INFERENCE IN POWER SERIES DISTRIBUTIONS

A dissertation submitted to
Kent State University in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy

by

Robert A. Korte

December, 2012
Dissertation written by

Robert A. Korte

B.S., Antioch College, 1989
M.A., Saint Louis University, 2005
Ph.D., Kent State University, 2012

Approved by

Dr. M. Kazim Khan, Chair, Doctoral Dissertation Committee

Dr. Hassan Allouba, Member, Doctoral Dissertation Committee

Dr. Oana Mocioalca, Member, Doctoral Dissertation Committee

Dr. Ye Zhao, Member, Outside Discipline

Dr. David Allender, Member, Graduate Faculty Representative

Accepted by

Dr. Andrew Tonge, Chair, Department of Mathematical Sciences

Dr. Raymond Craig, Associate Dean, College of Arts and Sciences
# TABLE OF CONTENTS

Acknowledgements ........................................................................ v

1 Introduction ............................................................................. 1

1.1 Early Experiments ................................................................... 1

1.1.1 Censored Observations ...................................................... 1

1.1.2 Truncated Observations-I .................................................. 3

1.1.3 Truncated Observations-II ............................................... 5

1.2 General Problem ..................................................................... 5

1.3 Examples of Power Series Distributions ................................. 6

2 Estimators and Their Shortcomings ........................................... 10

2.1 Maximum Likelihood (ML) and Method of Moments (MM) Estimation Results 10

2.1.1 Maximum Likelihood Estimation of \( \theta \) in the truncated and censored

Poisson distribution .............................................................. 10

2.1.2 MM Estimation of \( \theta C''(\theta)/C'(\theta) \) .............................. 11

2.2 Unbiased Estimators ............................................................. 15

2.2.1 Unbiased Estimation of \( P(Z = 0) \) .................................. 15

2.2.2 Unbiased Estimation of \( g(\theta) = \theta^p \) ............................. 16

2.3 Lack of Existence of C-R Estimators ..................................... 26

3 Optimality of the UMVU Estimator \( h_m(X) = \frac{aX - m}{aX} \) ............... 32

3.1 Asymptotic Normality of the UMVU Estimator ....................... 32

3.2 Limiting Absolute Mean Deviation ....................................... 35

3.3 Asymptotic Efficiency ......................................................... 42
Acknowledgements

First, I would like to thank my advisor, Dr. Mohammad Kazim Khan, for all his help, support and inspiration on this project and others, in this country and abroad.

Second, I need to thank my friend and colleague, Dr. John Axelopoulos, who told me, “You should buy a copy of Chung and take probability at main branch [Kent State University].” Also, I need to thank the late Dr. Vladimir Gurariy and Dr. Hassan Allouba, who taught the probability sequence, for their inspiration and encouragement.

I am grateful to Dr. Joe Diestel, Dr. Jing Li, Dr. Lothar Reichel and Dr. Eugene (Chuck) Gartland for their help and encouragement throughout my student career at Kent State University.

Finally, I would like to thank my family. I am very grateful for the support and encouragement that my parents, Dr. George Korte, MD AAOS (retired) and Christine Korte, have provided throughout my life. I would also like to thank my wife Julie for her support, and lastly, special thanks to Tenzing, Fritz and Willy.
CHAPTER 1

Introduction

1.1 Early Experiments

In the early twentieth century, especially from 1905 to 1910, Rutherford and Geiger, conducted a number experiments that counted the number of radioactive particles emitted by a substance. These fundamental experiments led to a number of advances, especially in developing the properties of the Poisson process and subsequent development of the subject of queueing and congestion theory. In 1910, Bateman showed "that if \( \alpha \) is the true average number of particles received by a phosphorescent screen in a given time, then the probability \( p \), that \( n \) particles will be observed in that time interval, is given by \( p = \frac{\alpha^n}{n!} e^{-\alpha} \) where \( n \) may be given the values 0, 1, 2, 3,..." (Makower and Geiger, 1912, also Bateman 1910).

Various scientific experiments have prompted mathematicians to consider discrete statistical experiments which have the following types of observations:

(i) censored observations,

(ii) truncated observations with known point of truncation,

(iii) truncated observations with unknown point of truncation.

A brief description of these types of data is presented in the following three subsections.

1.1.1 Censored Observations

In some early experiments conducted by Rutherford and Geiger, the particle counter had difficulty counting particles if there were too many of them. Below is a data set from Rutherford and Geiger published in 1910 (Moore 1952). In this experiment, the counter could detect the correct number of alpha particles during a time interval if the count was
less than or equal to 8. If there were more than 8 alpha particles, the counter would jam and the count became incorrect.

<table>
<thead>
<tr>
<th>i</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>&gt; 8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_i$</td>
<td>57</td>
<td>203</td>
<td>383</td>
<td>525</td>
<td>532</td>
<td>408</td>
<td>273</td>
<td>139</td>
<td>45</td>
<td>43</td>
<td>2608</td>
</tr>
</tbody>
</table>

Here $N_i$ stands for the number of time intervals during each of which $i$ alpha particles were counted. This later prompted the concept of censored observations.

The statistical nature of the issue is what is an estimate of $\theta$ in this context. It is clear that, even though the radio-active material is emitting particles by a Poisson distribution, the observed random variables are not Poisson due to technical restrictions.

This estimation aspect of the problem was first studied by Tippett in 1932. Tippett proposed using the $Poisson(\theta)$ distribution over this data set and using the maximum likelihood method and his nomograms to estimate the parameter. However, this requires the solution of a nonlinear equation and leads to extensive computations. Also, if there are more than four distinct frequencies, then one cannot use Tippett’s nomograms. Bliss (1948) gave approximation tables to the maximum-likelihood method to facilitate the computational process, but again the tables are only for cases involving four or fewer distinct frequencies.

In 1952, Moore eliminated this computational burden and gave an extremely simple way to get an estimator of $\theta$. Moore assumed that all the censored observations that fall in the set \{r + 1, r + 2, \cdots\}, for some positive integer $r$. Considering the fact that

$$\theta = \frac{\sum_{i=1}^{r} i e^{-\theta} \theta^{i} / i!}{\sum_{j=0}^{r-1} e^{-\theta} \theta^{j} / j!},$$

Moore took

$$\hat{\theta} = \frac{\sum_{i=1}^{r} i N_i}{\sum_{j=0}^{r-1} N_j},$$

(1.1)

as an estimate for $\theta$, where $N_i$ equal the number of observations in the sample that take the value $i$. Although the denominator can take the value zero with positive probability, the
probability of this happening may be quite small. When the sample size is reasonably large and the point of censoring does not miss the major bulk of the distribution, this division by zero will not take place.

Other than the MLE and Moore’s estimators, there are no other estimators available in the literature.

Later, we will consider this problem when viewed from a slightly more general context of censored samples from a power series distribution. We will give an alternative to the MLE and Moore’s estimators. The efficiency of these estimators will be provided. As we will see, these estimators can be sometimes quite inefficient. The problem still remains open as to what is an “optimal” estimator of $\theta$ in such situations? Does there exist an unbiased estimator of $\theta$?

1.1.2 Truncated Observations-I

In practice, there are distributions that are of a Poisson type process but zero group is not observed. Such distributions are called zero-truncated Poisson (usually written 0-truncated), and if $X$ is a zero-truncated Poisson random variable with parameter $\theta$, then

$$P(X = k) = \frac{e^{-\theta} \theta^k}{1 - e^{-\theta} k!}, \quad P(X = k) = \frac{1}{e^\theta - 1} \frac{\theta^k}{k!}, \quad k = 1, 2, \ldots .$$

An example from biology is counting the number of individuals in separate colonies of some species. By default, a colony must have at least one individual in it.

In 1952, F. N. David and N. L. Johnson were looking at the number of accidents per worker at a factory. For various reasons, every worker had at least one accident. This prompted them to examine the problem of estimating $\theta$ in the 0-truncated Poisson distribution. They proposed an estimator using the method of moments and compared its efficiency with respect to the maximum likelihood estimator. Once again, the computation of the MLE requires the solution of a nonlinear equation, and one needs to make a decision
as to use the MLE at the cost of computations or the estimator of David and Johnson at
the loss of some efficiency.

In 1953, Plackett showed the existence of an unbiased estimator of $\theta$ in the 0-truncated
Poisson distribution. Plackett showed that asymptotic relative efficiency of his estimator
compared to the MLE was very close to one. For $\theta \not\in (1, 5)$, his estimator’s efficiency is very
close to one. And for $\theta \in (1, 5)$ the efficiency does not go down below 0.95. The estimator
is computationally easy to calculate and gives a viable alternative to the MLE. The idea
behind Plackett’s estimator is also extremely simple. It is essentially taking Moore’s idea
and applying it to the 0-truncated Poisson distribution. That is, since

$$\theta = \frac{\sum_{i=2}^{\infty} i e^{-\theta} \theta^i / i!}{\sum_{j=1}^{\infty} e^{-\theta} \theta^j / j!},$$

estimate $\theta$ by

$$\hat{\theta} = \frac{\sum_{i=2}^{\infty} i N_i}{\sum_{j=1}^{\infty} N_j},$$  \hspace{1cm} (1.2)

where $N_i$ equal the number of observations in the sample that take the value $i$. In this case,
we can write this estimator in the simpler form

$$\hat{\theta} = \frac{1}{n} \sum_{i=1: Y_i \neq 1}^{n} Y_i = \frac{1}{n} (S_n - N_1),$$  \hspace{1cm} (1.3)

where $Y_i, i = 1, 2, \cdots, n$ form the random sample and $S_n = \sum_{i=1}^{n} Y_i$. Taking the expectation, it is an unbiased estimator of $\theta$.

Also in 1953, Rider provided an estimator based on the first two incomplete moments that compared favorably to the MLE.

In 1954, Cohen used tables and interpolation and iterative procedures using a set of
Poisson tables to give maximum likelihood estimators in truncated and censored forms of
the Poisson distribution.

In 1958, Tate and Goen were the first to prove that, in the 0-truncated and $\{0, 1, 2, \cdots, m\}$-
truncated Poisson case (called $m$-truncated case), there exists a uniformly minimum vari-
ance unbiased (UMVU) estimator of $\theta$. However, the calculation of the estimator requires
an $n$-fold Cauchy product of the series $a_k = \frac{1}{k^r}$ for $k = m, m+1, m+2, \cdots$. Again, there is the issue of the computational cost of the MLE or the UMVU versus the cost of loss of efficiency using Plackett's estimator.

Later, we will study the properties of such estimators in the context of power series distributions. In the context of power series distributions, such problems of truncation are absorbed into the sequence that generates the power series distribution. Hence, one does not need to include the prefix of truncation to such problems. The calculations of the estimators and their limiting behavior still pose problems that need to be resolved.

1.1.3 Truncated Observations-II

A different type of problem arises when the point of truncation is unknown. Klotz (1970) considered this problem for the geometric density. More precisely, let $X_1, X_2, \cdots, X_n$ form a random sample from the density

$$P(X = k) = (1-p)^{k-\nu}p, \quad k = \nu, \nu+1, \cdots,$$

where both $p$ and $\nu$ are unknown parameters. By using the theory of complete sufficient statistics, he gave uniformly minimum variance unbiased estimators of $\nu$ and $p$ and gave uniformly most powerful unbiased tests concerning these parameters.

Shortly afterwards, C. J. Park (1973) and C. Charalambides (1974) independently showed that the results of Klotz hold in the general context of power series distributions. Charalambides further considered their computational aspects and the links with the various types of number sequences that arise in combinatorial fields.

1.2 General Problem

The family of distributions that we will consider is defined as follows. Let $a_k, k = 0, 1, 2, \cdots$ be a fixed sequence of nonnegative numbers. Assume that the power series

$$G(\theta) = \sum_{k=0}^{\infty} a_k \theta^k$$
converges for all $\theta$ in the interval $[0, R)$, where $R$ is the radius of convergence. This gives rise to the power series distribution $(PSD(\theta))$

$$P(X = k) = \frac{a_k \theta^k}{G(\theta)}, \quad k = 0, 1, 2, \ldots.$$ 

Some questions of practical interest are:

- Let $X_1, X_2, \ldots, X_n \sim_{iid} PSD(\theta)$, how should one estimate $\theta^m$ for a given number $m$? Does there exist a UMVU estimator of $\theta^m$, and if so, what is its asymptotic distribution and the limiting variance? Does it achieve the Cramér-Rao lower bound?

- Let $A$ be a nonempty proper subset of $\{0, 1, 2, \ldots\}$. Let $X_1, X_2, \ldots, X_n \sim_{iid} PSD(\theta)$ such that all those observations that fall in the set $A$ get censored. Again, how should one estimate $\theta^m$ for a given number $m$ and what are the limiting properties of the estimator as the sample size gets large?

- Let $A$ be a nonempty proper subset of $\{0, 1, 2, \ldots\}$ such that $A$ is the set of consecutive numbers $\{0, 1, 2, \ldots, \nu\}$ or the right tail $\{\nu, \nu + 1, \ldots\}$. Let $PSD(\theta)$ be a power series distribution generated by the sequence $\{a_k\}$. Define a new sequence $c_k = a_k$ if $k \notin A$ and $c_k = 0$ for $k \in A$. The resulting power series distribution generated by $\{c_k\}$ is the truncated power series distribution of $\{a_k\}$. If the point $\nu$ is an unknown parameter, how should one estimate $\gamma = (\nu, \theta)$? Does there exist a UMVU estimator and if so what is its variance? Once again, what are the limiting properties of the estimator?

### 1.3 Examples of Power Series Distributions

Let $a_k \geq 0$, $k \in \mathbb{N}_0 = \{0\} \cup \mathbb{N} = \{0, 1, 2, \ldots\}$, be a sequence so that

$$G(\theta) := \sum_{k=0}^{\infty} a_k \theta^k < \infty; \quad \text{for } \theta \in [0, R),$$
for some $R \in (0, \infty]$. This generates a corresponding power series distribution, defined by

$$
P(X_{\theta} = k) = \frac{a_k \theta^k}{G(\theta)}; \quad k = 0, 1, 2, \cdots, \theta \in [0, R).
$$

Every power series distribution has a well defined moment generating function:

$$
\mathbb{E}(e^{tX_1}) = \frac{1}{G(\theta)} \sum_{k=0}^{\infty} e^{tk} a_k \theta^k = \frac{G(e^t \theta)}{G(\theta)},
$$

provided

$$
e^t \theta < R, \quad \text{i.e.,} \quad t < \log(R/\theta).
$$

It is also not difficult to show that the family of power series distributions belongs to the one-parameter exponential family. The mean and variance of $PSD(\theta)$ is

$$
\mu = \frac{\theta G'(\theta)}{G(\theta)}, \quad \sigma^2(\theta) = \frac{\theta G'(\theta)G(\theta) + \theta^2 G''(\theta)G(\theta) - \theta^2 (G'(\theta))^2}{(G(\theta))^2}
$$

The following are some common examples of PSDs.

**Example. (Poisson)** If $a_k = 1/k!$ for all $k \in \mathbb{N}_0$ then the resulting power series random variable $X_{\theta}$ has the Poisson($\theta$) distribution. If $X_1, X_2, \cdots, X_n \overset{iid}{\sim} X_{\theta}$ then $S_n = X_1 + \cdots + X_n$ is a power series random variable with distribution $S_n \sim \text{Poisson}(n\theta)$.

**Example. (Geometric)** If $a_k \equiv 1$ for all $k \in \mathbb{N}_0$ then $X_{\theta} \sim \text{Geometric}(1 - \theta)$. If $X_1, X_2, \cdots, X_n \overset{iid}{\sim} X_{\theta}$ then $S_n = X_1 + \cdots + X_n$ power series random variable with distribution $S_n \sim \text{N.B.}(n, 1 - \theta)$.

**Example. (Binomial)** Let $m \in \mathbb{N}$ be fixed and let $a_k = \binom{m}{k}$ for $k = 0, 1, 2, \cdots, m$, then $X_{\theta} \sim B(m, p)$ where $p = \theta/(1 + \theta)$. In this context, $\theta$ represents the usual odds ratio parameter, $\theta = p/(1 - p)$. If $X_1, X_2, \cdots, X_n \overset{iid}{\sim} X_{\theta}$ then $S_n = X_1 + \cdots + X_n$ is also a binomial power series random variable with distribution $S_n \sim B(mn, p)$. 
Example. (Zero-truncated Poisson) The zero-truncated Poisson family is one of the most commonly used examples of a nontrivial power series distribution. In this case, $G(\theta) = e^\theta - 1$, so $a_0 = 0$ and $a_k = 1/k!$ for $k \in \mathbb{N}$. The resulting density is
\[
P(X_\theta = k) = \frac{1}{e^\theta - 1} \frac{\theta^k}{k!}, \quad k = 1, 2, \ldots.
\]
If $X_1, X_2, \ldots, X_n \sim X_\theta$ then $S_n = X_1 + \cdots + X_n$ is a power series random variable generated by the series expansion of $(e^\theta - 1)^n$.

Example. (Left-truncated Poisson) The left-truncated Poisson family is a generalization of the zero-truncated Poisson. In the general case, $a_k = 0$ for all $k \in \{0, 1, 2, \ldots, \kappa - 1\}$ and $a_k = \frac{1}{k!}$ for $k \in \{\kappa, \kappa + 1, \ldots\}$. So
\[
G(\theta) = e^\theta - 1 - \theta - \frac{\theta^2}{2!} - \cdots - \frac{\theta^{\kappa-1}}{(\kappa-1)!} = \sum_{k=\kappa}^{\infty} \frac{\theta^k}{k!}
\]
and
\[
P(X_\theta = k) = \left( \sum_{k=\kappa}^{\infty} \frac{\theta^k}{k!} \right)^{-1} \frac{\theta^k}{k!}, \quad k = \kappa, \kappa + 1, \ldots.
\]
If $X_1, X_2, \ldots, X_n \sim X_\theta$ then $S_n = X_1 + \cdots + X_n$ is a power series random variable generated by the series expansion of $\left( \sum_{k=\kappa}^{\infty} \frac{\theta^k}{k!} \right)^n$.

Example. (Truncated Binomial) For a fixed $m \in \mathbb{N}$, let $a_0 = 0$ and let $a_k = \binom{m}{k}$ for $k = 1, 2, \ldots, m$. Then
\[
G_m(\theta) = \sum_{k=1}^{m} \theta^k \binom{n}{k} = (1 + \theta)^m - 1.
\]
Let $p = \theta/(1 + \theta)$ so that $1 - p = 1/(1 + \theta)$ or $1 + \theta = (1 - p)^{-1}$ and $\theta = p/(1 - p)$. The generated power series distribution is
\[
P(X = k) = \frac{\binom{m}{k} \theta^k}{(1 + \theta)^m - 1}
= \frac{1}{(1 - p)^{-m} - 1} \binom{m}{k} \left( \frac{p}{1 - p} \right)^k, \quad k = 1, 2, \ldots, m,
\]
\[ \frac{1}{(1-p)^m - 1} \binom{m}{k} p^k (1-p)^{-k}, \quad k = 1, 2, \ldots, m, \]

\[ \frac{1}{1 - (1-p)^m} \binom{m}{k} p^k (1-p)^{-k}, \quad k = 1, 2, \ldots, m. \]

Hence, \( X_\theta \) gives rise to the truncated binomial distribution, denoted by \( T.B.(m,p) \) where \( p = \theta/(1+\theta) \) is the odds ratio.

Note that if \( m = 1 \), then \( X_\theta = 1 \) for sure, so it becomes a degenerate random variable. Thus, in the 0-truncated binomial setup, \( m \) should be at least 2 to get a nontrivial distribution.

Also, if \( X_1, X_2, \ldots, X_n \overset{iid}{\sim} X_\theta \) then \( S_n = X_1 + \cdots + X_n \) is a power series random variable generated by the series expansion of

\[ (G_m(\theta))^n = ((1 + \theta)^m - 1)^n, \]

which is not going to be of the same form as that of \( X_\theta \), for \( n \geq 2 \).
CHAPTER 2

Estimators and Their Shortcomings

This chapter presents some previous results for inference in power series distributions and mentions some other important results. It is important to note that, unless specifically stated otherwise, the term "truncated" refers to a known point of left truncation.

2.1 Maximum Likelihood (ML) and Method of Moments (MM) Estimation Results

2.1.1 Maximum Likelihood Estimation of $\theta$ in the truncated and censored Poisson distribution

Most papers on estimation in PSD have focused on maximum likelihood estimation of $\theta$ in the truncated and censored $\text{Poisson}(\theta)$ distribution cases.

- Tippett (1932) discussed the estimation problem in the censored $\text{Poisson}(\theta)$ case from maximum likelihood point of view and presented nomograms to aid in computation.

- David and Johnson (1952) studied the truncated $\text{Poisson}(\theta)$ case from maximum likelihood point of view.

- Moore (1952) considered the censored $\text{Poisson}(\theta)$ case from an ad hoc estimation point of view.

- Moore (1954) gave some estimation procedures which are applicable from both censored and truncation points of view for the $\text{Poisson}(\theta)$ and binomial distributions.

- Rider (1953) gave a maximum likelihood based method of estimation in the zero-truncated and zero-one-truncated $\text{Poisson}(\theta)$ cases. He provided an estimator based
on the first two incomplete moments which proved better than the maximum likelihood estimator in his limited number of experiments.

- Cohen (1954) studied the properties of estimators in the censored and truncated $\text{Poisson}(\theta)$ case from the point of view of maximum likelihood estimation. His main thrust is to show how to calculate these MLE's by using iterative or interpolation procedures using a set of tables of Poisson distribution.

