  satisfy the following conditions:*

- *The likelihood functions  $l(\theta) = f_\theta(x)$ ,  $x$  fixed, are proportional to an element of  $\mathcal{U}$ , i.e. for every possible observation  $x \in A$ , there exists a  $u_x \in \mathcal{U}$ , such that  $u_x(\theta) = f_\theta(x)(\int f_\theta(x)d\theta)^{-1}$ .*
- *$\mathcal{U}$  is closed under the product operation, i.e. for every pair  $u, v \in \mathcal{U}$  we have that  $u(\cdot)v(\cdot)(\int u(\theta)v(\theta)d\theta)^{-1} \in \mathcal{U}$ .*

*Then it holds that  $\mathcal{U}$  is conjugate to  $\mathcal{F}$ .*

*Proof.* See Theorem 2.21 of Bühlmann and Gisler (2005). □

We are now ready to state our first important result of this chapter.

**Theorem 3.4.** *Suppose  $X_1|\theta, \dots, X_n|\theta$  are i.i.d. distributed random variables with  $PH(\boldsymbol{\alpha}, \mathbb{T})$ . Then a prior which is conjugate for the joint likelihood  $f(x_1, \dots, x_n|\theta)$  can be written as an infinite mixture of Erlang distributions with the probability density function specified as*

$$\pi(\theta) = \sum_{l=0}^{\infty} \zeta_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!}, \quad (3.18)$$

where  $\beta > 0$  and  $m$  is a non-negative integer.  $\zeta_l$  is a probability measure, i.e.  $\sum_{l=0}^{\infty} \zeta_l = 1$  with the possibility  $\zeta_l = 0$  for some  $l$ . The parameters  $\beta$ ,  $m$  and  $\{\zeta_l\}$  are considered as the “updated parameters”.

*Proof.* See section 3.6.3. □

Theorem 3.4 shows the class of all distributions that is trivially conjugate for the joint likelihood specified above. Observing equation (3.32), we notice that the class of conjugate prior distributions is defined as arbitrary countable mixtures of Erlang distributions with the same scale parameter. It could be an infinite summation or a finite one by setting for example  $\zeta_i = 0$  for all  $i > k$  with  $k$  being fixed. We also want to point out that an alternate expression for (3.32) can be obtained by viewing  $l + m$  as a whole entity and adjusting relevant parameters:

$$u(\theta) = \sum_{l=0}^{\infty} \omega_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^l}{l!},$$

where  $\omega_0 = \omega_1 = \dots = \omega_{m-1} = 0$ ,  $\omega_m = \zeta_0$ ,  $\omega_{m+j} = \zeta_j$  for  $j = 1, 2, 3, \dots$

Note that according to Hassan-Zadeh and Stanford (2013), exact credibility does occur for the phase-type case when parameters are set in a particular way which reduces it to a Gamma distribution, with the corresponding prior being another single Gamma distribution guaranteed by setting  $\{\zeta_l\}$  being the negative Binomial coefficients.

## 3.4 Bühlmann Premium for Phase-type Distributed Losses

As Latouche and Ramaswami (1999) pointed out, the uniformization technique which leads to the alternative expression of phase-type distributions allows one to interpret a continuous time Markov process as a discrete time Markov chain, for which one merely replaces the constant unit of time between any two transitions by independent exponential random variables with the same parameter. In this regard, suppose a phase-type distributed random variable  $X|\Theta$  describing the time to absorption of certain Markov process. It can be viewed alternatively as comprising a succession of i.i.d. exponential intervals at rate  $\Theta$ . That is,

$$X = Y_1 + Y_2 + \dots + Y_N$$

where  $Y_i \sim \exp(\Theta)$ ,  $i = 1, 2, \dots$ , with  $N$  representing the number of transitions required for the process to become absorbed. In this case  $N$  follows a discrete phase-type distribution, i.e.

$$N \sim PH_d(\boldsymbol{\alpha}, P), \tag{3.19}$$

where  $P$  is defined by (3.12).

We are now in the position to state our second result of this chapter.

**Theorem 3.5.** *Suppose  $X_1|\Theta, \dots, X_n|\Theta$  are i.i.d. distributed random variables with  $PH(\alpha, T)$ . If  $\Theta$  follows a distribution specified by (3.32), the Bühlmann premium estimator based on losses  $X_1, \dots, X_n$  is*

$$P_{Cr} = \frac{n}{n+k} \bar{X} + \frac{k}{n+k} \mu, \quad (3.20)$$

where  $N$  is defined by (3.19),  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  and

$$\begin{aligned} \mu &= E(N) \beta \sum_{l=0}^{\infty} \frac{\zeta_l}{l+m}, \\ k &= \frac{\text{Var}(N) + E(N)}{(E(N))^2} \cdot g(m, \zeta) \end{aligned} \quad (3.21)$$

with

$$g(m, \zeta) = \frac{\sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)}}{\sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)} - \left( \sum_{l=0}^{\infty} \frac{\zeta_l}{l+m} \right)^2}.$$

Here parameters  $m > 1$ ,  $m \in \mathbb{Z}$ ,  $\beta$  and  $\{\zeta_l\}$  follows the same definitions from Theorem 3.4.

*Proof.* See section 3. □

The above result gives us solution of the Bühlmann premium under phase-type distributed losses. Notice that the solution is not in a sense of “closed form” since the definitions of  $\mu$  and  $k$  involve infinite series. Those series are obviously convergent since their general terms are all bounded above by  $\zeta_l$ . However, since the choice of the measure  $\{\zeta_l\}$  is relatively flexible, the infinite series can yield a

number of different limits.

### 3.5 Examples

**Example 3.5.1.** We consider a simple case here when the density function of  $X|\Theta$  just follows some single Erlang distribution such that  $f(x|\theta) = \theta e^{-\theta x} \frac{(\theta x)^{k-1}}{(k-1)!}$ ,  $q_{n+1} = 1$  when  $n = k - 1$  and  $q_{n+1} = 0$  otherwise. Then the initial step for the recursive  $C$  function is

$$C(k-1, x_1) = q_k = 1 \text{ and } C(m, x_1) = 0 \text{ for } m \neq k-1$$

Recall that  $C(M, x_1, \dots, x_n) = \sum_{i=0}^M \binom{M}{i} q_{i+1} \left(\frac{x_n}{\sum_{j=1}^n x_j}\right)^i \left(1 - \frac{x_n}{\sum_{j=1}^n x_j}\right)^{M-i} C(M-i, x_1, \dots, x_{n-1})$  and so the formula for the case of two observations is

$$C(M, x_1, x_2) = \sum_{i=0}^M \binom{M}{i} q_{i+1} \left(\frac{x_2}{x_1 + x_2}\right)^i \left(\frac{x_1}{x_1 + x_2}\right)^{M-i} C(M-i, x_1),$$

and it would only have a valid result only if both  $q_{i+1}$  and  $C(M-i, x_1)$  equal to 1. We have already known that  $q_{i+1} = 1$  when  $i = k-1$  and  $C(M-i, x_1)$  when  $M-i = k-1$ . By combining those two pieces of information we realize that it will happen when  $i = k-1$  and  $M = 2k-2$ . So we have

$$C(2k-2, x_1, x_2) = \binom{2k-2}{k-1} \left(\frac{x_2}{x_1 + x_2}\right)^{k-1} \left(\frac{x_1}{x_1 + x_2}\right)^{k-1}$$



$$C(M, x_1, x_2) = 0 \text{ when } M \neq 2k - 2.$$

Then the recursive  $C$  function for three observations is

$$C(3k - 3, x_1, x_2, x_3) = \frac{(3k - 3)!}{3 \cdot (k - 1)!} (x_1 \cdot x_2 \cdot x_3)^{k-1} \left( \frac{1}{x_1 + x_2 + x_3} \right)^{3k-3}$$

$$C(M, x_1, x_2, x_3) = 0 \text{ when } M \neq 3k - 3.$$

Therefore, in general, the recursive  $C$  function for  $n$  observations is

$$C(nk - n, x_1, \dots, x_n) = \frac{(nk - n)!}{n \cdot (k - 1)!} \left( \prod_{i=1}^n x_i \right)^{k-1} \left( \frac{1}{\sum_{i=1}^n x_i} \right)^{nk-n}$$

$$C(M, x_1, \dots, x_n) = 0 \text{ when } M \neq nk - n.$$

Recalled from the previous section that the conjugate prior is

$$u(\theta) = \sum_{l=0}^{\infty} \left\{ \frac{C(l, x_1, \dots, x_n) \frac{(l+n)!}{l!}}{\sum_{k=0}^{\infty} C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!}} \right\} \cdot \left( \sum_{i=1}^n x_i \right) e^{-\theta \sum_{i=0}^n x_i} \frac{(\theta \sum_{i=1}^n x_i)^{l+n}}{(l+n)!}$$

By substituting the  $C$  function inside we have

$$u(\theta) = \frac{C(nk - n, x_1, \dots, x_n) \frac{(nk)!}{(nk-n)!}}{C(nk - n, x_1, \dots, x_n) \frac{(nk)!}{(nk-n)!}} \cdot \left( \sum_{i=1}^n x_i \right) e^{-\theta \sum_{i=0}^n x_i} \frac{(\theta \sum_{i=1}^n x_i)^{nk}}{(nk)!}$$

$$= \left( \sum_{i=1}^n x_i \right) e^{-\theta \sum_{i=0}^n x_i} \frac{(\theta \sum_{i=1}^n x_i)^{nk}}{(nk)!},$$

which follows a gamma distribution. This example actually verifies the general result in our case, which is the conjugate prior for an Erlang distribution is a Gamma distribution.

**Example 3.5.2.** In section 3.3 we have obtained the explicit solution of the Bühlmann premium estimator given phase-type distributed loss and the corresponding conjugate prior. Now we would like to design a few simulation experiments to investigate some of its properties.

Recall from Theorem 3.4, the conjugate prior for phase-type distributions is in the form of

$$\pi(\theta) = \sum_{l=0}^{\infty} \zeta_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!},$$

where  $\zeta_l$  represents some probability measure. In this experiment, we apply the geometric distribution due to its ability to take countably infinite many values, i.e.

$$\zeta_l = (1-p)^l p, \quad l = 0, 1, 2, 3, \dots$$

with  $p = 0.3$ . Note that another advantage for using this particular measure is that it leads to a closed form solution of the  $g(m, \zeta)$  in Theorem 3.5. We also assume that  $\beta = 20$  and  $m = 10$ .

Under these assumptions, the prior pdf can be explicitly expressed in a closed-

form formula such as

$$\begin{aligned}
\pi(\theta) &= \sum_{l=0}^{\infty} \zeta_l \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!} \\
&= \sum_{l=0}^{\infty} (1-p)^l p \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!} \\
&= \sum_{l=0}^{\infty} \frac{p}{(1-p)^m} \beta e^{-\beta\theta} \frac{(\beta(1-p)\theta)^{l+m}}{(l+m)!} \\
&= \frac{p}{(1-p)^m} \beta e^{-\beta\theta} \left( e^{\beta(1-p)\theta} - \sum_{k=0}^m \frac{(\beta(1-p)\theta)^k}{k!} \right).
\end{aligned}$$

On the other hand, we assume the losses satisfy that

$$X_i | \theta \stackrel{i.i.d.}{\sim} PH(\boldsymbol{\alpha}, \theta(P - I)),$$

with the values of parameters being

$$\boldsymbol{\alpha} = (1, 0)$$

and

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

This setting provides a relatively simple phase-type structure to conveniently study relevant properties.

Given the above parameters, the particular formulation for Bühlmann pre-

mium can be obtained based on Theorem 3.5. By equation (3.35) we have

$$\begin{aligned}
\sum_{l=0}^{\infty} \frac{\zeta_l}{l+m} &= \sum_{l=0}^{\infty} \frac{(1-p)^l p}{l+m} \\
&= \frac{p}{(1-p)^m} \sum_{l=0}^{\infty} \frac{(1-p)^{l+m}}{l+m} \\
&= \frac{p}{(1-p)^m} (-\log(1 - (1-p)) - \sum_{l=1}^{m-1} \frac{(1-p)^l}{l}) \\
&= \frac{p}{(1-p)^m} (-\log(p) - \sum_{l=1}^{m-1} \frac{(1-p)^l}{l}),
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)} \\
&= \sum_{l=0}^{\infty} \frac{(1-p)^l p}{(l+m)(l+m-1)} \\
&= p \sum_{l=0}^{\infty} (1-p)^l \left( \frac{1}{l+m-1} - \frac{1}{l+m} \right) \\
&= p \sum_{l=0}^{\infty} \frac{(1-p)^l}{l+m-1} - p \sum_{l=0}^{\infty} \frac{(1-p)^l}{l+m} \\
&= \frac{p}{(1-p)^{m-1}} \sum_{l=0}^{\infty} \frac{(1-p)^{l+m-1}}{l+m-1} - \frac{p}{(1-p)^m} (-\log(p) - \sum_{l=1}^{m-1} \frac{(1-p)^l}{l}) \\
&= \frac{p}{(1-p)^{m-1}} (-\log(p) - \sum_{l=1}^{m-2} \frac{(1-p)^l}{l}) - \frac{p}{(1-p)^m} (-\log(p) - \sum_{l=1}^{m-1} \frac{(1-p)^l}{l}) \\
&= \frac{p^2}{(1-p)^m} \log(p) - \frac{p}{(1-p)^{m-1}} \sum_{l=1}^{m-2} \frac{(1-p)^l}{l} + \frac{p}{(1-p)^m} \sum_{l=1}^{m-1} \frac{(1-p)^l}{l}.
\end{aligned}$$

Notice that several equalities above involve the application of the Taylor's expan-

tion  $\log(1 - x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}$ . We also know that

$$E(N) = \boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1}$$

$$\text{Var}(N) + E(N) = 2\boldsymbol{\alpha}'(I - P)^{-2}\mathbf{1} - (\boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1})^2$$

Substituting corresponding parameter values into the above results, we obtain that

$$E(N) = 2.5, \text{Var}(N) + E(N) = 5.25, k = 7.215939, \mu = 4.215061. \quad (3.22)$$

By its definition, we know that the Bühlmann estimator should yield the least squared error loss amongst all linear estimators. In the following experiment, we conduct a comparison between the Bühlmann premium estimator and the straight average  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . For the Bühlmann premium  $P_{Cr}$  we know that

$$\begin{aligned} \text{Bias} &= E(Z\bar{X} + (1 - Z)\mu - \mu(\theta)) \\ &= (Z - 1)E(N)\theta^{-1} + (1 - Z)\mu, \end{aligned} \quad (3.23)$$

$$\begin{aligned} \text{Variance} &= \text{Var}(Z\bar{X} + (1 - Z)\mu - \mu(\theta)) \\ &= Z^2\text{Var}(\bar{X}) \\ &= \frac{Z^2}{n}(\text{Var}(N) + E(N))\theta^{-2}, \end{aligned}$$

$$\text{MSE}_\theta(P_{Cr}) = \text{bias}^2 + \text{variance}, \quad (3.24)$$

where  $Z = n/(n + k)$ .

Similarly, for  $\bar{X}$  we have

$$\begin{aligned}
 \text{Bias} &= E(\bar{X} - \mu(\theta)) = 0, \\
 \text{Variance} &= \text{Var}(\bar{X} - \mu(\theta)) \\
 &= \text{Var}(\bar{X}) \\
 &= \frac{(\text{Var}(N) + E(N))\theta^{-2}}{n}, \\
 \text{MSE}_\theta(\bar{X}) &= \frac{(\text{Var}(N) + E(N))\theta^{-2}}{n}. \tag{3.25}
 \end{aligned}$$

Before comparing both estimators we examine some important properties of our prior distribution. In fact, we can easily calculate empirical quantiles along with other descriptive statistics, which are shown below

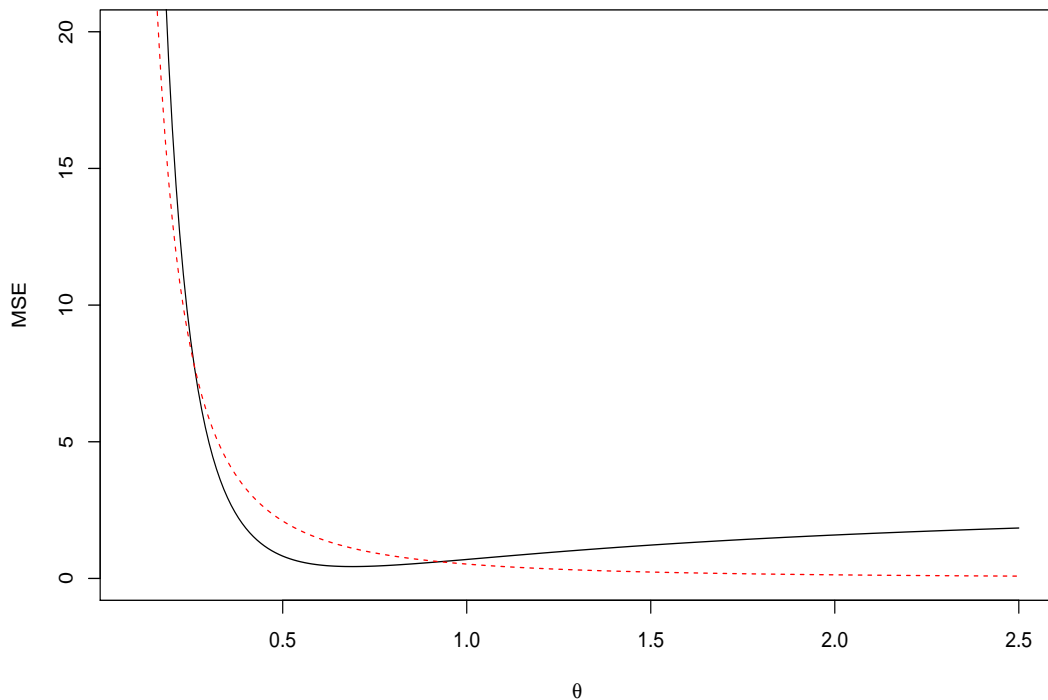
**Table 3.1:** Descriptive statistics for prior distribution

Min	2.5% Q	25% Q	Median	Mean	75% Q	97.5% Q	Max
0.1235	0.3133	0.5020	0.6322	0.6652	0.7940	1.1987	2.2370

We see that the area between 0.3133 and 1.1987 accounts for 95% of the probability mass.

To compare the MSEs of the Bühlmann estimator and the average  $\bar{X}$ , we conduct two sets of experiments. Firstly, we select a number of different  $\theta$ s following a scheme that ranging from 0.1 to 2.5 with an increment of 0.01. This gives rise to 241 equally spaced values. We set the total number of losses  $n = 10$ . The

MSEs can be evaluated based on (3.24) - (3.25). The following graph shows the MSEs at different values of  $\theta$ .



**Figure 3.1:** MSE comparison Bühlmann (solid line) vs.  $\bar{X}$  (dashed line)

We see that the MSE for Bühlmann estimator does not always yield a smaller MSE. Roughly speaking, when  $\theta < 0.3$  or  $\theta > 0.9$  the average estimator  $\bar{X}$  actually outperforms Bühlmann in the sense of MSE. If we calculate the respective averages over those 241 MSEs, Bühlmann estimator yields an average MSE of 3.448053 while  $\bar{X}$  yields an average MSE of 2.204005.

