8-2021

Empirical Bayes Estimates for the Reproduction Number of Epidemics

Elijah Lee Hight
The University of Texas Rio Grande Valley

EMPIRICAL BAYES ESTIMATES FOR THE REPRODUCTION NUMBER OF EPIDEMICS

A Thesis

by

ELIJAH LEE HIGHT

Submitted to the Graduate College of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2021

Major Subject: Mathematics

EMPIRICAL BAYES ESTIMATES FOR THE REPRODUCTION NUMBER OF EPIDEMICS

A Thesis
by
ELIJAH LEE HIGHT

COMMITTEE MEMBERS

Dr. George P. Yanev
Chair of Committee

Dr. Hansapani Rodrigo
Committee Member

Dr. Tamer F. Oraby
Committee Member

Dr. Santanu Chakraborty
Committee Member

August 2021

ABSTRACT

Hight, Elijah L., Empirical Bayes Estimates for the Reproduction Number of Epidemics. Master of Science (MS), August, 2021, 45 pp., 5 tables, 4 figures, 49 references, 22 titles.

Epidemic outbreaks can be modelled as a branching process in which the total progeny or outbreak size, follows a Borel-Tanner (BT) distribution. Following a procedure described by Liang [13], we construct empirical Bayes estimates for when the initial number of infected is a specified value r. Following the construction, we then simulate data and perform a numerical study, assuming BT distribution for the parameter θ, the reproduction number, with an initial outbreak size of three. Simulation results indicate that the empirical estimator suffers from "jumpiness." We then proceed to monotonize the empirical estimate via a method outlined by Houwelingen [8]. We then compared the regret risks of each of the estimators and found that the monotonized estimate is the superior estimator for θ. Testing for different values of b under the linex loss function seems to indicate negative b values produce better monotonized estimates. Lastly, we constructed an empirical Bayes estimate for the case when the initial number of infected is a Poisson random variable.

DEDICATION

To my husband, best friend, and partner in crime, Adrian. Thank you for your advice, patience and love, all of which has made me the man I am today.

To all of my family and friends for their affection and love that helped me through the toughest of times.

And to all my teachers, mentors and professors for their dedication and partnership to my academic and professional success.

ACKNOWLEDGMENTS

I am Presidential Graduate Reasearch Assistant (PGRA) award recipient; throughout my graduate studies I was also a Graduate Teaching Assistant (GTA) for the Department of Mathematics and Statistical Sciences (SMSS) at UTRGV. I want to thank the corresponding foundations and the math department for funding my education and providing me with invaluable experience.

Additionally, I would like to extend my sincere gratitude to my thesis advisor Dr. George Yanev, for making my research experience enriching and meaningful, for suggesting the area of interest, and for his unconditional assistance and patience throughout our collaboration. I want to acknowledge Dr. Tamer Oraby and Dr. Erwin Suazo who guided me during my undergraduate studies in two REUs which paved the foundation for my research success. I am also grateful to Dr. Hansapani Rodrigo, Dr. Santanu Chakraborty, and Dr. Tamer Oraby for forming my thesis committee. I want to acknowledge Dr. Timothy Huber, whom I have known since I was an undergraduate, for encouraging me to further my education and providing opportunities to further my professional and academic experience.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

1.1 Borel–Tanner Distribution

In 1942, French mathematician Emile Borel derived the first distribution that described the number of customers served in a single queue with constant service and an initial customer of one. Approximately eleven years later in 1953, Tanner generalized this to any positive integer r.

Definition 1. *For $r \in \mathbb{N}$ and $0 < \theta < 1$, the Borel–Tanner distribution is defined as*

$$p_r(x \mid \theta) = c_r(x)\theta^{x-r}e^{-\theta x}; \quad x = r, r+1, r+2, \dots , \tag{1.1}$$

where is given by $c_r(x) = \frac{rx^{x-r-1}}{(x-r)!}$.

Notably, the BT distribution has mean $\frac{r}{1-\theta}$ and variance $\frac{r\theta}{(1-\theta)^3}$.

The BT distribution has applications in various areas of science and real-world phenomena. In the original application of queuing theory, (1.1) is the probability that exactly x customers in a queue will be served before the first queue vanishes, assuming we start with r initial customers, a traffic rate θ, Poisson arrivals and constant service time [7]. This model has also arisen in coalescence models [3], self propagating internet viruses called worms, traffic flow [16], etc (for other applications, see [14],[9],[12],[6]). The motivation for this paper stems from the role of BT distribution in modeling epidemics.

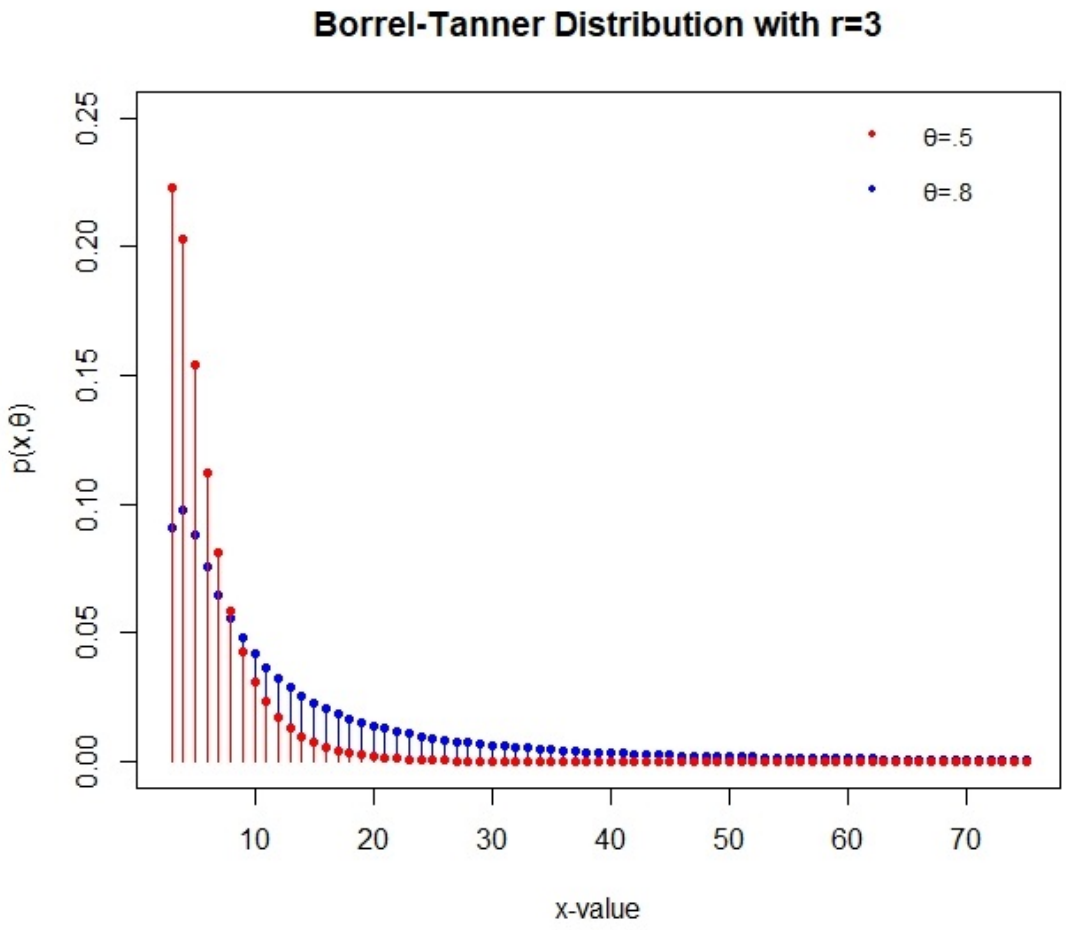

Figure 1.1: Borel–Tanner pmf with $r = 3$.

1.2 Total Progeny of Branching Processes

Sometime during the 19th century, the aristocratic families of Victorian England posed a question to mathematician Sir Francis Galton:

> *How many male children (on average) must each generation of a family have in order for the family name to continue in perpetuity?* [2]

The answer to this question became the oldest, and simplest branching process know as the Galton–Watson (GW) process. Additionally it is called the Bienayme–Galton–Watson process dating as far back as 1845 to the the work of statistician Bienayme. By definition, a system in which particles live for a random time and produce a random number of progenies is called a branching process.

Branching processes are useful in many applications, such as gene propagation, neutron chain reactions in nuclear fusion, biological cells, and epidemiology ([11], [21], [2], [10]). In this paper we consider its application to epidemiology by highlighting how the progeny, or total number of infected individuals of a communicable disease, is a BT random variable.

Definition 2. *For $i = 1,2,\ldots$ and $n = 0,1,2,\ldots$, the Galton-Watson process is defined by the recurrence formula*

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{i,n}, \quad Z_0 = 1,$$

where $X_{i,n}$ are independent and identically distributed (iid) random variables (rv) such that $X_{i,n} \in \mathbb{Z}^+$.

When looking at a branching process, there are two basic assumptions: (e.g. [21])

(i) The number of offspring $X_{i,n}$ produced by a single parent particle is independent of the history of the process, and of other particles existing at the present.

(ii) The offspring distribution is the same for all particles in all generations of the process.

Denote by $Z_0, Z_1, \ldots, Z_n$ the sizes of the first n generations in a simple branching process and let $Y_n := Z_0 + Z_1 + \ldots + Z_n$.

Proposition 1. *If the offspring variable X follows $Poi(\theta)$ and assuming $Y = Y_{n-1} = Y_n$ is the total progeny of a GW process, then the total progeny is a BT r.v. with pmf*

$$P(Y = y) = \frac{y_0 y^{y-y_0-1}}{(y-y_0)!}\theta^{y-y_0}e^{-\theta y}, \quad y = y_0, y_0+1, \ldots$$

Proof.

$$P(Z_1 = z|Z_0 = y) = \sum_{x_1+\ldots+x_y=z} \frac{\theta^{x_1}e^{-\theta}}{x_1}\frac{\theta^{x_2}e^{-\theta}}{x_2}\cdots\frac{\theta^{x_y}e^{-\theta}}{x_y}$$

$$= \theta^z e^{-\theta y} \sum_{x_1+\ldots+x_y=z} \frac{1}{x_1 x_2 \cdots x_y}.$$

Therefore,

$$P(Z_0 = z_0, Z_1 = z_1, \ldots, Z_n = z_n \mid \theta) = \prod_{k=1}^{n} P(Z_k = z_k | Z_{k-1} = z_{k-1}) = \prod_{k=1}^{n} P(Z_1 = z_k | Z_0 = z_{k-1})$$

$$= C(z_0, \ldots, z_n)\theta^{z_1+\ldots+z_n} e^{-\theta(z_0+\ldots+z_{n-1})}$$

$$= C(z_0, \ldots, z_n)\theta^{y_n - y_0} e^{-\theta y_{n-1}}.$$

By assuming that $Y = Y_{n-1} = Y_n$ is the total progeny of the process, we recognize the kernel of Borel-Tanner distribution. The coefficient C does not depend on θ and so must be equivalent to the normalization constant of the Borel-Tanner pmf. Hence,

$$P(Y = y) = \frac{y_0 y^{y-y_0-1}}{(y-y_0)!}\theta^{y-y_0} e^{-\theta y}, \quad y = y_0, y_0+1, \ldots$$

□

More importantly, we are concerned with the variable θ which characterizes the offspring distribution. We call this θ the reproduction number, or number of secondary infections caused by a parent (infected individual). In the sections that follow, we will provide several Bayesian estimators for the value of θ, perform a numerical study, and give a rough assessment of their admissibility.

CHAPTER II

BAYES ESTIMATORS

Bayesian statistics is a branch of statistics for making an inference on a parameter θ. The process starts by first formulating a prior distribution $G(\theta)$ based on a belief or knowledge of an observer. This distribution describes the variation in θ. After conducting an experiment, we observe data x, indexed by θ, taken from the population. The resulting sample distribution $p(x \mid \theta)$ illustrates the observer's belief of x given θ is true. Using the experimental data, we then update the prior and create a posterior distribution $G(\theta \mid x)$. The following result is known as the Bayes Rule:

$$G(\theta \mid x) = \frac{p(x \mid \theta)G(\theta)}{m(x)} \qquad \theta \in \Omega, \tag{2.1}$$

where $m(x)$ is the marginal distribution of X ($m(x) = \int_{\Omega} p(x,\theta)\mathrm{d}\theta$) and $p(x,\theta)$ is the joint probability mass function. The posterior distribution is now used to make inferences about θ.