2.1.2 MM Estimation of $\theta G''(\theta)/G'(\theta)$

In 1952, David and Johnson considered the 0-truncated $\text{Poisson}(\theta)$ distribution. Specifically, they considered the case when $X_1, X_2, \cdots, X_n \overset{iid}{\sim} X$, where

$$P(X = k) = \frac{1}{e^\theta - 1} \frac{\theta^k}{k!}, \quad k = 1, 2, \cdots.$$  

Their main interest was the maximum likelihood estimator of $\theta$, but they also considered a method of moments type estimator. Their reasoning was

$$\mathbb{E}(X_1^2) = \sum_{k=1}^{\infty} k^2 \frac{1}{e^\theta - 1} \frac{\theta^k}{k!}$$

$$= \frac{e^\theta}{e^\theta - 1} \mathbb{E}(Z^2), \quad \text{where } Z \sim \text{Poisson}(\theta),$$

$$= \frac{1}{1 - e^{-\theta}} (\text{Var}(Z) + \mathbb{E}(Z^2))$$

$$= \frac{1}{1 - e^{-\theta}} (\theta + \theta^2).$$

Similarly,

$$\mathbb{E}(X_1) = \frac{1}{1 - e^{-\theta}} \mathbb{E}(Z) = \frac{\theta}{1 - e^{-\theta}}.$$  

So that

$$\frac{\mathbb{E}(X_1^2)}{\mathbb{E}(X_1)} = 1 - \frac{\theta + \theta^2}{\theta} = 1 - \theta.$$  

Hence, they proposed the method of moments estimator for $\theta$ to be

$$\tilde{\theta} = \frac{\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} X_i} - 1.$$
They used some heuristic arguments to show that the asymptotic variance of this estimator is

$$n \text{Var}(\tilde{\theta} + 1) \sim (\theta + 2)(1 - e^{-\theta}).$$

And then, by using the Cramér-Rao lower bound (Fisher information) they state the asymptotic variance of the maximum likelihood estimator is

$$\frac{\theta(1 - e^{-\theta})^2}{1 - e^{-\theta} - \theta e^{-\theta}}.$$

They then establish that the asymptotic relative efficiency of their method of moments estimator compared to the maximum likelihood estimator is

$$A.R.E.(\tilde{\theta}, \hat{\theta}) = \frac{\theta}{(\theta + 2)(1 - \theta/(e^\theta - 1))},$$

and state that this function never goes below 0.70 and the minimum is achieved somewhere between 2 and 4. Due to this reason and computational ease of their method of moments estimator, their estimator has some merit for practical use. They also provide a table to simplify the computations involved in calculating the maximum likelihood estimator.

More generally, the David-Johnson method of moments estimator can be applied to estimating $g(\theta) := \theta G''(\theta)/G'(\theta)$. In some special cases, this function reduces to usual estimation problems. For instance, when the parent distribution is Poisson($\theta$), $G(\theta) = e^\theta$ and $g(\theta) = \theta$. When the parent distribution is 0-truncated Poisson($\theta$) then $G(\theta) = e^\theta - 1$ and, $g(\theta) = \theta$. When we are dealing with $\{0,1\}$-truncated Poisson($\theta$) then $G(\theta) = e^\theta - 1 - \theta$ and hence, $g(\theta) = \theta e^\theta/(e^\theta - 1)$. When $a_k = \binom{m}{k}$, and $m \geq 2$, then the parent distribution is $B(m, \theta/(1 + \theta))$ and $\theta$ is the odds ratio and $G(\theta) = (1 + \theta)^m$ so,

$$g(\theta) = \frac{\theta m(m-1)(1 + \theta)^{m-2}}{m(1 + \theta)^{m-1}} = (m - 1)\frac{\theta}{1 + \theta}.$$

Thus, it is essentially estimating the probability of heads in a coin toss.

Furthermore, function $g(\theta) = \frac{\theta G''(\theta)}{G'(\theta)}$ is related to some moments of any $X \sim PSD(\theta)$:

$$E(X^2) = \frac{\theta^2 G''(\theta) + \theta G'(\theta)}{G(\theta)},$$
\[ \mathbb{E}(X) = \theta G'(\theta) / G(\theta), \]

so

\[ \frac{\mathbb{E}(X^2)}{\mathbb{E}(X)} - 1 = \frac{\theta G''(\theta) + G'(\theta)}{G'(\theta)} - 1 = \frac{\theta G''(\theta)}{G'(\theta)} = g(\theta). \]

**Proposition 1.** The MM estimator \( \hat{g} \) of \( g(\theta) \) is asymptotically normal. That is,

\[ \sqrt{n}(\hat{g} - g(\theta)) \xrightarrow{\text{dist}} \mathcal{N}(0, \xi^2(\theta)), \]

where

\[ \xi^2(\theta) = \frac{1}{(\mathbb{E}(X_1))^2} \left\{ \mathbb{E}(X_1^4) \mathbb{E}(X_1)^2 + (\mathbb{E}(X_1^2))^3 - 2\mathbb{E}(X_1)\mathbb{E}(X_1^2)\mathbb{E}(X_1^3) \right\}. \]

**Proof:** The delta method provides a brief proof. Let \( g(\theta) \) be denoted by \( g \) to save space. Let \( X_1, X_2, ..., X_n \) be iid \( PSD(\theta) \). Slutsky’s theorem and the law of large numbers gives

\[ \sqrt{n}(\hat{g} - g(\theta)) = \sqrt{n} \left( \frac{\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} X_i} - 1 - g \right) \]

\[ = \frac{1}{\sum_{i=1}^{n} X_i} \sqrt{n} \left( \sum_{i=1}^{n} X_i^2 - (1 + g) \sum_{i=1}^{n} X_i \right) \]

\[ = \sqrt{n} \left( \frac{\sum_{i=1}^{n} X_i^2 - (1 + g) \sum_{i=1}^{n} X_i}{\mathbb{E}(X_1)^2} \right) \]

\[ \sim \frac{1}{\mathbb{E}(X_1) \sqrt{n}} \left( \sum_{i=1}^{n} (X_i^2 - (1 + g) X_i) \right) \]

\[ = \frac{G(\theta)}{\theta G'(\theta)} \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} (X_i^2 - (1 + g) X_i) \right). \]

Applying the central limit theorem, we have

\[ \mathbb{E}(X_1^2 - (1 + g) X_1) = \mathbb{E}(X_1^2) - (1 + g) \mathbb{E}(X_1) \]

\[ = \mathbb{E}(X_1^2) - \left( 1 + \frac{\mathbb{E}(X_1^2)}{\mathbb{E}(X_1)} - 1 \right) \mathbb{E}(X_1) = 0. \]

And

\[ \text{Var}(X_1^2 - (1 + g) X_1) = \text{Var}(X_1^2) + (1 + g)^2 \text{Var}(X_1) - 2(1 + g) \text{Cov}(X_1^2, X_1) \]
\[ \begin{align*}
\mathbb{E}(X_1^4) &= \mathbb{E}(X_1^2)^2 + (1 + g)^2 \left\{ \mathbb{E}(X_1^2) - (\mathbb{E}(X_1))^2 \right\} \\
&\quad - 2(1 + g) \left\{ \mathbb{E}(X_1^2) - \mathbb{E}(X_1) \mathbb{E}(X_1) \right\} \\
&= \mathbb{E}(X_1^2)^2 + \frac{(\mathbb{E}(X_1^2))^2}{(\mathbb{E}(X_1))^2} \left\{ \mathbb{E}(X_1^2) - (\mathbb{E}(X_1))^2 \right\} \\
&\quad - 2 \frac{\mathbb{E}(X_1^2)}{\mathbb{E}(X_1)} \left\{ \mathbb{E}(X_1^2) - \mathbb{E}(X_1) \mathbb{E}(X_1) \right\} \\
&= \mathbb{E}(X_1^2) + \frac{(\mathbb{E}(X_1^2))^3}{(\mathbb{E}(X_1))^2} - 2 \frac{\mathbb{E}(X_1^2) \mathbb{E}(X_1^2)}{\mathbb{E}(X_1)} \\
&\quad = \frac{1}{(\mathbb{E}(X_1))^2} \left\{ \mathbb{E}(X_1^4) \left( \mathbb{E}(X_1^2) \right)^2 + (\mathbb{E}(X_1^2))^3 - 2 \mathbb{E}(X_1) \mathbb{E}(X_1^2) \mathbb{E}(X_1^3) \right\}
\end{align*} \]

To compute the moments of the \(PSD(\theta)\) distribution, differentiating the moment generating function

\[ \phi(t) = \frac{G(\theta e^t)}{G(\theta)}, \]

gives

\[ \begin{align*}
\phi'(t) &= \frac{\theta e^t G'(\theta e^t)}{G(\theta)}, \\
\phi''(t) &= \frac{\theta e^t G''(\theta e^t) + e^t G'(\theta e^t)}{G(\theta)}, \\
\phi'''(t) &= \frac{\theta^3 e^{3t} G'''(\theta e^t) + 3\theta^2 e^{2t} G''(\theta e^t) + 3\theta e^t G'(\theta e^t)}{G(\theta)},
\end{align*} \]

and

\[ \phi'''(t) = \frac{\theta^4 e^{4t} G''''(\theta e^t) + 6\theta^3 e^{3t} G'''(\theta e^t) + 7\theta^2 e^{2t} G''(\theta e^t) + 7\theta e^t G'(\theta e^t)}{G(\theta)}. \]

Evaluating these expressions at \(t = 0\) we get

\[ \begin{align*}
\mathbb{E}(X_1) &= \frac{\theta G'(\theta)}{G(\theta)}, \\
\mathbb{E}(X_1^2) &= \frac{\theta^2 G''(\theta) + \theta G'(\theta)}{G(\theta)}, \\
\mathbb{E}(X_1^3) &= \frac{\theta^3 G'''(\theta) + 3\theta^2 G''(\theta) + 3\theta G'(\theta)}{G(\theta)}, \\
\mathbb{E}(X_1^4) &= \frac{\theta^4 G'''(\theta) + 6\theta^3 G''(\theta) + 7\theta^2 G'(\theta) + 7\theta G'(\theta)}{G(\theta)}.
\end{align*} \]

Substituting these expressions into the equation above, we get the asymptotic variance.
2.2 Unbiased Estimators

2.2.1 Unbiased Estimation of $P(Z = 0)$

Let $Z$ be a nonnegative integer valued random variable and let $P(Z = k) = a_k \theta^k / G(\theta)$ where $a_k > 0$ for all $k = 0, 1, 2, \ldots$, and suppose we truncate the distribution at zero. Let $X$ be a random variable such that

$$P(X = k) = \frac{1}{P(Z \neq 0) G(\theta)} a_k \theta^k, \quad k = 1, 2, \ldots.$$

**Proposition 2.** Unbiased estimators of $g(\theta) = P(Z = 0) = \frac{a_0}{G(\theta)}$ exist.

**Proof.** We need to show that there exist constants $h_k$, $k = 1, 2, \ldots$ with the property that

$$\sum_{k=1}^{\infty} h_k \frac{a_k \theta^k}{P(Z \neq 0) G(\theta)} = P(Z = 0), \quad \text{for each } \theta \in (0, R)$$

which is equivalent to asking that

$$\sum_{k=1}^{\infty} h_k a_k \theta^k = P(Z \neq 0) G(\theta) \frac{a_0}{G(\theta)}, \quad \text{for each } \theta \in (0, R).$$

$$= P(Z \neq 0) a_0$$

$$= a_0(1 - P(Z = 0))$$

$$= a_0(G(\theta) - a_0)$$

Now since $a_0 > 0$, we can find a sequence $b_k$ so that

$$\frac{1}{G(\theta)} = \sum_{j=0}^{\infty} b_j \theta^j.$$

Hence, we may write

$$\frac{(G(\theta) - a_0)}{G(\theta)} = \left( \sum_{i=1}^{\infty} a_i \theta^i \right) \left( \sum_{j=0}^{\infty} b_j \theta^j \right).$$

This is a product of two generating functions. The first one is the sequence $c = (0, a_1, a_2, \ldots)$ and the second one is the $b$ sequence. The product of the two generating functions is the generating function of the sequence $c \ast b$, so it must be that

$$\sum_{k=1}^{\infty} h_k a_k \theta^k = a_0 \sum_{k=1}^{\infty} (c \ast b)_k \theta^k.$$
and 
\[ h_k = \frac{a_0(c \ast b)_k}{a_k}, \quad k \geq 1. \]

Hence, we see that unbiased estimators of \( \mathbb{P}(Z = 0) \) can exist. \(\square\)

2.2.2 Unbiased Estimation of \( g(\theta) = \theta^m \)

Let 
\[ G(\theta) = a_0 + a_1 \theta + a_2 \theta^2 + ... \]

be a function such that:

1) There exists \( R > 0 \) such that \( G(\theta) \) converges on \((-R, R)\)

2) There exists known \( \kappa \in \mathbb{N}_0 \) and \( N \in \mathbb{N} \cup \{\infty\} \) such that \( a_k > 0, \forall k \in I = \{\kappa, \kappa + 1, \ldots, N\} \) and \( a_k = 0 \) for \( k \notin I \).

For \( m \in \mathbb{N} \), we define the estimator \( h_m(X) = \frac{a_{X-m}}{a_X} x \in \mathbb{N}_0 \) of \( \theta^m \). When \( m = 1 \), we will drop the subscript \( m \). Throughout this paper, when \( \frac{a_{k-m}}{a_k} = \frac{0}{0} \) and \( k < \kappa \), we will take \( \frac{a_{k-m}}{a_k} = 0 \). This is justified by Taylor’s theorem and repeated applications of L’Hôpital’s rule, and the fact that \( a_k > 0, \forall k \in I \) and \( a_k = 0, \forall k < \kappa \).

First, by Taylor’s theorem, we have 
\[ G(\theta) = a_0 + a_1 \theta + a_2 \theta^2 + ... = G(0) + \frac{G^{(1)}(0)}{1!} \theta + \frac{G^{(2)}(0)}{2!} \theta^2 + ..., \]

so 
\[ a_k = \frac{G^{(k)}(0)}{k!}, \]

and 
\[ \mathbb{P}(X = k) = \frac{a_k}{G(\theta)} \theta^k = \frac{G^{(k)}(0)}{k! \theta^k}. \]

Trivially, \( a_k = 0 \iff G^{(k)}(0) = 0 \), so \( G^{(k)}(0) > 0 \) for all \( k \geq \kappa \) and \( G^{(k)}(0) = 0 \) for all \( k < \kappa \).

If \( \frac{a_{k-m}}{a_k} = \frac{0}{0} \), then we have 
\[ \frac{0}{0} = \frac{a_{k-m}}{a_k} \]
\[
= \frac{G^{(k-m)}(0)/(k-m)!}{G^{(k)}(0)/k!}
= \frac{k! G^{(k-m)}(0)}{(k-m)! G^{(k)}(0)}.
\]

Taking the limit as \( \theta \to 0 \) of \( \frac{k!}{(k-m)!} \frac{G^{(k-m)}(\theta)}{G^{(k)}(\theta)} \) we have:

\[
\lim_{\theta \to 0} \frac{k!}{(k-m)!} \frac{G^{(k-m)}(\theta)}{G^{(k)}(\theta)} = \frac{k!}{(k-m)!} \lim_{\theta \to 0} \left( \frac{G^{(k-m)}(\theta)}{G^{(k)}(\theta)} \right)'
= \frac{k!}{(k-m)!} \lim_{\theta \to 0} \left( \frac{G^{(k-m)}(\theta)}{G^{(k)}(\theta)} \right)''
= \frac{k!}{(k-m)!} \lim_{\theta \to 0} \left( \frac{G^{(k-m)}(\theta)}{G^{(k)}(\theta)} \right)'''
= \ldots
= \frac{k!}{(k-m)!} \lim_{\theta \to 0} \left( \frac{G(\theta)^{(k-m)}}{G^{(k)}(\theta)} \right)^{(k-k)}
= \frac{k!}{(k-m)!} \lim_{\theta \to 0} \frac{G^{(k-m)}(\theta)}{G^{(k)}(\theta)}
= \frac{k!}{(k-m)!} \frac{G^{(k-m)}(0)}{G^{(k)}(0)}
= \frac{k!}{(k-m)!} \frac{0}{\kappa!a_{\kappa}} = 0.
\]

Previous Results

Tukey (1949) shows that truncation at a known point does not destroy the sufficiency of a statistic. However, if the point of truncation is unknown, then obviously the sufficiency of the statistic may be destroyed.

Tate and Goen (1958) gives a detailed analysis of minimum variance unbiased estimation of the parameter in the truncated Poisson distribution. They show that there is no unbiased estimator whose variance attains the Cramér-Rao lower bound in the truncated Poisson case. They also give upper and lower bounds for the variance of the UMVU estimator based on Plackett's unbiased estimator and the Fisher information inequality. However, they do not give any closed form expression for the actual variance of the UMVU estimator, nor do they
state what the limiting or asymptotic variance of the minimum variance unbiased estimator will be. The paper does provide a long discussion as to how one can calculate the estimator quickly, and it provides some tables for this purpose.

Lehmann (1983) shows that if \( a_k > 0 \) for all \( k = 0, 1, 2, \ldots \), then the UMVU estimator of \( \theta \) is

\[
h(X) = \begin{cases} 
0 & \text{if } X = 0 \\
\frac{a(X-1)}{a(X)} & \text{if } X = 1, 2, \ldots 
\end{cases}
\]

where \( a(j) := a_j \).

Later we will consider the same estimator in other situations when some of the \( a_k \) are zero.

Basic Properties

**Proposition 3.** When \( a_k > 0 \) for \( k = \kappa, \kappa + 1, \ldots, N \), the expectation of \( h(X) \) is

\[
\theta - \frac{a(\kappa + N)\theta^{\kappa+N+1}}{G(\theta)} = \theta (1 - P_\theta(X = \kappa + N)),
\]

and \( P_\theta(X = \kappa + N) \) is taken to be zero if \( N = \infty \).

**Proof.** Just note that

\[
\mathbb{E}_\theta \{ h(X) \} = 0 + \sum_{k=\kappa}^{\kappa+N} \frac{a_{k-1}}{a_k} \frac{a_k \theta^k}{G(\theta)} + 0
\]

\[
= \theta \sum_{k=\kappa+1}^{\kappa+N} \frac{a_{k-1} \theta^{k-1}}{G(\theta)}
\]

\[
= \theta \sum_{j=\kappa}^{\kappa+N-1} \frac{a_j \theta^j}{G(\theta)}
\]

This gives the result. \( \square \)

**Remark.** We need the second moment of this distribution in order to find the mean squared error of the estimator. The second moment can be written as

\[
E_\theta \{ h^2(X) \} = 0 + \sum_{k=\kappa}^{\kappa+N} \left( \frac{a_{k-1}}{a_k} \right)^2 \frac{a_k \theta^k}{G(\theta)} + 0
\]
\[
\begin{align*}
\theta^{k+N} \sum_{k=\kappa+1} a_{k-1} \theta^{k-1} \\
= \theta \sum_{j=\kappa} a_j \frac{a_j \theta^j}{G(\theta)}
\end{align*}
\]

Since, we are interested in the variance after selecting a sample of size \( n \), the \( a_j \) will depend on both \( j \) and \( n \) and it is with respect to the sample size \( n \) that we would like to find the rate of decrease of the variance.

**Proposition 4.** If \( a_k = 0 \) for \( k = 0, 1, \cdots, \kappa - 1 \) and \( a_k > 0 \) for all \( k = \kappa, \kappa + 1, \cdots \), then the estimator, \( h(X) \), is the only unbiased estimator for \( \theta \in (0, R) \) (and hence UMVU).

**Proof.** First note that

\[
\mathbb{E}_\theta(h(X)) = \sum_{k=\kappa+1} a_k \theta^k G(\theta)
\]

\[
= \sum_{j=\kappa} a_j \theta^j G(\theta)
\]

\[
= \theta \sum_{j=\kappa} a_j \theta^j G(\theta) = \theta.
\]

This shows that \( h(X) \) is an unbiased estimator of \( \theta \). To show that it is the only one, suppose that \( H(X) \) is unbiased estimator of \( \theta \). Then \( H(X) \) satisfies

\[
\mathbb{E}_\theta(H(X)) = \sum_{k=\kappa} H(k) \frac{a_k \theta^k}{G(\theta)} = \theta, \quad \theta \in (0, R).
\]

This implies that

\[
\sum_{k=\kappa} H(k) a_k \theta^k = \theta \sum_{k=\kappa} a_k \theta^k.
\]

Writing out the sums in expanded form, we have

\[
H(\kappa) a_\kappa \theta^\kappa + H(\kappa + 1) a_{\kappa+1} \theta^{\kappa+1} + \cdots = \theta a_\kappa \theta^\kappa + \theta a_{\kappa+1} \theta^{\kappa+1} + \cdots.
\]

Dividing by \( \theta^\kappa \) on both sides and letting \( \theta \to 0 \), gives

\[
H(\kappa) a_\kappa = 0.
\]
Since, $a_\kappa > 0$, it must be that $H(\kappa) = 0$. This then leads to the conclusion that

$$H(\kappa + 1) a_{\kappa+1} \theta^{\kappa+1} + H(\kappa + 2) a_{\kappa+2} \theta^{\kappa+2} + \cdots = \theta a_\kappa \theta^\kappa + \theta a_{\kappa+1} \theta^{\kappa+1} + \cdots.$$  

Dividing by $\theta^{\kappa+1}$ on both sides and letting $\theta \to 0$, gives

$$H(\kappa + 1) a_{\kappa+1} = a_\kappa.$$  

Hence,

$$H(\kappa + 1) = \frac{a_\kappa}{a_{\kappa+1}}.$$  

Plugging this back, we get

$$a_\kappa \theta^{\kappa+1} + H(\kappa + 2) a_{\kappa+2} \theta^{\kappa+2} + \cdots = a_\kappa \theta^{\kappa+1} + \theta a_{\kappa+1} \theta^{\kappa+1} + \cdots.$$  

Cancelling the first term on both sides, we have

$$H(\kappa + 2) a_{\kappa+2} \theta^{\kappa+2} + H(\kappa + 3) a_{\kappa+3} \theta^{\kappa+3} + \cdots = a_{\kappa+1} \theta^{\kappa+2} + a_{\kappa+2} \theta^{\kappa+3} + \cdots.$$  

Dividing by $\theta^{\kappa+2}$ on both sides and then letting $\theta \to 0$, gives that

$$H(\kappa + 2) a_{\kappa+2} = a_{\kappa+1}.$$  

Hence,

$$H(\kappa + 2) = \frac{a_{\kappa+1}}{a_{\kappa+2}}.$$  

Repeating the process concludes the proof. \qed

Remark. The UMVU estimator of $\theta$ exists if and only if for some $\kappa$, $a_k \geq 0$ for all $k \geq \kappa$ and $a_k = 0$ for all $k < \kappa$. The UMVU estimator is, by default, the best unbiased estimator. It does not tell us how much variance it has. Later we broaden the scope by asking what is the variance of $h(X)$ even when it is not unbiased.
Moments of $h_m (X)$

Let

$$G(\theta) = a_0 + a_1 \theta + a_2 \theta^2 + ...$$  \hfill (2.1)

be a function such that:

1) There exists $R > 0$ such that $G(\theta)$ converges on $(-R, R)$

2) There exists a minimum, known $\kappa \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ such that $a_k > 0$, $\forall k \geq \kappa$ and

$a_k = 0$, $\forall k < \kappa$.