The above observation may seem to be contradictive with the known property of Bühlmann estimator. The true reason that  $\bar{X}$  has a smaller average MSE is

*because the averaging process fails to take account of the distributional information implied by the prior distribution. This average is taken based on equally spaced values of  $\theta$  which were manually set hence does not reflect on the expectation of equation (3.2).*

*In order to correctly estimate the squared error losses, we now randomly sample 200  $\theta$ s from the prior distribution and evaluate the MSEs accordingly. The averages of the MSEs are now found to be 1.1570003 for the Bühlmann estimator and 1.757805 for  $\bar{X}$ , which is consistent with the theoretical result. This is because our prior distribution, according to Figure ??, has most of probability mass between 0.3 and 0.9, which is exactly the area that Bühlmann estimator yields lower MSEs.*

*We can repeat the above process from different sample sizes for  $\theta$ , denoted by  $n$ . Below shows the squared error losses estimated based on  $n = 100$  until 20000 different  $\theta$ s.*

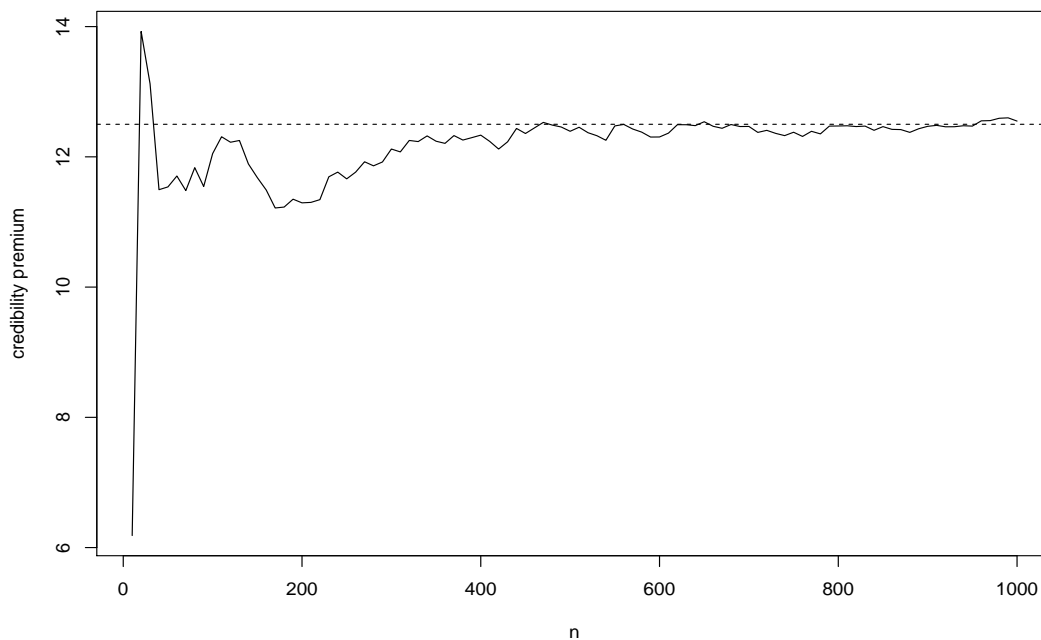


**Table 3.2:** Squared error losses (S.E.L): Bühlmann vs.  $\bar{X}$ 

$n$	S.E.L of $P_{Cr}$	S.E.L of $\bar{X}$	difference
100	1.0479763	1.706382	-0.6584057
200	1.1570003	1.757805	-0.6008050
300	1.0190901	1.708748	-0.6896583
500	0.9595561	1.680415	-0.7208591
1000	0.9983797	1.709539	-0.7111594
2000	1.0073427	1.740795	-0.7334527
3000	0.9655605	1.663279	-0.6977186
5000	1.0032764	1.715398	-0.7121216
10000	0.9872958	1.685964	-0.6986678
20000	0.9708508	1.677591	-0.7067405

From the above table, we see that even at the level  $n = 100$  the Bühlmann estimator shows less squared error loss. However, the difference between two S.E.L. are not quite stable at the start. For instance, there are noticeable fluctuations in the last column for the first few entries. When  $n$  reaches 3000, the difference is more steady and slightly fluctuates around  $-0.7$ .

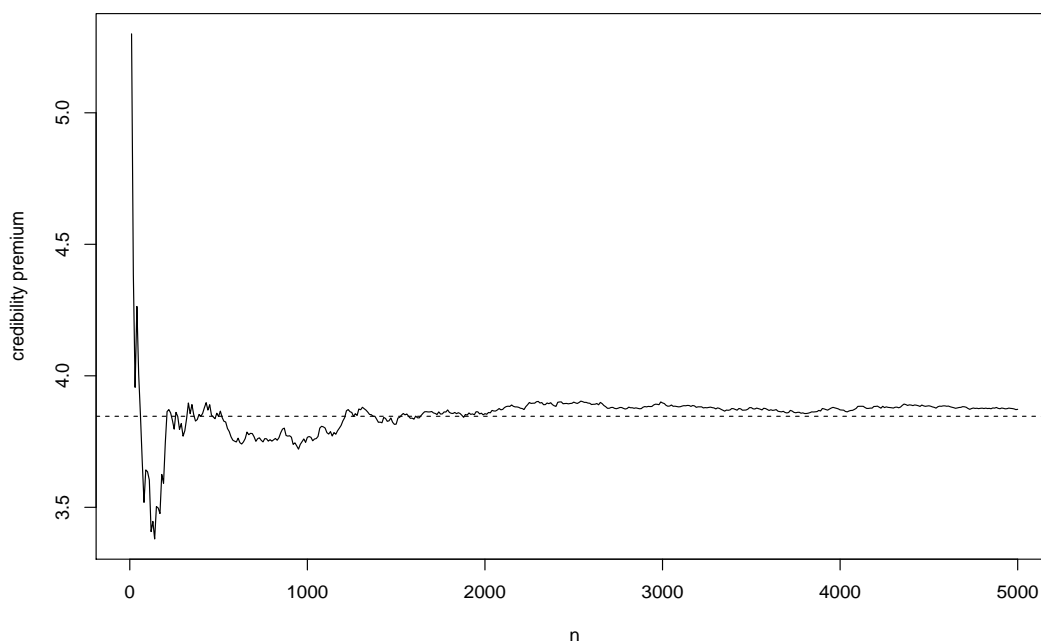
Lastly, we examine the Bühlmann estimator itself. According to (3.23), the bias of the Bühlmann estimator converges to 0 as the total number of losses goes to infinity. To investigate this property, we produce Bühlmann estimates based on different number of random samples drawn from  $X|\theta$  with  $\theta$  fixed. Below shows the Bühlmann premium estimates against the corresponding sample size with  $\theta = 0.2$ . For reference, a horizontal line  $y = \mu(\theta)$  is also plotted.



**Figure 3.2:** Bühlmann premium (solid line) vs.  $\mu(\theta)$  (dashed line) with  $\theta = 0.2$

From Figure 3.2, we can see that when  $\theta = 0.2$  the Bühlmann estimates converge fairly quickly to the hypothetical true value  $\mu(\theta)$ . The estimates become much more stable after  $n = 400$ . Only minor fluctuations present after the level  $n = 800$ .

One may also want to study the impact of the values of  $\theta$ . Figure 3.3 below shows the trending for premium estimates when  $\theta = 0.65$ .



**Figure 3.3:** Bühlmann premium (solid line) vs.  $\mu(\theta)$  (dashed line) with  $\theta = 0.65$

*Comparing Figure 3.3 with Figure 3.2, we notice the similarity between their overall convergence trends. However, the rate of convergence seems to be slower this time. The estimates are still notably volatile even at  $n = 1000$ . The fluctuation then weakens after  $n = 1500$ . This suggests that different choices of  $\theta$  may have an impact on the speed of convergence of the estimator.*

## 3.6 Proofs

### 3.6.1 Proof of Proposition 3.1

**Proposition 3.1** The density function of  $X|\theta \sim PH(\boldsymbol{\alpha}, T)$  can be rewritten as an infinite mixture of Erlang densities as follows:

$$f(x|\Theta = \theta) = \sum_{n=0}^{\infty} q_{n+1} \theta e^{-\theta x} (\theta x)^n / n! \quad (3.26)$$

where  $\theta = \max_i \{\theta_i\}$  with  $\theta_i$  specified in (3.9) and

$$q_{n+1} = \boldsymbol{\alpha}' P^n (I - P) \mathbf{1} \geq 0; \quad n = 0, 1, \dots, \quad (3.27)$$

with  $P$  specified by (3.12). The coefficients  $q_{n+1}$  satisfy  $\sum_{n=0}^{\infty} q_{n+1} = 1$ .

*Proof.* We have

$$\begin{aligned} \mathbf{t}_0 &= T \mathbf{1} \\ &= \theta (I - P) \mathbf{1}. \end{aligned} \quad (3.28)$$

From Proposition 2.2, after substitution of (3.13) and (3.28), the density function of the phase-type distribution can be written as

$$f(x|\Theta) = \boldsymbol{\alpha}' \exp(Tx) \mathbf{t}_0$$

$$\begin{aligned}
&= \alpha' \sum_{i=0}^{\infty} e^{-\theta x} (\theta x)^i / i! P^i \theta (I - P) \mathbf{1} \\
&= \sum_{i=0}^{\infty} \alpha' P^i (I - P) \mathbf{1} \theta e^{-\theta x} (\theta x)^i / i! \\
&= \sum_{i=0}^{\infty} q_{i+1} \theta e^{-\theta x} (\theta x)^i / i!,
\end{aligned}$$

where  $q_{i+1}$  is defined as

$$q_{i+1} = \alpha' P^i (I - P) \mathbf{1}, \quad i = 0, 1, 2, \dots$$

The following establish that  $q_{i+1}$  sum to unity.

$$\begin{aligned}
\sum_{i=0}^{\infty} q_{i+1} &= \sum_{i=0}^{\infty} \alpha' P^i (I - P) \mathbf{1} \\
&= \alpha' \sum_{i=0}^{\infty} P^i (I - P) \mathbf{1} \\
&= \alpha' (I - P)^{-1} (I - P) \mathbf{1} \\
&= \alpha' \mathbf{1} \\
&= 1.
\end{aligned}$$

Specifically,  $q_{i+1}$  represents the probability that absorption will occur at the  $(i + 1)$ th transition of the uniformized Markov chain. The following section establishes that  $q_{i+1}$  is non-negative for all possible. Define

$$f_0 = (I - P) \mathbf{1} \text{ and } f_i = P \cdot f_{i-1} \text{ for } i = 1, 2, \dots \quad (3.29)$$

The  $i$ th component of  $f_0$  is  $\frac{\theta_i}{\theta}(1 - \sum_{i \neq j} p_{ij})$  which is non-negative for all  $i$ 's. Thus,  $q_1 = \alpha' \cdot f_0$  is also non-negative. The non-negativity of  $P$  together with the form of the recursion in 3.29 ensures that  $f_i \geq 0$  for  $i = 1, 2, \dots$

By investigating the pattern and properties of the probability mass function, we recognize that  $q_{i+1}$  actually represents the density function of a discrete phase-type distribution, i.e.  $PH(\alpha, P)$ .  $\square$

### 3.6.2 Proof of Corollary 3.2

**Corollary 3.2** Assume that given  $\Theta = \theta$ ,  $X_1, \dots, X_n$  are random samples from (3.26). Then their joint probability density is

$$f(x_1, \dots, x_n | \Theta = \theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{l=0}^{\infty} \frac{(\theta \sum_{i=1}^n x_i)^l}{l!} C(l, x_1, \dots, x_n), \quad (3.30)$$

where

$$C(l, x_1, \dots, x_n) = \sum_{i=0}^l \binom{l}{i} q_{i+1} \left( \frac{x_n}{\sum_{j=1}^n x_j} \right)^i \left( 1 - \frac{x_n}{\sum_{j=1}^n x_j} \right)^{l-i} C(l-i, x_1, \dots, x_{n-1}) \quad (3.31)$$

for all  $l \in \mathbb{Z}^+ \cup \{0\}$  and  $n \in \mathbb{N}$ . The above iteration is initiated by

$$C(l, x_1) = q_{l+1}, \quad l = 0, 1, 2, \dots$$

*Proof.* The statement can be proved by induction. The statement holds for  $n = 1$ ,

which can be readily verified after substitution. Assuming that the statement holds for  $n$  observations, we show below that the statement also holds for  $n + 1$  observations.

$$\begin{aligned}
& f(x_1, \dots, x_{n+1} | \Theta) \\
&= f(x_1, \dots, x_n | \Theta) f(x_{n+1} | \Theta) \\
&= \theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{l=0}^{\infty} \frac{(\theta \sum_{i=1}^n x_i)^l}{l!} C(l, x_1, \dots, x_n) \cdot \sum_{k=0}^{\infty} q_{k+1} \theta e^{-\theta x_{n+1}} \frac{(\theta x_{n+1})^k}{k!} \\
&= \theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} \frac{(\theta \sum_{i=1}^{n+1} x_i)^{l+k}}{(l+k)!} \frac{(l+k)!}{l!k!} \frac{(\sum_{i=1}^n x_i)^l (x_{n+1})^k}{(\sum_{i=1}^{n+1} x_i)^{l+k}} q_{k+1} C(l, x_1, \dots, x_n) \\
&= \theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{s=0}^{\infty} \sum_{k=0}^s \frac{(\theta \sum_{i=1}^{n+1} x_i)^s}{s!} \frac{s!}{(s-k)!k!} \frac{(\sum_{i=1}^n x_i)^{s-k} (x_{n+1})^k}{(\sum_{i=1}^{n+1} x_i)^s} q_{k+1} C(s-k, x_1, \dots, x_n) \\
&\quad \text{by letting } s = l + k \\
&= \theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{s=0}^{\infty} \frac{(\theta \sum_{i=1}^{n+1} x_i)^s}{s!} \sum_{k=0}^s \binom{s}{k} \frac{(\sum_{i=1}^n x_i)^{s-k} (x_{n+1})^k}{(\sum_{i=1}^{n+1} x_i)^s} q_{k+1} C(s-k, x_1, \dots, x_n) \\
&= \theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{s=0}^{\infty} \frac{(\theta \sum_{i=1}^{n+1} x_i)^s}{s!} C(s, x_1, \dots, x_{n+1}) \\
&\quad \text{by the definition of } C(s, x_1, \dots, x_{n+1}) \text{ function,}
\end{aligned}$$

thereby showing that indeed  $f(x_1, \dots, x_{n+1} | \Theta)$  satisfies the stated forms.  $\square$

### 3.6.3 Proof of Theorem 3.4

**Theorem 3.4** Suppose  $X_1 | \theta, \dots, X_n | \theta$  are i.i.d. distributed random variables with  $PH(\alpha, T)$ . Then a prior which is conjugate for the joint likelihood  $f(x_1, \dots, x_n | \theta)$  can be written as an infinite mixture of Erlang distributions with the probability

density function specified as

$$\pi(\theta) = \sum_{l=0}^{\infty} \zeta_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!}, \quad (3.32)$$

where  $\beta > 0$  and  $m$  is a non-negative integer.  $\zeta_l$  is a probability measure, i.e.  $\sum_{l=0}^{\infty} \zeta_l = 1$  with the possibility  $\zeta_l = 0$  for some  $l$ . The parameters  $\beta$ ,  $m$  and  $\{\zeta_l\}$  are considered as the “updated parameters”.

*Proof.* In order to find the conjugate prior for phase-type distributions, we need to obtain a closed form for the joint density function of multiple claims conditional on a risk parameter, which has been introduced in the previous section. Based on Proposition 3.3, the conjugate prior could be derived as following

$$u(\theta) = \frac{f(x_1, \dots, x_n | \theta)}{\int f(x_1, \dots, x_n | \theta) d\theta}.$$

According to Corollary 3.2, the joint likelihood has the form of (3.30). Then the denominator on the right-hand side above becomes

$$\begin{aligned} \int f(x_1, \dots, x_n | \theta) d\theta &= \int_0^{\infty} \theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{l=0}^{\infty} \frac{(\theta \sum_{i=1}^n x_i)^l}{l!} C(l, x_1, \dots, x_n) d\theta \\ &= \sum_{l=0}^{\infty} C(l, x_1, \dots, x_n) \int_0^{\infty} \theta^n e^{-\theta \sum_{i=1}^n x_i} \frac{(\theta \sum_{i=1}^n x_i)^l}{l!} d\theta \\ &= \sum_{l=0}^{\infty} C(l, x_1, \dots, x_n) \frac{(l+n)!}{l!} \frac{1}{(\sum_{i=1}^n x_i)^{n+1}} \end{aligned}$$



$$\begin{aligned}
& \times \int_0^\infty \underbrace{\left( \sum_{i=1}^n x_i \right) e^{-\theta \sum_{i=1}^n x_i} \frac{(\theta \sum_{i=1}^n x_i)^{l+n}}{(l+n)!}}_{\text{Gamma}(\alpha = l+n+1, \beta = \sum_{i=1}^n x_i)} d\theta \\
& = \frac{1}{\left( \sum_{i=1}^n x_i \right)^{n+1}} \sum_{l=0}^\infty C(l, x_1, \dots, x_n) \frac{(l+n)!}{l!}
\end{aligned}$$

Hence,

$$\begin{aligned}
u(\theta) & = \frac{\theta^n e^{-\theta \sum_{i=1}^n x_i} \sum_{l=0}^\infty \frac{(\theta \sum_{i=1}^n x_i)^l}{l!} C(l, x_1, \dots, x_n)}{\frac{1}{\left( \sum_{i=1}^n x_i \right)^{n+1}} \sum_{k=0}^\infty C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!}} \\
& = \frac{\sum_{l=0}^\infty C(l, x_1, \dots, x_n) \frac{(l+n)!}{l!} \cdot \left( \sum_{i=1}^n x_i \right) e^{-\theta \sum_{i=1}^n x_i} \frac{(\theta \sum_{i=1}^n x_i)^{l+n}}{(l+n)!}}{\sum_{k=0}^\infty C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!}} \\
& = \sum_{l=0}^\infty \left\{ \frac{C(l, x_1, \dots, x_n) \frac{(l+n)!}{l!}}{\sum_{k=0}^\infty C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!}} \right\} \cdot \left( \sum_{i=1}^n x_i \right) e^{-\theta \sum_{i=0}^n x_i} \frac{(\theta \sum_{i=1}^n x_i)^{l+n}}{(l+n)!} \\
& = \sum_{l=0}^\infty \zeta_l \left( \sum_{i=1}^n x_i \right) e^{-\theta \sum_{i=0}^n x_i} \frac{(\theta \sum_{i=1}^n x_i)^{l+n}}{(l+n)!},
\end{aligned}$$

where  $\zeta_l = C(l, x_1, \dots, x_n) \frac{(l+n)!}{l!} / \sum_{k=0}^\infty C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!}$ . We need to check the convergence property of  $\zeta_l$  to ensure the validity of  $u(\theta)$ .  $C$  function presented in Corollary 3.2 is

$$\begin{aligned}
& C(M, x_1, \dots, x_n) \\
& = \sum_{i_1=0}^M \binom{M}{i_1} \left( \frac{x_n}{\sum x_j} \right)^{i_1} \left( 1 - \frac{x_n}{\sum x_j} \right)^{M-i_1} q_{i_1+1} C(M - i_1, x_1, \dots, x_{n-1}) \\
& \leq \sum_{i_1=0}^M q_{i_1+1} C(M - i_1, x_1, \dots, x_{n-1})
\end{aligned}$$

since the Binomial probability is less or equal to 1

$$\begin{aligned}
&= \sum_{i_1=0}^M q_{i_1+1} \sum_{i_2=0}^{M-i_1} \binom{M-i_1}{i_2} \left( \frac{x_{n-1}}{\sum x_j} \right)^{i_2} \left( 1 - \frac{x_{n-1}}{\sum x_j} \right)^{M-i_1-i_2} q_{i_2+1} C(M-i_1-i_2, x_1, \dots, x_{n-2}) \\
&\leq \sum_{i_1=0}^M q_{i_1+1} \sum_{i_2=0}^{M-i_1} q_{i_2+1} C(M-i_1-i_2, x_1, \dots, x_{n-2}) \\
&\quad \vdots \\
&\leq \sum_{i_1=0}^M q_{i_1+1} \sum_{i_2=0}^{M-i_1} q_{i_2+1} \cdots \sum_{i_{n-1}=0}^{M-\sum_{l=1}^{n-1} i_l} q_{i_{n-1}+1} \cdot q_{M+1-\sum_{l=1}^{n-1} i_l} \\
&= \sum \cdots \sum q_{i_1+1} \cdot q_{i_2+1} \cdots q_{M+1-\sum_{l=1}^{n-1} i_l} \\
&= Pr(N_1 + \cdots + N_n = M),
\end{aligned}$$

where  $N_i$  are i.i.d discrete phase-type distributions with the probability density function  $Pr(N_i = l) = q_{l+1}$  with  $q_{l+1}$  specified in Proposition 3.1. Then,

$$\sum_{k=0}^{\infty} C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!} \leq \sum_{k=0}^{\infty} Pr(N_1 + \cdots + N_n = k) (k+1)(k+2) \cdots (k+n).$$

According to the closure properties of phase-type distributions, the summation of independent discrete phase-type distributions,  $N_1 + \cdots + N_n$ , still follows a discrete phase-type distribution denoted by  $N$ . Then the inequality could be rewritten as

$$\begin{aligned}
\sum_{k=0}^{\infty} C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!} &\leq \sum_{k=0}^{\infty} Pr(N = k) (k+1)(k+2) \cdots (k+n) \\
&= E\left(\prod_{i=1}^n (N+i)\right),
\end{aligned}$$

which is the factorial moments of the discrete phase-type distribution  $N$  and its closed form is presented in Corollary 2.5. So  $\sum_{k=0}^{\infty} C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!}$  is bounded. As a consequence, the general term  $C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!}$  approaches 0 as  $k$  goes to infinity.

