

August 2015

Addressing Very Short Stimulus Encoding Times in Modeling Schizophrenia Cognitive Deficit

Colleen D. Cutler

The University of Western Ontario

Supervisor

Dr. R. W. J. Neufeld

The University of Western Ontario

Graduate Program in Psychology

A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

© Colleen D. Cutler 2015

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

 Part of the [Psychiatry and Psychology Commons](#)

Recommended Citation

Cutler, Colleen D., "Addressing Very Short Stimulus Encoding Times in Modeling Schizophrenia Cognitive Deficit" (2015). *Electronic Thesis and Dissertation Repository*. 2983.

<https://ir.lib.uwo.ca/etd/2983>

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 tadam@uwo.ca, wlsadmin@uwo.ca.

ADDRESSING VERY SHORT STIMULUS ENCODING TIMES IN
MODELING SCHIZOPHRENIA COGNITIVE DEFICITS
(Thesis format: Monograph)

by

Colleen Cutler

Graduate Program in Psychology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

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

© Colleen D. Cutler 2015

Abstract

It is well known that encoding times in persons with schizophrenia are longer than those of normal controls. Neufeld and others have argued that this is the consequence of additional subprocesses being executed during the encoding process in the case of schizophrenia. In general they expressed an encoding time as the sum of k' independent exponentially-distributed subprocesses, each executed with rate v . A troubling consequence of their application of this model to real data was that under some circumstances some individuals appeared to encode instantaneously. This was accommodated in Neufeld et al. by placing a Poisson distribution on k' . In this thesis the view is taken that $k' = 0$ is not realistic and an alternative model is developed in which k' is restricted to positive integers. This is made compatible with very short encoding times by introducing a task parameter α into the model. The problem of estimating α is addressed.

Keywords: encoding times, schizophrenia, subprocesses

Contents

Abstract	ii
List of Tables	iv
List of Appendices	v
1 Mathematical Models and Theoretical Results	1
1.1 Introduction	1
1.2 The Original Neufeld Model	7
1.3 A Competing Model	12
1.4 Estimating α	16
1.4.1 Estimating α from the x data	16
1.4.2 Obtaining the x data from encoding times	22
1.5 Maximum likelihood for the geometric distribution	24
2 Numerical Results	26
Bibliography	35
A Sampling Schematic	40
B Proofs of Results	41
Curriculum Vitae	48

List of Tables

2.1	$\alpha = .01, p = .8$	27
2.2	$\alpha = .01, p = .2$	29
2.3	$\alpha = .01, p = .05$	30
2.4	$\alpha = .17, p = .8$	31
2.5	$\alpha = .17, p = .2$	33
2.6	$\alpha = .17, p = .05$	34

List of Appendices

Appendix A Sampling Schematic	40
Appendix B Proofs of Results	41

Chapter 1

Mathematical Models and Theoretical Results

1.1 Introduction

Of considerable interest in cognitive science are the earmarks of psychopathology, in particular the key elements which might distinguish one disorder from another. One concept which has received considerable attention is that of *encoding* in the case of paranoid schizophrenia (for a review see Neufeld, 2007a). Encoding is the process by which an object or event is transformed mentally into a form which facilitates carrying out the task at hand. For example, consider a basic memory search task (Sternberg, 1975) where the participant is first presented with a collection of alphanumeric items (memory set) then later is presented with an item (probe item) which may or may not have been in the set. The total reaction time is the amount of time required for the participant to indicate whether or not the item belonged to the memory set. In order to accomplish this, the participant must first encode the probe item, i.e., extract its salient physical features (curves, lines, and intersections) in order to facilitate comparison with members of the memory set. Alternatively, if the task at hand were to have participants indicate whether the “real-life size” of a probe item matched with that of a member of a memory set,

then the relevant information to be encoded would be the size properties of the item. Tasks which involve sentence verification would require the encoding of linguistic properties, and so forth.

The reaction time z can then be decomposed into a sum

$$z = t + y + w \quad (1.1)$$

where t represents the encoding time, y consists of additional mental processes such as making comparisons with members of a memory set and making a decision, and w is the physical reaction time to render a yes-no response.

Converging evidence suggests that persons with schizophrenia display prolonged reaction times and that this prolongation is due specifically to a protracted encoding time t (the other terms y and w are spared; e.g., Neufeld, Vollick, & Highgate, 1993; Neufeld, 2007a). This elongation of encoding times is particularly apparent in persons with paranoid schizophrenia (Neufeld & Williamson, 1996). The paranoid subtype has been removed from the new DSM-V but there is evidence that such a subtype exists, e.g., Nicholson & Neufeld, 1993, and the previous DSM-IV. A person will be considered to have paranoid schizophrenia if he or she meets the criteria in DSM-IV. The elongation of encoding times is not only apparent in memory search tasks such as those described above, but is believed to be the reason behind the saccadic eye movements observed in persons with paranoid schizophrenia when they attempt to track a continuously moving object. Paranoid patients display the highest rate of saccades compared to non-paranoid patients, patients with affective disorders, and healthy controls (Neufeld & Williamson, 1996). The saccades have been viewed as the result of inefficient encoding of the visual stimulus properties required for smooth tracking.

Consequently, methods of determining parameters related to encoding times (particularly if they do not tax an already stressed patient) might provide a welcome window into the degree of “illness” of the patient (see Neufeld, 2007a, and the discussion in Section 5 of Neufeld, Boksman, Vollick, George, & Carter, 2010). A first step in this is to recognize that encoding times will vary from trial to trial and will be governed by a stochastic process. The goal is to develop a mathematical model that adequately captures the stochastic system that is the encoding process. The expression “the encoding process” is used because evidence suggests that the same model architecture holds in general. In the case of paranoid schizophrenic illness, a slippage in a parameter occurs, not a change in architecture (Neufeld, 2007a; Neufeld et al., 2010).

In order to motivate the models considered in this thesis, results from the analysis of some aspects of an experiment of George and Neufeld (2007) will first be described. Participants viewed a four-letter word in the central visual field for 1.5 seconds. This was immediately followed by the presentation of a probe item for 20 milliseconds; this probe item consisted of two words, one in the left visual field and one in the right visual field. Participants were instructed to press the “yes” key as quickly as possible if either of the probe words matched the first word; otherwise they were to press “no” as quickly as possible. In the case of a match, there was considered to be a differential encoding load (low vs. high) depending on the visual field in which the matching probe word was presented. A word presented in the right visual field should be processed more quickly (left hemispheric superiority for verbal stimuli) than one presented in the left visual field. For the purposes of this investigation only positive trials (i.e., trials on which a probe word matched the original word) were considered. In this way a 2 x 2 factorial design with four cells was created, where one factor was encoding load (low vs.

high) and one factor was health status (healthy control vs. paranoid schizophrenia). Multiple trials were run on each participant and the expression of the reaction time on a single trial took the form of (1.1). More specifically, if the reaction time on the j^{th} trial of the i^{th} participant in one of the four cells is denoted by z_{ij} , then

$$z_{ij} = t_{ij} + y_{ij} + w_{ij} \quad (1.2)$$

In their analysis of the above data, Neufeld et al. (2007) and Neufeld et al. (2010) first calculated the sample mean and the sample variance of the reaction times for each participant within a cell. Those values were then averaged over all the participants within the cell (see the sampling schematic in Appendix A where there are M participants per cell, each subject to N encoding trials.) The effect was to produce two statistics from each cell, an average mean \bar{z} and an average variance \bar{u}^2 . These quantities were the basic units of analysis and had the advantage of cancelling out much of the noise in the individual trials so that differences between cells could be observed. The basic results were as expected: reaction times were longer when the encoding load was higher or when the participant had paranoid schizophrenia. A particularly interesting result, however, was that means and variances were *additive* over the four cells (i.e., there was no interaction). Mean reaction times for both healthy and paranoid individuals increased by the same amount as the encoding load moved from low to high; similarly for mean variances. This observation is important as it restricts the possible valid models for encoding, a topic which has been addressed at length by Neufeld et al. (2010).

Since, as noted earlier, processes other than encoding seem to be spared in paranoid schizophrenia, it is not unreasonable to assume that over all participants the y_{ij} variables are independently and identically distributed with some mean $E(Y)$ and variance $\text{Var}(Y)$, and similarly that the w_{ij}

are independently and identically distributed with their own mean and variance. However it is expected that the distribution of the t_{ij} variables will depend on the cell to which the participant belongs, since encoding is affected by encoding load and health status. Simply put, if for a particular cell (*) we write

$$\bar{z} = \bar{t} + \bar{y} + \bar{w} \quad (1.3)$$

then

$$E(\bar{z}) = E(T)^* + E(Y) + E(W) \quad (1.4)$$

where only the first term on the right hand side of (1.4) depends on the cell (*). With somewhat more work it can be argued that

$$E(\bar{u}^2) = E(\text{Var}(T))^* + \text{Var}(Y) + \text{Var}(W) \quad (1.5)$$

where only the first term on the right hand side of (1.5) depends on the cell (*)¹. The initial work of this thesis is to examine models of encoding times whose parameters can be suitably adjusted for the cell (*) to which the participant belongs.

Returning to the problem of additivity, first consider the case of the four means in the four cells. Additivity in (1.4) is clearly unaffected by the “nuisance parameters” $E(Y)$ and $E(W)$. Suppose there exists a constant $c > 0$ such that $E(T)^*$ in the cell (*) takes the general form

$$E(T)^* = m^* c \quad (1.6)$$

for some $m^* > 0$ which depends on the cell. Evidently, based on the knowledge that encoding times increase under increased encoding load or under change from healthy status to paranoid

¹The expectation sign appears in the first element on the right hand side of (1.5) because ultimately the distribution of the encoding time T will depend on the individual and must be averaged out over individuals in the cell. See (1.15) and (1.16).

schizophrenia status, it can be deduced that there are parameters $m > 0$, $h > 0$, and $g > 0$ such that $m^* = m$ for normals under low-encoding load, $m^* = m + h$ for normals under high-encoding load, and $m^* = m + g$ for patients with paranoid schizophrenia under low encoding load. Additivity then results if and only if $m^* = m + g + h$ for patients with paranoid schizophrenia under high encoding load. Similar conclusions can be reached about the four average variances in the four cells if $E(\text{Var}(T))^*$ takes the form

$$E(\text{Var}(T))^* = n^* d \quad (1.7)$$

where the constants here do not necessarily take the same values as those in (1.6). (However a proliferation of different constants may result in a situation where it is not possible to estimate all parameters or to test model fit.) In Neufeld et al. (2007, 2010) the tendered model (to be described more fully in the next section) featured $m^* = n^*$ and both c and d were functions of the same parameters. Moreover additivity was “forced” in the estimation procedure by requiring that $m^* = m + g + h$ in the fourth cell (a tactic to be followed later in the procedure considered in this thesis). Model parameters in (1.6) and (1.7), as well as the nuisance parameters $E(Y)$ and $\text{Var}(Y)$ were estimated by fitting a loss function and using a minimization procedure (e.g., Neufeld & McCarty, 1994; Chandler, 1959). The quantities $E(W)$ and $\text{Var}(W)$ were directly replaced by the known experimental values 0.160 and 0.001296 respectively (Woodworth & Schlossberg, 1954). See (1.17) and (1.18) and the discussion in the next section.