2.1 Loss Functions–LINEX Loss

In the Bayesian framework, the unknown parameter θ, a r.v. with posterior distribution G, is a value drawn from $G(\theta \mid x)$, the posterior distribution, and is a possible realization of the true parameter. It is thus important to consider how accurate the estimation is by computing the expected loss of the given estimate. To do this, we use a loss function.

A loss function $L(\theta, \hat{\theta})$, is defined as the difference between the estimated and true value of a parameter. This function represents the "cost" associated with some random event. In most cases, errors are minimized and do not consider the loss associated with the error. And so there is a level of ignorance in one's sureness of the parameter. Bayesian estimation however, minimizes posterior

loss and so if one is to be unsure or wrong in their estimation, then it is best to ere on the side of *least* wrong. In this paper, we introduce the following Linear Exponential (LINEX) loss function defined as follows:

Definition 3. *Let θ be an unknown r.v. with posterior distribution $G(\theta)$. Then for any estimate $\hat{\theta}$ for θ we have*

$$L(\theta,\hat{\theta}) = e^{b(\hat{\theta}-\theta)} - b(\hat{\theta}-\theta) - 1, \quad b \neq 0 \tag{2.2}$$

(2.2) was first defined by financial economist Varian in 1975 and is an asymmetric loss function originally derived to determine the cost of a house depending on if you were either a buyer or seller [20]. The asymmetry is determined strictly by the value of b. When $b < 0$, underestimation is penalized more severely than overestimation. Conversely, when $b > 0$ overestimation is penalized more severely than underestimation. A visual depiction of this asymmetry is given in Figure 2.1b.

(a) $b < 0$ (b) $b > 0$

Figure 2.1: Note how in (a), when $\theta - a < 0$, the function is approximately exponential while when $\theta - a > 0$ the function is approximately linear. The converse is true for (b)

We chose the linex loss function particularly for its asymmetry. In an epidemiological model, estimates of the reproduction number are often used to advise public health officials on the possible severity of an outbreak. For example, if we underestimate the true parameter, we are underestimating the spread of the disease and the public may not be adequately prepared if public health officials use our estimate to make policy. This particular situation can be disastrous. Thus we want a loss function that allows us how to determine how severely we want to penalize our given estimate.

2.2 Classical Bayes Estimation

A more detailed Bayes mathematical framework consists of the following elements (e.g. Stijnen [19]). We shall instead summarize some important terms pertaining to Bayesian estimation.

(i) A sample space S of observations,with a $\sigma-$algebra on $\mathcal{S}$.

(ii) A collection $\mathcal{P}$ of probability measures on the space $(S,\mathcal{S})$. Usually, $\mathcal{P}$ is parametrized by some set suitable parameters $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$.

(iii) A set A of possible actions which can be taken by the statistician upon observing some $x \in S$. The set A, called the action space, is equipped with a $\sigma-$algebra on $\mathcal{A}$.

(iv) A collection D of decision rules. A decision rule is defined to be a $\mathcal{S}-\mathcal{A}$ measurable map from S into A. A decision rule is defined to be a $\mathcal{S}-\mathcal{A}$ measurable map from S into A When using the decision rule $d \in D$, the statistician will take action $d(x) \in A$ upon observing $x \in S$.

(v) A loss function $L : \Omega \times A \longrightarrow \mathbb{R}$. For each $\theta \in \Omega$, the function $L(\theta,\cdot)$ must be $\mathcal{A}$ measurable and bounded from below on A. When taking $d(x) \in A$, if θ is the true parameter value, the statistician will incur a loss function $L(\theta, d(x))$.

(vi) A probability measure G (called the prior distribution) on Ω, which is equipped with the $\sigma-$ algebra $\mathcal{W}$.

Consider a random variable or vector X, with distribution θ which is unknown. What is the correct decision to take concerning the true value of θ? To answer this question, we will adopt the Bayesian model; we define the following Bayes estimator θ_G. Suppose $\theta \in \Omega$ is a realization of a rv Θ. Under the linex loss function, with prior distribution G and Borel-Tanner pmf (1.1), it is well known [13] that the Bayes estimator θ_G for θ is

$$\theta_G(x) = -\frac{1}{b}\ln E\left[e^{-b\Theta}|X = x\right].$$

$E\left[e^{-b\Theta}|X=x\right]$ is the posterior expectation of $e^{-b\Theta}$ given $X=x$. Further it can be shown

$$E\left[e^{-b\Theta}|X=x\right]=\frac{\int_0^1 e^{-b\theta}\theta^{x-r}e^{-x\theta}\,dG(\theta)}{\int_0^1 \theta^{x-r}e^{-x\theta}\,dG(\theta)}=\frac{\psi_G(x,b)}{q_G(x)}.$$

Example 2.2.1. Let G be $Uni(c_1,c_2)$ for θ; we can then reduce the above equation to the following

$$\theta_{Uni}(x)=-\frac{1}{b}\ln\frac{\psi_G(x,b)}{q_G(x)}=-\frac{1}{b}\ln\frac{\int_m^n e^{-b\theta}\theta^{x-r}e^{-x\theta}\,d\theta}{\int_m^n \theta^{x-r}e^{-x\theta}\,d\theta}. \tag{2.3}$$

In the context of a possible pandemic, the parameter Θ represents the reproduction number of a current outbreak of a disease modeled by a Galton-Watson process. It is our goal to construct a quality estimator $\hat{\theta}$ to inform health officials to take appropriate measures in controlling the spread of the disease. If $\hat{\theta}$ is close to 0, public intervention may not be necessary as the outbreak will eventually go to extinction. If $\hat{\theta}$ is close to 1, then health measures must be taken to curb the spread of a disease. For this reason, we shall construct a series of quality estimates to make reliable inferences about the population parameter.

2.3 Empirical Bayes Estimation

In order to properly make an inference on the population, we need prior knowledge of the distribution. However in most cases, it is reasonable to assume that this prior distribution exists but is unknown. The empirical Bayes approach offers a solution when the experiment under the investigation has been preceded by a series of comparable experiments. Using these past observations, we can formulate information about the prior distribution. Consider a series of n independent copies of the random pair (X,Θ), where Θ has a (prior) distribution G, i.e.

$$(X_1,\Theta_1),(X_2,\Theta_2),\cdot,(X_n,\Theta_n).$$

Let's assume

(i) X_i, $i=1,2,\ldots,n+1$ are observable and parametrized by θ_i, $i=1,2,\ldots,n+1$.

(ii) Each θ_i is not observable and has unknown prior G.

(iii) Let X_{n+1} stand for the present observation and $\underline{X}(n) := (X_1, \ldots, X_n)$ denote the n past observations. And $X_1, X_2, \ldots, X_n, X_{n+1} \equiv X$

The question then becomes whether it is possible to infer from the set of values $X_1, X_2, \ldots, X_n$ the approximate form of the unknown G, or directly of the Bayes estimator $\theta_G(x)$ [15]. And the answer is yes.

In what follows, we will adopt the empirical Bayes method of estimation [4], which relies on the assumption for the existence of a prior that is unspecified except that it is also i.i.d. from an unknown distribution, with cumulative distribution function G. Since the prior is not known, the Bayes estimator (2.3) cannot be calculated directly. Suppose then that our estimation problem is one in a sequence of similar problems with the same prior distribution. Our goal is to construct a point estimate for θ given the sequence of past data. Such an estimator is called empirical Bayes (EB) estimator. We will seek a direct (independent of G) estimate of the Bayes estimator θ_G, i.e., the posterior mean. In case of BT distribution, Liang [13] proposed such an EB estimator, which we will present and use below.

Following Robbins, we consider the case where $X_1, X_2, \ldots, X_n$ is a sequence of independent random variables, independent from (X, Θ) and with the same BT marginal distribution as X. Consider past observed data $\underline{X}(n) := \{X_1, X_2, \ldots, X_n\}$ generated by an unobserved set of parameter values $\{\Theta_1, \Theta_2, \ldots, \Theta_n\}$ according to the BT p.m.f. $p_r(x;\theta)$ given in (1.1). Let x be the present observation and θ be the present parameter value of Θ. An EB estimator $\theta_n(\underline{X}(n);x) =: \theta_n(x)$ for the parameter θ is a function of the currently observed x and the past data $\underline{X}(n)$. Define for $x = r, r+1, \ldots$, $k = 1, 2, \ldots$, and each $j = 1, \ldots, n$ [13]

$$\psi_{nj0}(x) = q_{nj}(x) = \frac{I\{X_j = x\}}{c_r(X_j)},$$

$$\psi_{njk}(x) = c_k(X_j - x)\frac{I\{X_j \geqslant x + k\}}{c_r(X_j)}, \textit{and}$$

$$\psi_{nj}(x,b) = \sum_{k=0}^{\infty} \frac{(-b)^k}{k!} \psi_{njk}(x).$$

$I\{\cdot\}$ denotes an indicator function that returns a Boolean value, 1 if true and 0 if false. Thus for $\psi_G(x,b)$ and $q_G(x)$ we can define their empirical equivalents as

$$\psi_n(x,b) := \frac{1}{n}\sum_{j=1}^{n} \psi_{nj}(x,b) \quad \text{and} \quad q_n(x) := \frac{1}{n}\sum_{j=1}^{n} q_{nj}(x).$$

Definition 4. *Following Liang (2009) [13], consider the EB estimator θ_n given by*

$$\theta_n(x) := -\frac{1}{b}\ln\left[\left(a_1 \vee \frac{\psi_n(x,b)}{q_n(x)}\right) \wedge a_2\right], \qquad q_n(x) \neq 0, \quad x = r, r+1, \ldots \tag{2.4}$$

where $a_1 = \min(e^{-b}, 1)$ and $a_2 = \max(e^{-b}, 1)$.

For BT distribution there is an additional estimator. When r is fixed the maximum likelihood estimator (MLE) for the parameter θ is given by the method of moments estimator, i.e for $x > 0$

$$\theta_{mle}(x) = \frac{x-r}{x}.$$

2.4 Bayes Risk and Regret

The Bayes risk of an estimator $\hat{\theta}$ under linex loss $L(\theta, \hat{\theta}(x)) = e^{b(\hat{\theta}(x)-\theta)} - b(\hat{\theta}(x)-\theta) - 1$ is defined as

$$R(\theta, \hat{\theta}) = E\big[L(\theta, \hat{\theta})\big], \tag{2.5}$$

where the expectation is taken with respect to both x and θ. Then it follows that the minimum Bayes Risk is given by:

$$R(\theta, \theta_G) = E\big[L(\theta, \theta_G)\big] = E\big[e^{b(\theta_G(x)-\theta)} - b(\theta_G(x)-\theta) - 1\big].$$

By equation (2.3) and as shown in [13], we see $E\left[e^{b(\theta_G(x)-\theta)}-1\right]=E_X\left[E\left[e^{-b(\theta_G(X)-\theta)}-1|x\right]\right]=0$ so the minimum Bayes Risk becomes

$$R(\theta,\theta_G)=E\left[b(\theta-\theta_G(x)\right].$$

By definition the Bayes estimator θ_G for θ minimizes the risk (2.5) over the set of all estimators $\hat{\theta}$. Consider the EB estimator θ_n in (2.4). For fixed $X_1,X_2,\ldots,X_n$ the risk of $\theta_n(X_1,\ldots,X_n;X)$, denoted by $\tilde{R}(G,\theta_n)$, is given by

$$\tilde{R}(\theta,\theta_n)=E_nE_{(X_{n+1},\Theta_{n+1})}\left[e^{b(\theta_n(X_1,\ldots,X_n;X)-\theta)}-b(\theta_n(X_1,\ldots,X_n;X)-\theta)-1\mid X_1,\ldots,X_n\right].$$

Note that $\tilde{R}(\theta,\theta_n)$, which is called the conditional Bayes risk of θ_n, is a random variable. Then the overall Bayes risk of the EB estimator θ_n is defined by

$$R(\theta,\theta_n):=E\tilde{R}(\theta,\theta_n)=E_n\left[e^{b(\theta_n(x_1,\ldots,x_n;x)-\theta)}-b(\theta_n(x_1,\ldots,x_n;x)-\theta)-1\right],$$

where the expectation $E_n[\cdot]$ is taken with respect to $(x_1,\ldots,x_n)$.

In practice, various criteria are used to select estimators, which are optimal in some sense. For our study, we apply the concept of *regret risk*.

Definition 5. *The non-negative difference*

$$S(\theta_n):=R(\theta,\theta_n)-R(\theta,\theta_G)\geqslant 0,$$

called regret risk of θ_n, is a standard measure of the quality (optimality) of an EB estimator.

