2014

The Relative Advantage of Knowing an Initiation Time or a Mean Time When Estimating a Parameter from an Exponential Distribution Using Type-II Left Censored Data

Christopher Haines
Lehigh University

The Relative Advantage of Knowing an Initiation Time or a Mean Time When Estimating a Parameter from an Exponential Distribution Using Type-II Left Censored Data

by

Christopher A Haines

A Dissertation
Presented to the Graduate Committee
of Lehigh University
in Candidacy for the Degree of
Doctor of Philosophy
in
Department of Mathematics

Lehigh University
January 2014

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Christopher Haines
The Relative Advantage of Knowing an Initiation Time or a Mean Time When Estimating a Parameter from an Exponential Distribution Using Type-II Left Censored Data

____
Defense Date

____
Dr. Bennett Eisenberg
Dissertation Director

____
Approved Date

Committee Members:

____
Dr. Wei-Min Huang (Member)

____
Dr. Ping-Shi Wu (Member)

____
Dr. Gary Harlow (Outside Member)

Contents

List of Tables

List of Figures

Abstract

In this thesis, the model being considered is a left Type-II censoring scheme with the underlying density being of the form

$$f(x) = \frac{1}{\mu} \exp\left(-\frac{1}{\mu}(x-\theta)\right), \; x \geq \theta.$$

This is what is known as an exponential distribution with scale parameter μ and location parameter θ. Our main focus is on the advantage of knowing one of the parameters in point estimation. For example, we ask to what advantage does an observer have in knowing θ when estimating μ to one who does not know either of the parameters. Our criteria between comparisons is Mean Square Error. (MSE) One of the most interesting results is that the relative advantage in knowing θ when estimating μ is the same as the relative advantage in knowing μ when estimating θ. Essentially all of our work revolves around considering a given proportion of the data that is left censored and determining the asymptotic MSE of our estimators. In the work particularly done by Balakrishnan and Cohen, (See [4].) they derive formulas for the MSEs of the estimators with a much more general doubly Type-II censoring scheme, but they do not fix the proportion being censored and ask questions relating to the asymptotic nature of the MSEs for the estimators. The beauty of the ratio identity is that the limiting ratio turns out to be very close to the function $y = p$, where $1 - p$ is the proportion of the data that is censored. This is a consequence of two of our estimators nearly attaining the Cramer-Rao Lower Bound (CRLB) based on all the data for fairly small values of p. Another interesting result is how close the MSEs of these two estimators are as a function of p asymptotically.

Chapter 1

Introduction

1.1 The Model

Suppose we have a store which n customers all arrive at the same opening time θ. Not taking in to account the time each customer takes to pick out an item or items, or how many items the customer chooses, we simply assume that all n of them are in a line for check out at time θ. The time for each one to be checked out is an exponential random variable with mean $\mu > 0$. Let's call the service times $W_1, W_2, ..., W_n$ and accordingly, assume that they are *iid* $\exp(\mu)$ random variables. Denote the actual times that the customers check out as $X_1, X_2, ..., X_n$, and note that therefore the $X's$ are *iid* $\exp(\mu, \theta)$ random variables. We can think of each as $X_i = W_i + \theta$ for $i = 1, 2, ..., n$. Also denote the order statistics for the times of check out as $Y_1, Y_2, ..., Y_n$, where $Y_1 > Y_2 > .... > Y_{n-1} > Y_n$. As a result of some of the check out times being lost, we only have knowledge of the last k check out times $Y_1, Y_2, ..., Y_k$, where $2 \leq k \leq n-1$. We then have a Type-II left-censored model with underlying distribution function $F(x) = 1 - \exp\left(-\frac{x-\theta}{\mu}\right)$, $x \geq \theta$.

1.2 History of Point Estimation of Scale Parameter μ (mean μ in exponential case)

In the literature of two-parameter exponential distributions, the point estimation inference has always been focused on the parameter μ only. Balakrishnan among several other authors (See [1].) have done point estimation of μ with doubly Type-II censored models (a generalization to the model described in Section 0.1) for several two-parameter continuous distributions (including the two-parameter exponential distributions) such as normal, logistic, Gamma, Cauchy and Weibull as well as exponential, and the point estimation has either been the method of finding the Best Linear Unbiased Estimator (BLUE) or finding the Maximum Likelihood Estimator. (MLE) (See [1] and [2].) The meaning of μ and θ for these distributions varies. For instance in the case of normal, θ would be the mean, and μ would be the standard deviation. In all of the cases however, μ is the scale parameter, and θ is the location parameter. Of course in the case of the normal distribution, it is very common for the point estimation to be more focused on θ than μ. However as we stated in the opening sentence, point estimation in the two-parameter exponential distributions has always been stressed on the scale parameter (mean in this case) μ because it represents the mean failure time.

1.3 The Primary Theme and Objectives

Suppose we are estimating the mean service time for each customer, which is the parameter μ based on the uncensored data $Y_1, Y_2, ..., Y_k$. Define $T_j = j\left(Y_j - Y_{j+1}\right)$ for $j = 1, 2, ..., k-1$. Then $T_1, T_2, ..., T_{k-1}$ are *iid* $\exp\left(\mu\right)$ random variables, and we can refer to them as the standardized intervals between ordered times of service. (i.e., order statistics, alternatively) A natural unbiased estimator for μ is then

$$\overline{T} = \frac{\sum_{j=1}^{k-1} T_j}{k-1}. \quad (1.1)$$

Its variance is equal to

$$Var\left(\overline{T}\right) = \frac{\mu^2}{k-1} > \frac{\mu^2}{n} = Var\left(\overline{X}\right),$$

and μ^2/n is the Cramer-Rao Lower Bound (CRLB) based on the entire sample. Notice 1.1 does not include the value of Y_k, so all the data is not being used. However if the store opening time θ is unknown to the one reviewing the times of transaction, Y_k has no meaning as far as the estimation of μ is concerned. It turns out that when θ is unknown, (and therefore we are dealing with the two-parameter exponential distributions.) 1.1 is the BLUE based on the data $Y_1, Y_2, ..., Y_k$. If θ is known, then $Y_k - \theta$ does help in estimating μ, and therefore of course, so does Y_k. This is actually a consequence of the independence of the spacings for the exponential distribution. This then reduces to a one-parameter exponential distribution involving only the parameter μ, and in that case, the BLUE does take in to account Y_k as well as $Y_1, Y_2, ..., Y_{k-1}$. The BLUE for μ turns out to be a weighted linear combination (depending on k) of $\overline{T}$ and Y_k. So it has a lower variance than the case when θ is unknown simply because all of the data is being used. The MLEs have very similar expressions to the BLUEs as well as the same asymptotic variances as we will see in the latter chapters. The first question then is by how much does knowing θ reduce the variance of the estimator for μ. Now if we were upper management for the retail store, we would not only be concerned about the mean service time, but we would also want to ensure the store was opening precisely on time. This would be only of course if θ were unknown, and we were also interested in estimating it based on $Y_1, Y_2, ..., Y_k$. Not knowing μ would put us at a disadvantage in estimating θ in this case. If we did not know μ, we would be back to the case where neither of the parameters is known, so therefore the BLUE for μ would be $\overline{T}$ once again. Y_k is not used in the estimation of μ when both parameters are unknown, but since we can estimate μ, Y_k does come in to good use for estimating θ in conjunction with $\overline{T}$. It turns out that the BLUE for the two-parameter model of θ does depend on all of the data given, and this is intuitive, but only follows since the spacings between order statistics are independent. Otherwise, the use of the mean service time μ will not help. This can be seen merely by telescoping $Y_k - \theta$ as a sum of independent exponential random variables as follows.

$$\begin{aligned} Y_k &= Y_k - \theta + \theta \\ &= \sum_{j=k}^{n-1} (Y_j - Y_{j+1}) + Y_n - \theta + \theta \ (1.2) \end{aligned}$$

Taking the expectation on both sides of 1.2 and solving for θ, we have that

$$\theta = E(Y_k) - \mu \sum_{j=k}^{n} \frac{1}{j}.$$

Substituting Y_k for $E(Y_k)$ and $\overline{T}$ for μ, we arrive at a logical guess of

$$Y_k - \overline{T} \sum_{j=k}^{n} \frac{1}{j} \ (1.3)$$

as the BLUE for θ, which in fact it is. For distributions other than exponential however, it would not follow because of the nonzero covariance terms between the $n-k+1$ random variables $Y_k - Y_{k+1}, Y_{k+1} - Y_{k+2}, ..., Y_{n-1} - Y_n$ and $Y_n - \theta$. By the logic of 1.2, we would be looking at an expression such as

$$E(Y_k) - \sum_{j=k}^{n-1} E(Y_j - Y_{j+1}) - E(Y_n - \theta),$$

and while the calculation of the expectations can be done rather tediously, the substitution would not be immediate since the variance of the sum

$$Y_k - \sum_{j=k}^{n-1} (Y_j - Y_{j+1}) - Y_n - \theta$$

may not be minimized due to the nonzero covariances of the spacings. With the normal distribution for example, 1.3 is *not* the BLUE for the location parameter θ in the two-parameter normal case. When μ is known, 1.3 obviously should be replaced by

$$Y_k - \mu \sum_{j=k}^{n} \frac{1}{j}, \ (1.4)$$

and just at a glance, we can see that the variance of 1.4 is lower than the variance of 1.3. Again, the question is to how much of a benefit does knowing μ have in estimating θ, which can be seen by comparing the variance of 1.4 to the variance of 1.3. Both 1.3 and 1.4 can be understood as using the lowest of the order statistics Y_k and projecting down based on the expected times between failures leading up to the k^{th} to last failure Y_k. Just as with the two estimators for μ, the MLE for θ in each case appears similar to 1.3 and 1.4, respectively. In summary, when estimating the mean μ, the variance of the BLUE and AMLE for μ are lower when knowing θ than when not knowing θ. Similarly, the variance of the BLUE and MLE for θ are lower when knowing μ than when not knowing μ.

1.4 The Unexpected Results

Below are all the results which after careful analysis were discovered, and we will mention them again in the chapters to come. All of these results came on an unexpected notice, but we stress on all of them quite a bit throughout the thesis.

Primary Results

i. If we are estimating μ and θ is known, the variance of the BLUE comes fairly close to the Cramer Rao Lower Bound of μ^2/n (when all of the data is observed) when only 20% or more of the upper order statistics are observed. We illustrate this by an example at the conclusion of this section in the case of $n = 4$ and $k = 2$. This appears in Chapter 2, and we give an explanation for the reason for this in Section 4.6.

ii. The ratio of the variance of the BLUE of μ when θ is known to when θ unknown has an identical formula to the ratio of the variance of the BLUEs of θ when μ is known to when μ is unknown. (Known divided by unknown in both cases) This appears in Chapter 3. There is an important interpretation of symmetry for this result which we illustrate in Chapter 3.

iii. We use two methods of point estimation in this thesis. They are the methods of Maximum Likelihood Estimation and Best Linear Unbiased Estimation. (MLE and BLUE, respectively) In Chapter 4, we question how our simpler BLUEs compare to their respective MLEs with respect to Mean Square Error. (MSE) As it turns out, the BLUEs have asymptotically the same MSEs as their corresponding MLEs at the order of $1/n$ and at the order of $1/n^2$, two out of four cases, the MLE is superior, one out of the four, we conjecture that the BLUE is superior, and the fourth case which one has the lower MSE depends on the value of p.

1.5 Example of Point Estimation

Suppose there are $n = 4$ customers in the store, and we would like to estimate μ based on Y_1 and Y_2 using the method of finding the BLUE based on Y_1 and Y_2.

In one scenario, we are given the opening time, which we can assume without loss of generality to be $\theta = 0$, while in the other case, θ is unknown. We ask how much is the variance reduced if we knew that $\theta = 0$ as in the former case. In this example, we introduce some notation. Let $\widehat{\mu}_B$ and $\widetilde{\mu}_B$ be the BLUEs when θ is known (as zero) and θ is unknown, respectively. We first want to calculate $\widehat{\mu}_B = a_1 Y_1 + a_2 Y_2$, where a_1 and a_2 are constants which minimize the variance of $\widehat{\mu}_B$ with the constraint that $E\left(\widehat{\mu}_B\right) = \mu$. We can take advantage of the fact that $(Y_1 - Y_2, Y_2)$ span the same subspace (or have exactly the same information) as (Y_1, Y_2) and the fact that $Y_1 - Y_2$ and Y_2 are independent random variables. First note that clearly $Var\left(Y_1 - Y_2\right) = \mu^2$, and

$$\begin{aligned} Var\left(Y_2\right) &= Var\left(Y_2 - Y_3\right) + Var\left(Y_3 - Y_4\right) + Var\left(Y_4\right) \\ &= \mu^2 \left(\frac{1}{4} + \frac{1}{9} + \frac{1}{16}\right) \\ &= \mu^2 \left(\frac{36 + 16 + 9}{144}\right) \\ &= \frac{61}{144}\mu^2. \end{aligned}$$

As we will see in Chapter 3, we can apply a theorem (which we state in Chapter 3) to calculate the BLUE for μ. The two conditions required in the theorem is for $Y_1 - Y_2$ and $Y_2^* = \frac{Y_2}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}}$ to be independent and both unbiased for μ, and in fact with these two random variables, those two conditions are satisfied. Note that $Y_2^* = \frac{Y_2}{\frac{12+8+6}{24}} = \frac{12}{13}Y_2$, and so

$$Var\left(Y_2^*\right) = \frac{144}{169}\frac{61}{144}\mu^2 = \frac{61}{169}\mu^2.$$

Since $Y_1 - Y_2$ and Y_2^* are unbiased for μ, $\widehat{\mu}_B = c\left(Y_1 - Y_2\right) + \left(1 - c\right)Y_2^*$ for some $c \in \mathcal{R}$. To minimize the variance of $\widehat{\mu}_B$, we simply differentiate with respect to c and solve for c when the derivative is set to zero. First note that

$$Var\left(\widehat{\mu}_B\right) = c^2\mu^2 + \left(1-c\right)^2\frac{61}{169}\mu^2,$$

and then differentiating with respect to c and setting the resulting expression equal to zero, we have that

$$\begin{aligned}\mu^2\left(2c - \frac{122}{169}\left(1-c\right)\right) &= 0, \text{ or}\\ \frac{460}{169}c &= \frac{122}{169},\\ c &= \frac{61}{230}.\end{aligned}$$

So therefore,

$$\begin{aligned}\widehat{\mu}_B &= \frac{61}{230}\left(Y_1 - Y_2\right) + \frac{169}{230}Y_2^*\\ &= \frac{61}{230}Y_1 + \frac{169}{230}\frac{156}{169}Y_2\\ &= \frac{61}{230}Y_1 + \frac{78}{115}Y_2,\end{aligned}$$

and

$$
\begin{aligned}
Var\left(\widehat{\mu}_B\right) &= \left(\frac{61}{230}\right)^2 \mu^2 + \left(\frac{169}{230}\right)^2 \frac{61}{169}\mu^2 \\
&= \frac{61^2 + (169)(61)}{(230)^2}\mu^2 \\
&= \frac{61(230)}{(230)^2}\mu^2 \\
&= \frac{61}{230}\mu^2.
\end{aligned}
$$

It can be shown that $\widetilde{\mu}_B = Y_1 - Y_2$, and this is intuitive because we have no idea what $Y_2 - \theta$ is, since in this case, θ is unknown. Therefore, the ratio of the two BLUEs is equal to $r = \frac{61}{230}$, which is much less than expected, which we will later see to be around 0.5. This is only because n is so small. That is an implication of i. in Section 1.3, so we would expect the variance of $\widehat{\mu}_B$ to not be anywhere near the CRLB based on all the data. However,

$$
\begin{aligned}
\frac{\frac{\mu^2}{4}}{Var\left(\widehat{\mu}_B\right)} &= \frac{\frac{230}{920}}{\frac{244}{920}} \\
&= \frac{230}{244} \\
&\approx 0.942\,62,
\end{aligned}
$$

which relatively speaking is fairly close to the CRLB based on all the data, but not close enough to make the ratio $Var\left(\widehat{\mu}_B\right)/Var\left(\widetilde{\mu}_B\right)$ near 0.5 because n is small. ($k/n = 0.5$ in that example) Although, this does exemplify what is claimed in i. of Section 1.3.

1.6 Notation

We will use the following notation for each of the three scenarios for the BLUEs and/or MLEs of μ and θ. The subscript 'M' similarly stands for MLE, and the subscript "AM" stands for Approximate Maximum Likelihood Estimator. (AMLE) The BLUE and AMLE for μ when θ is known we denote as $\widehat{\mu}_B$ and $\widehat{\mu}_{AM}$, respectively.

The BLUE and MLE for μ when θ is unknown we denote as $\widetilde{\mu}_B$ and $\widetilde{\mu}_M$, respectively. When both parameters are unknown, the BLUE and MLE for θ we denote as $\widetilde{\theta}_B$ and $\widetilde{\theta}_M$, respectively. Finally when θ is unknown and μ is known, we denote the BLUE and MLE for θ as $\widehat{\theta}_B$ and $\widehat{\theta}_M$, respectively.

Chapter 2

The Advantage of Knowing θ When Estimating μ

2.1 Introduction:

Recall the store opening model described in Chapter 1. In probabalistic terminology, we have n *iid* random variables $X_1, X_2, ..., X_n$ which are of the form $X_i = W_i + \theta$ for $1 \leq i \leq n$, where $\theta \in \mathcal{R}$, and $W_1, W_2, ..., W_n$ are *iid* $\exp(\mu)$ random variables. As stated in Chapter 1 if θ is known, then the last k times of service *since the store opened* can be determined by simply subtracting θ from each of the order statistics Y_j, $1 \leq j \leq k$. Even though θ may not be zero, we may assume it to be zero in the case of θ being known without loss of generality as we discussed in Chapter 1. On the contrary if θ is unknown, then the last k failure times $Y_1, Y_2, ..., Y_k$ do not *necessarily* represent the last k failure times unless $\theta = 0$. We would like to estimate μ in both situations of θ being known and unknown, and thus the question of how much the variance is reduced if we are given the value of θ immediately arises. We would expect the variance of the estimator for μ to be reduced since if θ is unknown, it would only makes sense to use the standardized $k-1$ spacings defined as $T_j = j(Y_j - Y_{j+1})$ for $1 \leq j \leq k-1$. This is not equivalent to the complete data set we have, and by intuition, what should be excluded is Y_k, since it does not represent the k^{th} to last failure time-it only represents the k^{th} to last time on the clock from the point of view of someone who does not know θ. On the contrary if θ is known, the entire recorded data set should be taken in to account since then $Y_k - \theta$ does represent the k^{th} to last failure time (or just Y_k if we assume $\theta = 0$). In Section 2.2, we derive formulas for the BLUEs and their variances under the assumption of θ being known and unknown. In Section 2.3 we give asymptotic formulas for these variances in the case of $k = [np]$. The reason we choose $k = [np]$ is to acquire insight of how the variances behave as a function of the sample percentiles for finite n. For moderately large values of n, these variances are fairly close to their respective limits for each value of p. The purpose of Section 2.3 is to compare the asymptotic normalized variances of the two BLUEs graphically and numerically, and as expected the BLUE of μ when θ is known has a much lower variance.

The surprising result (which we mentioned in Section 1.4, part i.) is how small a percentage of the upper order statistics we can observe and still come relatively close to the CRLB based on all the $X's$. This is suggested in Section 2.3, and thus Section 2.3 is really a preview of what is to come in Section 2.4. Roughly speaking, if the upper 20% of the data is all we have recorded, the variance of the BLUE when θ is known is less than 18% above the CRLB based on $X_1, X_2, X_3, ..., X_n$. This is the loosely stated version to the result given in Corollary 4. The glitch to this of course is that we are observing the upper 20% or more of the upper order statistics. We could not come nearly as close to the CRLB if we chose to censor out a *specific* 80% of the $X's$. (thus observing only 20%) For example if $n = 5$, and we choose to *leave* out four out of five (80%) of the data by selecting X_1, X_2, X_3, X_4, then the variance of the BLUE based on that one observation X_5 would be μ^2, which is five times greater than the CRLB based on all the $X's$. Practically speaking, a random collection of the $X's$ would be lost. It is not as if we would be choosing which ones to be lost as in times of transactions in a retail store. So in the real world, the lower 80% of the data would be lost, which was what we really had in mind when introducing the model in Chapter 1. If the upper 80% of the observations would have been lost, (which is Type-II right censoring) then we would be left with just

Y_5, (Assuming $\theta = 0$) and since the failure times are exponential, the BLUE would be Y_5 and so there again, the variance of the BLUE would be five times greater. Type-II left censoring yields a much less trivial result than Type-II right censoring does. Another reason why this result that becomes apparent in Section 2.3 is so intriguing is in one case, we would have lost a certain percentage of the data, whereas in the latter one explained, we would have been intentionally neglecting a certain percentage of the data. Yet for some counterintuitive reason if we pre-determined which of the $X's$ were to be left out, the variance of the BLUE is much higher-in the case of 80% neglected, five times higher as we exemplified. The contribution of the knowledge of the k^{th} to last order statistic Y_k has much to do with why it would be advantagous to have accidentally misplaced/lost records as opposed to just simply ignoring records on purpose. In Section 2.4, we come to the main result which is the limiting ratio of the variances with the variance when θ is known in the numerator and the other variance being in the denominator. This turns out to be a function of the proportion p of the upper order statistics recorded. It also turns out to be an increasing function which becomes very close to the function p for fairly low values of $p = 0.3$, and the function becomes closer to the line p as p increases. This actually is a consequence of Corollary 4. One would expect an increasing ratio function, since the higher percentage of order statistics we have available, the less important it should be to know θ when estimating the mean server time μ. For example, the BLUE for μ when θ is unknown as we will see does not account for the value of Y_k, but if the upper 90% of the order statistics is observed, the BLUE's variance would be approximately $\mu^2/0.9n$, which is only 11% higher than the variance of the sample mean for all the transaction times $X_1, X_2, ..., X_n$. ($Var\left(\overline{X}\right) = \frac{\mu^2}{n}$.) Of course, the variance of all the data is not the variance of the BLUE when θ is known, but it becomes quickly close to it as we see. One would think that the variance of the BLUE when θ is known should come close to the variance of the sample mean based on all the data if $p \approx 1$, but not nearly for such small values of p as it does, which as we will also see in Section 2.4 will explain that not only is the ratio function an increasing function in p, but also one that quickly becomes close to the function p as we mentioned before. We refer back to Corollary 4 in Section 2.4 to justify why the ratio function has its properties of being close to the function p. So also in Section 2.4, we give some probabilistic reasoning as to why the variance of the BLUE when θ is known approaches one as quickly as it does. In Section 2.3, we see numerically and graphically how the asymptotic normalized variance of the BLUE for μ when θ is known behaves as a function of p, and in Section 2.4 we give some explanation for why it behaves the way it does.

2.2 Derivation of BLUEs and Variances

In this section, we will derive the following formulas for the BLUEs and their corresponding variances which are given in Table 2.1.