As a consequence of the above estimation procedure, an anomalous value was found for m that cast some doubt upon the overall correctness of the model. This provided the impetus for the research in this thesis. In the next section the original Neufeld model is described in detail and the anomaly discussed. Subsequently, a modified new model is then developed, along

with estimation procedures and a discussion of the attendant difficulties. Finally, in the second chapter, some numerical techniques and results are provided.

1.2 The Original Neufeld Model

The first step in model construction occurs under the assumption that two basic parameters are involved in the distribution of the encoding time T : the number k' of subprocesses that must be executed in order for a probe item to be encoded, and the rate v at which each subprocess can be completed (sometimes called *capacity*). This construct must take place within an architectural framework that embraces additivity as discussed in the Introduction. See Neufeld and Williamson (1996), Neufeld et al. (1993), Neufeld et al. (2007), and Neufeld et al. (2010) for detailed discussions on architectures that support additivity. The basic conclusion here is that additivity can be obtained with a model when capacity behaves the same from cell to cell (i.e., remains constant under a change in encoding load or health status) but the number of subprocesses increases as encoding load increases or as health status switches from healthy to paranoid schizophrenia. An increase in number of subprocesses is to be expected with an increase in encoding load, but it is not known why individuals with paranoid schizophrenia show an increase in the number of subprocesses required to encode an item. This may occur for various reasons; the person may simply fail at navigating a subprocess step (requiring repetition), may fail to tag the step as completed and unnecessarily repeat it, or may be disturbed by the intrusion of irrelevant subprocesses (see Cromwell & Dokecki (1968), Hemsley (1993, 1994), Steffy and Galbraith (1980), and Steffy and Waldman (1993) for discussion). Russell and Knight (1977) have suggested that there is insertion of “priming” or activating subprocesses prior to the start of actual encoding. This is somewhat supported by recent fMRI results

(Bluhm et al., 2007; Murphy, Birn, Handwerker, & Bandettini, 2009) that suggest there is an abnormal “resting state” neurocircuitry in individuals with paranoid schizophrenia. Braver and Barch (2006) have also noted that anomalous default-system connectivity may reduce the deployment of resources once encoding is initiated.

The Neufeld model (Neufeld et al., 2002, Neufeld et al., 2007; Neufeld et al., 2010) can be built as follows. Suppose that for a given individual the time ℓ required to complete a subprocess is exponentially distributed with rate parameter ν , i.e., it has probability density function (pdf)

$$g(\ell | \nu) = \nu e^{-\nu \ell} \quad \text{for } \ell > 0 \quad (1.8)$$

The quantity ν is the capacity or rate at which the individual can execute a subprocess. Townsend and Ashby (1983) discuss some empirical evidence in support of exponentially-distributed subprocess times but this is by no means definitive and will be modified later in this thesis. If the individual must complete k' subprocesses in order to encode an item (where k' is a positive integer) then his or her encoding time is given by

$$t = \ell_1 + \dots + \ell_{k'} \quad (1.9)$$

It has been emphasized that k' should be a positive integer as it is central to this thesis that in order for encoding to occur at least one subprocess must be executed (this will be expanded upon shortly). Further assuming that the subprocessing times are mutually independent (again a defensible assumption, Neufeld et al. 2010), it follows that t will have an Erlang distribution with parameters k' and ν , and pdf given by

$$f(t | k', \nu) = \frac{\nu^{k'} t^{k'-1}}{(k' - 1)!} e^{-\nu t} \quad \text{for } t > 0 \quad (1.10)$$

The Erlang is a special case of the gamma distribution and it is well-known (Hogg, McKean, & Craig, 2005) that its mean and variance are given by

$$\mu = \frac{k'}{v} \quad \text{and} \quad \sigma^2 = \frac{k'}{v^2} \quad (1.11)$$

Note that the simplest implementation of this model would be to keep v identical over all four cells while keeping k' fixed within a cell but allowing it to vary between cells, reflecting a change in the number of subprocesses executed as encoding load or health status changes. Then (1.6) and (1.7) hold with $m^* = n^* = k'$, $c = 1/v$, and $d = 1/v^2$ and additivity can be obtained by a simple constraint on the fourth cell as discussed in the second last paragraph of the Introduction.

The shortcoming of the above implementation, however, is that it assumes that all individuals have the same capacity v and that all individuals within a cell require the same number of subprocesses k' to encode an item. This is unlikely to hold in practice. An innovation of Neufeld et al. (2002), Neufeld et al. (2007), and Neufeld et al. (2010) was to assign each participant his or her own particular values of k' and v , then allow these values to vary over each cell according to some distribution (to be discussed shortly). In keeping with the evidence that capacity does not depend on encoding load or health status, the distribution for v was deemed to be the same in each cell. However, the distribution of k' shifted from cell to cell to accommodate the expected changes in the average number of required subprocesses.

The distribution assigned to v was that of a gamma distribution with shape parameter $k > 0$ and rate parameter $r > 0$ (note that k need not be an integer here), thus having pdf

$$f(v|k, r) = \frac{r^k v^{k-1}}{\Gamma(k)} e^{-rv} \quad \text{for } v > 0 \quad (1.12)$$

In this context k and r are sometimes called *hyperparameters*. The parameter k is related

to performer competence on the task at hand whereas r is determined by the participant's susceptibility to stress effects (Neufeld, 2007b).

The distribution assigned to k' in the cell (*) was Poisson with parameter λ^* . This is a discrete distribution on the nonnegative integers $k' = 0, 1, 2, \dots$ with probability mass function (pmf) given by

$$P(k') = \frac{e^{-\lambda^*} (\lambda^*)^{k'}}{(k')!} \quad (1.13)$$

The parameter λ^* is the mean of the distribution, i.e., $E(k') = \lambda^*$. Note that $P(k' = 0) = e^{-\lambda^*} > 0$ so this violates the original assumption that $k' \geq 1$. This yields both a conceptual problem (some individuals must be able to encode without executing any subprocesses) and a mathematical problem (the Erlang distribution (1.10) is not defined when $k' = 0$) although the latter can be resolved by setting the encoding time $t = 0$ when $k' = 0$. This situation will be elaborated upon further when estimation of parameters is discussed.

Now let T_i denote the encoding distribution of the i^{th} participant in cell (*). It follows from the model assumptions and (1.11) that there exist parameters k'_i and v_i such that

$$E(T_i) = \frac{k'_i}{v_i} \quad \text{and} \quad \text{Var}(T_i) = \frac{k'_i}{v_i^2} \quad (1.14)$$

Assume now that the distributions of k' and v are independent. Mixing over all participants in cell (*) (i.e, averaging over all individuals in the cell with respect to the Poisson and gamma distributions) thus yields

$$E(T)^* = E(E(T_i)) = E\left(\frac{k'_i}{v_i}\right) = E(k')E(1/v) \quad (1.15)$$

and

$$E(\text{Var}(T))^* = E(\text{Var}(T_i)) = E\left(\frac{k'_i}{v_i^2}\right) = E(k')E(1/v^2) \quad (1.16)$$

Remark 1.2.1 Note that regardless of the distributions placed on k' and v , (1.15) and (1.16) hold and take the form of (1.6) and (1.7) with $m^* = n^* = E(k')$, $c = E(1/v)$, and $d = E(1/v^2)$, so that additivity can be obtained even with changes in the distributions provided that the distribution of k' has the property that $E(k')$ can be adjusted to be any value $a > b_0$ (by a change in its parameter(s)) where b_0 is the infimum over all values that $E(k')$ can take on. This will be called the *fill property* for convenience. To see the connection of the fill property with additivity, note that if $E(k') = m$ in the first cell, then the fill property implies that, for the subsequent cells, $E(k') = m + h$, $E(k') = m + g$ and $E(k') = m + g + h$ are possible since $h, g > 0$. The Poisson of course has the fill property since $b_0 = 0$ in this case and $E(k') = \lambda^*$ is possible for any $\lambda^* > 0$.

For the gamma distribution it can easily be shown that $E(1/v) = r/(k - 1)$ for $k > 1$ and $E(1/v^2) = r^2/(k - 1)(k - 2)$ for $k > 2$. Neufeld et al. (2007) and Neufeld et al. (2010) obtained the equations

$$E(\bar{z}) = \frac{\lambda^* r}{k - 1} + E(Y) + .160 \quad \text{in cell (*)} \quad (1.17)$$

and

$$E(\bar{u}^2) = \frac{\lambda^* r^2}{(k - 1)(k - 2)} + \text{Var}(Y) + .001296 \quad \text{in cell (*)} \quad (1.18)$$

The parameter k was set to $k = 30$ as an appropriate choice for the encoding task at hand (Neufeld, 2007b; Neufeld et al., 2007) so it did not require estimation. The value of λ^* was set to m , $m + h$, $m + g$, or $m + g + h$ depending on the cell (*). This resulted in six parameters to be estimated: m , g , h , r , $E(Y)$, and $\text{Var}(Y)$ using eight observations (the cell means and cell variances). The estimated value of m turned out to be $\hat{m} = .0971 < 1$. Since m is the expected number of subprocesses $E(k')$ executed by healthy controls under the low encoding load, this implies that a segment of that population was encoding instantaneously without exe-

cutting any subprocesses (i.e. $k' = 0$ for some individuals). In fact $\hat{h} = .08175$ so the expected number of subprocesses $m + h$ executed by healthy controls even under the high encoding load was only 0.1789, again implying that some individuals must be encoding without executing subprocesses. The estimates for the participants with paranoid schizophrenia exceeded 1 but this was not reassuring; instantaneous encoding could be occurring for some individuals while being compensated for by particularly poor performances by others.

It is important to note, as indicated earlier in the Introduction, that the parameters in (1.17) and (1.18) were estimated by minimizing a loss function in which the Poisson played no role other than to ensure that additivity could be obtained. It would be correct to say that the Poisson did not *cause* the anomalous values for \hat{m} and \hat{h} but rather that it *accommodates* them by virtue of the fact that $P(k' = 0) > 0$ renders instantaneous encoding mathematically possible. In this sense, the small values for \hat{m} and \hat{h} can be viewed as providing evidence for instantaneous encoding and supporting a Poisson model. In this thesis a competing model is developed which does not permit instantaneous encoding but can still account for the small estimates \hat{m} and \hat{h} .

1.3 A Competing Model

In this section it will be assumed that any viable candidate for the distribution of the number of subprocesses k' must take values only on the positive integers $k' = 1, 2, 3, \dots$. For convenience such distributions will be called *positive distributions* and they necessarily feature $E(k') \geq 1$. In order to be able to account for fractional values of m such as that obtained in the preceding section, the processing time of subprocesses will be speeded up by introducing a parameter α . (In principle any value of $\alpha > 0$ is possible but what will be of interest here is the case $0 < \alpha < 1$.) The time to execute a subprocess will no longer be exponentially-distributed as in

(1.8) but rather will follow a gamma distribution with shape parameter α and rate parameter ν

$$g(\ell | \alpha, \nu) = \frac{\nu^\alpha \ell^{\alpha-1}}{\Gamma(\alpha)} e^{-\nu \ell} \quad (1.19)$$

Again letting the encoding time be the sum of the subprocess execution times

$$t = \ell_1 + \dots + \ell_k \quad (1.20)$$

it follows that the distribution of t is gamma with shape parameter $k'\alpha$ and rate ν (see Hogg et al. 2005)

$$f(t | \alpha, k', \nu) = \frac{\nu^{k'\alpha} t^{k'\alpha-1}}{\Gamma(k'\alpha)} e^{-\nu t} \quad (1.21)$$

The mean and variance are respectively

$$\mu = \frac{k'\alpha}{\nu} \quad \text{and} \quad \sigma^2 = \frac{k'\alpha}{\nu^2} \quad (1.22)$$

Note that the mean and variance go to zero as $\alpha \rightarrow 0$. The parameter α has an interpretation as a *task parameter*. For very simple encoding tasks, subprocess execution might be dispatched very quickly (small α) and hence encoding times would be very short (but not instantaneous).