**Proposition 5.** If $R < \infty$, then all moments of the estimator $h_m (X)$ are finite.

*Proof.* First, if $m = 1$, applying the ratio test to

$$\mathbb{E}(h^r(X)) = \sum_{k=0}^{\infty} (h(k))^r \frac{a_k}{G(\theta)} \theta^k$$

gives:

$$\lim_{k \to \infty} \frac{\left( \frac{a_k}{a_{k+1}} \right)^r \frac{a_{k+1}}{G(\theta)} \theta^{k+1}}{\left( \frac{a_{k-1}}{a_k} \right)^r \frac{a_k}{G(\theta)} \theta^k}$$

$$= \lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \right)^r \frac{a_k}{a_{k-1}} \frac{a_{k+1}}{a_k} \theta^{r+1} \left( \frac{a_k}{a_{k+1}} \right)^r \frac{a_k}{a_{k-1}} \theta^{r}$$

$$= \lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \right)^{r-1} \frac{a_k}{a_{k+1}} \left( \frac{a_k}{a_{k-1}} \right)^{r-1} \frac{a_k}{a_{k-1}} \theta.$$  

Since

$$0 < R = \frac{1}{\lim_{k \to \infty} \frac{a_{k+1}}{a_k}} < \infty,$$

applying the limit laws for sequences gives

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \right)^{r-1} \frac{a_k}{a_{k+1}} \left( \frac{a_k}{a_{k-1}} \right)^{r-1} \frac{a_k}{a_{k-1}} \theta$$
\[
\theta \cdot \lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \right)^{r-1} \times \lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \right)^{r-1} \times \lim_{k \to \infty} \left( \frac{a_k}{a_{k-1}} \right)^{r-1} = \theta \cdot \frac{1}{R^r}
\]

which gives convergence for \( \theta < R \), and \( \mathbb{E}(h^r(X)) \) converges for the same parameter space as the original density.

If \( m > 1 \),
\[
\mathbb{E}(h_m^r(X)) = \sum_{k=0}^{\infty} (h_m(k))^r \frac{a_k}{G(\theta)} \theta^k
\]

\[
= \sum_{k \geq \kappa + m}^{\infty} \left( \frac{a_k}{a_{k-m}} \right)^r \frac{a_k}{G(\theta)} \theta^k.
\]

Applying the ratio test gives
\[
\lim_{k \to \infty} \left( \frac{a_{k+1-m}}{a_{k+1}} \right)^r \frac{a_{k+1}}{G(\theta)} \theta^k = \lim_{k \to \infty} \left( \frac{a_{k+1-m}}{a_{k+1}} \right)^r \frac{a_k}{a_{k-m}} \theta.
\]

Now
\[
\frac{a_k}{a_{k-m}} = \left( \frac{a_k}{a_{k-1}} \right) \left( \frac{a_{k-1}}{a_{k-2}} \right) \cdots \left( \frac{a_{k-m+1}}{a_{k-m}} \right),
\]
and
\[
\frac{a_{k+1-m}}{a_{k+1}} = \frac{a_{k+1-m}}{a_{k+2-m}} \frac{a_{k+2-m}}{a_{k+3-m}} \cdots \frac{a_{k+1}}{a_{k+1}}.
\]
and
\[
0 < \lim_{k \to \infty} \frac{a_{k+1}}{a_k} = \frac{1}{R} < \infty,
\]
so
\[
\lim_{k \to \infty} \left( \frac{a_k}{a_{k-m}} \right) = \lim_{k \to \infty} \left( \frac{a_k}{a_{k-1}} \right) \left( \frac{a_{k-1}}{a_{k-2}} \right) \cdots \left( \frac{a_{k-m+1}}{a_{k-m}} \right)
\]
\[
= \lim_{k \to \infty} \frac{a_k}{a_{k-1}} \cdot \lim_{k \to \infty} \frac{a_{k-1}}{a_{k-2}} \cdot \cdots \cdot \lim_{k \to \infty} \frac{a_{k-m+1}}{a_{k-m}} = \left( \frac{1}{R} \right)^m,
\]
and
\[
\lim_{k \to \infty} \left( \frac{a_{k+1-m}}{a_{k+1}} \right) = \lim_{k \to \infty} \left( \frac{a_{k+1-m}}{a_{k+2-m}} \right) \left( \frac{a_{k+2-m}}{a_{k+3-m}} \right) \cdots \left( \frac{a_k}{a_{k+1}} \right)
\]
\[
= \lim_{k \to \infty} \frac{a_{k+1-m}}{a_{k+2-m}} \cdot \lim_{k \to \infty} \frac{a_{k+2-m}}{a_{k+3-m}} \cdot \ldots \cdot \lim_{k \to \infty} \frac{a_k}{a_{k+1}} = (R)^m
\]

Consequently,
\[
\lim_{k \to \infty} \left( \frac{a_{k+1-m}}{a_{k+1}} \right)^r \left( \frac{a_k}{a_{k-m}} \right)^r \frac{a_{k+1}}{a_k} = \theta \lim_{k \to \infty} \left( \frac{a_{k+1-m}}{a_{k+1}} \right)^r \lim_{k \to \infty} \left( \frac{a_k}{a_{k-m}} \right)^r \lim_{k \to \infty} \frac{a_{k+1}}{a_k} = \theta (R^m)^r \left( \frac{1}{R^m} \right)^r \frac{1}{R} = \theta \frac{1}{R}
\]

and \( \mathbb{E}(h_m^r(X)) \) converges for the same parameter space as the original density. \( \square \)

Now, if \( R = \infty \), higher moments may not converge for the same parameter space as the original density.

**Example.** Consider the series \( G(\theta) = \sum_{k=0}^{\infty} c^k \theta^k \) for some positive constant \( c \). If \( c = 1 \), then \( G(\theta) = \sum_{k=0}^{\infty} \theta^k \) and \( X \sim \text{Geom}(1 - \theta) \) which is discussed later. If \( c \neq 1 \), then applying the ratio test to \( G(\theta) \) gives:
\[
\lim_{k \to \infty} \frac{c^{2k+1} \theta^{k+1}}{c^{2k} \theta^k} = \lim_{k \to \infty} c^{2k+1-2k} \theta = \lim_{k \to \infty} c^{2(2k-1)} \theta = \lim_{k \to \infty} c^2 \theta,
\]
so if \( 0 < c < 1 \), \( G(\theta) \) converges for \( \theta \in [0, \infty) \) and if \( c > 1 \), \( G(\theta) \) converges only for \( \theta = 0 \).

Applying the ratio test to \( \mathbb{E}(h^2(X)) \):
\[
\mathbb{E}(h^2(X)) = \sum_{k=1}^{\infty} \left( \frac{c^{2k-1}}{c^{2k}} \right)^2 \frac{c^{2k}}{G(\theta)} \theta^k
= \sum_{k=1}^{\infty} \left( \frac{c^{2k-1}}{c^{2k}} \right)^2 \frac{c^{2k}}{c^{2k}} \frac{1}{G(\theta)} \theta^k
= \sum_{k=1}^{\infty} \frac{c^{2(2k-1)}}{c^{2k}} \frac{1}{G(\theta)} \theta^k
= \sum_{k=1}^{\infty} \frac{c^2}{c^{2k}} \frac{1}{G(\theta)} \theta^k
= \sum_{k=1}^{\infty} \frac{1}{c^{2k} G(\theta)} \theta^k,
\]
gives \( \mathbb{E}(h^2(X)) \) converges only for \( \theta \in [0, 1) \).
If a function $G(\theta) = a_0 + a_1 \theta + a_2 \theta^2 + \ldots$ as above (2.1) with radius of convergence $R = \infty$, and 

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right) = L < \infty,$$

then all moments of the estimator $h_m(X) = \frac{aX-m}{aX}$ exist for the same parameter space as the original density. Fortunately, most of the more useful power series distributions satisfy 

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right) = L < \infty,$$

in fact, they satisfy 

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right) = 1.$$

If 

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right) = \infty,$$

then $\frac{a_{k+1}}{a_k}$ must converge to zero faster than 

$$\left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right) \text{ converges to } \infty$$

in order for some higher moments of $h(X)$ to exist.

A sufficient condition for these moments to exist is 

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right) = L < \infty.$$ 

Clearly, if a function $G(\theta) = a_0 + a_1 \theta + a_2 \theta^2 + \ldots$ has radius of convergence $R = \infty$, and 

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right) = L < \infty,$$

then all moments of the estimator $h(X) = \frac{a(X-1)}{a(X)}$ exist for the same parameter space as the original density:

First, for $m = 1$, applying the ratio test to $E(h^r(X)) = \sum_{k=1}^{\infty} \left( \frac{a_{k-1}}{a_k} \right)^r \frac{a_k}{G(\theta)} \theta^k$ gives:

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a\k_{-1}} \right)^r \frac{a_{k+1}}{G(\theta)} \theta^k = \lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \right)^r \left( \frac{a_k}{a_{k-1}} \right)^r \frac{a_{k+1}}{a_k} \theta$$

$$= \lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right)^r \frac{a_{k+1}}{a_k} \theta.$$

Since 

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right) = L < \infty,$$

$$\lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right)^r \frac{a_{k+1}}{a_k} = \lim_{k \to \infty} \left( \frac{a_k}{a_{k+1}} \frac{a_k}{a_{k-1}} \right)^r \lim_{k \to \infty} \frac{a_{k+1}}{a_k}$$

$$= L^r 0 = 0$$

and $E(h^r(X))$ exist for the same parameter space as the original density.

If $m > 1$, applying the ratio test to $E(h_m^r(X)) = \sum_{k=1}^{\infty} \left( \frac{a_{k-1}}{a_k} \right)^r \frac{a_k}{G(\theta)} \theta^k$ gives:

$$\lim_{k \to \infty} \left( \frac{a_{k-m+1}}{a_{k+1}} \frac{a_{k-m}}{a_k} \right)^r \frac{a_{k-1}}{G(\theta)} \theta^k = \lim_{k \to \infty} \left( \frac{a_{k-m+1}}{a_{k+1}} \right)^r \left( \frac{a_{k-m}}{a_k} \right)^r \frac{a_{k+1}}{a_k} \theta$$
\[
\lim_{k \to \infty} \left( \frac{a_{k-m+1}}{a_{k+1}} \frac{a_k}{a_{k-m}} \right)^r \frac{a_{k+1}}{a_k}.
\]

Now
\[
= \lim_{k \to \infty} \left( \frac{a_{k-m+1}}{a_{k+1}} \frac{a_k}{a_{k-m}} \right)^r \frac{a_{k+1}}{a_k}
= \lim_{k \to \infty} \left( \frac{a_{k-m+1}}{a_{k+1}} \left( \frac{a_{k-m+1}}{a_{k-m+1}} \right)^2 \frac{a_{k-m+2}}{a_{k-m+2}} \left( \frac{a_{k-m+3}}{a_{k-m+3}} \right)^2 \cdots \right)
= \lim_{k \to \infty} \left[ \left( \frac{a_{k-m+1}}{a_{k-m+2}} \frac{a_k}{a_{k-m}} \right)^2 \left( \frac{a_{k-m+1}}{a_{k-m+1}} \frac{a_k}{a_{k-m}} \right)^2 \frac{a_{k+1}}{a_k} \right]
= \lim_{k \to \infty} \left( \frac{a_{k-m+1}}{a_{k-m+2}} \frac{a_k}{a_{k-m}} \right)^r \lim_{k \to \infty} \left( \frac{a_{k-m+2}}{a_{k-m+3}} \frac{a_k}{a_{k-m+1}} \right)^r \cdots
= L^r \cdot L^r \cdots L^r \cdot 0 = 0.
\]

So all moments of \( h_m(X) \) exist for the same parameter space as the original density.

Examples

**Example.** If \( a_k > 0 \) for all \( k = 0, 1, 2, \ldots, \ell - 1 \) and \( a_k = 0 \) for all \( k = \ell, \ell + 1, \ldots, \), then the estimator does not make sense at \( k = \ell \) since then \( a(\ell - 1)/a(\ell) \) becomes infinite. In this case, there is no unbiased estimator of \( \theta \). In this case, we will consider the same estimator \( h(X) \) after redefining it to equal zero for \( X = \ell \). This is not an unrealistic approach since \( \mathbb{P}(X = \ell) = 0 \). With this stipulation, \( h(X) \) is a well defined estimator of \( \theta \).

**Example.** (Odds Ratio) If we take \( a = (1, 1, 0, 0, \ldots) \), then we get \( X \sim B(1, \theta/(1 + \theta)) \). In this case there is no unbiased estimator of \( \theta \).

**Proof.** Let \( \delta(X) \) be any estimator of \( \theta \). Then we have
\[
\frac{\delta(0) + \delta(1)\theta}{1 + \theta} = \theta; \quad \theta \in (0, \infty)
\]
implies that \( \delta(0) = 0 \) by taking \( \theta \to 0 \). But then, dividing by \( \theta \) on both sides and letting \( \theta \) go to zero again gives that \( \delta(1) = 1 \). This estimator cannot be unbiased. \( \square \)

**Example.** (Zero-truncated Poisson) If \( G(\theta) = e^\theta - 1 \), \( a_k = 1/k! \) for \( k \in \mathbb{N} \), and \( h(k) = k \) for \( k > 1 \). \( \mathbb{E}(h(k)) = \theta \), and \( \text{Var}(h(X)) = \theta \frac{1 + \theta e^{-\theta}}{(1 - e^{-\theta})^2} \)

### 2.3 Lack of Existence of C-R Estimators

When a UMVU estimator of \( \theta \) exists, the variance of the estimator is often difficult to compute and sometimes impossible to write in closed form. Tate and Goen (1958) showed that the UMVU estimator of \( \theta \) for the zero-truncated Poisson case does not achieve the Cramér-Rao lower bound for its variance. Wijsman (1973) provided a rigorous proof that if an unbiased estimator of a function of a real parameter attains the Cramér-Rao lower bound, then the family of distributions must be a one-parameter exponential family. That is: let \( X \) be a random variable with density \( f_\theta(X) \), an unbiased estimator \( \delta(X) \) of a real-valued function \( g(\theta) \) attains its Cramér-Rao lower bound if and only if \( \frac{\partial}{\partial \theta} \log f_\theta(x) = a(\theta) + h(\theta)\delta(x) \) for some functions \( a(\theta) \) and \( h(\theta) \).

First recall the C-R inequality. Let \( L(\theta, x) \) be the likelihood function of \( X_1, X_2, \cdots, X_n \) which are iid random variables with common density \( f_\theta(x) \) that satisfies the two regularity conditions:

(i) The set \( D = \{ x : f_\theta(x) > 0 \} \) does not depend on \( \theta \), and for all \( x \in D \) and \( \theta \in \Theta \), \( \frac{\partial}{\partial \theta} \log f_\theta(x) \) exists and is finite.

(ii) If \( T(X) \) is any statistic such that \( \mathbb{E}_\theta(T(X)) < \infty \) for all \( \theta \in \Theta \), then the operations of expectation and differentiation with respect to \( \theta \) can be interchanged.

Let

\[
U(X) = \frac{\partial}{\partial \theta} \log L(\theta, X)
\]

\[
= \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f_\theta(X_i)
\]
\[ = \sum_{i=1}^{n} \frac{\partial f_{\theta}(X_i)}{\partial \theta} f_{\theta}(X_i). \]

Then,

\[
\mathbb{E}(U(X)) = n \mathbb{E} \left\{ \frac{\partial f_{\theta}(X_1)}{\partial \theta} \right\} \\
= n \int_{\mathbb{R}} \left\{ \frac{\partial f_{\theta}(x)}{\partial \theta} \right\} f_{\theta}(x) \, d\mu(x) \\
= n \int_{\mathbb{R}} \frac{\partial f_{\theta}(x)}{\partial \theta} \, d\mu(x) \\
= n \frac{\partial}{\partial \theta} \int_{\mathbb{R}} f_{\theta}(x) \, d\mu(x) \\
= n \frac{\partial}{\partial \theta} (1) = 0.
\]

and

\[
\text{Var}(U(X)) = n \text{Var} \left\{ \frac{\partial f_{\theta}(X_1)}{\partial \theta} \right\} \\
= n \mathbb{E} \left\{ \frac{\partial f_{\theta}(X_1)}{\partial \theta} \right\}^2 \\
= n \mathbb{E} \left\{ \frac{\partial f_{\theta}(X_1)}{f_{\theta}(X_1)} \right\} \\
= : nI(\theta).
\]

Let \( \delta(X) \) be an unbiased estimator of \( g(\theta) \). Then, since \( \mathbb{E}(U(X)) = 0 \),

\[
\text{Cov}(\delta(X), U(X)) = \mathbb{E}(\delta(X) \cdot U(X)) \\
= \int_{\mathbb{R}^n} \delta(x) \left\{ \frac{\partial L(\theta, x)}{\partial \theta} \right\} \frac{L(\theta, x)}{L(\theta, x)} \, d\mu(x) \\
= \int_{\mathbb{R}^n} \delta(x) \frac{\partial L(\theta, x)}{\partial \theta} \, d\mu(x) \\
= \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} \delta(x)L(\theta, x) \, d\mu(x) \\
= \frac{\partial}{\partial \theta} \mathbb{E}(\delta(X)) \\
= \frac{\partial g(\theta)}{\partial \theta} = g'(\theta).
\]

By the correlation inequality, we have

\[
\{\text{Cov}(\delta(X), U(X))\}^2 \leq \text{Var}(\delta(X) \cdot \text{Var}(U(X)).
\]
Which gives

\[ \text{Var}(\delta(X)) \geq \frac{(g'(\theta))^2}{n I(\theta)}. \]

For equality to hold, we have the following proposition. As we do not need to worry about null sets depending on \( \theta \), we may simplify Wijsman’s proof. For a detailed proof of this proposition, see Wijsman (1973).

**Proposition 6.** (Wijsman) When the necessary regularity conditions are satisfied, the necessary and sufficient condition for the existence of an unbiased estimator (of \( \theta \)) to have its variance equal the Cramér-Rao lower bound is that there should exist a function \( h(\theta) \) such that

\[ \theta + h(\theta) \left\{ \frac{\partial \log L(\theta, X)}{\partial \theta} \right\} \]

does not depend on \( \theta \) for all values of \( x \).

The regularity conditions that concern us are:

(i) \( 0 < \text{Var}_\theta \left( \frac{\partial}{\partial \theta} \log f_\theta(X) \right) < \infty \) for every \( \theta \in \Theta \)

(ii) we can interchange the summation and differentiation with respect to \( \theta \).

**Proof.** Suppose there exists a function \( h(\theta) \) such that for each value of \( x \), 6 does not depend on \( \theta \). That is,

\[ \delta(X) = \theta + h(\theta) \left\{ \frac{\partial \log L(\theta, X)}{\partial \theta} \right\} \]

can be considered an estimator of \( \theta \). Then

\[ \mathbb{E}(\delta(X)) = \theta + h(\theta) \mathbb{E} \left\{ \frac{\partial \log L(\theta, X)}{\partial \theta} \right\} \]

\[ = \theta + h(\theta) \theta = \theta \]

and,

\[ \text{Var}(\delta(X)) = h^2(\theta) \text{Var} \left\{ \frac{\partial \log L(\theta, X)}{\partial \theta} \right\} \]
\[ = h^2(\theta) \text{Var}(U(X)). \]

Also,

\[
\text{Cov}(\delta(X), U(X)) = h(\theta) \text{Cov}(U(X), U(X))
\]

\[ = h(\theta) \text{Var}(U(X)). \]

so,

\[
\{\text{Cov}(\delta(X), U(X))\}^2 = h^2(\theta) \{\text{Var}(U(X))\}^2.
\]

On the other hand

\[
\text{Var}(\delta(X)) \cdot \text{Var}(U(X)) = h^2(\theta) \text{Var}(U(X)) \cdot \text{Var}(U(X)).
\]

Hence, an unbiased estimator exists for which the equality holds in the C-R bound.

Conversely, suppose there exists an unbiased estimator \( \delta(X) \) of \( \theta \) whose variance equals the lower bound in the C-R inequality for each value of \( \theta \) in the parameter space. Hence, by the equality part of the correlation inequality, it must be that \( \delta(X) \) is a linear function of \( U(X) \). That is, there exist functions \( a(\theta) \) and \( h(\theta) \) such that

\[
\delta(X) = a(\theta) + h(\theta)U(X).
\]

Since the estimator is unbiased for \( \theta \),

\[
\theta = \mathbb{E}(\delta(X)) = a(\theta) + h(\theta)\mathbb{E}(U(X)) = a(\theta).
\]

So, the estimator must be

\[
\delta(X) = \theta + h(\theta)U(X).
\]

And since it is an estimator, this function must not depend on \( \theta \) for each value of \( X \). This completes the proof.

\[ \square \]

**Proposition 7.** For the zero-truncated Poisson distribution, there does not exist any unbiased estimator whose variance equals the Cramér-Rao lower bound.
Proof. By the last proposition we need to check if there exists a function $h(\theta)$ such that

$$\theta + h(\theta) \left( \frac{S_n}{\theta} - \frac{n}{1 - e^{-\theta}} \right)$$

does not depend on $\theta$ for each value of $S_n$. One can show that there is no such function by taking two values of $S_n = n, 2n$ and show that that leads to a contradiction.