Table 2.1: *Formulas for BLUEs of μ and their variances*

Formulas	θ Known	θ Unknown
BLUE	$\widehat{\mu}_B = \frac{\sum_{j=1}^{k-1} T_j + S_k Y_k}{R_k + k - 1} - \theta$ (See 2.3.)	$\widetilde{\mu}_B = \overline{T}$ (See 2.6.)
Variance	$Var\left(\widehat{\mu}_B\right) = \frac{\mu^2}{R_k + k - 1}$ (See 2.4.)	$Var\left(\widetilde{\mu}_B\right) = \frac{\mu^2}{k-1}$ (See 2.7.)

where $R_k = \frac{\left(\sum_{j=k}^{n} \frac{1}{j}\right)^2}{\sum_{j=k}^{n} \frac{1}{j^2}}$ and $S_k = \frac{\sum_{j=k}^{n} \frac{1}{j}}{\sum_{j=k}^{n} \frac{1}{j^2}}$.

Theorem 1. *Let $X_1, X_2, ..., X_n$ be independent random variables with $E\left(X_j\right) = \mu$ for all j, $1 \le j \le n$ and with $Var\left(X_j\right) = \sigma_j^2$ for $1 \le j \le n$. Let $c_j = \sigma_j^{-2}$. Then, the BLUE $\widehat{\mu}_B$ for μ and its variance are given by the following formulas.*

i.

$$\widehat{\mu}_B = \frac{\sum_{j=1}^{n} c_j X_j}{\sum_{j=1}^{n} c_j}$$

ii.

$$Var\left(\widehat{\mu}_B\right) = \frac{1}{\sum_{j=1}^{n} c_j}.$$

θ **known:**

Assume $\theta = 0$ without loss of generality. The problem of finding the BLUE based on $Y_1, Y_2, ..., Y_k$ means finding the coefficients $a_1, a_2, ..., a_k$ such that the linear combination

$$\sum_{j=1}^{k} a_j Y_j \ (2.1)$$

has minimal variance subject to the condition

$$E\left(\sum_{j=1}^{k} a_j Y_j\right) = \mu.$$

This is just the same as finding the coefficients $b_1, b_2, ..., b_k$ such that the variance of

$$\sum_{j=1}^{k-1} b_j T_j + b_k Y_k^* \ (2.2)$$

is minimized subject to being an unbiased estimator of μ, where

$$Y_k^* = \frac{Y_k}{\sum_{j=k}^{n} \frac{1}{j}}.$$

We can apply Theorem 1 to derive the formulas for the BLUE and its variance as they are given in Table 2.1. This is because $T_1, T_2, ..., T_{k-1}$ and Y_k^* are unbiased estimators for μ, and they are independent random variables. Let $T_1, T_2, ..., T_{k-1}$ and Y_k^* correspond to $X_1, X_2, ..., X_k$ respectiely as in Theorem 1. Then the respective coefficients are $c_j = \mu^{-2}$ for $1 \leq j \leq k-1$, and

$$c_k = \mu^{-2} R_k. \text{ (as } R_k \text{ is defined in Table 2.1)}$$

Therefore by Theorem 1,

$$\begin{aligned}
\widehat{\mu}_k &= \frac{\sum_{j=1}^{k-1} \mu^{-2} T_j + \mu^{-2} R_k \frac{Y_k}{\sum_{j=k}^{n} \frac{1}{j}}}{\mu^{-2}\left(k-1+R_k\right)} \\
&= \frac{\sum_{j=1}^{k-1} T_j + S_k Y_k}{R_k + k - 1}, \text{ (2.3, as } S_k \text{ is defined in Table 2.1)}
\end{aligned}$$

as claimed in Table 2.1. Also by Theorem 1,

$$Var\left(\widehat{\mu}_k\right) = \frac{1}{(k-1)\,\mu^{-2} + \mu^{-2}R_k}$$
$$= \frac{\mu^2}{R_k + k - 1}, \; (2.4)$$

also as claimed in Table 2.1.

θ **Unknown:**

Note that in this case

$$E\left(Y_k\right) = \mu \sum_{j=k}^{n} \frac{1}{j} + \theta,$$

so if we used Y_k^* as defined in the previous case, its expectation would depend on an unknown parameter θ, so it would not be unbiased unless $\theta = 0$, which we cannot assume like we did in the previous case because θ is unknown. Since the BLUE for μ must be of the form

$$\sum_{j=1}^{k-1} b_j T_j + b_k Y_k, \; (2.5)$$

we must have that $b_k = 0$, since the expectation of Y_k as we saw depends on θ. So we can apply Theorem 1 to $T_1, T_2, ..., T_{k-1}$. In doing so, we immediately arrive at

$$\widetilde{\mu}_B = \frac{\sum\limits_{j=1}^{k-1} \mu^{-2} T_j}{(k-1)\,\mu^{-2}}$$
$$= \overline{T}. \; (2.6)$$

Clearly without even using Theorem 1,

$$Var\left(\widetilde{\mu}_B\right) = \frac{\mu^2}{k-1}. \; (2.7)$$

We have now verified all four formulas as given in Table 2.1.

2.3 Asymptotic Limits of the Variances

We have derived in Section 2.2 two estimators for two different situations. One would be when the opening store time θ is known, while the other would be when the opening store time θ is unknown. The primary objectives in this section are as follows.

Objectives:

1. Gain insight on how $Var(\widehat{\mu}_B)$ and $Var(\widetilde{\mu}_B)$ behave for sufficiently large n when $k = [np]$ as a function of p.

2. Demonstrate numerically and graphically just how superior $\widehat{\mu}_B$ an estimator is compared to $\widetilde{\mu}_B$, or in other words in the sense of how much more the variance is reduced if we knew θ than opposed to the contrary case of not knowing θ.

We will be able to achieve Objectives 1 and 2 with the help of an important theorem involving the asymptotic variances of $\widehat{\mu}_B$ and $\widetilde{\mu}_B$. In order to prove the theorem, we first prove the following lemma.

Lemma 2. *Let R_k be defined as in Table 2.1. Then for all $p \in (0, 1)$,*

$$\frac{R_{[np]}}{np} \to \frac{\log^2 p}{1-p} \text{ as } n \to \infty.$$

Proof. By definition given in Table 2.1 and the fact that (See A1 for a proof on the following two limits.)

$$\frac{\sum\limits_{j=[np]}^{n} \frac{1}{j}}{\int\limits_{np}^{n} \frac{1}{x} dx} \to 1,$$

and

$$\frac{\sum\limits_{j=[np]}^{n} \frac{1}{j^2}}{\int\limits_{np}^{n} \frac{1}{x^2} dx} \to 1$$

as $n \to \infty$, we have that

$$\lim_{n\to\infty} \frac{R_{[np]}}{np} = \lim_{n\to\infty} \frac{\frac{\left(\sum\limits_{j=[np]}^{n} \frac{1}{j}\right)^2}{\sum\limits_{j=[np]}^{n} \frac{1}{j^2}}}{np}$$

$$= \lim_{n\to\infty} \frac{\left(\int\limits_{np}^{n} \frac{1}{x} dx\right)^2}{np \int\limits_{np}^{n} \frac{1}{x^2} dx}$$

$$= \lim_{n\to\infty} \frac{\log^2 p}{np\left(\frac{1}{np} - \frac{1}{n}\right)}$$

$$= \frac{\log^2 p}{1-p}$$

□

Theorem 3. *Let $Y_1, Y_2, ..., Y_k$ be the highest $k = [np]$ order statistics for the iid $X's$, where $X_i = W_i + \theta$, and $W_i \stackrel{d}{=} \exp(\mu)$. Then we have the following limiting results valid for all $p \in (0,1)$.*

i.

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\mu}_B\right) = \frac{1}{p\left(1 + \frac{\log^2 p}{1-p}\right)}$$

ii.

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widetilde{\mu}_B\right) = \frac{1}{p}$$

Proof. We refer the reader to Table 2.1 for the two formulas of the variances. To prove i., we use the results for variances given in that table and Lemma 3 to obtain

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\mu}_B\right) = \lim_{n\to\infty} \frac{n}{R_{[np]} + [np] - 1}$$

$$= \lim_{n\to\infty} \frac{1}{p\left(\frac{R_{[np]}}{np} + \frac{[np]-1}{np}\right)}$$

$$= \frac{1}{p\left(1 + \frac{\log^2 p}{1-p}\right)}.$$

Proving ii. is very straight forward. It's very clear that by Table 2.1,

$$\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widetilde{\mu}_B\right) = \lim_{n\to\infty}\frac{n}{[np]-1}$$
$$= \frac{1}{p}.$$

□

As a consequence of Theorem 3, we have the following aforementioned corollary. This gives an indication of how quickly $Var\left(\widehat{\mu}_B\right)$ approaches the CRLB based on all the $X's$.

Corollary 4. *Let $Y_1, Y_2, ..., Y_k$ be the highest $k = [np]$ order statistics for the iid $X's$, where $X_i = W_i + \theta$, and $W_i \overset{d}{=} \exp(\mu)$. Then for $p \geq 0.2$,*

$$\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widehat{\mu}_B\right) \leq 1.18.$$

Proof. When $p = 0.2$, we have by Theorem 3 that

$$\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widehat{\mu}_B\right) = \frac{1}{0.2\left(1+\frac{\log^2 0.2}{1-0.2}\right)}$$
$$= 1.1798$$

Since it can be readily verified that

$$\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}$$

is a decreasing function, the result follows. □

Note that by Corollary 4

$$\lim_{n\to\infty}\frac{Var\left(\widehat{\mu}_B\right)}{\frac{\mu^2}{n}} \leq 1.18 \text{ for } p \geq 0.2,$$

which essentially means that for large n when 20% or more of the upper order statistics are observed, the variance of $\widehat{\mu}_B$ is within 18% of the CRLB based on all the $X's$. We discuss this in greater detail in Section 2.4 and even more so in Section 4.6. So the use of Corollary 5 does not come in to our analysis until Section 2.4.

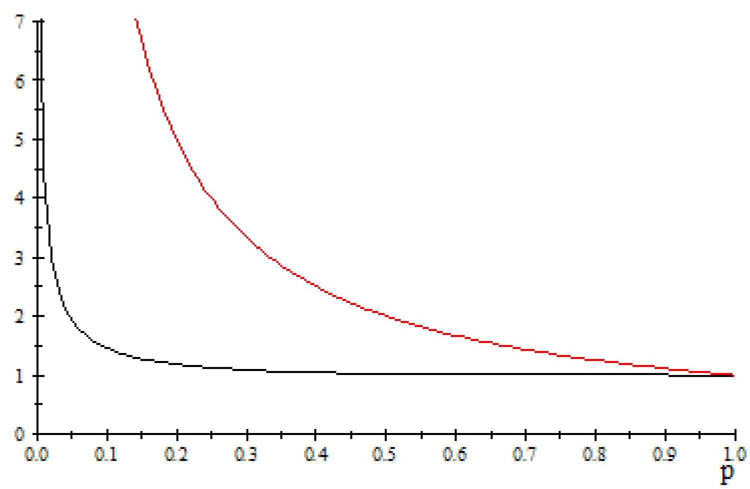
Figure 2.1 and Table 2.2 show the comparison between the functions

$$\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}$$

and (Later in this chapter, we will define $s(p) = p^{-1}\left(1+\frac{\log^2 p}{1-p}\right)^{-1}$.)

$$\frac{1}{p}.$$

Figure 2.1: $\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}$, $\frac{1}{p}$ $0 < p < 1$



We can see from Figure 2.1 that it is very clear that knowing θ has an advantage, although as p increases, the two plots come closer together, meaning knowing θ has less of an advantage as p increases. In other words, the higher the percentage of upper order statistics we have available, the less difference it makes if we know θ.

Table 2.2: *Approximate Variances:* $\mu = 1$, $n = 1000$

p	$Var\left(\widehat{\mu}_B\right)$	$\frac{1}{1000p\left(1+\frac{\log^2 p}{1-p}\right)}$	$Var\left(\widetilde{\mu}_B\right)$	$\frac{1}{p}$
0.1	1.4543×10^{-3}	1.4512×10^{-3}	1.0101×10^{-2}	10
0.2	1.1808×10^{-3}	1.1798×10^{-3}	5.0251×10^{-3}	5
0.5	1.0201×10^{-3}	1.0199×10^{-3}	2.004×10^{-3}	2
0.75	1.0017×10^{-3}	1.0017×10^{-3}	1.3351×10^{-3}	1.3333

We can see from Table 2.2 that numerically $Var\left(\widehat{\mu}_B\right)$ both by the exact formula and approximate formula that the CRLB based on $X_1, X_2, ..., X_n$ is nearly achieved even for $p = 0.2$ in accordance with Corollary 5 (relatively speaking, of course). $Var\left(\widetilde{\mu}_B\right)$ tends to 1/1000 much slower than $Var\left(\widehat{\mu}_B\right)$, but it is essentially only the CRLB based on $T_1, T_2, ..., T_{[np]-1}$ for each $p \in (0,1)$, so that is the reason. Since there is not much difference between $Var\left(\widetilde{\mu}_B\right)$ and $1/np$ for moderately large values of n and for any value of p, it is very obvious that the rate of convergence for $nVar\left(\widetilde{\mu}_B\right)/\mu^2$ to the function $1/p$ is so fast that there is no need to discuss that issue. However, the rate at which the function (of p) $nVar\left(\widehat{\mu}_B\right)/\mu^2$ converges to $1/p\left(1+\frac{\log^2 p}{1-p}\right)$ is questionable, so we conclude this section to numerically demonstrate that rate of convergence. Note that $nVar\left(\widehat{\mu}_B\right)/\mu^2$ for finite n and fixed p is a function of n and p, so we can write $nVar\left(\widehat{\mu}_B\right)/\mu^2 = s(n,p)$ and the limiting function we can denote as $s(p)$. We will see this function $s(p)$ in the next section as it has some probabilistic motivation, and Corollary 5 will come in to good use.

From Table 2.3, we can see that as p increases, the rate of convergence increases. For instance when $p = 0.9$, even when n is as small as 100, the limit is practically attained. At the other extreme when $p = 0.05$, $s(100, 0.05)$ (i.e., $n = 100$ and $p = 0.05$) is roughly greater than $s(0.05)$ by as much as 5%. As we stated in the next section that $s(p)$ has probabilistic meaning, the same is true for $s(n,p)$ for finite n. $s(n,p)$ since it represents finite n, is more practical than the limit $s(p)$ as we will briefly discuss in Section 2.4. Table 2.3 we refer back to in Section 2.4 for a specific reason involving the CRLB based on all the $X's$, so that table actually serves more than one purpose other than giving us an idea of the rate of convergence

Table 2.3: $s(n,p)$, $s(p)$

p	$s(10,p)$	$s(20,p)$	$s(100,p)$	$s(1000,p)$	$s(p)$
0.05	N/A	2.4663	2.0168	1.9243	1.9145
0.1	1.8065	1.6234	1.4833	1.4543	1.4512
0.2	1.2873	1.2312	1.1896	1.1808	1.1798
0.3	1.1348	1.1090	1.0900	1.0859	1.0855
0.4	1.0680	1.0543	1.0443	1.0422	1.0420
0.5	1.0343	1.0267	1.0212	1.0201	1.0199
0.6	1.0164	1.0123	1.0093	1.0087	1.0087
0.7	1.0071	1.0050	1.0035	1.0032	1.0032
0.8	1.0025	1.0016	1.0010	1.0008	1.0008
0.9	1.0006	1.0003	1.0001	1.0001	1.0001

of $s(n,p)$ to $s(p)$. Corollary 5 which we stated and proved in Section 2.3 is stressed in Section 2.4 and much later on in Section 4.6 as we mentioned before.

2.4 The Finite and Asymptotic Ratio of the Variances

Let $r_n(p)$ be defined as the ratio of the variance of the BLUE for μ when θ is known to the variance of the BLUE for μ when θ is unknown for finite n. In other words,

$$r_n(p) = \frac{Var(\widehat{\mu}_B)}{Var(\widetilde{\mu}_B)}. \quad (2.8)$$

One of the main results of discussion for this section is the following.

$$\lim_{n\to\infty} r_n(p) = \frac{1}{1+\frac{\log^2 p}{1-p}} = r(p). \quad (2.9)$$

We will see that the limiting ratio function in 2.9 is very close to the function p for values of $p \geq 0.5$ and even fairly close for $p \geq 0.2$. $r(p)$ becomes closer to the function p as p increases, and $r(p) > p$ for all $p \in (0,1)$. For $n \geq 10$, all the same is essentially true for the functions $r_n(p)$ defined in 2.8 as we will numerically see in this section. One would expect $r(p)$ to be an increasing function between zero and one, and according to 2.9, that is in fact the case. This is because we would think that knowing θ would reduce the variance for the BLUE of μ, and the higher proportion p of order statistics we have, the less important it would be to know θ. (therefore a higher ratio) To prove that the ratio functions defined in 2.8 and 2.9 are close to p and lie above p on the interval $(0,1)$ is not so clear from a pure mathematical perspective. We will discuss this later in the section and will find the $s(n,p) = nVar(\widehat{\mu}_B)/\mu^2$ (which when not tabulating its values we will denote as $s_n(p)$ from this point forward) and $s(p) = \lim_{n\to\infty} s_n(p)$ functions to be quite useful in intuitively justifying the two assertions of closeness to p and $r(p) > p$. Because of some minor discrepancy between the definitions of $r_n(p)$ and $s_n(p)$, it is quite rigorous to show that $r_n(p) > p$ for all n, so we negate that in this section. If $s_n(p)$ is defined a certain way, it is possible. We will find that using the $s(p)$ function, we can justify that $r(p) > p$ much more easily. First we prove 2.9 in the following theorem.

Theorem 5. *Let $Y_1, Y_2, ..., Y_k$ be the highest $k = [np]$ order statistics for the iid $X's$, where $X_1 = W_1 + \theta$, and $W_1 \stackrel{d}{=} \exp(\mu)$. Then we have the following limiting result valid for all $p \in (0,1)$.*

$$\lim_{n\to\infty} r_n(p) = \frac{1}{1 + \frac{\log^2 p}{1-p}} = r(p)$$

Proof. Based on Theorem 3, it is immediate that as $n \to \infty$,

$$\begin{aligned} r_n(p) &= \frac{nVar(\widehat{\mu}_B)}{nVar(\widetilde{\mu}_B)} \\ &\to \frac{\frac{\mu^2}{p\left(1+\frac{\log^2 p}{1-p}\right)}}{\frac{\mu^2}{p}} \\ &= \frac{1}{1 + \frac{\log^2 p}{1-p}} \\ &= r(p) \text{ for } 0 < p < 1. \end{aligned}$$

□

The following table gives values for $r_n(p)$ for the same values of n as in Table 2.3 and the limit function $r(p)$. To be consistent with Table 2.3, we denote in the table $r_n(p)$ as $r(n,p)$. By definition 2.8, 2.4 and 2.7,

$$r_n(p) = \frac{[np]-1}{R_{[np]}+[np]-1} \quad (2.10)$$

to justify the calculations that were done in Table 2.4.

Table 2.4: *Values for $r(p)$*

p	$r(10,p)$	$r(20,p)$	$r(100,p)$	$r(1000,p)$	$r(p)$
0.05	N/A	0.123 32	0.100 84	0.09622	0.09572
0.10	0.180 65	0.162 34	0.148 33	0.145 43	0.145 12
0.20	0.257 46	0.246 24	0.237 92	0.236 16	0.235 97
0.30	0.340 44	0.332 70	0.32700	0.325 77	0.325 65
0.40	0.427 20	0.421 72	0.417 72	0.416 88	0.416 79
0.50	0.517 15	0.513 35	0.510 60	0.510 05	0.509 97
0.60	0.609 84	0.607 38	0.605 58	0.605 22	0.605 20
0.70	0.704 97	0.703 50	0.702 45	0.702 24	0.702 22
0.80	0.80200	0.801 28	0.800 80	0.800 64	0.800 66
0.90	0.900 54	0.900 27	0.900 09	0.900 09	0.900 08

The values in Table 2.4 for finite n and in the limit are close to p for roughly $p \geq 0.2$. Also much like $s_n(p)$ in Table 2.3, the rate of convergence to $r(p)$ increases with p. Also, it is suggested in Table 2.4 that the functions $r_n(p)$ and $r(p)$ are increasing and also lie above the function p. There is a relationship between the values in Tables 2.3 and 2.4 as well. One will notice that for each n and p in the tables, $r_n(p) \approx ps_n(p)$, (Even for small n, there is hardly no difference though.) and $r(p) = ps(p)$. The approximation is very close, and the exact equality $r(p) = ps(p)$ directly follows by how we defined $s_n(p)$ as $nVar(\widehat{\mu}_B)/\mu^2$, since by Theorem 3 and

Theorem 5,

$$s(p) = \lim_{n\to\infty} s_n(p)$$
$$= \frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}$$
$$= \frac{r(p)}{p}. \quad (2.11)$$

Recognize the statistical meaning of $s_n(p)$, which we were alluding to before. By its definition,

$$s_n(p) = \frac{Var(\widehat{\mu}_B)}{Var(\overline{X})}. \quad (2.12)$$

Intuitively from 2.12, it seems clear that $s_n(p) > 1$, and by Theorem 3, we know that $s(p) > 1$, implying from 2.11 that $r(p) > p$. We give an even more meaningful, statistical argument which shows $r(p) > p$ in Section 4.6. That argument will also show how close $r(p)$ is to p for high enough values of p. For now, we informally address these questions by referring back to Table 2.3 in the previous section. Essentially, $s_n(p) = nVar(\widehat{\mu}_B)/\mu^2$ tends to a function $s(p)$ given in Theorem 3 that as we have seen graphically and numerically approaches from above one for fairly low values of p and greater. The function $s(p)$ is less than or equal to 1.18 for $p \geq 0.2$ by Corollary 4. This means the CRLB based on all the $X's$ is nearly reached (See Table 2.3.) with 80% of the earlier times of the transactions in the store lost. At $p = 0.3$, the variance of the BLUE when θ is known is only about 8.55% above the CRLB as seen in Table 2.3, and it just gets closer as p increases. This fact as we stated in Chapter 1 is one of the secondary results that we mistakenly discovered along the way in our analysis. Our real objective in this chapter was to study the advantage of knowing θ when estimating μ. Since the limiting ratio function $r(p)$ is increasing, the higher percentage of order statistics we have available, the less of an advantage there is by knowing θ. If we have only lost the lower 10% of the data, ($p = 0.9$) the ratio is roughly 0.9, which means that the variance when knowing θ is 90% of the variance without knowing θ. If we have lost 90% of the data, then knowing θ according to Table 2.4 means that the variance of the BLUE for μ is significantly lower-only 14.5% of the variance in the scenario when we do not know θ. We conclude this chapter with a plot of $r(p)$ versus p to demonstrate how $r(p)$ increases and its closeness to p on the interval $(0.4, 1)$. We also include a plot on the interval $(0, 0.5)$ to show how much of a difference there is between the two functions for smaller values of p.