If T_i is the encoding time of the i^{th} participant in cell (*) then

$$E(T_i) = \frac{k'_i \alpha}{\nu_i} \quad \text{and} \quad \text{Var}(T_i) = \frac{k'_i \alpha}{\nu_i^2} \quad (1.23)$$

Now place a positive probability distribution $P(k')$ on k' with parameter(s) that vary over cells and satisfies the fill property in Remark 1.2.1. Place the same gamma distribution (1.12) on ν .

Then mixing over participants in a cell (*) with respect to these distributions yields

$$E(T)^* = E(E(T_i)) = E\left(\frac{k'_i \alpha}{\nu_i}\right) = \alpha E(k') E(1/\nu) \quad (1.24)$$

and

$$E(\text{Var}(T))^* = E(\text{Var}(T_i)) = E\left(\frac{k'_i \alpha}{\nu_i^2}\right) = \alpha E(k') E(1/\nu^2) \quad (1.25)$$

Thus $m^* = \alpha E(k')$ rather than $E(k')$ as in (1.15) and (1.16), and the estimates of m^* could be very small if α is very small. Specifically, that which is actually being estimated by the loss function is the product $\alpha E(k')$ rather than $E(k')$.

Two common distributions that are candidates for $P(k')$ are now described.

Example 1.3.1 Consider the geometric distribution $P(k')$ given by

$$P(k') = (1 - p)^{k'-1} p \quad \text{for } k' = 1, 2, 3 \dots \quad (1.26)$$

where the parameter $p > 0$ represents the probability of “success” on a trial. It is well known that $E(k') = 1/p$ and it follows that the fill property in Remark 1.2.1 is satisfied because here $b_0 = 1$ (occurring when $p = 1$) and obviously any value of $E(k') > 1$ can be obtained by the correct selection of p . This fact will be utilized later when developing a constraint on the parameter p_4 in the fourth cell (as a function of the parameters p_1, p_2, p_3 in the first three cells) to obtain additivity as discussed in the Introduction. The geometric distribution has a nice physical interpretation here as each trial (each “subprocess”) can be viewed as an attempt to encode the item. The value of k' corresponds to the number of attempts required to successfully encode the item.

Example 1.3.2 Consider the truncated Poisson distribution $P(k')$ given by

$$P(k') = \frac{e^{-\lambda} \lambda^{k'}}{(1 - e^{-\lambda})(k')!} \quad \text{for } k' = 1, 2, 3 \dots \quad (1.27)$$

The truncated Poisson is obtained from the standard Poisson by removing zero from the range of possible values and redistributing the mass over the positive integers. The distribution behaves much like the standard Poisson when λ is reasonably large. It can be shown that $E(k') = \lambda/(1 - e^{-\lambda})$. The infimum b_0 occurs when $\lambda \rightarrow 0$; specifically $\lim_{\lambda \rightarrow 0} E(k') = \lim_{\lambda \rightarrow 0} \lambda/(1 - e^{-\lambda}) = 1$ by L'Hôpital's rule. The fill property of Remark 1.2.1 thus holds

because $E(k')$ is a continuous function of λ which diverges to ∞ as $\lambda \rightarrow \infty$ and hence by the Intermediate Value Theorem must take on every value between 1 and ∞ .

Now suppose a positive distribution $P(k')$ satisfying the fill property has been selected. Then (1.17) and (1.18) take the form

$$E(\bar{z}) = \frac{m^* r}{k-1} + E(Y) + .160 \quad \text{in cell (*)} \quad (1.28)$$

and

$$E(\bar{u}^2) = \frac{m^* r^2}{(k-1)(k-2)} + \text{Var}(Y) + .001296 \quad \text{in cell (*)} \quad (1.29)$$

where $m^* = \alpha E(k')$ in cell (*). Note that it is not possible to separate α and $E(k')$ in the above. If α is adjusted downward then $E(k')$ can be adjusted upward so that the product m^* is unchanged (this uses the fill property). Formally this nonuniqueness problem is stated as the following:

Proposition 1.3.3. Suppose α, m, g, h have been found so that $m_1^* = \alpha m$, $m_2^* = \alpha(m + h)$, $m_3^* = \alpha(m + g)$ and $m_4^* = \alpha(m + g + h)$. Then there exist alternate solutions α_0, m_0, h_0 , and g_0 such that $m_1^* = \alpha_0 m_0$, $m_2^* = \alpha_0(m_0 + h_0)$, $m_3^* = \alpha_0(m_0 + g_0)$ and $m_4^* = \alpha_0(m_0 + g_0 + h_0)$.

See Appendix B for a proof.

As a consequence of the above, another method must be found for estimating α . The key here is to *not* average reaction time data across participants in a cell as done in (1.28) and (1.29). Rather, the differences between reaction times of individuals in a cell provide information about k' and α . This involves looking at the data on a microscopic level instead of through summary statistics. This is the topic of the next section.

1.4 Estimating α

In this section an additional assumption will be made on the distribution $P(k')$ in order to facilitate understanding and simplify proofs. A positive probability distribution $P(k')$ will be called *complete* if $P(k') > 0$ for each integer $k' \geq 1$. (Note that the geometric and truncated Poisson distributions are examples of complete distributions.)

1.4.1 Estimating α from the x data

Suppose a sequence of values k'_1, k'_2, k'_3, \dots has been generated from a complete distribution $P(k')$ (for example by sampling over individuals within a cell). This is considered to be a “hidden” sequence in that it is not observed by the experimenter. Suppose there is also an unknown value of $\alpha > 0$ and the actual observed data consists of the sequence x_1, x_2, x_3, \dots where $x_i = k'_i \alpha$ for each i . This is called the x data and may seem mysterious at this juncture. However, obtaining the x data from encoding time data is discussed in the next subsection. At present it is simply assumed that the x data has been obtained by some procedure, and the question of estimating α from it is addressed.

The first step is to verify the uniqueness of α .

Theorem 1.4.1.1 Suppose x_1, x_2, \dots has two representations

$$x_1 = k'_1 \alpha, x_2 = k'_2 \alpha, \dots, x_m = k'_m \alpha, \dots \tag{1.30}$$

$$x_1 = k''_1 \beta, x_2 = k''_2 \beta, \dots, x_m = k''_m \beta, \dots$$

where $\alpha > 0$, $\beta > 0$, and k'_1, k'_2, k'_3, \dots and $k''_1, k''_2, k''_3, \dots$ are independent and identically-distributed sequences from the same positive distribution $P(k')$. Then $\alpha = \beta$ and thus necessarily $k'_i = k''_i$ for every i . (Note: this result does not require $P(k')$ to be complete; see proof in Appendix B.)

Whereas uniqueness of α holds for an infinite sequence, this is not the case in a finite sample x_1, x_2, \dots, x_M from a complete distribution. Suppose

$$x_1 = k'_1 \alpha, x_2 = k'_2 \alpha, \dots, x_M = k'_M \alpha \quad (1.31)$$

Then the x sample can also be expressed as

$$x_1 = (nk'_1)(\alpha/n), x_2 = (nk'_2)(\alpha/n), \dots, x_M = (nk'_M)(\alpha/n) \quad (1.32)$$

for any positive integer n , so there are infinitely many representations. Thus finding α is not simply a matter of constructing integers from the x sample, i.e., selecting some α such that $x_1/\alpha, x_2/\alpha, \dots, x_M/\alpha$ produces integers. There are infinitely many such α and the goal is to find the one “true” α as described in Theorem 1.4.1.1. From this point onward, in order to avoid confusion, the symbol α will be reserved for the true α . An estimate of α will now be developed.

Proposition 1.4.1.2 Let x_1, x_2, \dots, x_M be an x sample arising from a complete distribution.

Define

$$A = \left\{ \beta > 0 \mid \frac{x_i}{\beta} \text{ is a positive integer for all } i = 1, 2, \dots, M \right\} \quad (1.33)$$

Then A is not empty and it has a largest member which will be denoted by α_{\max} .

Note that A is of course not empty because $\alpha \in A$. The quantity α_{\max} can be expressed as $\alpha_{\max} = \sup A$. The work of the proof of Proposition 1.4.1.2 (in Appendix B) is to show that α_{\max} belongs to A . It then follows that the constructed set of integers

$$j_1 = \frac{x_1}{\alpha_{\max}}, j_2 = \frac{x_2}{\alpha_{\max}}, \dots, j_M = \frac{x_M}{\alpha_{\max}} \quad (1.34)$$

are the smallest integers that can be constructed from the x sample. In fact even more can be said about this constructed set.

Proposition 1.4.1.3 The set of integers j_1, \dots, j_M in (1.34) is *coprime*, i.e., if there exists an integer $n \geq 1$ and integers j'_1, \dots, j'_M such that $j_i = nj'_i$ for all $i = 1, \dots, M$, then $n = 1$. (Equivalently, $\gcd\{j_1, \dots, j_M\} = 1$.)

To see that Proposition 1.4.1.3 must be true, suppose that $j_i = nj'_i$ for all $i = 1, \dots, M$. Then $x_i = j_i \alpha_{\max} = j'_i (n \alpha_{\max})$ for $i = 1, \dots, M$, yielding $x_i / (n \alpha_{\max}) = j'_i$ for $i = 1, \dots, M$. Thus $\beta = n \alpha_{\max}$ belongs to the set A in (1.33). If $n \geq 2$ then this would imply that α_{\max} is not the largest member of A . Hence $n = 1$.

The quantity α_{\max} is a viable candidate for an estimate of α based on the following Proposition and Theorem.

Proposition 1.4.1.4 Let $k'_1, k'_2, \dots, k'_m, \dots$ be an independent and identically-distributed sequence from a complete distribution $P(k')$. Then there exists $M \geq 1$ such that k'_1, k'_2, \dots, k'_M is coprime.

The above result can be seen to hold because $P(k' = 1) > 0$ and hence the value 1 must appear at some point, say M , in the sequence. It follows then that k'_1, k'_2, \dots, k'_M is coprime. In fact k'_1, k'_2, \dots, k'_M will tend to be coprime even for rather small M since the completeness of $P(k')$ implies that two coprime integers k'_i and k'_j are likely to be generated early in the sequence (it is generally not necessary to wait for a 1).

Theorem 1.4.1.5 Suppose k'_1, k'_2, \dots, k'_M is generated from a complete distribution $P(k')$ and $x_1 = k'_1 \alpha, x_2 = k'_2 \alpha, \dots, x_M = k'_M \alpha$ is the observed x sample. Then $\alpha_{\max} = \alpha$ if and only if k'_1, k'_2, \dots, k'_M is coprime. Moreover, as a consequence of Proposition 1.4.1.4, there will be some M for which $\alpha_{\max} = \alpha$.