Therefore, according to the first condition in Proposition 3.3 we have

$$u(\theta) = \sum_{l=0}^{\infty} \zeta_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+n}}{(l+n)!},$$

where  $\zeta_l = C(l, x_1, \dots, x_n) \frac{(l+n)!}{l!} / \sum_{k=0}^{\infty} C(k, x_1, \dots, x_n) \frac{(k+n)!}{k!}$ .

In order to construct a conjugate prior for the PH distribution, we also need to check the second condition of Proposition 3.3, which is the prior set should be closed under the product operation.

$$\begin{aligned} u(\theta)v(\theta) &= \sum_{l=0}^{\infty} \zeta_l \beta_1 e^{-\beta_1\theta} \frac{(\beta_1\theta)^{l+n}}{(l+n)!} \cdot \sum_{m=0}^{\infty} \zeta_m \beta_2 e^{-\beta_2\theta} \frac{(\beta_2\theta)^{m+n}}{(m+n)!} \\ &= \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \zeta_l \zeta_m \beta_1^{l+n+1} \beta_2^{m+n+1} e^{-(\beta_1+\beta_2)\theta} \frac{\theta^{l+m+2n}}{(l+n)!(m+n)!} \\ &= \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \zeta_l \zeta_m (\beta_1 + \beta_2) e^{-(\beta_1+\beta_2)\theta} \frac{[(\beta_1 + \beta_2)\theta]^{l+m+2n}}{(l+m+2n)!} \\ &\quad \times \frac{(l+m+2n)!}{(l+n)!(m+n)!} \left(\frac{\beta_1}{\beta_1 + \beta_2}\right)^{l+n} \left(\frac{\beta_2}{\beta_1 + \beta_2}\right)^{m+n} \left(\frac{\beta_1\beta_2}{\beta_1 + \beta_2}\right) \end{aligned}$$

By changing the variables,  $s = l + m$ , we get

$$u(\theta)v(\theta) = \sum_{s=0}^{\infty} \sum_{l=0}^s \zeta_l \zeta_{s-l} (\beta_1 + \beta_2) e^{-(\beta_1+\beta_2)\theta} \frac{[(\beta_1 + \beta_2)\theta]^{s+2n}}{(s+2n)!}$$

$$\begin{aligned}
& \times \frac{(s+2n)!}{(l+n)!(s-l+n)!} \left(\frac{\beta_1}{\beta_1+\beta_2}\right)^{l+n} \left(\frac{\beta_2}{\beta_1+\beta_2}\right)^{s-l+n} \left(\frac{\beta_1\beta_2}{\beta_1+\beta_2}\right) \\
& = \sum_{s=0}^{\infty} \underbrace{(\beta_1+\beta_2)e^{-(\beta_1+\beta_2)\theta} \frac{[(\beta_1+\beta_2)\theta]^{s+2n}}{(s+2n)!}}_{\text{Gamma}(\alpha=s+2n+1, \beta=\beta_1+\beta_2)} \\
& \quad \times \sum_{l=0}^s \zeta_l \zeta_{s-l} \frac{(s+2n)!}{(l+n)!(s-l+n)!} \left(\frac{\beta_1}{\beta_1+\beta_2}\right)^{l+n} \left(\frac{\beta_2}{\beta_1+\beta_2}\right)^{s-l+n} \left(\frac{\beta_1\beta_2}{\beta_1+\beta_2}\right)
\end{aligned}$$

The above equation can be rewritten as

$$u(\theta)v(\theta) = \frac{\beta_1\beta_2}{\beta_1+\beta_2} \sum_{s=0}^{\infty} \xi_s (\beta_1+\beta_2) e^{-(\beta_1+\beta_2)\theta} \frac{[(\beta_1+\beta_2)\theta]^{s+2n}}{(s+2n)!},$$

where  $\xi_s = \sum_{l=0}^s \zeta_l \zeta_{s-l} \frac{(s+2n)!}{(l+n)!(s-l+n)!} \left(\frac{\beta_1}{\beta_1+\beta_2}\right)^{l+n} \left(\frac{\beta_2}{\beta_1+\beta_2}\right)^{s-l+n}$ .

Therefore,

$$\frac{u(\theta)v(\theta)}{\int u(\theta)v(\theta)d\theta} = \sum_{s=0}^{\infty} \frac{\xi_s}{\sum_{k=0}^{\infty} \xi_k} (\beta_1+\beta_2) e^{-(\beta_1+\beta_2)\theta} \frac{[(\beta_1+\beta_2)\theta]^{s+2n}}{(s+2n)!}.$$

We also need to check that  $\xi_s$  and  $\sum_{s=0}^{\infty} \xi_s$  are convergent as  $s$  goes to infinity.

We have that

$$\begin{aligned}
\lim_{l \rightarrow \infty} \sum_{s=0}^l \xi_s &= \lim_{l \rightarrow \infty} \sum_{s=0}^l \sum_{n=1}^s \zeta_n \zeta_{s-l} \frac{(s+2n)!}{(l+n)!(s-l+n)!} \left(\frac{\beta_1}{\beta_1+\beta_2}\right)^{l+n} \left(\frac{\beta_2}{\beta_1+\beta_2}\right)^{s-l+n} \\
&\leq \lim_{l \rightarrow \infty} \sum_{s=0}^l \sum_{n=0}^s \zeta_n \zeta_{s-l+n} \\
&= \lim_{l \rightarrow \infty} \sum_{s=0}^l P(Z_1 + Z_2 = s+2n) \\
&= 1.
\end{aligned}$$

Therefore the series  $\sum_{s=0}^{\infty} \xi_s$  is convergent. As a necessary condition, we know that the general term  $\xi_s$  should approach 0 as  $s$  increases. Thus, the prior set is closed under the product operation.

Therefore, by natural extension,

$$u(\theta) = \sum_{l=0}^{\infty} \zeta_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!}, \quad (3.33)$$

is a suitable form of the conjugate prior for phase-type distributions according to Proposition 3.3, with  $\zeta_l$  following a discrete distribution.  $\square$

### 3.6.4 Proof of Theorem 3.5

**Theorem 3.5** Suppose  $X_1|\Theta, \dots, X_n|\Theta$  are i.i.d. distributed random variables with  $PH(\alpha, T)$ . If  $\Theta$  follows a distribution specified by (3.32), the Bühlmann premium estimator based on losses  $X_1, \dots, X_n$  is

$$P_{Cr} = \frac{n}{n+k} \bar{X} + \frac{k}{n+k} \mu, \quad (3.34)$$

where  $N$  is defined by (3.19),  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  and

$$\begin{aligned} \mu &= E(N)\beta \sum_{l=0}^{\infty} \frac{\zeta_l}{l+m}, \\ k &= \frac{Var(N) + E(N)}{(E(N))^2} \cdot g(m, \zeta) \end{aligned} \quad (3.35)$$

with

$$g(m, \zeta) = \frac{\sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)}}{\sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)} - \left( \sum_{l=0}^{\infty} \frac{\zeta_l}{l+m} \right)^2}.$$

Here parameters  $m > 1$ ,  $m \in \mathbb{Z}$ ,  $\beta$  and  $\{\zeta_l\}$  follows the same definitions from Theorem 3.4.

*Proof.* The conditional mean and variance of  $X|\Theta$  can be expressed as

$$\begin{aligned} \mu(\Theta) &= E(X|\Theta) = E(Y_1|\Theta)E(N) \\ &= E(N)\Theta^{-1}, \end{aligned}$$

and

$$\begin{aligned} v(\Theta) &= Var(X|\Theta) \\ &= E^2(Y_1|\Theta)Var(N) + E(N)Var(Y_1|\Theta) \\ &= (Var(N) + E(N))\Theta^{-2}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \mu &= E(\mu(\Theta)) \\ &= E(N)E(\Theta^{-1}) \\ &= E(N) \int_0^{\infty} \theta^{-1} \sum_{l=0}^{\infty} \zeta_l \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!} d\theta \end{aligned}$$

$$\begin{aligned}
&= E(N) \sum_{l=0}^{\infty} \zeta_l \frac{\beta}{l+m} \int_0^{\infty} \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m-1}}{(l+m-1)!} d\theta \\
&= E(N) \sum_{l=0}^{\infty} \zeta_l \frac{\beta}{l+m},
\end{aligned}$$

$$\begin{aligned}
v &= E\{(Var(N) + E(N))\Theta^{-2}\} \\
&= [Var(N) + E(N)] \int_0^{\infty} \theta^{-2} \sum_{l=0}^{\infty} \zeta_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!} d\theta \\
&= [Var(N) + E(N)] \sum_{l=0}^{\infty} \zeta_l \frac{\beta^2}{(l+m)(l+m-1)} \int_0^{\infty} \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m-2}}{(l+m-2)!} d\theta \\
&= [Var(N) + E(N)] \beta^2 \sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)},
\end{aligned}$$

Similarly,

$$\begin{aligned}
a &= Var\{E(N)\Theta^{-1}\} \\
&= (E(N))^2 \{E(\Theta^{-2}) - (E(\Theta^{-1}))^2\} \\
&= (E(N))^2 \beta^2 \left\{ \sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)} - \left( \sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)} \right)^2 \right\}.
\end{aligned}$$

Therefore, the Bühlmann coefficient  $k$  is given by

$$k = \frac{Var(N) + E(N)}{(E(N))^2} \frac{\sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)}}{\sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)} - \left( \sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)} \right)^2}.$$

As the number of transitions until absorption  $N$  follows a discrete phase-type

distribution with parameters  $(\boldsymbol{\alpha}, P)$ , we have the first two factorial moments from Latouche and Ramaswami (1999) as

$$\begin{aligned} E(N) &= \boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1}, \\ E[N(N - 1)] &= 2\boldsymbol{\alpha}'P(I - P)^{-2}\mathbf{1}. \end{aligned}$$

Then we can obtain that

$$\begin{aligned} \text{Var}(N) + E(N) &= E(N^2) - (E(N))^2 + E(N) \\ &= E[N(N - 1)] + 2E(N) - (E(N))^2 \\ &= 2\boldsymbol{\alpha}'P(I - P)^{-1}\mathbf{1} + 2\boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1} - (\boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1})^2 \\ &= 2\boldsymbol{\alpha}'(P + (I - P))(I - P)^{-2}\mathbf{1} - (\boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1})^2 \\ &= 2\boldsymbol{\alpha}'(I - P)^{-2}\mathbf{1} - (\boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1})^2. \end{aligned}$$

□



# Chapter 4

## Bayesian Premium for

## Phase-type Distributed Losses

In the class of all linear estimators, the Bühlmann premium we considered in Chapter 3 provides the minimum squared error loss  $Q$  defined in (3.2). However, the Bühlmann model is restrictive in the sense of the linear structure it relied upon. By relaxing this linearity restriction we can study the more general Bayesian premium which offers the minimum squared error loss amongst all estimators.

This chapter starts by a rigorous account of the mathematical formulation of Bayesian premiums. We will then derive the estimator under phase-type distributed losses using the same conjugate prior distribution outlined in Theorem 3.4. We will also investigate some interesting properties of the marginal distribution of the loss.

## 4.1 The Bayesian Premium

In this section we give a detailed review of the formulation of Bayesian premium to make the thesis self-contained. The materials are largely based on the presentation from Klugman et al. (2008).

To construct the Bayesian premium estimator, suppose that for a particular policyholder, we have observed  $n$  losses  $\mathbf{x} = (x_1, \dots, x_n)$ , which is viewed as the observations from the random vector  $\mathbf{X} = (X_1, \dots, X_n)$  just as before. We are interested in setting a rate to cover the future loss  $X_{n+1}$ .

Similar to Chapter 3, we continue to assume that the sampling distributions  $X_1, \dots, X_n$  given the risk parameter  $\theta$  are independent and identically distributed. This independence can be interpreted as the relative irrelevance between the experiences of the policyholder during different exposure periods.

As already been argued, the most ideal premium estimator is the risk premium  $\mu_{n+1}(\theta) = E(X_{n+1}|\theta)$  but it is not tractable. Fortunately by applying Bayesian inference we can infer plausible values of  $\theta$  based on historical data  $x_i$ . In other words, the Bayesian premium focuses on  $E(X_{n+1}|\mathbf{X})$  instead of  $E(X_{n+1}|\theta)$ .

Now we start to derive the Bayesian premium. Suppose the risk parameter  $\Theta$  has a prior distribution  $\pi(\theta)$  and the sampling distribution of  $X_i|\theta$  has a density  $f_{X_j|\Theta}(x_j|\theta)$ . By the independence of the losses we have the joint likelihood:

$$f(x_1, \dots, x_n|\theta) = \prod_{j=1}^n f_{X_j|\Theta}(x_j|\theta),$$

based on which the posterior distribution can be derived:

$$\pi(\theta|x_1, \dots, x_n) = h \cdot f(x_1, \dots, x_n|\theta)\pi(\theta) = h \cdot \left( \prod_{j=1}^n f_{X_j|\Theta}(x_j|\theta) \right) \pi(\theta),$$

with the normalization factor given by

$$h^{-1} = \int \left( \prod_{j=1}^n f_{X_j|\Theta}(x_j|\theta) \right) \pi(\theta) d\theta.$$

The conditional density of  $X_{n+1}|\mathbf{X}$  can thus be obtained by

$$f(x_{n+1}|x_1, \dots, x_n) = \int f(x_{n+1}|\theta)\pi(\theta|x_1, \dots, x_n) d\theta. \quad (4.1)$$

Based on the above results the Bayesian premium can be derived as follows

$$\begin{aligned} E(X_{n+1}|X_1, \dots, X_n) &= \int x_{n+1}f(x_{n+1}|x_1, \dots, x_n) dx_{n+1} \\ &= \int x_{n+1} \left( \int f(x_{n+1}|\theta)\pi(\theta|x_1, \dots, x_n) d\theta \right) dx_{n+1} \\ &= \int \left( \int x_{n+1}f(x_{n+1}|\theta) dx_{n+1} \right) \pi(\theta|x_1, \dots, x_n) d\theta \\ &= \int \mu_{n+1}(\theta)\pi(\theta|x_1, \dots, x_n) d\theta, \end{aligned} \quad (4.2)$$

where the integral is evaluated with regard to all possible values of  $\theta$ .

Equation (4.2) shows that the Bayesian premium considers the risk premium  $\mu_{n+1}(\theta)$  and the posterior density  $\pi(\theta|\mathbf{x})$  collectively. In a certain sense, the

Bayesian model provides an indirect approach to evaluate the hypothetical object  $\mu_{n+1}(\theta)$  while the knowledge of  $\theta$  is gained from the posterior.

## 4.2 Bayesian Premium for Phase-type Distributed Losses

Now we are ready to proceed with the derivation of the Bayesian premium estimator given phase-type distributed losses. The same set of distribution assumptions are applied as the Bühlmann estimator: the sampling distribution of  $X|\Theta$  is assumed to be  $PH(\alpha, T)$  with its conjugate prior distribution assigned to  $\Theta$ . The characterization of the resulting posterior distribution is given by the following lemma.

**Lemma 4.1.** *Suppose  $X_1|\Theta, \dots, X_n|\Theta$  are i.i.d. distributed random variables with  $PH(\alpha, T)$ . If  $\Theta$  follows a distribution specified by (3.32), the corresponding posterior distribution is*

$$\pi(\theta|x_1, \dots, x_n) = \sum_{s=0}^{\infty} \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} \frac{B(s, x_1, \dots, x_n, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)}, \quad (4.3)$$

where the parameters  $\beta$ ,  $m$  and  $\{\zeta_l\}$  are specified in Proposition 3.3.  $B$  is defined

by a recursive relation

$$\begin{aligned}
B(s, x_1, \dots, x_n, m, \beta) &= (s + m + n) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum_{i=0}^s q_{i+1} B(s - i, x_1, \dots, x_{n-1}, m, \beta) \\
&\times \binom{s + m + n - 1}{i} \left(1 - \frac{x_n}{\beta + \sum_{j=1}^n x_j}\right)^{s-i+m+n-1} \left(\frac{x_n}{\beta + \sum_{j=1}^n x_j}\right)^i, \quad (4.4)
\end{aligned}$$

with initial conditions

$$B(s, x_1, m, \beta) = (s+m+1) \frac{\beta}{(\beta + x_1)^2} \sum_{i=0}^s \binom{s+m}{i} \left(\frac{\beta}{\beta + x_1}\right)^{s+m-i} \left(\frac{x_1}{\beta + x_1}\right)^i q_{i+1} B(s-i).$$

and

$$B(l) = \zeta_l.$$

*Proof.* See section 4.5.1. □

From Lemma 4.1, we notice that the posterior distribution is also expressed as infinite mixtures of Erlang distributions with common scale parameters, which follows the same pattern of the prior distribution but with different “updated parameters”.

We want to make a note on the structure of the object  $B(s, x_1, \dots, x_n, m, \beta)$  defined above. We may loosely call it as “function”  $B$  but technically, each iteration defines a different function. The first one starts with  $\zeta_l$  with no  $x_i$  involved, then each time a iteration is applied to bring in one additional  $x_i$ . The iteration hence produces a sequence of functions which we denote by  $B(l)$ ,

$B(s, x_1, m, \beta)$ ,  $B(s, x_1, x_2, m, \beta)$  and so on.

Now we present the most important result of this chapter.

**Theorem 4.2.** *Suppose  $X_1|\Theta, \dots, X_n|\Theta$  are i.i.d. distributed random variables with  $PH(\alpha, T)$ . If  $\Theta$  follows a distribution specified by (3.32), the Bayesian premium estimator is*

$$E(X_{n+1}|X_1, \dots, X_n) = E(N)\left(\beta + \sum_{j=1}^n x_j\right) \frac{\sum_{s=0}^{\infty} \frac{1}{s+m+n} B(s, x_1, \dots, x_n, m, \beta)}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)}, \quad (4.5)$$

where  $N$  is defined by (3.19), parameters  $m$  and  $\beta$  follows the same definitions from Theorem 3.4 and  $B(s, x_1, \dots, x_n, m, \beta)$  follows the same definition from Lemma 4.1.

*Proof.* See section 4.5.2. □

Equation (4.16) incorporates complicated mathematical structures. It also involves infinite series as equation (3.34) but are far more difficult to handle. The general term contains a function  $B(s, x_1, \dots, x_n, m, \beta)$  which is defined by a recursive relation. Therefore it is quite difficult to obtain a closed form solution of the Bayesian premium. In section 4.5.2 we will prove the convergence of the relevant infinite series, which is not as trivial as in the Bühlmann case.

### 4.3 Marginal and Predictive Distributions for the Losses

Sometimes it is interesting to know the marginal distribution of losses which is independent of risk parameter  $\theta$ . It provides valuable insight around the risk pattern of a policyholder. The other motivation is that knowing the explicit form of the marginal distribution of losses is an important element for empirical Bayes estimation of prior parameters. In this section we intend to investigate this problem by applying some known results obtained previously.

From the proof of Theorem 4.2, we know that under our setting the joint density function of  $X_1, \dots, X_n$  is

$$f(x_1, \dots, x_n) = \sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta),$$

where related objects  $m$ ,  $\beta$  and  $B$  are defined in Theorem 4.2.

Analyzing the above form is difficult due to the algebraic complexity introduced by function  $B$ . Nevertheless, we can still study some special cases. The following corollary contains the result derived from one particular case.