Figure 2.2: $r(p), p, 0.4 < p < 1$

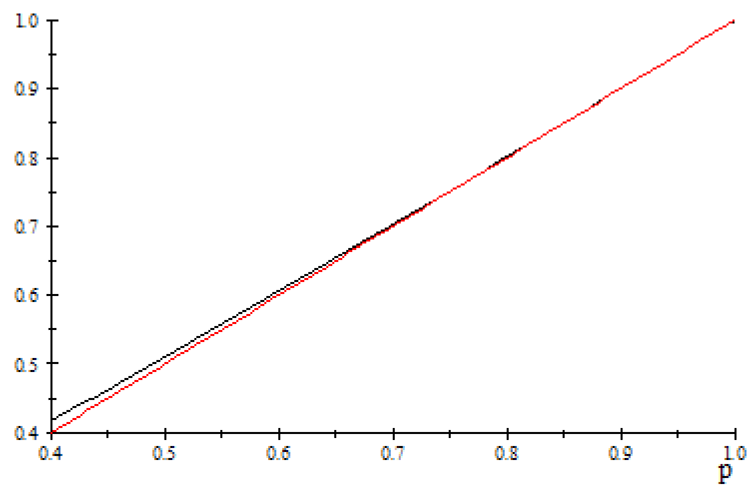
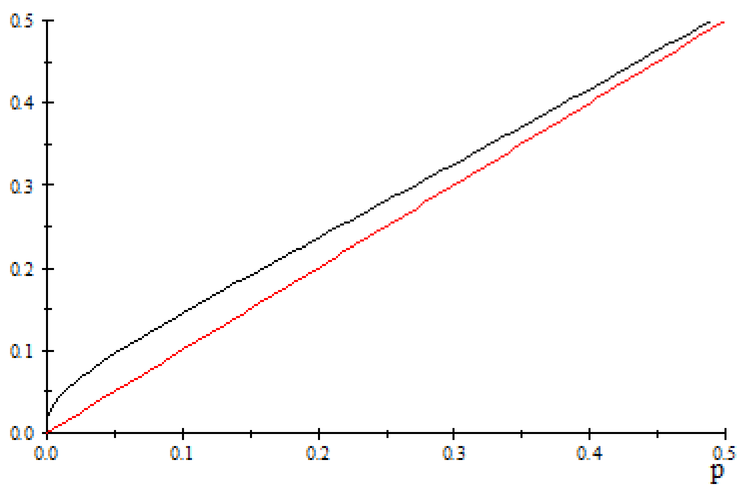


Figure 2.3: $r(p), p, 0 < p < 0.5$



Chapter 3

The Advantage of Knowing μ when Estimating θ and the Ratio Identity

3.1 Introduction:

Having access to only the last k transaction times $Y_1 > Y_2 > Y_3 > .... > Y_k$, one might also question at what time the store opened, which we call θ. Suppose we have all the transaction times so that $k = n$. If the mean server time μ is known, it seems natural that the only transaction time that is relevant in estimating θ among all n transaction times would be the first one Y_n, since the mean server time in that case would not have to be estimated. For the sake of argument, we are suppressing the notation of 'B' for BLUE and replacing the subscript of the estimators (including the one for μ) with the number of order statistics available. We are only doing that in this introductory section. Authors that we have cited have not considered the case where μ is known and θ is unknown perhaps because θ is quite often just a nuisance parameter. Thus, we have to make the following adjustment in order to construct a linear estimator that is a function of $T_1, T_2, T_3, ..., T_{n-1}, Y_n$. Unlike all the other cases, (with μ being known) there is no unbiased estimator for θ of the form $\sum_{j=1}^{n-1} a_j T_j + a_n Y_n$, so we have to consider the BLUE to be of the form $\sum_{j=1}^{n-1} a_j T_j + a_n (Y_n + c)$, where c is some *known* constant possibly depending on μ. Since $E(T_i) = \mu$, $\sum_{i=1}^{n-1} a_i = 0$, and because the variance of this estimator would be of the form

$$\mu^2 \sum_{j=1}^{n-1} a_j^2 + a_n^2 \frac{\mu^2}{n^2},$$

$a_1 = a_2 = ... = a_{n-1} = 0$, since otherwise it would not minimize the variance. The reason why we cannot have the BLUE to be of the form aY_n is because $E(Y_n) = E(Y_n - \theta + \theta) = \theta + \mu/n \neq \theta$. So the BLUE must be of the form $a(Y_n + c)$. Taking the expectation of this estimator and equating it to θ, we have

$$a\left(\theta + \frac{\mu}{n} + c\right) = \theta,$$

so in order for this to hold for all θ, we must have $a = 1$ and $c = -\mu/n$ which gives the BLUE as

$$\widehat{\theta}_n = Y_n - \frac{\mu}{n}.$$

For general $k < n$, the BLUE must be of the form $\sum_{j=1}^{k-1} a_j T_j + a_k (Y_k + c)$, and we remind the reader of this in a derivation which is analogous to the derivation we just did for the case of $k = n$. We point this out now for complete clarification, since in all the references cited, the BLUE has not been derived (In fact, not even the MLE has either.). If μ is unknown, we have to estimate μ. Since θ is also unknown, we cannot use Y_n to arrive at an estimator of μ. We saw this in Chapter 2. Y_n by itself gives no indication of the mean server time μ unless θ is known. So it is intuitive that we should exclude Y_n in estimating μ to arrive at

$$\widetilde{\theta}_n = Y_n - \frac{\widetilde{\mu}_n}{n},$$

where $\widetilde{\mu}_n = \overline{T}_{n-1}$ with

$$\overline{T}_{n-1} = \frac{\sum_{j=1}^{n-1} T_j}{n-1}.$$

Obviously, $Var\left(\widetilde{\theta}_n\right) > Var\left(\widehat{\theta}_n\right)$ by the two formulas we derived and when μ is unknown, we are faced with the problem of estimating μ first before we estimate θ. Now suppose that one of the transaction times is missing so that we only have $Y_1, Y_2, ..., Y_{n-1}$. In this situation, we are missing the time interval between the first customer's transaction time and the opening of the store and the time between the second transaction and the first transaction. In arriving at an estimator for θ in both cases of μ being known and unknown, it seems most logical to analogously take the least recorded time Y_{n-1} and note the following.

$$Y_{n-1} - \theta = Y_{n-1} - Y_n + Y_n - \theta$$

So we have written the difference between the second transaction time and the opening time as a sum of two time intervals, which consist of the time between the second and first transaction times and the first transaction time and the opening store time. So therefore when μ is known, we would expect our estimator to be

$$\widehat{\theta}_{n-1} = Y_{n-1} - \mu\left(\frac{1}{n-1} + \frac{1}{n}\right).$$

If μ is unknown, then in order to estimate θ, we have to first estimate μ using $Y_1, Y_2, ..., Y_{n-1}$, but here again, Y_{n-1} only represents a time on the clock, so it should be excluded in our estimation of μ. Clearly then we should have as our BLUE

$$\widetilde{\theta}_{n-1} = Y_{n-1} - \widetilde{\mu}_{n-1}\left(\frac{1}{n-1} + \frac{1}{n}\right),$$

where

$$\widetilde{\mu}_{n-1} = \frac{\sum_{j=1}^{n-2} T_j}{n-2}.$$

One question in this chapter is the difference in knowing μ when estimating θ with a certain proportion of upper order statistics p being available. The derivations can be carried out for any k in exactly the same manner as we just did with $k = n$ and $k = n - 1$. It is intuitively clear that the ratio of the variances as in Chapter 2 for the two estimators is less than one because if both parameters are unknown, we have to estimate μ before we arrive at an estimate of θ. That can easily be verified by glancing at the formulas for the BLUEs and taking their variances. In Section 3.2, we will derive the BLUEs and their variances. Also in Section 3.2, we will establish that

$$\frac{Var\left(\widehat{\theta}_B\right)}{Var\left(\widetilde{\theta}_B\right)} = \frac{Var\left(\widehat{\mu}_B\right)}{Var\left(\widetilde{\mu}_B\right)}.$$

This is an identity which we focus on for the bulk of this chapter, and we call it the ratio identity. (Sections 3.3 and 3.4) In Section 3.3, we give one obvious interpretation, which is the percentage of the variance in the estimator reduced in knowing each respective parameter is the same in both cases. For estimation of μ, we would like to know θ so that we can have the luxury of being able to use $\overline{T}$ and Y_k in our estimation of μ. We do not want to be forced to leave out Y_k and then be reduced to estimating μ with only $\overline{T}$. Similarly as shown earlier in this section, we do not want to be forced to estimate μ when trying to estimate θ. Clearly, the variance is lower if we only have to use Y_k to estimate θ in which case μ is known. In both estimation scenarios, how much it matters when having to resort to only using $\overline{T}$ or having to estimate μ with $\overline{T}$ depends on the ratio given in 3.1, or otherwise known in 2.8 as $r_n(p)$. The higher the value of $r_n(p)$, the greater the reliability of $\overline{T}$ for both estimating of μ and θ, and that is because of 3.1. In Section 3.4, we look at it from the estimation of μ perspective. We encourage the reader to imitate our methods in Section 3.4 for the estimation of θ. It seems far from trivial, so we avoid taking on that task. In Section 3.5, we demonstrate a trial run study of $n = 500$ demonstrating how $\overline{T}$ is weighted, and finally in Section 3.6, we derive the asymptotic normalized variance for $\widehat{\theta}_B$ and $\widetilde{\theta}_B$.

3.2 Derivation of BLUEs and Their Variances

We will derive the following formulas for the BLUEs given in Table 3.1 in this section.

Case 1: μ Known

As we explained in Section 3.1, the BLUE for θ in this case must be of the form

μ	known	unknown
BLUE	$\widehat{\theta}_B = Y_k - \mu \sum\limits_{j=k}^{n} \frac{1}{j}$ (See 3.5.)	$\widetilde{\theta}_B = Y_k - \widetilde{\mu}_B \sum\limits_{j=k}^{n} \frac{1}{j}$ (See 3.8.)
Variance	$Var\left(\widehat{\theta}_B\right) = \mu^2 \sum\limits_{j=k}^{n} \frac{1}{j^2}$ (See 3.6.)	$Var\left(\widetilde{\theta}_B\right) = \mu^2 \left(\sum\limits_{j=k}^{n} \frac{1}{j^2} + \frac{1}{k-1} \left(\sum\limits_{j=k}^{n} \frac{1}{j} \right)^2 \right)$ (See 3.9.)

Table 3.1: *Blues of θ and Their Variances*

$$\widehat{\theta}_B = \sum_{j=1}^{k-1} b_j T_j + b_k \left(Y_k + c\right) \ (3.1)$$

for constants $b_1, b_2, ..., b_k$. Now, $E\left(T_j\right) = \mu$ for $1 \leq j \leq k-1$, and

$$E\left(Y_k\right) = \mu \sum_{j=k}^{n} \frac{1}{j} + \theta.$$

Also observe by 3.1,

$$Var\left(\widehat{\theta}_B\right) = \sum_{j=1}^{k-1} b_j^2 \mu^2 + b_k^2 \mu^2 \sum_{j=k}^{n} \frac{1}{j^2}. \ (3.2)$$

If $b_k = 0$, it is impossible for the estimator in 3.1 to be unbiased. The estimator's expectation would then just be a constant $\sum\limits_{j=1}^{k-1} b_j \mu$ independent of θ. So $b_k \neq 0$. Further by taking expectations of both sides of 3.1, it is clear that $\sum\limits_{j=1}^{k-1} b_j = 0$. So to minimize 3.2, we would have to set $b_1 = b_2 = ... = b_{k-1} = 0$. Therefore in accordance with 3.1, we are now left with $b_k \left(Y_k + c\right)$ as the possible estimator for some constant b_k. Taking the expected value of that function of Y_k and setting it equal to θ, we have that

$$b_k \left(\sum_{j=k}^{n} \frac{\mu}{j} + \theta + c \right) = \theta. \ (3.3)$$

For 3.3 to follow *for all* θ, we must have that $b_k = 1$ and $c = -\mu \sum\limits_{j=k}^{n} \frac{1}{j}$. So

$$\widehat{\theta}_B = Y_k - \mu \sum_{j=k}^{n} \frac{1}{j}, \; (3.5)$$

and clearly it follows from this that

$$Var\left(\widehat{\theta}_B\right) = \mu^2 \sum_{j=k}^{n} \frac{1}{j^2}. \; (3.6)$$

Case 2: μ Unknown

The BLUE for θ here must be of the form

$$\widetilde{\theta}_B = a_1 \widetilde{\mu}_B + a_2 Y_k.$$

For the unbiased condition to be satisfied, we must have

$$a_1 \mu + a_2 \left(\mu \sum_{j=k}^{n} \frac{1}{j} + \theta \right) = \theta, \text{ or}$$

$$\mu \left(a_1 + a_2 \sum_{j=k}^{n} \frac{1}{j} \right) + a_2 \theta = \theta. \; (3.7)$$

3.7 must hold for all μ and θ, so we immediately have that $a_2 = 1$, and $a_1 = -\sum\limits_{j=k}^{n} \frac{1}{j}$. So in agreement with Table 3.1,

$$\widetilde{\theta}_B = Y_k - \widetilde{\mu}_B \sum_{j=k}^{n} \frac{1}{j}, \; (3.8)$$

and by independence,

$$Var\left(\widetilde{\theta}_B\right) = \mu^2 \sum_{j=k}^{n} \frac{1}{j^2} + \frac{\mu^2}{k-1} \left(\sum_{j=k}^{n} \frac{1}{j} \right)^2. \; (3.9)$$

The following theorem establishes the ratio identity result we mentioned in Section 3.1.

Theorem 6. *Let $X_i = W_i + \theta$, where $W_1, W_2, ..., W_n$ are iid $\exp(\mu)$ random variables. Let $Y_1 > Y_2 > .... > Y_k$ be the upper k order statistics. Then the following identity holds.*

$$\frac{Var(\widehat{\mu}_B)}{Var(\widetilde{\mu}_B)} = \frac{Var\left(\widehat{\theta}_B\right)}{Var\left(\widetilde{\theta}_B\right)}$$

Proof. By Table 2.1, we know that

$$\frac{Var(\widehat{\mu}_B)}{Var(\widetilde{\mu}_B)} = \frac{k-1}{k-1+R_k}.$$

Now observe that by Table 3.1,

$$\begin{aligned}\frac{Var\left(\widehat{\theta}_B\right)}{Var\left(\widetilde{\theta}_B\right)} &= \frac{\sum_{j=k}^{n} \frac{1}{j^2}}{\sum_{j=k}^{n} \frac{1}{j^2} + \frac{1}{k-1}\left(\sum_{j=k}^{n} \frac{1}{j}\right)^2} \\ &= \frac{k-1}{k-1+R_k} \\ &= \frac{Var(\widehat{\mu}_B)}{Var(\widetilde{\mu}_B)}\end{aligned}$$

□

We now try to give some intuition behind the identity derived in Theorem 6.

3.3 The Interpretation and Concept of the Ratio Variance Identity

Theorem 6 simply validates the ratio identity, which from now on we refer to as 3.10.

$$\frac{Var\left(\widehat{\theta}_B\right)}{Var\left(\widetilde{\theta}_B\right)} = \frac{Var(\widehat{\mu}_B)}{Var(\widetilde{\mu}_B)} \quad (3.10)$$

In Chapter 2, we desired to estimate the mean μ, while in this chapter we desire to estimate θ, but because of the identity in 3.10, we are compelled to consistently remind the reader of some of our results in Chapter 2. This we suggested in Section 3.1. What the result in 3.10 means is simply this. There's either one or two parameters unknown. If we were in the situation of θ being unknown, the *relative* reduction of the variance for the BLUE of θ in having the advantage when knowing μ is equal to the *relative* reduction of the variance of the BLUE for μ when knowing θ. For example, if $Var\left(\widetilde{\mu}_B\right) = 100$, and $Var\left(\widehat{\mu}_B\right) = 20$ and if $Var\left(\widetilde{\theta}_B\right) = 5$, then that would all mean $Var\left(\widehat{\theta}_B\right) = 1$. They both would reduce by 80%, and always by the same percentage in any case. There are three possible situations as we explained in Chapter 1. They can be interpreted as three different perceptions to the observer. We list them here along with their (possible) objective(s).

Cases for Estimation:

1. μ known, θ unknown; Objective: Estimate μ

2. μ unknown, θ unknown; Objective: Estimate μ and/or θ (In this thesis, exactly one of the two though.)

3. θ unknown, μ known; Objective: Estimate θ

We summarize the formulas for the BLUEs in the following comprehensive table of Chapters 2 and 3.

Table 3.2: *BLUEs in three different situations*

Parameter known?	μ known	μ unknown
θ known	N/A	$\widehat{\mu}_B = \frac{\sum_{j=1}^{k-1} T_j + S_k Y_k}{R_k + k - 1} - \theta$
θ unknown	$\widehat{\theta}_B = Y_k - \mu \sum_{j=k}^{n} \frac{1}{j}$	$\widetilde{\mu}_B = \overline{T}$, $\widetilde{\theta}_B = Y_k - \widetilde{\mu}_B \sum_{j=k}^{n} \frac{1}{j}$

Consider the following two cases of these last two chapters. In Chapter 2, μ was assumed to be unknown, while in Sections 3.1 and 3.2, θ was assumed to be unknown. $\widehat{\mu}_B$ uses all the order statistics $Y_1, Y_2, ..., Y_k$, while $\widetilde{\mu}_B$ in its formula does not take in to account the value of Y_k. So in Chapter 2 when estimating μ, what is excluded in the θ unknown case is Y_k. As a visual aid for Chapter 2, one can compare the two BLUEs for μ by glancing down the second column of Table 3.2 to see the difference. On the contrary when looking at the difference between the two estimators of θ, we can look across the second row of Table 3.2. In a sense, the exact opposite happens for the estimation of θ as compared to the estimation of μ. Looking at the second row and first column, only Y_k is used in the formula for $\widehat{\theta}_B$, and all the upper $k-1$ spacings are ignored. Looking at the second row and second column, $\widetilde{\theta}_B$ uses all the information in its formula. In other words, $\widehat{\mu}_B$ and $\widetilde{\theta}_B$ both use all the order statistics, while $\widehat{\theta}_B$ only uses Y_k, and $\widetilde{\mu}_B$ uses everything *except* Y_k. The natural logic in all four estimators we have discussed, so all four formulas are consistent with our intuition. In essence, the explanation for the equality in 3.10 is that to estimate μ, all the k upper order statistics are desired, but to estimate θ, we only want Y_k. When estimating μ, one does not want to be in the situation where θ is unknown because then we cannot use all the data-we then have to leave out Y_k. When estimating θ, of course we need Y_k, but we do not want to be forced in to having to come up with an estimator of μ. This is essentially the meaning of 3.10. We depict this concept in Figure 3.1.

Figure 3.1: *Venn Diagram for Estimation Scenarios*

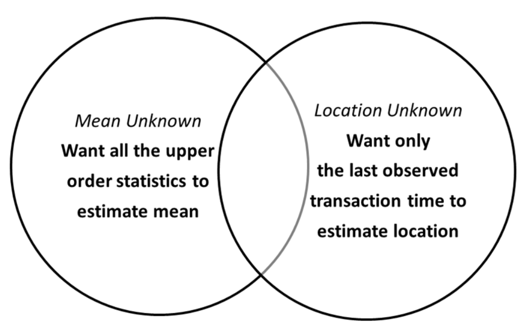


Figure 3.1 not only captures the three aforementioned situations above, but in the two non-intersection regions, in one region, it shows that we want all the data to estimate μ, while in the other for the estimation of θ, we are saying we only want one piece of the data which is Y_k. Whichever parameter is unknown, we do not want to be caught up in the intersection region where both parameters are unknown whether our goal is to estimate μ or θ. The intersection region goes against the will of both of the non-intersection regions. For the estimation of μ, we are forced to leave out Y_k in that region, while for the estimation of θ, we have to use all of the data as opposed just having to use Y_k. That we have just reiterated for the purpose of emphasis. We will find in Section 3.4 that (as defined in 2.8) $r_n(p) = Var(\widehat{\mu}_B)/Var(\widetilde{\mu}_B)$ and (as defined in 2.9) $r(p) = \lim_{n\to\infty} r_n(p)$ give an indication of how much it matters to be in the danger zone shown in Figure 3.1, which is the intersection region of the two circles. This is due to the ratio identity equality given in 3.10. For estimating μ, the greater the value of $Var(\widehat{\mu}_B)/Var(\widetilde{\mu}_B)$, the more accurate an estimator $\widetilde{\mu}_B$ is, so we can more freely go against the recommendation to use all the last k transaction times and exclude Y_k. At the same time by the identity in 3.10, the greater

$r_n(p) = Var\left(\widehat{\theta}_B\right)/Var\left(\widetilde{\theta}_B\right)$ is, the more accurate an estimator $\widetilde{\mu}_B$ is, so we can more comfortably go against our will and use $\widetilde{\mu}_B$ in addition to Y_k in the estimation of θ. The greater the value of $r_n(p)$, the less advantage there is in knowing μ and θ, and once again by 3.10, the relative advantage in knowing either of the parameters is exactly the same. (percentage reduction from not knowing either of the two parameters) Section 3.4 demonstrates the μ unknown situation shown in Figure 3.1.