Remark 1.4.1.6 It must be reiterated that the integer sample k'_1, k'_2, \dots, k'_M is unobserved so the

value of M at which it becomes coprime will be unknown. Thus the point at which $\alpha_{\max} = \alpha$ is also unknown. A guess at the latter can be made by plotting α_{\max} vs. M and locating a plateau (however, a plateau may be temporary). The utility of the procedure lies in the fact that the necessary value of M is likely to be relatively small.

Example 1.4.1.7 The present example illustrates the consequences of k'_1, k'_2, \dots, k'_M not being coprime. Suppose $M = 3$, $k'_1 = 6$, $k'_2 = 2$, and $k'_3 = 4$ with $\alpha = 1/8$. Then the observed x sample is $x_1 = 3/4$, $x_2 = 1/4$, and $x_3 = 1/2$. It follows that $\alpha_{\max} = 1/4$ (in particular, $\alpha_{\max} \neq \alpha$) and the constructed integers (1.34) are $j_1 = 3$, $j_2 = 1$, and $j_3 = 2$. Thus $\alpha_{\max} \neq \alpha$ and produces a coprime set (as it must, cf. Proposition 1.4.1.3) which disagrees with the true hidden integer set k'_1, k'_2, k'_3 .

An algorithm for finding α_{\max} : In the preceding example it was easy to see what α_{\max} must be. However, for a large x sample (or for one in which the x_i are expressed to several decimal digits) it may not be so straightforward. The key to developing a simple algorithm for finding α_{\max} is to realize that $\alpha_{\max} \leq x_{\min}$ (where x_{\min} is the smallest value in the x sample) and that in fact $\alpha_{\max} = x_{\min}/j$ for some positive integer j . Under the important restriction that there be *no rounding error* in the members of the x sample, the algorithm proceeds as follows:

1. Let $\alpha_1 = x_{\min}$. If $x_1/\alpha_1, x_2/\alpha_1, \dots, x_M/\alpha_1$ are integers, then $\alpha_{\max} = \alpha_1$ and the algorithm stops. Otherwise proceed to Step 2.
2. Let $\alpha_2 = x_{\min}/2$. If $x_1/\alpha_2, x_2/\alpha_2, \dots, x_M/\alpha_2$ are integers, then $\alpha_{\max} = \alpha_2$ and the algorithm stops. Otherwise proceed to Step 3.
3. Let $\alpha_3 = x_{\min}/3$. If $x_1/\alpha_3, x_2/\alpha_3, \dots, x_M/\alpha_3$ are integers, then $\alpha_{\max} = \alpha_3$ and the algo-

rithm stops. Otherwise proceed to Step 4.

Ultimately the algorithm stops at the first integer j for which division of the x sample by $\alpha_j = x_{\min}/j$ produces integers.

Example 1.4.1.8 A hidden sample of 15 integers k'_1, \dots, k'_{15} was generated from a truncated Poisson distribution $P(k')$ and multiplied by an unknown α to produce the observed x sample

$$.135 \ .090 \ .120 \ .120 \ .090 \ .150 \ .135 \ .180 \ .105 \ .120 \ .150 \ .180 \ .195 \ .135 \ .240 \quad (1.35)$$

Note that $x_{\min} = .090$ so $\alpha_1 = .090$. However, $.135/.090 = 1.5$ which is not an integer, so the algorithm proceeds to Step 2. Setting $\alpha_2 = .090/2 = .045$, note that $.120/.045 = 2.67$ which is not an integer, so proceed to Step 3. Set $\alpha_3 = .090/3 = .030$. Then $.135/.030 = 4.5$ is not an integer. Carrying on in this way, it can be seen that the first integer j for which division by α_j produces a sample of integers is $j = 6$. Thus $\alpha_{\max} = x_{\min}/6 = .090/6 = .015$ and this is the suggested estimate of α . The constructed integer sample (1.34) from $\alpha_{\max} = .015$ is

$$9 \ 6 \ 8 \ 8 \ 6 \ 10 \ 9 \ 12 \ 7 \ 8 \ 10 \ 12 \ 13 \ 9 \ 16 \quad (1.36)$$

In practice it cannot be known for certain if $\alpha_{\max} = \alpha$ and if the correct integer sample has been constructed because it cannot be known for certain if k'_1, \dots, k'_{15} is coprime. However, in the case of this example, (1.35) was created by generating a random sample k'_1, \dots, k'_{15} from a truncated Poisson distribution with parameter $\lambda = 10$ (using the statistical software program **R**(2013)) and multiplying by $\alpha = .015$. Hence the correct values were recovered. This should not be a surprise because $P(k')$ is complete. In fact, because $k'_1 = 9, k'_2 = 6, k'_3 = 8$ form a coprime set, the correct value $\alpha_{\max} = \alpha = .015$ would have been reached at $M = 3$. (A plateau in α_{\max} would have been observed for $M \geq 3$.)

Example 1.4.1.9: Now suppose k'_1, \dots, k'_{15} have been sampled from a geometric distribution and multiplied by some α to obtain the following x sample

0.270000000 0.203333333 0.205000000 0.165000000 0.501666667 0.231666667 0.096666667
 0.145000000 0.133333333 0.101666667 0.056666667 0.135000000 0.235000000 0.028333333
 0.003333333

Nine-digit decimal accuracy has deliberately been taken in order to avoid significant rounding error. (The problem of rounding error is discussed below in Remark 1.4.1.10, and a method for coping with it is addressed in Chapter 2.) Note that $\alpha_1 = x_{\min} = .003333333$ and that division by this value does not produce a sample of integers because, for example, $.028333333/.003333333 = 8.5$ is not an integer. However, setting $\alpha_2 = x_{\min}/2 = .001666667$ produces a sample of integers (to within rounding error in the 9th decimal place). Thus $\alpha_{\max} = .001666667$ and the constructed sample j_1, \dots, j_{15} of integers (1.34) is

$$162 \ 122 \ 123 \ 99 \ 301 \ 139 \ 58 \ 87 \ 80 \ 61 \ 34 \ 81 \ 141 \ 17 \ 2 \quad (1.37)$$

In fact the original k' sample was generated using the geometric random number generator in \mathbf{R} with $p = .01$ and coincided with (1.37). The multiplier was $\alpha = 1/600 = \alpha_{\max} = .001666667$. Again it is not a surprise that the correct values were recovered because the geometric distribution rapidly generates a coprime set.

Remark 1.4.1.10 Although the above method is a particularly powerful theoretical tool for determining α and recovering the original integer sample, it is deeply adversely affected by rounding error in the x sample. Rounding error and/or noise is of course ubiquitous in real-world data. In Chapter 2 a statistical method will be developed to estimate α for noisy x data and applied to the case where $P(k')$ is geometric.

1.4.2 Obtaining the x data from encoding times

At this point a crucial simplifying assumption is made, namely that the actual encoding times t (not just the reaction times z) can be observed. Since $z_{ij} = t_{ij} + y_{ij} + w_{ij}$ as before in (1.2), this implies that either some method of observing the t_{ij} directly must exist or the terms y_{ij} and w_{ij} can be satisfactorily estimated and subtracted from z_{ij} . At this point an adequate method has not been determined (although see Donders (1969) and Townsend and Wenger (2004) for a discussion). It should be noted that an attempt by the present author to obtain encoding times for the George and Neufeld (1987) data by subtracting the estimates of $E(Y)$ and $E(W)$ utilized by Neufeld et al. (2007) resulted in failure as the estimates were too crude to apply to individual trials rather than means. Having said that, it will nonetheless be assumed that the encoding times t_{ij} have been obtained.

Consider the M participants within a given cell, and assume that the competing model holds for some unknown α . Associated with the i^{th} participant are parameters k'_i and v_i where k'_i is generated according to some complete distribution $P(k')$ which satisfies the fill property, and v_i is generated according to the gamma distribution (1.12). The distribution of the encoding time T_i of the i^{th} participant is gamma with mean and variance (see (1.23))

$$E(T_i) = \frac{k'_i \alpha}{v_i} \quad \text{and} \quad \text{Var}(T_i) = \frac{k'_i \alpha}{v_i^2} \quad (1.38)$$

The goal here is to estimate the product $x_i = k'_i \alpha$ for each participant, thereby generating an x sample so that the method of estimating α of the previous subsection can be applied. The key to estimating the x_i values lies in the following theorem:

Theorem 1.4.2.1 Suppose $P(k')$ has the additional property that its mean and variance are finite. Let $t_{i1}, t_{i2}, \dots, t_{iN}$ denote the sample of encoding times of the i^{th} participant, and let \bar{t}_i

and s_i^2 denote the sample mean and sample variance. Then

$$\lim_{N \rightarrow \infty} \bar{t}_i = \frac{k'_i \alpha}{v_i} = \frac{x_i}{v_i} \quad \text{almost surely} \quad (1.39)$$

and

$$\lim_{N \rightarrow \infty} s_i^2 = \frac{k'_i \alpha}{v_i^2} = \frac{x_i}{v_i^2} \quad \text{almost surely} \quad (1.40)$$

The above theorem may at first seem a simple consequence of the Strong Law of Large Numbers. However, the sequence of encoding times is not independently and identically-distributed; rather, it forms an *exchangeable* sequence. See Appendix B for a discussion of exchangeability and a proof of this theorem.

Theorem 1.4.2.1 implies that, for sufficiently large N , \bar{t}_i may be equated with x_i/v_i and s_i^2 may be equated with x_i/v_i^2 . This makes it possible to solve for an estimate \hat{x}_i of x_i . Specifically

$$\hat{x}_i = \frac{\bar{t}_i^2}{s_i^2} \quad (1.41)$$

Thus an estimated x sample $\hat{x}_1, \dots, \hat{x}_M$ is generated for the participants in a cell. It is important, however, to reiterate that these are estimates and hence contaminated by noise. An approach to dealing with noise is given in Chapter 2. The idea is to estimate α and the parameter(s) of $P(k')$ directly within each cell. This will not only provide direct estimates of $\alpha E(k')$ within each cell, but permit testing of the quality of the estimate of α by testing goodness-of-fit to the putative distribution. The technique will be illustrated by application to the geometric distribution. To this end, estimation of the parameters α and p when $P(k')$ follows a geometric distribution is discussed in the next section.

1.5 Maximum likelihood for the geometric distribution

Let $\alpha > 0$ and suppose k'_1, \dots, k'_M are integers generated according to a geometric distribution $P(k')$ with parameter $0 < p < 1$. Let $x_1 = k'_1\alpha, \dots, x_M = k'_M\alpha$ be the corresponding x sample. Note that it is possible to directly consider the likelihood function (Hogg et al., 2005) of α and p as a function of x_1, \dots, x_M . Specifically

$$L(\alpha, p; x_1, \dots, x_M) = \prod_{i=1}^M (1-p)^{\frac{x_i}{\alpha}-1} p = (1-p)^{\frac{\sum x_i}{\alpha}-M} p^M \quad (1.42)$$

Since α is unknown it may be considered to vary at this point, and (1.42) will hold only for those α for which $x_1/\alpha, \dots, x_M/\alpha$ produces positive integers. Such an α will be called *admissible*. (If α is not admissible the likelihood will equal zero.) This means that for a given x sample there is a discrete set of candidates for the value of α ; this set coincides with the set A described in (1.33). The quantity p , on the other hand, can take on any value $0 \leq p \leq 1$ and there is no range restriction based on the x sample. The following theorem is proved in Appendix B:

Theorem 1.5.1 The values $\hat{\alpha}$ and \hat{p} which maximize the likelihood function (1.42) are

$$\hat{\alpha} = \alpha_{\max} \quad \text{and} \quad \hat{p} = \alpha_{\max}/\bar{x} \quad (1.43)$$

It follows then that there exists M for which $\hat{\alpha} = \alpha$ (the true α) by applying Proposition 1.4.1.4 and Theorem 1.4.1.5, and that $\hat{p} \rightarrow p$ (the true p) in probability as $M \rightarrow \infty$ by applying Theorem 6.1.3 of Hogg et al. (2005).