**Corollary 4.3.** *Suppose a loss  $X$  has a likelihood  $X|\Theta \sim PH(\alpha, \theta(P - I))$ . If  $\Theta$  follows a distribution specified by (3.32) with  $m = 0$  and  $\{\zeta_l\}$  being a probability*

measure representing a Geometric distribution, i.e.

$$\zeta_l = (1 - p)^l p, \quad l \in \mathbb{Z}^+ \cup \{0\},$$

where  $0 < p < 1$ . Then the marginal distribution of  $X$  is

$$f(x) = \sum_{l=0}^{\infty} q_{l+1} \frac{(l+1) \cdot \beta p \cdot x^l}{(x + \beta p)^{l+2}}, \quad (4.6)$$

where  $q_{l+1}$  is defined in Proposition 3.1.

*Proof.* See section 4.5.3. □

The term  $\frac{(l+1) \cdot \beta p \cdot x^l}{(x + \beta p)^{l+2}}$  is actually the density function of an inverse Pareto distribution with shape and scale parameters being  $l + 1$  and  $\beta p$  respectively. The corollary tells us under this special setting, marginal distribution for one loss is in fact an infinite mixture of inverse Pareto distributions.

Equation (4.19) also has an equivalent closed form. To derive this, first notice that we can rewrite (4.19) as

$$\begin{aligned} f(x) &= \frac{\beta p}{(x + \beta p)^2} \sum_{l=0}^{\infty} q_{l+1} \frac{d}{dz} (z^{l+1}) \Big|_{z=\frac{x}{x+\beta p}} \\ &= \frac{\beta p}{(x + \beta p)^2} \frac{d}{dz} \left\{ \sum_{l=0}^{\infty} q_{l+1} z^{l+1} \right\} \Big|_{z=\frac{x}{x+\beta p}}, \end{aligned}$$

where  $q_{l+1}$  is defined in Corollary 3.1 and can be considered as the probabil-



ity density function of  $PH_d(\boldsymbol{\alpha}, P)$ . We also know from Corollary 2.5 that the probability generating function of this discrete phase-type distribution is

$$P(z) = z\alpha(I - zP)^{-1}w,$$

where  $w = (I - P)\mathbf{1}$  and  $I$  is an identity matrix of appropriate dimension. Therefore

$$\begin{aligned} \frac{d}{dz} \left\{ \sum_{l=0}^{\infty} q_{l+1} z^{l+1} \right\} &= \frac{d}{dz} P(z) \\ &= \frac{d}{dz} z\alpha(I - zP)^{-1}w \\ &= \alpha(I - zP)^{-1}w + z\alpha \frac{d}{dz} (I - zP)^{-1}w \\ &= \alpha(I - zP)^{-1}w + z\alpha(I - zP)^{-1}P(I - zP)^{-1}w \\ &= \alpha(I - zP)^{-1}(I + zP(I - zP)^{-1})w \\ &= \alpha(I - zP)^{-1}(I + zP(I - zP)^{-1})(I - zP)(I - zP)^{-1}w \\ &= \alpha(I - zP)^{-1}(I - zP + zP)(I - zP)^{-1}w \\ &= \alpha(I - zP)^{-2}w. \end{aligned}$$

So the marginal density function  $f(x)$  can be also expressed as

$$f(x) = \frac{\beta p}{(x + \beta p)^2} \alpha \left( I - \frac{x}{x + \beta p} P \right)^{-2} (I - P)\mathbf{1}.$$

The following Corollary is to show the explicit expression of the predictive

distribution  $X_{n+1}|X_1, \dots, X_n$ , which is the relevant distribution to be employed for risk management purposes. Mathematically, the derivation of the predictive distribution's density function is virtually identical to the marginal loss distribution's, except that the posterior is used as mixing distribution rather than the prior. So it will be a natural extension to discuss the predictive distribution in conjunction with the marginal loss distribution.

**Corollary 4.4.** *Suppose a loss  $X$  has a likelihood  $X|\Theta \sim PH(\boldsymbol{\alpha}, \theta(P - I))$ . If the posterior  $\Theta|X_1, \dots, X_n$  follows a distribution specified in Lemma 4.1, then the density function of the predictive distribution is given by*

$$f(x_{n+1}|x_1, \dots, x_n) = \frac{\sum_{k=0}^{\infty} B(k, x_1, \dots, x_{n+1}, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)}, \quad (4.7)$$

where the function  $B(k, x_1, \dots, x_n, m, \beta)$  is also specified in Lemma 4.1.

*Proof.* See section 4.5.4. □

## 4.4 Examples

**Example 4.4.1.** *From the previous chapter, the prior can be expressed as*

$$\sum_{l=0}^{\infty} \zeta_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!},$$

where  $\beta > 0$ ,  $m$  is a non-negative integer and  $\sum \zeta_l = 1$ . We would like to construct a single Gamma prior and compute the Bayesian premium accordingly by setting  $\beta = \gamma$ ,  $m = 0$ , and

$$\zeta_l = \begin{cases} 1, & l = K - 1; \\ 0, & \text{otherwise.} \end{cases}$$

We also know that

$$B(s, x_1, \dots, x_n, m, \beta) = \sum_{l=0}^s \frac{(s+m+n)!}{l!(s-l+m)!} \frac{\beta^{s-l+m+1} (\sum_{j=1}^n x_j)^l}{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}} \zeta_{s-l} C(l, x_1, \dots, x_n).$$

So,

$$B(s, x_1, \dots, x_n, 0, \gamma) = \begin{cases} 0, & \text{when } s < K - 1; \\ \frac{(s+n)!}{(s-K+1)!(K-1)!} \frac{\gamma^K (\sum_{j=1}^n x_j)^{s-K+1}}{(\gamma + \sum_{j=1}^n x_j)^{s+n+1}} 1 \\ \times C(s - K + 1, x_1, \dots, x_n), & \text{when } s > K - 1. \end{cases}$$

The Bayesian premium for Phase-type distributions is

$$E(X_{n+1} | X_1, \dots, X_n) = E(N) \left( \gamma + \sum_{j=1}^n x_j \right) \frac{\sum_{s=0}^{\infty} \frac{1}{s+n} B(s, x_1, \dots, x_n, 0, \gamma)}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, 0, \gamma)}.$$

Then the denominator of the Bayesian premium is

$$\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, 0, \gamma)$$

$$\begin{aligned}
&= \sum_{s=K-1}^{\infty} B(s, x_1, \dots, x_n, 0, \gamma) \\
&= \sum_{s=K-1}^{\infty} \frac{(s+n)!}{(s-K+1)!(K-1)!} \frac{\gamma^K (\sum_{j=1}^n x_j)^{s-K+1}}{(\gamma + \sum_{j=1}^n x_j)^{s+n+1}} {}_1C(s-K+1, x_1, \dots, x_n) \\
&= \sum_{l=0}^{\infty} \frac{(l+K+n-1)!}{l!(K-1)!} \frac{\gamma^K (\sum_{j=1}^n x_j)^l}{(\gamma + \sum_{j=1}^n x_j)^{l+K+n}} C(l, x_1, \dots, x_n), \text{ by letting } l = s - K + 1 \\
&= \frac{\gamma^K}{(K-1)!} \sum_{l=0}^{\infty} \frac{(l+K+n-1)!}{(\gamma + \sum_{j=1}^n x_j)^{l+K+n}} \frac{(\sum_{j=1}^n x_j)^l}{l!} C(l, x_1, \dots, x_n).
\end{aligned}$$

Similarly, the numerator of the Bayesian premium is

$$\begin{aligned}
&\sum_{s=0}^{\infty} \frac{1}{s+n} B(s, x_1, \dots, x_n, 0, \gamma) \\
&= \sum_{s=K-1}^{\infty} \frac{1}{s+n} B(s, x_1, \dots, x_n, 0, \gamma) \\
&= \sum_{s=K-1}^{\infty} \frac{1}{s+n} \frac{(s+n)!}{(s-K+1)!(K-1)!} \frac{\gamma^K (\sum_{j=1}^n x_j)^{s-K+1}}{(\gamma + \sum_{j=1}^n x_j)^{s+n+1}} {}_1C(s-K+1, x_1, \dots, x_n) \\
&= \sum_{s=K-1}^{\infty} \frac{(s+n-1)!}{(s-K+1)!(K-1)!} \frac{\gamma^K (\sum_{j=1}^n x_j)^{s-K+1}}{(\gamma + \sum_{j=1}^n x_j)^{s+n+1}} {}_1C(s-K+1, x_1, \dots, x_n) \\
&= \sum_{l=0}^{\infty} \frac{(l+K+n-2)!}{l!(K-1)!} \frac{\gamma^K (\sum_{j=1}^n x_j)^l}{(\gamma + \sum_{j=1}^n x_j)^{l+K+n}} C(l, x_1, \dots, x_n), \text{ by letting } l = s - K + 1 \\
&= \frac{\gamma^K}{(K-1)!} \frac{1}{\gamma + \sum_{j=1}^n x_j} \sum_{l=0}^{\infty} \frac{(l+K+n-2)!}{(\gamma + \sum_{j=1}^n x_j)^{l+K+n-1}} \frac{(\sum_{j=1}^n x_j)^l}{l!} C(l, x_1, \dots, x_n).
\end{aligned}$$

Hassan-Zadeh and Stanford (2013) also calculated the Bayesian premium of Phase-type distributions with a Gamma prior and introduced a “d function” such that

$$d(K, \gamma, x_1, \dots, x_n) = \sum_{l=0}^{\infty} \frac{\Gamma(l+K+n)}{(n\bar{x} + \gamma)^{l+K+n}} \frac{(n\bar{x})^l}{l!} C(l, x_1, \dots, x_n).$$

Therefore, the denominator and numerator can be rewritten as

$$\begin{aligned} \sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, 0, \gamma) &= \frac{\gamma^K}{(K-1)!} d(K, \gamma, x_1, \dots, x_n) \\ \sum_{s=0}^{\infty} \frac{1}{s+n} B(s, x_1, \dots, x_n, 0, \gamma) &= \frac{1}{\gamma + \sum_{j=1}^n x_j} \frac{\gamma^K}{(K-1)!} d(K-1, \gamma, x_1, \dots, x_n) \end{aligned}$$

Hence, the Bayesian premium is

$$E(X_{n+1} | X_1, \dots, X_n) = E(N) \frac{d(K-1, \gamma, x_1, \dots, x_n)}{d(K, \gamma, x_1, \dots, x_n)},$$

which is consistent with the result derived by Hassan-Zadeh and Stanford (2013).

**Example 4.4.2.** We have now obtained solutions for both Bühlmann and Bayesian premiums in Theorem 3.5 and Theorem 4.2 respectively, under the setting of phase-type distributed losses with the corresponding conjugate prior. In this experiment, we conduct a number of simulations to investigate and compare their properties.

The first step is to set up the conjugate prior distribution. From Theorem 3.4, we know the density of this distribution is given by

$$\pi(\theta) = \sum_{l=0}^{\infty} \zeta_l \cdot \beta e^{-\beta\theta} \frac{(\beta\theta)^{l+m}}{(l+m)!},$$

where  $\zeta_l$  represents some probability measure. In this example, we apply a prior

by imposing the following parameter setting:

$$\zeta_l = \begin{cases} \frac{1}{5}, & l = 0; \\ \frac{3}{5}, & l = 10; \\ \frac{1}{5}, & l = 40; \\ 0, & \text{otherwise,} \end{cases} \quad (4.8)$$

as well as  $m = 2$  and  $\beta = 8$ . Substituting those values into  $\pi(\theta)$  above gives us

$$\pi(\theta) = \frac{1}{5}\beta e^{-\beta\theta} \frac{(\beta\theta)^2}{2!} + \frac{3}{5}\beta e^{-\beta\theta} \frac{(\beta\theta)^{12}}{12!} + \frac{1}{5}\beta e^{-\beta\theta} \frac{(\beta\theta)^{42}}{42!}, \quad (4.9)$$

which is essentially a mixture of  $\text{Gamma}(3, 8)$ ,  $\text{Gamma}(13, 8)$  and  $\text{Gamma}(43, 8)$  densities<sup>1</sup>. We also set up the distribution of losses  $X_i|\theta \sim PH(\boldsymbol{\alpha}, \theta(P - I))$ , where  $\boldsymbol{\alpha} = (1, 0)$  and the matrix

$$P = \begin{pmatrix} 0 & 0.4 \\ 0.8 & 0 \end{pmatrix}.$$

Now we proceed with the calculation of Bühlmann premium. Suppose we observe losses  $x_1, \dots, x_n$ . By Theorem 3.5 we have

$$f = \sum_{l=0}^{\infty} \frac{\zeta_l}{l+m} = \frac{1}{5} \left( \frac{1}{m} + \frac{1}{10+m} + \frac{1}{40+m} \right),$$

<sup>1</sup>We adopt the conventions that  $\text{Gamma}(\alpha, \beta)$  indicated the Gamma density with  $\alpha$  being the shape parameter and  $\beta$  being the rate parameter

$$h = \sum_{l=0}^{\infty} \frac{\zeta_l}{(l+m)(l+m-1)} = \frac{1}{5} \left( \frac{1}{m(m-1)} + \frac{3}{(10+m)(9+m)} + \frac{1}{(40+m)(39+m)} \right),$$

$$E(N) = \boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1},$$

$$\text{Var}(N) + E(N) = 2\boldsymbol{\alpha}'(I - P)^{-2}\mathbf{1} - (\boldsymbol{\alpha}'(I - P)^{-1}\mathbf{1})^2.$$

Based on the above results we can obtain  $\mu = 2.54902$  and  $k = 1.508471$ . Therefore the Bühlmann premium can be calculated as

$$P_{Cr} = \frac{n}{n + 1.508471} \bar{x} + \frac{3.845122}{n + 1.508471}, \quad (4.10)$$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ .

To calculate the Bayesian premium we follow Theorem 4.2. The first two terms on the right-hand side of equation (4.16) are not difficult to handle since  $E(N)$  and  $\beta + \sum_{i=1}^n x_i$  can be easily calculated based on the parameter setting. The last term is more difficult to evaluate algebraically since the recursive relation (4.15) that defines function  $B$  has very complex structure. For this reason, we target to find an approximation of the Bayesian premium instead of a closed form solution as (4.10).

Given observations  $x_1, \dots, x_n$  we can initiate the iteration (4.15) by setting  $B(l)$  according to (4.8) along with  $m$  and  $\beta$  specified in the example. Keep applying (4.15) we eventually move to the level of  $B(s, x_1, \dots, x_n, m, \beta)$ . Denote  $S(k) = B(k, x_1, \dots, x_n, m, \beta)$ , for some small number  $\epsilon > 0$  we find a positive integer  $N$

such that

$$\left| \sum_{k=0}^N S(k) - \sum_{k=0}^{N-1} S(k) \right| < \epsilon, \quad (4.11)$$

$$\left| \sum_{k=0}^N \frac{S(k)}{k+m+n} - \sum_{k=0}^{N-1} \frac{S(k)}{k+m+n} \right| < \epsilon. \quad (4.12)$$

Now define

$$\hat{B} = \sum_{k=0}^N \frac{\frac{1}{k+m+n} S(k)}{\sum_{t=0}^N S(t)}$$

then we approximate the Bayesian premium by:

$$P_{Ba} = E(N) \left( \beta + \sum_{i=1}^n x_i \right) \hat{B} = 2.058824 \left( 8 + \sum_{i=1}^n x_i \right) \hat{B}. \quad (4.13)$$

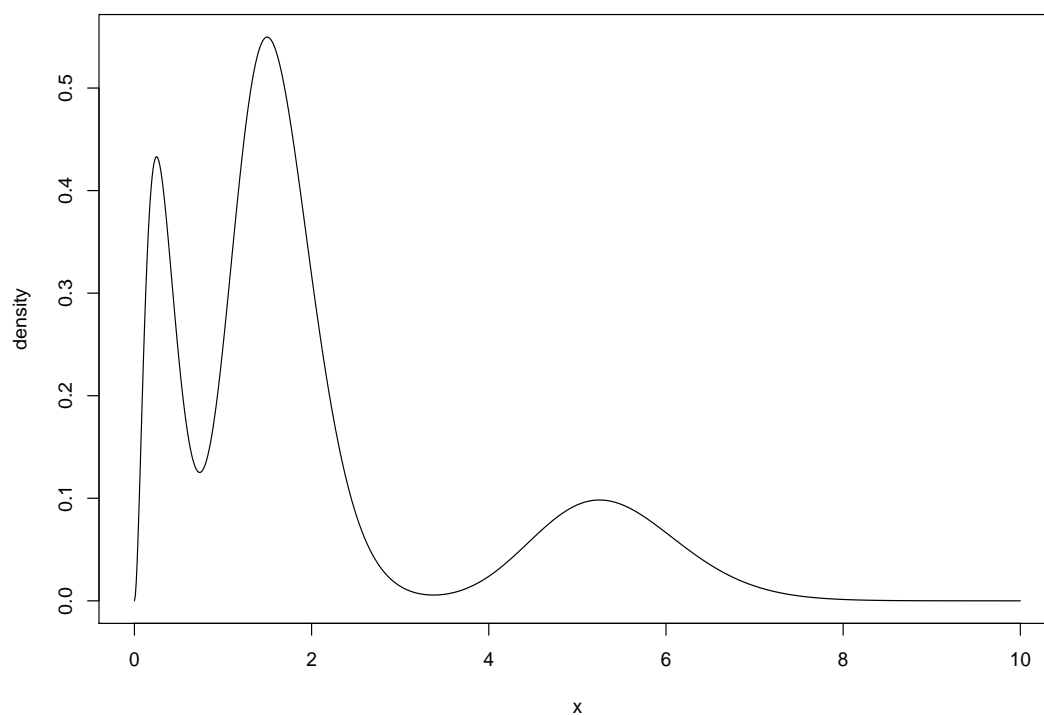
To compare Bayesian and Bühlmann estimators we conduct a simulation experiment based on the following algorithm:

1. Draw 200 values of  $\theta$  from the prior density (4.9), denote by  $\theta_1, \dots, \theta_{200}$ .
2. For  $\theta_1$ , generate  $x_1, \dots, x_{30}$  based on the  $PH(\boldsymbol{\alpha}, \theta_1(P-I))$  with parameters specified previously. Estimate Bühlmann and Bayesian premiums according to (4.10) and (4.13).
3. Repeat steps 2 for 10000 times obtaining 10000 Bühlmann premiums  $P_{Cr_1}^{(j)}$  and Bayesian premiums  $P_{Ba_1}^{(j)}$ ,  $j = 1, \dots, 10000$ .
4. Repeat step 3 for  $\theta_2, \dots, \theta_{200}$ . We finally end up with Bühlmann premi-



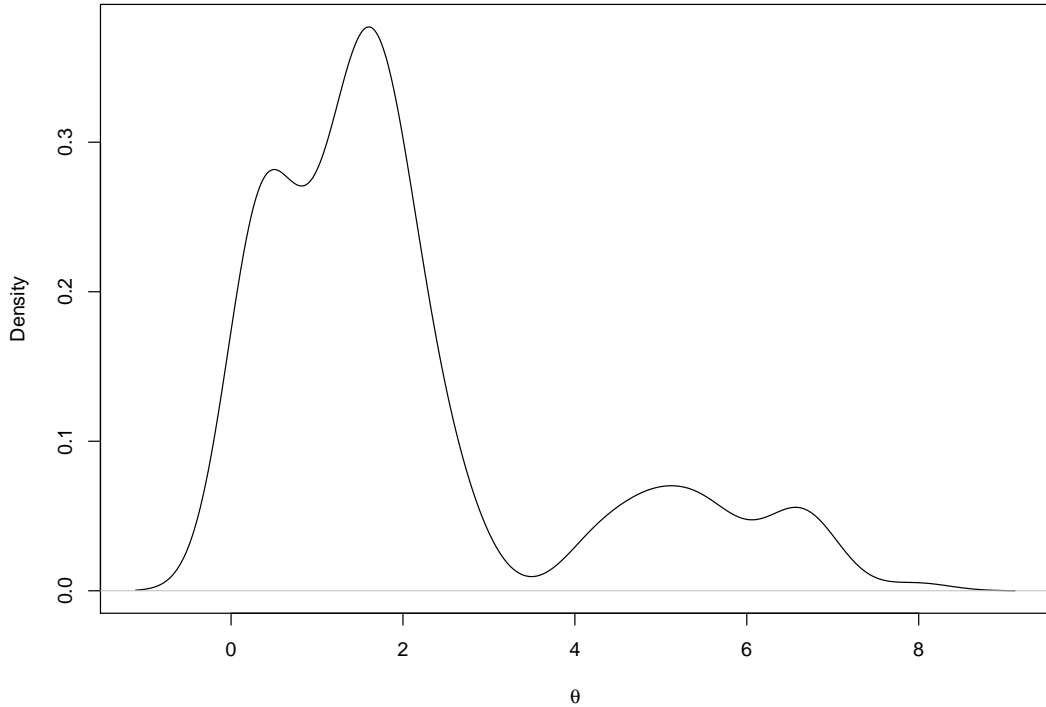
ums  $P_{Cr_i}^{(j)}$  and Bayesian premiums  $P_{Ba_i}^{(j)}$ , where  $i = 1, \dots, 200$  and  $j = 1, \dots, 10000$ .