3.4 The Reliance on Y_k on the Estimation of μ When θ is Known

In this section, we show how important it is to include Y_k in the estimation of μ when θ is known. This would be how important it is to be out of the intersection region given in Figure 3.1, and that depends on the proportion p of upper order statistics we have observed. Setting $k = [np]$, $Y_{[np]}$ is in high demand for small values of p

to estimate μ. As p increases, $\widetilde{\mu}_B = \overline{T}$ becomes a more improved estimator, (i.e., variance decreases) and we do not have to rely on $Y_{[np]}$ as much. When p is near one, $\widetilde{\mu}_B$ nearly attains the CRLB for all the $X's$, and thus it is intuitive that the weighting we place on $Y_{[np]}$ approaches zero as $p \to 1$ and $n \to \infty$. Referring back to Table 3.2 in the case where μ is unknown and θ is known we can rewrite $\widehat{\mu}_B$ as a linear combination of two unbiased estimators of μ given by

$$\widehat{\mu}_B = c_1(p)\,\overline{T} + c_2(p)\,Y^*_{[np]} \quad (3.11)$$

for some constants c_1, c_2 which depend on p, and where (assuming without loss of generality $\theta = 0$)

$$Y^*_{[np]} = \frac{Y_{[np]}}{\sum\limits_{j=[np]}^{n} \frac{1}{j}}.$$

Now since both estimators in the linear combination of 3.11 are unbiased, $c_2(p) = 1 - c_1(p)$. The question is what is this function $c_1(p)$? Interestingly enough it turns out that (See 2.8.)

$$c_1(p) = r_n(p) = \frac{Var(\widehat{\mu}_B)}{Var(\widetilde{\mu}_B)}. \quad (3.12)$$

From the numerical results given in Table 2.4, $r_n(p) \approx p$ for moderately large n and $p \geq 0.2$, which further suggests that

$$\widehat{\mu}_B \approx p\overline{T} + (1-p)\,Y^*_{[np]} \text{ for large } n \text{ and } p \geq 0.2. \quad (3.13)$$

The approximation in 3.13 from our observations in Table 2.4 becomes more accurate as p increases. $r_n(0.5) \approx 0.5$, and that approximation is extremely accurate for values of n as low as 20. So with the last 10 of 20 transaction times available, $\overline{T}$ and Y^*_{10} are just about equally weighted. If we had more than the last 10 transaction times, then $\overline{T}$ would be the more heavily weighted and dominant estimator. When we only have the last two transaction times, the accuracy of the approximation in 3.13 is poor, (upper 10%) but Y^*_2 dominates in the estimation of μ. Consider once again Figure 3.1 of the last section. If we have the last 18 out of 20 transaction times, then $p = 0.9$, and the advantage of being in the non-overlapping portion of the μ unknown circle (i.e., not θ) is not nearly as high as it would be if $p = 0.4$ and only eight of the last transaction times were available. That is why in the former case, the weighting on $\overline{T}$ would be 0.9, while in the latter aforementioned case, the weighting on $\overline{T}$ would only be $p = 0.4$. $r_n(p)$ also actually represents the proportion of the variance of $Y^*_{[np]}$ out of the variance of $\overline{T} + Y^*_{[np]}$, so with $p = 0.3$, $Y^*_{[0.3n]}$ only accounts for about 30% of the variance of $\overline{T} + Y^*_{[0.3n]}$. When $p = 0.9$, $Y^*_{[0.9n]}$ accounts for 90% of the variance of $\overline{T} + Y^*_{[0.9n]}$. All this is a consequence of the following identity given in 3.14, which we now prove using Theorem 1.

$$\widehat{\mu}_B = r_n(p)\,\overline{T} + (1 - r_n(p))\,Y^*_{[np]}. \quad (3.14)$$

Theorem 7. *Let $\widehat{\mu}_B$ be the BLUE based on $Y_1 > Y_2 > .... > Y_{[np]}$ where the Y's are the order statistics from $X_1, X_2, ..., X_n$ that are iid $\exp(\mu)$. Then $\widehat{\mu}_B$ can be written as the expression given in 3.14.*

Proof. We apply Theorem 1 to the estimator $\overline{T}$ in order to compute its coefficient $c_1(p)$ given in 3.12.

$$\begin{aligned} c_1(p) &= \frac{\dfrac{\mu^2 \sum_{j=[np]}^{n} \frac{1}{j^2}}{\left(\sum_{j=[np]}^{n} \frac{1}{j}\right)^2}}{\dfrac{\mu^2 \sum_{j=[np]}^{n} \frac{1}{j^2}}{\left(\sum_{j=[np]}^{n} \frac{1}{j}\right)^2} + \frac{\mu^2}{[np]-1}} \\ &= \frac{[np]-1}{[np]-1+R_{[np]}} \\ &= r_n(p). \end{aligned}$$

□

This establishes the validity of 3.14.

By 2.9,

$$\lim_{n\to\infty} c_1(p) = \frac{1}{1+\frac{\log^2 p}{1-p}}$$
$$= r(p). \quad (3.15)$$

and therefore

$$\lim_{n\to\infty} c_2(p) = 1 - r(p).$$

Recall in 2.9 we defined the limiting ratio function as

$$r(p) = \lim_{n\to\infty} \frac{Var(\widehat{\mu}_B)}{Var(\widetilde{\mu}_B)}.$$

Now we would like to compare the variances of the BLUE for μ and its two weighted unbiased estimators of μ, $\overline{T}$ and $Y^*_{[np]}$. The most convenient way to do this is to study them asymptotically just as we did in Chapter 2 using Theorem 3. This way we can have all three continuous functions on the same graph and see it visually. Then we can confirm the graphical results numerically in a table as we did throughout Chapter 2. By Theorem 3, we already know the asymptotic normalized variances of $\widehat{\mu}_B$ and $\overline{T}$. Recall from that theorem,

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var(\widehat{\mu}_B) = \frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}, \quad (3.16)$$

and

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\overline{T}\right) = \frac{1}{p}. \quad (3.17)$$

Now we show that

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(Y^*_{[np]}\right) = \frac{1-p}{p\log^2 p} \quad (3.18)$$

in the following theorem.

Theorem 8. *Let $Y_1 > Y_2 > .... > Y_{[np]}$ be the upper $[np]$ order statistics from an iid sequence of* $\exp(\mu)$ *random variables $X_1, X_2, ..., X_n$. Then,*

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(Y^*_{[np]}\right) = \frac{1-p}{p\log^2 p}.$$

Proof.

$$Var\left(Y^*_{[np]}\right) = \frac{\mu^2 \sum\limits_{j=[np]}^{n} \frac{1}{j^2}}{\left(\sum\limits_{j=[np]}^{n} \frac{1}{j}\right)^2},$$

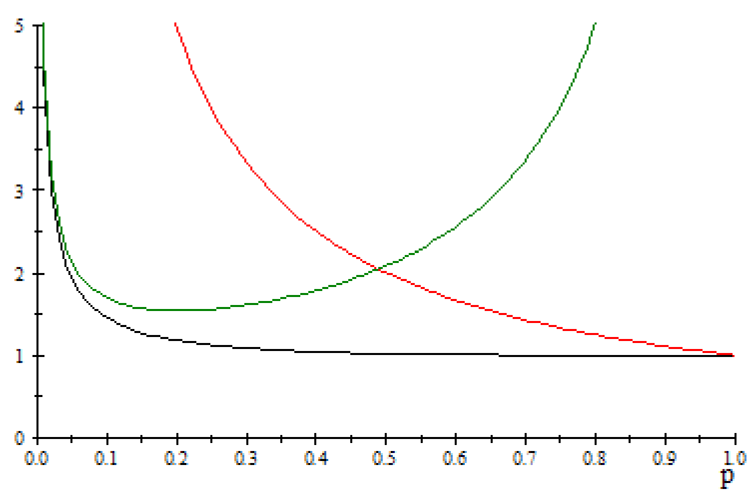
so

$$\begin{aligned} &\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(Y^*_{[np]}\right) \\ &= \lim_{n\to\infty} \frac{n}{\mu^2} \frac{\mu^2\left(\frac{1}{np} - \frac{1}{n}\right)}{\log^2 p} \\ &= \frac{1-p}{p\log^2 p} \end{aligned}$$

□

Figure 3.2 shows the plots of the three functions given in 3.16, 3.17 and 3.18.

Figure 3.2: $\lim\limits_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\mu}_B\right)$, (lowest) $\lim\limits_{n\to\infty} \frac{n}{\mu^2} Var\left(Y^*_{[np]}\right)$, (middle until $p = 0.5$) $\frac{1}{p}$



Based on Figure 3.2, (which also includes the plots of the other two functions from Figure 2.1) we can see that $Y^*_{[np]}$ is never a decent estimator for μ. Its variance seems to minimize at around $p = 0.2$, (See [1] for the CRLB function of $Y^*_{[np]}$.) and even there it appears to be about 60% (since the asymptotic normalized variance seems to be around 1.6) above the CRLB based on all the $X's$, and for $p > 0.2$, that asymptotic normalized variance for $Y^*_{[np]}$ tends to infinity as $p \to 1$. When $p \approx 0$, $Y^*_{[np]}$ is about as good an estimator as $\widehat{\mu}_B$ itself, although that is not saying very much because both asymptotic normalized variances tend to infinity as $p \to 0$, meaning they are both significantly higher than μ^2/n, practically speaking for finite n. When $p \approx 0.5$, the weighting on $\overline{T}$ and $Y^*_{[np]}$ is about equal as suggested by Theorem 7 and the fact that

$$\begin{aligned} r\left(0.5\right) &= \frac{1}{1+\frac{\log^2(0.5)}{0.5}} \\ &= 0.509\,97. \end{aligned}$$

Actually as alluded to earlier in this section, that means that

$$\lim_{n\to\infty} \frac{Var\left(Y^*_{[0.5n]}\right)}{Var\left(Y^*_{[0.5n]}\right) + Var\left(\overline{T}\right)} = 0.50997,$$

so for finite n, roughly 51% of the variance of $Var\left(Y^*_{[0.5n]}\right) + Var\left(\overline{T}\right)$ is accounted for by $Var\left(Y^*_{[0.5n]}\right)$. As shown in Figure 3.2, the variances of the two unbiased estimators are roughly equal near 0.5. (where the two plots intersect) If θ is known to be zero, we note that by 3.16 and 3.17 with $p = 0.5$

$$\begin{aligned} \lim_{n\to\infty} \frac{Var\left(\overline{T}\right)}{Var\left(\widehat{\mu}_B\right)} &= \frac{\frac{1}{p}}{\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}} \\ &= 1 + \frac{\log^2 0.5}{0.5} \\ &= 1.960\,9, \end{aligned}$$

which means from the perception of the observer who does not know θ, the variance of his BLUE is about 96% greater than the one who does know $\theta = 0$. If $p = 0.95$,

$$\lim_{n\to\infty} \frac{Var\left(Y^*_{[0.95n]}\right)}{Var\left(Y^*_{[0.95n]}\right) + Var\left(\overline{T}\right)} = \frac{1}{1 + \frac{\log^2 0.95}{0.05}} = 0.950\,01,$$

and

$$\lim_{n\to\infty} \frac{Var\left(\overline{T}\right)}{Var\left(\widehat{\mu}_B\right)} = 1 + \frac{\log^2 0.95}{0.05} = 1.0526,$$

and so for the observer not knowing θ, the variance of his estimator of $\overline{T}$ when 95% of the upper order statistics are observed is only about 5.26% greater than the observer who knows θ. For $p > 0.5$, $\overline{T}$ is the better estimator, and it finally hits the CRLB based on all the $X's$ when $p = 1$. As $p \to 1$, we know that $r(p) \to 1$ so using the fact that

$$\lim_{n\to\infty} \frac{Var\left(Y^*_{[np]}\right)}{Var\left(Y^*_{[np]}\right) + Var\left(\overline{T}\right)} = r(p),$$

$Var\left(Y^*_{[np]}\right)$ accounts for nearly 100% of $Var\left(Y^*_{[np]}\right) + Var\left(\overline{T}\right)$ when $p \approx 1$. The most critical question is this. If in Figure 3.1, we are in the non-overlapping μ unknown region for estimating μ, how much better off are we using the BLUE $\widehat{\mu}_B$ than just using $\overline{T}$ as our estimator of μ? That depends on the value of p. The greater the value of p, the greater the weighting on $\overline{T}$ which is $r_n(p)$, and the less of an advantage there is to the observer who knows θ. We should point out that as poor as an estimator $Y^*_{[np]}$ becomes as $p \to 1$, it always lowers the variance of the BLUE. Going back to the identity 3.10

$$r_n(p) = \frac{Var\left(\widehat{\theta}_B\right)}{Var\left(\widetilde{\theta}_B\right)} = \frac{Var\left(\widehat{\mu}_B\right)}{Var\left(\widetilde{\mu}_B\right)},$$

a similar idea applies to the estimation of θ. Based on 3.10 and Table 3.2, we have that

$$r_n(p) = \frac{Var\left(Y_{[np]}\right)}{Var\left(Y_{[np]}\right) + Var\left(\overline{T}\right)\left(\sum_{j=[np]}^{n} \frac{1}{j}\right)^2},$$

so here again the existence of $\overline{T}$ in $\widetilde{\theta}_B$ and lack thereof in $\widehat{\theta}_B$ is what increases the variance for the estimator of θ when not knowing μ. If we did not know μ with θ being unknown, the only possible estimator to use would be $\overline{T}$. If we knew μ and were estimating θ, the relative advantage (in terms of percentage) would depend on the variance of $\overline{T}$ just like it would if we knew θ and were estimating μ, and the point is that those relative advantages are equal. (i.e., ratio identity exemplified from a different perspective) So there is a sense of symmetry which can be seen from Figure 3.1. The ratio $r_n(p)$ can be looked at as a reliability measure of $\overline{T}$ from estimation scenarios of μ and θ and relatively speaking, an equal reliability measure in from both directions from the non-overlapping regions in Figure 3.1 to the intersection region. Now, it is important to understand that even though

$$\widehat{\mu}_B = r_n(p)\,\overline{T} + (1 - r_n(p))\,Y^*_{[np]},$$

if you take the limit of both sides, you have almost sure convergence of both sides to μ. (and thus also in distribution to a point mass μ) To obtain something nontrivial such as convergence to a standard normal random variable, you would have to rewrite 3.11 as

$$\frac{\widehat{\mu}_B - \mu}{\frac{\mu}{\sqrt{n}}} = r_n(p)\,\frac{\left(\overline{T} - \mu\right)}{\frac{\mu}{\sqrt{n}}} + (1 - r_n(p))\,\frac{\left(Y^*_{[np]} - \mu\right)}{\frac{\mu}{\sqrt{n}}}. \quad (3.19)$$

It can be shown that 3.17 converges to (See [6].)

$$Z(p) = r(p)\,Z_1(p) + (1 - r(p))\,Z_2(p),$$

where for each p, $Z_1(p) \stackrel{d}{=} N\left(0, \frac{1}{p}\right)$, $Z_2(p) \stackrel{d}{=} N\left(0, \frac{1-p}{p\log^2 p}\right)$and $Z_1(p)$ and $Z_2(p)$ are independent. However μ is unknown in this problem, so 3.19 would only be applicable for hypothesis testing or for confidence intervals for μ. This is well outside the scope of this thesis and open for discussion. We conclude this section with numerical values for the normalized variances of $\widehat{\mu}_B$, $\overline{T}$ and $Y^*_{[np]}$ for $n = 100$ and asymptotically. We also give values for $r_{100}(p)$ and $r(p)$ to show the weighting on $\overline{T}$.

Table 3.3: *Normalized Variances for Estimators of* μ; $n = 100$

p	$\frac{100}{\mu^2}Var(\widehat{\mu}_B)$	$\frac{100}{\mu^2}Var(\overline{T})$	$\frac{100}{\mu^2}Var\left(Y^*_{[100p]}\right)$	$r_{100}(p)$
0.1	1.4833	11.111	1.7119	0.1335
0.25	1.1293	4.1667	1.5491	0.27103
0.5	1.0212	2.0408	2.0441	0.50040
0.75	1.0019	1.3514	3.8749	0.74143

Table 3.4: *Asymptotic Normalized Variances for Estimators of* μ

p	$\lim_{n\to\infty}\frac{n}{\mu^2}Var(\widehat{\mu}_B)$	$\lim_{n\to\infty}\frac{n}{\mu^2}Var(\overline{T})$	$\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(Y^*_{[100p]}\right)$	$r(p)$
0.1	1.4512	10	1.6975	0.14512
0.25	1.1228	4	1.561	0.28071
0.5	1.0199	2	2.0814	0.50997
0.75	1.0017	1.3333	4.0277	0.75129

3.5 A Simulation Study for Estimation of μ and θ

As we saw in the last section, the coefficient $r_n(p)$ is a measure of the accuracy of $\overline{T}$ for the estimation of μ and for the estimation of θ in the case when both parameters are unknown. (in the intersection region of Figure 3.1) That follows by the ratio identity given in 3.10. When estimating μ, one observer knows θ while the other does not know θ. The lower the variance of $\overline{T}$, or equivalently the higher percentage of order statistics we have available, the less important it is to know θ. Similarly, the lower the variance of $\overline{T}$, the less of an advantage there is in knowing μ when estimating θ. The critical point is that the marginal advantages are equal because both observers who are at a disadvantage in not knowing θ or not knowing μ in their respective estimation scenarios both rely on $\overline{T}$. We simulated one random sample of $n = 500$ exponential random variables with $\mu = 1$ and $\theta = -5$. We selected certain values of p and calculated the statistics shown below in the following table.

Table 3.5: *Simulation,* $n = 500$, $\mu = 1$, $\theta = -5$

p	$\overline{T}$	$Y^*_{[np]}$	$\widehat{\mu}_B$	$\widetilde{\theta}_B$	$\widehat{\theta}_B$
0.1	0.9143	0.9388	0.935 3	−4.943	−5.141
0.4	0.8948	1.022	0.969 23	−4.882	−4.979
0.5	0.9242	1.016	0.969 36	−4.935	−4.988
0.7	0.9415	1.040	0.971 02	−4.964	−4.985
0.9	0.9328	1.358	0.976 13	−4.954	−4.961
0.98	0.9628	1.534	0.974 22	−4.987	−4.988

We should keep in mind that this is only one simulation run and referring to Figure 3.2, both the variances of $\overline{T}$ and $Y^*_{[np]}$ are high for small values of p, and not to mention the variance of $Y^*_{[np]}$ is also very high near $p = 1$. The values as p increases for $\widehat{\mu}_B$ are far more consistent from $p = 0.4$ and greater quite possibly because the variance for the BLUE of μ is essentially a constant of $1/500$ for values of $p \geq 0.2$. Generally the same is true for the θ estimators, and as we will see in Section 3.6, they are better estimators than those of μ for moderately large enough values of p.

3.6 On the Variances of Estimators for θ

Although the ratio identity holds, the BLUE estimators of μ and θ have very different variance formulas. In particular, take the case where $k = n$. Then by Table 3.1,

$$Var\left(\widehat{\theta}_B\right) = \frac{\mu^2}{n^2}, \quad (3.20)$$

and

$$\begin{aligned} Var\left(\widetilde{\theta}_B\right) &= \frac{\mu^2}{n^2} + \frac{\mu^2}{(n-1)}\frac{1}{n^2} \\ &= \frac{\mu^2}{n^2}\left(1 + \frac{1}{n-1}\right) \\ &= \frac{\mu^2}{n^2}\frac{n}{n-1} \\ &= \frac{\mu^2}{n\,(n-1)}. \quad (3.21) \end{aligned}$$

The variances in 3.20 and 3.21 tend to zero at the rate of $1/n^2$, while the variances for both μ BLUE estimators which are when $k = n$

$$\begin{aligned} Var\left(\widehat{\mu}_B\right) &= Var\left(\overline{X}\right) \\ &= \frac{\mu^2}{n}, \quad (3.22) \end{aligned}$$

and

$$Var\left(\widetilde{\mu}_B\right) = \frac{\mu^2}{n-1}. \quad (3.23)$$

One can easily verify that the identity ratio holds here as it should because we proved it in Theorem 6 for general k. The μ estimators are bounded in variance below by the CRLB of μ^2/n, while the θ estimators, the Fisher Information does not exist, and no such bound exists other than the MLE variance bounds which we will see in Chapter 4.

Remark 9. *When both parameters are unknown, the CRLB based on all n observations still makes sense when estimating μ, since after all the random variables are still shifted by a constant even though it is unknown. That is, whether θ is known or unknown, $Var\left(\overline{X}\right) = \mu^2/n$, and that is the CRLB. Now for the other unknown parameter θ, the CRLB does not exist because we cannot define its Fisher Information. The same is true when θ is unknown and μ is known.*

In light of 3.20 and 3.21, we have the following theorem on the asymptotic normalized variances for the two BLUE θ estimators. This is for general percentiles with $k = [np]$.

Theorem 10. *Let $Y_1 > Y_2 > .... > Y_{[np]}$ be the upper $[np]$ order statistics from an iid sequence of* $\exp(\mu)$ *random variables $X_1, X_2, ..., X_n$. Then,*

i.

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\theta}_B\right) = \frac{1-p}{p}$$

ii.

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widetilde{\theta}_B\right) = \frac{1-p}{p}\left(1 + \frac{\log^2 p}{1-p}\right)$$

Proof. i.

When $k = [np]$, we have from Table 3.1 that

$$\begin{aligned}
\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\theta}_B\right) &= \lim_{n\to\infty} \frac{n}{\mu^2}\mu^2 \sum_{j=[np]}^{n} \frac{1}{j^2} \\
&= \lim_{n\to\infty} n \int_{[np]}^{n} \frac{1}{x^2} dx \\
&= \lim_{n\to\infty} n \left(\frac{1}{np} - \frac{1}{n}\right) \\
&= \frac{1-p}{p}.
\end{aligned}$$

ii.

Also from Table 3.1, we have that

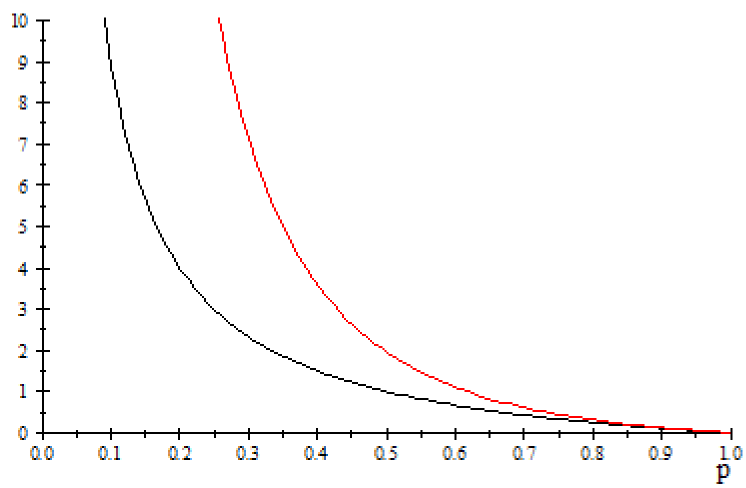
$$\begin{aligned}
&\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widetilde{\theta}_B\right) \\
&= \lim_{n\to\infty} n \left(\sum_{j=[np]}^{n} \frac{1}{j^2} + \frac{1}{[np]-1}\left(\sum_{j=[np]}^{n} \frac{1}{j}\right)^2\right) \\
&= \frac{1-p}{p} + \lim_{n\to\infty} \frac{n}{np} \log^2 p \\
&= \frac{1-p}{p}\left(1 + \frac{\log^2 p}{1-p}\right).
\end{aligned}$$

□

Remark 11. *It should be noted that both the functions in Theorem 10 have asymptotic normalized variances that approach zero as $p \to 1$, indicating that they must cross the line $y = 1$ for some value of p. The asymptotic variance of $\widehat{\theta}_B$ crosses one at $p = 0.5$, making it a better estimator than $\widehat{\mu}_B$ for $p \geq 0.5$. The other asymptotic variance intersects $y = 1$ for some value of $p > 0.5$. The point is they are both better estimators if a high enough percentage of the upper order statistics is retained.*

We conclude this chapter with a plot of the two variance functions for the BLUEs of θ.