Thus the maximum likelihood method provides an alternative method for deriving α_{\max} as an estimator of α in the case of the geometric. However, it also provides an optimal estimator of p . This allows estimation of parameters within a cell.

Now consider the four cell model of Neufeld et al. (2007), Neufeld et al. (2010). The task

is the same in each of the four cells and so the task parameter α is assumed to be the same in all four cells. However, the probability p of successful encoding varies within each cell. Let Cell 1 denote the healthy-low load (with probability p_1), Cell 2 denote the healthy-high load (with probability p_2), Cell 3 denote the paranoid schizophrenia-low load (with probability p_3), and Cell 4 denote the paranoid schizophrenia-high load (with probability p_4). The parameters p_1, p_2, p_3 are not related to one another although it is expected that $p_2 < p_1$ and $p_3 < p_1$ as the first cell produced the shortest encoding times. However, p_4 is constrained through additivity. Specifically, the fourth cell is not independent of the first three cells. This yields the following

Theorem 1.5.2 Let $\bar{x}_1, \bar{x}_2,$ and $\bar{x}_3,$ denote the respective x sample means within each of the first three cells. Let $x_1^{(1)}, \dots, x_M^{(1)}$ denote the x sample from the first cell, $x_1^{(2)}, \dots, x_M^{(2)}$ denote the x sample from the second cell, and $x_1^{(3)}, \dots, x_M^{(3)}$ denote the sample from the third cell. Then

$$\hat{\alpha} = \max \left(\alpha > 0 \mid \frac{x_1^{(1)}}{\alpha}, \dots, \frac{x_M^{(1)}}{\alpha}, \frac{x_1^{(2)}}{\alpha}, \dots, \frac{x_M^{(2)}}{\alpha}, \frac{x_1^{(3)}}{\alpha}, \dots, \frac{x_M^{(3)}}{\alpha} \text{ are integers} \right) \quad (1.44)$$

$$\hat{p}_1 = \hat{\alpha} / \bar{x}_1, \quad \hat{p}_2 = \hat{\alpha} / \bar{x}_2, \quad \hat{p}_3 = \hat{\alpha} / \bar{x}_3$$

and

$$\hat{p}_4 = \frac{\hat{p}_1 \hat{p}_2 \hat{p}_3}{\hat{p}_1 \hat{p}_2 + \hat{p}_1 \hat{p}_3 - \hat{p}_2 \hat{p}_3} \quad (\text{additivity}) \quad (1.45)$$

(See Appendix B).

These results are applied in the next chapter.

Chapter 2

Numerical Results

In this chapter the problem of noisy data is addressed. $P(k')$ is assumed to be geometric and two examples are presented, one where $\alpha = .01$ and the other where $\alpha = .17$. In each example the four-cell model is considered, where data is generated and parameters estimated from each of the first three cells, the parameters in the fourth cell then being estimated by the additivity constraint. The geometric probabilities in each of the first three cells are $p_1 = .8$, $p_2 = .2$, and $p_3 = .05$, respectively. There were assumed to be 30 participants in each cell (the value $M = 30$ was chosen in order to give the chi-square tests considered herein sufficient power). Within each cell a hidden geometric sample k'_1, \dots, k'_{30} was generated using the geometric random number generator from \mathbf{R} with the appropriate value of p . A random sample v_1, \dots, v_{30} was generated according to the gamma distribution (1.12) in \mathbf{R} with shape parameter $k = 30$ and rate parameter $r = 10$. (The value $k = 30$ was used by Neufeld et al. (2007), Neufeld et al. (2010), and a value of r approximately equal to 10 was estimated in those references.) For the i^{th} participant in a cell, N encoding times t_{i1}, \dots, t_{iN} were then generated according to the gamma distribution (1.21) with shape parameter $k'_i \alpha$ and rate parameter v_i . The sample mean \bar{t}_i , the sample variance s_i^2 , and then the estimate $\hat{x}_i = \bar{t}_i^2 / s_i^2$ (see (1.41)) were then computed for each participant in the cell. The number of encoding trials per participant was taken to be very

large; specifically $N = 10,000$. This is not a reasonable number in a real-world experiment. However, the estimates \hat{x}_i can be viewed as having two components, a “signal” component (usually confined to the first few decimal places) and a “noise” component (the remainder of the estimate). Large values of N were required to separate the signal from the noise in some cases. The method of signal detection (and the subsequent estimation of α) is best illustrated by example.

Example 2.1: $\alpha = .01$: Estimation within each of the first three cells will proceed sequentially.

1. Cell One: $p = .8$: The estimated values \hat{x} for each of the 30 participants are displayed in Table 2.1 to four decimal digits. The rounding of \hat{x} to one decimal digit is also displayed (columns three and seven) and to two decimal digits (columns four and eight).

Table 2.1: $\alpha = .01, p = .8$

n	\hat{x}	1 dec	2 dec	n	\hat{x}	1 dec	2 dec
1	.0185	0	.02	16	.0082	0	.01
2	.0100	0	.01	17	.0114	0	.01
3	.0270	0	.03	18	.0106	0	.01
4	.0339	0	.03	19	.0098	0	.01
5	.0099	0	.01	20	.0183	0	.02
6	.0103	0	.01	21	.0117	0	.01
7	.0110	0	.01	22	.0417	0	.04
8	.0096	0	.01	23	.0118	0	.01
9	.0105	0	.01	24	.0177	0	.02
10	.0106	0	.01	25	.0110	0	.01
11	.0092	0	.01	26	.0317	0	.03
12	.0094	0	.01	27	.0079	0	.01
13	.0102	0	.01	28	.0105	0	.01
14	.0220	0	.02	29	.0083	0	.01
15	.0094	0	.01	30	.0100	0	.01

Note that rounding to one decimal digit produces only 0s, which implies that $\alpha < .1$. Thus the values rounded to two decimal places are considered. Here $x_{\min} = .01$ and division by .01

obviously produces integers for all of the rounded values, so $\alpha_{\max} = .01$. Consider the resulting constructed integers j_1, \dots, j_{30} (see (1.34)):

2 1 3 3 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 4 1 2 1 3 1 1 1 1

If these integers appear to fit a geometric distribution, this supports the idea that $\alpha = .01$ (in other words, rounding to two decimal digits captures the signal). In order to determine whether the integers fit a geometric distribution, a chi-square goodness-of-fit test was employed (Wonnacott & Wonnacott, 1972). The estimate \hat{p}_1 based on these integers was $\hat{p}_1 = 1/\bar{j} = .698$. Three bins were used for the chi-square: $\{X = 1\}$, $\{X = 2\}$, and $\{X \geq 3\}$. The expected values in the bins (using \hat{p}_1) are $E_1 = 20.94$, $E_2 = 6.32$, and $E_3 = 2.74$. (Note that normally it is desired that expected counts in a bin be at least 5, but in this case having one very small bin was required in order not to have 0 degrees of freedom (df) for the chi-square test.) The observed bin counts are $O_1 = 22$, $O_2 = 4$, and $O_3 = 4$. Thus

$$\chi_{\text{obs}}^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{(22 - 20.94)^2}{20.94} + \frac{(4 - 6.32)^2}{6.32} + \frac{(4 - 2.74)^2}{2.74} = 1.485 \quad (2.1)$$

Since $1.485 < 3.84 = \chi_{1,.05}^2$ it follows that the hypothesis that the integers are geometric is not rejected. In other words, there appears to be a good fit to the geometric. At this stage, rounding to three decimal places can be considered to see if a larger value $\alpha > .01$ will also provide a good geometric fit. However, inspecting the data to three decimal places shows several observations which are less than .01 in size, meaning that a larger α cannot work. Thus it is sufficient to stop at two decimal places and set $\hat{\alpha} = .01$ and $\hat{p}_1 = .698$.

2. Cell Two: $p = .2$: The assumption that α is the same in all cells implies that a starting point for $\hat{\alpha}$ in Cell Two is $\hat{\alpha} = .01$. Theorem 1.5.2 indicates that it is possible that $\hat{\alpha}$ may

decrease as a result of information from Cell Two. Ideally, however, Cell Two will support the conclusion from Cell One. The table below shows the estimates \hat{x} and their rounded values.

Table 2.2: $\alpha = .01, p = .2$

n	\hat{x}	1 dec	2 dec	n	\hat{x}	1 dec	2 dec
1	.0297	0	.03	16	.0183	0	.02
2	.0395	0	.04	17	.0104	0	.01
3	.0083	0	.01	18	.0629	.1	.06
4	.0628	.1	.06	19	.0379	0	.04
5	.0481	0	.05	20	.0845	.1	.08
6	.0721	.1	.07	21	.0431	0	.04
7	.0379	0	.04	22	.0110	0	.01
8	.0392	0	.04	23	.1740	.2	.17
9	.0081	0	.01	24	.1034	.1	.10
10	.0781	.1	.08	25	.0132	0	.01
11	.0313	0	.03	26	.0937	.1	.09
12	.0163	0	.02	27	.0670	.1	.07
13	.0608	.1	.06	28	.1293	.1	.13
14	.1229	.1	.12	29	.0296	0	.03
15	.0496	0	.05	30	.0423	0	.04

Note $\alpha_{\max} = .01$ for the values rounded to two decimal places, and the constructed set of integers (1.34) is

3 4 1 6 5 7 4 4 1 8 3 2 6 12 5 2 1 6 4 8 4 1 17 10 1 9 7 13 3 4

Then $\hat{p}_2 = 1/\bar{j} = .186$. Five bins were selected $\{X = 1\}$, $\{X = 2, 3\}$, $\{X = 4, 5\}$, $\{X = 6, 7, 8, 9\}$, and $\{X \geq 10\}$. Expected bin counts are $E_1 = 5.58$, $E_2 = 8.24$, $E_3 = 5.46$, $E_4 = 6.01$, and $E_5 = 4.7$, whereas observed bin counts are $O_1 = 5$, $O_2 = 5$, $O_3 = 8$, $O_4 = 8$, and $O_5 = 4$. Then

$$\chi_{\text{obs}}^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = 3.279 \quad (2.2)$$

Since $3.279 < \chi_{3,.05}^2 = 7.81$ the data supports the geometric distribution and $\hat{\alpha} = .01$. Since $\hat{\alpha}$ cannot exceed .01 (from the results of Cell One) it is possible to stop at this step. However,

this conclusion is also supported by considering the points $\hat{x}_3 = .0083$ and $\hat{x}_9 = .0081$. Thus $\hat{\alpha} = .01$ and $\hat{p}_2 = .186$.