We start with investigating the prior distribution. The following Figure 4.1 shows the kernel density plot of the prior distribution.



**Figure 4.1:** Density plot of prior distribution: finite mixtures of 3 Gammas

This density has most of the probability mass focused roughly between 1 and 3. The rest of the probabilities are spread upon two areas:  $(0, 1)$  and  $(4, 7)$ . Figure 4.2 shows the density estimation based on the 200  $\theta$ s drawn.



**Figure 4.2:** Density estimation based on 200 samples of  $\theta$

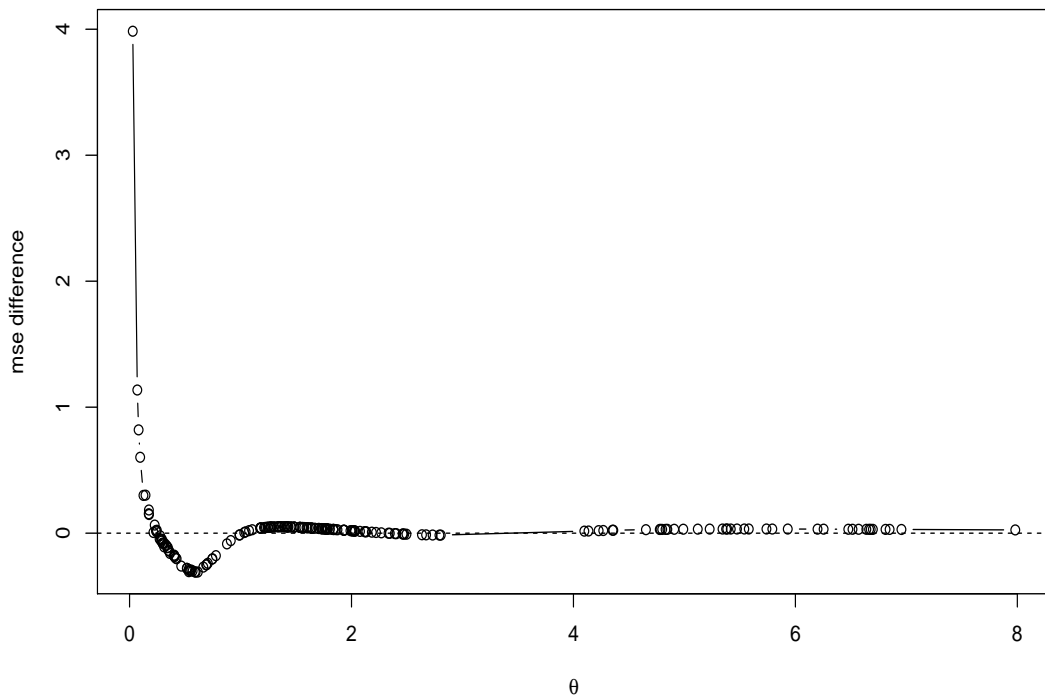
Comparing with Figure 4.1, we see our 200  $\theta$ s approximate the true density well in general. However, we found that the left tail has been considerably over-estimated as we are not seeing an obvious drop around  $\theta = 1$ .

One of the most important comparison between Bühlmann and Bayesian estimators is around the mean squared error. Considering  $\mu(\theta_i) = E(N)/\theta_i$ , the MSEs for Bühlmann and Bayesian premiums can be estimated by

$$MSE_{\theta_i}(P_{Cr}) = \frac{1}{10000} \sum_{j=1}^{10000} (P_{Cr_i}^{(j)} - E(N)/\theta_i)^2,$$

$$MSE_{\theta_i}(P_{Ba}) = \frac{1}{10000} \sum_{j=1}^{10000} (P_{Ba_i}^{(j)} - E(N)/\theta_i)^2.$$

The graph below shows respective MSEs against 200 different values of  $\theta$ .



**Figure 4.3:**  $MSE(P_{Cr})$  minus  $MSE(P_{Ba})$ : finite mixtures of 3 Gammas

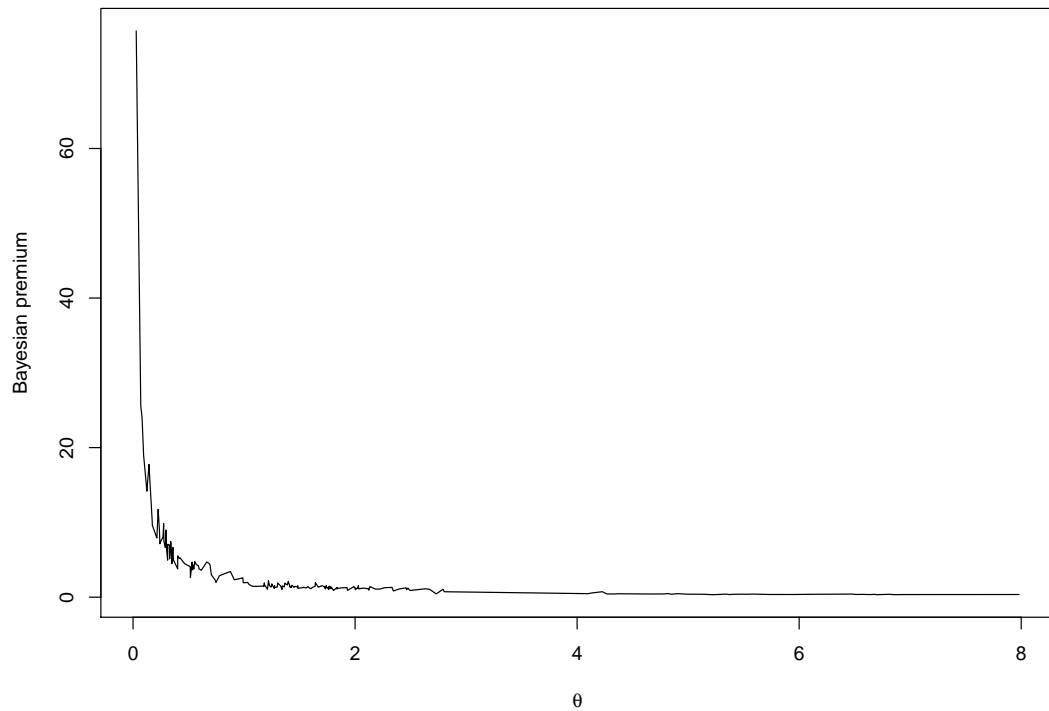
From Figure 4.3 we see that the Bayesian estimator does not always yield less MSE. Roughly speaking on the region  $(0.1, 1)$ , we see higher MSEs for the Bayesian estimator and vice versa on the rest of the support. However, according to Figure 4.1, we know the area  $(0.1, 1)$  does not take account of most probability mass. Therefore, we still have good reason to believe that the Bayesian premium will yield less squared error loss than the Bühlmann premium. Taking the average

---

over the 200 MSEs, our result has confirmed that the Bayesian premium yields a lower average MSE of 1.778677 while the Bühlmann premium yields 1.799254. We also noticed that for extremely small  $\theta$  the Bayesian premium seems to have much greater advantage over the Bühlmann premium by yielding smaller values of MSE.

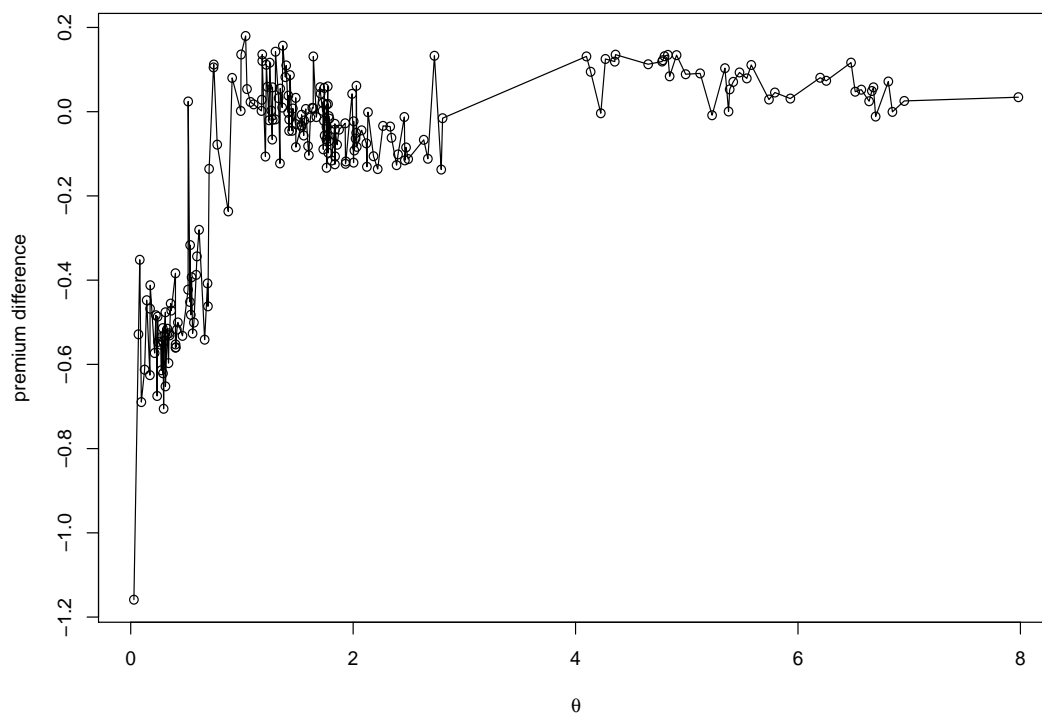
Ideally, in order to confirm the fact that Bayesian premium does yield lower squared error loss, we wish to produce a chart in the same fashion as Table 3.2. Unfortunately, our algorithm for computing the Bayesian premium turned out to be quite time consuming: the MSEs obtained for 200  $\theta$ s shown in Figure 4.3 caused a computational time of more than 2 weeks. Further increasing the number of  $\theta$ s involved will lead to even greater computational time.

Besides MSEs, we also want to investigate the premium itself. Below shows the Bayesian premium estimates given different values of  $\theta$ .



**Figure 4.4:** Bayesian premiums at different  $\theta$ : finite mixtures of 3 Gammas

*From Figure 4.4 we see that the Bayesian premiums gradually decreases as the value of  $\theta$  get larger. Investigating Bühlmann premiums we found it shows an almost identical pattern. The figure below presents the difference between those two.*



**Figure 4.5:** Difference in premiums,  $P_{Cr} - P_{Ba}$ : finite mixtures of 3 Gammas

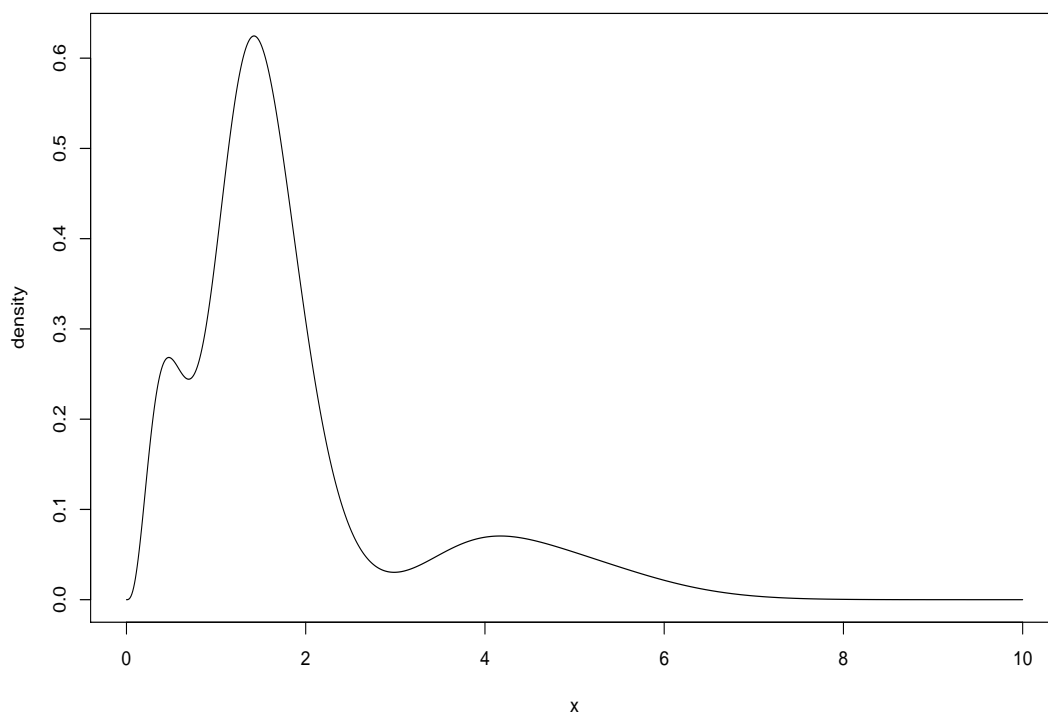
*The result shows the difference between the two premiums are very small at most of the time. We found one occasion that the Bayesian premium is larger by about 1.1. Comparing all 200 cases we found the Bayesian premium is on average larger than the Bühlmann premium by only 0.12. Considering their computational costs the Bühlmann estimator might be a more practical choice for our case.*

**Example 4.4.3.** *In this example, we would like to investigate the possible impact of the prior distribution. We change our conjugate prior to include more Gamma*

densities by assuming:

$$\zeta_l = \begin{cases} \frac{1}{18}, & l = 1; \\ \frac{1}{9}, & l = 2; \\ \frac{1}{3}, & l = 9; \\ \frac{1}{3}, & l = 10; \\ \frac{1}{9}, & l = 30; \\ \frac{1}{18}, & l = 40; \\ 0, & \text{otherwise} \end{cases}$$

The other parameters are assumed to be the same, i.e  $m = 2$  and  $\beta = 8$ . In this case the prior is in the form of a mixture of 6 Gamma distributions. The density plot of this prior distribution is shown below.



**Figure 4.6:** Density plot of prior distribution: finite mixtures of 6 Gammas

*Comparing to Figure 4.1 this prior looks slightly different. The right tail is relatively flat and left tail becomes more smooth.*

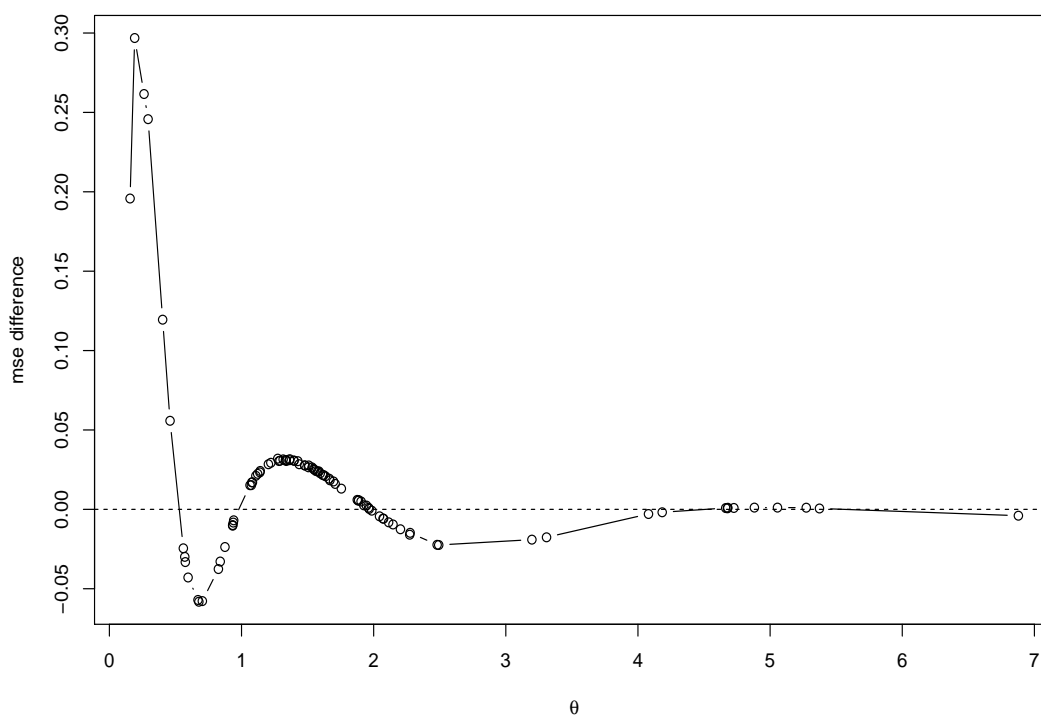
*As in the last example we still assume 30 losses. For Bühlmann premiums, the credibility factor  $Z$  is found to be 0.9521249 and  $\mu$  is equal to 2.54902. Therefore it can be estimated by:*

$$P_{Cr} = 0.9521249\bar{x} + 0.1220346.$$

*The Bayesian premium can be estimated following the similar logic to (4.11)*



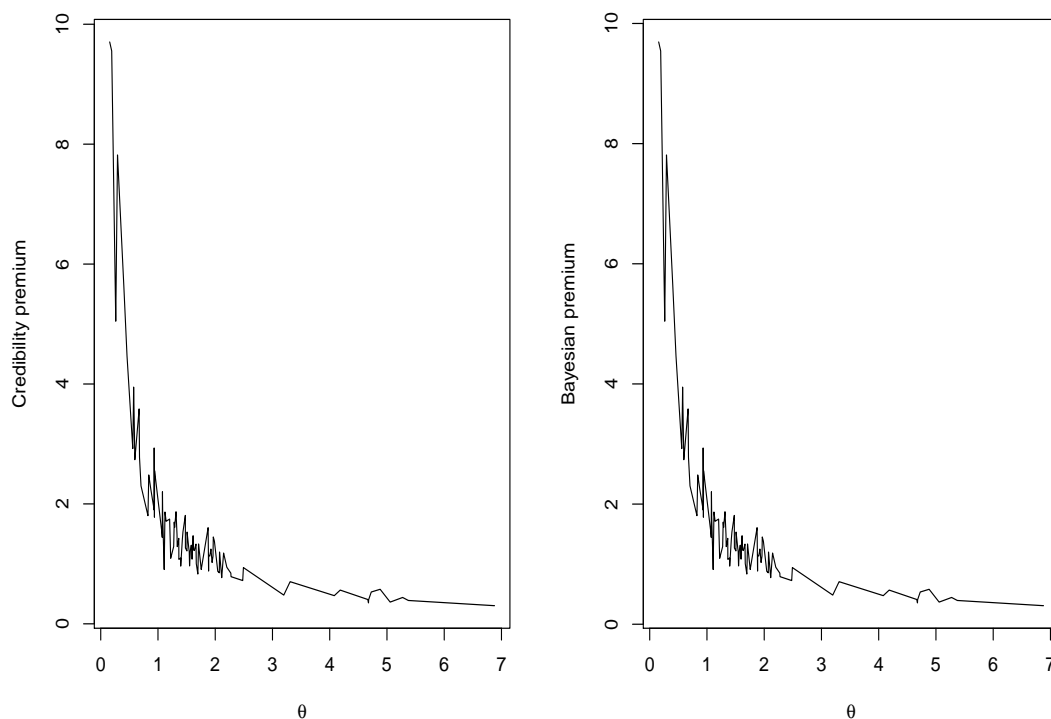
- (4.13). Then we apply the same algorithm as the previous example to compare the MSEs of both estimators at different values of  $\theta$ .