Figure 3.3: $\frac{1-p}{p}$, $\frac{1-p}{p}\left(1 + \frac{\log^2 p}{1-p}\right)$



Chapter 4

Maximum Likelihood Estimation and the Minimum Mean Square Error Conjecture

4.1 Introduction

Up until this point in the thesis, we have only been dealing with the BLUEs for each of the three estimation scenarios. We have made several observations in regards to the advantages of knowing a certain parameter when estimating the other parameter. (e.g., The marginal advantage of knowing θ for the estimation of μ) One of the primary objectives in this chapter is to compare the four BLUE estimators' Mean Square Errors (MSE) with the MSEs of their corresponding Maximum Likelihood Estimators. (MLE) In this Type-II censoring scheme that we have used, it is not possible to derive an explicit formula for the MLE of μ when θ is known, but surprisingly, when both μ and θ are unknown, it is possible to find the MLEs for both μ and θ. (See [4].) As we mentioned in Chapter 2, in all of the references cited in the bibliography section, no one has considered estimating θ with μ being known, so the derivation of that MLE is original, yet easily follows from the case when both parameters are unknown. So unlike in Chapter 2 where we derived the BLUEs and their variances, we do not derive the MLEs and their MSEs and biases squared in this chapter, but rather reserve those derivations for *A2* in the appendix. It is also worth mentioning that our derivations of the BLUEs in Chapter 3 is original too for the very same reason as the derivations of the BLUEs from Chapter 2. For the case of where there is no explicit solution of the MLE, we use in *A2* the method of Approximate Maximum Likelihood Estimation (AMLE) introduced by Balakrishnan and Cohen. (See [4].) In Chapter 2, we used a transformation on the vector $(Y_1, Y_2, ..., Y_k)$ to a vector with k independent random variables, and in that regard, our work there is distinguishable from any of the authors who have derived the BLUEs in all of the estimation scenarios. In Section 4.2, we simply state the formulas for the MLEs and their respective MSEs and bias squares and note the similarity of the formulas of the MLEs to those of the BLUEs. In Section 4.3, we come to one of the main results in this chapter which is a theorem stating that asymptotically, the MSEs of the MLEs and the BLUEs are equal. The critical point of Section 4.3 is that the BLUEs may in fact be asymptotically the minimal variance unbiased estimators based on the data recorded, and this is because typically, the MLEs are asymptotically the minimum variance unbiased estimators. In Section 4.4, we show that the bias squares for the MLEs are increasingly small compared to their variances as n increases. For all four MLEs, the variances tend to zero at the rate of $1/n$, while the bias squares tend to zero at the rate of $1/n^2$. Since the bias squares when multiplied by n^2/μ^2 actually have limits which are functions of p for each of the MLEs, it would at first glance seem conceivable that the BLUEs may possibly for large n have lower MSEs than those of the MLEs. In fact it is clear that $MSE\left(\widehat{\theta}_B\right) < MSE\left(\widehat{\theta}_M\right)$ for all n, since it is always the case that $Var\left(\widehat{\theta}_B\right) = Var\left(\widehat{\theta}_M\right)$. However, the difference of the variance of the BLUE minus the variance of the MLE when multiplied by n^2/μ^2 we prove has a limiting function in three out of the four cases, (particularly the one we mentioned in the previous sentence) and as for the other case, there is probably a limiting function there as well for which we leave as an open question. The main premise of Section 4.4 is that the difference in the MSEs for the BLUEs and their respective MLEs tends to a function in at least three out of four cases, while in Section 4.3, we show that the ratio of the MSEs for the BLUEs and their respective MLEs tends to one. So Section 4.4 shows Section 4.3 to be a bit misleading, yet also shows that the BLUEs in conceivably three out of the four cases can have a lower MSE than their respective MLEs. One of those cases depends on value of p, in particular. In Section 4.5, we show that not only asymptotically $\widehat{\mu}_B$ and $\widehat{\mu}_M$ are unbiased, but they also asymptotically achieve the CRLB based on our data $\left(Y_1, Y_2, ..., Y_{[np]}\right)$ so that they are asymptotically the minimum variance unbiased estimators. For the purpose of clarity, we give a definition to the meaning of asymptotically attaining the CRLB based on the data recorded also in Section 4.5. The purpose of Section 4.5 is to claim minimum variance among all asymptotically unbiased estimators. Unfortunately, we can only prove this in one out of the three estimation scenarios, since in the other two, the support depends on an unknown parameter θ. The other three BLUE and three MLE estimators we conjecture in Chapter 5 to be the asymptotically minimum variance unbiased estimators. Finally in Section 4.6, we regress back to Section 2.3 and Theorem 3 and revisit the function $s(p) = r(p)/p$. We use a CRLB argument to support the reasons for why $s(p) > 1$ and also why $s(p)$ tends to one as quickly as it does. At $p = 0.2$, $s(p) \approx 1.18$, which means that when only 20% of the upper order statistics are observed, the variances of $\widehat{\mu}_B$ and $\widehat{\mu}_M$ are only 18% above the CRLB based on all the $X's$. Section 4.6 is really a continuation of Section 4.5 because we continue to use the Fisher Information for the entire data set and that for $Y_{[np]}$ by itself.

4.2 Formulas for MLEs, Variances and Bias Squares

In Tables 4.1 and 4.2, the formulas for the MLEs, their variances and their bias squares are given. One can determine the MSE for each of the MLEs simply by adding the variance to the bias squared. In Tables 4.1 and 4.2,

$$q_k = \frac{k}{k + R_k^*},$$

$$Y_k^{**} = \frac{Y_k - \theta}{\log\left(\frac{n}{k}\right)}, \text{ (with } \theta \text{ being known)}$$

and

$$R_k^* = \frac{\left(\int_k^n \frac{1}{x} dx\right)^2}{\int_k^n \frac{1}{x^2} dx}$$
$$= \frac{nk}{n-k} \log^2\left(\frac{n}{k}\right).$$

Table 4.1: *MLE and Variance*

MLE	*Expression*	*Variance*
$\widehat{\mu}_{AM}$	$q_k \frac{\sum_{j=1}^{k-1} T_j}{k} + (1-q_k) Y_k^{**}$	$\mu^2 q_k^2 \left(\frac{k-1}{k^2} + \frac{n^2}{(n-k)^2} \log^2\left(\frac{n}{k}\right) \sum_{j=k}^{n} \frac{1}{j^2} \right)$
$\widetilde{\mu}_M$	$\frac{\sum_{j=1}^{k-1} T_j}{k}$	$\mu^2 \frac{k-1}{k^2}$
$\widetilde{\theta}_M$	$Y_k + \widetilde{\mu}_M \log\left(\frac{k}{n}\right)$	$\mu^2 \left(\sum_{j=k}^{n} \frac{1}{j^2} + \frac{(k-1)\log^2\left(\frac{k}{n}\right)}{k^2} \right)$
$\widehat{\theta}_M$	$Y_k + \mu \log\left(\frac{k}{n}\right)$	$\mu^2 \sum_{j=k}^{n} \frac{1}{j^2}$

By Theorem 7 and 3.14, one can see that when $\theta = 0$, $\widehat{\mu}_{AM}$ and $\widehat{\mu}_B$ are both weighted linear combinations of two independent estimators involving $\sum_{j=1}^{k-1} T_j$ and Y_k. The difference is that $\widehat{\mu}_M$ is a weighted linear combination of two biased estimators which are $\widetilde{\mu}_M$ and Y_k^{**}. For both the estimators of θ, $\sum_{j=k}^{n} \frac{1}{j}$ is replaced by $\log\left(\frac{n}{k}\right)$, and obviously $\widetilde{\mu}_M$ appears nearly identical to $\widetilde{\mu}_B$. Also notice the resemblance of q_k to r_k. Adopting the same notation we did in Chapter 2, (See 2.8.) when $k = [np]$, we change q_k to $q_n(p)$, and by Lemma 3 and *A1*, it is clear that $q_n(p) \to r(p)$ as $n \to \infty$. We use that result in several theorems/corollaries throughout this chapter.

Table 4.2: *Bias Squares of MLEs*

MLE	*Bias Squared*
$\widehat{\mu}_{AM}$	$\mu^2\left(q_k\left(\frac{(k-1)}{k}-\frac{\sum_{j=k}^{n}\frac{1}{j}}{\log\left(\frac{n}{k}\right)}\right)+\left(\frac{\sum_{j=k}^{n}\frac{1}{j}}{\log\left(\frac{n}{k}\right)}-1\right)\right)^2$
$\widetilde{\mu}_M$	$\frac{\mu^2}{k^2}$
$\widetilde{\theta}_M$	$\frac{\mu^2\left((k-1)\log\left(\frac{k}{n}\right)+k\sum_{j=k}^{n}\frac{1}{j}\right)^2}{k^2}$
$\widehat{\theta}_M$	$\mu^2\left(\sum_{j=k}^{n}\frac{1}{j}+\log\left(\frac{k}{n}\right)\right)^2$

4.3 The Asymptotic MSEs of the MLEs

It is often the case that the MLEs are optimal in the sense that asymptotically, they have minimal variance. They are almost always asymptotically unbiased. Surprisingly, the BLUEs have asymptotically the same MSEs as the MLEs as we prove in Theorem 12. This suggests that perhaps the BLUEs are asymptotically the minimum variance unbiased estimators. Theorem 12 is as follows, and we immediately prove the result using the formulas given in Tables 4.1 and 4.2..

Theorem 12. *Let $X_1, X_2, ..., X_n$ be iid random variables such that $X_i = W_i + \theta$ for $1 \le i \le n$ with $W_1 \overset{d}{=} \exp(\mu)$ and $Y_1 > Y_2 > ... > Y_{[np]}$ be the upper $[np]$ order statistics of the $X's$. Then for all $p \in (0,1)$,*

i.

$$\lim_{n\to\infty}\frac{n}{\mu^2}MSE\left(\widehat{\mu}_{AM}\right)=\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widehat{\mu}_{AM}\right)=\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}.$$

ii.

$$\lim_{n\to\infty}\frac{n}{\mu^2}MSE\left(\widetilde{\mu}_{M}\right)=\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widetilde{\mu}_{M}\right)=\frac{1}{p}$$

iii.

$$\lim_{n\to\infty}\frac{n}{\mu^2}MSE\left(\widetilde{\theta}_{M}\right)=\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widetilde{\theta}_{M}\right)=\frac{1-p}{p}\left(1+\frac{\log^2 p}{1-p}\right)$$

iv.

$$\lim_{n\to\infty}\frac{n}{\mu^2}MSE\left(\widehat{\theta}_M\right)=\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widehat{\theta}_M\right)=\frac{1-p}{p}$$

Proof. i.

Note that by Table 4.1,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widehat{\mu}_{AM}\right)\\
&=\lim_{n\to\infty}\frac{n}{\mu^2}\mu^2q_n^2\left(p\right)\left(\frac{[np]-1}{[np]^2}+\frac{n^2}{\left(n-[np]\right)^2}\log^2\left(\frac{n}{[np]}\right)\sum_{j=[np]}^{n}\frac{1}{j^2}\right)\\
&=r^2\left(p\right)\lim_{n\to\infty}\left(\frac{n}{np}+\frac{n}{\left(1-p\right)^2}\log^2 p\left(\frac{1}{np}-\frac{1}{n}\right)\right)\\
&=r^2\left(p\right)\left(\frac{1}{p}+\frac{\log^2 p}{p\left(1-p\right)}\right)\\
&=\frac{1}{p}r^2\left(p\right)r^{-1}\left(p\right)\\
&=\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}.
\end{aligned}$$

Now by Table 4.2,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n}{\mu^2}Bias^2\left(\widehat{\mu}_{AM}\right)\\
&=\lim_{n\to\infty}\frac{n}{\mu^2}\mu^2\left(q_n\left(p\right)\left(\frac{\left([np]-1\right)}{[np]}-\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}\right)+\left(\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}-1\right)\right)^2\\
&=\lim_{n\to\infty}\left(q_n\left(p\right)\sqrt{n}\left(\frac{\left([np]-1\right)}{[np]}-\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}\right)+\sqrt{n}\left(\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}-1\right)\right)^2
\end{aligned}$$

Note that

$$\lim_{n\to\infty}\sqrt{n}\left(\frac{([np]-1)}{[np]}-1\right)$$
$$=-\lim_{n\to\infty}\sqrt{n}\frac{1}{np}$$
$$=0.$$

So therefore by A1,

$$\lim_{n\to\infty}\frac{n}{\mu^2}Bias^2\left(\widehat{\mu}_{AM}\right)$$
$$=\left(r\left(p\right)\lim_{n\to\infty}\sqrt{n}\left(\frac{([np]-1)}{[np]}-\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}\right)+\lim_{n\to\infty}\sqrt{n}\left(\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}-1\right)\right)^2$$
$$=0.$$

ii. Also by Table 4.1,

$$\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widetilde{\mu}_M\right)$$
$$=\lim_{n\to\infty}\frac{n}{\mu^2}\frac{([np]-1)\,\mu^2}{[np]^2}$$
$$=\frac{1}{p}.$$

By Table 4.2,

$$\lim_{n\to\infty}\frac{n}{\mu^2}Bias^2\left(\widetilde{\mu}_M\right)$$
$$=\lim_{n\to\infty}\frac{n}{\mu^2}\frac{\mu^2}{[np]^2}$$
$$=0.$$

iii.

By Table 4.1,

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widetilde{\theta}_M\right) \\
&= \lim_{n\to\infty} \frac{n}{\mu^2}\mu^2 \left(\sum_{j=[np]}^{n} \frac{1}{j^2} + \frac{[np]-1}{[np]^2} \log^2\left(\frac{[np]}{n}\right)\right) \\
&= \lim_{n\to\infty} n \left(\frac{1}{np} - \frac{1}{n} + \frac{[np]-1}{[np]^2} \log^2\left(\frac{[np]}{n}\right)\right) \\
&= \frac{1-p}{p} + \frac{1}{p}\log^2 p \\
&= \frac{1-p}{p}\left(1 + \frac{\log^2 p}{1-p}\right).
\end{aligned}
$$

Now by Table 4.2,

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{n}{\mu^2} Bias^2\left(\widetilde{\theta}_M\right) \\
&= \lim_{n\to\infty} \left(\sqrt{n}\frac{([np]-1)}{[np]} \log\left(\frac{[np]}{n}\right) + \sqrt{n} \sum_{j=[np]}^{n} \frac{1}{j}\right)^2 \\
&= \lim_{n\to\infty} \left(\sqrt{n} \left(\sum_{j=[np]}^{n} \frac{1}{j} - \log\left(\frac{n}{[np]}\right)\right) - \frac{\sqrt{n}}{[np]} \log\left(\frac{[np]}{n}\right)\right)^2 \\
&= 0.
\end{aligned}
$$

again using A1.

iv. By Table 4.1,

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\theta}_M\right) \\
&= \lim_{n\to\infty} \frac{n}{\mu^2}\mu^2 \left(\frac{1}{np} - \frac{1}{n}\right) \\
&= \frac{1-p}{p},
\end{aligned}
$$

and by Table 4.2,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n}{\mu^2}Bias^2\left(\widehat{\theta}_M\right)\\
&=\lim_{n\to\infty}\frac{n}{\mu^2}\mu^2\left(\log\left(\frac{[np]}{n}\right)+\sum_{j=[np]}^{n}\frac{1}{j}\right)^2\\
&=0,
\end{aligned}$$

once again by A1. □

4.4 The Asymptotic Normalized Difference Between the MSEs

Although by Theorem 12, it is clear that the ratio of the MSEs for the MLEs to their corresponding BLUEs approaches one, that does not mean that the MSEs of the MLEs are lower or higher than the MSEs of the BLUEs asymptotically. By multiplying the variance and the bias squared of each MLE by n/μ^2, we saw in Theorem 12 that the asymptotic normalized variance tends to a function of p, while the corresponding bias squared portion tends to zero. It turns out that in at least three out of four cases, the difference in the MSEs of the BLUEs and the MLEs does have a limiting function, only to obtain it, we need multiply by n^2/μ^2. The bias squares of the MLEs are a multiple of $1/n$ times smaller than their variances, and each bias squared when multiplied by n^2/μ^2 does tend to a limiting function of p. This is established in Theorem 14. Theorem 15 is a bit more restricted in that we try to obtain a limiting function by multiplying the difference of the variances of the MLEs and their respective BLUEs by n^2/μ^2, but for the case of θ being known and μ being unknown, we were unable to determine the existence of a limiting function simply because of the algebraic complexity involved. Thus, we only conclude with three functions for the difference of the MSEs. When we say difference of the MSEs, we mean the MSE of the MLE minus the MSE of the BLUE. In two out of four cases, the BLUE actually has the lower asymptotic MSE. In one of the cases, the difference of the MSEs depends on p. When we say "equal asymptotically," we really mean the ratio of the MSEs tends to one. In this section, we are putting a more powerful microscope on the comparison by multiplying the difference of the MSEs by n^2/μ^2. By multiplying each MSE by just n/μ^2 we cannot see any difference, which is what is hidden in Theorem 12. Our objective in this section is to find the limiting functions of the following as $n \to \infty$.

i. $\dfrac{n^2}{\mu^2}\left(MSE\left(\widehat{\mu}_{AM}\right) - MSE\left(\widehat{\mu}_B\right)\right)$

ii. $\dfrac{n^2}{\mu^2}\left(MSE\left(\widetilde{\mu}_M\right) - MSE\left(\widetilde{\mu}_B\right)\right)$

iii. $\dfrac{n^2}{\mu^2}\left(MSE\left(\widetilde{\theta}_M\right) - MSE\left(\widetilde{\theta}_B\right)\right)$

iv. $\dfrac{n^2}{\mu^2}\left(MSE\left(\widehat{\theta}_M\right) - MSE\left(\widehat{\theta}_B\right)\right)$

Part i. is the one we only were able to obtain the first part of given in Theorem 14, assuming of course there is a limit to the second part of the difference. (when decomposed as a difference of two different terms) Theorem 14 captures the limiting functions for the bias squared terms, while Theorem 15 takes care of the rest of the expression in parts ii. through iv..

Lemma 13. *For each $p \in (0,1)$,*

$$\lim_{n\to\infty} n\left(\sum_{j=[np]}^{n} \frac{1}{j} - \log\left(\frac{n}{[np]}\right)\right) = \frac{1+p}{2p}.$$

Theorem 14. *Let $X_1, X_2, ..., X_n$ be iid random variables such that $X_i = W_i + \theta$ for $1 \le i \le n$ with $W_1 \overset{d}{=} \exp(\mu)$ and $Y_1 > Y_2 > ... > Y_{[np]}$ be the upper $[np]$ order statistics of the $X's$. Then for all $p \in (0,1)$,* *i.*

$$\lim_{n\to\infty}\frac{n^2}{\mu^2}Bias^2\left(\widehat{\mu}_{AM}\right)=\left(\frac{1+p}{2p}\left(\frac{\log p}{1-p+\log^2 p}\right)+\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}\right)^2$$

ii.

$$\lim_{n\to\infty}\frac{n^2}{\mu^2}Bias^2\left(\widetilde{\mu}_{M}\right)=\frac{1}{p^2}$$

iii.

$$\lim_{n\to\infty}\frac{n^2}{\mu^2}Bias^2\left(\widetilde{\theta}_{M}\right)=\left(\frac{1+p}{2p}-\frac{\log p}{p}\right)^2$$

iv.

$$\lim_{n\to\infty}\frac{n^2}{\mu^2}Bias^2\left(\widehat{\theta}_{M}\right)=\frac{(1+p)^2}{4p^2}$$

Proof. i. From Table 4.2 and Lemma 13,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n^2}{\mu^2}Bias^2\left(\widehat{\mu}_{AM}\right)\\
&=\lim_{n\to\infty}\frac{n^2}{\mu^2}\mu^2\left(q_n(p)\left(\frac{([np]-1)}{[np]}-\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}\right)+\left(\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}-1\right)\right)^2\\
&=\lim_{n\to\infty}\left(nq_n(p)\left(1-\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}\right)+nq_n(p)\left(\frac{[np]-1}{[np]}-1\right)+n\left(\frac{\sum_{j=[np]}^{n}\frac{1}{j}}{\log\left(\frac{n}{[np]}\right)}-1\right)\right)^2\\
&=\left(r(p)\frac{1+p}{2p\log p}-\frac{r(p)}{p}-\frac{1+p}{2p\log p}\right)^2\\
&=\left(-\frac{1+p}{2p\log p}\left(\frac{\log^2 p}{1-p+\log^2 p}\right)-\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}\right)^2\\
&=\left(\frac{1+p}{2p}\left(\frac{\log p}{1-p+\log^2 p}\right)+\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}\right)^2
\end{aligned}$$

ii. By Table 4.2,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n^2}{\mu^2}Bias^2\left(\widetilde{\mu}_M\right)\\
&=\lim_{n\to\infty}\frac{n^2}{\mu^2}\frac{\mu^2}{[np]^2}\\
&=\frac{1}{p^2}.
\end{aligned}$$

iii. By Table 4.2 and Lemma 13,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n^2}{\mu^2}Bias^2\left(\widetilde{\theta}_M\right)\\
&=\lim_{n\to\infty}\left(n\frac{([np]-1)}{[np]}\log\left(\frac{[np]}{n}\right)+n\sum_{j=[np]}^{n}\frac{1}{j}\right)^2\\
&=\lim_{n\to\infty}\left(n\left(\sum_{j=[np]}^{n}\frac{1}{j}-\log\left(\frac{n}{[np]}\right)\right)-\frac{n}{[np]}\log\left(\frac{[np]}{n}\right)\right)^2\\
&=\left(\frac{1+p}{2p}-\frac{\log p}{p}\right)^2.
\end{aligned}$$

iv. Also by Table 4.2 and Lemma 13,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{n^2}{\mu^2}Bias^2\left(\widehat{\theta}_M\right)\\
&=\lim_{n\to\infty}\frac{n^2}{\mu^2}\mu^2\left(\log\left(\frac{[np]}{n}\right)+\sum_{j=[np]}^{n}\frac{1}{j}\right)^2\\
&=\frac{(1+p)^2}{4p^2}.
\end{aligned}$$

□

As a consequence of Theorem 14, we have the following, which is our main result for this section.