3. Cell Three: $p = .05$: The table below has the \hat{x} values and their rounded values:

Table 2.3: $\alpha = .01, p = .05$

n	\hat{x}	1 dec	2 dec	n	\hat{x}	1 dec	2 dec
1	.0114	0	.01	16	.3647	.4	.36
2	.0094	0	.01	17	.4701	.5	.47
3	.9505	1	.95	18	.1082	.1	.11
4	.0478	0	.05	19	.0297	0	.03
5	.0116	0	.01	20	.2735	.3	.27
6	.0375	0	.04	21	.4535	.5	.45
7	.1687	.2	.17	22	.1793	.2	.18
8	.1297	.1	.13	23	.3704	.4	.37
9	.1333	.1	.13	24	.5620	.6	.56
10	.1461	.1	.15	25	.1431	.1	.14
11	.0494	0	.05	26	.0323	0	.03
12	.1177	.1	.12	27	.1110	.1	.11
13	.3650	.4	.36	28	.4198	.4	.42
14	.2053	.2	.21	29	.0194	0	.02
15	.6412	.6	.64	30	.3370	.3	.34

Again rounding to two decimal places gives $\alpha_{\max} = .01$, and the constructed integers are

1 1 95 5 1 4 17 13 13 15 5 12 36 21 64 36 47 11 3 27 45 18 37 56 14 3 11 42 2 34

Here $\hat{p}_3 = 1/\bar{j} = .044$. The selected bins were $\{1 \leq X \leq 5\}$, $\{6 \leq X \leq 15\}$, $\{16 \leq X \leq 35\}$, and $\{X \geq 36\}$. Expected bin counts are $E_1 = 6.03$, $E_2 = 8.67$, $E_3 = 9.06$, and $E_4 = 6.21$, whereas the observed bin counts are $O_1 = 9$, $O_2 = 7$, $O_3 = 5$, and $O_4 = 9$. This yields $\chi_{\text{obs}}^2 = 4.857 < 5.99 = \chi_{2,.05}^2$ and so this supports a good fit to the geometric distribution. Thus $\hat{\alpha} = .01$ and $\hat{p}_3 = .044$.

4. Cell Four: additivity Here the estimate of p_4 would be generated according to Theorem

1.5.2; specifically

$$\hat{p}_4 = \frac{(.698)(.186)(.044)}{(.698)(.186) + (.698)(.044) - (.186)(.044)} = .037 \quad (2.3)$$

Example 2.2: $\alpha = .17$: The following example shows that this choice for α leads to some different steps.

1. Cell One: $p = .8$: Considering the table below, there are no 0s when \hat{x} is rounded to one decimal place, so a large value of α (i.e. $\alpha \geq .1$) is possible. However, given the observations rounded to one decimal place, it is evident that $\alpha_{\max} = .1$.

Table 2.4: $\alpha = .17, p = .8$

n	\hat{x}	1 dec	2 dec	target	n	\hat{x}	1 dec	2 dec	target
1	.3394	.3	.34		16	.1848	.2	.18*	(.17)
2	.1688	.2	.17		17	.1696	.2	.17	
3	.1668	.2	.17		18	.1689	.2	.17	
4	.1709	.2	.17		19	.1694	.2	.17	
5	.3353	.3	.34		20	.1737	.2	.17	
6	.1698	.2	.17		21	.5065	.5	.51	
7	.3438	.3	.34		22	.1703	.2	.17	
8	.1653	.2	.17		23	.1666	.2	.17	
9	.1676	.2	.17		24	.1581	.2	.16*	(.17)
10	.1687	.2	.17		25	.1700	.2	.17	
11	.3489	.3	.35*	(.34)	26	.1702	.2	.17	
12	.3449	.3	.34		27	.1675	.2	.17	
13	.1737	.2	.17		28	.1639	.2	.16*	(.17)
14	.1735	.2	.17		29	.3444	.3	.34	
15	.1622	.2	.16*	(.17)	30	.1741	.2	.17	

The constructed sample using $\alpha_{\max} = .1$ is

3 2 2 2 3 2 3 2 2 2 3 3 2 2 2 2 2 2 2 2 5 2 2 2 2 2 2 3 2

It would not be a surprise if this sample were not geometric because it clearly has a small variance but a 1 does not appear anywhere in the sequence. Computing gives $\hat{p}_1 = \bar{j} = .435$.

Using this and the bins $\{X = 1\}$, $\{X = 2\}$, and $\{X \geq 3\}$, yields the expected values $E_1 = 13.05$, $E_2 = 7.38$, $E_3 = 9.57$, and the observed bin counts $O_1 = 0$, $O_2 = 23$, and $O_3 = 7$. The resulting $\chi_{\text{obs}}^2 = 46.8 \gg 10.8 = \chi_{1,.001}^2$ which leads to strong rejection of the geometric distribution. Thus $\alpha = .1$ is rejected and the \hat{x} are now considered rounded to two decimal places. At this point the reader may notice a slightly different format to Table 2.4 than in the previous example; specifically, values marked with an asterisk (*) and a column of “target” values. These values were constructed by noting that the vast majority of entries in the 2-decimal column were multiples of .17. The few that were not (those marked with an asterisk) were very close to being multiples of .17. The target values were the closest values that were actual multiples of .17, and the deviation from them was considered to be rounding error.

The constructed sample of integers was based on those 2-decimal values that were multiples of .17 and on the target values (that is, the starred values were replaced by the target values in construction of the integers). Thus the constructed sample is

2 1 1 1 2 1 2 1 1 1 2 2 1 1 1 1 1 1 1 3 1 1 1 1 1 1 2 1

Here $\hat{p}_1 = 1/\bar{j} = .789$. Using the bins $\{X = 1\}$, $\{X = 2\}$, and $\{X \geq 3\}$ yields $E_1 = 23.67$, $E_2 = 4.98$, and $E_3 = 1.35$. (Again a very small bin has been used in order to avoid 0 df.) The observed counts are $O_1 = 23$, $O_2 = 6$, and $O_3 = 1$. The resulting $\chi_{\text{obs}}^2 = .319 \ll 3.84 = \chi_{1,.05}^2$ which supports the geometric distribution. Inspecting the third decimal place shows that α cannot exceed .17 so the process can stop here.

2. Cell Two: $p = .2$: Note the increase in size of the entries (as well as the increase in the number of starred values) in Table 2.5. This is a consequence of a combination of the decrease in p with a rather large value of α , thereby shifting the decimal place. The results of

the previous cell suggest that $\alpha = .17$ and strongly reject the possibility that $\alpha = .1$. Therefore it would not be necessary (and even appropriate) to look at rounding to one decimal place and trying $\alpha_{\max} = .1$. However, it is worth noting that if one does this, the constructed sample is

2 20 7 2 5 4 5 9 5 8 10 10 10 3 9 3 2 12 2 12 2 27 7 5 7 14 8 12 10 7

and the geometric is *not* rejected for the bins $\{X = 1, 2\}$, $\{3 \leq X \leq 5\}$, $\{6 \leq X \leq 10\}$, and $\{X \geq 11\}$. It produces $\chi_{\text{obs}}^2 = 3.776 < 5.99 = \chi_{2,.05}^2$. This illustrates the importance of moving to the next decimal place and seeing if an adequate fit can be found with a larger α . In this example, the value $\alpha = .17$ has already been suggested by Cell One.

Table 2.5: $\alpha = .17, p = .2$

n	\hat{x}	1 dec	2 dec	target	n	\hat{x}	1 dec	2 dec	target
1	.1718	.2	.17		16	.3439	.3	.34	
2	1.9780	2	1.98*	(2.04)	17	.1689	.2	.17	
3	.6893	.7	.69*	(.68)	18	1.2104	1.2	1.21*	(1.19)
4	.1668	.2	.17		19	.1639	.2	.16*	(.17)
5	.5149	.5	.51		20	1.1975	1.2	1.20*	(1.19)
6	.3501	.4	.35*	(.34)	21	.1669	.2	.17	
7	.5105	.5	.51		22	2.7043	2.7	2.70*	(2.72)
8	.8523	.9	.85		23	.6800	.7	.68	
9	.5117	.5	.51		24	.5073	.5	.51	
10	.8482	.8	.85		25	.6753	.7	.68	
11	1.0375	1	1.04*	(1.02)	26	1.3575	1.4	1.36	
12	1.0338	1	1.03*	(1.02)	27	.8394	.8	.84*	(.85)
13	1.0134	1	1.01*	(1.02)	28	1.1878	1.2	1.19	
14	.3416	.3	.34		29	1.0199	1	1.02	
15	.8520	.9	.85		30	.6869	.7	.69*	(.68)

Replacing the starred values by the target values yields the constructed sample

1 12 4 1 3 2 3 5 3 5 6 6 6 2 5 2 1 7 1 7 1 16 4 3 4 8 5 7 6 4

Then $\hat{p}_3 = .214$ and using the bins $\{X = 1\}$, $\{X = 2, 3\}$, $\{X = 4, 5, 6\}$, $\{X \geq 7\}$ shows a good fit to the geometric. Thus $\hat{\alpha} = .17$.

3. Cell Three: $p = .05$: Table 2.6 clearly shows that there is more rounding error in the \hat{x} data. Nonetheless, by using the fact that the previous cells suggested $\alpha = .17$ and inspecting the rounded data closely, the pattern of multiples of $.17$ can be seen. Obtaining the constructed sample (not shown here as the method is the same as in the other cells) produces a good fit to the geometric distribution for $\hat{p} = .048$ and $\hat{\alpha} = .17$.

Table 2.6: $\alpha = .17, p = .05$

n	\hat{x}	1 dec	2 dec	target	n	\hat{x}	1 dec	2 dec	target
1	2.0388	2	2.04		16	6.4932	6.5	6.49*	(6.46)
2	6.8843	6.9	6.88*	(6.80)	17	1.1897	1.2	1.19	
3	2.8786	2.9	2.88*	(2.89)	18	.6805	.7	.68	
4	3.4337	3.4	3.43*	(3.40)	19	4.7299	4.7	4.73*	(4.76)
5	2.3750	2.4	2.38		20	.5090	.5	.51	
6	3.2403	3.2	3.24*	(3.23)	21	8.9972	9.0	9.00*	(9.01)
7	4.4132	4.4	4.41*	(4.42)	22	4.0587	4.1	4.06*	(4.08)
8	.3432	.3	.34		23	1.5416	1.5	1.54*	(1.53)
9	3.0509	3.1	3.05*	(3.06)	24	10.1482	10.1	10.15*	(10.20)
10	3.4346	3.4	3.43*	(3.40)	25	4.5766	4.6	4.58*	(4.59)
11	.8494	.8	.85		26	3.2146	3.2	3.21*	(3.23)
12	.8456	.8	.85		27	2.2134	2.2	2.21	
13	1.5450	1.5	1.55*	(1.53)	28	1.6987	1.7	1.70	
14	2.1734	2.2	2.17*	(2.21)	29	9.1720	9.2	9.17*	(9.18)
15	1.0206	1.0	1.02		30	9.0183	9.0	9.02*	(9.01)

Concluding Remark: Assuming that sufficient data can be obtained so that some of the signal of the x data can be isolated, the above technique (which utilizes careful data inspection plus the algorithm to obtain α_{\max}) can recover an accurate estimate of α .