In particular, a sequence $\{\theta_n\}_{n=1}^{\infty}$ of EB estimators is called asymptotically optimal for G if $\lim_{n\to\infty}S(\theta_n)=0$. It is proved in [13] that, under certain conditions, θ_n given by (2.4) is asymptotically optimal with rate of convergence $O\left(n^{-\lambda/2}\right)$ for some $\lambda\in(0,2)$. We will use this measure of estimator quality to determine the best estimator for θ

CHAPTER III

MONOTONIZING THE EMPIRICAL BAYES ESTIMATOR

3.1 Randomized Testing

In statistics, randomization is essential to the experimental process. Arbitrarily choosing samples from a population may seem "random" but this method is prone to hidden biases. For example, if I were to select a sample of student names from a class roster, I might be more inclined to choose names that are familiar to me or perhaps my friends or colleagues. Because these biases are often overlooked, it is important to incorporate randomization into the experimental design which serve to "control" or reduce bias by all means. Such considerations provide legitimacy to both the researchers and the study. For more of a basic introduction to randomization, see ∼[1].

Example 3.1.1 (Randomized Test)**.** Let X_1, X_2, X_3 be a sample from $Bernoulli(1, p)$ where $0 \leqslant p \leqslant 1$ and p is unknown. Let x be the number of successes in 3 independent trials. Consider the null hypothesis that $H_0 : p = \frac{1}{3}$ and the alternative that $H_1 : p = \frac{2}{3}$ set at a significance level of $\alpha = 0.05$. The probabilities are shown in the table below: Clearly $P(X = 3)$ fully falls in the rejection region

Table 3.1: Probability values for $\{X_1, X_2, X_3\} \sim Bernoulli(1, p)$

x	$P(X = x)$
0	$\frac{3!}{3!0!}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^3 = \frac{8}{27} \approx 0.296$
1	$\frac{3!}{2!1!}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2 = \frac{4}{9} \approx 0.444$
2	$\frac{3!}{1!2!}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1 = \frac{2}{9} \approx 0.222$
3	$\frac{3!}{0!3!}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^0 = \frac{1}{27} \approx 0.037$

and $P(X=2), P(X=1), P(X=0)$ do not. The issue is that we are not using $\alpha = 0.05$ as our exact critical value; this leads us to establishing a bias for certiain vlaues of x in the decision process. A way to fix this is to randomize the test. In this case we will add a weight c to $X=2$, that is, we partially include the point $X=2$ so that we obtain the exact critical value $\alpha = 0.05$,

$$\begin{aligned} P(X=3) + cP(X=2) &= 0.05 \\ \frac{1}{27} + c\frac{2}{9} &= 0.05 \\ c &= \frac{7}{120} \end{aligned}$$

Thus, the optimal test of size $\alpha = 0.05$ is given by

$$\Phi_\theta(x) := \begin{cases} 0 & \text{if } x < 2 \\ \dfrac{7}{120} & \text{if } x = 2 \\ 1 & \text{if } x = 3. \end{cases}$$

As you can see, this type of randomization procedure assigns values by chance not by choice. In the above example the weight c is used to obtain the exact α value and eliminate bias. There are other ways to randomize an experiment; in the section that follows we use a function, namely $D(a \mid x)$, to randomize the EB estimator θ_n as seen in [20].

3.2 Van-Houwilengen's Monotonization Procedure

The EB estimator is not monotone with respect to x. We provide an illustration of θ_n in the Chapter IV numerical study. This is unwanted behavior for an estimator following BT distribution.

Proposition 2. *The BT distribution, (1.1) has monotone likelihood ratio (MLR), i.e.,*

$$q(x) = \frac{p_r(x \mid \theta_2)}{p_r(x \mid \theta_1)} \tag{3.1}$$

is increasing with respect to x whenever $0 < \theta_1 < \theta_2 < 1$.

The above proposition is proved in Soltero [18] but we will prove it here for easy reference.

Proof. Let g be the natural logarithm of the likelihood ratio q. $0 < \theta_1 < \theta_2 < 1$ and $r \in \mathbb{N}$. Then,

$$q(x) = \frac{p_r(x;\theta_2)}{p_r(x;\theta_1)} = \left(\frac{\theta_2}{\theta_1}\right)^{x-r} e^{-(\theta_2-\theta_1)}$$

is an increasing function of x for $0 < \theta_1 < \theta_2 < 1$. Indeed,

$$g(x) = \log \frac{p_r(x;\theta_2)}{p_r(x;\theta_1)} = \log\left[\left(\frac{\theta_2}{\theta_1}\right)^{x-r} e^{-x(\theta_2-\theta_1)}\right] = (x-r)(\log\theta_2 - \log\theta_1) - x(\theta_2-\theta_1).$$

Thus, for $0 < \theta_1 < \theta_2 < 1$

$$g'(x) = \frac{d}{dx}\log\frac{p_r(x;\theta_2)}{p_r(x;\theta_1)} = \log\left(\theta_2 e^{-\theta_2}\right) - \log\left(\theta_1 e^{-\theta_1}\right) > 0,$$

because $\log\left(\theta e^{-\theta}\right)$ is an increasing function for $0 < \theta < 1$. This completes the proof. □

Seeing as how the monotonicity property of the BT distribution is desirable, it is also desirable in the estimates we compute for θ; however as Houwalingen [20] points out, this is not the case for the EB estimator; for this reason, he outlined a classical approach for monotonizing the EB estimator. In addition to monotonizing the θ_n, Houwalingen also shows that the monotonized EB estimator, θ_n^* has a smaller Regret risk than the EB estimator θ_n, i.e., θ_n^* is a "better" estimator than θ_n. A procedure for constructing a monotone estimator that dominates an EB estimator for distributions with MLR is given. In his paper, Houwalingen also provides examples of this estimator for the Geometric and Poisson distributions. In Chapter IV, we discuss yet another example to this classical construction by monotonizing the EB estimator for BT distribution.

Estimators for discrete distributions with MLR can be made monotone applying a procedure developed in [20] (see also [22]). Consider a simple randomized version of the estimator $\theta_n(x)$

represented by the following function $D(a \mid x)$ for $a \in (0,1)$:

$$D(a \mid x) := \begin{cases} 0 & \text{if } \theta_n(x) > a, \\ 1 & \text{if } \theta_n(x) \leqslant a. \end{cases}$$

The number $D(a \mid x)$ is the probability that an estimate $\theta_n(x)$ less than or equal to a is selected given $X = x$. In other words, $D(a \mid x)$ is a cdf on the action space $(0,1)$ for every $X = x$. Then define for $a \in (0,1)$

$$\alpha(a) := E(D(a \mid X)) = \sum_{\{x:\ \theta_n(x) \leqslant a\}} p_r(x \mid a). \tag{3.2}$$

Denote $F(x \mid \theta) := \sum_{k=r}^{x} p_r(k \mid \theta)$ for $x \geqslant r$ and assume $F(r-1 \mid \theta) = 0$. Now, we can construct a randomized estimator with $D^*(a \mid x)$ as follows

$$D^*(a \mid x) := \begin{cases} 0 & \text{if } \alpha(a) < F(x-1 \mid a) \\ \dfrac{\alpha(a) - F(x-1 \mid a)}{F(x \mid a) - F(x-1 \mid a)} & \text{if } F(x-1 \mid a) \leqslant \alpha(a) \leqslant F(x \mid a) \\ 1 & \text{if } F(x \mid a) < \alpha(a), \end{cases} \tag{3.3}$$

with $D^*(1 \mid x) = 1$, and $D^*(0 \mid x) = \lim_{a \downarrow 0} D^*(a \mid x)$. Let $a \in (\theta_0, \theta_1)$ be fixed. It follows from the construction of D^*, that $E_a D^*(a \mid X) = E_a D(a \mid X)$.

The next proposition shows that, using the monotone estimator D^*, one can construct another (non-random) monotone estimator θ_n^*, say, with risk less than or equal to the risk of the θ_n.

Proposition 3. *Let $D^*(a \mid x)$ be the monotone estimator constructed in (3.3).Define*

$$\theta_n^*(x) := \int_0^1 a \, \mathrm{d}D^*(a \mid x). \tag{3.4}$$

Then the monotone non-random estimator $\theta_n^*(x)$ dominates $D^*(a \mid x)$, which itself dominates the

initial estimator $D(a \mid x)$, i.e.,

$$R(\theta, \theta_n^*) \leqslant R(\theta, D^*) \leqslant R(\theta, D). \tag{3.5}$$

Proof. The proposition follows from the theorem in [20]. It suffices to verify that BT distribution satisfies all assumptions of the theorem. In particular, it has a monotone likelihood ratio as it was shown in (3.1). Therefore, the second inequality in (3.5) follows as in [20]. That is D^* represents a monotone estimator, which dominates the initial estimator represented by D for all $\theta \in (0,1)$. It is not difficult to see that, under the linex loss function, D^* itself is dominated by the non-random monotone estimator θ_n^*. Indeed, using Jensen's inequality, we have

$$\begin{aligned}
R(\theta, \theta_n^*(X)) &= E[L(\theta - \theta_n^*(X))] \\
&= E\left[e^{b(\theta_n^*(X)-\theta)} - b(\theta_n^* - \theta) - 1\right] \\
&= E\left[-b\left(\int_0^1 a\, dD^*(a,X) - \theta\right)\right] \\
&\leqslant E\left(\int_0^1 -b(a-\theta)\, dD^*(a,X)\right) \\
&= R(\theta, D^*(a,X)).
\end{aligned}$$

□

CHAPTER IV

NUMERICAL STUDY

In this section, using simulations, we compare the performance of the monotone EB estimator θ_n^* with the initial EB estimator θ_n and the maximum likelihood estimator θ_{mle}. We also consider the effect that the value of b has on the performance of each of the estimates. The algorithm for the simulations is provided in Appendix A. In applications, there is a compelling argument [13] for θ to take on values in a sub-interval of $(0,1)$. Additionally, we want to consider the values of r which represents some real quantity. Let's consider r then to be the initial number of infected individuals coming into a country with a communicable disease. Lastly, we consider values of $b < 0$ which penalizes underestimation more severely, because in an epidemiological framework, estimates that are smaller than the parameter can lead to health officials being under prepared for a potential outbreak. To carry out the simulation, we assume $Uni(0.5, 0.8)$ prior. We choose this range of values because at the low end, a reproductive number of $\theta = 0.5$ is indicative of a healthy epidemic that should die out on its own. On the high end, a reproductive number of $\theta = 0.8$ is indicative of a viral outbreak that could become epidemic. Setting $r = 3$ and $b = -1$, we evaluate $\theta_G(x)$ and calculate the minimum Bayes risk $R(\theta, \theta_G) = E[L(\theta_G(x) - \theta)] = 0.003299071$. The maximum likelihood estimator $\theta_{mle}(x) = (x-3)/x$ has regret risk $R(\theta_{mle}) = R(\theta, \theta_{MLE}) - R(\theta, \theta_G) = 0.05529896$.

4.1 The Effect of n in the Empirical Bayes Scheme

Within the empirical Bayes framework, consider $n = 80$ independent copies

$$(X_1, \Theta_1), (X_2, \Theta_2), \ldots, (X_{80}, \Theta_{80}) \tag{4.1}$$

of the random pair (X, Θ), where Θ is $Uni(0.5, 0, 8)$ variable and, given Θ, X follows the BT distribution (2.3). Assume that X_i are observable, but Θ_i are not. For our simulation study, we simulate 80 pairs like in (4.1). For the j^{th} pair $1 \leqslant j \leqslant 80$, we calculate the EB estimate $\theta_{80}^{(j)}(x)$. We estimate $S(\theta_{80})$ by the average (for the 80 pairs) $\hat{S}(\theta_{80}) := \dfrac{1}{80}\displaystyle\sum_{j=1}^{80} S(\theta_{80}^{(j)})$ and calculate the standard error. Next, we monotonise the EB estimator and compute the estimate $\theta_{80}^{*(k)}(x)$. Similarly to $S(\theta_{80})$, we estimate $S(\theta_{80}^*)$ by the average $\hat{S}(\theta_{80}^*)$. We repeat the entire procedure for $n = 20$, 40, and 60 as well. Once complete, we report the numerical results for the regret risks ratios w.r.t θ_{mle} in Table 4.1 below. The improvement of θ_n^* over θ_n and θ_{mle} is quite substantial.

Table 4.1: Estimates for the Regret Risk Ratios of θ_n, and θ_n^* w.r.t θ_{mle}

n	$\hat{S}(\theta_{mle})$	$\hat{S}(\theta_n)$	$\hat{S}(\theta_n^*)$
20	0.05529896	▲ 31.6%	▼ -77.3%
40	0.05529896	▼ -12.4%	▼ -82.0%
60	0.05529896	▼ -39.0%	▼ -85.6%
80	0.05529896	▼ -52.4%	▼ -86.5%

All standard errors are less than 10^{-4}; $b = -1$ and $r = 3$.