Theorem 15. *Let $X_1, X_2, ..., X_n$ be iid random variables such that $X_i = W_i + \theta$ for $1 \leq i \leq n$ with $W_1 \overset{d}{=} \exp(\mu)$ and $Y_1 > Y_2 > ... > Y_{[np]}$ be the upper $[np]$ order statistics of the $X's$. Then for all $p \in (0,1)$,*

i.

$$\lim_{n\to\infty} \frac{n^2}{\mu^2} \left(MSE\left(\widetilde{\mu}_M\right) - MSE\left(\widetilde{\mu}_B\right) \right) = -\frac{1}{p^2}.$$

ii.

$$\lim_{n\to\infty} \frac{n^2}{\mu^2} \left(MSE\left(\widetilde{\theta}_M\right) - MSE\left(\widetilde{\theta}_B\right) \right) = -\frac{\log^2 p}{p^2} + \frac{(1+p)^2}{4p^2}$$

iii.

$$\lim_{n\to\infty} \frac{n^2}{\mu^2} \left(MSE\left(\widehat{\theta}_M\right) - MSE\left(\widehat{\theta}_B\right) \right) = \frac{(1+p)^2}{4p^2}$$

Proof. i. From Tables 4.1 and 2.1,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{n^2}{\mu^2} \left(Var\left(\widetilde{\mu}_B\right) - Var\left(\widetilde{\mu}_M\right) \right) \\
&= \lim_{n\to\infty} n^2 \left(\frac{1}{[np]-1} - \frac{[np]-1}{[np]^2} \right) \\
&= \lim_{n\to\infty} n^2 \left(\frac{[np]^2 - [np]^2 + 2[np] - 1}{n^3 p^3} \right) \\
&= \frac{2}{p^2}.
\end{aligned}$$

Combining this with part ii. of Theorem 14, we arrive at the result.

ii. From Tables 4.1 and 2.1, it follows from Lemma 13 that

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{n^2}{\mu^2} \left(Var\left(\widetilde{\theta}_B\right) - Var\left(\widetilde{\theta}_M\right)\right) \\
&= \lim_{n\to\infty} n^2 \left(\frac{1}{[np]-1} \left(\sum_{j=[np]}^{n} \frac{1}{j} \right)^2 - \frac{[np]-1}{[np]^2} \log^2\left(\frac{n}{[np]}\right) \right) \\
&= \lim_{n\to\infty} n^2 \left(\frac{1}{[np]-1} \left(\left(\sum_{j=[np]}^{n} \frac{1}{j} \right)^2 - \frac{([np]-1)^2}{[np]^2} \log^2\left(\frac{n}{[np]}\right) \right) \right) \\
&= \lim_{n\to\infty} n^2 \left(\frac{1}{[np]-1} \left(\left(\sum_{j=[np]}^{n} \frac{1}{j} \right)^2 - \log^2\left(\frac{n}{[np]}\right) \right) + \frac{2}{n^2p^2} \log^2(p) - \frac{1}{n^3p^3} \log^2 p \right) \\
&= \lim_{n\to\infty} \frac{n}{p} \left(\left(\sum_{j=[np]}^{n} \frac{1}{j} \right)^2 - \log^2\left(\frac{n}{[np]}\right) \right) + 2\frac{\log^2 p}{p^2} \\
&= \frac{1}{p} \lim_{n\to\infty} n \left(\sum_{j=[np]}^{n} \frac{1}{j} - \log\left(\frac{n}{[np]}\right) \right) \left(\sum_{j=[np]}^{n} \frac{1}{j} + \log\left(\frac{n}{[np]}\right) \right) + 2\frac{\log^2 p}{p^2} \\
&= -\frac{2\log p}{p}\frac{1+p}{2p} + 2\frac{\log^2 p}{p^2} \\
&= -\frac{(1+p)\log p}{p^2} + 2\frac{\log^2 p}{p^2}.
\end{aligned}
$$

Combining this with part iii. of Theorem 14, we have

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{n^2}{\mu^2} \left(MSE\left(\widetilde{\theta}_M\right) - MSE\left(\widetilde{\theta}_B\right)\right) \\
&= \left(\frac{1+p}{2p} - \frac{\log p}{p}\right)^2 + \frac{(1+p)\log p}{p^2} - 2\frac{\log^2 p}{p^2} \\
&= \frac{(1+p)^2}{4p^2} - \frac{\log^2 p}{p^2}
\end{aligned}
$$

iii. This is immediate from Theorem 14.

□

Tables 4.3 through Tables 4.10 numerically give the asymptotic normalized difference in the MSEs of the MSE of the MLE minus the MSE of the BLUE. Based on the values given in the third column of Tables 4.3 and 4.4, it seems apparent that convergence of $n^2/\mu^2\left(Var\left(\widehat{\mu}_B\right)-Var\left(\widehat{\mu}_{AM}\right)\right)$ does occur. (to some function of p) Whatever function it is, it would be very small compared to $n^2/\mu^2\left(Bias\left(\widehat{\mu}_B\right)-Bias\left(\widehat{\mu}_{AM}\right)\right)$ for moderately large values of p. In fact, it would be practically zero for $p \geq 0.5$. This would make the MSE of $\widehat{\mu}_B$ lower than $\widehat{\mu}_{AM}$, (as can be seen in the fourth column of Tables 4.3 and 4.4) but the difference becomes less as p increases.

Tables 4.3-4.10: Asymptotic Difference of MSEs

Table 4.3: *μ unknown, θ known,* $n = 500$

p	$\frac{500^2}{\mu^2}Bias^2\left(\widehat{\mu}_{AM}\right)$	$\frac{500^2}{\mu^2}\left(Var\left(\widehat{\mu}_B\right)-Var\left(\widehat{\mu}_{AM}\right)\right)$	$\frac{500^2}{\mu^2}\left(MSE\left(\widehat{\mu}_{AM}\right)-MSE\left(\widehat{\mu}_B\right)\right)$
0.1	0.356 35	-1.75	2.106 4
0.25	3.0839×10^{-2}	-0.4	0.430 84
0.50	0.001 67	-0.1	0.101 67
0.75	4.9156×10^{-5}	0	4.9156×10^{-5}
0.90	9.2556×10^{-7}	0	9.2556×10^{-7}

Table 4.4: *μ unknown, θ known,* $n = 1000$ *

p	$\frac{1000^2}{\mu^2}B^2\left(\widehat{\mu}_{AM}\right)$	$\frac{1000^2}{\mu^2}\left(V\left(\widehat{\mu}_B\right)-V\left(\widehat{\mu}_{AM}\right)\right)$	$\frac{1000^2}{\mu^2}\left(MSE\left(\widehat{\mu}_{AM}\right)-MSE\left(\widehat{\mu}_B\right)\right)$
0.1	0.352 70	-1.7	2.052 7
0.25	3.0612×10^{-2}	-0.4	0.430 61
0.50	1.6556×10^{-3}	0	1.6556×10^{-3}
0.75	4.8374×10^{-5}	-0.1	0.100 05
0.90	8.9023×10^{-7}	0	8.9023×10^{-7}

*-For $p = 0.75$, we most likely have a round-off error in the third column.

Table 4.5: μ *unknown*, θ *unknown*, $n = 500$

p	$\frac{500^2}{\mu^2} Bias^2 (\widetilde{\mu}_M)$	$\frac{500^2}{\mu^2} (Var (\widetilde{\mu}_B) - Var (\widetilde{\mu}_M))$	$\frac{500^2}{\mu^2} (MSE (\widetilde{\mu}_M) - MSE (\widetilde{\mu}_B))$
0.1	100.0	202.0	-102
0.25	16.0	32.15	-16.15
0.50	4.0	8.0	-4
0.75	1.7778	3.55	-1.7722
0.90	1.2346	2.5	-1.2654

Table 4.6: μ *unknown*, θ *unknown*, $n = 1000$

p	$\frac{1000^2}{\mu^2} Bias^2 (\widetilde{\mu}_M)$	$\frac{1000^2}{\mu^2} (Var (\widetilde{\mu}_B) - Var (\widetilde{\mu}_M))$	$\frac{1000^2}{\mu^2} (MSE (\widetilde{\mu}_M) - MSE (\widetilde{\mu}_B))$
0.1	100.0	201.0	-101
0.25	16.0	32.1	-16.1
0.50	4.0	8.0	-4
0.75	1.7778	3.5	-1.7222
0.90	1.2346	2.4	-1.1654

We conclude this section by graphing the limiting function given in part ii. of Theorem 15 for the purpose of illustration. $MSE \left(\widetilde{\theta}_B\right) > MSE \left(\widetilde{\theta}_M\right)$ for roughly $p < 0.5$ based on Figure 4.1, and the inequality reverses for $p > 0.5$. The other two functions are monotonically decreasing or increasing and strictly positive or negative on the entire interval $(0, 1)$. Once again, we leave it as a conjecture in Chapter 5 to show the existence and determine the limiting function (if the limit exists) of

$$\frac{n^2}{\mu^2} (Var (\widehat{\mu}_B) - Var (\widehat{\mu}_M))$$

as $n \to \infty$.

4.5 The Asymptotic Optimality of $\widehat{\mu}_B$

With the help of Theorem 12 and Balakrishnan (See [1].), we can claim that $\widehat{\mu}_B$ is optimal in the sense that

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var (\widehat{\mu}_B) = \frac{r (p)}{p}$$

Table 4.7: μ *unknown,* θ *unknown,* $n = 500$

p	$\frac{500^2}{\mu^2} Bias^2 \left(\widetilde{\theta}_M\right)$	$\frac{500^2}{\mu^2} \left(Var\left(\widetilde{\theta}_B\right) - Var\left(\widetilde{\theta}_M\right)\right)$	$\frac{500^2}{\mu^2} \left(MSE\left(\widetilde{\theta}_M\right) - MSE\left(\widetilde{\theta}_B\right)\right)$
0.1	814.67	1331.0	−516.33
0.25	64.765	89.5	−24.735
0.50	8.3336	8.0	0.3336
0.75	2.4037	1.195	1.2087
0.90	1.3751	0.28	1.0951

Table 4.8: μ *unknown,* θ *unknown,* $n = 1000$

p	$\frac{1000^2}{\mu^2} Bias^2 \left(\widetilde{\theta}_M\right)$	$\frac{1000^2}{\mu^2} \left(Var\left(\widetilde{\theta}_B\right) - Var\left(\widetilde{\theta}_M\right)\right)$	$\frac{1000^2}{\mu^2} \left(MSE\left(\widetilde{\theta}_M\right) - MSE\left(\widetilde{\theta}_B\right)\right)$
0.1	814.19	1323.0	−508.81
0.25	64.745	90.0	−25.255
0.50	8.3321	8.0	0.3321
0.75	2.4035	1.2	1.2035
0.90	1.3751	0.28	1.0951

has the smallest asymptotic variance for all p among all asymptotically unbiased estimators (for all p asymptotically unbiased) by establishing in Theorem 17 that

$$\lim_{n\to\infty} \frac{Var\left(\widehat{\mu}_B\right)}{I^{-1}_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}\left(\mu\right)} = 1 \text{ for all } p \in (0,1)\,. \quad (4.1)$$

(See Definition 16.)

Among the three estimation scenarios, this is the only one where the Fisher

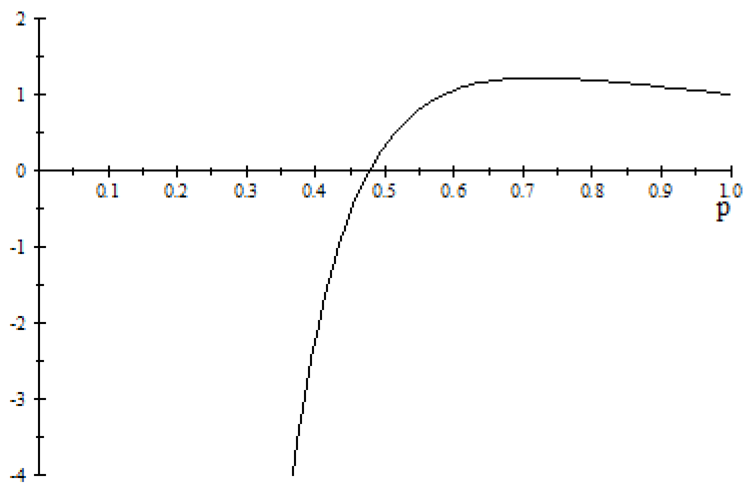
Table 4.9: μ *known,* θ *unknown,* $n = 500$

p	$\frac{500^2}{\mu^2} \left(MSE\left(\widehat{\theta}_M\right) - MSE\left(\widehat{\theta}_B\right)\right)$
0.1	30.432
0.25	6.2625
0.50	2.2515
0.75	1.3614
0.90	1.1143

Table 4.10: *μ known, θ unknown, $n = 1000$*

p	$\frac{1000^2}{\mu^2}\left(MSE\left(\widehat{\theta}_M\right) - MSE\left(\widehat{\theta}_B\right)\right)$
0.1	30.341
0.25	6.2563
0.50	2.2508
0.75	1.3613
0.90	1.1142

Figure 4.1: $\lim_{n\to\infty} \frac{n^2}{\mu^2}\left(MSE\left(\widetilde{\theta}_M\right) - MSE\left(\widetilde{\theta}_B\right)\right)$, $0 < p < 1$



Information exists for the vector $\left(T_1, T_2, ..., T_{[np]-1}, Y_{[np]}\right)$. We leave the other two cases of such optimality as two additional conjectures in Chapter 5. We formally define what it means for an estimator to asymptotically achieve its CRLB (provided the Fisher Information exists).

Definition 16. *Let X_n be a random vector indexed by a parameter θ and having Fisher Information $I_{X_n}(\theta)$. (See [7].) We say that an estimator $T(X_n)$ of θ asymptotically attains the CRLB based on X_n if the following criteria are satisfied.*

i.

$$\lim_{n\to\infty}\frac{Bias^2\left(T\left(X_n\right)\right)}{Var\left(T\left(X_n\right)\right)}=0$$

ii.

$$Bias^2\left(T\left(X_n\right)\right),Var\left(T\left(X_n\right)\right)\to 0\ \ as\ n\to\infty$$

iii.

$$\lim_{n\to\infty}\frac{Var\left(T\left(X_n\right)\right)}{I_{X_n}^{-1}\left(\theta\right)}=1$$

Note that Definition 16 can be reduced to 4.1 for $\widehat{\mu}_B$. Also by Theorems 12 and 14, $\widehat{\mu}_{AM}$ also satisfies these three conditions given in Definition 16. For this one parameter case, the Fisher Information for the k^{th} highest order statistic (See [1].) is given by

$$\begin{aligned}
I_{Y_k}\left(\mu\right) &= \frac{1}{\mu^2}+\frac{1}{\mu^2}\frac{n\left(n-\left(n-k+1\right)+1\right)}{n-k-1}\left(\left(\sum_{j=k-1}^{n-1}\frac{1}{j}\right)^2+\sum_{j=k-1}^{n-1}\frac{1}{j^2}\right)\\
&= \frac{1}{\mu^2}+\frac{1}{\mu^2}\frac{n\left(n-\left(n-k+1\right)+1\right)}{n-k-1}\left(\left(\sum_{j=k-1}^{n-1}\frac{1}{j}\right)^2+\sum_{j=k-1}^{n-1}\frac{1}{j^2}\right)\\
&= \frac{1}{\mu^2}+\frac{1}{\mu^2}\frac{nk}{n-k-1}\left(\left(\sum_{j=k-1}^{n-1}\frac{1}{j}\right)^2+\sum_{j=k-1}^{n-1}\frac{1}{j^2}\right). \quad (4.2)
\end{aligned}$$

With $k=[np]$, we will recognize the

$$\lim_{n\to\infty}\frac{n}{\mu^2}I_{Y_{[np]}}^{-1}\left(\mu\right)$$

function to be the same as 3.18, which shows that

$$Y_{[np]}^{*}=\frac{Y_{[np]}}{\sum\limits_{j=[np]}^{n}\frac{1}{j}}$$

is asymptotically optimal in the same sense as 4.1. Now since $T_1\overset{d}{=}\exp\left(\mu\right)$, we have that

$$\begin{aligned} &I_{T_1}(\mu) \\ &= Var\left(\frac{d}{d\mu}\log\left(\frac{1}{\mu}\exp\left(-\frac{1}{\mu}T_1\right)\right)\right) \\ &= Var\left(\frac{d}{d\mu}\left(-\frac{T_1}{\mu}\right)\right) \\ &= \frac{1}{\mu^4}\mu^2 \\ &= \frac{1}{\mu^2}. \ (4.3) \end{aligned}$$

So by independence it follows that

$$\begin{aligned} &I_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}(\mu) \\ &= \frac{[np]-1}{\mu^2} + I_{Y_{[np]}}(\mu). \ (4.4) \end{aligned}$$

In light of 4.2 and 4.4, we can now establish 4.1 as given in Theorem 17.

Theorem 17. *Let $X_1, X_2, ..., X_n$ be iid $\exp(\mu)$ random variables and $Y_1 > Y_2 > .... > Y_{[np]}$ be the upper $[np]$ order statistics. Then,*

$$i. \ \lim_{n\to\infty}\frac{n}{\mu^2}I^{-1}_{Y_{[np]}}(\mu) = \frac{1-p}{p\log^2 p}$$

and consequently,

$$ii. \ \lim_{n\to\infty}\frac{Var(\widehat{\mu}_B)}{I^{-1}_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}(\mu)} = 1.$$

Proof. i. By 4.2,

$$\begin{aligned} &\lim_{n\to\infty}\frac{\mu^2}{n}I_{Y_{[np]}}(\mu) \\ &= \lim_{n\to\infty}\left(\frac{1}{n} + \frac{[np]}{n-[np]-1}\left(\left(\sum_{j=[np]-1}^{n-1}\frac{1}{j}\right)^2 + \sum_{j=[np]-1}^{n-1}\frac{1}{j^2}\right)\right) \\ &= \frac{p}{1-p}\log^2 p + \frac{p}{1-p}\lim_{n\to\infty}\left(\frac{1}{np} - \frac{1}{n}\right) \\ &= \frac{p}{1-p}\log^2 p, \end{aligned}$$

and then taking the inverse of this, we arrive at i..

ii. By 4.4 and part i.,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{\mu^2}{n}I_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}\left(\mu\right)\\
&=p+\frac{p}{1-p}\log^2 p\\
&=\frac{p}{r\left(p\right)}.
\end{aligned}$$

So by part i. of Theorem 12,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{Var\left(\widehat{\mu}_B\right)}{I^{-1}_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}\left(\mu\right)}\\
&=\lim_{n\to\infty}\frac{\frac{n}{\mu^2}Var\left(\widehat{\mu}_B\right)}{\frac{n}{\mu^2}I^{-1}_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}\left(\mu\right)}\\
&=\frac{\frac{r(p)}{p}}{\frac{r(p)}{p}}\\
&=1.
\end{aligned}$$

□

As we stated in Section 4.1, we cannot establish optimality in the sense of 4.1 in all three estimation scenarios, so we leave that generalization as a conjecture in Chapter 5.

4.6 The Relevance of Theorem 17 to the Limiting Ratio Function $r\left(p\right)$

Recall the definition of $s\left(p\right)=r\left(p\right)/p$ given in Section 2.3. We have already seen from Theorem 3, Table 2.3 and Figure 2.1 that $s\left(p\right)>1$ and also have seen how quickly $s\left(p\right)$ approaches one. There is practically a horizontal asymptote at $y=1$ for that function, and that limit is nearly attained at $p=0.5$. This illustrates how high a percentage of the lower order statistics can be censored before the function

$$\lim_{n\to\infty}\frac{n}{\mu^2}Var\left(\widehat{\mu}_B\right)=s\left(p\right)$$

approaches one, or in more practical terms how high a percentage of the lower order statistics can be lost before $Var\left(\widehat{\mu}_B\right)$ is to within a certain level of accuracy of the CRLB based on all the $X's$, μ^2/n. We give intuition of these properties for the ratio function using the Fisher Information introduced in Section 4.5. Other than by using the formulas, there is no real probabilistic and/or statistical proof for these properties. However they are essentially intuitively obvious, and so our only objective in this section is to further illustrate where we left off in Section 2.3. Note that for finite n and $p\in(0,1)$ it *should* follow that,

$$Var\left(\widehat{\mu}_B\right)\geq I^{-1}_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}\left(\mu\right)>I^{-1}_{X_1,X_2,\ldots,X_n}\left(\mu\right)=\frac{\mu^2}{n}. \quad (4.5)$$

Since $p<1$, one would then think by 4.5 and as a consequence of part ii. of Theorem 17 that

$$s\left(p\right)=\lim_{n\to\infty}\frac{n}{\mu^2}I^{-1}_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}\left(\mu\right)>1,$$

which would suggest that $r\left(p\right)>p$, and indeed this is the case since

$$\begin{aligned}s^{-1}\left(p\right)&=\frac{p}{r\left(p\right)}\\&=p+\frac{p}{1-p}\log^2 p\end{aligned}$$

can be shown to be an increasing function which increases from zero to one on the interval $(0,1)$. (This means that $s\left(p\right)$ decreases from infinity to one on the interval $(0,1)$.) More interestingly, the reason why $s\left(p\right)$ approaches one as quickly as it does is really because of the help of the last order statistic $Y_{[np]}$. The contribution of information from $Y_{[np]}$ to

$$s^{-1}\left(p\right)=\lim_{n\to\infty}\frac{\mu^2}{n}I_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}\left(\mu\right)$$

is by part i. of Theorem 17

$$\lim_{n\to\infty}\frac{\mu^2}{n}I_{Y_{[np]}}\left(\mu\right)=\frac{p}{1-p}\log^2 p. \quad (4.6)$$

Note that

$$\frac{d}{dp}\frac{p}{1-p}\log^2 p$$
$$= \frac{1-p+p}{(1-p)^2}\log^2 p + \frac{2\log p}{1-p},$$

or rearranging and setting the derivative to zero, we have

$\log^2 p + 2\log p\,(1-p) = 0$, and so this happens when

$\log p\,(\log p + 2\,(1-p)) = 0$, or when

$$\log p = -2\,(1-p)\,. \quad (4.7)$$

The solution to 4.7 is approximately $p \approx 0.1953$, and 4.7 agrees with the result given by Balakrishnan for where the asymptotic Fisher Information of $Y_{[np]}$ maximizes. (See [1].) So when $p \approx 0.1953$,

$$\lim_{n\to\infty}\frac{\mu^2}{n}I_{Y_{[np]}}(\mu)$$

maximizes. At that value where it is maximized,

$$\lim_{n\to\infty}\frac{\mu^2}{n}I_{Y_{[np]}}(\mu) = 0.647\,38,$$

so therefore,

$$\begin{aligned} s^{-1}(0.1953) &= 0.1953 + 0.647\,38 \\ &= 0.842\,68, \text{ or} \\ s(0.1953) &= (0.1953 + 0.647\,38)^{-1} \\ &= 1.\,186\,7. \end{aligned}$$

So because of the fact that $I_{Y_{[np]}}(\mu)$ maximizes at such a low value of $p = 0.1953$, this causes the asymptotic normalized Fisher Information

$$s^{-1}(p) = \lim_{n\to\infty}\frac{\mu^2}{n}I_{T_1,T_2,\ldots,T_{[np]-1},Y_{[np]}}(\mu)$$

to already increase to over 84% of its maximum value of one, or in other words drop the variance of $\widehat{\mu}_B$ to just 18.67% above the CRLB based on all the data. Since by Theorems 8 and Theorem 17 $\widehat{\mu}_B$ and $Y^*_{[np]}$ asymptotically attain the CRLB with respect to their own data, (i.e., $Y^*_{[np]}$ with respect to the CRLB based on $Y_{[np]}$) the inverses of their respective Fisher Informations are asymptotically equal to their asymptotic variances. So it does not matter whether we study the inverse of their Fisher Informations or their variances asymptotically. They are the same according to Definition 16. This is the statistical reasoning for why the variance of $\widehat{\mu}_B$ approaches one as quickly as it does. Corollary 4 from Chapter 2 is simply a numerical proof based on the asymptotic normalized variance of $\widehat{\mu}_B$. For finite n, $\widehat{\mu}_B$ does not quite reach the CRLB based on just $T_1, T_2, ..., T_{[np]-1}$ and $Y_{[np]}$ because the unbiased estimator $Y^*_{[np]}$ does not achieve the CRLB based on $Y_{[np]}$ for finite n. It is considerably close though according to Balakrishnan's approximation for the Fisher Information of $Y_{[np]}$. (See [1].)