Bibliography

- Bluhm, R. L., Miller, J., Lanius, R. A., Osuch, E. A., Boksman, K., & Neufeld, R. W. J. (2007). Spontaneous low frequency fluctuations in the BOLD signal in schizophrenia patients: Anomalies in the default network. *Schizophrenia Bulletin*, *33*, 1004-1012.
- Braver, T. S., & Barch, D. M. (2006). Extracting core components of cognitive control. *Trends in Cognitive Science*, *10*, 529-532.
- Chandler, J. P. (1959). STEPIT: Finds local minima of a smooth function of several parameters. *Behavioral Science*, *14*, 81-82.
- Chow, Y. S., & Teicher, H. (1988). *Probability theory: Independence, interchangeability, martingales, 2nd Ed.* New York: Springer-Verlag.
- Cromwell, R. L., & Dokecki, P. (1968). Schizophrenic language: A disattention interpretation. In S. Rosenberg & J. H. Koplín (Eds.), *Developments in applied psycholinguistic research* (pp. 209-260). New York: Macmillan.
- Donders, F. C. (1969). Over die snelheid van psychische processen. [On the rapidity of mental processes] (W. G. Koster, Trans.) *Acta Psychologica*, *30*, 412-431. (Originally published in 1868).
- George, L., & Neufeld, R. W. J. (1987). Attentional resources and hemispheric functional asymmetry in schizophrenia. *British Journal of Clinical Psychology*, *26*, 35-45.
- Hemsley, D. R. (1993). Perception and cognition in schizophrenia. In R. L. Cromwell & C. R. Snyder (Eds.), *Schizophrenia: Origins, processes, treatment and outcome* (pp. 135-150).

New York: Oxford University Press.

Hemsley, D. R. (1994). Perceptual and cognitive normality as the bases for schizophrenic symptoms. In A. S. David & J. Cutting (Eds.), *The neuropsychology of schizophrenia* (pp. 97-116). Hillsdale, NJ: Erlbaum.

Highgate, S., & Neufeld, R. W. J. (1986). Schizophrenic memory-search performance involving nonverbal stimulus properties. *Journal of Abnormal Psychology, 95*, 67-73.

Hogg, R. V., McKean, J. W., & Craig, A. T. (2005). *Introduction to mathematical statistics, 6th Ed.* Upper Saddle River, NJ: Pearson Prentice Hall.

Kuritsyn, Y. G. (1987). On monotonicity in the law of large numbers for exchangeable random variables. *Theory of Probability and its Applications, 29*(1), 150-153.

Murphy, K., Birn, R. M., Handwerker, D. A., & Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *Neuroimage, 44*, 893-905.

Neufeld, R. W. J. (2007a). On the centrality and significance of encoding deficit in schizophrenia. *Schizophrenia Bulletin, 33*, 982-983.

Neufeld, R. W. J. (2007b). Composition and uses of formal clinical cognitive science. In B. Shuart, W. Spaulding, & J. Poland (Eds.), *Modeling complex symptoms: Nebraska symposium on motivation: Vol 52* (pp. 1-83). Lincoln, Nebraska: University of Nebraska.

Neufeld, R. W. J., Boksman, K., Vollick, D., George, L., & Carter, J. R. (2010). Stochastic dynamics of stimulus encoding in schizophrenia: Theory, testing, and application. *Journal*

of Mathematical Psychology, 54, 90-108.

Neufeld, R. W. J., & McCarty, T. S. (1994). A formal analysis of stressor and stress-proneness effects on simple information processing. *British Journal of Mathematical and Statistical Psychology*, 47, 193-226.

Neufeld, R. W. J., Vollick, D., Carter, J. R., Boksman, K., & Jetté, J. (2002). Applications of stochastic modelling to the assessment of group and individual differences in cognitive functioning. *Psychological Assessment*, 14, 279-298.

Neufeld, R. W. J., Vollick, D., Carter, J. R., Boksman, K., Levy, L. R., George, L., & Jetté, J. (2007). A mathematical process account of group and individual differences in memory-search facilitative stimulus encoding, with application to schizophrenia. In R. W. J. Neufeld (Ed.), *Advances in clinical cognitive science: Formal modeling of processes and symptoms* (pp. 147-177). Washington, DC: American Psychological Association.

Neufeld, R. W. J., Vollick, D., & Highgate, S. (1993). Stochastic modelling of stimulus encoding and memory search in paranoid schizophrenia: clinical and theoretical implications. In R. L. Cromwell & C. R. Snyder (Eds.), *Schizophrenia: Origins, processes, treatment, and outcome* (pp. 176-198). New York: Oxford University Press.

Neufeld, R. W. J., & Williamson, P. C. (1996). Neuropsychological correlates of positive symptoms: Delusions and hallucinations. In C. Pantelis, H. E. Nelson, & T. R. E. Barnes (Eds.), *Schizophrenia: A neuropsychological perspective* (pp. 205-235). New York: John Wiley & Sons Ltd.

Nicholson, I. R., & Neufeld, R. W. J. (1993). The classification of the schizophrenias accord-

- ing to symptomatology: A two-factor model. *Journal of Abnormal Psychology*, *102*, 259-270.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>
- Russell, P. N., & Knight, R. G. (1977). Performance of process schizophrenics on tasks involving visual search. *Journal of Abnormal Psychology*, *86*, 16-26.
- Steffy, R. A., & Galbraith, K. (1980). Relation between latency and redundancy-associated deficit in schizophrenic reaction time performance. *Journal of Abnormal Psychology*, *89*, 419-427.
- Steffy, R. A., & Waldman, I. (1993). Schizophrenics' reaction time: North star or shooting star? In R. L. Cromwell & C. R. Snyder (Eds.), *Schizophrenia: Origins, processes, treatment and outcome* (pp. 111-134). New York: Oxford University Press.
- Sternberg, S. (1975). Memory and scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, *27*, 1-32.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modelling of elementary psychological processes*. Cambridge, England: Cambridge University Press.
- Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, *11*(3), 391-418.
- Wonnacott, T. H., & Wonnacott, R. J. (1972). *Introductory statistics, 2nd Ed.* New York: John Wiley & Sons, Inc.

Woodworth, R. S., & Schlossberg, H. (1954). *Experimental psychology*. New York: Holt, Rinehart & Winston.

Appendix A

Sampling Schematic

Sampling Method For One Cell of Participants

	Trial 1	Trial 2	...	Trial N		Sample mean	Sample variance
Participant 1	z_{11}	z_{12}	...	z_{1N}	→	\bar{z}_1	u_1^2
Participant 2	z_{21}	z_{22}	...	z_{2N}	→	\bar{z}_2	u_2^2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Participant M	z_{M1}	z_{M2}	...	z_{MN}	→	\bar{z}_M	u_M^2

Then average the sample means and sample variances over the participants to obtain

$$\bar{z} = \frac{1}{M} \sum_{i=1}^M \bar{z}_i$$

and

$$\bar{u}^2 = \frac{1}{M} \sum_{i=1}^M u_i^2.$$

Appendix B

Proofs of Results

Proof of Proposition 1.3.3: Recall that m^* takes the form $m^* = \alpha E(k')$ within cell (*), where $E(k')$ varies over the cells. Suppose solutions α, m, g, h have been found such that $m_1^* = \alpha m$, $m_2^* = \alpha(m + h)$, $m_3^* = \alpha(m + g)$ and $m_4^* = \alpha(m + g + h)$ solves (1.28) and (1.29). Then, using the fill property of $E(k')$, there exists $m_0 > m$, $h_0 > h$, $g_0 > g$ and $0 < \alpha_0 < \alpha$ such that $m_1^* = \alpha_0 m_0$, $m_2^* = \alpha_0(m_0 + h_0)$, and $m_3^* = \alpha_0(m_0 + g_0)$. Moreover, additivity holds since $m_4^* = m_3^* + m_2^* - m_1^* = \alpha_0(m_0 + g_0) + \alpha_0(m_0 + h_0) - \alpha_0 m_0 = \alpha_0(m_0 + g_0 + h_0)$. ■

Proof of Theorem 1.4.1.1: Suppose there exists $\alpha > 0$, $\beta > 0$ and sequences k'_1, k'_2, \dots and k''_1, k''_2, \dots satisfying the assumptions of Theorem 1.4.1.1. Then without loss of generality it can be assumed that $\alpha \leq \beta$. Since $P(k')$ is a positive distribution, there exists a smallest integer $k'_{\min} = \min\{k' \geq 1 \mid P(k') > 0\}$. With probability one this integer k'_{\min} must eventually appear in the k'_1, k'_2, \dots sequence, say, $k'_m = k'_{\min}$. Then x_m has the two representations

$$x_m = k'_{\min} \alpha = k''_m \beta. \quad (\text{B.1})$$

Since $k'_{\min} \leq k''_m$ and $\alpha \leq \beta$, the only way the equality in (B.1) can hold is if $\alpha = \beta$, which completes the proof. ■

Proof of Proposition 1.4.1.2: Let A be as defined in (1.33). Then $A \neq \emptyset$ since $\alpha \in A$. Moreover,

A actually contains infinitely many elements since if $\beta \in A$ then $\beta/n \in A$ for each positive integer n . Let $\alpha_{\max} = \sup A$ where “sup” denotes the supremum over all the elements of A . The goal is to show that $\alpha_{\max} \in A$. Since $\alpha_{\max} = \sup A$ it follows that there exists a sequence $\beta_n \in A$ such that $\beta_n \uparrow \alpha_{\max}$. Thus x_1/β_n forms a decreasing sequence of integers bounded below by 1. Since the integers are well-ordered, there exists an integer $i_0 \geq 1$ and an integer $n_0 \geq 1$ such that $x_1/\beta_n = i_0$ for all $n \geq n_0$. Hence $\beta_n = \beta_{n_0}$ for all $n \geq n_0$. It follows that $\alpha_{\max} = \beta_{n_0}$ and hence belongs to A . ■

Proof of Theorem 1.4.1.5: The “only if” part of this theorem follows from Proposition 1.4.1.3 which states that the set (1.34) is always coprime, and hence $\alpha_{\max} = \alpha$ implies that k'_1, \dots, k'_M are coprime. Now consider the “if” part of the proof. It is automatic that $\alpha_{\max} \geq \alpha$. The goal is to show that if k'_1, \dots, k'_M is coprime then $\alpha_{\max} = \alpha$. The procedure will be a proof by contradiction; namely that if $\alpha_{\max} > \alpha$ then k'_1, \dots, k'_M is *cannot* be coprime. Now under the assumption $\alpha_{\max} > \alpha$ there exist positive integers m and n such that

$$\frac{x_1}{\alpha_{\max}} = n \quad \text{and} \quad \frac{x_1}{\alpha} = m \quad \text{where } m > n \quad (\text{B.2})$$

Consequently $\alpha_{\max}/\alpha = m/n$. Now writing the two representations

$$\begin{aligned} x_1 &= k'_1 \alpha, \dots, x_M = k'_M \alpha \\ x_1 &= k''_1 \alpha_{\max}, \dots, x_M = k''_M \alpha_{\max} \end{aligned} \quad (\text{B.3})$$

it follows that $k'_i \alpha = k''_i \alpha_{\max}$ and hence $k'_i = (\alpha_{\max}/\alpha)k''_i = (m/n)k''_i$ for all $i = 1, \dots, M$. Now divide out the greatest common divisor (gcd) from m and n to produce integers $m_0 > n_0$ such that $\gcd(m_0, n_0) = 1$. Thus $k'_i = (m_0/n_0)k''_i$ for all $i = 1, \dots, M$. Since m_0 and n_0 are coprime and n_0 divides the product $m_0 k''_i$ it follows that n_0 divides k''_i . Hence there exist integers $d_i \geq 1$,