**Figure 4.7:**  $MSE(P_{Cr})$  minus  $MSE(P_{Ba})$ : finite mixtures of 6 Gammas

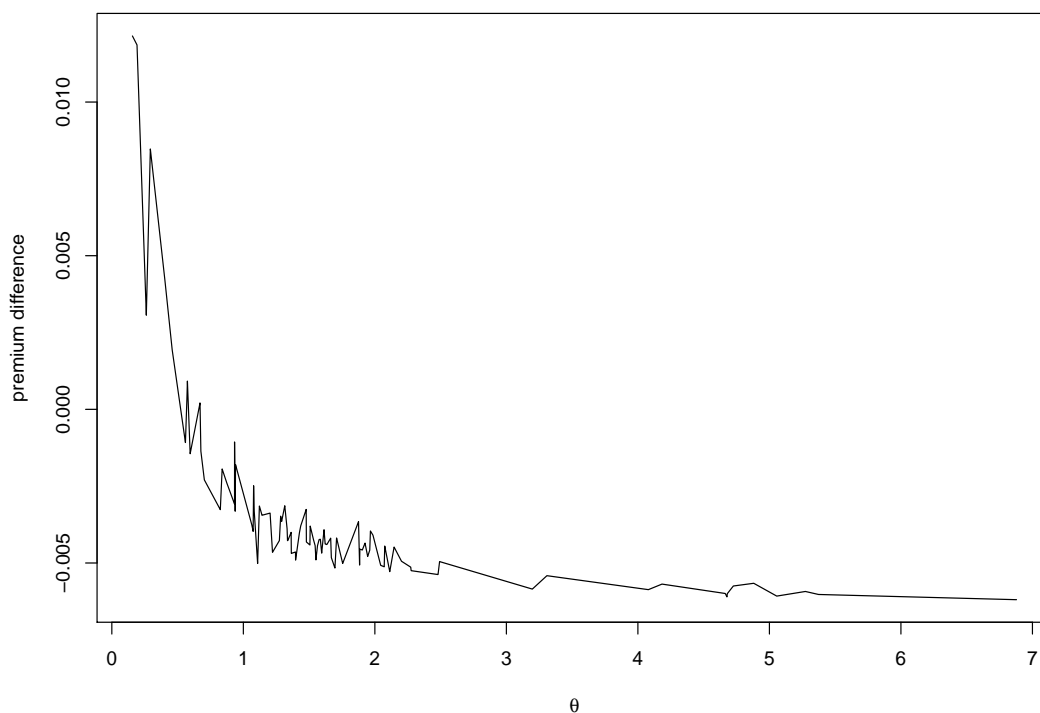
According to Figure 4.6 most of the probability mass is distribution between 1 and 2, which is the exact region that Bayesian premiums show less MSEs. Overall, the average mean squared error for Bühlmann premiums is found to be 0.2699196 and the average MSE for Bayesian premiums is 0.2523452, which is consistent with the theoretical result. Nevertheless, they are quite close to each other indicating very limited advantage from the Bayesian estimator.

Next we compare the premium estimates.



**Figure 4.8:** Bühlmann premiums (left) and Bayesian (right) premiums at different  $\theta$ : finite mixtures of 6 Gammas

*We found that the Bayesian and Bühlmann estimates are almost identical given various value of  $\theta$ . To further investigate their difference, we present the plot of  $P_{Cr} - P_{Ba}$  against  $\theta$  in the graph below.*



**Figure 4.9:** Difference in premiums,  $P_{Cr} - P_{Ba}$ : finite mixtures of 6 Gammas

Comparing to Figure 4.5, we see the difference has become even more negligible. The greatest difference is only about 0.01. On average the Bayesian premium is only larger than Bühlmann premium by 0.003390951. The conjecture of exact credibility seems possible.

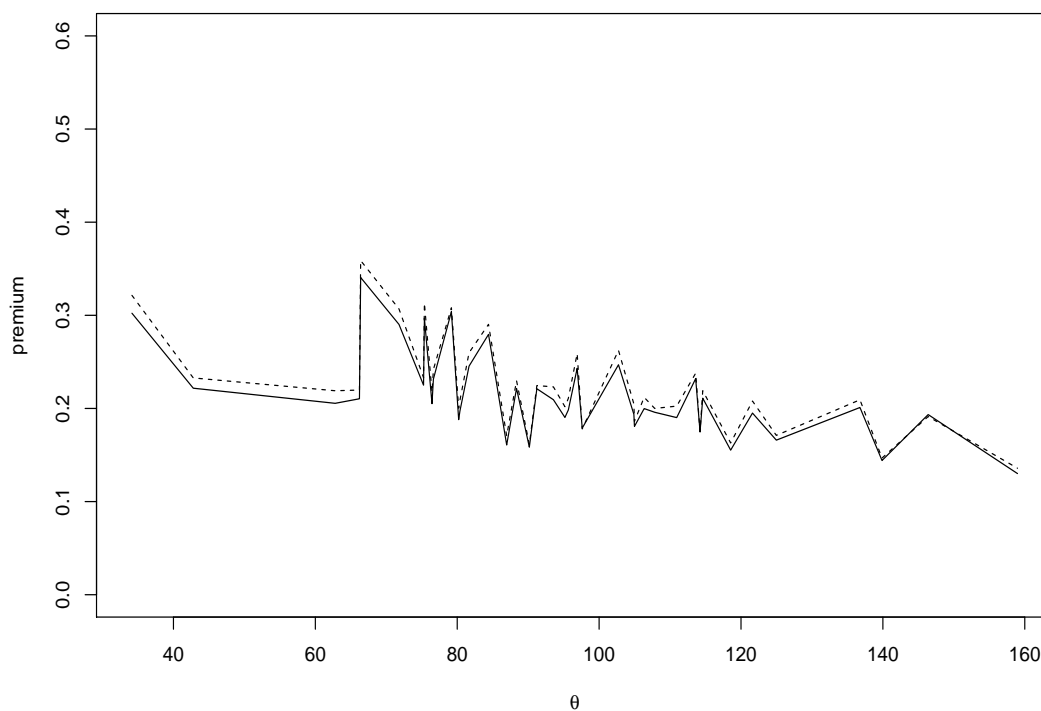
**Example 4.4.4.** In examples 4.4.2 and 4.4.3, we assume the losses  $X_i|\theta \sim PH(\boldsymbol{\alpha}, \theta(P - I))$  with  $P$  being a  $2 \times 2$  matrix. In this example, we extend to the case that  $P$  being a  $3 \times 3$  matrix. To be more specific, we set  $\boldsymbol{\alpha} = (1, 0, 0)'$

and

$$P = \begin{pmatrix} 0.1 & 0.8 & 0.1 \\ 0.8 & 0.1 & 0.0 \\ 0.8 & 0.0 & 0.1 \end{pmatrix}.$$

We assume that a total number of 10 losses are observed. The conjugate prior we are using in this example is a single Gamma distribution with the shape parameter 10 and the rate parameter 0.1. For the Bühlmann premium, the credibility factor can then be calculated as 0.5405 and  $\mu$  to be 0.2222. For the Bayesian premium, we estimate the premium based on the approach used in previous examples, outlined by equations (4.11) - (4.13). However, we are no longer able to implement 10000 repetitions for each  $\theta$  we drawn from the prior. As the dimension of the matrix  $P$  increases, the speed of the convergence for function  $B$  defined in Theorem 4.2 becomes much slower than the  $2 \times 2$  case, leading to a significant increase in the computational cost of Bayesian premium estimation. Due to this reason we are not able to obtain the MSE estimates this time.

Figure 4.10 below shows both estimates at a total number of 45 different values of  $\theta$ .



**Figure 4.10:** Bühlmann (solid line) and Bayesian premiums (dashed line) vs.  $\theta$ : higher dimension PH

*We see that in general, Bayesian premiums are slightly larger than Bühlmann premiums. However, the difference is almost negligible considering the scale of y-axis. Calculating the averages over all premium estimates indicates that on average, Bayesian premium is larger by only 0.009494021. Our finding that the Bühlmann and Bayesian premium yield relatively close values is consistent with what we see from the previous two examples. It is possible that this observation is due to exact credibility.*

*Table 4.1 below shows the details of the 45 different values of  $\theta$  and their corresponding Bühlmann and Bayesian estimates.*

**Table 4.1:** Bühlmann and Bayesian Premiums

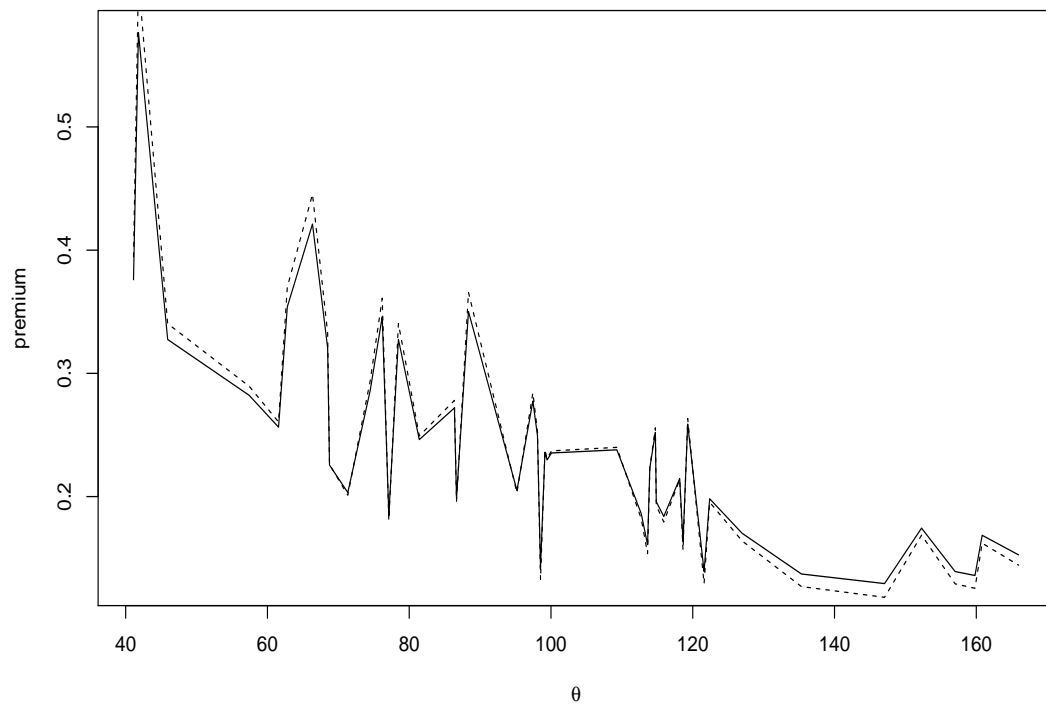
$\theta$	Bühlmann premium	Bayesian premium
34.14869713	0.302236082	0.321404998
42.79435118	0.22174486	0.232775557
62.79605163	0.205496365	0.219002058
66.21883728	0.210393949	0.219750605
66.38865763	0.340599342	0.358810239
71.79211913	0.290199231	0.306828659
75.254049	0.225038727	0.233054627
75.37344014	0.296995508	0.313194236
76.44335718	0.205121471	0.217401828
76.64149874	0.231554936	0.241876828
79.17830488	0.303965402	0.30807168
80.2033774	0.187972165	0.200248442
80.38552405	0.197382249	0.208832867
81.66550168	0.245409197	0.259834256
84.40946205	0.279504545	0.290290129
86.97170852	0.160756745	0.170526562
88.36015902	0.221526611	0.229836998
90.16527715	0.158431359	0.161234834
91.22303124	0.221165798	0.224578943
93.56016068	0.209358413	0.223151974
95.18084294	0.190346468	0.201876015
95.66358591	0.198440525	0.210852743
96.90740593	0.243118101	0.258997664
97.61937869	0.178894147	0.177991448
102.7290665	0.246916946	0.262413317
104.9009097	0.193639087	0.204858129
104.9810904	0.180718995	0.186188564
106.3523824	0.199853468	0.211932295
107.8381893	0.195895189	0.200011474
110.9361579	0.190167546	0.20251323
113.6102975	0.232032506	0.237829989
114.2259155	0.174818388	0.175022853
114.6073869	0.211028513	0.219069725
118.5417648	0.15524825	0.162929053
121.6205557	0.195016711	0.207934545
124.9809513	0.165947348	0.171050802
136.7691784	0.201026129	0.209277404
139.8760172	0.144104739	0.146708442
146.3851159	0.193410778	0.19133235
158.9753741	0.130072604	0.135813872
Mean	0.2133887	0.2228828

Now we change the prior from a single Gamma distribution to a mixture of 3 Gamma distributions by assigning weight according to

$$\zeta_l = \begin{cases} \frac{1}{5}, & l = 0; \\ \frac{3}{5}, & l = 50; \\ \frac{1}{5}, & l = 100; \\ 0, & \text{otherwise.} \end{cases}$$

We keep all the assumptions the same as before and also set  $m = 2$ . For the Bühlmann premium, the credibility factor is found to be 0.8918911 and  $\mu$  is 0.3039216. For the Bayesian premium, we estimate it by equations (4.11) - (4.13) and we are only producing one each estimate for a specific  $\theta$ , for the reason we have explained in the last experiment.

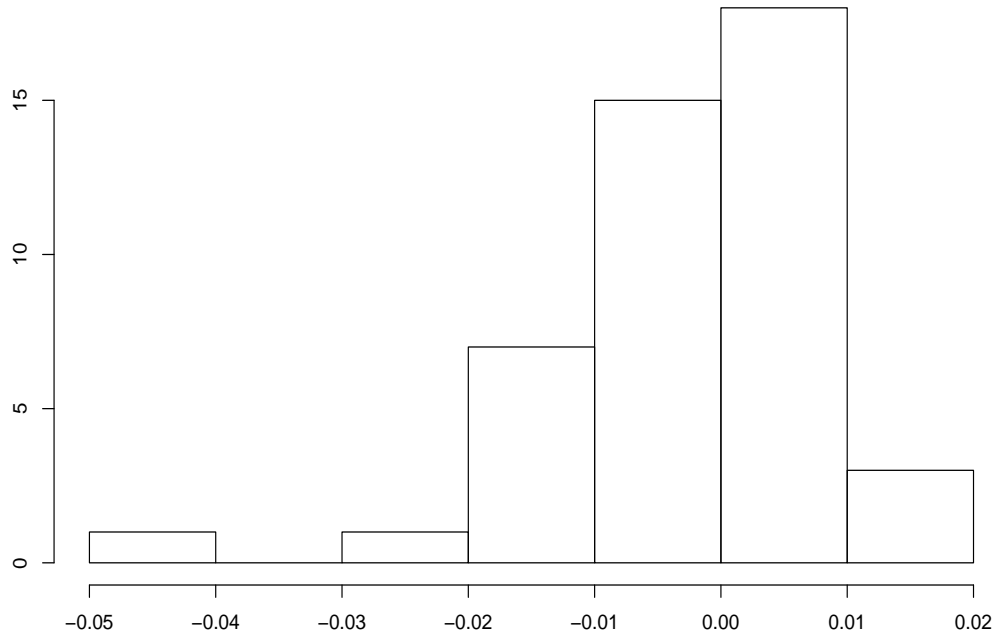
The following graph shows both estimates for a number of 45 different  $\theta$ s.



**Figure 4.11:** Bühlmann (solid line) and Bayesian premiums (dashed line) vs.  $\theta$ : higher dimension PH

*We see that the Bayesian premium does not consistently stay above or below the Bühlmann premium as  $\theta$  changes. To help us understand the difference between both estimators, one may also study a histogram.*





**Figure 4.12:** Histogram for the difference  $P_{Cr} - P_{Ba}$ : higher dimension PH

*The histogram shows the differences are not exactly distributed symmetrically. On a slightly larger proportion of  $\theta$ 's we see a larger Bayesian premium. This fact can also be confirmed by closer examining the data. The average Bayesian premium is found to be just 0.002123647 larger than the average Bühlmann premium.*

*The details around the values of  $\theta$  and respective Bayesian and Bühlmann estimates are given in Table 4.2 below.*

**Table 4.2:** Bühlmann and Bayesian Premiums for Figure 4.11

$\theta$	Bühlmann premium	Bayesian premium
41.10726654	0.375972285	0.394705777
41.80179183	0.57655457	0.619601266
45.92609184	0.327485316	0.340341551
57.4171651	0.282197734	0.28956452
61.56653536	0.256337413	0.260569589
62.80264632	0.353965143	0.37003108
66.3542512	0.421203464	0.445419568
68.49175297	0.320679198	0.332710442
68.75480924	0.225583821	0.226088259
71.33682846	0.20327754	0.201078164
74.49887597	0.28617303	0.294021674
76.19256169	0.345916541	0.361006882
77.12285506	0.184966179	0.180547225
78.47737041	0.327581696	0.340449614
81.42280203	0.246252327	0.249262058
86.37466356	0.271922209	0.278043467
86.67245391	0.19873366	0.195983506
88.35969813	0.350207618	0.365818095
95.19895927	0.20502298	0.203035173
97.44486379	0.277491412	0.28428773
98.09599325	0.249252716	0.252626134
98.52242607	0.142271362	0.132677236
99.14934166	0.235974915	0.237738889
99.45254008	0.229798406	0.230813707
100.0737574	0.235384426	0.237076825
109.2923823	0.237946879	0.239949881
112.7388241	0.186075625	0.18179115
113.6257764	0.160975167	0.153648187
113.9449118	0.223252298	0.223474124
114.7240078	0.252097043	0.255815231
114.8545051	0.195811349	0.192706972
115.9135617	0.183904541	0.179356903
118.1572945	0.214650687	0.213829885
118.6268251	0.163449729	0.156422699
119.2860557	0.258851232	0.263388116
121.6002771	0.139258186	0.129298823
122.3832646	0.198247401	0.195438305
126.9599809	0.170245488	0.164042193
135.3311894	0.13718506	0.126974408
147.05884	0.129404236	0.118250445
152.2790966	0.174410076	0.168711584
157.0170884	0.139161133	0.129190007
159.8129401	0.135966286	0.125607902
160.8287531	0.168598357	0.162195409
165.9622627	0.152701949	0.144372147
Mean	0.238942193	0.24106584

## 4.5 Proofs

### 4.5.1 Proof of Lemma 4.1

**Lemma 4.1** Suppose  $X_1|\Theta, \dots, X_n|\Theta$  are i.i.d. distributed random variables with  $PH(\alpha, T)$ . If  $\Theta$  follows a distribution specified by (3.32), the corresponding posterior distribution is

$$\pi(\theta|x_1, \dots, x_n) = \sum_{s=0}^{\infty} \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} \frac{B(s, x_1, \dots, x_n, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)}, \quad (4.14)$$

where the parameters  $\beta$ ,  $m$  and  $\{\zeta_l\}$  are specified in Proposition 3.3.  $B$  is defined by a recursive relation

$$\begin{aligned} B(s, x_1, \dots, x_n, m, \beta) &= (s+m+n) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum_{i=0}^s q_{i+1} B(s-i, x_1, \dots, x_{n-1}, m, \beta) \\ &\times \binom{s+m+n-1}{i} \left(1 - \frac{x_n}{\beta + \sum_{j=1}^n x_j}\right)^{s-i+m+n-1} \left(\frac{x_n}{\beta + \sum_{j=1}^n x_j}\right)^i, \end{aligned} \quad (4.15)$$

with initial conditions

$$B(s, x_1, m, \beta) = (s+m+1) \frac{\beta}{(\beta + x_1)^2} \sum_{i=0}^s \binom{s+m}{i} \left(\frac{\beta}{\beta + x_1}\right)^{s+m-i} \left(\frac{x_1}{\beta + x_1}\right)^i q_{i+1} B(s-i).$$

and

$$B(l) = \zeta_l.$$

*Proof.* Recall from Corollary 3.2 the joint density for losses is given by:

$$f(x_1, \dots, x_n | \theta) = \theta^n e^{-\theta \sum_{j=1}^n x_j} \sum_{l=0}^{\infty} \frac{(\theta \sum_{j=1}^n x_j)^l}{l!} C(l, x_1, \dots, x_n),$$

where  $C(M, x_1, \dots, x_n) = \sum_{i=0}^M \binom{M}{i} q_{i+1} \left(\frac{x_n}{\sum_{j=1}^n x_j}\right)^i \left(1 - \frac{x_n}{\sum_{j=1}^n x_j}\right)^{M-i} C(M-i, x_1, \dots, x_{n-1})$

for  $M = 0, 1, 2, \dots$  initialized by  $C(l, x_1) = q_{l+1}$ ,  $l = 0, 1, 2, \dots$

The conjugate prior  $\pi(\theta)$  for the joint density has been constructed in the form of infinite mixtures of Erlang distributions, given by Theorem 3.4:

$$\pi(\theta) = \sum_{k=0}^{\infty} \zeta_k \beta e^{-\beta \theta} \frac{(\beta \theta)^{k+m}}{(k+m)!},$$

where  $\sum_{k=0}^{\infty} \zeta_k = 1$ ,  $\beta > 0$  and  $m$  is non-negative.