Notice that the monotonized estimator outperforms the other estimators, especially as sample size increases, indicated by the larger decrease in the regret risk ratios in Table 4.1. In other words, the monotone empirical estimate is closest to the minimum Bayes Risk and is better estimate in comparison to the maximum likelihood and Empirical Bayes estimates. Additionally we present the complete results – we average the monotone estimate and empirical estimate across all 10 realizations of size $n = 80$ – and compare them to the maximum likelihood and Bayes estimate in Figure 4.1 to illustrate the estimators' behaviour.

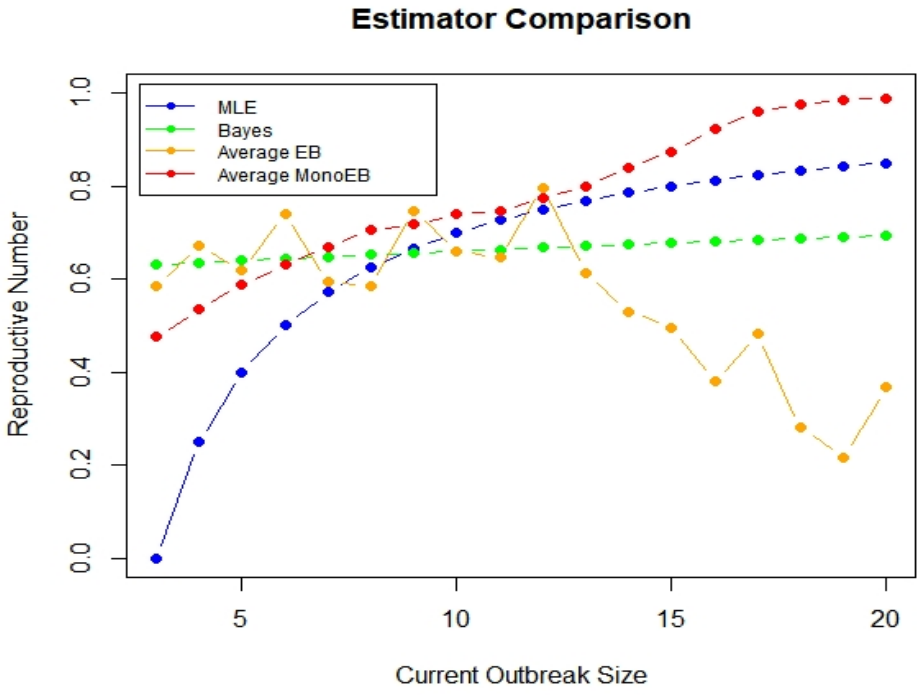

Figure 4.1: Average estimations given $n = 80$, $r = 3$, $b = -1$, and prior $U(0.5, 0.8)$.

4.2 Investigating the Effect of the Penalization Weight b

Next we investigated the effects of changing the penalization weight b. As discussed previously, it is more likely that the consequences of underestimating the reproduction number θ are more severe than overestimating. If public health officials use underestimated reproduction numbers to establish and drive health mandates, then the community may be under preparing for a potentially more infectious disease. Hence we prefer to consider values of $b < 0$. However, we want to verify whether this is supported by our numerical study. Essentially, we are curious to see if the estimates perform better when $b < 0$. To test this, we computed the regret risk ratios for each of the three estimates θ_{MLE}, θ_n and θ_n^*, for a simulation where we consider $n = 20$ independent copies, with $r = 3$ and $b = \{-2, -1.5, -1, -.5, .5, 1, 1.5, 2\}$. The following results were organized in Table 4.2 below.

Table 4.2: Regret Ratios of θ_n, and θ_n^* w.r.t θ_{mle}

b	$\hat{S}(\theta_{mle})$	$\hat{S}(\theta_n)$	$\hat{S}(\theta_n^*)$
-2	0.27568647	▲ 21.7%	▼ -85.6%
-1.5	0.13844631	▲ 25.6%	▼ -82.1%
-1	0.05529896	▲ 31.6%	▼ -77.3%
-.5	0.01250272	▲ 39.6%	▼ -71.7%
.5	0.01040670	▲ 6.00%	▼ -39.2%
1	0.03828749	▲ 32.5%	▼ -25.3%
1.5	0.07963463	▲ 66.7%	▼ -7.69%
2	0.13149226	▲ 107.3%	▲ 15.7%

Note for $b < 0$, the monotonized EB has lower regret than when $b > 0$.

As suspected, the regret risk ratios are lower when b takes on negative values. Indeed, the harsher we penalize underestimation, the lower the regret risk is in comparison to θ_{mle}, indicating that the best estimator from the group is the monotone empirical estimate θ_n^*.

When looking at Figure 4.1, we observed that although the monotone estimate has a lower regret risk ratio than the maximum likelihood estimator, the MLE followed more closely the behavior of the Bayes estimate than the monotone estimate. In an attempt to "correct" the behavior of the empirical estimates to follow more closely that of the Bayes estimate θ_G, we take the average between θ_n and θ_n^* for each simulation, and then average across the 10 simulations. The result is an estimate, denoted θ_n^{ave}. Most notably, this new estimate, much like the empirical Bayes estimate θ_n, is not monotone, however its behavior heavily resembles that of the Bayes estimate. This should indicate a very small regret risk. We compared θ_n^{ave} to the MLE and Bayes estimate graphically in Figure 4.2. Additionally, for $n = 20, 40, 60, 80$ we computed the regret risk ratios w.r.t the MLE, with $r = 3$, $b = -1$, and prior $U(0.5, 0.8)$. The results were compiled in Table 4.3 below.

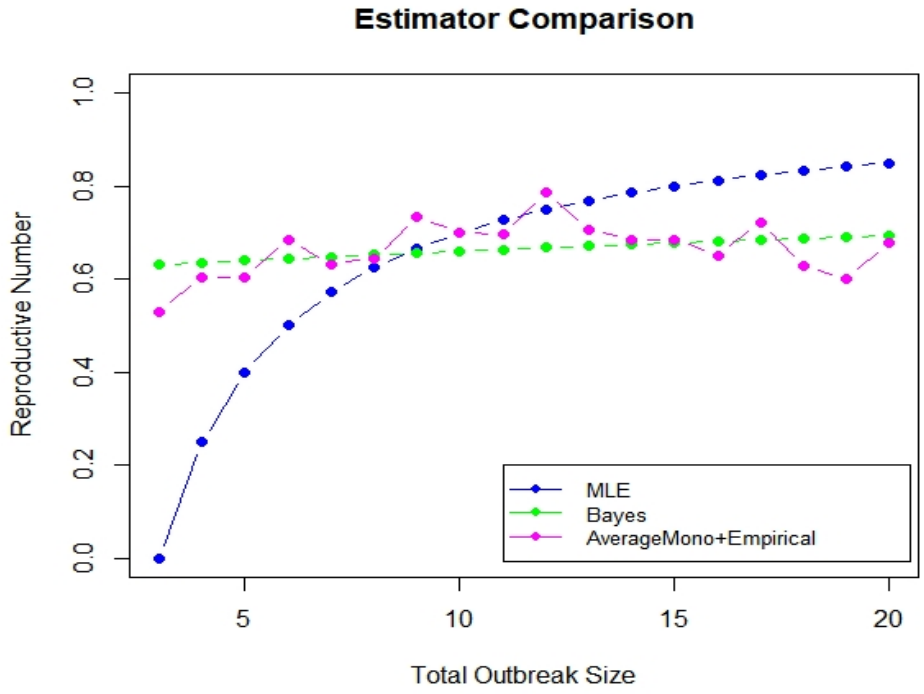

Figure 4.2: Comparison between θ_G, θ_{mle} and θ_n^{ave} given $n = 80$, $r = 3$, $b = -1$, and prior $U(0.5, 0.8)$.

Table 4.3: Regret Risk Ratios of θ_n, θ_n^*, and θ_n^{ave} w.r.t θ_{mle}

n	$\hat{S}(\theta_{mle})$	$\hat{S}(\theta_n)$	$\hat{S}(\theta_n^*)$	$\hat{S}(\theta_n^{ave})$
20	0.05529896	▲ 31.6%	▼ -77.3%	▼ -81.3%
40	0.05529896	▼ -12.4%	▼ -82.0%	▼ -94.7%
60	0.05529896	▼ -39.0%	▼ -85.6%	▼ -97.7%
80	0.05529896	▼ -52.4%	▼ -86.5%	▼ -97.3%

All standard errors are less than 10^{-4}; $b = -1$ and $r = 3$.

CHAPTER V

GENERALIZED POISSON DISTRIBUTION

The Poisson distribution has commonly been used to describe the behavior of living beings and other phenomena over a period of time and is dependent on a constant rate λ; however, the randomness of the Poisson distribution, while ideal, is not always practical even in "natural" events. For example, consider the number of automobile accidents that occur with constant rate λ. Factors such as experience, road condition, weather, etc. "effect the personal liability of an individual" [5] which lead to different rates in accidents among drivers. Moreover, this can result in unequal mean and variance in the numerical data,

Out of all the power series distributions, the Poisson distribution is uniquely characterized as having equal mean and variance [5]; however, in populations that are supposed to be Poissonian, researchers have observed that this is not always the case. In addressing these particular issues, Consul and Jain, in 1970 [5], introduced a generalized Poisson distribution (GPD) which has two parameters, say λ and τ. Let X be a discrete r.v. and let $P_X(\tau,\lambda)$ denote the probability that $X = x$ such that $x \geqslant 0$. The GPD has pmf:

Definition 6.

$$P_X(\tau,\lambda) = \begin{cases} \frac{\tau}{x!}(\tau+\lambda x)^{x-1}e^{-(\tau+\lambda x)}, & x = 0,1,2,\ldots \\ 0, & x > M,\ \lambda < 0 \end{cases}$$

where $\tau > 0$, $max(1,-\tau/M) < \lambda \leqslant 1$ *and* $M(\geqslant 4)$ *is the largest positive integer for which* $\tau + M\lambda > 0$ *when* λ *is negative. [5]*

5.1 One Subfamily of Generalized Poisson Distribution

Let's now consider a particular subfamily of the generalized Poisson distribution (PGPD) as it pertains to Borel-Tanner distribution. Assume that the number of ancestors $Z_0 = r$ follows $Pois(\tau)$ for $\tau > 0$, that is

$$P(Z_0 = r) = \frac{\tau^r e^{-\tau}}{r!}, \quad r = 0, 1, \ldots$$

As it was shown in Section 1.2, if $Z_0 = r$ and the offspring distribution is $Pois(\lambda)$ with $0 < \lambda < 1$, then the total progeny Y follows BT distribution

$$P(Y = y \mid Z_0 = r) = \frac{r y^{y-r-1}}{(y-r)!} \lambda^{y-r} e^{-\lambda y}, \quad y = r, r+1, \ldots$$

Therefore, since $0 \leqslant r \leqslant y$ we can obtain

$$\begin{aligned} P(Y = y) &= \sum_{r=0}^{y} P(Y = y \mid Z_0 = r) P(Z_0 = r) \qquad (5.1) \\ &= \sum_{r=0}^{y} \frac{r y^{y-r-1}}{(y-r)!} \lambda^{y-r} e^{-\lambda y} \frac{\tau^r e^{-\tau}}{r!} \\ &= \frac{e^{-(\tau+\lambda y)}}{y!} \sum_{r=1}^{y} \frac{(y-1)!}{(y-r)!(r-1)!} (\lambda y)^{y-r} \tau^r \quad (set\ k = r-1) \\ &= \frac{\tau e^{-(\tau+\lambda y)}}{y!} \sum_{k=0}^{y-1} \frac{(y-1)!}{(y-1-k)!k!} (\lambda y)^{y-1-k} \tau^k \\ &= \frac{\tau(\tau+\lambda y)^{y-1}}{y!} e^{-(\tau+\lambda y)}. \end{aligned}$$

Next we further look at a subfamily of the PGPD discussed in ([5] pg44). Let the ancestors follow $Poi(\tau)$ with probability generating function (pgf) $f(t) = exp\{\tau(t-1)\}$ for $\tau > 0$ and the offspring distribution is $Poi(\lambda)$ with pgf $g(t) = exp\{\lambda(t-1)\}$ for $0 < \lambda < 1$. Lastly we assume that there exists a φ such that

$$\tau = \varphi\lambda, \quad \varphi \geqslant 1. \qquad (5.2)$$

It follows then from (5.1) and (5.2), changing Y to X for $0 < \lambda < 1$ and $\varphi \geqslant 1$ that

$$P(X = x) = a_\varphi(x)\lambda^x e^{-\lambda(\varphi+x)}, \quad x = 0, 1, \ldots \tag{5.3}$$

where $a_\varphi(x) = \frac{1}{x!}\varphi(\varphi + x)^{x-1}$.