$i = 1, \dots, M$ such that

$$k'_1 = m_0 d_1, k'_2 = m_0 d_2, \dots, k'_M = m_0 d_M. \quad (\text{B.4})$$

Note that the initial assumption $m > n$ implies $m_0 \geq 2$. Thus from (B.4) it is evident that k'_1, \dots, k'_M are not coprime. ■

Proof of Theorem 1.4.2.1: It is important to note that the sample of encoding times $t_{i1}, t_{i2}, \dots, t_{iN}$ does not consist of independent random variables because the variables are connected through the specific and unknown values k'_i and v_i corresponding to the i^{th} participant. In general, if t_1, t_2, \dots is a random sequence of encoding times obtained from a single individual (with unknown parameters k' and v), then t_1, t_2, \dots comprise what is called an *exchangeable* sequence of random variables. A sequence t_1, t_2, \dots is exchangeable if, for each positive integer n , the joint distribution of any subset t_{j_1}, \dots, t_{j_n} depends only on n and not on the subset (e.g., Chow & Teicher, 1988). (In other words, the finite-dimensional distributions of the sequence are invariant under permutations.) A sequence of random variables is exchangeable if and only if there exists a σ -algebra \mathcal{G} of events such that the random variables are conditionally independent given \mathcal{G} (de Finetti's theorem; see Corollary 4, p. 226 of Chow & Teicher, 1988). In the present case \mathcal{G} is clearly the σ -algebra generated by k' and v , i.e., if k_i and v_i are known, then the variables become independent. Strong laws of large numbers hold for exchangeable sequences as they do for independent identically-distributed sequences. However, a significant difference between the results for exchangeable sequences and those for independent sequences is that the former do not converge to constants but rather to values that depend on the specific sequence. Combining the first theorem of Kuritsyn (1987) and Theorem 2 (p. 224) of Chow and Teicher (1988) yields the following

Lemma Let t_1, t_2, \dots be an exchangeable sequence of random variables which is conditionally independent given the σ -algebra \mathcal{G} . Suppose $E(|t_1|^n) < \infty$ where $n \geq 1$. Let

$$\bar{t}_N(\omega) = \frac{1}{N} \sum_{i=1}^N t_i(\omega) \quad (\text{B.5})$$

Then

$$\lim_{N \rightarrow \infty} \bar{t}_N(\omega) = E(t_1 | \mathcal{G})(\omega) \quad \text{almost surely} \quad (\text{B.6})$$

The above result can be applied to prove Theorem 1.4.2.1 by noting that the assumption that $P(k')$ has finite mean and variance can be used to prove that t_1 has finite mean and variance.

Specifically, the independence of k' and v yields

$$E(t_1) = E[E(t_1 | k', v)] = E[k'/v] = E(k')E(1/v) \quad (\text{B.7})$$

and

$$\begin{aligned} \text{Var}(t_1) &= E[\text{Var}(t_1 | k', v)] + \text{Var}[E(t_1 | k', v)] \\ &= E[k'/v^2] + \text{Var}(k'/v) \\ &= E(k')E(1/v^2) + E((k')^2)E(1/v^2) - [E(k')E(1/v)]^2 \end{aligned} \quad (\text{B.8})$$

Thus both t_1 and t_1^2 are integrable. Now noting that the sequence t_1^2, t_2^2, \dots is also exchangeable and that the sample variance can be expressed as $s_N^2 = \frac{1}{N-1}(\sum_{i=1}^N t_i^2 - N\bar{t}_N^2)$, the Lemma can be applied to both \bar{t} and s^2 to conclude that (1.39) and (1.40) hold, thereby proving the theorem.

■

Proof of Theorem 1.5.1: The maximum likelihood estimate (MLE) of p is first considered.

It will be simpler to maximize the log likelihood $\log L$ rather than L itself. Note that for an admissible α

$$\log L = \left(\frac{\sum x_i}{\alpha} - M \right) \log(1-p) + M \log p \quad (\text{B.9})$$

Thus

$$\frac{\partial \log L}{\partial p} = \left(\frac{\sum x_i}{\alpha} - M \right) \left(\frac{-1}{1-p} \right) + \frac{M}{p}. \quad (\text{B.10})$$

Setting $\frac{\partial \log L}{\partial p} = 0$ gives

$$\frac{M(1-p) - p \frac{\sum x_i}{\alpha} + pM}{p(1-p)} = 0 \quad (\text{B.11})$$

which easily yields the MLE of p in terms of α as

$$\hat{p} = \frac{\alpha}{\bar{x}} \quad (\text{B.12})$$

Now substitute $\hat{p} = \alpha/\bar{x}$ back into $\log L$ and maximize over α . In this case L is only nonzero for admissible values of α but it will be convenient to temporarily treat L as continuous in α .

Note that

$$\log L = \left(\frac{M\bar{x}}{\alpha} - M \right) \log \left(1 - \frac{\alpha}{\bar{x}} \right) + M \log \left(\frac{\alpha}{\bar{x}} \right) \quad (\text{B.13})$$

Then

$$\begin{aligned} \frac{\partial \log L}{\partial \alpha} &= M \left\{ \left(\frac{\bar{x}}{\alpha} - 1 \right) \frac{1}{\left(1 - \frac{\alpha}{\bar{x}} \right)} \left(-\frac{1}{\bar{x}} \right) + \log \left(1 - \frac{\alpha}{\bar{x}} \right) \left(-\frac{\bar{x}}{\alpha^2} \right) + \frac{1}{\bar{x}} \frac{1}{\bar{x}} \right\} \\ &= -M \frac{\bar{x}}{\alpha^2} \log \left(1 - \frac{\alpha}{\bar{x}} \right) \end{aligned} \quad (\text{B.14})$$

Note that any admissible value of α must satisfy $\alpha \leq x_{\min}$ (otherwise $L = 0$). Note from the second line of (B.14) that $\frac{\partial \log L}{\partial \alpha}$ is positive for $0 < \alpha \leq x_{\min}$. Hence L is increasing as a function of (admissible) α and so

$$\hat{\alpha} = \sup\{\alpha \mid \alpha \text{ is admissible, } \alpha \leq x_{\min}\} \quad (\text{B.15})$$

The right hand side of (B.15) is clearly the same as (1.33) and so $\hat{\alpha} = \alpha_{\max}$. ■

Proof of Theorem 1.5.2: First consider just the first two cells. Let $x_1^{(1)}, \dots, x_M^{(1)}$ denote the x sample in the first cell, and let $x_1^{(2)}, \dots, x_M^{(2)}$ denote the x sample in the second cell. Note that from independence of the cells the joint likelihood factors as the product of the individual likelihoods

$$L(\alpha, p_1, p_2; x_1^{(1)}, \dots, x_M^{(1)}, x_1^{(2)}, \dots, x_M^{(2)}) = (1 - p_1)^{\frac{\sum x_i^{(1)}}{\alpha} - M} p_1^M (1 - p_2)^{\frac{\sum x_i^{(2)}}{\alpha} - M} p_2^M \quad (\text{B.16})$$

and hence

$$\log L = \left(\frac{\sum x_i^{(1)}}{\alpha} - M \right) \log(1 - p_1) + M \log p_1 + \left(\frac{\sum x_i^{(2)}}{\alpha} - M \right) \log(1 - p_2) + M \log p_2 \quad (\text{B.17})$$

Differentiating $\log L$ with respect to p_1 does not involve p_2 and produces the same solution \hat{p}_1 obtained in (B.12), i.e., $\hat{p}_1 = \alpha/\bar{x}_1$. Similarly, differentiating with respect to p_2 yields the solution $\hat{p}_2 = \alpha/\bar{x}_2$. Now note that a value of α is admissible only if it is admissible for both $x_1^{(1)}, \dots, x_M^{(1)}$ and $x_1^{(2)}, \dots, x_M^{(2)}$ (otherwise $L = 0$). Differentiating with respect to α produces

$$\frac{\partial \log L}{\partial \alpha} = -M \frac{\bar{x}_1}{\alpha^2} \log \left(1 - \frac{\alpha}{\bar{x}_1} \right) - M \frac{\bar{x}_2}{\alpha^2} \log \left(1 - \frac{\alpha}{\bar{x}_2} \right) \quad (\text{B.18})$$

Thus L is an increasing function of (admissible) α for $0 < \alpha \leq \min(x_{\min}^{(1)}, x_{\min}^{(2)})$. It follows that

$$\hat{\alpha} = \max \left(\alpha > 0 \mid \frac{x_1^{(1)}}{\alpha}, \dots, \frac{x_M^{(1)}}{\alpha}, \frac{x_1^{(2)}}{\alpha}, \dots, \frac{x_M^{(2)}}{\alpha} \text{ are integers} \right) \quad (\text{B.19})$$

where the maximum exists and belongs to the set by the same argument used in the proof of Proposition 1.4.1.2. The result extends by induction to any number of independent cells, so this proves the first part of the theorem.

Now consider additivity. $E(k') = 1/p$ for the appropriate p in a given cell. Since $p_2 < p_1$ and $p_3 < p_1$ (reflecting a larger number of subprocesses in Cells Two and Three as compared to Cell One), there exist $m > 0$, $h > 0$, and $g > 0$ such that

$$m = \frac{1}{p_1}, \quad m + h = \frac{1}{p_2}, \quad m + g = \frac{1}{p_3} \quad (\text{B.20})$$

Additivity requires that $m + g + h = 1/p_4$. Now

$$h = \frac{1}{p_2} - \frac{1}{p_1} \quad \text{and} \quad g = \frac{1}{p_3} - \frac{1}{p_1} \quad (\text{B.21})$$

Therefore, additivity will hold if

$$\frac{1}{p_1} + \left(\frac{1}{p_3} - \frac{1}{p_1} \right) + \left(\frac{1}{p_2} - \frac{1}{p_1} \right) = \frac{1}{p_4} \quad (\text{B.22})$$

that is, if

$$\frac{1}{p_3} + \frac{1}{p_2} - \frac{1}{p_1} = \frac{1}{p_4} \quad (\text{B.23})$$

Solving for p_4 yields

$$p_4 = \frac{p_1 p_2 p_3}{p_1 p_2 + p_1 p_3 - p_2 p_3} \quad (\text{B.24})$$

which completes the proof of the theorem. (Note that the fill property of the geometric has been used in order to be able to conclude that any value $E(k') = 1/p$ is possible.) ■

Curriculum Vitae

Name: Colleen Cutler

Post-Secondary Education and Degrees University of Guelph
2013 BA (Hons) (Psychology)

Carleton University
1985 PhD (Mathematics)

Honours and Awards: OGS Sep 2013- Aug 2015
SSC-CRM Prize in Mathematics 2001

Related Work Experience: Teaching Assistant
Psych 2810 (Statistics) and Psych 2800 (Research Methods)
2014-2015 and 2013-2014 respectively

Professor, University of Waterloo
Department of Statistics and Actuarial Science
July 1999 - Aug 2009

Publications Relevant to Psychology:

Cutler, C.D. (2009). Repeat sampling of extreme observations: regression to the mean revisited. *Extremes*. **12**, 297–325. (also available online)