Thus the density of the posterior distribution  $\pi(\theta | x_1, \dots, x_n)$  can be derived using the Bayesian methodology so that:

$$\pi(\theta | x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n | \theta) \pi(\theta)}{f(x_1, \dots, x_n)}.$$

The numerator on the right hand side is

$$\begin{aligned} & f(x_1, \dots, x_n | \theta) \pi(\theta) \\ &= \left[ \theta^n e^{-\theta \sum_{j=1}^n x_j} \sum_{l=0}^{\infty} \frac{(\theta \sum_{j=1}^n x_j)^l}{l!} C(l, x_1, \dots, x_n) \right] \left[ \sum_{k=0}^{\infty} \zeta_k \beta e^{-\beta \theta} \frac{(\beta \theta)^{k+m}}{(k+m)!} \right] \\ &= \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{l+k+m+n} \beta^{k+m+1} \left( \sum_{j=1}^n x_j \right)^l \frac{\zeta_k C(l, x_1, \dots, x_n)}{l!(k+m)!} \end{aligned}$$

$$\begin{aligned}
&= \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{l+k+m+n}}{(l+k+m+n)!} (\beta + \sum_{j=1}^n x_j)^{l+k+m+n+1} \\
&\times \frac{(l+k+m+n)!}{l!(k+m)!} \frac{\beta^{k+m+1} (\sum_{j=1}^n x_j)^l}{(\beta + \sum_{j=1}^n x_j)^{l+k+m+n+1}} \zeta_k C(l, x_1, \dots, x_n)
\end{aligned}$$

Letting  $s = l + k$ ,

$$\begin{aligned}
&= \sum_{s=0}^{\infty} \sum_{l=0}^s \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} \\
&\times \frac{(s+m+n)!}{l!(s-l+m)!} \frac{\beta^{s-l+m+1} (\sum_{j=1}^n x_j)^l}{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}} \zeta_{s-l} C(l, x_1, \dots, x_n) \\
&= \sum_{s=0}^{\infty} \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} B(s, x_1, \dots, x_n, m, \beta),
\end{aligned}$$

where  $B(s, x_1, \dots, x_n, m, \beta) = \sum_{l=0}^s \frac{(s+m+n)!}{l!(s-l+m)!} \frac{\beta^{s-l+m+1} (\sum_{j=1}^n x_j)^l}{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}} \zeta_{s-l} C(l, x_1, \dots, x_n)$ .

Similarly, for the denominator part we have

$$\begin{aligned}
f(x_1, \dots, x_n) &= \int_0^{\infty} f(x_1, \dots, x_n | \theta) \pi(\theta) d\theta \\
&= \int_0^{\infty} \sum_{s=0}^{\infty} \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} B(s, x_1, \dots, x_n, m, \beta) d\theta \\
&= \sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta) \int_0^{\infty} \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} d\theta \\
&= \sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta).
\end{aligned}$$

The last equality is due to the fact that the integrand can be viewed as a Gamma density.

From the above results, we have the posterior

$$\pi(\theta|x_1, \dots, x_n) = \sum_{s=0}^{\infty} \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} \frac{B(s, x_1, \dots, x_n, m, \beta)}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)},$$

which shows that the posterior is another infinite mixture of Erlang distributions and thus follows the same distribution as the prior, but with different parameters.

We realize that the function  $B(s, x_1, \dots, x_n, m, \beta)$  is expressed in terms of the coefficient  $C(l, x_1, \dots, x_n)$ . We next establish an equivalent recursive formula for  $B(s, x_1, \dots, x_n, m, \beta)$  which is independent of the  $C$  coefficients. Obtaining such representation allows us to develop a more efficient algorithm updating the premium based on the new  $(n+1)$ th observation without resorting to the recalculation of all  $C$  coefficients from the beginning. We know the original specification of  $B$  is

$$B(s, x_1, \dots, x_n, m, \beta) = \sum_{l=0}^s \frac{(s+m+n)!}{l!(s-l+m)!} \frac{\beta^{s-l+m+1} (\sum_{j=1}^n x_j)^l}{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}} \zeta_{s-l} C(l, x_1, \dots, x_n).$$

By expanding  $C(l, x_1, \dots, x_n)$  we have

$$\begin{aligned} & B(s, x_1, \dots, x_n, m, \beta) \\ &= \sum_{l=0}^s \frac{(s+m+n)!}{l!(s-l+m)!} \frac{\beta^{s-l+m+1} (\sum_{j=1}^n x_j)^l}{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}} \zeta_{s-l} \\ & \quad \times \sum_{i=0}^l \binom{l}{i} q_{i+1} \left( \frac{x_n}{\sum_{j=1}^n x_j} \right)^i \left( 1 - \frac{x_n}{\sum_{j=1}^n x_j} \right)^{l-i} C(l-i, x_1, \dots, x_{n-1}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=0}^s \sum_{k=0}^{s-i} \frac{(s+m+n)!}{(k+i)!(s-k-i+m)!} \frac{\beta^{s-k-i+m+1}}{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}} \zeta_{s-k-i} \\
&\quad \times \frac{(k+i)!}{k!i!} q_{i+1} (x_n)^i \left( \sum_{j=1}^{n-1} x_j \right)^k C(k, x_1, \dots, x_{n-1}) \\
&= \sum_{i=0}^s \frac{(s+m+n)!}{i!(s-i+m+n-1)!} q_{i+1} \frac{(\beta + \sum_{j=1}^{n-1} x_j)^{s-i+m+n}}{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}} (x_n)^i \\
&\quad \times \sum_{k=0}^{s-i} \frac{(s-i+m+n-1)!}{k!(s-i-k+m)!} \zeta_{s-i-k} \frac{\beta^{s-i-k+m+1} (\sum_{j=1}^{n-1} x_j)^k}{(\beta + \sum_{j=1}^{n-1} x_j)^{s-i+m+n}} C(k, x_1, \dots, x_{n-1}) \\
&= \sum_{i=0}^s \frac{(s+m+n)!}{i!(s-i+m+n-1)!} q_{i+1} \frac{(\beta + \sum_{j=1}^{n-1} x_j)^{s-i+m+n}}{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}} (x_n)^i \\
&\quad \times B(s-i, x_1, \dots, x_{n-1}, m, \beta) \\
&= \frac{s+m+n}{\beta + \sum_{j=1}^n x_j} \sum_{i=0}^s \frac{(s+m+n-1)!}{i!(s-i+m+n-1)!} q_{i+1} \left( \frac{\beta + \sum_{j=1}^{n-1} x_j}{\beta + \sum_{j=1}^n x_j} \right)^{s-i+m+n} \\
&\quad \times \left( \frac{x_n}{\beta + \sum_{j=1}^n x_j} \right)^i B(s-i, x_1, \dots, x_{n-1}, m, \beta) \\
&= (s+m+n) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum_{i=0}^s \binom{s+m+n-1}{i} q_{i+1} \\
&\quad \times \left( 1 - \frac{x_n}{\beta + \sum_{j=1}^n x_j} \right)^{s-i+m+n-1} \left( \frac{x_n}{\beta + \sum_{j=1}^n x_j} \right)^i B(s-i, x_1, \dots, x_{n-1}, m, \beta).
\end{aligned}$$

Note that the second equality is obtained by setting  $k = l - i$ .

The above result shows the formula to calculate function  $B$  recursively. The initial condition can also be easily obtained and it is given by:

$$B(s, x_1, m, \beta) = (s+m+1) \frac{\beta}{(\beta + x_1)^2} \sum_{i=0}^s \binom{s+m}{i} q_{i+1} \left( \frac{\beta}{\beta + x_1} \right)^{s+m-i} \left( \frac{x_1}{\beta + x_1} \right)^i \zeta_{s-i}.$$

Lastly we need to prove the convergence of  $\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)$  so that

the posterior density is well defined. From the above result we know

$$\begin{aligned}
& B(s, x_1, \dots, x_n, m, \beta) \\
&= (s + m + n) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum_{i_1=0}^s \binom{s + m + n - 1}{i_1} q_{i_1+1} \\
&\times \left(1 - \frac{x_n}{\beta + \sum_{j=1}^n x_j}\right)^{s-i_1+m+n-1} \left(\frac{x_n}{\beta + \sum_{j=1}^n x_j}\right)^{i_1} B(s - i_1, x_1, \dots, x_{n-1}, m, \beta) \\
&\leq (s + m + n) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum_{i_1=0}^s q_{i_1+1} B(s - i_1, x_1, \dots, x_{n-1}, m, \beta) \\
&\leq (s + m + n)(s + m + n - 1) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum_{i_1=0}^s q_{i_1+1} \sum_{i_2=0}^{s-i_1} q_{i_2+1} B(s - i_1 - i_2, x_1, \dots, x_{n-2}, m, \beta) \\
&\vdots \\
&\leq (s + m + n) \cdots (s + m + 1) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum_{i_1=0}^s q_{i_1+1} \sum_{i_2=0}^{s-i_1} q_{i_2+1} \cdots \sum_{i_n=0}^{s-i_1-\cdots-i_{n-1}} q_{i_n+1} \\
&= (s + m + n) \cdots (s + m + 1) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum \cdots \sum_{i_1+\cdots+i_n=s} q_{i_1+1} q_{i_2+1} \cdots q_{i_n+1} \\
&= (s + m + n) \cdots (s + m + 1) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} Pr(N_1 + \cdots + N_n = s) \\
&= (s + m + n) \cdots (s + m + 1) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} Pr(N = s),
\end{aligned}$$

where  $N_k$ ,  $k = 1, \dots, n$  are independent discrete phase-type distributions with p.m.f  $q_{i_k+1}$  and  $N = \sum_{k=1}^n N_k$ . From the above result we have:

$$\begin{aligned}
\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta) &\leq \sum_{s=0}^{\infty} (s + m + n) \cdots (s + m + 1) \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} Pr(N = s) \\
&= \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} \sum_{s=0}^{\infty} (s + m + n) \cdots (s + m + 1) Pr(N = s)
\end{aligned}$$



$$= \frac{\beta + \sum_{j=1}^{n-1} x_j}{(\beta + \sum_{j=1}^n x_j)^2} E\left(\prod_{k=1}^n (N + m + k)\right).$$

Since the summation of independent discrete phase-type distributions is still a discrete phase-type distribution, we know  $N$  follows a discrete phase-type distribution. From Corollary 2.5 we know the factorial moment  $E(\prod_{k=1}^n (N + m + k))$  is finite. Hence,  $\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)$  is convergent.  $\square$

#### 4.5.2 Proofs of Theorem 4.2

**Theorem 4.2** Suppose  $X_1|\Theta, \dots, X_n|\Theta$  are i.i.d. distributed random variables with  $PH(\boldsymbol{\alpha}, T)$ . If  $\Theta$  follows a distribution specified by (3.32), the Bayesian premium estimator is

$$E(X_{n+1}|X_1, \dots, X_n) = E(N) \left( \beta + \sum_{j=1}^n x_j \right) \frac{\sum_{s=0}^{\infty} \frac{1}{s+m+n} B(s, x_1, \dots, x_n, m, \beta)}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)}, \quad (4.16)$$

where  $N$  is defined by (3.19), parameters  $m$  and  $\beta$  follows the same definitions from Theorem 3.4 and  $B(s, x_1, \dots, x_n, m, \beta)$  follows the same definition from Lemma 4.1.

*Proof.* Here we will show two different approaches to derive the Bayesian premium. The simpler approach proceeds as following:

$$E(X_{n+1}|X_1, \dots, X_n) = \int \mu(\theta) \pi(\theta|x_1, \dots, x_n) d\theta$$

$$\begin{aligned}
&= \int_0^\infty E(N)\theta^{-1} \sum_{s=0}^\infty \frac{e^{-\theta(\beta+\sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} \\
&\quad \times \frac{B(s, x_1, \dots, x_n, m, \beta)}{\sum_{s=0}^\infty B(s, x_1, \dots, x_n, m, \beta)} d\theta \\
&= E(N) (\beta + \sum_{j=1}^n x_j) \frac{\sum_{s=0}^\infty \frac{1}{s+m+n} B(s, x_1, \dots, x_n, \beta)}{\sum_{s=0}^\infty B(s, x_1, \dots, x_n)}.
\end{aligned}$$

This result can be also verified following an alternative approach outlined below.

$$\begin{aligned}
E(X_{n+1}|X_1, \dots, X_n) &= \int_0^\infty x_{n+1} f(x_{n+1}|x_1, \dots, x_n) dx_{n+1} \\
&= \int_0^\infty x_{n+1} \frac{f(x_1, \dots, x_{n+1})}{f(x_1, \dots, x_n)} dx_{n+1} \tag{4.17}
\end{aligned}$$

We have already known that  $f(x_1, \dots, x_n) = \sum_{s=0}^\infty B(s, x_1, \dots, x_n, m, \beta)$ . The substitution of the numerator and denominator of (4.17) yields

$$\begin{aligned}
E(X_{n+1}|X_1, \dots, X_n) &= \int_0^\infty x_{n+1} \frac{\sum_{s=0}^\infty B(s, x_1, \dots, x_{n+1}, m, \beta)}{\sum_{s=0}^\infty B(s, x_1, \dots, x_n, m, \beta)} dx_{n+1} \\
&= \frac{1}{\sum_{s=0}^\infty B(s, x_1, \dots, x_n, m, \beta)} \int_0^\infty x_{n+1} \sum_{s=0}^\infty B(s, x_1, \dots, x_{n+1}, m, \beta) dx_{n+1} \\
&= \frac{1}{\sum_{s=0}^\infty B(s, x_1, \dots, x_n, m, \beta)} \int_0^\infty x_{n+1} \sum_{s=0}^\infty (s+m+n+1) \frac{\beta + \sum_{j=1}^n x_j}{(\beta + \sum_{j=1}^{n+1} x_j)^2} \\
&\quad \times \sum_{i=0}^s \binom{s+m+n}{i} q_{i+1} \left(1 - \frac{x_n}{\beta + \sum_{j=1}^{n+1} x_j}\right)^{s-i+m+n} \left(\frac{x_{n+1}}{\beta + \sum_{j=1}^{n+1} x_j}\right)^i \\
&\quad \times B(s-i, x_1, \dots, x_n, m, \beta).
\end{aligned}$$

After simplification the numerator of the above formula becomes

$$\begin{aligned}
& \sum_{s=0}^{\infty} \sum_{i=0}^s \frac{(s+m+n+1)!}{i!(s+m+n-i)!} q_{i+1} B(s-i, x_1, \dots, x_n, m, \beta) \\
& \times \int_0^{\infty} x_{n+1} \frac{(\beta + \sum_{j=1}^n x_j)^{s-i+m+n+1} x_{n+1}^i}{(\beta + \sum_{j=1}^{n+1} x_j)^{s+m+n+2}} dx_{n+1} \\
& = \sum_{s=0}^{\infty} \sum_{i=0}^s \frac{(s+m+n+1)!}{i!(s+m+n-i)!} q_{i+1} B(s-i, x_1, \dots, x_n, m, \beta) (\beta + \sum_{j=1}^n x_j)^{s-i+m+n+1} \\
& \times \int_0^{\infty} \frac{x_{n+1}^{i+1}}{(\beta + \sum_{j=1}^{n+1} x_j)^{s+m+n+2}} dx_{n+1}. \tag{4.18}
\end{aligned}$$

To solve the integral part, we may assume that  $x = x_{n+1}$ ,  $a = i + 1$ ,  $b = s + m + n + 2$ , i.e  $b > a$ , and  $c = \beta + \sum_{j=1}^n x_j$  to make the calculation more convenient. So

$$\begin{aligned}
& \int_0^{\infty} \frac{x_{n+1}^{i+1}}{(\beta + \sum_{j=1}^{n+1} x_j)^{s+m+n+2}} dx_{n+1} \\
& = \int_0^{\infty} \frac{x^a}{(c+x)^b} dx \\
& = \frac{1}{-b+1} x^a (x+c)^{-b+1} \Big|_0^{\infty} - \int_0^{\infty} \frac{a}{-b+1} x^{a-1} (x+c)^{-b+1} dx \\
& = \int_0^{\infty} \frac{a}{b-1} x^{a-1} (x+c)^{-b+1} dx \\
& = \int_0^{\infty} \frac{a(a-1)}{(b-1)(b-2)} x^{a-2} (x+c)^{-b+2} dx \\
& \vdots \\
& = \int_0^{\infty} \frac{a(a-1)\cdots 1}{(b-1)(b-2)\cdots(b-a)} (x+c)^{-b+a} dx \\
& = \frac{a!}{(b-1)(b-2)\cdots(b-a)} \int_0^{\infty} (x+c)^{-b+a} dx
\end{aligned}$$

$$\begin{aligned}
&= \frac{a!}{(b-1)(b-2)\cdots(b-a-1)} c^{-b+a+1} \\
&= \frac{(i+1)!}{(s+m+n+1)\cdots(s-i+m+n)} \frac{1}{(\beta + \sum_{j=1}^n x_j)^{s-i+m+n}}.
\end{aligned}$$

Substituting the above result back into (4.18) we have

$$\begin{aligned}
&E(X_{n+1}|X_1, \dots, X_n) \\
&= \frac{1}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)} \\
&\times \sum_{s=0}^{\infty} \sum_{i=0}^s \frac{(s+m+n+1)!}{i!(s+m+n-i)!} q_{i+1} B(s-i, x_1, \dots, x_n, m, \beta) (\beta + \sum_{j=1}^n x_j)^{s-i+m+n+1} \\
&\times \frac{(i+1)!}{(s+m+n+1)\cdots(s-i+m+n)} \frac{1}{(\beta + \sum_{j=1}^n x_j)^{s-i+m+n}} \\
&= \frac{\sum_{s=0}^{\infty} \sum_{i=0}^s (i+1) q_{i+1} B(s-i, x_1, \dots, x_n, m, \beta) \frac{1}{s-i+m+n} (\beta + \sum_{j=1}^n x_j)}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)} \\
&= \frac{\sum_{i=0}^{\infty} \sum_{s=i}^{\infty} (i+1) q_{i+1} B(s-i, x_1, \dots, x_n, m, \beta) \frac{1}{s-i+m+n} (\beta + \sum_{j=1}^n x_j)}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)} \\
&= \frac{\sum_{i=0}^{\infty} (i+1) q_{i+1} \sum_{s=i}^{\infty} B(s-i, x_1, \dots, x_n, m, \beta) \frac{1}{s-i+m+n} (\beta + \sum_{j=1}^n x_j)}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)} \\
&= E(N) (\beta + \sum_{j=1}^n x_j) \frac{\sum_{s=0}^{\infty} \frac{1}{s+m+n} B(s, x_1, \dots, x_n, \beta)}{\sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)}.
\end{aligned}$$

We have showed that the Bayesian premium is consistent through different methodologies.  $\square$

### 4.5.3 Proof of Corollary 4.3

**Corollary 4.3** Suppose a loss  $X$  has a likelihood  $X|\Theta \sim PH(\alpha, \theta(P-I))$ . If  $\Theta$  follows a distribution specified by (3.32) with  $m=0$  and  $\{\zeta_i\}$  being a probability

measure representing a Geometric distribution, i.e.

$$\zeta_l = (1 - p)^l p, \quad l \in \mathbb{Z}^+ \cup \{0\},$$

where  $0 < p < 1$ . Then the marginal distribution of  $X$  is

$$f(x) = \sum_{l=0}^{\infty} q_{l+1} \frac{(l+1) \cdot \beta p \cdot x^l}{(x + \beta p)^{l+2}}, \quad (4.19)$$

where  $q_{l+1}$  is defined in Proposition 3.1.