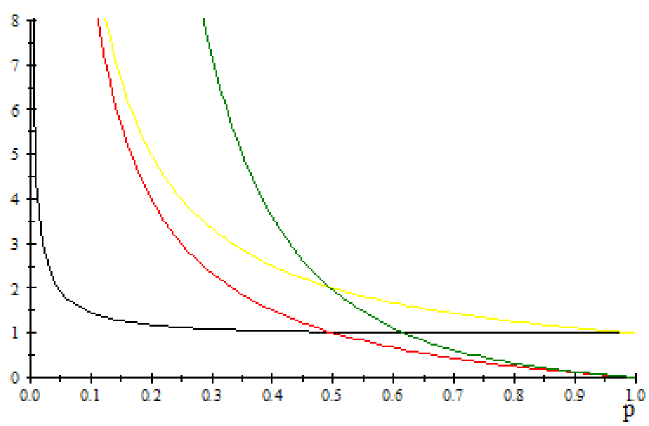
Chapter 5

Further Ideas to Consider

5.1 The Decision of Which Parameter to Estimate When Neither is Known

Recall the limiting functions of p given in Theorem 3 and Theorem 12. For the sake of argument in this section, we have plot in Figure 5.1 the asymptotic normalized variances for the four estimators $\widehat{\mu}_B$, $\widehat{\theta}_B$, $\widetilde{\mu}_B$ and $\widetilde{\theta}_B$.

Figure 5.1: $\frac{1}{p\left(1+\frac{\log^2 p}{1-p}\right)}$, $\frac{1}{p}$, $\frac{1-p}{p}$, $\frac{1-p}{p}\left(1+\frac{\log^2 p}{1-p}\right)$



The estimators of μ have curves bounded below by one which is of course because of the CRLB based on all the $X's$. The θ estimator functions are not bounded below by any positive constant because the CRLB does not apply to them. Recall the three estimation scenarios.

i. Know θ

ii. Do not know either of the parameters

iii. Know μ

You can only be in exactly one of those three situations, and in ii., which parameter being targeted for estimation is questionable. (assuming we are only interested in one of them) That being said, there are six comparisons here, and one of them is built in to ii.. The reason one of these comparisons is built in to ii. is because we might ask are we better off estimating μ or θ in situation ii.. As it turns out for $p < 0.5$, it is better to be estimating μ when in ii., and when $p > 0.5$, it is better to be estimating θ when in ii., and that is evident by the plots in Figure 5.1. A fourth comparison (since we have discussed two of them in this thesis) is i. and iii., or in other words, are we better off knowing θ or μ? The estimator for μ when knowing θ is better than the estimator for θ when knowing μ for $p < 0.5$, and for $p > 0.5$, it is the exact opposite. So being in i. is better for $p < 0.5$, and being in iii. is better for $p > 0.5$. (comparing i. and iii.) Rewriting the ratio identity we have that

$$\frac{Var\left(\widehat{\mu}_B\right)}{Var\left(\widehat{\theta}_B\right)} = \frac{Var\left(\widetilde{\mu}_B\right)}{Var\left(\widetilde{\theta}_B\right)}, \quad (5.1)$$

which gives yet another interpretation we are hinting towards. In particular, at $p \approx 0.5$, these two ratios given in 5.1 are both equal to one. So two comparisons we never mentioned also have a ratio identity. The difference between the comparison on the left of 5.1 and the right of 5.1 is that on the left, these are two different situations, while on the right, they are the same situation. If we were in situation ii., we would have the option of estimating μ or θ, and our judgement would be based on the value of p. If we are in situations i. or iii., we only have one of the two parameters to estimate, and since we can only be in one of those two situations, we do not have a choice of which parameter to estimate contrary to being in situation ii.. If $p < 0.5$ and if we knew one parameter, the more helpful parameter to know would be θ. If $p > 0.5$, the more helpful parameter to know is μ. That is the disadvantage of being in i. and iii. as opposed to ii. because we cannot in reality decide which parameter to know. So for instance if $p < 0.5$ and we know neither of the parameters, we would choose $\widetilde{\mu}_B$ over $\widetilde{\theta}_B$. Suppose that another observer knows μ with the same proportion $p < 0.5$ of upper order statistics available so that he is estimating θ using $\widehat{\theta}_B$. Recall that

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\theta}_B\right) = \frac{1-p}{p}, \ (5.2)$$

and

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widetilde{\mu}_B\right) = \frac{1}{p}, \ (5.3)$$

so that 5.2 is just 5.3 shifted down by one. This difference is insignificant if $p << 0.5$ and close to zero as can be seen from Figure 5.1. Just consider the fact that

$$\lim_{p\to 0} \frac{\frac{1}{p}}{\frac{1-p}{p}}$$

$$= 1.$$

The point is if p is small, $\widehat{\theta}_B$ is not that much of a better estimator than $\widetilde{\mu}_B$, and when in situation ii., we do have the option of choosing between $\widetilde{\mu}_B$ over $\widetilde{\theta}_B$. (or targeting μ as the parameter of interest when in ii.) When $p \approx 0.5$, the ratio of 5.2 divided by 5.3 is 0.5. In other words near $p = 0.5$, $Var\left(\widetilde{\mu}_B\right) \approx 2Var\left(\widehat{\theta}_B\right)$, and $Var\left(\widetilde{\mu}_B\right) \approx Var\left(\widetilde{\theta}_B\right)$. So near $p = 0.5$, it really does not matter which parameter you choose to estimate when you do not know either of the parameters. (i.e., situation ii.) The question is how small does p have to be so that $Var\left(\widetilde{\mu}_B\right)/Var\left(\widehat{\theta}_B\right)$ is "close" enough to one. How "close" is close in other words? You as the one who knows neither of the parameters is trying to outmatch someone who knows μ, and it is no contest if p is not small or anywhere near 0.5 for that matter. You ideally want a very small percentage of upper order statistics available if you are the blinded observer who does not know the value of either of the parameters versus someone who knows μ. That covers the fifth comparison. Now suppose $p > 0.5$ once again. Then in situation ii., we would select $\widetilde{\theta}_B$. Suppose another man knows θ so he has nothing better to do with his time than to estimate μ. This would be situation i. for this guy. Now recall that

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widetilde{\theta}_B\right) = \frac{1-p}{p}\left(1 + \frac{\log^2 p}{1-p}\right), \quad (5.4)$$

and

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\mu}_B\right) = \frac{1}{p\left(1 + \frac{\log^2 p}{1-p}\right)}. \quad (5.5)$$

Equating 5.4 to 5.5 and solving for p, we obtain $p = 0.615\,47$. Unlike the previous example, for $p > 0.61547$, $\widetilde{\theta}_B$ is in fact a better estimator than $\widehat{\mu}_B$, so why not choose $\widetilde{\theta}_B$ over $\widetilde{\mu}_B$ for that interval of $(0.61547, 1)$? If we choose to estimate θ when $p > 0.61547$, we would out perform the one using $\widehat{\mu}_B$, and that is the sixth comparison between the four estimators. There may be an advantage in knowing a parameter, but there is also a possible disadvantage in knowing a parameter because the one who knows neither parameter can choose. However, the one who knows neither parameter in reality has no control over how high a percentage of the lower order statistics are misplaced. In conclusion, which parameter do we estimate? It seems that the answer would depend on the practicallity of the problem. In what real world situation would p be small, intermediate or high? Two of the six comparisons between the estimators we covered, and they are independent of p. However with those two which we discussed, we were a bit tunnel visioned in only being open to estimating one parameter. We summarize our analysis in Table 5.1 and leave it up to the reader to determine in what particular instances p is small, large or in between. That will determine in the other four cases which parameter is better to estimate. Cases i. through iii. are just as above.

Key Questions: 1. When is p in the last four cases in Table 5.1 applicable?

2. How small does p have to be in the last case? (case six)

3. Is it better to know a parameter or to not know either of the parameters?

Table 5.1: *Estimator scenarios from two different observers*

Case	Observer A (Estimator 1)	Observer B (Estimator 2)
$\widehat{\mu}_B$, $\widetilde{\mu}_B$	Wins for all p	Loses for all p
$\widehat{\theta}_B$, $\widetilde{\theta}_B$	Wins for all p	Loses for all p
$\widehat{\mu}_B$, $\widehat{\theta}_B$	Wins for $p < 0.5$	Wins for $p > 0.5$
$\widetilde{\mu}_B$, $\widetilde{\theta}_B$	Wins for $p < 0.5$	Wins for $p > 0.5$
$\widehat{\mu}_B$, $\widetilde{\theta}_B$	Wins for $p < 0.61547$	Wins for $p > 0.61547$
$\widehat{\theta}_B$, $\widetilde{\mu}_B$*	Wins for all p	Loses for all p

*-Observer A always has a lower variance, but for small p, he is not much better off than Observer B!

5.2 Type-II Right Censoring and the CRLB Parodox

Consider as in this thesis $X_1, X_2, ..., X_n$ being *iid* random variables where $X_i = W_i + \theta$, where $W_1, W_2, ..., W_n$ are *iid* $\exp(\mu)$ random variables. We did Type-II Left Censoring, and the difference maker in the estimation of μ was the k^{th} highest order statistic Y_k when determining the observer knowing θ over another who does not know θ. Consider only having the lowest k order statistics instead given by $Y_1 < Y_2 < ..... < Y_k$. Suppose we know θ and would like to estimate μ. If $k < n$, we cannot use Theorem 1 to derive the BLUE because Y_k is not independent of the lower $k-1$ spacings. The BLUE for μ is (See [1].)

$$\widehat{\mu}_B = \frac{\sum_{j=1}^{k-1} (n-j)(Y_{j+1} - Y_j) + nY_1}{k},$$

and

$$Var(\widehat{\mu}_B) = \frac{\mu^2}{k}. \quad (5.6)$$

If θ is unknown, the BLUE is

$$\widetilde{\mu}_B = \frac{\sum_{j=1}^{k-1} (n-j)(Y_{j+1} - Y_j)}{k-1},$$

and

$$Var\left(\widetilde{\mu}_B\right) = \frac{\mu^2}{k-1}, \ (5.7)$$

so if we set $k = [np]$,

$$\frac{Var\left(\widehat{\mu}_B\right)}{Var\left(\widetilde{\mu}_B\right)} \to 1,$$

and further $Var\left(\widetilde{\mu}_B\right) - Var\left(\widehat{\mu}_B\right) \to 0$ at the rate of $1/n$. Note that no information was given up when not knowing θ contrary to when we were observing the highest k order statistics. Knowing θ would only have a significant advantage if k were fixed and small. In fact, 5.7 is exactly the same as its upper k order statistics analog. (without knowing θ) The difference here is 5.6. Note that

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(\widehat{\mu}_B\right) = \frac{1}{p},$$

so unlike where we were looking at the upper k order statistics, $Var\left(\widehat{\mu}_B\right)$ does not approach the CRLB based on all the $X's$ fast at all contrary to its upper order statistics analog which has an asymptotic expression of (when also multiplied by n/μ^2)

$$\frac{1}{p\left(1 + \frac{\log^2 p}{1-p}\right)}.$$

Whether θ is known or not here, the BLUEs for μ have variances which only (for the most part, the latter) reach the CRLB as if we just observed any of the k random iid random variables such as $X_1, X_2, ..., X_k$. So when observing the lower k order statistics, there is no significant advantage in knowing θ, which is not surprising because after all, we have the minimum recorded Y_1. What is surprising is the fact that having the upper say 50% order statistics we nearly approach the CRLB of μ^2/n, while if we have the lower 50% of the order statistics, the asymptotic normalized variance is nearly twice as high as the CRLB of μ^2/n. We compare the two asymptotic normalized variances here in the following table.

The only way the lower and upper $100p\%$ variances for the one parameter BLUE of μ cross over is if in fact $p = 1$. Otherwise, the BLUE based on the upper $100p\%$ always has the lower variance than the lower $100p\%$.

Table 5.2: *Lower versus upper asymptotically*

p	Lower $100p\%$	Upper $100p\%$
0.1	10	1.4512
0.5	2	1.0199
0.75	1.3333	1.0017
0.95	1.0526	1.0

Key Questions: 1. Does the ratio identity hold when we only have the lower $100p\%$ of the data recorded?

2. In what real world applications do either of these situations arise?

Lower $100p\%$ lost or upper $100p\%$?

3. Why is it better to be observing the upper $100p\%$ order statistics when estimating μ?

5.3 The CRLB Based on the Data Given

The CRLB based on the k upper order statistics can be calculated by using the independence of Y_k and the upper $k-1$ spacings. To explain why in Chapter 3 the variance of $Y^*_{[np]}$ minimized when $p = 0.2$ seems to be consistent with Balakrishnan's statement of the Fisher Information of an order statistic being maximized when looking at the 80^{th} sample percentile. (See [1].) We invite the reader to compare the asymptotic normalized variance of $Y^*_{[np]}$ with the approximation of the inverse of the Fisher Information given by Balakrishnan, which is

$$\begin{aligned} I^{-1}_{Y_{[np]}}(\mu) &\approx \left(\frac{2}{\mu^2} + \frac{np}{\mu^2(1-p)}\log^2 p\right)^{-1} \\ &= \frac{\mu^2}{n}\left(\frac{2}{n} + \frac{p}{1-p}\log^2 p\right)^{-1} \quad (5.8) \end{aligned}$$

Recall from Theorem 8 that

$$\lim_{n\to\infty} \frac{n}{\mu^2} Var\left(Y^*_{[np]}\right) = \frac{1-p}{p\log^2 p}. \quad (5.9)$$

Rewriting 5.9 as an approximation, we have

$$Var\left(Y^*_{[np]}\right) \approx \frac{\mu^2}{n}\frac{1-p}{p\log^2 p}. \quad (5.10)$$

For finite n, 5.8 and 5.10 are practically identical. Why is $Y^*_{[np]}$ not perfectly efficient though? This would prevent the estimator of $\widehat{\mu}_B$ from being 100% efficient as well as $\widehat{\mu}_{AM}$. The problem does not lie in the upper $[np]-1$ spacings. It lies in $Y_{[np]}$, and the question is what is the reason for it. Further, in the other two estimation scenarios, is it possible to obtain an unbiased estimator which at least asymptotically achieves the CRLB?

5.4 Approximate Maximum Likelihood Estimation of μ (one parameter exponential)

How accurate are our variance and bias squared calculations for our AMLE $\widehat{\mu}_{AM}$? Further, does the function (of p)

$$\frac{n^2}{\mu^2}\left(MSE\left(\widehat{\mu}_{AM}\right) - MSE\left(\widehat{\mu}_B\right)\right)$$

converge? We conjecture that it does based on the consistency of the values in Tables 4.3 and 4.4. We would imagine some significant numerical error in the variance and bias squared calculations because we do not have the MLE, but a mere approximation of it based on a Taylor Series expansion. (See [4].)

5.5 A Generalization to Non-Exponential Distributions

Without the independence of the *upper* spacings, other authors still managed to derive BLUEs and MLEs using other methods. How much of these results apply to other distributions if any? The conjecture we have is none at all unless they are close to being exponential. For example, a more practical model than having a distribution function being $F(x) = 1 - \exp\left(-\frac{1}{\mu}(x-\theta)\right)$, $x \geq \theta$ would be one satisfying the condition $F(x) = 1-\exp\left(-\frac{1}{\mu}(x-\theta)\right)$ for sufficiently large x. Here in addition to being interested in the mean μ, one might want to estimate the projected start time θ. An even more general distribution would be one that simply satisfies the condition

$$\frac{1 - F(x)}{\exp\left(-\frac{1}{\mu}(x-\theta)\right)} \to 1$$

as $x \to \infty$.

Appendix A

Limiting Results, MLE Derivations

A1: Limiting Results

Theorem 18. *Let $p \in (0,1)$. Then,*

i.

$$\lim_{n\to\infty} \frac{\sum\limits_{j=[np]}^{n} \frac{1}{j}}{\log\left(\frac{n}{[np]}\right)} = 1$$

ii.

$$\lim_{n\to\infty} \frac{\sum\limits_{j=[np]}^{n} \frac{1}{j^2}}{\int\limits_{[np]}^{n} \frac{1}{x^2} dx} = 1$$

iii.

$$\lim_{n\to\infty} \frac{\sum\limits_{j=[np]}^{n} \frac{1}{j^3}}{\int\limits_{[np]}^{n} \frac{1}{x^3} dx} = 1$$

iv.

$$\lim_{n\to\infty} n \left(\sum_{j=[np]}^{n} \frac{1}{j} - \log\left(\frac{n}{[np]}\right) \right) = \frac{1+p}{2p}$$

Proof. i. To prove i., first note by over approximating the integral, (by left endpoints) it is clear that

$$\sum_{j=[np]}^{n-1} \frac{1}{j} - \log\left(\frac{n}{[np]}\right) \geq 0. \ (1)$$

By under approximating the integral, it is clear that

$$\sum_{j=[np]+1}^{n} \frac{1}{j} - \log\left(\frac{n}{[np]}\right) \leq 0. \ (2)$$

So by (2)

$$\begin{aligned}
&\sum_{j=[np]}^{n-1} \frac{1}{j} - \log\left(\frac{n}{[np]}\right) \\
&\leq \sum_{j=[np]}^{n-1} \frac{1}{j} - \sum_{j=[np]+1}^{n} \frac{1}{j} \\
&= \frac{1}{[np]} - \frac{1}{n} \\
&\leq \frac{1}{[np]},
\end{aligned}$$

and therefore by (1),

$$0 \leq \sum_{j=[np]}^{n} \frac{1}{j} - \log\left(\frac{n}{[np]}\right) \leq \frac{1}{[np]} + \frac{1}{n}$$

So it follows from this that

$$\lim_{n\to\infty} \frac{\sum_{j=[np]}^{n} \frac{1}{j}}{\log\left(\frac{n}{[np]}\right)} = 1.$$

ii. Observe by over approximating the integral, we arrive at

$$\sum_{j=[np]}^{n-1} \frac{1}{j^2} - \int_{[np]}^{n} \frac{1}{x^2} dx \geq 0. \ (3)$$

Also by under approximating, we obtain

$$\sum_{j=[np]+1}^{n} \frac{1}{j^2} - \int\limits_{[np]}^{n} \frac{1}{x^2} dx \leq 0. \ (4)$$

So

$$\begin{aligned} &\sum_{j=[np]}^{n} \frac{1}{j^2} - \int\limits_{[np]}^{n} \frac{1}{x^2} dx \\ &\leq \sum_{j=[np]}^{n} \frac{1}{j^2} - \sum_{j=[np]+1}^{n} \frac{1}{j^2} \\ &= \frac{1}{[np]^2} \ (5) \end{aligned}$$

Rewriting (5) and using (3) we have that

$$0 \leq \frac{\sum\limits_{j=[np]}^{n} \frac{1}{j^2}}{\int\limits_{[np]}^{n} \frac{1}{x^2} dx} - 1 \leq \frac{1}{[np]^2} \frac{1}{\frac{1}{[np]} - \frac{1}{n}} = \frac{1}{[np] - \frac{[np]^2}{n}} \to 0$$

as $n \to \infty$.

iii. Observe by over approximating the integral, we arrive at

$$\sum_{j=[np]}^{n-1} \frac{1}{j^3} - \int\limits_{[np]}^{n} \frac{1}{x^3} dx \geq 0. \ (6)$$

Also by under approximating, we obtain

$$\sum_{j=[np]+1}^{n} \frac{1}{j^3} - \int\limits_{[np]}^{n} \frac{1}{x^3} dx \leq 0. \ (7)$$

So

$$\sum_{j=[np]}^{n} \frac{1}{j^3} - \int_{[np]}^{n} \frac{1}{x^3} dx$$
$$\leq \sum_{j=[np]}^{n} \frac{1}{j^3} - \sum_{j=[np]+1}^{n} \frac{1}{j^3}$$
$$= \frac{1}{[np]^3}. \quad (8)$$

Rewriting (8) and using (6) we have that

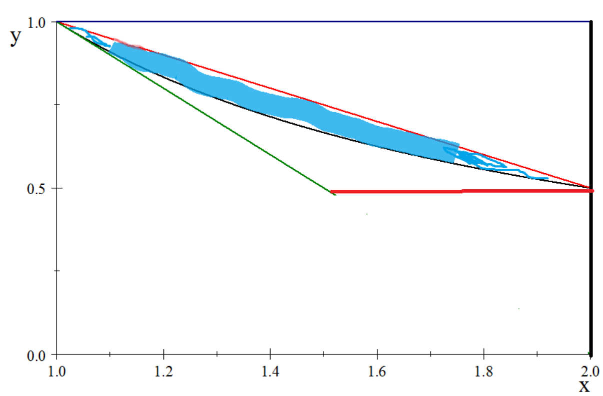
$$0 \leq \frac{\sum_{j=[np]}^{n} \frac{1}{j^3}}{\int_{[np]}^{n} \frac{1}{x^3} dx} - 1 \leq \frac{1}{[np]^3} \frac{1}{\frac{1}{2[np]^2} - \frac{1}{2n^2}} \to 0$$

as $n \to \infty$.

This establishes iii..

iv.