Remark. For an arbitrary PSD, $X_1, X_2, \cdots, X_n \overset{iid}{\sim} PSD(\theta)$, we have:

$$L(\theta, X) = \frac{A(n, S_n)\theta^{S_n}}{(G(\theta))^n}$$

$$\log L(\theta, X) = \log \frac{A(n, S_n)\theta^{S_n}}{(G(\theta))^n}$$

$$\frac{\partial}{\partial \theta} \log L(\theta, X) = \frac{S_n}{\theta} - \frac{nG'(\theta)}{G(\theta)}.$$

So, a UMVU estimator of $\theta$ which achieves the C-R lower bound will exist if and only if there exists a function $h(\theta)$ so that

$$\theta + h(\theta) \left( \frac{S_n}{\theta} - \frac{nG'(\theta)}{G(\theta)} \right)$$

does not depend on $\theta$ for each value of $S_n$. That is,

$$\delta(X) = \theta + h(\theta)\frac{S_n}{\theta} - \frac{nh(\theta)G'(\theta)}{G(\theta)} = h(\theta)\frac{S_n}{\theta} + \left( \theta - \frac{nh(\theta)G'(\theta)}{G(\theta)} \right)$$

does not depend on $\theta$ for each value of $S_n$. Since $h(\theta)$ cannot depend on $S_n$, $\delta(X)$ is a function of $X$, and $\left( \theta - \frac{nh(\theta)G'(\theta)}{G(\theta)} \right)$ is a function $\theta$, we must have $h(\theta) = c\theta$ for some constant $c$. Note that $c$ cannot be zero because $c = 0$ gives $\delta(X) = \theta$ and $\delta(X)$ is a statistic.

Working backwards, if $h(\theta) = c\theta$,

$$\delta(X) = \theta + h(\theta)\left( \frac{S_n}{\theta} - \frac{nG'(\theta)}{G(\theta)} \right)$$

$$= \theta + c\theta \left( \frac{\partial}{\partial \theta} \log L(\theta, X) \right)$$

and

$$\frac{\partial}{\partial \theta} \log L(\theta, X) = \frac{\delta(x) - \theta}{c\theta}.$$
\[ \frac{1}{c} \left( \frac{\delta(x)}{\theta} - 1 \right). \]

Integrating both sides with respect to \( \theta \) gives:

\[
\int \frac{\partial}{\partial \theta} \log L(\theta, x) d\theta = \int \frac{1}{c} \left( \frac{\delta(x)}{\theta} - 1 \right) d\theta
\]

\[
\log L(\theta, x) = \frac{1}{c} (\delta(x) \log(\theta) - \theta) + c_i(x)
\]

\[
= \frac{1}{c} (\log(\theta^\delta(x)) - \log(e^\theta) + \log(e^{c_i(x)}))
\]

\[
= \log \left( \frac{e^{c_i(x)} \theta^\delta(x)}{e^\theta} \right)^{1/c}
\]

where \( c_i \) is some constant of integration. So

\[
L(\theta, x) = \left( \frac{e^{c_i(x)} \theta^\delta(x)}{e^\theta} \right)^{1/c} = \frac{e^{c_i(x)} \theta^\delta(x)/c}{e^{\theta/c}},
\]

and

\[
L(\theta, X) = A(n, S_n) \frac{\theta^{S_n}}{(G(\theta))^n} = e^{c_i(x)} \frac{\theta^\delta(x)/c}{e^{\theta/c}}
\]

for all values of \( S_n \) and all values of \( \theta \) in some open interval of \( \mathbb{R} \).

Now,

\[
A(n, S_n) \frac{\theta^{S_n}}{(G(\theta))^n} = e^{c_i(x)} \frac{\theta^\delta(x)/c}{e^{\theta/c}} \neq 0
\]

implies

\[
(G(\theta))^n = \frac{A(n, S_n) \theta^{S_n - \delta(x)/c} e^{\theta/c}}{e^{c_i(x)}}.
\]

That is, \( (G(\theta))^n = \beta \theta^\alpha e^{\theta/c} \) for some constants \( \alpha \) and \( \beta \). This gives that the distribution generated \( (G(\theta))^n \) is of the form

\[
\beta \frac{\theta^{\alpha+k}}{c^\alpha (k!) \beta^\alpha e^{\theta/c}}, \quad k = 0, 1, 2, \ldots
\]

which is a Poisson distribution.
CHAPTER 3

Optimality of the UMVU Estimator $h_m(X) = \frac{aX - m}{aX}$

In this chapter, we consider the optimality of the UMVU estimator $h_m(X) = \frac{aX - m}{aX}$. The Poisson distribution is the only power series distribution whose UMVU estimator for $\theta$ equals the Cramér-Rao lower bound. This chapter discusses the properties of $h_m(S_n)$ as $n \to \infty$.

Let $X_1, X_2, \ldots, X_n$ be iid with power series distribution defined by $G(\theta) = a_0 + a_1 \theta + \ldots$ where $a_k > 0$ for $k = \kappa, \kappa + 1, \ldots$ for some positive integer $\kappa$ and let $S_n = X_1 + X_2 + \cdots + X_n$. Let $A(n, k)$ denote the $k^{th}$ term of the $n$-fold convolution sequence of the original sequence $(a_0, a_1, \cdots)$, so $P(S_n = k) = \frac{A(n, k)}{(G(\theta))^n} \theta^k$. (Note that for PSD, convolution is the same as exponentiation, so $A(n, k)$ is the $k^{th}$ term of the power series of $(G(\theta))^n$.) The UMVU estimator of $\theta^m$ becomes

$$h_m(S_n) = \frac{A(n, S_n - m)}{A(n, S_n)}.$$

3.1 Asymptotic Normality of the UMVU Estimator

Theorem 1. When $\sigma^2 = \text{Var}_1(X_1) < \infty$, the UMVU estimator $h_m(S_n) = \frac{A(n, S_n - m)}{A(n, S_n)}$ is asymptotically normal and

$$\sqrt{n} \left( \frac{A(n, S_n - m)}{A(n, S_n)} - \theta^m \right) \xrightarrow{\text{dist}} N \left( 0, \frac{m^2 \theta^{2m}}{\sigma^2} \right).$$

(3.1)

Proof. Let $T_n \sim S_n$ and the two random variables be independent of each other for all values of $n$. We may write

$$\sqrt{n} \left( \frac{A(n, S_n - m)}{A(n, S_n)} - \theta^m \right) = \theta^m \sqrt{n} \left( \frac{A(n, S_n - m)}{A(n, S_n)} \theta^{S_n - m} - 1 \right).$$
\[ a.s. \quad \theta^n \sqrt{n} \left[ \mathbb{P}(T_n = S_n - m | S_n) - \mathbb{P}(T_n = S_n | S_n) \right] / \mathbb{P}(T_n = S_n | S_n). \]

Now we use the following result of Petrov (1975),
\[ \sigma \sqrt{n} \mathbb{P}(T_n = k) = \phi(t_{nk}) + \frac{\phi(t_{nk})q(t_{nk})}{\sqrt{n}} + \frac{o(1)}{\sqrt{n}(1 + |t_{nk}|^3)}, \]
where \( \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \) and \( q(t) \) is a polynomial of degree three.

Applying this result gives:
\[
\sqrt{n} \left( \frac{A(n, S_n - m)}{A(n, S_n)} - \theta^n \right) = \theta^n \sqrt{n} \left\{ \begin{array}{c}
\frac{\phi(Z_n - \alpha_n) - \phi(Z_n)}{\phi(Z_n)} + \frac{\phi(Z_n)q(Z_n)\sqrt{n}}{\sqrt{n}(1 + |Z_n|^3)} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \\
+ \phi(Z_n - \alpha_n)q(Z_n - \alpha_n) - \phi(Z_n)q(Z_n) + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} - \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \\
+ \frac{\phi(Z_n - \alpha_n)q(Z_n - \alpha_n)\sqrt{n}}{\sqrt{n}(1 + |Z_n|^3)} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \\
+ \phi(Z_n) + \frac{\phi(Z_n)q(Z_n)\sqrt{n}}{\sqrt{n}(1 + |Z_n|^3)} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \end{array} \right\}
\]
\[ = A_n + B_n, \]

where
\[ Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}}, \quad \alpha_n = \frac{m}{\sigma \sqrt{n}}, \quad \mu = \mathbb{E}(X_1), \quad \sigma^2 = \text{Var}(X_1). \]

Considering the \( B_n \) term, asymptotically for \( n \), we have
\[
B_n = \theta^n \sqrt{n} \left\{ \begin{array}{c}
\frac{\phi(Z_n - \alpha_n)q(Z_n - \alpha_n) - \phi(Z_n)q(Z_n)}{\phi(Z_n)} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} - \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \\
+ \phi(Z_n) + \frac{\phi(Z_n)q(Z_n)\sqrt{n}}{\sqrt{n}(1 + |Z_n|^3)} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \\
+ \phi(Z_n - \alpha_n)\sqrt{n} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} - \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \\
+ \phi(Z_n) + \frac{\phi(Z_n)q(Z_n)\sqrt{n}}{\sqrt{n}(1 + |Z_n|^3)} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \\
+ \phi(Z_n - \alpha_n)\sqrt{n} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} - \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)} \end{array} \right\}
\]
\[ = \theta^n \frac{\alpha_n}{\phi(Z_n)} \left( \frac{\phi(Z_n - \alpha_n)q(Z_n - \alpha_n) - \phi(Z_n)q(Z_n)}{\alpha_n} + \frac{o(1)}{(1 + |Z_n|^3)} \right) \]
\[ = \theta^n \frac{\alpha_n}{\phi(Z_n)} \left( \frac{\phi(Z_n - \alpha_n)q(Z_n - \alpha_n) - \phi(Z_n)q(Z_n)}{\alpha_n} + \frac{o(1)}{(1 + |Z_n|^3)} \right) \]
\[ = \frac{\alpha_n}{\phi(Z_n)} \left( \frac{\phi(Z_n - \alpha_n)q(Z_n - \alpha_n) - \phi(Z_n)q(Z_n)}{\alpha_n} + \frac{o(1)}{(1 + |Z_n|^3)} \right) \]
\[ = \frac{\alpha_n}{\phi(Z_n)} \left( \frac{\phi(Z_n - \alpha_n)q(Z_n - \alpha_n) - \phi(Z_n)q(Z_n)}{\alpha_n} + \frac{o(1)}{(1 + |Z_n|^3)} \right) \]
\[ = \frac{\alpha_n}{\phi(Z_n)} \left( 1 + \frac{\phi(Z_n)q(Z_n)\sqrt{n}}{(1 + |Z_n|^3)} + \frac{o(1)}{(1 + |Z_n|^3)} \right) \]
\[ = \frac{\alpha_n}{\phi(Z_n)} \left( 1 + \frac{\phi(Z_n)q(Z_n)\sqrt{n}}{(1 + |Z_n|^3)} + \frac{o(1)}{(1 + |Z_n|^3)} \right) \]

Since, \( q \) is a polynomial of degree three and since \( Z_n \overset{\text{dist}}{\rightarrow} Z \sim N(0, 1) \), we have
\[ q(Z_n) \overset{\text{dist}}{\rightarrow} q(Z), \quad \phi(Z_n) \overset{\text{dist}}{\rightarrow} \phi(Z). \]
Also, for any \( \varepsilon > 0 \),

\[
\mathbb{P} \left( \left| \frac{o(1)}{\phi(Z_n)} - 0 \right| > \varepsilon \right) = \mathbb{P} \left( \phi(Z_n) < \frac{|o(1)|}{\varepsilon} \right) \\
\leq \mathbb{P} \left( \phi(Z_n) < \frac{\delta}{\varepsilon} \right)
\]

for any small positive \( \delta \) and all large enough \( n \). This gives

\[
\limsup_{n} \mathbb{P} \left( \left| \frac{o(1)}{\phi(Z_n)} - 0 \right| > \varepsilon \right) \leq \mathbb{P} \left( \phi(Z) \leq \frac{\delta}{\varepsilon} \right).
\]

Since \( \delta \) is arbitrary and \( \phi(Z) \) is a positive random variable, the last probability must be zero. Therefore,

\[
\frac{o(1)}{\phi(Z_n)} \xrightarrow{\mathbb{P}} 0.
\]

Therefore,

\[
\frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)\phi(Z_n)} \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{\alpha_n}{\phi(Z_n)} \xrightarrow{\mathbb{P}} 0.
\]

The mean value theorem gives

\[
\frac{\phi(Z_n - \alpha_n)q(Z_n - \alpha_n) - \phi(Z_n)q(Z_n)}{\alpha_n} = (\phi q)'(\eta_n) \quad \text{for some } \eta_n \in (Z_n - \alpha_n, Z_n).
\]

Since \( Z_n \) and \( Z_n - \alpha_n \) both converge to \( Z \sim N(0, 1) \), \( \eta_n \xrightarrow{\text{dist}} N(0, 1) \), and \( (\phi q)'(\eta_n) \xrightarrow{\mathbb{P}} (\phi q)'(Z) \). Lastly, \( (\phi q)'(\eta_n) \) is bounded as

\[
\lim_{Z \to \pm \infty} (\phi q)'(Z) = \lim_{Z \to \pm \infty} (\phi'q + \phi q')(Z) \\
= \lim_{Z \to \pm \infty} \left( \text{degree 4 polynomial} + \text{degree 2 polynomial} \right)(Z) \\
= 0,
\]

so \( \frac{\alpha_n}{\phi(Z_n)} (\phi q)'(\eta_n) \xrightarrow{\mathbb{P}} 0 \). Therefore \( B_n \xrightarrow{\mathbb{P}} 0 \).

In a similar fashion,

\[
A_n = 0^n \sqrt{n} \frac{\alpha_n}{\phi(Z_n)} \frac{\phi(Z_n - \alpha_n) \cdot \phi(Z_n)}{\alpha_n} \\
= \frac{o(1)}{\sqrt{n}} + \frac{o(1)}{\sqrt{n}(1 + |Z_n|^3)\phi(Z_n)}
\]
\[
= \frac{m \theta^m}{\sigma} \left( \frac{1}{\phi(Z_n)} \left( \frac{\phi(Z_n - \alpha_n) - \phi(Z_n)}{\alpha_n} \right) \right).
\]

By Taylor’s Theorem and The Mean Value Theorem,

\[
\phi(Z_n - \alpha_n) = \phi(Z_n) - \alpha_n \phi'(Z_n) + \frac{\alpha_n^2}{2} \phi''(\xi_n),
\]

where \(\xi_n\) is between \(Z_n\) and \(Z_n - \alpha_n\). Hence,

\[
\frac{\phi(Z_n - \alpha_n) - \phi(Z_n)}{\alpha_n} = -\phi'(Z_n) + \frac{\alpha_n}{2} \phi''(\xi_n).
\]

Since,

\[
\phi'(t) = -t\phi(t), \quad \text{and} \quad \phi''(t) = -\phi(t) + t^2 \phi(t),
\]

we have

\[
\frac{1}{\phi(Z_n)} \left( \frac{\phi(Z_n - \alpha_n) - \phi(Z_n)}{\alpha_n} \right) = -Z_n \phi(Z_n) + \frac{\alpha_n}{2} \phi''(\xi_n).
\]

Also, \(\xi_n\) lies between \(Z_n - \alpha_n\) and \(Z_n\) and both of the last two random variables converge to \(N(0, 1)\), so \(\xi_n\) must converge to \(N(0, 1)\). Thus,

\[
-\frac{\phi(\xi_n)}{\phi(Z_n)} + \frac{\xi_n^2 \phi(\xi_n)}{2 \phi(Z_n)} \xrightarrow{d} -1 + Y,
\]

where \(Y\) is a chi-square random variable with one degree of freedom. Hence,

\[
-Z_n + \frac{\alpha_n}{2} \left( -\frac{\phi(\xi_n)}{\phi(Z_n)} + \frac{\xi_n^2 \phi(\xi_n)}{\phi(Z_n)} \right) \xrightarrow{d} N(0, 1).
\]

This gives that \(A_n \xrightarrow{d} N(0, \frac{m^2 \theta^2}{\sigma^2})\).

\[\square\]

3.2 Limiting Absolute Mean Deviation

**Theorem 2.** For any \(\theta \in (0, R)\), we have

\[
E_{\phi} |h_n(S_n) - \theta^m| = O(n^{-1/2}).
\]

(3.2)
Proof. First note that

$$\mathbb{E}_\theta |h_n(S_n) - \theta^m| = \sum_{k=\kappa}^{\kappa+N} \frac{A(n, k - m)}{A(n, k)} \theta^m \mathbb{P}(S_n = k) - \theta^m \mathbb{P}(S_n = k)$$

$$= \sum_{k=\kappa}^{\kappa+N} \theta^m \mathbb{P}(S_n = k - m) \mathbb{P}(S_n = k) - \theta^m \mathbb{P}(S_n = k)$$

$$= \theta^m \sum_{k=\kappa}^{\kappa+N} \mathbb{P}(S_n = k - m) - \mathbb{P}(S_n = k).$$

A result of Bikelis and Jasjuna (1967) says that if $X_i$ are iid integer-valued and aperiodic (which is true in our case by assumptions on $G(\theta)$) random variables with mean $\mu$ and variance $\sigma^2$ and finite third moment, then

$$\sum_j (1 + |t_{nj}|^3) \mathbb{P}(S_n = j) \frac{\phi(t_{nj})}{\sigma \sqrt{n}} = O(n^{-1/2}),$$

where $\phi(t)$ is the standard normal density and $t_{nj} = (j - n\mu)/\sigma \sqrt{n}$. This gives that

$$\sum_j \mathbb{P}(S_n = j) \frac{\phi(t_{nj})}{\sigma \sqrt{n}} = O(n^{-1/2}).$$

So, by the triangle inequality, we need only show that

$$\sum_j \frac{\phi(t_{nj})}{\sigma \sqrt{n}} - \frac{\phi(t_{n,j+m})}{\sigma \sqrt{n}} \leq m \sum_j \frac{\phi(t_{nj})}{\sigma \sqrt{n}} - \frac{\phi(t_{n,j+1})}{\sigma \sqrt{n}} = O(n^{-1/2}).$$

Now, monotonicity of $t_{nj}$ as a function of $j$ and fixed $n$ gives that the first term is smaller than the second term for $j + 1 \leq n\mu$, and the converse holds if $j \geq n\mu$. So

$$\sum_j \left| \frac{\phi(t_{nj})}{\sigma \sqrt{n}} - \frac{\phi(t_{n,j+1})}{\sigma \sqrt{n}} \right|$$

$$= \frac{1}{\sigma \sqrt{n}} \left\{ \sum_{j<n\mu} (\phi(t_{nj}) - \phi(t_{n,j})) + \sum_{j \geq n\mu} (\phi(t_{nj}) - \phi(t_{n,j+1})) \right\}$$

$$\leq \frac{1}{\sigma \sqrt{n}} \left\{ \phi(t_{n,n\mu}) + \phi(t_{n,n\mu}) \right\}$$

$$= \frac{2}{\sigma \sqrt{n} \sqrt{2\pi}} = O(n^{-1/2}).$$

We can also evaluate the above sum as follows. Note that,

$$\frac{\phi(t_{nj})}{\sigma \sqrt{n}} - \frac{\phi(t_{n,j+1})}{\sigma \sqrt{n}}$$
\[
\phi(t_{nj}) = \frac{\phi(t_{nj})}{\sigma \sqrt{n}} \left(1 - \exp\{2(j - n\mu) - 1\}/(2\sigma^2 n)\right)
\]
\[
= \frac{\phi(t_{nj})}{\sigma \sqrt{n}} \left(1 - \sum_{k=1}^{\infty} \left(\frac{t_{nj}}{\sigma \sqrt{n}} - \frac{1}{2\sigma^2 n}\right)^k \frac{1}{k!}\right).
\]

This gives that for large enough \(n\), we have
\[
\sum_j \left|\frac{\phi(t_{nj})}{\sigma \sqrt{n}} - \frac{\phi(t_{nj+1})}{\sigma \sqrt{n}}\right| \leq \sum_j \frac{\phi(t_{nj})}{\sigma \sqrt{n}} \left(\sum_{k=1}^{\infty} \left(\frac{|t_{nj}|}{\sigma \sqrt{n}} - \frac{1}{2\sigma^2 n}\right)^k \frac{1}{k!}\right)
\]
\[
= \frac{1}{\sigma \sqrt{n}} \sum_j \frac{\phi(t_{nj})}{\sigma \sqrt{n}} \left(\sum_{k=1}^{\infty} \left(\frac{1}{\sigma \sqrt{n}}\right)^{k-1} \left(|t_{nj}| - \frac{1}{2\sigma \sqrt{n}}\right)^k \frac{1}{k!}\right)
\]
\[
\leq \frac{1}{\sigma \sqrt{n}} \sum_j \frac{\phi(t_{nj})}{\sigma \sqrt{n}} \left(\sum_{k=1}^{\infty} (|t_{nj}| - 1)^k \frac{1}{k!}\right),
\]

Therefore,
\[
\sum_j \left|\frac{\phi(t_{nj})}{\sigma \sqrt{n}} - \frac{\phi(t_{nj+1})}{\sigma \sqrt{n}}\right| \leq \frac{1}{\sigma \sqrt{n}} \sum_j \frac{\phi(t_{nj})}{\sigma \sqrt{n}} \exp\{|t_{nj}| - 1\} = O(n^{-1/2}),
\]
since \(\mathbb{E}(e^{|Z|}) < \infty\) for \(Z \sim N(0, 1)\).