*Proof.* When there is only one observation available and the parameter  $m$  equal to 0 the  $B$  function can be expressed as

$$B(s, x, m = 0, \beta) = \sum_{l=0}^s \frac{(s+1)!}{l!(s-l)!} \frac{\beta^{s+1-l} x^l}{(\beta+x)^{s+2}} \zeta_{s-l} q_{l+1}.$$

We also know that the joint density function of  $X_1, \dots, X_n$  is  $f(x_1, \dots, x_n) = \sum_{s=0}^{\infty} B(s, x_1, \dots, x_n, m, \beta)$  from the previous section. Then,

$$\begin{aligned} f(x) &= \sum_{s=0}^{\infty} B(s, x, 0, \beta) \\ &= \sum_{s=0}^{\infty} \sum_{l=0}^s \frac{(s+1)!}{l!(s-l)!} \frac{\beta^{s+1-l} x^l}{(\beta+x)^{s+2}} \zeta_{s-l} q_{l+1} \\ &= \frac{\beta}{(\beta+x)^2} \sum_{s=0}^{\infty} \sum_{l=0}^s \frac{(s+1)!}{l!(s-l)!} \frac{\beta^{s-l} x^l}{(\beta+x)^s} \zeta_{s-l} q_{l+1} \\ &= \frac{\beta}{(\beta+x)^2} \sum_{l=0}^{\infty} \sum_{s=l}^{\infty} \frac{(s+1)!}{l!(s-l)!} \frac{\beta^{s-l} x^l}{(\beta+x)^s} \zeta_{s-l} q_{l+1} \end{aligned}$$

$$= \frac{\beta}{(\beta+x)^2} \sum_{l=0}^{\infty} q_{l+1} \sum_{k=0}^{\infty} \frac{(k+l+1)!}{l!k!} \frac{\beta^k x^l}{(\beta+x)^{k+l}} \zeta_k$$

By considering the case when  $\zeta_k$  follows a Geometric distribution so that  $\zeta_k = (1-p)^k p$ ,  $k = 0, 1, 2, 3, \dots$ , we have

$$\begin{aligned} f(x) &= \frac{\beta}{(\beta+x)^2} \sum_{l=0}^{\infty} q_{l+1} \sum_{k=0}^{\infty} \frac{(k+l+1)!}{l!k!} \frac{\beta^k x^l}{(\beta+x)^{k+l}} (1-p)^k p \\ &= \beta p \sum_{l=0}^{\infty} q_{l+1} \sum_{k=0}^{\infty} (1-p)^k \underbrace{\frac{(k+l+1)!}{(l+1)!k!} \frac{\beta^k x^{l+2}}{(\beta+x)^{k+l+2}}}_{NB(r=l+2, p=\frac{\beta}{\beta+x})} \frac{(l+1)}{x^2} \\ &= \beta p \sum_{l=0}^{\infty} q_{l+1} \frac{(l+1)}{x^2} \sum_{k=0}^{\infty} \underbrace{(1-p)^k f_{NB}(k)}_{\text{p.g.f of NB, } p(z)|_{z=1-p}} \\ &= \beta p \sum_{l=0}^{\infty} q_{l+1} \frac{(l+1)}{x^2} \left( \frac{1 - \frac{\beta}{\beta+x}}{1 - \frac{\beta}{\beta+x}(1-p)} \right)^{l+2} \\ &= \beta p \sum_{l=0}^{\infty} q_{l+1} \frac{(l+1)}{x^2} \left( \frac{\beta+x-\beta}{\beta+x-\beta(1-p)} \right)^{l+2} \\ &= \beta p \sum_{l=0}^{\infty} q_{l+1} \frac{(l+1)}{x^2} \left( \frac{x}{x+\beta p} \right)^{l+2} \\ &= \beta p \sum_{l=0}^{\infty} q_{l+1} (l+1) \frac{x^l}{(x+\beta p)^{l+2}} \\ &= \sum_{l=0}^{\infty} q_{l+1} \frac{(l+1) \cdot \beta p \cdot x^l}{(x+\beta p)^{l+2}} \end{aligned}$$

□

#### 4.5.4 Proofs of Corollary 4.4

**Corollary 4.4** Suppose a loss  $X$  has a likelihood  $X|\Theta \sim PH(\boldsymbol{\alpha}, \theta(P - I))$ . If the posterior  $\Theta|X_1, \dots, X_n$  follows a distribution specified in Lemma 4.1, then the density function of the predictive distribution is given by

$$f(x_{n+1}|x_1, \dots, x_n) = \frac{\sum_{k=0}^{\infty} B(k, x_1, \dots, x_{n+1}, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)}, \quad (4.20)$$

where the function  $B(k, x_1, \dots, x_n, m, \beta)$  is also specified in Lemma 4.1.

*Proof.* We have that

$$\begin{aligned} & f(x_{n+1}|x_1, \dots, x_n) \\ &= \int_0^{\infty} f(x_{n+1}|\theta)\pi(\theta|x_1, \dots, x_n)d\theta \\ &= \int_0^{\infty} \sum_{l=0}^{\infty} q_{l+1}\theta e^{-\theta x_{n+1}} \frac{(\theta x_{n+1})^l}{l!} \sum_{s=0}^{\infty} \frac{e^{-\theta(\beta + \sum_{j=1}^n x_j)} \theta^{s+m+n}}{(s+m+n)!} (\beta + \sum_{j=1}^n x_j)^{s+m+n+1} \\ &\quad \times \frac{B(s, x_1, \dots, x_n, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)} d\theta \\ &= \int_0^{\infty} \sum_{l=0}^{\infty} \sum_{s=0}^{\infty} q_{l+1} \theta^{s+l+m+n+1} e^{-\theta(\beta + \sum_{j=1}^{n+1} x_j)} \frac{(x_{n+1})^l}{l!} \frac{(\beta + \sum_{j=1}^n x_j)^{s+m+n+1}}{(s+m+n)!} \\ &\quad \times \frac{B(s, x_1, \dots, x_n, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)} d\theta \\ &= \int_0^{\infty} \sum_{l=0}^{\infty} \sum_{s=0}^{\infty} (\beta + \sum_{j=1}^{n+1} x_j) e^{-\theta(\beta + \sum_{j=1}^{n+1} x_j)} \frac{[(\beta + \sum_{j=1}^{n+1} x_j)\theta]^{s+l+m+n+1}}{(s+l+m+n+1)!} \frac{(s+l+m+n+1)!}{(s+m+n)!l!} q_{l+1} \\ &\quad \times \left( \frac{x_{n+1}}{\beta + \sum_{j=1}^{n+1} x_j} \right)^l \left( \frac{\beta + \sum_{j=1}^n x_j}{\beta + \sum_{j=1}^{n+1} x_j} \right)^{s+m+n+1} \frac{1}{\beta + \sum_{j=1}^{n+1} x_j} \frac{B(s, x_1, \dots, x_n, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)} d\theta \\ &= \sum_{l=0}^{\infty} \sum_{s=0}^{\infty} q_{l+1} \frac{(s+l+m+n+1)!}{(s+m+n)!l!} \left( \frac{x_{n+1}}{\beta + \sum_{j=1}^{n+1} x_j} \right)^l \left( \frac{\beta + \sum_{j=1}^n x_j}{\beta + \sum_{j=1}^{n+1} x_j} \right)^{s+m+n+1} \frac{1}{\beta + \sum_{j=1}^{n+1} x_j} \end{aligned}$$

$$\begin{aligned}
& \times \frac{B(s, x_1, \dots, x_n, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)} \\
& = \sum_{k=0}^{\infty} \sum_{l=0}^k q_{l+1} \frac{\beta + \sum_{j=1}^n x_j}{(\beta + \sum_{j=1}^{n+1} x_j)^2} (k + m + n + 1) \frac{(k + m + n)!}{l!(k - l + m + n)!} \left( \frac{x_{n+1}}{\beta + \sum_{j=1}^{n+1} x_j} \right)^l \\
& \times \left( \frac{\beta + \sum_{j=1}^n x_j}{\beta + \sum_{j=1}^{n+1} x_j} \right)^{k-l+m+n} \frac{B(k - l, x_1, \dots, x_n, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)} \\
& = \frac{1}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)} \sum_{k=0}^{\infty} (k + m + n + 1) \frac{\beta + \sum_{j=1}^n x_j}{(\beta + \sum_{j=1}^{n+1} x_j)^2} \sum_{l=0}^k \binom{k + m + n}{l} \\
& \times \left( \frac{x_{n+1}}{\beta + \sum_{j=1}^{n+1} x_j} \right)^l \left( \frac{\beta + \sum_{j=1}^n x_j}{\beta + \sum_{j=1}^{n+1} x_j} \right)^{k-l+m+n} B(k - l, x_1, \dots, x_n, m, \beta) \\
& = \frac{\sum_{k=0}^{\infty} B(k, x_1, \dots, x_{n+1}, m, \beta)}{\sum_{t=0}^{\infty} B(t, x_1, \dots, x_n, m, \beta)}.
\end{aligned}$$

The last step is achieved by applying the recursive relation specified by Equation 4.15 in Lemma 4.1. □



## Chapter 5

# Conclusions and Future Work

In this thesis, we derived original credibility models under the assumption that individual losses depending on his risk parameter value  $\theta$  follows a phase-type distribution. Since the risk parameter  $\theta$  for a policyholder is never known, we constructed premium estimators following Bayesian inference techniques. By imposing a prior distribution on  $\Theta$ , we are able to probabilistically describe the risk structure for the entire rating class. In practice, the choice of this prior distribution is subjective to personal judgements or induced from historical data of the corresponding group.

The focus of Chapter 3 is to determine Bühlmann credibility when losses follows a phase-type distribution. The problem is well-understood for linear exponential family, wherein exact credibility occurs. We are interested to see if it's still the case under the phase-type text. In section 3.2, we constructed a family of conjugate priors for phase-type distributions. The prior was found to be in

the form of a possibly infinite mixture of Erlang distributions. One motive for using this distribution is to investigate the exact credibility property which is well known for linear exponential family with corresponding conjugate priors. Lastly, the explicit form of the Bühlmann premium estimator is obtained based on this prior.

The Bayesian premium was treated in Chapter 4 based on the same distributional assumptions. The solution obtained was much more complicated than the Bühlmann estimator and involved certain infinite series with the general term expressed by a recursive relation. We were able to show that the Bayesian estimator was well defined but unfortunately, this infinite series portion was very difficult to evaluate analytically due to its algebraic complexity. Therefore a numeric algorithm was designed to approximate this quantity and simulation studies were also performed for both Bayesian and Bühlmann estimators.

One important experiment was to compare the squared error losses yielded by both estimators: we want to know how much difference there is between the accuracy of the Bühlmann case and the Bayesian case. Theoretically the Bayesian estimator should yield lower squared error loss than the Bühlmann one. We performed a series of comparisons based on conjugate priors comprising different numbers of Erlang distributions. Our findings are consistent with the theory, although the differences between the differences in the squared error losses of the Bayesian and Bühlmann premiums were often found to be small. We'd also like to point out that our squared error losses were not obtained by algebraic

---

means, but rather estimated based on a number of simulated losses depending on different values of  $\theta$ . By the law of large numbers, such estimation improves as more repeated samplings of  $\theta$  and losses are involved. However, our numeric algorithm for Bayesian premium estimation is quite time consuming, which makes the computational cost for large scale experimentation fairly high. For instance, it can take more than 2 weeks for one machine to complete the computation of Bayesian premiums for 200 different  $\theta$ s. To reduce the running time, we leveraged high performance computing platforms with MPI parallel computing package, by which most of our experiments were completed within 5 days.

To continue this work we list a few directions to pursue. One interesting problem would be to investigate if Corollary 4.3, which studies the marginal distribution of one loss under a specific prior/likelihood setting, can be extended by considering the marginal joint distribution of multiple losses with more relaxed assumption on  $\{\zeta_i\}$ . We can also revisit the Bayesian premium problem but from a simulation based approach. If we can obtain reliable samplings from the conditional density of future loss  $f(x_{n+1}|x_1, \dots, x_n)$  via some algorithms, for instance MCMC, a Bayesian premium approximation can be easily obtained. This may offer better computational efficiency than the current algorithm introduced in section 4.4. We also wish to dedicate more effort to better understanding the algebraic structure of the infinite series in equation (4.16). The function denoted  $B$  incorporated quite complex structures and our algorithm determines its status of convergence by considering the sequential difference of two consecutive terms.

We are wondering if there is another termination condition we can apply in terms of determining how many terms to calculate. Lastly we could investigate further in an even more important model, the Bühlmann-Straub model, which is a more generalized model of the Bühlmann premium estimator.

# Bibliography

- Asmussen, S. (2000). *Ruin probabilities*, Volume 2 of *Advance series on statistical science & applied probability*. River Edge, N.J.: World Scientific. 7
- Asmussen, S., O. Nerman, and M. Olsson (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics* 23, 419–441. 6
- Bailey, A. (1945). An generalized theory of credibility. *Proc. Cas. Act. Soc.* 32, 13–20. 5
- Bailey, A. (1950). Credibility procedures - Laplace's generalisation of bayes' rule and the combination of collateral knowledge with observed data. *Proc. Cas. Act. Soc.* 37, 7–23, 94–115. 5
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2 ed.). Springer Series in Statistics. New York: Springer-Verlag. 5
- Bladt, M. (2005). A review on phase-type distributions and their use in risk theory. *ASTIN Bulletin* 35, 145–161. 7

- Bolch, G., S. Greiner, H. de Meer, and K. S. Trivedi (2001). *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*, Chapter Steady-state solutions of Markov chains. New York, USA: John Wiley & Sons. 6
- Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin* 4. 5, 14, 24
- Bühlmann, H. (1969). Experience rating and credibility. *ASTIN Bulletin* 5, 157–165. 5, 14, 24
- Bühlmann, H. and A. Gisler (2005). *A Course in Credibility Theory and its Applications*. Heidelberg, Germany: Springer-Verlag Berlin Heidelberg. 2, 6, 15, 34
- Bühlmann, H. and E. Straub (1990). Glaubwürdigkeit für Schadensätze (credibility for loss ratios). *Mitteilungen der Vereinigung Schweizerischer Versicherungs-Mathematiker* 70. 15
- Cox, D. (1955). A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society* 51(2), 313–319. 6
- Diaconis, P. and D. Ylvisaker (1979). Conjugate priors for exponential families. *Annals of Statistics* 7, 269–281. 15

- Erlang, A. (1909). Sandsynlighedsregning og telefonsamtaler. *Nyt tidsskrift for Matematik* 20, 33–39. 16
- Fackrell, M. (2009). Modelling healthcare systems with phase-type distributions. *Health Care Management Science* 12(1), 11–26. 7
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). *Bayesian Data Analysis* (2 ed.). Boca Raton, FL: Chapman and Hall/CRC. 5
- Goel, P. (1982). On implications of credible means being exact Bayesian. *Scand. Act. J.*, 41–46. 15
- Goulet, V. (1998). Principles and application of credibility theory. *Journal of Actuarial Practice* 6. 15
- Goulet, V. (2008). *Encyclopedia of Quantitative Risk Analysis and Assessment*, Volume 1. John Wiley & Sons. 5, 12
- Hachemeister, C. (1975). *Credibility, Theory and Applications, Proceedings of the Berkeley Actuarial Research Conference on Credibility*, Chapter Credibility for regression models with application to trend, pp. 129–163. New York, USA: Academic Press. 15
- Hassan-Zadeh, A. and D. A. Stanford (2013). Bayesian and Bühlmann credibility for phase-type distributions with a univariate risk parameter. Working Paper, University of Western Ontario. 32, 35, 74, 75

- Heilmann, W. (1989). Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics* 77-95, 3. 5
- Herbertsson, A. (2011). Modelling default contagion using multivariate phase-type distributions. *Review of Derivatives Research* 14(1), 1–36. 7
- Herzog, T. (1990). Credibility: the Bayesian model versus Bühlmann's model. *Trans. Soc. Act.* 41, 43–88. 14
- Hewitt(Jr.), C. (1970). Credibility for severity. *Proceedings of the Casualty Actuarial Society* 57, 148–171. 14
- Jensen, A. (1953). Markoff chains as an aid in the study of markoff processes. *Skand. Aktuarietidskr* 36, 87–91. 30
- Jewell, W. (1974a). Credibility means are exact bayesian for exponential families. *ASTIN Bulletin* 4, 77–90. 7, 14
- Jewell, W. (1974b). Regularity conditions for exact credibility. *ASTIN Bulletin* 8, 336–341. 14
- Johnson, M. and M. R. Taaffe (1988). The denseness of phase distributions. Working Paper, Purdue University. 30
- Klugman, S. (1992). *Bayesian statistics in actuarial sciences with emphasis on credibility*. Boston, MA: Kluwe. 5



- Klugman, S., T. Rhodes, M. Purushotham, and S. Gill (2009). Credibility theory practice. *MIB Solutions*. 6
- Klugman, S. A., H. H. Panjer, and G. E. Willmot (2008). *Loss Models from Data to Decisions*. Hoboken, New Jersey: John Wiley and Sons. Inc. 5, 8, 12, 24, 28, 29, 64
- Landsman, Z. and U. Makov (1998). Exponential dispersion models and credibility. *Scandinavian Actuarial Journal* 1, 89–96. 5
- Landsman, Z. and U. Makov (2000). On credibility evaluation and the tail area of the exponential dispersion family. *Insurance: Mathematics and Economics* 27, 277–283. 6
- Latouche, G. and V. Ramaswami (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics. 16, 21, 36, 62
- Longley-Cook, L. H. (1962). An introduction to credibility theory. *Proceedings of the Casualty Actuarial Society* 49, 194–221. 12
- Makov, U., A. Smith, and Y. Liu (1996). Bayesian methods in actuarial science. *Journal of the Royal Statistical Society* 45, 503–515. 15
- Mayerson, A. L. (1964). A Bayesian view of credibility. *Proc. Cas. Act. Soc.* 51, 85–104. 5

- Meester, L. E. and M. Sander (2007). Stochastic delay propagation in railway networks and phase-type distributions. *Transportation Research* 41(2), 218–230. 7
- Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society I*, 24–30. 3
- Neuts, M. (1981). *Matrix-geometric solutions in stochastic models*, Volume 2 of *John Hopkins Series in the Mathematical Sciences*. Baltimore, Md: Johns Hopkins University Press. 6, 16, 18, 19, 20
- Neuts, M. (1989). *Structured stochastic matrices of the M/G/1 type and their applications*, Volume 5 of *Probability: Pure and Applied*. New York: marcel Dekker Inc. 16
- Neuts, M. (1995). *Algothmic probability*. Stochastic Modeling Series. London: Chapman & Hall. 16
- Norberg, R. (1979). The credibility approach to experience rating. *Scandinavian Actuarial Journal*, 181–221. 2
- Norberg, R. (2004). Credibility theory. *Encyclopedia of Actuarial Science*, 398–406. 12, 13
- Schmidt, K. (1980). Convergence of bayes and credibility premium. *Astin Bull.* 10, 167–172. 15

- 
- Shanthikumar, J. (1985). Bilateral phase-type distributions. *Naval Research Logistics Quarterly* 32, 119–135. 30
- Stanford, D. (2011). Discussion on the paper “Risk modelling with the mixed erlang distribution” by G.E. Willmot and X.S. Lin. *Appl. Stochastic Models Bus.* 21, 17–18. 30
- Tschupp, M. (2011). Application of credibility theory to group life insurance. *Milliman Research Report*. 6
- Whitney, A. W. (1918). The theory of experience rating. *Proceedings of the Casualty Actuarial Society* 4, 274–292. 5, 11

# Curriculum Vitae

**Name:** Yanyan Zang

**Post-Secondary Education and Degrees:** University of Western Ontario  
London, ON, Canada  
2009 - 2013, Ph.D Candidate

University of Western Ontario  
London, ON  
2008 - 2009, M.Sc

**Honours and Awards:** Western Graduate Research Scholarship  
2008 - 2013

**Related Work Experience:** Teaching Assistant and Research Assistant  
University of Western Ontario  
2008 - 2013