Figure A.1: Plot on the interval $(1, 2)$



Consider the function $f(x) = 1/x$ on the interval $(1, 2)$. The function in the middle is $f(x)$. The other two are lines where the upper line is formed by connecting the coordinates $(1,1)$ to $(2,1/2)$, while the other is the tangent line running through the point $(1,1)$ on the graph of $f(x)$. If we draw two horizontal lines at $y=1$ and $y=1/2$ with the second horizontal line being cut off where the line intersects the tangent line, we form two triangles and a trapezoid as depicted in Figure A1.The area of the upper triangle is

$$\begin{aligned} A_1 &= \frac{1}{2}\left(1-\frac{1}{2}\right) \\ &= \frac{1}{4}. \end{aligned}$$

The area of the lower triangle is $A_2 = A_T - A_1$, where A_T is the area of the trapezoid. The slope of the tangent line is equal to $m = -1/1^2 = -1$, so the equation for the tangent line is $y = 2 - x$. This means that when $x = 3/2$, the line intersects the horizontal line $y = 1/2$. So the length of the upper base of the trapezoid is one, while the length of the lower base is $1/2$. So

$$\begin{aligned} A_T &= \frac{1}{2}\left(1+\frac{1}{2}\right)\frac{1}{2} \\ &= \frac{3}{8}. \end{aligned}$$

And therefore, $A_2 = 1/8$. More generally, consider the interval $(j, j+1)$, where j is a positive integer. We can form a line from $(j, 1/j)$ to $(j+1, 1/(j+1))$ and create a second lower line which is tangent to the curve at $x = j$. The slope of this tangent line would then be $m = -1/j^2$. The line must satisfy the equation

$$\frac{1}{j} = -\frac{1}{j^2}j + b,$$

so

$$b = \frac{2}{j}.$$

The equation of the tangent line is then

$$y = -\frac{1}{j^2}x + \frac{2}{j}.$$

Let A_1, A_2 and A_T be the analogs of the previous example on the current interval we are considering. Then,

$$A_1 = \frac{1}{2}\left(\frac{1}{j} - \frac{1}{j+1}\right).$$

We would like to determine where the tangent line intersects the horizontal line $y = 1/(j+1)$. Substituting in to the linear equation, we have that

$$\begin{aligned} \frac{1}{j+1} &= -\frac{1}{j^2}x + \frac{2}{j}, \text{ or} \\ x &= 2j - \frac{j^2}{j+1} \\ &= \frac{j^2+2j}{j+1}. \end{aligned}$$

This means that the length of the lower base of the trapezoid is

$$\begin{aligned} & j + 1 - \frac{j^2+2j}{j+1} \\ &= \frac{j^2+2j+1-j^2-2j}{j+1} \\ &= \frac{1}{j+1}. \end{aligned}$$

The area of the trapezoid is then

$$\begin{aligned} A_T &= \frac{1}{2}\left(1 + \frac{1}{j+1}\right)\left(\frac{1}{j} - \frac{1}{j+1}\right) \\ &= \frac{1}{2}\frac{j+2}{j+1}\frac{1}{j(j+1)} \\ &= \frac{1}{2}\frac{j+2}{j(j+1)^2}. \end{aligned}$$

So,

$$\begin{aligned} A_2 &= \frac{1}{2}\frac{j+2}{j(j+1)^2} - \frac{1}{2}\left(\frac{1}{j} - \frac{1}{j+1}\right) \\ &= \frac{1}{2j(j+1)}\left(\frac{j+2}{j+1} - 1\right) \\ &= \frac{1}{2j(j+1)^2}. \end{aligned}$$

Now note by the concavity of the function $f(x) = 1/x$,

$$A_1 \leq \frac{1}{j} - \int_j^{j+1} \frac{1}{x} dx \leq A_1 + A_2, \text{ or}$$

$$\frac{1}{2}\left(\frac{1}{j} - \frac{1}{j+1}\right) \leq \frac{1}{j} - \int_j^{j+1} \frac{1}{x} dx \leq \frac{1}{2}\left(\frac{1}{j} - \frac{1}{j+1}\right) + \frac{1}{2j(j+1)^2} \text{ for } 1 \leq j \leq n-1. \quad (9)$$

Adding the terms up in (9) from $j = k$ to $j = n-1$ we obtain

$$\frac{1}{2}\left(\frac{1}{k} - \frac{1}{n}\right) \leq \sum_{j=k}^{n-1} \frac{1}{j} - \log\left(\frac{n}{k}\right) \leq \frac{1}{2}\left(\frac{1}{k} - \frac{1}{n}\right) + \sum_{j=k}^{n-1} \frac{1}{2j(j+1)^2}, \text{ or}$$

$$\frac{n}{2}\left(\frac{1}{k} + \frac{1}{n}\right) \leq n\left(\sum_{j=k}^{n} \frac{1}{j} - \log\left(\frac{n}{k}\right)\right) \leq \frac{n}{2}\left(\frac{1}{k} + \frac{1}{n}\right) + n\sum_{j=k}^{n-1} \frac{1}{2j(j+1)^2}. \quad (10)$$

Let $k = [np]$. Then,

$$\lim_{n\to\infty} \frac{n}{2}\left(\frac{1}{k} + \frac{1}{n}\right)$$
$$= \lim_{n\to\infty} \frac{n}{2}\left(\frac{1}{[np]} + \frac{1}{n}\right)$$
$$= \frac{1+p}{2p},$$

and note that this is the limit of the left-side of (10). Now by part iii.,

$$\lim_{n\to\infty} n \sum_{j=[np]}^{n-1} \frac{1}{2j(j+1)^2}$$
$$\leq \lim_{n\to\infty} n \sum_{j=[np]}^{n} \frac{1}{2j^3}$$
$$= \lim_{n\to\infty} \frac{n}{2} \int_{[np]}^{n} \frac{1}{x^3} dx$$
$$= \lim_{n\to\infty} \frac{n}{4}\left(\frac{1}{[np]^2} - \frac{1}{n^2}\right)$$
$$= 0.$$

So by letting $k = [np]$ and taking the limit of the right-side of (10) as $n \to \infty$, we establish iv.. $\square$

A2: MLE Derivations

Table A.1: *MLE Formulas*

Parameter Known?	μ known	μ unknown
θ known	N/A	$\widehat{\mu}_M = q_k \frac{\sum_{j=1}^{k-1} T_j}{k} + (1 - q_k)\, Y_k^{**}$
θ unknown	$\widehat{\theta}_M = Y_k + \mu \log\left(\frac{k}{n}\right)$	$\widetilde{\mu}_M = \frac{\sum_{j=1}^{k-1} T_j}{k}$, $\widetilde{\theta}_M = Y_k + \widetilde{\mu}_M \log\left(\frac{k}{n}\right)$

where

$$q_k = \frac{k}{R_k^* + k},$$

$$R_k^* = \frac{\left(\int_k^n \frac{1}{x} dx\right)^2}{\int_k^n \frac{1}{x^2} dx} = \frac{nk}{n-k} \log^2\left(\frac{n}{k}\right)$$

and

$$Y_k^{**} = \frac{Y_k - \theta}{\log\left(\frac{n}{k}\right)}.$$

The variance and bias squared for the MLEs shown in Table A1 are given by the following, which once again in addition to the formulas in Table A1, we will also derive. The formulas for the MSEs of each estimator can be determined simply by adding its variance and its biased squared.

Variances and Squared Biases for MLEs for μ and θ

i. *μ unknown, θ known*

$$Var\left(\widehat{\mu}_M\right) = \mu^2 q_k^2 \left(\frac{k-1}{k^2} + \frac{n^2}{\left(n-k\right)^2} \log^2\left(\frac{n}{k}\right) \sum_{j=k}^{n} \frac{1}{j^2}\right)$$

$$Bias^2\left(\widehat{\mu}_M\right) = \mu^2 \left(q_k \left(\frac{(k-1)}{k} - \frac{\sum\limits_{j=k}^{n} \frac{1}{j}}{\log\left(\frac{n}{k}\right)} \right) + \left(\frac{\sum\limits_{j=k}^{n} \frac{1}{j}}{\log\left(\frac{n}{k}\right)} - 1 \right) \right)^2$$

ii. *μ unknown, θ unknown*

$$Var\left(\widetilde{\mu}_M\right) = \frac{(k-1)\,\mu^2}{k^2}$$

$$Bias^2\left(\widetilde{\mu}_M\right) = \frac{\mu^2}{k^2}.$$

$$Var\left(\widetilde{\theta}_M\right) = \mu^2 \left(\sum_{j=k}^{n} \frac{1}{j^2} + \frac{k-1}{k^2} \log^2\left(\frac{k}{n}\right)\right)$$

$$Bias^2\left(\widetilde{\theta}_M\right) = \frac{\mu^2 \left((k-1)\log\left(\frac{k}{n}\right) + k \sum\limits_{j=k}^{n} \frac{1}{j}\right)^2}{k^2}$$

iii. *θ unknown, μ known*

$$Var\left(\widehat{\theta}_M\right) = \mu^2 \sum_{j=k}^{n} \frac{1}{j^2}$$

$$Bias^2\left(\widehat{\theta}_M\right) = \mu^2 \left(\sum_{j=k}^{n} \frac{1}{j} + \log\left(\frac{k}{n}\right)\right)^2$$

Case 1: μ and θ Unknown

The general likelihood function takes the form

$$L\left(\lambda, \theta\right) = C\left(1 - \exp\left(-\lambda\left(Y_k - \theta\right)\right)\right)^{n-k} \lambda^k \exp\left(-\lambda \sum_{j=1}^{k} \left(Y_j - \theta\right)\right), Y_k \geq \theta,$$

where C is a constant which is independent of $\lambda = \mu^{-1}$ and θ. We are using λ here instead of μ to simplify the calculus and algebra to follow. The log likelihood function then takes the form

$$l(\lambda, \theta) = K + (n-k)\log(1 - \exp(-\lambda(Y_k - \theta))) + k\log\lambda - \lambda\sum_{j=1}^{k}(Y_j - \theta)$$

$$l(\lambda, \theta) = K + (n-k)\log(1 - \exp(-\lambda(Y_k - \theta))) + k\log\lambda - \lambda k(Y_k - \theta) - \lambda\sum_{j=1}^{k-1}T_j, \theta \leq Y_k \quad (11)$$

To find the MLEs in this case, it is easiest to derive the MLE of θ in terms of λ. Notice in 11, we have singled out Y_k from the sum so that we have a function in the sum of one variable, which is

$$x = Y_k - \theta, Y_k \geq \theta \quad (12)$$

We can then call the function which only involves θ and rewrite it in terms of x.

$$h(x) = (n-k)\log(1 - \exp(-\lambda x)) - \lambda kx, x \geq 0. \quad (13)$$

Note that if $k = n$ in 13, the maximum is clearly achieved when $x = 0$, or when $\theta = Y_n$-that is, in the full sample case, the MLE for θ would be the upper bound which in the case of $k = n$ is the sample minimum Y_n. When $k < n$, h is a smooth function which comes up from $-\infty$ and increases to a unique maximum and then decreases back down to $-\infty$. Therefore, to find the maximum, we differentiate the function with respect to x and set the derivative equal to zero and then solve for x.

$$\begin{aligned}
h(x) &= (n-k)\log(1 - \exp(-\lambda x)) - \lambda kx \\
h'(x) &= \frac{\lambda(n-k)\exp(-\lambda x)}{1 - \exp(-\lambda x)} - \lambda k = 0 \\
\lambda(n-k)\exp(-\lambda x) - \lambda k + \lambda k\exp(-\lambda x) &= 0 \\
n\exp(-\lambda x) &= k \\
x &= -\frac{1}{\lambda}\log\left(\frac{k}{n}\right), \text{ for all } \lambda > 0. \quad (14)
\end{aligned}$$

Substituting 14 back in to 13 and in turn 13 back in to 11, we can then solve for the MLE of λ. Then by substitution back in to 12, we can determine the MLE of θ.

$$l\left(\lambda,\widetilde{\theta}_M\right) = K + (n-k)\log\left(1-\frac{k}{n}\right) + k\log\lambda + k\log\left(\frac{k}{n}\right) - \lambda\sum_{j=1}^{k-1} T_j$$

$$\frac{\partial}{\partial\lambda} l\left(\lambda,\widetilde{\theta}_M\right) = \frac{k}{\lambda} - \sum_{j=1}^{k-1} T_j = 0$$

$$\widetilde{\lambda}_M = \frac{k}{\sum_{j=1}^{k-1} T_j} \quad (15)$$

$$Y_k - \widetilde{\theta}_M = -\frac{\sum_{j=1}^{k-1} T_j}{k}\log\left(\frac{k}{n}\right)$$

$$\widetilde{\theta}_M = Y_k + \frac{\sum_{j=1}^{k-1} T_j}{k}\log\left(\frac{k}{n}\right). \quad (16)$$

Since our focus on estimation of the parameter μ, we should mention that it immediately follows from 15 that

$$\widetilde{\mu}_M = \frac{\sum_{j=1}^{k-1} T_j}{k}. \quad (17)$$

In this case of both parameters being unknown, note the similarity of the MLEs for μ and θ to the BLUEs of μ and θ given in Table 3.2. Now by 17,

$$\begin{aligned} Var\left(\widetilde{\mu}_M\right) &= Var\left(\widetilde{\mu}_M\right) \\ &= \frac{(k-1)\,\mu^2}{k^2}, \text{ and } (18) \\ E^2\left(\frac{\sum_{j=1}^{k-1} T_j}{k} - \mu\right) &= \left(\frac{k-1}{k}\mu - \mu\right)^2 \\ &= \frac{\mu^2}{k^2}. \quad (19) \end{aligned}$$

Also by 16 we have,

$$Var\left(\widetilde{\theta}_M\right) = \mu^2 \sum_{j=k}^{n} \frac{1}{j^2} + \frac{k-1}{k^2}\mu^2 \log^2\left(\frac{k}{n}\right), \text{ (20) and}$$

$$Bias^2\left(\widetilde{\theta}_M\right) = E^2\left(Y_k + \frac{\sum\limits_{j=1}^{k-1} T_j}{k}\log\left(\frac{k}{n}\right) - Y_k + \frac{\sum\limits_{j=1}^{k-1} T_j}{k-1}\sum_{j=k}^{n}\frac{1}{j}\right)$$

$$= \left(\frac{\log\left(\frac{k}{n}\right)}{k} + \frac{\sum\limits_{j=k}^{n}\frac{1}{j}}{k-1}\right)^2 (k-1)^2\mu^2. \text{ (21)}$$

Case 2: μ unknown, θ known

With θ being known, once again without loss of generality assume that $\theta = 0$. Once again to simplify the differentiation, we use the parameter $\lambda = \mu^{-1}$. Then the likelihood function takes the form

$$L(\lambda) = C\left(1 - \exp\left(-\lambda Y_k\right)\right)^{n-k} \lambda^k \exp\left(-\lambda \sum_{j=1}^{k} Y_j\right). \text{ (22)}$$

The log-likelihood function is

$$l(\lambda) = K + (n-k)\log\left(1 - \exp\left(-\lambda Y_k\right)\right) + k\log\lambda - \lambda\sum_{j=1}^{k} Y_j. \text{ (23)}$$

As $\lambda \to 0+$, $l(\lambda)$ tends to $-\infty$, increases up to a unique maximum $l\left(\widehat{\lambda}_M\right)$ and then goes right back down to $-\infty$ as $\lambda \to \infty$ as can be seen from 23. So differentiating and setting the derivative equal to zero, we obtain the equality

$$l'(\lambda) = \frac{(n-k)\,Y_k \exp\left(-\lambda Y_k\right)}{1 - \exp\left(-\lambda Y_k\right)} + \frac{k}{\lambda} - \sum_{j=1}^{k} Y_j = 0. \text{ (24)}$$

As we said in the introduction, the solution of 24 does not have an explicit form. However by the very argument preceding 24 describing the behavior of the the log-likelihood function, there is a unique solution to 24. Balakrishnan uses a linear

Taylor Series approximation of the random variable

$$h\left(Y_k\right)=\frac{Y_k \exp\left(-\lambda Y_k\right)}{1-\exp\left(-\lambda Y_k\right)} \quad (25)$$

around the $\frac{n-k+1}{n}100^{th}$ percentile of the $\exp(\lambda)$ distribution. Our work here is actually a special case of his in his paper. (See [4].) We will follow his exact steps. Considering the function

$$h(x)=\frac{x\exp(-\lambda x)}{1-\exp(-\lambda x)},$$

note that

$$h^{'}(x)=\frac{-\exp(-2\lambda x)+\exp(-\lambda x)-\lambda x\exp(-\lambda x)}{\left(1-\exp(-\lambda x)\right)^2}. \quad (26)$$

Setting $p=\exp(-\lambda x_p)$, we can find the *upper* $100p^{th}$ percentile of the $\exp(\lambda)$ distribution. The solution can easily be determined and is well known to be

$$x_p=-\frac{\log p}{\lambda}.$$

Following the same method of Balakrishnan for this type of problem, (See [4].) we should now expand the function $h(x)$ about the point

$$x=-\frac{\log\left(\frac{k}{n}\right)}{\lambda},$$

since Y_k should be around the upper $100\frac{k}{n}^{th}$ percentile. We then obtain a linear approximation to the random function $h(Y_k)$ given in 25. By 25 and 26,

$$h\left(-\frac{\log\left(\frac{k}{n}\right)}{\lambda}\right)=\frac{-\frac{\log\left(\frac{k}{n}\right)}{\lambda}\frac{k}{n}}{\frac{n-k}{n}}$$

$$=-\frac{k\log\left(\frac{k}{n}\right)}{\lambda(n-k)}, \quad (27)$$

and

$$h^{'}\left(-\frac{\log\left(\frac{k}{n}\right)}{\lambda}\right)=\frac{-\left(\frac{k}{n}\right)^2+\frac{k}{n}+\log\left(\frac{k}{n}\right)\frac{k}{n}}{\left(\frac{n-k}{n}\right)^2}. \quad (28)$$

So the random variable in 25 can be approximated by the first order Taylor Series expansion

$$h\left(Y_k\right) \approx \frac{\frac{k}{n}\left(\log\left(\frac{k}{n}\right)\right)^2}{\lambda\left(1-\frac{k}{n}\right)^2} + \frac{kY_k}{n-k}\left(1+\left(\frac{n}{n-k}\right)\log\left(\frac{k}{n}\right)\right). \quad (29)$$

Substituting 29 in to 24, we obtain

$$(n-k)\left(\frac{\frac{k}{n}\left(\log\left(\frac{k}{n}\right)\right)^2}{\lambda\left(1-\frac{k}{n}\right)^2} + \frac{kY_k}{n-k}\left(1+\left(\frac{n}{n-k}\right)\log\left(\frac{k}{n}\right)\right)\right) + \frac{k}{\lambda} - \sum_{j=1}^{k} Y_j = 0. \quad (30)$$

It is understood that 30 is only an approximate equality, but now we can obtain the Approximate Maximum Likelihood Estimator (AMLE) for μ. Rearranging 30 and solving for μ, we have

$$\frac{1}{\lambda}\left((n-k)\frac{\frac{k}{n}\left(\log\left(\frac{k}{n}\right)\right)^2}{\left(1-\frac{k}{n}\right)^2} + k\right) = \sum_{j=1}^{k} Y_j - kY_k\left(1+\left(\frac{n}{n-k}\right)\log\left(\frac{k}{n}\right)\right)$$

$$\widehat{\mu}_M = \frac{\sum_{j=1}^{k-1} T_j + \left(\frac{n}{n-k}\right)\log\left(\frac{n}{k}\right)kY_k}{k\left(\frac{n}{n-k}\left(\log\left(\frac{k}{n}\right)\right)^2 + 1\right)}$$

$$= q_k\frac{\sum_{j=1}^{k-1} T_j}{k} + \left(1-q_k\right)Y_k^{**}. \quad (31)$$

By 31,

$$
\begin{aligned}
Var\left(\widehat{\mu}_M\right) &= q_k^2 \frac{(k-1)\,\mu^2}{k^2} + (1-q_k)^2 \frac{\sum_{j=k}^{n} \frac{1}{j^2}}{\log^2\left(\frac{n}{k}\right)} \\
&= \mu^2 q_k^2 \left(\frac{k-1}{k^2} + \frac{n^2}{(n-k)^2} \log^2\left(\frac{n}{k}\right) \sum_{j=k}^{n} \frac{1}{j^2} \right), \text{(32) and} \\
Bias^2\left(\widehat{\mu}_M\right) &= E^2\left(\widehat{\mu}_M - \widehat{\mu}_B\right) \\
&= E^2 \left(q_k \frac{\sum_{j=1}^{k-1} T_j}{k} + (1-q_k)\, Y_k^{**} - r_k \overline{T} - (1-r_k)\, Y_k^{*} \right) \\
&= E^2 \left(\sum_{j=1}^{k-1} T_j \left(\frac{q_k}{k} - \frac{r_k}{k-1} \right) + Y_k \left(\frac{1-q_k}{\log\left(\frac{n}{k}\right)} - \frac{1-r_k}{\sum_{j=k}^{n} \frac{1}{j}} \right) \right) \\
&= \mu^2 \left(\frac{(k-1)\, q_k}{k} - r_k + \frac{\sum_{j=k}^{n} \frac{1}{j}}{\log\left(\frac{n}{k}\right)} (1-q_k) - (1-r_k) \right)^2 \\
&= \mu^2 \left(q_k \left(\frac{(k-1)}{k} - \frac{\sum_{j=k}^{n} \frac{1}{j}}{\log\left(\frac{n}{k}\right)} \right) + \left(\frac{\sum_{j=k}^{n} \frac{1}{j}}{\log\left(\frac{n}{k}\right)} - 1 \right) \right)^2 . \text{ (33)}
\end{aligned}
$$

Case 3: θ unknown, μ known

In this case, the likelihood function takes the form

$$
L(\theta) = C \left(1 - \exp\left(-\frac{1}{\mu} (Y_k - \theta) \right) \right)^{n-k} \exp\left(-\frac{1}{\mu} \sum_{j=1}^{k} (Y_j - \theta) \right), \theta \leq Y_k.
$$

This can be rewritten as

$$
L(\theta) = C \left(1 - \exp\left(-\frac{1}{\mu} (Y_k - \theta) \right) \right)^{n-k} \exp\left(-\frac{k}{\mu} (Y_k - \theta) \right) \exp\left(-\frac{1}{\mu} \sum_{j=1}^{k-1} T_j \right), \theta \leq Y_k.
$$

The log-likelihood is then

$$l(\theta) = K + (n-k)\log\left(1 - \exp\left(-\frac{1}{\mu}(Y_k - \theta)\right)\right) - \frac{k}{\mu}(Y_k - \theta) - \frac{1}{\mu}\sum_{j=1}^{k-1} T_j,\ \theta \leq Y_k. \quad (34)$$

We have already obtained the MLE for θ when μ is unknown. The part of the function in 34 where θ appears is the same function we rewrote as the expression in 13. In this case however, μ is a known constant. 14 leads directly to the solution for the MLE of θ upon substituting it in to 13 in the case where μ is known. So therefore,

$$\widehat{\theta}_M = Y_k + \mu\log\left(\frac{k}{n}\right). \quad (35)$$

By 35,

$$Var\left(\widehat{\theta}_M\right) = \mu^2 \sum_{j=k}^{n} \frac{1}{j^2}, \quad (36) \text{ and}$$

$$E^2\left(Y_k + \mu\log\left(\frac{k}{n}\right) - \theta\right) = \left(\mu\sum_{j=k}^{n}\frac{1}{j} + \theta + \mu\log\left(\frac{k}{n}\right) - \theta\right)^2$$
$$= \mu^2\left(\sum_{j=k}^{n}\frac{1}{j} + \log\left(\frac{k}{n}\right)\right)^2. \quad (37)$$

Bibliography

[1] C.B. Arnold, N. Balakrishnan, and H.N. Nagaraja, *A First Course in Order Statistics*, Wiley, New York, New York, 1992.

[2] N. Balakrishnan, *On the Maximum Likelihood Estimation of the Location and Scale Parameters of Exponential Distribution Based on Multiply Type II Censored Samples*, **17** (1990), 55–61.

[3] N. Balakrishnan and K. Balasubramanian, *Estimation for One and Two Parameter Exponential Distributions Under Multiple Type II Censoring*, **33** (1991), 203–216.

[4] N. Balakrishnan and A. Clifford Cohen, *Order Statistics and Inference Estimation Methods*, Academic Press, San Diego, California, 1991.

[5] P. Billingsley, *Probability and Measure: Anniversay Edition*, Wiley, Hoboken, New Jersey, 2012.

[6] H.A. David, *Order Statistics*, Wiley, New York, New York, 1970.

[7] Robert V. Hogg and Allen T. Craig, *Introduction to Mathematical Statistics: Fifth Edition*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1995.

[8] A. K. Md. Ehsanes Saleh, *Estimation of the Parameters of the Exponential Distribution Based on Optimum Order Statistics in Censored Samples*, **37** (1966), 1717–1735.

[9] A.E. Sarhan and B.G. Greenberg, *Estimation of Location and Scale Parameters by Order Statistics from Singly and Doubly Censored Samples Part I. The Normal Distrbution up to Samples of Size 10*, (1955), 427–451.

Vita

Education:

Towson University, B.S., Mathematics 2001

Lehigh University, M.S., Statistics 2004

Lehigh University, Ph.D., Applied Mathematics 2014

Employment History:

Lehigh University: 2003-2009, Teaching Assistant

Bloomsburg University: Summer 2006, Instructor

Lehigh Carbon Community College: Summer 2007, Instructor

Research Interests:

General Statistical Inference

Type-II Censored Data

Characteristics of Exponential Distribution

Statistical Inference for Exponential Distribution

Statistical Inference for Distribution Functions Similar to Exponential Distribution

Asymptotic Statistical Inference

Simulated Inferential Studies