The mean and variance of (5.3) ([5] pg.13) is

$$E[X] = \frac{\varphi\lambda}{(1-\lambda)} \quad \text{and} \quad Var[X] = \frac{\varphi\lambda}{(1-\lambda)^2}.$$

5.2 Bayesian Estimators for λ

In this section we will estimate λ using the the empirical Bayes approach as seen in Chapter II, under squared error loss function. Squared error loss for a parameter λ is defined as $L(\lambda, \hat{\lambda}) = (\lambda - \hat{\lambda})^2$. The Bayes estimator for λ, under squared error loss and assuming that φ is known, equals the posterior mean:

$$\lambda_G(x) = \frac{\int_0^1 a_\varphi(x)\lambda^{x+1}e^{-\lambda(\varphi+x)}dG(\lambda)}{\int_0^1 a_\varphi(x)\lambda^x e^{-\lambda(\varphi+x)}dG(\lambda)} = \frac{\int_0^1 \lambda^{x+1}e^{-\lambda(\varphi+x)}dG(\lambda)}{\int_0^1 \lambda^x e^{-\lambda(\varphi+x)}dG(\lambda)} = \frac{\psi_G(x)}{q_G(x)}. \tag{5.4}$$

Note that since $0 < \lambda < 1$, we have $0 \leqslant \psi_G(x)/q_G(x) \leqslant 1$. Next we construct an empirical Bayes estimate by modifying the procedure in Section (2.3), in Chapter II. Define for $j = 1, 2, \ldots, n$ and $\varphi \geqslant 1$

$$\psi_{nj}(x) = \frac{a_1(X_j - x - 1)}{a_\varphi(X_j)} I(X_j \geqslant x + 1) \quad \text{and} \quad \psi_n(x) = \frac{1}{n}\sum_{j=1}^{n} \psi_{nj}(x).$$

Define also for $j = 1, 2, \ldots, n$ and $\varphi \geqslant 1$,

$$q_{nj}(x) = \frac{1}{a_\varphi(x)} I(X_j = x) \quad \text{and} \quad q_n(x) = \frac{1}{n}\sum_{j=1}^{n} q_{nj}(x).$$

Motivated by [13], we will prove the following proposition. Recall the assumptions in 2.3 and that the expectation E_n is taken with respect to $X_1, X_2, \ldots, X_n$.

Proposition 4.

$$(i) \quad E_n[\psi_n(x)] = \psi_G(x) \quad \textit{and} \quad E_n[q_n(x)] = q_G(x).$$

$$(ii) \quad Var_n[\psi_n(x)] \leqslant \frac{\psi_G(x)}{n} \quad \textit{and} \quad Var_n[q_n(x)] \leqslant \frac{q_G(x)}{a_\varphi(x)} \frac{1}{n}.$$

Proof. We begin by looking at the first moment of $\psi_n(x)$. Making the index change $y = t - x - 1$

$$\begin{aligned} E_n[\psi_{nj}(x)] &= \sum_{t=x+1}^{\infty} \frac{a_1(t-x-1)}{a_\varphi(t)} \int_0^1 a_\varphi(t) \lambda^t e^{-\lambda(t+\varphi)} \, dG(\lambda) \\ &= \sum_{y=0}^{\infty} a_1(y) \int_0^1 \lambda^{y+x+1} e^{-\lambda(y+x+1+\varphi)} \, dG(\lambda) \\ &= \int_0^1 \lambda^{x+1} e^{-\lambda(x+\varphi)} \sum_{y=0}^{\infty} a_1(y) \lambda^y e^{-\lambda(y+1)} \, dG(\lambda) \\ &= \psi_G(x). \end{aligned}$$

Since $\sum_{y \geqslant 0} a_1(y) \lambda^y e^{-\lambda(y+1)} = 1$, therefore

$$E_n[\psi_n(x)] = E_n \left[\frac{1}{n} \sum_{j=1}^{n} \psi_{nj}(x) \right] = \psi_G(x).$$

Next, we look at the first moment of $q_n(x)$. We have

$$\begin{aligned} E_n \left[q_{nj}(x) \right] &= \frac{1}{a_\varphi(x)} \int_0^1 a_\varphi(x) \lambda^x e^{-\lambda(x+\varphi)} \, dG(\lambda) \\ &= \int_0^1 \lambda^x e^{-\lambda(x+\varphi)} \, dG(\lambda) \\ &= q_G(x). \end{aligned}$$

Thus,

$$E_n[q_n(x)] = E_n \left[\frac{1}{n} \sum_{j=1}^{n} q_{nj}(x) \right] = q_G(x).$$

This proves part (i) of Proposition 4. Let us turn to the variances of $q_n(x)$ and $\psi_n(x)$. We have

$$\begin{aligned} Var_n\Big[q_{nj}(x)\Big] &= Var_n\left[\frac{1}{a_\varphi(x)}I(X_j=x)\right] \\ &= \frac{1}{a_\varphi^2(x)}Var_n\big[I(X_j=x)\big] \\ &= \frac{1}{a_\varphi^2(x)}P(X_j=x)(1-P(X_j=x)) \\ &\leqslant \frac{1}{a_\varphi^2(x)}\int_0^1 a_\varphi(x)\lambda^x e^{-\lambda(\varphi+x)} = \frac{q_G(x)}{a_\varphi(x)}. \end{aligned}$$

Therefore,

$$Var_n\Big[q_n(x)\Big] = Var_n\left[n^{-1}\sum_{j=1}^n q_{nj}(x)\right] \leqslant \frac{q_G(x)}{a_\varphi(x)}\frac{1}{n}.$$

Let us prove that for $j=1,2,\ldots,n$, $\varphi\geqslant 1$ and $x\geqslant 0$

$$0\leqslant \psi_{nj}(x) = \frac{a_1(X_j-x-1)}{a_\varphi(X_j)}I(X_j\geqslant x+1)\leqslant 1,$$

where, as before, $a_t(x)=t(t+x)^{x-1}/x!$. Denoting $z:=x_j-x-1$, we obtain

$$\begin{aligned} \frac{a_1(z)}{a_\varphi(z+x+1)} &= \frac{(1+z)^{z-1}}{z!}\frac{(z+x+1)!}{\varphi(\varphi+z+x+1)^{z+x}} \\ &= \frac{(1+z)^{z-1}(z+x+1)(z+x)\ldots(z+2)(z+1)}{\varphi(\varphi+z+x+1)^{z+x}} \\ &= \frac{1}{\varphi}\frac{(1+z)^z}{(1+z+x+\varphi)^z}\frac{(z+x+1)(z+x)\ldots(z+2)}{(z+x+1+\varphi)^x} \\ &\leqslant 1, \end{aligned}$$

since all three factors are less than one. Thus,

$$Var_n\big[\psi_{nj}(x)\big]\leqslant E_n\big[\psi_{nj}^2(x)\big]-\big(E_n\big[\psi_{nj}(x)\big]\big)^2\leqslant E_n\big[\psi_{nj}^2(x)\big]\leqslant E_n\big[\psi_{nj}(x)\big]=\psi_G(x).$$

Thus,

$$Var_n\Big[\psi_n(x)\Big] = Var_n\left[\frac{1}{n}\sum_{j=1}^{n}\psi_{nj}(x)\right] \leqslant \frac{1}{n}\psi_G(x)$$

This proves part (ii). Hence the proof of Proposition 4 is complete. □

Notably Proposition 4 implies $\psi_n(x)$ and $q_n(x)$ are both unbiased and consistent estimators for $\psi_G(x)$ and $q_G(x)$ respectively. Following similar constructions in Singh and Wei [17] and Liang [13], we define the EB estimator $\tilde{\lambda}_n$ for λ as

Definition 7.

$$\tilde{\lambda}_n(x) = \frac{\psi_n(x)}{q_n(x)} \wedge 1$$

where $a \wedge b = min\{a,b\}$.

Note that for fixed $X = x$

$$0 \leqslant \frac{\psi_n(x)}{q_n(x)} \wedge 1, \quad \frac{\psi_G(x)}{q_G(x)} \leqslant 1.$$

Under the squared error loss function, the Bayes risk of the EB estimator $\tilde{\lambda}_n$ is given by

$$R(G,\tilde{\lambda}_n) = E_n E_{(X,\Lambda)}[\tilde{\lambda}_n(x) - \Lambda]^2,$$

where Λ is the r.v. with value λ. Therefore the regret risk of $\tilde{\lambda}_n$ can be written as

$$S(\tilde{\lambda}_n,\lambda_G) = R(G,\tilde{\lambda}_n) - R(G,\lambda_G) = E_n E_{(X,\Lambda)}[\tilde{\lambda}_n(x) - \lambda_G(x)]^2.$$

Indeed,

$$\begin{aligned}
R(G,\tilde{\lambda}_n) &= E_n E_{(X,\Lambda)}[\tilde{\lambda}_n(x) - \Lambda]^2 \\
&= E_n E_{(X,\Lambda)}[\tilde{\lambda}_n(x) - \lambda_G(x) + \lambda_G(x) - \Lambda]^2 \\
&= E_n E_{(X,\Lambda)}[\tilde{\lambda}_n(x) - \lambda_G(x)]^2 + E_n E_{(X,\Lambda)}[\lambda_G(x) - \Lambda]^2 - 2E_n E_{(X,\Lambda)}[(\tilde{\lambda}_n(x) - \lambda_G(x))(\lambda_G(x) - \Lambda)] \\
&= E_n E_{(X,\Lambda)}[\tilde{\lambda}_n(x) - \lambda_G(x)]^2 + R(G,\lambda_G).
\end{aligned}$$

Next we consider if the EB estimator $\tilde{\lambda}_n$ is asymptotically optimal for every prior G and at what rate.

Theorem 1. *For each prior distribution G, the EB estimator $\tilde{\lambda}_n$ is asymptotically optimal.*

Proof. Let $X = x$ be fixed. Then

$$S(\tilde{\lambda}_n(x), \lambda_G(x)) = \sum_{x=0}^{\infty} E_n[\tilde{\lambda}_n(x) - \lambda_G(x)]^2 p_G(x),$$

where $\sum_{x=0}^{\infty} p_G(x) = 1$. Following Robbins [15], it is sufficient to show that:

$$\lim_{n\to\infty} E_n[\tilde{\lambda}_n(x) - \lambda_G(x)]^2 = 0. \tag{5.5}$$

Recall that the second moment of a non negative r.v. Z is given by

$$E[Z^2] = \int_0^{\infty} 2t(1 - P(Z \leqslant t))dt.$$

It follows then

$$E_n[\tilde{\lambda}_n(x) - \lambda_G(x)]^2 = \int_0^{\infty} 2tP\Big(\Big|\tilde{\lambda}_n(x) - \lambda_G(x)\Big| > t\Big)dt,$$

$$= \int_0^{\lambda_G(x)} 2tP(\tilde{\lambda}_n(x) - \lambda_G(x) < -t)dt + \int_0^{1-\lambda_G(x)} 2tP(\tilde{\lambda}_n(x) - \lambda_G(x) > t)dt.$$

It suffices then, in order to prove (5.5), to show that $\forall\ t > 0$ both

$$\lim_{n\to\infty} P(\tilde{\lambda}_n(x) - \lambda_G(x) < -t) = 0 \text{ and } \lim_{n\to\infty} P(\tilde{\lambda}_n(x) - \lambda_G(x) > t) = 0.$$

Without loss of generality, let's consider the limit of the right tail probability. For $t > 0$, we rearrange the terms to get

$$P(\tilde{\lambda}_n(x) - \lambda_G(x) > t) = P\Big(\frac{\psi_n(x)}{q_n(x)} \wedge 1 - \frac{\psi_G(x)}{q_G(x)} > t\Big)$$

$$\leqslant P\Big(\frac{\psi_n(x)}{q_n(x)} - \frac{\psi_G(x)}{q_G(x)} > t\Big) = P\Big(\psi_n(x) - \Big(t + \frac{\psi_G(x)}{q_G(x)}\Big)q_n(x) > 0\Big)$$

$$= P\left([\psi_n(x) - \psi_G(x)] - \left(t + \frac{\psi_G(x)}{q_G(x)}\right)[q_n(x) - q_G(x)] > t q_G(x) \right). \tag{5.6}$$

Next we use the following inequality. For any r.v. V and W, and $c > 0$, then

$$P(V - W > c) \leqslant P\Big(V > \frac{c}{2}\Big) + P\Big(W < -\frac{c}{2}\Big). \tag{5.7}$$