**Theorem 3.** For any \(\theta \in (0, R)\), we have
\[
\lim_{n \to \infty} \sqrt{n} \mathbb{E}_\theta |h_m(S_n) - \theta^m| = \frac{m \theta^m}{\sigma} \sqrt{\frac{2}{\pi}}.
\] (3.3)

**Proof.** We will use the result from (3.2). Fix a positive number \(M\) taking any one of the values \(\{j + \frac{1}{2m}, j = 1, 2, \cdots\}\). Next, take \(n\) large enough so that \(\frac{m}{\sigma \sqrt{n}} < \frac{1}{2}\). This ensures that if
\[
|t_{nk}| = \left|\frac{k - n\mu}{\sigma \sqrt{n}}\right| < M,
\]
then
\[
|t_{n,k-m} - t_{n,k} - \frac{m}{\sigma \sqrt{n}}| \leq M + 1,
\]
and similarly, if \(|t_{nk}| \geq M\) then \(|t_{n,k-m}| \geq M - 1\).
As in the proof of the last theorem, note that

\[
\mathbb{E}[h_n(S_n) - \theta^n] = \theta^n \sum_{k=0}^{\infty} |\mathbb{P}(S_n = k) - \mathbb{P}(S_n = k)|
\]

\[
= \theta^n \sum_{k : |t_{nk}| \geq M} |\mathbb{P}(S_n = k) - \mathbb{P}(S_n = k)|
+ \theta^n \sum_{k : |t_{nk}| < M} |\mathbb{P}(S_n = k) - \mathbb{P}(S_n = k)|.
\]

The result of Bikelis and Jaisohn applied to the first sum gives

\[
\sum_{k : |t_{nk}| \geq M} |\sigma \sqrt{n} \mathbb{P}(S_n = k) - \phi(t_{nk})|
\]

\[
= \frac{1}{1 + M} \sum_{k : |t_{nk}| \geq M} \left(1 + M\right) |\sigma \sqrt{n} \mathbb{P}(S_n = k) - \phi(t_{nk})|
\]

\[
\leq \frac{1}{1 + M} \sum_{k : |t_{nk}| \geq M} \left(1 + |t_{nk}|^3\right) |\sigma \sqrt{n} \mathbb{P}(S_n = k) - \phi(t_{nk})|
\]

\[
\leq \frac{1}{1 + M} \sum_{k} (1 + |t_{nk}|^3) |\sigma \sqrt{n} \mathbb{P}(S_n = k) - \phi(t_{nk})|
= O_n(1)
\]

Similarly, we have

\[
\sum_{k : |t_{nk}| \geq M} |\sigma \sqrt{n} \mathbb{P}(S_n = k - m) - \phi(t_{nk,k-m})|
\]

\[
= \frac{1}{1 + M} \sum_{k : |t_{nk}| \geq M} \left(1 + M\right) |\sigma \sqrt{n} \mathbb{P}(S_n = k - m) - \phi(t_{nk,k-m})|
\]

\[
= \frac{1}{1 + M} \sum_{k : |t_{nk}| \geq M} \left(2 + (M - 1)\right) |\sigma \sqrt{n} \mathbb{P}(S_n = k - m) - \phi(t_{nk,k-m})|
\]

\[
\leq \frac{1}{1 + M} \sum_{k : |t_{nk}| \geq M} (2 + |t_{nk,k-m}|^3) |\sigma \sqrt{n} \mathbb{P}(S_n = k - m) - \phi(t_{nk,k-m})|
\]

\[
\leq \frac{1}{1 + M} \sum_{k} (2 + |t_{nk,k-m}|^3) |\sigma \sqrt{n} \mathbb{P}(S_n = k - m) - \phi(t_{nk,k-m})|
= O_n(1)
\]
Now consider the following term.

\[
\sum_{k:|t_n-k| \geq M} |\phi(t_{nk}) - \phi(t_{n,k-m})| \\
\leq m \sum_{k:|t_n-k| \geq M} |\phi(t_{nk}) - \phi(t_{n,k-1})| \\
= m \sum_{k:|t_n-k| \leq -M} |\phi(t_{nk}) - \phi(t_{n,k-1})| + m \sum_{k:|t_n-k| \geq M} |\phi(t_{nk}) - \phi(t_{n,k-1})| \\
= m \sum_{k:|t_n-k| \leq -M} (\phi(t_{nk}) - \phi(t_{n,k-1})) + m \sum_{k:|t_n-k| \geq M} (\phi(t_{n,k-1}) - \phi(t_{nk})) \\
\leq m\phi(-M) + m\phi(M-1) \\
\leq 2m\phi(M-1).
\]

Here we used the fact that \( t_{nk} \) is a monotonic increasing sequence in \( k \) and that \( \phi(t) \) is bell shaped curve. Note that all the above three bounds go to zero as \( M \) gets large.

To consider the other term, we use the result of Petrov (1975). That is,

\[
\sigma \sqrt{n} \mathbb{P}(S_n = k) = \phi(t_{nk}) + \frac{\phi(t_{nk})q(t_{nk})}{\sqrt{n}} + \frac{o_n(1)}{\sqrt{n}(1 + |t_{nk}|^3)}.
\]

For the second term, we have \( |t_{nk}| < M \), and \( \phi(t_{nk}) > \phi(M) > 0 \). Thus, we may write

\[
\sigma \sqrt{n} \mathbb{P}(S_n = k) = \phi(t_{nk}) + \frac{\phi(t_{nk})q(t_{nk})}{\sqrt{n}} + \frac{o_n(1)\phi(t_{nk})}{\phi(M)\sqrt{n}(1 + |t_{nk}|^3)}.
\]

Which is the same as

\[
\sigma \sqrt{n} \mathbb{P}(S_n = k) = \phi(t_{nk}) + \frac{\phi(t_{nk})q(t_{nk})}{\sqrt{n}} + \frac{o_n(1)\phi(t_{nk})}{\sqrt{n}},
\]

where \( o_n(1) \) term does not depend on \( k \) or \( t_{nk} \). Similarly, we have

\[
\sigma \sqrt{n} \mathbb{P}(S_n = k - m) = \phi(t_{n,k-m}) + \frac{\phi(t_{n,k-m})q(t_{n,k-m})}{\sqrt{n}} + \frac{o_n(1)\phi(t_{n,k-m})}{\sqrt{n}},
\]

where \( o_n(1) \) term does not depend on \( k \) or \( t_{n,k-m} \). So

\[
\sum_{k:|t_{n,k}| < M} |\sigma \sqrt{n} \mathbb{P}(S_n = k) - \sigma \sqrt{n} \mathbb{P}(S_n = k - m)|
\]
\[
= \sum_{k:|t_{n,k}|<M} \left| \phi(t_{nk}) - \phi(t_{n,k-m}) + \frac{\phi(t_{nk})q(t_{nk}) - \phi(t_{n,k-m})q(t_{n,k-m})}{\sqrt{n}} \right.
+ \frac{o_n(1)(\phi(t_{nk}) + \phi(t_{n,k-m}))}{\sqrt{n}} \biggr|.
\]

We will show that this expression converges to
\[
\int_{-M}^{M} |t| \phi(t) dt.
\]
Note that this is the usual Riemann integral and (hence the same as the) Lebesgue integral.

For our purposes, we will invoke the definition of the Riemann integral. First note that

\[
\phi(t_{nk}) - \phi(t_{n,k-m}) = \phi(t_{nk}) - \phi(t_{nk} - m/(\sigma \sqrt{n}))
= \phi(t_{nk}) - \phi(t_{nk}) + m\phi'(t_{nk})/(\sigma \sqrt{n}) + O_n(1/n)
= - \frac{mt_{nk} \phi(t_{nk})}{\sigma \sqrt{n}} + O_n(1/n).
\]

Again, the \(O_n(1/n)\) term does not depend on \(k\) or \(t_{nk}\) since \(|t_{nk}| \leq M\). By a similar argument,

\[
\frac{\phi(t_{nk})q(t_{nk}) - \phi(t_{n,k-m})q(t_{n,k-m})}{\sqrt{n}} = O_n(1/n),
\]

where the \(O_n(1/n)\) term does not depend on \(k\) or \(t_{nk}\). Finally, the error terms can be written as

\[
\frac{o_n(1)(\phi(t_{nk}) + \phi(t_{n,k-m}))}{\sqrt{n}} = o_n(1/\sqrt{n}),
\]

and it does not depend on \(k\) and \(t_{nk}\), since \(t_{n,k}\) are bounded by \(M\). Thus, we have

\[
= \sum_{k:|t_{n,k}|<M} \left| \sigma \sqrt{n} \mathbb{P}(S_n = k) - \sigma \sqrt{n} \mathbb{P}(S_n = k - m) \right|
= \sum_{k:|t_{n,k}|<M} \left| \phi(t_{nk}) - \phi(t_{n,k-m}) + \frac{\phi(t_{nk})q(t_{nk}) - \phi(t_{n,k-m})q(t_{n,k-m})}{\sqrt{n}} \right.
+ \frac{o_n(1)(\phi(t_{nk}) + \phi(t_{n,k-m}))}{\sqrt{n}} \biggr|.
\]

\[
= \sum_{k:|t_{n,k}|<M} \left| \frac{mt_{nk} \phi(t_{nk})}{\sigma \sqrt{n}} + O_n(1/n) + o_n(1/\sqrt{n}) \right|
\]
\[
= m \sum_{k: |t_{n,k}| < M} \frac{\phi(t_{nk})}{\sigma/\sqrt{n}} |t_{nk} + O_n(1/\sqrt{n}) + o_n(1)|.
\]

Again, we used that \( \phi(t_{nk}) \) remains bounded away from zero by \( \phi(M) \) for all \( k \) and all \( t_{nk} \) in the range of summation. The term in the absolute value is bounded by

\[
|t_{nk}| - \varepsilon < |t_{nk} + O_n(1/\sqrt{n}) + o_n(1)| < |t_{nk}| + \varepsilon
\]

for any fixed positive \( \varepsilon \) and all sufficiently large values of \( n \). By the definition of proper Riemann integral, we see that

\[
\lim_{n \to \infty} \sum_{k: |t_{n,k}| < M} \frac{\phi(t_{nk})}{\sigma/\sqrt{n}} (|t_{nk}| \pm \varepsilon) = \int_{-M}^{M} (|t| \pm \varepsilon) \phi(t) \, dt.
\]

Since \( \varepsilon > 0 \) is arbitrarily small, we have

\[
\lim_{n \to \infty} \sum_{k: |t_{n,k}| < M} |\sigma/\sqrt{n} \mathbb{P}(S_n = k) - \sigma/\sqrt{n} \mathbb{P}(S_n = k - m)|
\]

\[
= \lim_{n \to \infty} m \sum_{k: |t_{n,k}| < M} \frac{\phi(t_{nk})}{\sigma/\sqrt{n}} |t_{nk} + O_n(1/\sqrt{n}) + o_n(1)|
\]

\[
= m \int_{-M}^{M} |t| \phi(t) \, dt.
\]

In other words, for any fixed large value of \( M \), we have

\[
\sum_{k: |t_{n,k}| < M} |\sigma/\sqrt{n} \mathbb{P}(S_n = k) - \sigma/\sqrt{n} \mathbb{P}(S_n = k - m)| = m \int_{-M}^{M} |t| \phi(t) \, dt + o_n(1).
\]

Also, for any fixed value of \( M \), we have already shown that

\[
\sum_{k: |t_{nk}| \geq M} |\sigma/\sqrt{n} \mathbb{P}(S_n = k - m) - \sigma/\sqrt{n} \mathbb{P}(S_n = k)|
\]

\[
\leq \sum_{k: |t_{nk}| \geq M} |\sigma/\sqrt{n} \mathbb{P}(S_n = k) - \phi(t_{nk})| + \sum_{k: |t_{nk}| \geq M} |\phi(t_{nk}) - \phi(t_{n,k-m})|
\]

\[
+ \sum_{k: |t_{nk}| \geq M} |\sigma/\sqrt{n} \mathbb{P}(S_n = k - m) - \phi(t_{n,k-m})|
\]

\[
= \frac{O_n(1)}{1 + M} + 2m\phi(M - 1).
\]
Letting $n$ go to infinity gives that
\[
\lim_{n \to \infty} \sum_{k} |\sigma \sqrt{n} \mathbb{P}(S_n = k) - \sigma \sqrt{n} \mathbb{P}(S_n = k - m)| = m \int_{-M}^{M} |t| \phi(t) \, dt + O(1/M) + O(\phi(M - 1)).
\]
The error terms on the right involving $M$ have upper bounds that go to zero as $M$ gets large. Indeed,
\[
\phi(M - 1) = \frac{1}{\sqrt{2\pi}} e^{-(M-1)^2/2} \to 0,
\]
as $M$ gets large. Hence, we have shown that
\[
\lim_{n \to \infty} \sum_{k} |\sigma \sqrt{n} \mathbb{P}(S_n = k) - \sigma \sqrt{n} \mathbb{P}(S_n = k - m)| = m \int_{-\infty}^{\infty} |t| \phi(t) \, dt = m \sqrt{\frac{2}{\pi}}.
\]
This then gives that
\[
\lim_{n \to \infty} \sqrt{n} |h_m(S_n) - \theta^m| = \lim_{n \to \infty} \sqrt{n} \theta^m \sum_{k} |\mathbb{P}(S_n = k) - \mathbb{P}(S_n = k - m)| = \lim_{n \to \infty} \theta^m \sum_{k} |\sqrt{n} \mathbb{P}(S_n = k) - \sqrt{n} \mathbb{P}(S_n = k - m)| = \lim_{n \to \infty} \frac{\theta^m}{\sigma} \sum_{k} |\sigma \sqrt{n} \mathbb{P}(S_n = k) - \sigma \sqrt{n} \mathbb{P}(S_n = k - m)| = \frac{m\theta^m}{\sigma} \sqrt{\frac{2}{\pi}}.
\]
This completes the proof. \qed

3.3 Asymptotic Efficiency

In section 3.1, we have shown that
\[
\sqrt{n} \left( \frac{A(n, S_n - m)}{A(n, S_n)} - \theta^m \right) \xrightarrow{\text{dist}} N \left( 0, \frac{m^2 \theta^{2m}}{\sigma^2} \right).
\]
In section 3.2, we have shown that
\[
\lim_{n \to \infty} \sqrt{n} \mathbb{E}_\phi |h_m(S_n) - \theta^m| = \frac{m\theta^m}{\sigma} \sqrt{\frac{2}{\pi}}.
\]
These two results suggest that a result concerning uniform integrability for the sequence
\[ \{(\sqrt{n}(h_m(S_n) - \theta))^r, r \geq 1\} \] may hold for some \( r \geq 1 \). We will show that indeed this is the case for \( 1 < r < 2 \). Unfortunately, the case \( r = 2 \) is still open. First, we will collect some preliminary results.

3.3.1 Odds Ratio

Before proceeding to the general case, it is worthwhile to consider the special case of odd’s ratio. We should particularly mention two techniques that are useful for this particular case. One is due to R. A. Khan and the other is due to Biancamaria Della Vecchia.

In the odd’s ratio case, \( a_0 = 1, a_1 = 1 \) and \( a_k = 0 \), for all \( k = 2, 3, \cdots \). This gives rise to the binomial density. That is,

\[ P(S_n = k) = \binom{n}{k} \frac{\theta^k}{(1 + \theta)^n}, \quad k = 0, 1, 2, \cdots, n. \]

Note that in our traditional notation,

\[ p = \frac{\theta}{1 + \theta}, \quad \text{or} \quad \theta = \frac{p}{1 - p} = \frac{p}{q} \]

is the usual odds ratio. So, the estimator of odds ratio is

\[ \frac{A(n, S_n - 1)}{A(n, S_n)} = \frac{\binom{n}{S_n - 1}}{\binom{n}{S_n}} = \frac{S_n}{n - S_n + 1}. \]

The question becomes, what is the limiting variance of this estimator? R. A. Khan (1988) and Della Vecchia (1991) answered this question in two different ways in the context of approximation theory. Both of the techniques seem to have some promise when our estimator simplifies out to something manageable. Unfortunately, odds ratio seems to be the only case when this happens. Here we use their results to obtain the limiting value of the variance.

R. A. Khan proved that

\[ \mathbb{E} \left( \frac{S_n}{n - S_n + 1} \right)^2 \]
\[
\begin{align*}
&= \theta^2 - \theta^2 \left( \frac{\theta}{1 + \theta} \right)^n - n\theta \left( \frac{\theta}{1 + \theta} \right)^n + \frac{\theta}{n(1 + \theta)^n} + R_n \\
&= \theta^2 + O \left( n \max \{1, (\theta/(1 + \theta))^n\} \right) + R_n,
\end{align*}
\]

where

\[
R_n = \sum_{k=0}^{n-2} \frac{(n+1)!p^{k+2}q^{n-k-2}}{(k+1)! (n-k)! (n-k-1)!}.
\]

We already know that

\[
\mathbb{E} \left( \frac{S_n}{n - S_n + 1} \right) = \theta - \theta p^n.
\]

Therefore, we see that

\[
n\mathbb{E} \left( \frac{S_n}{n - S_n + 1} - \theta \right)^2 = O(n^2(\theta/(1 + \theta))^n) + nR_n.
\]

Hence, we need only find the limiting value of \(nR_n\). For this purpose, we will first convert \(R_n\) into an appropriate expectation and then we will use the extended Lebesgue dominated convergence theorem. Note that

\[
nR_n = n \sum_{j=1}^{n-1} \binom{n+1}{j} \frac{p^{j+1}q^{n-j-1}}{(n-j)}
\]

\[
= \frac{p}{q^2} \sum_{j=1}^{n-1} \frac{n}{n-j} \binom{n+1}{j} p^j q^{n+1-j}
\]

\[
= \frac{p}{q^2} \sum_{j=1}^{n-1} \frac{n}{n-j} \mathbb{P}(T_{n+1} = j), \quad T_n \sim B(n, p)
\]

\[
= \frac{p}{q^2} \left\{ \sum_{j=1}^{n-1} \frac{n}{n - \min \{j, n-1\}} \mathbb{P}(T_{n+1} = j) \right\}
\]

\[
= \frac{p}{q^2} \left\{ \sum_{j=1}^{n+1} \frac{n}{n - \min \{j, n-1\}} \mathbb{P}(T_{n+1} = j) - n\mathbb{P}(T_{n+1} \geq n) \right\}
\]

\[
= \frac{p}{q^2} \left\{ \mathbb{E} \left( \frac{n}{n - \min \{T_{n+1}, n-1\}} \right) - n\mathbb{P}(T_{n+1} \geq n) \right\}.
\]

The last term goes to zero exponentially fast since

\[
n\mathbb{P}(T_{n+1} = n) = n(n+1)p^n q, \quad n\mathbb{P}(T_{n+1} = n+1) = np^n.
\]
So, we need only find the limiting value of

\[ \frac{p}{q^2} \mathbb{E} \left( \frac{n}{n - \min\{T_{n+1}, n - 1\}} \right). \]

Note that

\[ \frac{n}{n - \min\{T_{n+1}, n - 1\}} \leq \frac{3n}{(n + 1 - T_{n+1}) + 1}. \]

This can be verified for two cases: the case when \( T_{n+1} \geq n \) and the case when \( T_{n+1} < n \).

A result from Chao and Strawderman (1972) allows us to evaluate its expectation for the upper bound exactly. They showed that if \( U \sim B(n, r) \) then

\[ \mathbb{E} \left( \frac{1}{U + 1} \right) = \frac{1 - (1 - r)^{n+1}}{r(n + 1)}. \]

In our case, \( U = n + 1 - T_{n+1} \sim B(n + 1, q) \), Therefore,

\[ \mathbb{E} \left( \frac{3n}{(n + 1 - T_{n+1}) + 1} \right) = \frac{3n}{q(n + 2)} \rightarrow \frac{3}{q}. \]

Also, by the strong law of large numbers, we have

\[ \left( \frac{3n}{n + 1 - T_{n+1} + 1} \right) \xrightarrow{a.s.} \frac{3}{q}. \]

So if

\[ g_n(t) = \frac{3n}{(n + 1 - t) + 1} \]

then

\[ g_n(T_{n+1}) \xrightarrow{a.s.} \frac{3}{q} \quad \text{and} \quad \lim_{n \to \infty} \mathbb{E}g_n(T_{n+1}) = \frac{3}{q} = \mathbb{E} \left( \lim_{n \to \infty} g_n(T_{n+1}) \right). \]

That is, we can interchange the limit and the expectation operations for \( g_n \). Hence, we can interchange the limit and expectation operations for any sequence of functions which are bounded by \( g_n \). That is,

\[ \lim_{n \to \infty} nR_n = \frac{p}{q^2} \lim_{n \to \infty} \mathbb{E} \left( \frac{n}{n - \min\{T_{n+1}, n - 1\}} \right) \]

\[ = \frac{p}{q^2} \mathbb{E} \left( \lim_{n \to \infty} \frac{n}{n - \min\{T_{n+1}, n - 1\}} \right). \]
\[
= \frac{p}{q^2} \mathbb{E} \left( \lim_{n \to \infty} \frac{1}{1 - \min\{T_{n+1}/n, 1 - 1/n\}} \right) \\
= \frac{p}{q^2} \mathbb{E} \left( \frac{1}{1 - \min\{p, 1\}} \right) \\
= \frac{p}{q^3} = \frac{\theta}{1 + \theta} (1 + \theta)^3 = \theta(1 + \theta)^2 = \frac{\theta^2}{\sigma^2}.
\]

So in the case of the odd’s ratio,

\[
\lim_{n \to \infty} n\mathbb{E}(h_n(S_n) - \theta)^2 = \frac{\theta^2}{\sigma^2}.
\]

Hence, \( \{n(h_n(S_n) - \theta)^2, \ n \geq 1\} \) is a collection of uniformly integrable random variables.

3.3.2 General Case

For the general case, we can state the following theorem.