Indeed, for $c > 0$

$$P(V - W > c) = P\Big(V - W > c, V > \frac{c}{2}\Big) + P\Big(V - W > c, V \leqslant \frac{c}{2}\Big)$$

$$\leqslant P\Big(V > \frac{c}{2}\Big) + P\Big(V - W > c, V \leqslant \frac{c}{2}, W < -\frac{c}{2}\Big) + P\Big(V - W > c, V \leqslant \frac{c}{2}, W \geqslant -\frac{c}{2}\Big).$$

But $P\Big(V - W > c, V \leqslant \frac{c}{2}, W \geqslant -\frac{c}{2}\Big) = 0$ so we see

$$P(V - W > c) \leqslant P\Big(V > \frac{c}{2}\Big) + P\Big(W < -\frac{c}{2}\Big).$$

Applying equation (5.7) to equation (5.6), we obtain

$$P(\tilde{\lambda}_n(x) - \lambda_G(x) > t) \leqslant P\Big(\psi_n(x) - \psi_G(x) > \frac{t q_G(x)}{2}\Big) + P\left(q_n(x) - q_G(x) < \frac{-t q_G(x)}{2(t + \frac{\psi_G(x)}{q_G(x)})}\right). \tag{5.8}$$

Now by Proposition 4, we have $E[\psi_n(x)] = \psi_G(x)$ and $E[q_n(x) = q_G(x)]$; applying Chebysher inequality and Proposition 4(ii), we have for $t > 0$

$$P(\psi_n(x) - \psi_G(x)) > \frac{t q_G(x)}{2} \leqslant \frac{Var[\psi_n(x)]^2}{t q_G(x)/2}$$

$$\leqslant \frac{4}{t^2 q_G^2(x)} \frac{\psi_G(x)}{n} \to 0 \text{ as } n \to \infty. \tag{5.9}$$

Similarly,

$$P(q_n(x) - q_G(x)) \leqslant \frac{-t q_G(x)}{2\Big(t + \frac{\psi_G(x)}{q_G(x)}\Big)} \leqslant \frac{4\Big(t + \frac{\psi_G(x)}{q_G(x)}\Big)^2}{t^2 q_G^2(x)} Var\Big[q_n(x)\Big]$$

$$\leqslant \frac{4\Big(t+\frac{\psi_G(x)}{q_G(x)}\Big)^2}{t^2 q_G^2(x)} \frac{q_G(x)}{a_\varphi(x)} \frac{1}{n} \to 0 \text{ as } n \to \infty. \tag{5.10}$$

Therefore by (5.6)-(5.10), for any $t > 0$ we conclude

$$\lim_{n\to\infty} P(\tilde{\lambda}_n(x) - \lambda_G(x) > t) = 0.$$

Following a similar process, one can show that for any $t > 0$

$$\lim_{n\to\infty} P(\tilde{\lambda}_n(x) - \lambda_G(x) < -t) = 0,$$

the details of which are left to the reader. Hence the proof is complete. □

Theorem 2. *Given* $0 \leqslant \frac{\psi_n(x)}{q_n(x)} \wedge 1, \ \frac{\psi_G(x)}{q_G(x)} \leqslant 1$, *suppose* $0 < \mu < 2$. *If the prior distribution G satisfies:*

$$\sum_{x=0}^{\infty} a_\varphi(x)[q_G(x)]^{1-\mu/2} < \infty,$$

then the EB estimator $\tilde{\lambda}_n(x)$ *is asymptotically optimal at a rate* $O(n^{-\mu/2})$.

Proof. Recall that

$$0 \leqslant \frac{\psi_n(x)}{q_n(x)} \wedge 1, \ \frac{\psi_G(x)}{q_G(x)} \leqslant 1.$$

Next, let $0 < \mu < 2$. Then,

$$E_n\left[\Big(\frac{\psi_n(x)}{q_n(x)} \wedge 1\Big) - \frac{\psi_G(x)}{q_G(x)}\right]^2 \leqslant E_n\left[\left|\frac{\psi_n(x)}{q_n(x)} - \frac{\psi_G(x)}{q_G(x)}\right| \wedge 1\right]^2$$

$$\leqslant E_n\left[\left|\frac{\psi_n(x)}{q_n(x)} - \frac{\psi_G(x)}{q_G(x)}\right| \wedge 1\right]^\mu. \tag{5.11}$$

Using Lemma 3.1 of Singh and Wei [17], we obtain

$$\leqslant E_n\left[\left|\frac{\psi_n(x)}{q_n(x)} - \frac{\psi_G(x)}{q_G(x)}\right| \wedge 1\right]^\mu$$

$$\leqslant 2[q_G(x)]^{-\mu}\left(E_n\Big[\big|\psi_n(x)-\psi_G(x)\big|\Big]^{\mu}+\left|\frac{\psi_G(x)}{q_G(x)}+1\right|^{\mu}E_n\Big[\big|q_n(x)-q_G(x)\big|\Big]^{\mu}\right)$$

$$\leqslant \frac{2}{q_G^{\mu}(x)}E_n\Big[\big|\psi_n(x)-\psi_G(x)\big|^{\mu}\Big]+\frac{2^{\mu+1}}{q_G^{\mu}(x)}E_n\Big[\big|q_n(x)-q_G(x)\big|^{\mu}\Big]. \tag{5.12}$$

Applying Proposition 4 we have for $0<\mu<2$

$$\begin{aligned}\frac{2}{q_G^{\mu}(x)}E_n\Big[\big|\psi_n(x)-\psi_G(x)\big|^{\mu}\Big] &\leqslant \frac{2}{q_G^{\mu}(x)}E_n\Big[\Big(\psi_n(x)-\psi_G(x)\Big)^{2\frac{\mu}{2}}\Big]\\ &\leqslant \frac{2}{q_G^{\mu}(x)}\left(E_n\Big[\big(\psi_n(x)-\psi_G(x)\big)^2\Big]\right)^{\frac{\mu}{2}} \quad \text{[by Jensen's inequality]}\\ &= \frac{2}{q_G^{\mu}(x)}\Big(Var_n\big[\psi_n(x)\big]\Big)^{\frac{\mu}{2}} \leqslant \frac{2}{q_G^{\mu}(x)}\Big(\frac{\psi_G(x)}{n}\Big)^{\frac{\mu}{2}} \sim n^{-\frac{\mu}{2}} \text{ as } n\to\infty.\end{aligned} \tag{5.13}$$

Similarly we have for $0<\mu<2$

$$\begin{aligned}\frac{2^{\mu+1}}{q_G^{\mu}(x)}E_n\Big[\big|q_n(x)-q_G(x)\big|^{\mu}\Big] &\leqslant \frac{2^{\mu+1}}{q_G^{\mu}(x)}E_n\Big[\big(q_n(x)-q_G(x)\big)^2\Big]^{\frac{\mu}{2}}\\ &= \frac{2^{\mu+1}}{q_G^{\mu}(x)}\Big(Var_n\big[q_n(x)\big]\Big)^{\frac{\mu}{2}} \leqslant \frac{2^{\mu+1}}{q_G^{\mu}(x)}\left(\frac{q_G(x)}{a_\varphi(x)n}\right)^{\frac{\mu}{2}} \sim n^{-\frac{\mu}{2}} \text{ as } n\to\infty.\end{aligned} \tag{5.14}$$

Substituting the inequalities (5.13) and (5.14) into (5.12), we obtain

$$\begin{aligned}E_n\left[\left(\frac{\psi_n(x)}{\psi_G(x)}\wedge 1\right)-\frac{\psi_G(x)}{q_G(x)}\right]^2 &\leqslant \frac{2(\psi_G(x))^{\mu/2}}{q_G^{\mu}(x)}\ \frac{1}{n^{\mu/2}}+\frac{2^{\mu+1}}{(q_G(x))^{\mu/2}(a_\varphi(x))^{\mu/2}}\ \frac{1}{n^{\mu/2}}\\ &\leqslant \frac{c(\varphi,\mu)}{n^{\mu/2}}\big(q_G(x)\big)^{-\mu/2},\end{aligned} \tag{5.15}$$

where $c(\varphi,\mu)=2\Big(1+\frac{2^{\mu}}{(a_\varphi(x))^{\mu/2}}\Big)$. Recall that for $X=x$,

$$S(\tilde{\lambda}_n(x)-\lambda_G(x))=\sum_{x=0}^{\infty}E_n\Big[\tilde{\lambda}_n(x)-\lambda_G(x)\Big]^2 p_G(x),$$

where $p_G(x) = a_\varphi(x)q_G(x)$. Substituting the inequality (5.15) into the above, we see that

$$\begin{aligned} S(\tilde{\lambda}_n(x) - \lambda_G(x)) &\leqslant \frac{c(\varphi,\mu)}{n^{\mu/2}} \sum_{x=0}^{\infty} a_\varphi(x) q_G(x) (q_G(x))^{\mu/2} \\ &= \frac{c(\varphi,\mu)}{n^{\mu/2}} \sum_{x=0}^{\infty} a_\varphi(x) (q_G(x))^{1-\mu/2} = O(n^{-\mu/2}), \end{aligned} \tag{5.16}$$

by our initial assumption that $\sum_{x=0}^{\infty} a_\varphi(x)[q_G(x)]^{1-\mu/2} < \infty$. Thus the proof of Theorem 2 is complete. □

Example 5.2.1. Epidemics can me modelled as a branching process as seen in Chapter I where the total progeny of the process is a Borel-Tanner r.v. Indeed, a viral outbreak starts with an initial number of infected, each of which go on to possibly spread the disease. This generates the process and the secondary infections become the new generation for the next branch of the process. In some case the number of initial infected is a known constant. Let's consider the case where the original number of infected is a random variable with distribution $Poisson(\tau)$. Then let the offspring (secondary infections) distribution be $Poisson(\lambda)$.

In most cases, the rate of secondary infections λ, is proportional to τ [5]. Hence we can reasonably assume that

$$\tau = \lambda\varphi, \text{ for some } \varphi \geqslant 0 \tag{5.17}$$

This results in a restricted generalized Poisson distribution with p.m.f. (5.3), which represents the probability that the total progeny of the branching process is equal to x infected individuals. It should be noted that when $\varphi = 0$ or $\tau = 0$, the restricted gpd becomes the standard Poisson distribution [5]. As $\lambda \to 1$, the disease is likely to become pandemic; otherwise the disease will die out. Let's consider then the sub-critical process where $\lambda < 1$. That is the total progeny of the process is finite with probability 1. It is therefore reasonable to also assume that $\exists\ \delta_0,\ \lambda_0$ such that $0 < \delta_0 \leqslant \lambda \leqslant \lambda_0 < 1$ and a prior distribution G such that $G(\lambda_0) = 1$. We will now show that the empirical Bayes estimate $\tilde{\lambda}_n(x)$ of λ in this case is asymptotically optimal, converging at a rate $O(n^{-\mu/2})$, as defined in Theorem 2.