**Theorem 4.** Let \( G(\theta) = a_0 + a_1 \theta + a_2 \theta^2 + \ldots \) be a function with radius of convergence \( R > 0 \) and \( a_k > 0 \ \forall k \geq \kappa \in \mathbb{N}_0 \) and \( a_k = 0, \ \forall k < \kappa \). Let \( X_1, X_2, \ldots \overset{iid}{\sim} PSD(\theta) \). When \( \sigma^2 = \text{Var}_\theta(X_1) < \infty \), if \( \text{Var} \left( \frac{\alpha - \mu}{\sigma \alpha} \right) \) exists for \( \theta \in (0, R) \) and

\[
\sqrt{n} \frac{\mathbb{P}(S_n - k - 1) - \mathbb{P}(S_n - k)}{\mathbb{P}(S_n = k)}
\]

is bounded by a measurable function on \((0, R)\) for all \( n \in \mathbb{N} \) and \( k \geq \kappa n \), then for any \( \theta \in (0, R) \),

\[
\lim_{n \to \infty} n\mathbb{E}_{\theta}(h(S_n) - \theta)^2 = \frac{\theta^2}{\sigma^2}.
\]

**Proof.** By expanding the left hand side, we see that we need only show that

\[
n \left( \sum_{k \geq \kappa n} \frac{A(n, k - 1)^2}{A(n, k)} \frac{\theta^k}{(G(\theta))^n} - \theta^2 \right) \to \frac{\theta^2}{\sigma^2}
\]

We can rewrite the summation as follows

\[
= \sum_{k \geq \kappa n} \left( \frac{A(n, k - 1)}{A(n, k)} \right) \left( \frac{A(n, k - 1)\theta^{k-1}}{(G(\theta))^n} \right) \theta - \theta^2
\]
\[ \begin{align*}
&= \sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1)}{\mathbb{P}(S_n = k)} \right) \mathbb{P}(S_n = k - 1) - \theta^2. \\
&= \theta^2 \left( \sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1)}{\mathbb{P}(S_n = k)} \right) \mathbb{P}(S_n = k - 1) - 1 \right).
\end{align*} \]

Now, \( \sum_{k \geq n\kappa} \mathbb{P}(S_n = k - 1) = 1 \), so we can write

\[ \begin{align*}
\theta^2 &\left( \sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1)}{\mathbb{P}(S_n = k)} \right) \mathbb{P}(S_n = k - 1) - 1 \right) \\
&= \theta^2 \sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1)}{\mathbb{P}(S_n = k)} - 1 \right) \mathbb{P}(S_n = k - 1) \\
&= \theta^2 \sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)} \right) \mathbb{P}(S_n = k - 1) \\
\end{align*} \]

Hence, we need only show that

\[ \begin{align*}
n \sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)} \right) \mathbb{P}(S_n = k - 1) \to \frac{1}{\sigma^2}.
\end{align*} \]

Now, note that

\[ \begin{align*}
n &\sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)} \right) \mathbb{P}(S_n = k - 1) \\
&= n \sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)} \right) \{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)\} \\
&\quad + n \sum_{k \geq n\kappa} \left( \frac{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)} \right) \mathbb{P}(S_n = k) \\
&= n \sum_{k \geq n\kappa} \frac{\{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)\}^2}{\mathbb{P}(S_n = k)} \\
&\quad + n \sum_{k \geq n\kappa} (\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)).
\end{align*} \]

The last term is zero

\[ n \sum_{k \geq n\kappa} (\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)) = n\mathbb{P}(S_n = n\kappa - 1) = n0 \]
So, we need to show that

\[
\begin{align*}
    &\sum_{k : A(n,k) > 0} \frac{\{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)\}^2}{\mathbb{P}(S_n = k)} \\
    &= \frac{1}{\sigma^2} \sum_{k : A(n,k) > 0} \left( \frac{\sigma \sqrt{n} \mathbb{P}(S_n = k - 1) - \sigma \sqrt{n} \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)} \right)^2 \mathbb{P}(S_n = k) \\
    &= \frac{1}{\sigma^2} \mathbb{E} \left( \frac{\sigma \sqrt{n} \mathbb{P}(T_n = S_n - 1|S_n) - \sigma \sqrt{n} \mathbb{P}(T_n = S_n|S_n)}{\mathbb{P}(T_n = S_n|S_n)} \right)^2 \mathbb{P}(S_n = k)
\end{align*}
\]

goes to \( \frac{1}{\sigma^2} \) as \( n \) gets large.

We know that

\[
\frac{\sigma \sqrt{n} \mathbb{P}(T_n = S_n - 1|S_n) - \sigma \sqrt{n} \mathbb{P}(T_n = S_n|S_n)}{\mathbb{P}(T_n = S_n|S_n)} \xrightarrow{\text{dist}} Z \sim N(0, 1)
\]

and that

\[
\frac{\sigma \sqrt{n} \mathbb{P}(S_n = k - 1) - \sigma \sqrt{n} \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)} = \frac{\sigma \sqrt{n} \mathbb{P}(T_n = S_n - 1|S_n) - \sigma \sqrt{n} \mathbb{P}(T_n = S_n|S_n)}{\mathbb{P}(T_n = S_n|S_n)}
\]

is bounded, so Lebesgue dominated convergence gives

\[
\mathbb{E} \left( \frac{\sigma \sqrt{n} \mathbb{P}(T_n = S_n - 1|S_n) - \sigma \sqrt{n} \mathbb{P}(T_n = S_n|S_n)}{\mathbb{P}(T_n = S_n|S_n)} \right)^2 = \mathbb{E}(Z)^2 = 1
\]

\( \square \)

The problem in the general case is trying to find a bound for

\[
\frac{\sqrt{n} \mathbb{P}(S_n = k - 1) - \sqrt{n} \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)} = \sqrt{n} \left( \frac{\mathbb{P}(S_n = k - 1)}{\mathbb{P}(S_n = k)} - 1 \right)
\]

Looking at this expression a different way, by Taylor's theorem,

\[
\frac{\mathbb{P}(S_n = k - 1)}{\mathbb{P}(S_n = k)} = \frac{A_{n,k}^{-1}}{A_{n,k}} \frac{\partial^{k-1}}{(\partial G(\theta))^{k-1}} + \frac{A_{n,k}^{-1}}{A_{n,k}} \frac{\partial^k}{(G(\theta))^{k}}
\]

\[
= \frac{((G(\theta))^{(k-1)}(0))}{(k-1)!} \frac{1}{(G(\theta))^{(k)}(0)} \theta
\]
\begin{align*}
&= k \left( \frac{(G(\theta)^n (k-1))}{(G(\theta)^n (k))} \right) \frac{1}{\theta} \\

\text{This means we need to find a bound for}

\sqrt{n} \left( k \left( \frac{(G(\theta)^n (k-1))}{(G(\theta)^n (k))} \right) \frac{1}{\theta} - 1 \right)

\text{that holds for all } n, k \in \mathbb{N}.

\text{Without making any assumptions about this term, we can state the following theorem when } h(S_n) \text{ is the U MVU estimator for } \theta.

\textbf{Theorem 5.} If } h(S_n) \text{ is the U MVU estimator of } \theta \text{ and if } Y = \frac{a(X_i - 1)}{a(X)} \text{ has finite variance, the for any } 1 \leq r < 2, \text{ we have}

\lim_{n \to \infty} \mathbb{E} (\sigma \sqrt{n} |h(S_n) - \theta|)^r = \frac{\theta^r}{\sigma^r} \int_{-\infty}^{\infty} |t|^r \phi(t) \, dt.

\text{Therefore, } (\sigma \sqrt{n} |h(S_n) - \theta|)^r \text{ with } n \geq 1 \text{ is uniformly integrable.}

\textbf{Proof.} Recall the Plackett estimator } \bar{Y}_n = \sum_{i=1}^{n} Y_i \text{ where}

Y_i = \frac{a(X_i - 1)}{a(X_i)}, \quad i = 1, 2, \ldots, n.

\text{If its variance is}

\text{Var}(\bar{Y}_n) = \frac{\tau^2(\theta)}{n},

\text{then the U MVU estimator of } \theta \text{ should have lower variance. That is,}

\mathbb{E}(h(S_n) - \theta)^2 \leq \text{Var}(\bar{Y}_n) = \frac{\tau^2(\theta)}{n}.

\text{Taking } n \text{ to the other side, we get that}

\mathbb{E} (\sigma \sqrt{n} (h(S_n) - \theta))^2 \leq \tau^2 \sigma^2, \quad \text{for all } n = 1, 2, \ldots.

\text{For } 1 \leq r < 2, \text{ take a positive number } \alpha \text{ so that } ra = 2, \text{ that is, } \alpha = \frac{2}{r} > 1. \text{ Then find a } \beta > 1 \text{ so that}

\frac{1}{\alpha} + \frac{1}{\beta} = 1.
That is, \( \beta = \frac{\alpha}{\alpha - 1} > 1 \). Let \( T = \frac{S_n - n\mu}{\sigma \sqrt{n}} \), for any positive number \( M \), the Hölder inequality gives that

\[
\sum_{k: |t| > M} (\sigma \sqrt{n}(h(k) - \theta))^r \mathbb{P}(S_n = k) = \mathbb{E} \left( \frac{\sigma \sqrt{n}(h(S_n) - \theta)}{\mathbb{E}(X_{(|T| > M})^2} \right)^r \\
\leq \left\{ \mathbb{E} \left( \frac{\sigma \sqrt{n}(h(S_n) - \theta)}{\mathbb{E}(X_{(|T| > M})^2} \right)^r \right\}^{1/\alpha} \cdot \left\{ \mathbb{E} \left( \mathbb{E}(X_{(|T| > M}) \right)^{1/\beta} \\
= \left\{ \frac{r^2}{\sigma^2} \right\}^{1/\alpha} \cdot \left\{ \mathbb{E} \left( \mathbb{E}(X_{(|T| > M}) \right)^{1/\beta} \\
= \left\{ \frac{r^2}{\sigma^2} \right\}^{1/\alpha} \cdot \left\{ \mathbb{P} \left( \frac{|S_n - n\mu|}{\sigma \sqrt{n}} > M \right) \right\}^{1/\beta} .
\]

By the central limit theorem, we have

\[
\limsup_n \sum_{k: |t| > M} (\sigma \sqrt{n}(h_n(k) - \theta))^r \mathbb{P}(S_n = k) \leq K(\theta, r) \left\{ \int_{|t| > M} \phi(t) \, dt \right\}^{1/\beta}
\]

where \( K(\theta, r) \) is a positive constant. The result follows. \( \square \)
CHAPTER 4

Applications

4.1 Inference in Censored PSD

4.1.1 Introduction

Let $X$ have the power series distribution generated by the sequence $\{a_k\}$. Let $C$ be a nonempty proper subset of the nonnegative integers. We say that $C$ is a censoring set if observations of the random variable that fall in $C$ cannot be distinguished from each other. Let $X^*$ denote the censored version of the random variable. Pick any fixed value, say $c$, from the set $C$ and assign that value to $X^*$ if $X$ falls in $C$. The density of $X^*$ is

$$
\mathbb{P}(X^* = k) = \begin{cases} 
\frac{a_k \theta^k}{G(\theta)} & \text{if } k \notin C, \\
\mathbb{P}(X \in C) =: p & \text{if } k = c.
\end{cases}
$$

Hence, we may write the density of $X^*$ as follows.

$$
\mathbb{P}(X^* = k) = \left(\frac{a_k \theta^k}{G(\theta)}\right)^{\chi(k \notin C)} p^{\chi(k \in C)}.
$$

Let the observations of $X$ that we know did not get censored be denoted by $X_i^*$, i.e., $X_i^*$ is the observation that we know did not get censored. Hence, for a random sample of $n'$ observations of $X$, the joint density of $X_1, \ldots, X_n$ is

$$
\prod_{i=1}^{n'} \left(\frac{a(x_i)\theta^{x_i}}{G(\theta)}\right)^{\chi(x_i \notin C)} p^{\chi(x_i \in C)} = \left(\frac{\prod_{i=1}^{n'} a(x_i^*)}{G(\theta)^n}\right) p^{N_c},
$$

where $N_c$ is the number of observations that were censored and $n = n' - N_c$. This shows that the sufficient statistic is

$$
\left(\sum_{i=1}^{n'} X_i \chi(x_i \notin C), \ N_c\right),
$$

51
or equivalently
\[
\left(\sum_{i=1}^{n'} X_i X_i I(x_i \notin C), n\right).
\]

To find the conditional distribution of \(X_1\) when we are given that it was not censored, note that
\[
\mathbb{P}(X_1 = k|X_1 \notin C) = \begin{cases} 
\frac{\mathbb{P}(X_1 = k \cap X_1 \notin C)}{\mathbb{P}(X_1 \notin C)} & \text{if } k \notin C, \\
0 & \text{if } k \in C,
\end{cases}
\]
\[
= \begin{cases} 
\frac{\mathbb{P}(X_1 = k)}{\mathbb{P}(X_1 \notin C)} & \text{if } k \notin C, \\
0 & \text{if } k \in C,
\end{cases}
\]
\[
= \begin{cases} 
\frac{a_k \theta^k}{\mathbb{P}(X_1 \notin C) \theta^k} & \text{if } k \notin C, \\
0 & \text{if } k \in C,
\end{cases}
\]

Hence, the conditional distribution of \(X_1\) given that it was not censored is the \(C\)-truncated distribution of \(X_1\).

Define the sequence \(\{b_k\}\) by \(b_k = a_k\) if \(k \in C\) and \(b_k = 0\) if \(k \in C\). Let \(S^*_n\) denote the sum of those observations that did not get censored, that is,
\[
S^*_n = \sum_{i=1}^{n'} X_i X_i I(x_i \notin C).
\]

Consider the estimator
\[
h(S^*_n) := \frac{A(n, S^*_n - 1)}{A(n, S^*_n)},
\]
where \(A(n, k)\) is the \(n\)-fold Cauchy product of the sequence \(\{b_k\}\). Hence, with the usual \(\theta^0\) convention, the estimator \(h(S^*_n)\) is well defined. The censoring set \(C\) could be the left tail, the right tail, or both, or even a mid interval.

**Proposition 8.** If the estimator \(h(S_n)\) is unbiased for \(\theta\) in the noncensored version then the bias of the estimator \(h(S^*_n)\) goes to zero exponentially fast in the censored version.

**Proof:** Let \(B\) to denote the information (sigma field) generated by
\[
X_{(X_1 \notin C)}, X_{(X_2 \notin C)}, \ldots, X_{(X_{n'} \notin C)}.
\]
and let $I$ denote the information (sigma field) generated by the indicator function $\chi_{(n>0)}$.

So either $I = 0$, which happens with probability $p^{n'}$, or $I = 1$. The theorem of total expectations gives

$$
\mathbb{E}(h(S^*_n)) = \mathbb{E}\left\{\mathbb{E}\left(h(S^*_n)\mid \chi_{(X_1 \notin C)} \chi_{(X_2 \notin C)} \cdots \chi_{(X_n \notin C)}, n\right)\right\}
$$

$$
= 0 + \mathbb{E}\{\mathbb{E}(h(S^*_n)\mid B \cap \{I = 1\})\}.
$$

The event $\{I = 1\}$ consists of $n'$ number of mutually exclusive sets, namely when $n = 1, n = 2, \ldots , n = n'$. Denote their probabilities by $\mathbb{P}(1), \mathbb{P}(2), \ldots , \mathbb{P}(n')$.

Since the conditional density of $X_i$ given $\chi_{(X_i \notin C)} = 1$ has the same density as the $C$-truncated PSD, when $n$ is a positive integer, $h(S^*_n)$ has the same density as $h(T_n)$ where $T_n$ is the sum of $n$ iid $C$-truncated random variables. However, if $n = 0$, which happens with probability $p^{n'}$, $h(S^*_n) = 0$. Consequently,

$$
\mathbb{E}(h(S^*_n)) = \sum_{i=1}^{n'} \mathbb{P}(i)\mathbb{E}\{\mathbb{E}(h_i(T_i)\mid B \cap \{n = i\})\}
$$

$$
= \sum_{i=1}^{n'} \mathbb{P}(i)\mathbb{E}\{\theta \mid B \cap \{n = i\}\}
$$

$$
= \theta \sum_{n=i}^{n'} \mathbb{P}(i)
$$

$$
= \theta(1 - p^{n'}).
$$

4.1.2 Estimation for Right Tail Censored Data

Let $X_1, X_2, \cdots , X_{n'}, iid PSD(\theta)$. Some of these observations may fall into the a censoring set $C = \{r + 1, r + 2, \cdots \}$ and therefore will be considered censored (that is, we cannot use their actual values but we can use the information contained in the number of such observations). The sample size $n'$ is a fixed positive integer, and the number of observations which did not get censored, denoted by $n$, is a random variable. Define

$$
N_i = \text{number of } X_i \text{ that are equal to } i, \ i = 0, 1, 2, \cdots , r,
$$
with \( N_{r+1} = n' - \sum_{i=0}^{r} N_i \). In this setup, note that

\[
N_i \sim B(n', p_i), \quad p_i = \frac{a_i \theta^i}{G(\theta)},
\]
i = 0, 1, \ldots, r, and

\[
N_{r+1} \sim B(n', p^*), \quad p^* = \sum_{i=r+1}^{\infty} \frac{a_i \theta^i}{G(\theta)}.
\]

Instead of estimating \( \theta \), we consider estimating \( \theta^m \) for some given positive integer \( m < r \).

First note that

\[
\sum_{i=0}^{r} \frac{a(i-m)}{a(i)} \frac{a_i \theta^i}{G(\theta)} = \sum_{i=m}^{r} \frac{a(i-m)}{a(i)} \frac{a_i \theta^i}{G(\theta)} = \theta^m \sum_{i=m}^{r} \frac{a(i-m) \theta^{i-m}}{G(\theta)} = \theta^m \sum_{j=0}^{r-m} \frac{a_j \theta^j}{G(\theta)}.
\]

So

\[
\theta^m = \frac{\sum_{i=m}^{r} \frac{a(i-m)}{a(i)} \frac{a_i \theta^i}{G(\theta)}}{\sum_{j=0}^{r-m} \frac{a_j \theta^j}{G(\theta)}} = \frac{\sum_{i=m}^{r} a(i-m)}{\sum_{j=0}^{r-m} \mathbb{E}(N_j)} \frac{\mathbb{E}(N_i)}{\mathbb{E}(\sum_{j=0}^{r-m} N_j)}.
\]

That is, if \( g(\theta) = \theta^m \), then its estimator is

\[
\hat{g} := \left( \sum_{i=m}^{r} a(i-m) \frac{N_i}{\mathbb{E}(N_j)} \right) / \left( \sum_{j=0}^{r-m} N_j \right).
\]

Unfortunately, this estimator is that its denominator can equal zero with positive probability. Instead, we define the estimator of \( g(\theta) = \theta^m \) to be

\[
\tilde{g} := \frac{\sum_{i=m}^{r} a(i-m) N_i}{1 + \sum_{j=0}^{r-m} N_j}.
\]

This estimator is extremely easy to calculate when compared to the maximum likelihood estimator which requires a numerical solution of an equation. The important thing to note is that in this situation there is no unbiased estimator due to the censoring of the right tail.

Before proving a few facts about this estimator, first note a few things.

\[
\sum_{j=0}^{r-m} N_j = \sum_{j=0}^{r} N_j - \sum_{j=r-m+1}^{r} N_j = n - \sum_{j=r-m+1}^{r} N_j.
\]
where $n$ is precisely the number of terms which were not censored. Also, if we take $a_i = 0$, for $i = -m, -m + 1, \ldots, -1$, then

$$\sum_{i=m}^{r} \frac{a(i - m)}{a(i)} N_i = \sum_{i=0}^{r} \frac{a(i - m)}{a(i)} N_i.$$

And

$$\sum_{i=0}^{r} \frac{a(i - m)}{a(i)} N_i = \sum_{k=1, X_k \leq r}^{n'} \frac{a(X_k - m)}{a(X_k)} \chi_{\{X_k \leq r\}}.$$

This is because $N_i$ number of $X_k$'s will be contributing to the left sum for each value of $i$ and they will all contribute the same amount. Each of these $X_k$'s will give the amount $a(X_k - m)/a(X_k)$. Hence, all these $X_k$'s will contribute the correct total amount in the sum on the right side. Then summing over all possible values of $i$, namely $i = 0, 1, \ldots, r$, exhausts all the possibilities. Therefore, the estimator is the same as

$$\frac{1}{1 + \sum_{j=0}^{r-m} N_j} \sum_{k=1, X_k \leq r}^{n'} \frac{a(X_k - m)}{a(X_k)} + \frac{1}{1 + \sum_{j=0}^{r-m} N_j} \sum_{k=1}^{n'} \frac{a(X_k - m)}{a(X_k)} \chi_{\{X_k \leq r\}}.$$

Note that in the truncated case, the indicator terms all become 1.

**Proposition 9.** For $g(\theta) = \theta^m$,

$$\mathbb{E}(\hat{\theta}) = \theta^m + \frac{1}{(n' + 1)q} \left\{ \sum_{j=r-m+1}^{r} \frac{a(j - m)p_j}{a(j)q} - \theta^m \right\} + \frac{\theta^m s^{n' + 1}}{(n' + 1)q} \frac{1 + s}{1 + \frac{s}{(n' + 1)q}} \sum_{j=r-m+1}^{r} \frac{a(j - m)p_j}{a(j)}.
$$

where $1 - s = q = p_0 + p_1 + \cdots + p_{r-m}$.

**Proof.** First,

$$(N_0, N_1, \ldots, N_r, N_{r+1}) \sim \text{Multinomial}(n', p_0, p_1, \cdots, p_r, p_{r+1}).$$

Let $M = N_0 + N_1 + \cdots + N_{r-m}$. For any $j > r - m$,

$$(N_j, M) \sim \text{Trinomial}(n', p_j, q), \quad \text{where } q = \sum_{j=0}^{r-m} p_j.$$
We know that, when \( j > r - m \), the conditional distribution of \( N_j \) given \( M \) is \( B(n' - M, \frac{p_j}{1 - q}) \).

Hence, the expectation of the \( j \)-th term in the estimator is

\[
\mathbb{E} \left( \frac{N_j}{\sum_{j=0}^{r-m} N_j + 1} \right) = \mathbb{E} \left( \frac{N_j}{M + 1} | M \right) \\
= \mathbb{E} \left( \frac{1}{M + 1} \mathbb{E} (N_j | M) \right) \\
= \mathbb{E} \left( \frac{1}{M + 1} (n' - M) p_j \right) \\
= \frac{p_j}{1 - q} \mathbb{E} \left( \frac{n' - M}{M + 1} \right) \\
= \frac{p_j}{1 - q} \left( \mathbb{E} \left( \frac{n' + 1}{M + 1} \right) - 1 \right) \\
= \frac{p_j}{1 - q} \left( (n' + 1) \frac{1 - (1 - q)^{n' + 1}}{q(n' + 1)} - 1 \right) \\
= \frac{p_j}{1 - q} \left( \frac{1 - (1 - q)^{n' + 1}}{q} - 1 \right) \\
= \frac{p_j}{q(1 - q)} \left( (1 - q) - (1 - q)^{n' + 1} \right) \\
= \frac{p_j}{q} \left( 1 - (1 - q)^{n'} \right).
\]

The \( i \)-th terms, \( i \leq r - m \), in the estimator are done in a similar fashion. Let \( Z = n' - M = N_{r-m+1} + \cdots + N_r + N_{r+1} \). Then for any \( i \in \{0, 1, \cdots, r - m\} \),

\((N_i, Z) \sim \text{Trinomial}(n', p_i, s), \) where \( s = \sum_{j \geq r - m + 1} p_j = 1 - q. \)

We want

\[
E \left( \frac{N_i}{M + 1} \right) = E \left( \frac{N_i}{n' - Z + 1} \right).
\]

The conditional distribution of \( N_i \) given \( Z \) is \( B \left( n' - Z, \frac{p_i}{1 - s} \right) \). Therefore,

\[
\mathbb{E} \left( \frac{N_i}{n' - Z + 1} \right) = \mathbb{E} \left( \frac{N_i}{n' - Z + 1} | Z \right) \\
= \mathbb{E} \left( \frac{1}{n' - Z + 1} \mathbb{E} (N_i | Z) \right) \\
= \mathbb{E} \left( \frac{1}{n' - Z + 1} \frac{(n' - Z)p_i}{1 - s} \right).
\]
\[ \begin{align*}
\mathbb{E}(q) & = \frac{p_i}{1-s} \mathbb{E} \left\{ \frac{n' - Z}{n' - Z + 1} \right\} \\
& \quad + \frac{p_i}{1-s} \mathbb{E} \left\{ 1 - \frac{1}{n' - Z + 1} \right\} \\
& \quad + \frac{p_i}{1-s} \left( 1 - \frac{1 - s^{n'+1}}{(n' + 1)(1-s)} \right) \\
& \quad + \frac{p_i}{q} \left( 1 - \frac{1 - s^{n'+1}}{(n' + 1)q} \right).
\end{align*} \]

This gives

\[ \begin{align*}
\mathbb{E}(q) & = \sum_{i=0}^{r-m} \frac{a(i-m)p_i}{a(i)q} \left( 1 - \frac{1 - s^{n'+1}}{(n' + 1)q} \right) + \sum_{j=r-m+1}^{r} \frac{a(j-m)p_j}{a(j)q} \left( 1 - s^{n'} \right) \\
& \quad - \sum_{j=r-m+1}^{r} \frac{a(j-m)p_j}{a(j)q} \left( 1 - \frac{1 - s^{n'+1}}{(n' + 1)q} \right) \\
& \quad + \theta^m \left( 1 - \frac{1 - s^{n'+1}}{(n' + 1)q} \right) + \sum_{j=r-m+1}^{r} \frac{a(j-m)p_j}{a(j)q} \left( -s^{n'} + \frac{1 - s^{n'+1}}{(n' + 1)q} \right) \\
& \quad + \theta^m + \frac{1}{(n' + 1)q} \sum_{j=r-m+1}^{r} \frac{a(j-m)p_j}{a(j)q} - \theta^m \frac{1 - s^{n'+1}}{(n' + 1)q} \\
& \quad - \frac{s^{n'}}{q} \left( 1 + \frac{s}{(n' + 1)q} \right) \sum_{j=r-m+1}^{r} \frac{a(j-m)p_j}{a(j)} \\
& \quad + \theta^m + \frac{1}{(n' + 1)q} \left\{ \sum_{j=r-m+1}^{r} \frac{a(j-m)p_j}{a(j)q} - \theta^m \right\} + \frac{\theta^m s^{n'+1}}{(n' + 1)q} \\
& \quad - \frac{s^{n'}}{q} \left( 1 + \frac{s}{(n' + 1)q} \right) \sum_{j=r-m+1}^{r} \frac{a(j-m)p_j}{a(j)}.
\end{align*} \]

This ends the proof. \(\square\)

**Remark.** For the Poisson case with \(m = 1\), Moore gave the approximation to the expectation
of his estimator to be
\[
\theta + \frac{1}{n'} \left( \frac{rp_r}{q^2} - \frac{\theta(1 - q)}{q} \right),
\]
which is quite close to ours.

**Proposition 10.** The limiting distribution of the estimator \( \hat{g} \) is
\[
\sqrt{n'}(\hat{g} - \theta_m) \xrightarrow{dist} N(0, \xi^2(\theta)),
\]
where
\[
\xi^2(\theta) = \frac{1}{q^2} \left\{ \sum_{i=1}^{r-m} (b_i - \theta_m)^2 p_i + \sum_{i=r-m+1}^{r} b_i^2 p_i \right\},
\]
where \( b_i = \frac{a(i-m)}{a(i)} \) and \( q = \sum_{j=0}^{r-m} p_j \).

**Proof.** First
\[
\sqrt{n'}(\hat{g} - \theta_m) = \frac{\sqrt{n'}}{\sum_{j=0}^{r-m} N_j + 1} \left( \sum_{i=0}^{r} \frac{a(i-m)}{a(i)} N_i - \theta_m \left( \sum_{j=0}^{r-m} N_j + 1 \right) \right)
\]
\[
= \frac{\sqrt{n'}}{\sum_{j=0}^{r-m} \frac{N_j}{n'} + \frac{1}{n'}} \left( \sum_{i=0}^{r} \frac{a(i-m)}{a(i)} \frac{N_i}{n'} - \theta_m \left( \sum_{j=0}^{r-m} \frac{N_j}{n'} + \frac{1}{n'} \right) \right)
\]
\[
\sim A_n \sqrt{n'} \left( \sum_{i=0}^{r} \frac{a(i-m)}{a(i)} \frac{N_i}{n'} - \theta_m \left( \sum_{j=0}^{r-m} \frac{N_j}{n'} + \frac{1}{n'} \right) \right),
\]
where \( A_n \) is a random variable which converges to \( (\sum_{j=0}^{r-m} p_j)^{-1} = \frac{1}{q} \) almost surely. Therefore, by Slutsky’s theorem we need only find the limiting distribution of
\[
\frac{\sqrt{n'}}{\sum_{j=0}^{r-m} p_j} \left( \sum_{i=0}^{r} \frac{a(i-m)}{a(i)} \frac{N_i}{n'} - \theta_m \left( \sum_{j=0}^{r-m} \frac{N_j}{n'} + \frac{1}{n'} \right) \right).
\]
This is a linear combination of the components of a multinomial random vector. By the multivariate normal central limit theorem, we know that
\[
\left( \begin{array}{c}
\sqrt{n'}((N_0/n') - p_0) \\
\sqrt{n'}((N_1/n') - p_1) \\
\vdots \\
\sqrt{n'}((N_r/n') - p_r)
\end{array} \right) \xrightarrow{dist} MVN(0, \Sigma),
\]
where $\Sigma = [\sigma_{ij}]$ is the $(r + 1) \times (r + 1)$ variance covariance matrix consisting of

$$\sigma_{ii} = \text{Var}(\sqrt{n'}((N_i/n') - p_i) = p_i(1 - p_i), \quad i = 0, 1, \ldots, r.$$ 

And for $i \neq j$,

$$\sigma_{ij} = \frac{\text{Cov}(N_i, N_j)}{n'} = -p_ip_j.$$ 

Let $q = \sum_{j=0}^{r-m} p_j$. We may safely throw away the last $\theta^m/n'$ term in 4.1.2 since when it is multiplied by $\sqrt{n'}$ it still goes to zero by the rate of $1/\sqrt{n'}$. So, we need the limiting distribution of

$$\frac{\sqrt{n'}}{q} \left( \sum_{i=0}^{r} \frac{a(i - m)}{a(i)} \frac{N_i}{n'} - \theta^m \sum_{j=0}^{r-m} \frac{N_j}{n'} \right),$$

Its expectation and its variance will be the mean and variance of the limiting random variable. Note that its mean is zero. Indeed, the expectation is

$$\frac{\sqrt{n'}}{q} \left( \sum_{i=0}^{r} \frac{a(i - m)}{a(i)} p_i - \theta^m \sum_{j=0}^{r-m} p_j \right) = 0.$$ 

Now we find the variance. To save space we will denote $\frac{a(i - m)}{a(i)}$ by the symbol $b_i$.

$$\text{Var} \left\{ \frac{\sqrt{n'}}{q} \left( \sum_{i=0}^{r} b_i \frac{N_i}{n'} - \theta^m \sum_{j=0}^{r-m} \frac{N_j}{n'} \right) \right\}$$

$$= \frac{n'}{q^2} \left( \text{Var} \left( \sum_{i=0}^{r} b_i \frac{N_i}{n'} \right) + \theta^{2m} \text{Var} \left( \sum_{j=0}^{r-m} \frac{N_j}{n'} \right) - 2\theta^m \text{Cov} \left( \sum_{i=0}^{r} b_i \frac{N_i}{n'}, \sum_{j=0}^{r-m} \frac{N_j}{n'} \right) \right).$$

We take care of these terms one at a time. Note that

$$\text{Var} \left( \sum_{i=0}^{r} b_i \frac{N_i}{n'} \right) = \sum_{i=0}^{r} b_i^2 p_i(1 - p_i) + \sum_{i \neq j}^{r} b_i b_j - p_i p_j.$$ 

Also, since $\sum_{j=0}^{r-m} N_j \sim B(n', q)$, we see that

$$\text{Var} \left( \sum_{j=0}^{r-m} \frac{N_j}{n'} \right) = \frac{q(1 - q)}{n'}.$$ 

Finally,

$$\text{Cov} \left( \sum_{i=0}^{r} b_i \frac{N_i}{n'}, \sum_{j=0}^{r-m} \frac{N_j}{n'} \right)$$
\[
\begin{align*}
&= \sum_{i=0}^{r} \sum_{j=0}^{r-m} b_i \frac{Cov(N_i, N_j)}{(n')^2} \\
&= \sum_{i=r-m+1}^{r} b_i \sum_{j=0}^{r-m} \frac{Cov(N_i, N_j)}{(n')^2} + \sum_{i=0, i \neq j}^{r-m} b_i \sum_{j=0}^{r-m} \frac{Cov(N_i, N_j)}{(n')^2} + \sum_{i=0}^{r-m} b_i \frac{Cov(N_i, N_i)}{(n')^2} \\
&= \sum_{i=r-m+1}^{r} b_i \sum_{j=0}^{r-m} \frac{-p_ip_j}{n'} + \sum_{i=0, i \neq j}^{r-m} b_i \sum_{j=0}^{r-m} \frac{-p_ip_j}{n'} + \sum_{i=0}^{r-m} b_i p_i (1 - p_i). \\
\end{align*}
\]

Plugging these expressions back, we get the asymptotic variance to be

\[
\frac{1}{q^2} \left\{ \sum_{i=0}^{r} b_i^2 p_i (1 - p_i) - \sum_{i \neq j}^{r} b_i b_j p_i p_j + \theta^{2m} q (1 - q) \right\} \\
- 2\theta^m \left[ - \sum_{i=r-m+1}^{r} b_i p_i \sum_{j=0}^{r-m} p_j - \sum_{i=0, i \neq j}^{r-m} b_i p_i p_j + \sum_{i=0}^{r-m} b_i p_i (1 - p_i) \right].
\]

We can simplify this intimidating looking expression. Keep in mind that \( m \) is a positive integer and \( b_i = 0 \) for all \( i < m \). So, we may start the subscripts at \( i = 1 \) or \( i = 0 \) in those sums which involve \( b_i \) as a multiple. The above expression is equal to

\[
\frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \sum_{i=1}^{r} (b_i p_i)^2 - \sum_{i \neq j}^{r} b_i b_j p_i p_j + \theta^{2m} q (1 - q) \right\} \\
+ 2\theta^m \left[ q \sum_{i=r-m+1}^{r} b_i p_i + \sum_{i=0}^{r-m} b_i p_i (q - p_i) - \sum_{i=0}^{r-m} b_i p_i + \sum_{i=0}^{r-m} b_i p_i^2 \right]
\]

\[
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 + \theta^{2m} q (1 - q) \right\} \\
+ 2\theta^m \left[ q \sum_{i=r-m+1}^{r} b_i p_i + q \sum_{i=0}^{r-m} b_i p_i - \sum_{i=0}^{r-m} b_i p_i^2 - \sum_{i=0}^{r-m} b_i p_i + \sum_{i=0}^{r-m} b_i p_i^2 \right]
\]

\[
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 + \theta^{2m} q (1 - q) \right\} \\
+ 2\theta^m \left[ q \sum_{i=r-m+1}^{r} b_i p_i - (1 - q) \sum_{i=0}^{r-m} b_i p_i \right]
\]

\[
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 + \theta^{2m} q (1 - q) \right\} \\
+ 2\theta^m \left[ q \sum_{i=r-m+1}^{r} b_i p_i - (1 - q) \left( \sum_{i=0}^{r-m} b_i p_i - \sum_{i=r-m+1}^{r} b_i p_i \right) \right]
\]
\[
\frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 + \theta^{2m} q (1 - q) \right. \\
+ 2\theta^m \left\{ q \sum_{i=r-m+1}^{r} b_i p_i - (1 - q) \left( \theta^m q - \sum_{i=r-m+1}^{r} b_i p_i \right) \right\} \\
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 + \theta^{2m} q (1 - q) \right. \\
+ 2\theta^m \left\{ \sum_{i=r-m+1}^{r} b_i p_i - \left( \theta^m q - \sum_{i=r-m+1}^{r} b_i p_i \right) \right\} \\
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 + \theta^{2m} q (1 - q) \right. \\
+ 2\theta^m \left\{ \sum_{i=r-m+1}^{r} b_i p_i - \theta^m q + \theta^m q^2 \right\} \\
= \left\{ \frac{1}{q^2} \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 + \theta^{2m} q (1 - q) \right. \\
+ 2\theta^m \left\{ \sum_{i=r-m+1}^{r} b_i p_i - \theta^m q (1 - q) \right\} \\
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 - \theta^{2m} q (1 - q) + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \right\}.
\]

To show that our expression is nonnegative, we once again use the fact that
\[
\sum_{i=0}^{r} b_i p_i = \theta^m \sum_{j=0}^{r-m} p_j = \theta^m q.
\]

Hence, our expression can be written as
\[
\frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \left( \sum_{i=1}^{r} b_i p_i \right)^2 - \theta^{2m} q (1 - q) + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \right\}
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \theta^{2m} q^2 - \theta^{2m} q + \theta^{2m} q^2 + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \right\}
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} b_i^2 p_i - \theta^{2m} q + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \right\}
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} (b_i^2 + \theta^{2m} - \theta^{2m} b_i) p_i - \theta^{2m} \sum_{i=0}^{r} p_i + 2\theta^m \sum_{i=0}^{r} b_i p_i \right\}.
\[-\theta^{2m} q + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \]

\[
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} (b_i^2 + \theta^{2m} - 2\theta^m b_i) p_i - \theta^{2m} \sum_{i=0}^{r} p_i + 2\theta^m \theta^m q \right. \\
- \theta^{2m} q + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \right\} 
\]

\[
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} (b_i^2 + \theta^{2m} - 2\theta^m b_i) p_i - \theta^{2m} \left( q + \sum_{i=r-m+1}^{r} p_i \right) + 2\theta^2 q \right. \\
- \theta^{2m} q + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \right\} 
\]

\[
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r} (b_i - \theta^m)^2 p_i + \sum_{i=r-m+1}^{r} (b_i - \theta^m)^2 p_i \right. \\
- \theta^{2m} \sum_{i=r-m+1}^{r} p_i + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \right\} 
\]

\[
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r-m} (b_i - \theta^m)^2 p_i + \sum_{i=r-m+1}^{r} b_i^2 p_i - 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i + \theta^{2m} \sum_{i=r-m+1}^{r} p_i \right. \\
- \theta^{2m} \sum_{i=r-m+1}^{r} p_i + 2\theta^m \sum_{i=r-m+1}^{r} b_i p_i \right\} 
\]

\[
= \frac{1}{q^2} \left\{ \sum_{i=1}^{r-m} (b_i - \theta^m)^2 p_i + \sum_{i=r-m+1}^{r} b_i^2 p_i \right\} .
\]

Hence, we have

\[ \sqrt{n'} (\hat{\alpha} - \theta^m) \overset{\text{dist}}{\Rightarrow} N(0, \xi^2(\theta)), \]

where

\[ \xi^2(\theta) = \frac{1}{q^2} \left\{ \sum_{i=1}^{r-m} (b_i - \theta^m)^2 p_i + \sum_{i=r-m+1}^{r} b_i^2 p_i \right\} , \]

where \( b_i = \frac{a(i-m)}{a(i)} \) and \( q = \sum_{j=0}^{r-1} p_j \). \qed
**Remark.** For the Poisson case, $p_j = \frac{e^{-\theta} \theta^j}{j!}$. Moore considered the case when $m = 1$. We can rewrite $\xi^2(\theta)$ in a bit simplified form as

\[
\begin{align*}
\xi^2(\theta) &= \frac{1}{(\sum_{j=0}^{r-1} e^{-\theta} \frac{\theta^j}{j!})^2} \left\{ \sum_{i=1}^{r-1} (i - \theta)^2 \frac{e^{-\theta} \theta^i}{i!} + \frac{r^2 e^{-\theta} \theta^r}{r!} \right\} \\
&= \frac{e^\theta}{(\sum_{j=0}^{r-1} \frac{\theta^j}{j!})^2} \left\{ \sum_{i=1}^{r-1} (i - \theta)^2 \frac{\theta^i}{i!} + \frac{r^2 \theta^r}{r!} \right\}.
\end{align*}
\]

It is interesting to note that Moore gives an expression that is similar to the above but does not have the terms $-\theta^2 q(1 - q) + 2\theta r p$, at all.

**Remark.** As right-tailed censored PSD are actually multinomial distributions, one can perform a Pearson’s goodness of fit test to check the accuracy of the estimates.

### 4.2 UMVU and Testing in Left Tail T-PSD

Previously, the expression truncated power series distributions referred to situations when the generating sequence $a_k$ had a fixed, known $\kappa$ such that $a_k > 0$ for all $k \geq \kappa$ and $a_k = 0$ for all $k < \kappa$. From now on, the word truncated used without any prefix will mean that the set of integers over which $a_k > 0$ depends on an unknown parameter. Such distributions will be denoted by T-PSD. In this situation, the parameter $\theta = (\nu, p)$ is a vector. This section presents some of the main results that have been proved for such distributions.

**Park (1973)**

Klotz (1970) considers UMVU and testing with the truncated geometric model. To be specific, $a_k = (1 - p)^{k-\nu} p$ for $k = \nu, \nu + 1, \cdots$ and $a_k = 0$ otherwise, where $\nu$ is an unknown parameter. Many of these results were later extended by Park (1973) to arbitrary PSD.

Consider any arbitrary T-PSD. That is, let $\nu$ be an unknown integer valued parameter and

\[
\mathbb{P}(X = k) = \frac{a_k \theta^k}{G(\nu, \theta)}, \quad k = \nu, \nu + 1, \nu + 2, \cdots
\]
To emphasize that $\nu$ is unknown, we use the notation $G(\nu, \theta)$ instead of $G(\theta)$ which we use in the ordinary PSD case. This gives the power series,

$$G(\nu, \theta) = \sum_{k=0}^{\infty} a_k \theta^k,$$

where $\theta \in (0, R)$. Note that $R$ does not depend on $\nu$ since the above series converges for a value of $\nu$ if and only if it converges for all values of $\nu$.

Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} X$. The joint density is

$$P(X_1 = x_1, \cdots, X_n = x_n) = \prod_{i=1}^{n} \frac{a(x_i) \theta^{x_i}}{G(\nu, \theta)^n} X(x_i \geq \nu)$$

$$= \theta^{\sum x_i} \prod_{i=1}^{n} \frac{a(x_i)}{(G(\nu, \theta))^n} X_{\{\min(x_i) \geq \nu\}}.$$  

By the factorization theorem, $(X_{(1)}, \sum_{i=1}^{n} X_i)$ a minimal sufficient statistic. This statistic is equivalent to

$$\left( X_{(1)}, \sum_{i=1}^{n} (X_i - X_{(1)}) \right).$$

The following proposition gives the UMPU tests for hypotheses concerning $\nu$ and $\theta$.

**Proposition 11.** (Park 1973) Let $H : \nu \leq \nu_0$ versus $K : \nu > \nu_0$ where $\nu_0$ is a given positive real number. Let $Y$ be the minimum of the sample and let $Z$ be the sum of the sample. Then the UMPU test of size $\alpha$ is given by

$$\phi(Y) = \begin{cases} 1 & \text{if } Y > C(Z), \\ \gamma(Z) & \text{if } Y = C(Z), \\ 0 & \text{if } Y < C(Z), \end{cases}$$

where $C(Z)$ and $\gamma(Z)$ are uniquely determined from the equation

$$\sum_{y=\nu_0}^{\infty} \phi(y) \frac{A(n, y, z) - A(n, y + 1, z)}{A(n, \nu_0, z)}$$

$$= \sum_{y>C(z)} \frac{A(n, y, z) - A(n, y + 1, z)}{A(n, \nu_0, z)} + \gamma(z) \frac{A(n, C(z), z) - A(n, C(z) + 1, z)}{A(n, \nu_0, z)}.$$
\[
= \alpha,
\]
and \(\{A(n, y, z)\}_{z=0}^{\infty}\) is the \(n\)-fold Cauchy product sequence \(c_y\) where \(c_y(k) = 0\) if \(k < y\) and \(c_y(k) = a_k\) for \(k \geq y\).

**Proposition 12.** (Park 1973) Let \(H : \theta \leq \theta_0\) versus \(K : \theta > \theta_0\) where \(\theta_0\) is a given positive real number. Let \(Y\) be the minimum of the sample and let \(Z\) be the sum of the sample. Then the UMPU test of size \(\alpha\) is given by

\[
\phi(Z) = \begin{cases} 
1 & \text{if } Z > C(Y), \\
\gamma(Y) & \text{if } Z = C(Y), \\
0 & \text{if } Z < C(Y), 
\end{cases}
\]

where \(C(Y)\) and \(\gamma(Y)\) are uniquely determined from the equation

\[
\sum_{z=0}^{\infty} \phi(z) \frac{(A(n, y, z) - A(n, y + 1, z))\theta_0^z}{(G(y, \theta_0))^n - (G(y + 1, \theta_0))^n} = \sum_{z > C(y)} (A(n, y, z) - A(n, y + 1, z))\theta_0^z \frac{(A(n, y, C(y)) - A(n, y + 1, C(y)))\theta_0^{C(y)}}{(G(y, \theta_0))^n - (G(y + 1, \theta_0))^n} + \gamma(y) \frac{(A(n, y, C(y)) - A(n, y + 1, C(y)))\theta_0^{C(y)}}{(G(y, \theta_0))^n - (G(y + 1, \theta_0))^n} = \alpha,
\]

where \(\{A(n, y, z)\}_{z=0}^{\infty}\) is the \(n\)-fold Cauchy product sequence \(c_y\) where \(c_y(k) = 0\) if \(k < y\) and \(c_y(k) = a_k\) for \(k \geq y\) and \(G(y, \theta)\) is the generating function of \(c_y\) sequence.