We need to show that $\sum_{x=0}^{\infty} a_\varphi(x)\big[q_G(x)\big]^{2-\frac{\mu}{2}} < \infty$. Recall that

$$a_\varphi(x) = \frac{\varphi(x+\varphi)^{x-1}}{x!}.$$

By our assumption in 5.17, $\exists\, M \in \mathbb{Z}$ such that $0 \leqslant \varphi \leqslant M$. Then,

$$a_\varphi(x) = \frac{\varphi(x+\varphi)^{x-1}}{x!} \leqslant \frac{M(x+M)^{x-1}}{x!}.$$

Letting $y = x+M \Rightarrow x = y-M$, we see that,

$$\begin{aligned} a_\varphi(x) &\leqslant \frac{M(x+M)^{x-1}}{x!} \\ &= \frac{My^{y-M-1}}{(y-M)!} \\ &= \frac{My^y}{y^{M+1}} \cdot \frac{y(y-1)\cdots(y-M+1)}{y!}. \end{aligned} \tag{5.18}$$

Next, recall the Stirling Formula:

$$x! = x^{x+1/2}e^{-x+\varepsilon_x}\sqrt{2\pi},$$

where $\frac{1}{12x+1} < \varepsilon_x < \frac{1}{12x}$. Using the above formula, we can see that (5.18) becomes

$$\begin{aligned} a_\varphi(x) &= \frac{M}{y^{M+1}} \cdot \frac{y(y-1)\cdots(y-M+1)}{y^{1/2}e^{-y+\varepsilon_y}\sqrt{2\pi}} \\ &= \frac{Me^y}{y^{3/2}} \cdot \frac{y(y-1)\cdots(y-M+1)}{y^M e^{\varepsilon_y}\sqrt{2\pi}} \\ &= \frac{Me^{y-M+M}}{y^{3/2}} \cdot \frac{y(y-1)\cdots(y-M+1)}{y^M e^{\varepsilon_y}\sqrt{2\pi}} \\ &= \frac{Me^{y-M+M}}{y^{3/2}} \cdot \frac{y(y-1)\cdots(y-M+1)}{y^M e^{\varepsilon_y}\sqrt{2\pi}} \\ &\leqslant \frac{Me^M}{y^{3/2}e^{\varepsilon_y}\sqrt{2\pi}} e^{y-M}. \end{aligned} \tag{5.19}$$

Additionally, note that $\lambda e^{-\lambda}$ is an increasing function for $0 \leqslant \lambda \leqslant \lambda_0 < 1$. Indeed,

$$\frac{d}{d\lambda}[\lambda e^{-\lambda}] = e^{-\lambda}(1-\lambda^2) > 0 \text{ iff } 0 < \lambda < 1.$$

Since $\lambda \leqslant \lambda_0$ by our initial assumption, then

$$\begin{aligned} \lambda e^{-\lambda} &\leqslant \lambda_0 e^{-\lambda_0} \\ &\leqslant e^{2/(2-\mu)} \lambda e^{-\lambda} \leqslant e^{2/(2-\mu)} \lambda_0 e^{-\lambda_0} \\ &= 1. \end{aligned} \tag{5.20}$$

Note that the identity presented in (5.20) is true for $\mu = 2\big[1 - 1/(\lambda_0 - \ln \lambda_0)\big]$.Indeed,

$$e^{2/(2-\mu)} = \frac{e_0^{\lambda}}{\lambda_0}$$

$$\Rightarrow \ln e^{2/(2-\mu)} = \ln \frac{e_0^{\lambda}}{\lambda_0}$$

$$\Rightarrow \frac{2}{2-\mu} = \ln e^{\lambda} - \ln \lambda_0$$

$$\Rightarrow \mu = 2\big[1 - 1/(\lambda_0 - \ln \lambda_0)\big].$$

Suppose $0 < \mu < 2$. It follows that

$$\begin{aligned} [a_{\varphi}(x)]^{2/(2-\mu)} (\lambda e^{-\lambda})^x &\leqslant \left[\frac{Me^M}{y^{3/2} e^{\varepsilon_x} \sqrt{2\pi}}\right]^{2/(2-\mu)} (\lambda e^{-\lambda})^x \qquad \text{by (5.19)} \\ &\leqslant \left[\frac{Me^M}{y^{3/2} e^{\varepsilon_x} \sqrt{2\pi}}\right]^{2/(2-\mu)} (e^{2/(2-\mu)} \lambda e^{-\lambda})^x \qquad \text{by (5.20)} \\ &\leqslant \left[\frac{Me^M}{y^{3/2} e^{\varepsilon_x} \sqrt{2\pi}}\right]^{2/(2-\mu)} \end{aligned} \tag{5.21}$$

Lastly, recalling that the support of G is $[\delta_0, \lambda_0]$ such that $0 \leqslant \delta_0 \leqslant \lambda \leqslant \lambda_0 < 1$, we obtain

$$\begin{aligned}
\sum_{x=0}^{\infty} a_\varphi(x)[q_G(x)]^{1-\mu/2} &= \sum_{x=0}^{\infty} a_\varphi(x)\left[\int_{\delta_0}^{\lambda_0} \lambda^x e^{-\lambda(\varphi+x)} dG(\lambda)\right]^{1-\mu/2} \\
&= \sum_{x=0}^{\infty} a_\varphi(x)\left[\int_{\delta_0}^{\lambda_0} (\lambda e^{-\lambda})^x e^{-\lambda\varphi} dG(\lambda)\right]^{1-\mu/2} \\
&\leqslant \sum_{x=0}^{\infty} a_\varphi(x)\left[\int_{\delta_0}^{\lambda_0} (\lambda e^{-\lambda})^x dG(\lambda)\right]^{1-\mu/2} \\
&\leqslant \sum_{x=0}^{\infty}\left[\int_{\delta_0}^{\lambda_0} [a_\varphi(x)]^{2/(2-\mu)} (\lambda e^{-\lambda})^x dG(\lambda)\right]^{1-\mu/2} \\
&\leqslant \sum_{x=0}^{\infty}\left[\int_{\delta_0}^{\lambda_0} \left[\frac{Me^M}{y^{3/2} e^{\varepsilon_x}\sqrt{2\pi}}\right]^{2/(2-\mu)} dG(\lambda)\right]^{1-\mu/2} \qquad \text{by (5.20)} \\
&\leqslant \sum_{y=M}^{\infty} \frac{Me^M}{y^{3/2} e^{\varepsilon_x}\sqrt{2\pi}}\left(\int_{\delta_0}^{\lambda_0} \lambda^{-\varphi} dG(\lambda)\right)^{(2-\mu)/2} \qquad \text{where the integral is } < c < \infty \\
&\leqslant \frac{cMe^M}{\sqrt{2\pi}} \sum_{y=M}^{\infty} \frac{1}{y^{3/2}} < \infty \qquad \text{by p-series test.} \qquad (5.22)
\end{aligned}$$

Hence, we that Theorem 2 is satisfied. Therefore we conclude that the EB estimator $\tilde{\lambda}_n$ is asymptotically optimal at a rate of $O(n^{-\mu/2})$, where $0 < \mu = 2[1 - 1/(\lambda_0 - \ln\lambda_0)] < 2$.

CHAPTER VI

CLOSING REMARKS

In this paper, we were able to successfully develop Empirical Bayes estimates for the reproduction parameter θ with Borel-Tanner distribution. Our interest stemmed from applying branching processes as models of epidemic outbreaks where θ equals the average number of secondary infections caused by a host. More importantly, our Empirical Bayes estimate are useful in that they bypass the knowledge of an existing prior. This aligns with real life circumstances since due to the random nature of disease spread and the many variables that contribute to a pandemic, it is often difficult or impossible to ascertain a prior. In addition, our we were able to improve upon the empirical Bayes estimates by monotonizing; indeed the monotone the new monotone estimator is strictly better than the original empirical estimation, as evident by the smaller regret risk as seen in Chapter IV. The non-monotone empirical Bayes estimator θ_n turns out to be quite jumpy (see Figure 4.1) and does not having good small sample properties (see Table 4.1). Simulation results show that θ_n^* performs much better than θ_n, especially when the number of past observations and/or the epidemic size are small. This confirms the major positive effect of the monotonization procedure. One should also note that the choice of loss function, specifically linex loss, is incredibly powerful for epidemic analysis. Due to its asymmetric nature, we can choose how severely we penalize our estimation, especially when considering a framework where underestimating could have dire consequences. This is again supported in Chapter IV, Table 4.2, where the regret risk is improved upon for values of $b < 0$. This corresponds to penalizing underestimation more which in the context of a pandemic is where we want to be with the estimates.

But this was just the case where the initial number of infected, r, was known.To go one step further, we also consider the case where the initial number of infected was random. One

could relate such a situation to infected individuals entering a country with a communicable disease from different places. In such a case, it is quite difficult to determine the number of progenitors. When the initial number of ancestors is a Poisson r.v. and the offspring distribution is $Poisson(\lambda)$, then we end up with a restricted general Poisson distribution (GDP) as introduced in V. We then constructed an empirical Bayes estimate $\tilde{\lambda}_n$ for λ, using square error loss this time around for its simplicity. Additionally, we were able to show that $\psi_n(x)$ and $q_n(x)$ were both unbiased and consistent estimators for $\psi_G(x)$ and $q_G(x)$ respectively. Lastly we proved that the empirical Bayes estimate, $\tilde{\lambda}_n$, is asymptotically optimal at a rate $O(n^{-\mu/2})$ (see example at end of Chapter V).

BIBLIOGRAPHY

[1] *Randomization*. `https://explorable.com/randomization`, 8 2010. Accessed: Aug. 5 2019.

[2] K. ALBERTSEN, J. STEFFENSEN, AND E. KIRSTENSEN, *THREE PAPERS ON THE HISTORY OF BRANCHING PROCESSES*, Tech. Rep. 242, Department of Statistics, GN-22 University of Washington, Seattle, Washington, 11 1992. Translated from Danish by Peter Guttorp.

[3] D. J. ALDOUS, *Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists*, Bernoulli, 5 (1999), pp. 3–48.

[4] P. CARLIN, BRADLEY, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall,, 2nd ed., 2000.

[5] P. CONSUL, C., *Generalized Poisson Distribution Properties and Applications*, Marcel Dekker Inc, 1989.

[6] P. G. GIPPS, *An Abbreviated Procedure for Estimating Equilibrium Queue Lengths in Rural Two Lane Traffic.*, Transportation Science, 10 (1976), p. 337.

[7] F. A. HAIGHT AND M. A. BREUER, *The Borel-Tanner distribution*, Biometrika, 47 no. 1/2 (1960), p. 143.

[8] C. C. HEYDE, *ON ASSESSING THE POTENTIAL SEVERITY OF AN OUTBREAK OF A RARE INFECTIOUS DISEASE: A BAYESIAN APPROACH*, Australian Journal of Statistics, 21 (1979), pp. 282–292.

[9] T. IEŠMANTAS AND R. ALZBUTAS, *Bayesian assessment of electrical power transmission grid outage risk*, International Journal of Electrical Power & Energy Systems, 58 (2014), pp. 85 – 90.

[10] C. JACOB, *Branching Processes: Their Role in Epidemiology*, International Journal of Environmental Research and Public Health, 7 (2010), pp. 1186–1204.

[11] M. KIMMEL AND D. AXELROD, *Branching Processes in Biology*, Springer Publishing Company, Incorporated, 2nd ed., 2015.

[12] G. KOOREY, *Passing Opportunities at Slow-Vehicle Bays*, Journal of Transportation Engineering, 133 (2007), pp. 129–137.

[13] T. LIANG, *Empirical Bayes estimation for Borel–Tanner distributions*, Statistics & Probability Letters, 79 (2009), pp. 2212 – 2219.

[14] M. NIREI, T. STAMATIOU, AND V. SUSHKO, *Stochastic Herding in Financial Markets Evidence from Institutional Investor Equity Portfolios*, Tech. Rep. 371, Bank for International Settlements, 2 2012.

[15] H. ROBBINS, *The empirical Bayes approach to statistical decision problems*, The Annals of Mathematical Statistics, 35 (1964), pp. 1–20.

[16] S. H. SELLKE, N. B. SHROFF, AND S. BAGCHI, *Modeling and Automated Containment of Worms*, IEEE Transactions on Dependable and Secure Computing, 5 (2008), pp. 71–86.

[17] S. R. SINGH AND L. WEI, *Empirical Bayes with rates and best rates of convergence in $u(x)C(\theta)\exp -x/\theta$ family*, Annals of the Institute of Statistical Mathematics, 44 (1992), pp. 435–449.

[18] R. SOLTERO, CELESTINA, *Empirical Bayes Estimators and Borel-Tanner Distribution*, 2019.

[19] T. STIJNEN, *On the asymptotic behaviour of monotonized empirical Bayes rules*, PhD thesis, University of Utrecht, The Netherlands, 1980.

[20] J. C. VAN HOUWELINGEN, *Monotonizing empirical Bayes estimators for a class of discrete distributions with monotone likelihood ratio*, Statistica Neerlandica, 31 (1977), pp. 95–104.

[21] G. P. YANEV, *Statistical Modeling of Epidemic Disease Propagation via Branching Processes and Bayesian Inference*, PhD thesis, University of South Florida, Tampa, Florida, 2001.

[22] G. P. YANEV AND R. COLSON, *Monotone Empirical Bayes Estimators for the Reproduction Number in Borel-Tanner Distribution*, Pliska Studia Mathematica, 27 (2017), pp. 115–122.

APPENDIX A

APPENDIX A

```r

if(!require(VGAM)) install.packages("VGAM",repos = "http://cran.us.r-project.org") #install required package
library(VGAM)
options(scipen = 999) # suppresses scientific notation in the output

######################################## Predefined Items ##################################################
# Define the lower and upper bounds of G prior ~ Uniform(a,b)
a=.5
b=.8

# Initialize Parameters
B=2 # B>0, the loss function penalizes overestimation more than underestimation
r=3 # initial outbreak size
a1=min(c(exp(-B),1)) # lower bound for EB estimate
a2=max(c(exp(-B),1)) # upper bound for EB estimate
Xmax=20 # max no. current OB size
n=20 # no. of past observations
kmax=100 # k=1,2,3,...

# Create storage vars
x=r:Xmax # current observations
X=matrix(0,n,1) # past observations
Cx=matrix(0,Xmax-r+1,1) # BT c_r(x) coefficient for current OBS
CX=matrix(0,n,1) # BT c_r(x) coefficient for past OBS
Cxk=0 # BT c_k(X_j-x)
thetaG=matrix(0,n,1) # theta from G: Uni(a,b)

BE=matrix(0,Xmax-r+1,1) # Bayes Estimate for x=r,r+1,...,Xmax

mle=0 # maximum likelihood estimator (MLE)
Rmle=0 # risk of MLE
RBE=0 # Risk of Bayesian Estimate

######################## Bayesian Estimate and MLE Estimate ################

## Bayes Estimate
for (i in 1:length(x)) {
  integrand1<-function(theta){theta^(x[i]-r)*exp(-x[i]*theta)}
  integrand2<-function(theta){exp(-B*theta)*theta^(x[i]-r)*exp(-x[i]*theta)}
  qGx=(1/(b-a))*integrate(integrand1,lower=a,upper=b)$value                  # compute q_g(x)
  pGx=(1/(b-a))*integrate(integrand2,lower=a,upper=b)$value                  # compute psi_G(x)
  BE[i]=-(1/B)*log(pGx/qGx)                                                  # Bayes Estimate for x=r,r+1,...
}

# Graphing the Bayes Estimate
plot(x,BE,type="b",main="Bayes Estimate",xlab="Current Outbreak Size",ylab="Reproductive Number",ylim=c(.4,.8),col="blue",pch=16)

# Compute C_r(x) for currOBS
for(i in 1:length(x)){
  Cx[i]=r*x[i]^(x[i]-r-1)/factorial(x[i]-r)
}

# Compute the Risks for MLE and BE
for (i in 1:length(x)) {
  mle[i]=(x[i]-r)/x[i]
  intTEMP1=function(theta){(exp(B*(mle[i]-theta))-B*(mle[i]-theta)-1)*theta^(x[i]-r)*exp(-x[i]*theta)}
  intTEMP2=function(theta){(exp(B*(BE[i]-theta))-B*(BE[i]-theta)-1)*theta^(x[i]-r)*exp(-x[i]*theta)}
  intTEMP3=function(theta){(exp(B*(BE[i]-theta))-B*(BE[i]-theta)-1)*theta^(x[i]-r)*exp(-x[i]*theta)}
  Rmle[i]=(Cx[i]/(b-a))*integrate(intTEMP1,a,b)$value
  RBE[i]=(Cx[i]/(b-a))*integrate(intTEMP2,a,b)$value
}
```

```r
Rmle=sum(Rmle)      # Bayes risk of MLE
RBE=sum(RBE)        # Minimum Bayes risk
Regretmle=Rmle-RBE  # Regret risk of MLE

#################################### Empirical Bayes ################################################

## Simulate Data
# Seed RNG
seeds=c(9,28,54,55,976,442,28,1,151,38) #can use any seed to produce different realizations of empirical data

#Create storage vars related to Empirical Estimate
qI=matrix(0,length(x),n) # q_n(x) Indicator for Empirical Bayes Estimate
qn=matrix(0,length(x),1) # q_n(x) denominator for Empirical Bayes Estimate
psin=matrix(0,length(x),1) # psi_n(x,b) numerator for Empirical Bayes Estimate
EBEtemp=matrix(0,1,length(x))
EBE=matrix(0,length(seeds),length(x)) # EB estimate
EBERisks=0 # Bayes risk of Empirical estimate

m=1  # set counter for while loop
while (m<=length(seeds)) {
  set.seed(seeds[m])  # sets current seed for the m'th iteration
  # Simulate past observations and draw random theta from G-prior
  for (i in 1:n){
    TEMP=runif(1,a,b) # draws random theta from G: Uni(a,b)
    thetaG[i]=TEMP # matrix of random theta values
    X[i]=rbort(1, Qsize=r, a=TEMP) # observed past total OB sizes
    CX[i]=r*X[i]^(X[i]-r-1)/factorial(X[i]-r) #BT C_r(x) coefficient for pastOBS
  }

  ## Empirical Bayes Estimate
  # Compute q_{n}(x)
  for (i in 1:length(x)){
    for (j in 1:n) {
      qI[i,j]=as.numeric(I(X[j]==x[i]))/CX[i]  # check Indicator to compute q_n(x)
    }
    qn[i]=sum(qI[i,])/n  # compute q_n(x)
  }

  # Compute psi_nj(x,B)
  psik=matrix(0,n,kmax) #matrix for psi_njk values
  psinj=matrix(0,length(x),n) #matrix for psi_nj(x,b) values
  for (i in 1:length(x)) {
    for (j in 1:n) {                  #compute psi_njk
      for (k in 1:kmax) {
        DIFF=X[j]-x[i]
        if(DIFF>=k){Cxk=k*DIFF^(DIFF-k-1)/factorial(DIFF-k)}       # check Indicator for instances when X>= x+k
        else{Cxk=0}                                                # if not true, set coefficient to 0
        result=((-B)^k/factorial(k))*Cxk*as.numeric(I(X[j]>=x[i]+k))/CX[j]
        psik[j,k]=result
      }
      psinj[i,j]=qI[i,j]+sum(psik[j,])   # compute psi_nj
    }
  }

  # Compute psi_n(x,b) and the EB estimate
  ratio=matrix(0,length(x),1)
  for (i in 1:length(x)) {
    psin[i]=sum(psinj[i,])/n
    if (qn[i]==0){ratio[i]=0}     # to avoid undefined results, set ratio=0 if q_n(x)=0
    else {ratio[i]=psin[i]/qn[i]}
    EBEtemp[i]=(-1/B)*log(min(a2,(max(a1,ratio[i]))))
  }
```

```
  EBE[m,]=EBEtemp  # Empirical Estimate for each x=r,r+1,...

  # Graphing the Empirical Bayes Estimate
  #plot(x,EBEtemp,type="b",main="Empirical Bayes Estimate",xlab="Current Outbreak Size",ylab="Reproductive Number",ylim=c(0,1),col="orange",pch=16)

  # Risks of BE and MLE
  REB=0
  for (i in 1:length(x)) {
    intEBE=function(theta){(exp(B*(EBEtemp[i]-theta))-B*(EBEtemp[i]-theta)-1)*theta^(x[i]-r)*exp(-x[i]*theta)}
    REB[i]=(Cx[i]/(b-a))*integrate(intEBE,a,b)$value
  }
  EBERisks[m]=sum(REB)
  m=m+1      # update counter to next iteration
}

REBE=sum(EBERisks)/length(seeds) # compute the average Bayes risk of the Empirical estimate across all simulations
RREBE=REBE-RBE      # Regret risk of the Empirical estimate

############ Monotonized Empirical Bayes Estimator ##################
pmax=100
alpha=matrix(0,length(seeds),pmax)
aG=seq(0,1,length=pmax+2)  # create a grid for a-G
aG=aG[-1]
aG=aG[-101]

for (j in 1:length(seeds)) {
  for (i in 1:pmax) {
    for (k in 1:length(x)) {
      if(EBE[j,k]<=aG[i]){      # verify if EBE/,= a-grid value
        alpha[j,i]=alpha[j,i]+sum(dbort(x[k],r,aG[i])) # compute alpha as outlined in Houwalingen
        }
      else{alpha[j,i]=alpha[j,i]}
    }
  }
}

FBT=matrix(0,length(x),pmax)
for (i in 1:pmax) {
  FBT[1,i]=dbort(x[1],r,aG[i])                # BT cdf for x=r
  for (k in 2:length(x)) {
    FBT[k,i]=FBT[k-1,i]+dbort(x[k],r,aG[i])   # BT cdf for x>r
  }
}

Dstar=matrix(0,length(x),pmax)
Dtail=matrix(0,length(x),pmax)
tempmEB=matrix(0,1,length(x))
mEB=matrix(0,length(seeds),length(x))
RmEB=matrix(0,1,length(x))
mEBRisks=0

# Define D^*(a,x)
j=1
while (j<=length(seeds)) {
  for (i in 1:pmax) {
    if(alpha[j,i]>FBT[1,i]){Dstar[1,i]=1} # case for x=r
    else{Dstar[1,i]=alpha[j,i]/FBT[1,i]}
    for (k in 2:length(x)) {
      if(FBT[k-1,i]>alpha[j,i]){Dstar[k,i]=0} # case for x>r
      else if(FBT[k,i]<alpha[j,i]){Dstar[k,i]=1}
      else{Dstar[k,i]=(alpha[j,i]-FBT[k-1,i])/(FBT[k,i]-FBT[k-1,i])}
    }
```

```
  }

  for (l in 1:length(x)) {
    for (i in 1:pmax) {
      Dtail[l,i]=1-Dstar[l,i]
      tempmEB[l]=sum(Dtail[l,])/pmax  # define monotonized Empirical estimate
    }
  }

  for (i in 1:length(x)) {
    intEBE=function(theta){(exp(B*(tempmEB[i]-theta))-B*(tempmEB[i]-theta)-1)*theta^(x[i]-r)*exp(-x[i]*theta)}
    RmEB[i]=(Cx[i]/(b-a))*integrate(intEBE,a,b)$value
  }
  mEBRisks[j]=sum(RmEB) # compute the Bayes risk for the monotonized estimate
  mEB[j,]=tempmEB       # monotonized Empirical Bayes Estimate
  j=j+1  # update counter
}

RiskmEB=sum(mEBRisks)/length(seeds)  # compute average Bayes risk for the monotonized EBE
RRmEB=RiskmEB-RBE # compute the regret risk for the monotonized EBE

########### Comparing Regret Risks of estmators MLE, EB, and monoEB #################
Regretmle
RREBE
RRmEB

# newstored=cbind.data.frame("n"=n,"RegretMLE"=Regretmle,"RatioEBE"=(RREBE-Regretmle)/Regretmle*100,"RatioMEB"=(RRmEB-Regretmle)/Regretmle*100)
# stored=rbind(stored,newstored)
#
# bnewstored=cbind.data.frame("b"=B,"RegretMLE"=Regretmle,"RatioEBE"=(RREBE-Regretmle)/Regretmle*100,"RatioMEB"=(RRmEB-Regretmle)/Regretmle*100)
# bstored=rbind(bstored,bnewstored)

# Compute average estimate for x=r,r+1,... for the Empirical Bayes and monotone estimates
aveMEB=matrix(0,1,length(x))
aveEBE=matrix(0,1,length(x))
for (z in 1:length(x)) {
  aveEBE[z]=sum(EBE[,z])/length(seeds)
  aveMEB[z]=sum(mEB[,z])/length(seeds)
}

# Average Estimate Comparison plot
plot(x,mle,type="b",main="Estimate Comparison",xlab="Current Outbreak Size",ylab="Reproductive Number",ylim=c(0,1),col="blue",pch=16)
lines(x,aveEBE,type="b",col="orange",pch=16)
lines(x,aveMEB,type="b",col="red",pch=16)
legend(14, .25, legend=c("MLE", "Empricial", "Monotonized"),col=c("blue", "orange", "red"), lty=c(1,1,1), cex=0.8)
```

Table 1.1: References on notation

Notation	**Description**
Θ	unknown rv parametrizing X; the reproduction number
θ	a realization of the reproduction parameter Θ
$\hat{\theta}$	refers to any estimator
θ_{mle}	Maximum likelihood estimator for BT distribution
θ_n	Empirical Bayes estimator for BT based on the procedure found in [13]
θ_n^*	Monotonized EB estimator for BT based on the procedure found in [20]
$\underline{X}$	the set of n observations $X_1, X_2, \ldots, X_n$
$Poi(\lambda)$	Poisson distribution with parameter λ
φ	a weight $\geqslant 1$ of the Poissonian parameter λ
$R(G, \hat{\theta})$	Bayes risk for estimator $\hat{\theta}$ under $G-$prior
$R(\hat{\theta})$	Regret risk for estimator $\hat{\theta}$
$\hat{S}(\hat{\theta})$	Average regret risk for estimator $\hat{\theta}$
$Uni(a,b)$	Uniform distribution with parameters (a,b)
BT	Borel–Tanner distribution
cdf	cummulative distribution function
EB	Empirical Bayes
GW	Galton–Watson also known as Bienaymé–Galton–Watson
iid	independent identically distributed
MLE	maximum likelihood estimator
MLR	monotone likelihood ratio
pmf	probability mass function
rv	random variable

BIOGRAPHICAL SKETCH

Elijah Hight was born in Bluefield, West Virginia and spent his childhood loving all that was science, math and art. He received a Bachelor of Science in Applied Mathematics in 2019. In the Summer of 2021 he completed a Master of Science in Mathematics-Statistics at University of Texas Rio Grande Valley. He will begin work for the Policy and Analytics Branch of the Headquarters Marine Corps, Marine and Family Programs Division in Quantico, VA.

Elijah may be contacted via email at elijah.hight0228@outlook.com.