Charalambides (1974)

There is some overlap between the work of Park (1973) and that of Charalambides (1974). It was shown that \((X_{(1)}, \sum_{i=1}^{n} X_i)\) is a minimal sufficient statistic. This statistic is also complete. To prove this, the following proposition gives the joint density of \((X_{(1)}, \sum_{i=1}^{n} X_i)\).

**Proposition 13.** (Charalambides 1974) Now let \(X_1, X_2, \ldots, X_n \overset{iid}{\sim} X\), where \(X\) is a left
tail-truncated PSD random variable. Then the joint density of

\[ Y_n = X_{(1)}, \quad Z_n = \sum_{i=1}^{n} X_i \]

is given as follows:

\[ \mathbb{P}(Y_n = y, Z_n = z) = \frac{A(n, y, z) - A(n, y + 1, z)}{(G(\nu, \theta))^n} \]

for \( y = \nu, \nu + 1, \ldots \) and \( z = ny, ny + 1, \ldots \) where

\[ A(n, y, z) = \frac{(c_y * c_y * \cdots * c_y)_z}{n - \text{fold}}, \]

and \( c_y(k) = 0 \) for \( k < y \) and \( c_y(k) = a_k \) for \( k \geq y \). Yet another way of writing \( A(n, y, z) \) is

\[ A(n, y, z) = \sum a(k_1)a(k_2) \cdots a(k_n), \]

where the sum is over all nonnegative integers \( k_1, k_2, \ldots, k_n \) so that \( k_1 + \cdots + k_n = z \) and \( k_i \geq y \), for each \( i = 1, 2, \ldots, n \).

Proposition 14. (Charalambides 1974) Now let \( X_1, X_2, \ldots, X_n \overset{iid}{\sim} X \), where \( X \) is a left tail-truncated PSD random variable with parameters \( (\nu, \theta) \). For any given fixed positive integer \( m \), the UMVU estimators of \( g(\nu, \theta) = \nu^m \) and \( h(\nu, \theta) = \theta^m \) are as follows

\[ g_m(Y, Z) = \hat{g} = Y^m - \frac{A(n, Y, Z)}{A(n, Y, Z) - A(n, Y + 1, Z)} \sum_{k=0}^{m-1} \binom{m}{k} Y^k \]

\[ h_m(Y, Z) = \hat{h} = \frac{A(n, Y, Z - m) - A(n, Y + 1, Z - m)}{A(n, Y, Z) - A(n, Y + 1, Z)} \]

where \((Y, Z)\) is the complete sufficient statistic as derived earlier.

Proof. Since \((Y, Z)\) form a complete sufficient statistic and the given estimators are functions of \( Y \) and \( Z \) only, it is enough to show that the given estimators are unbiased for \( \nu^m \) and \( \theta^m \) respectively. For an unbiased estimator of \( \nu^m \), we need to find a sequence \( b(y, z) \) which has the property that

\[ \sum_{y=\nu}^{\infty} \sum_{z=ny}^{\infty} b(y, z) \mathbb{P}(Y = y, Z = z) = \nu^n, \quad \text{for all } \nu = 1, 2, \ldots, \theta \in (0, R). \]
The left side is equal to

\[
\sum_{y=\nu}^{\infty} \sum_{z=ny}^{\infty} b(y, z) \left( A(n, y, z) - A(n, y + 1, z) \right) \theta^z \\
= \sum_{z=n\nu}^{\infty} \left( \sum_{y=\nu}^{\lfloor z/n \rfloor} b(y, z) \left( A(n, y, z) - A(n, y + 1, z) \right) \right) \theta^z \\
= \nu^m (G(\nu, \theta))^n.
\]

The outer series represents a generating function of the inner terms. The last expression can also be written in the form of a generating function. This is because all the probabilities of the joint density of \((Y, Z)\) must add up to one, giving that

\[
\sum_{z=n\nu}^{\infty} \sum_{y=\nu}^{\lfloor z/n \rfloor} \frac{A(n, y, z) - A(n, y + 1, z)}{(G(\nu, \theta))^n} \theta^z = 1.
\]

In other words,

\[
(G(\nu, \theta))^n = \sum_{z=n\nu}^{\infty} \sum_{y=\nu}^{\lfloor z/n \rfloor} (A(n, y, z) - A(n, y + 1, z)) \theta^z.
\]

Using this observation, we need to find constants \(b(y, z)\) having the property that

\[
\sum_{z=n\nu}^{\infty} \sum_{y=\nu}^{\lfloor z/n \rfloor} b(y, z) \left( A(n, y, z) - A(n, y + 1, z) \right) \theta^z \\
= \nu^m \sum_{z=n\nu}^{\infty} \sum_{y=\nu}^{\lfloor z/n \rfloor} \left( A(n, y, z) - A(n, y + 1, z) \right) \theta^z.
\]

The power series being equal for all \(\theta \in (0, R)\) forces that the coefficients must be the same. That is,

\[
\sum_{y=\nu}^{\lfloor z/n \rfloor} b(y, z) \left( A(n, y, z) - A(n, y + 1, z) \right) \\
= \nu^m \sum_{y=\nu}^{\lfloor z/n \rfloor} \left( A(n, y, z) - A(n, y + 1, z) \right) \\
= \nu^m \left( A(n, \nu, z) - A(n, \lfloor z/n \rfloor + 1, z) \right) \\
= \nu^m A(n, \nu, z), \quad \text{since } A(n, \lfloor z/n \rfloor + 1, z) = 0.
\]
This is supposed to hold for all values of \( \nu = 1, 2, \cdots \). (Note that \( \nu = 0 \) means that there is no truncation at all.) Looking at two successive values, namely \( \nu \) and \( \nu + 1 \), we have that

\[
\sum_{y=\nu+1}^{[z/n]} b(y, z) (A(n, y, z) - A(n, y + 1, z)) = (\nu + 1)^m A(n, \nu + 1, z).
\]

Subtracting this condition from the condition corresponding to \( \nu \), the left side yields

\[
\sum_{y=\nu}^{[z/n]} b(y, z) (A(n, y, z) - A(n, y + 1, z)) - \sum_{y=\nu+1}^{[z/n]} b(y, z) (A(n, y, z) - A(n, y + 1, z))
\]

\[
= b(\nu, z) (A(n, \nu, z) - A(n, \nu + 1, z))
\]

and the right side yields

\[
\nu^m A(n, \nu, z) - (\nu + 1)^m A(n, \nu + 1, z)
\]

\[
= \nu^m A(n, \nu, z) - \nu^m A(n, \nu + 1, z) - A(n, \nu + 1, z) \sum_{k=0}^{m-1} \binom{m}{k} \nu^k.
\]

So

\[
b(\nu, z) (A(n, \nu, z) - A(n, \nu + 1, z))
\]

\[
= \nu^m (A(n, \nu, z) - A(n, \nu + 1, z)) - A(n, \nu + 1, z) \sum_{k=0}^{m-1} \binom{m}{k} \nu^k.
\]

for all values of \( \nu = 1, 2, \cdots \) and \( z = n\nu, n\nu + 1, \cdots \) and

\[
b(\nu, z) = \nu^m \frac{A(n, \nu + 1, z)}{A(n, \nu, z) - A(n, \nu + 1, z)} \sum_{k=0}^{m-1} \binom{m}{k} \nu^k.
\]

Therefore, the UMVU estimator of \( \nu^m \) is

\[
g_m(Y, Z) = Y^m \frac{A(n, Y + 1, Z)}{A(n, Y, Z) - A(n, Y + 1, Z)} \sum_{k=0}^{m-1} \binom{m}{k} Y^k.
\]

In a similar fashion, we need to find constants \( b(y, z) \) so that

\[
\sum_{y=\nu}^{\infty} \sum_{z=ny}^{\infty} b(y, z) (A(n, y, z) - A(n, y + 1, z)) \theta^z
\]
\[
= \sum_{z=n\nu}^{\infty} \sum_{y=\nu}^{\lfloor z/n \rfloor} b(y, z) (A(n, y, z) - A(n, y + 1, z)) \theta^z
\]

\[
= \theta^m (G(\nu, \theta))^n.
\]

This is equivalent to saying
\[
\sum_{z=n\nu}^{\infty} \sum_{y=\nu}^{\lfloor z/n \rfloor} b(y, z) (A(n, y, z) - A(n, y + 1, z)) \theta^z
\]

\[
= \sum_{z=n\nu}^{\infty} \sum_{y=\nu}^{\lfloor z/n \rfloor} (A(n, y, z) - A(n, y + 1, z)) \theta^{z+n}.
\]

Matching the coefficients of like powers of \( \theta \) on both sides, we have
\[
b(y, z + m) = \frac{A(n, y, z) - A(n, y + 1, z)}{A(n, y, z + m) - A(n, y + 1, z + m)}.
\]

This then gives that the UMVU of \( \theta^m \) is
\[
h_m(Y, Z) = \frac{A(n, Y, Z - m) - A(n, Y + 1, Z - m)}{A(n, Y, Z) - A(n, Y + 1, Z)}.
\]

This completes the proof. \( \Box \)

**Remark.** It should be noted that even though the above estimators are UMVU, it is a nontrivial matter to find the actual variance of these estimators. One way to conjecture that the variance of the UMVU estimator of \( \theta^m \) is of the order of \( O(n^{-1}) \) is to consider another unbiased estimator of \( \theta^m \) (which we may call as a Moore-Plackett form of estimator) as follows.

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{A(1, X_i, X_i - m) - A(1, X_i + 1, X_i - m)}{A(1, X_i, X_i) - A(1, X_i + 1, X_i)}.
\]

That is, we take the sample average of the UMVU estimator of \( \theta^m \) obtained by using each \( X_i \) separately. In this case \( Y = Z = X_i \) for each \( i = 1, 2, \cdots, n \). But the problem with this approach is that \( A(1, y, z) = 0 \) if \( y > z \). And \( A(1, y, y) = a_y \). This makes the numerator equal to zero for each \( i = 1, 2, \cdots, n \). Hence, this shows that there cannot exist a UMVU estimator for \( n = 1 \). It must be that \( n \geq 2 \). So, modify our argument and try pairing up
the observations into two consecutive ones (and at the end, the last pair may have two (if  
\( n \) is even) or three observations if \( n \) is odd). In that case, (for instance, when \( n \) is even)  
our estimator is  
\[
\frac{2}{n} \sum_{i=1}^{n/2} \frac{A(2, Y_i, Z_i - m) - A(2, Y_i + 1, Z_i - m)}{n} \approx \frac{A(2, Y_i, Z_i) - A(2, Y_i + 1, Z_i)}{n}.
\]
Now there is positive probability that \( Z_i - m \geq Y_i \) where \( Y_i \) is the minimum of the \( i \)-th pair  
and \( Z_i \) is the sum of the observations in the \( i \)-th pair. This unbiased estimator has variance  
\( \tau^2(\nu, \theta) / n \). Hence, our UMVU estimator must have variance that is of the order of \( O(n^{-1}) \).  

**Remark.** As we mentioned in the last remark, the actual variance of the UMVU estimator  
of \( \theta^m \) is difficult to find. However, it is very easy to give a UMVU estimator of the variance  
of the UMVU estimator of \( \theta^m \). This is because we note that  
\[
Var(h_m(Y_n, Z_n)) = E(h_m^2(Y_n, Z_n)) - (\theta^m)^2 = E(h_m^2(Y, Z)) - \theta^{2m}.
\]
But a UMVU estimator of \( \theta^{2m} \) is \( h_{2m}(Y_n, Z_n) \) and an unbiased estimator of \( E(h_m^2(Y_n, Z_n)) \)  
is \( h_m^2(Y_n, Z_n) \). Hence, an unbiased estimator of \( Var(h_m(Y_n, Z_n)) \) is  
\[
h_m^2(Y_n, Z_n) - h_{2m}(Y_n, Z_n).
\]
This being an unbiased estimator of \( Var(h_m(Y_n, Z_n)) \) and a function of the complete sufficient statistic, must itself be UMVU. Same argument can be used to give a UMVU estimator  
of the \( Var(g_m(Y_n, Z_n)) \).  

**Remark.** One needs fast algorithms to compute these estimators. Furthermore, the asymptotic properties of these estimators are still not fully known.

4.3 Relation to Approximation Theory  
This section will present some of the relations to approximation theory.
4.3.1 Background

We will be concerned with the approximation properties of the following positive linear operator sequence called power series operator

$$L_n(f, \theta) = \mathbb{E}_\theta \{ f(h(S_n)) \}, \quad n = 1, 2, \cdots,$$

defined for each $f \in C_B[0, R]$ — the space of bounded uniformly continuous functions over $[0, R)$, or defined over a more general space $C_B(I)$ of Herrmann (1990).

Let $I$ be an interval, and for each $x \in I$, let $\{ \mu_{n,x} \, n = 1, 2, \cdots \}$ be a sequence of finite measures concentrating on $I$, and let $D(I)$ be the linear space of continuous functions over $I$ so that $f \in D(I)$ if and only if for each $\alpha > 0$,

$$\sup_{t \in I} \frac{|f(t)|}{e^{\alpha|t|}} < \infty.$$ 

Note that this class contains all polynomials.

Let $g(t)$ be a non-negative continuous function increasing to infinity as $t \to \pm \infty$. Define the space $D_g(I)$ to be the linear space of continuous functions over $I$ so that $f \in D_g(I)$ if and only if for each $\alpha > 0$,

$$\sup_{t \in I} \frac{|f(t)|}{e^{\alpha g(t)}} < \infty.$$ 

When $I$ is unbounded, we will assume that $g$ grows faster than $\log(|t| + 1)$ to avoid trivialities. For all our applications, $g(t)$ will grow at least as fast as $|t|$, so, we will be working with those $g$ for which $D(I) \subseteq D_g(I)$. We should remark that T. Herrmann (1990) considered the space

$$C_E := \{ f \in C[0, \infty) : \log(|f(t)| + 1) = o(t), \text{ as } t \to \infty \}.$$ 

It is not difficult to see that our $D(I) = C_E$ when $I = [0, \infty)$.

If we restrict our attention to the space $C_B[0, R]$ of bounded, uniformly continuous functions over $[0, R)$, then

$$L_n(f, \theta) \to f(\theta)$$
for each $f \in C_B[0, R)$. As $f$ is continuous, given $\epsilon > 0$, there exists a $\delta > 0$ such that $|x - \theta| < \delta$ implies $|f(x) - f(\theta)| < \epsilon$. This gives

$$
\lim_{n \to \infty} |L_n(f, \theta) - f(\theta)| \leq \lim_{n \to \infty} \mathbb{E}|f(h(S_n)) - f(\theta)| \\
= \lim_{n \to \infty} \left\{ \mathbb{E}|f(h(S_n)) - f(\theta)|_{\delta} + \mathbb{E}|f(h(S_n)) - f(\theta)|_{\delta \leq \delta} \right\} \\
\leq \epsilon + 2M\mathbb{P}(\|S_n - \theta\| \geq \delta) \\
\leq \epsilon + \frac{2M}{\sqrt{n}} \lim_{n \to \infty} \mathbb{E}\frac{\sqrt{n}|h(S_n) - \theta|}{\delta} \\
= \epsilon,
$$

where $M$ is the bound of $f$. That is, given $\epsilon > 0$, there exists an $n \in \mathbb{N}$ such that was arbitrary. Furthermore, its rate of convergence is

$$
|L_n(f, \theta) - f(\theta)| \leq C\|f\| \omega(f, n^{-1/2}),
$$

for some constant $C$ depending on $\theta$ only and for each $f \in C_B[0, R)$. This result yields a large number of known results for special cases. Later, we provide a few examples of such results. Here we present a general result of Hermann (1977) concerning the approximation of functions.

**Theorem 6.** (Hermann 1977) Let $L_n$ be a sequence of positive linear operators defined over the space of bounded continuous functions over an interval $I$, so that for each bounded continuous function $f$,

$$
L_n(f, x) \to f(x); \quad \text{for each } x \in I.
$$

Then for any continuous $F(x) > 0$ for which

$$
L_n(F, x) \to F(x); \quad x \in D
$$

where $D \subseteq I$ and $D$ is an open set, and for each continuous function $f$ with $f(x) = O(F(x))$ we have

$$
L_n(f, x) \to f(x); \quad x \in D.
$$
Proof. Let \( x \in D \). Since \( D \) is open, there exists a \( \delta > 0 \) so that \( [x - 2\delta, x + 2\delta] \subseteq D \). Define a function \( h_x(t) \) which equals 1 over \( [x - \delta, x + \delta] \) and equals zero over the complement of \( [x - 2\delta, x + 2\delta] \) and is linear and continuous otherwise. And let \( g_x(t) = 1 - h_x(t) \) so that
\[
h_x(t) + g_x(t) = 1; \quad \text{for all } x \in I.
\]
Now let \( f \) be any continuous function so that \( f(x) = O(F(x)) \). Then be the linearity of \( L_n \) we have
\[
L_n(f, x) = L_n(f h_x, x) + L_n(f g_x, x).
\]
Clearly \( f(t)h_x(t) \) is zero beyond the compact set \( [x - 2\delta, x + 2\delta] \) and continuous (and hence bounded continuous). The given hypothesis gives that
\[
L_n(f h_x, x) \to f(x)h_x(x) = f(x).
\]
So we need only show that \( L_n(f g_x, x) \to 0 \). For any \( x \in D \), we are given that
\[
|f(t)g_x(t)| \leq K_x F(t)g_x(t); \quad \text{for all } t \in I.
\]
Since \( f g_x \) is a continuous function, by the positivity of \( L_n \) we have
\[
-K_x L_n(F g_x, x) \leq K_x L_n(f g_x, x) \leq K_x L_n(F g_x, x).
\]
Hence,
\[
|L_n(f g_x, x)| \leq K_x L_n(F g_x, x).
\]
However,
\[
L_n(F g_x, x) = L_n(F - F h_x, x)
\]
\[
= L_n(F, x) - L_n(F h_x, x)
\]
\[
\to F(x) - F(x)h_x(x) = 0; \quad x \in D.
\]
The last conclusion comes by the given hypothesis and the fact that \( F h_x \) is a bounded continuous function.
4.3.2 Examples

We list a few typical examples of power series operators.

Szász Operator

Let $P(X_1 = j) = e^{-x}x^j(j!)^{-1}$, $j = 0, 1, \ldots$. The power series operator reduces to the well known Szász operator

$$Z_n(f, x) = e^{-nx} \sum_{k=0}^{\infty} f \left( \frac{k}{n} \right) \frac{(nx)^k}{k!}, \quad n = 1, 2, \ldots,$$

where $f(t)$ is defined over $[0, \infty)$. In this case, since all the $a_k > 0$, for $k = 0, 1, 2, \ldots$, our UMVU estimator for $\theta$ is

$$h(S_n) = \frac{A(n, S_n - 1)}{A(n, S_n)} = \frac{S_n}{n},$$

the sample mean. It is known that such an operator approximates functions that do not grow more than the first order exponential rate. In fact, Hermann (1977) showed that $Z_n(f, x) \to f(x)$ for each $f \in D_g(I)$ where $g(t) = t \log(t + 1)$.

Bleimann, Butzer and Hahn Operator

Let $a_0 = a_1 = 1$ and $a_k = 0$ for all $k \geq 2$. The power series operator reduces to the Bleimann, Butzer and Hahn operator

$$B_n(f, \theta) = (1 + \theta)^{-n} \sum_{k=0}^{n} f \left( \frac{k}{n-k+1} \right) \left( \frac{n}{k} \right) \theta^k, \quad n = 1, 2, \ldots.$$  

This operator was further studied by a number of other authors. The three most relevant ones are Totik (1984), R. A. Khan (1988) and Hermann (1990). In 1990, Hermann showed that $B_n(f, x) \to f(x)$ for any $f \in D(I)$. 
Meyer-König and Zeller Operator

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Geometric}(1 - \theta)$, i.e., $P(X_i = j) = (1 - \theta)(\theta)^j, \ j = 0, 1, \ldots, \theta \in [0, 1)$. The power series operator reduces to the modified Meyer-König and Zeller operator

$$M_n(f, \theta) = (1 - \theta)^n \sum_{k=0}^{\infty} f \left( \frac{k}{n+k-1} \right) \left( \frac{n+k-1}{k} \right) \theta^k, \ n \geq 1,$$

where for $n = 1$ we will take $f(0) = f(1)$. If we take $\theta = x/(1 + x)$ then it is easy to see that $\sigma^2 = x(1 + x)$ and

$$\mathbb{E}|X_1 - x|^3 = 2x^3 + 3x^2 + x + \frac{2}{1 + x} \sum_{j=0}^{[x]} (x - j)^3 \left( \frac{x}{1 + x} \right)^j$$

This operator was studied by Cheney and Sharma (1964) for continuous functions and by M. K. Khan (1989) for functions of bounded variations.

Voronovskaya-Type Results

Perhaps the closest relation to general PSD is general Voronovskaya-type results. Let $I$ be a fixed interval in $\mathbb{R}$, bounded or not, and for each fixed $n \in \mathbb{N}$ let $\Gamma_n = (\nu_{n,k})_{k \in \mathbb{N}_0} \subset I$ be a sequence such that

$$0 < \nu_{n,k+1} - \nu_{n,k} \leq \lambda_n$$

where $\lambda_n$ is a sequence of positive real numbers converging to zero. Consider the sequence of operators $L = (L_n)$ where

$$(L_n f)(t) = \sum_{k=0}^{\infty} K_n(t, \nu_{n,k}) f(\nu_{n,k}), \ n \in N, \ t \in I$$

where the $K_n(t, \nu_{n,k}) : I \times \Gamma_n \to \mathbb{R}$ is a sequence of functions satisfying the following assumptions:

1. $\sum_{k=0}^{\infty} K_n(t, v_{n,k}) = 1$, for every $n \in N$ and $t \in \hat{I}$.

2. Define

$$m_j(K_n, t) := \sum_{k=0}^{\infty} K_n(t, v_{n,k})(v_{n,k} - t)^j.$$
For $j = 1, 2$,

$$-\infty < m_j(K_n, t) < \infty$$

for all $t \in I$, $n \in N$ and there are $\alpha > 0$ and real numbers $\ell_j(t)$ such that

$$\lim_{n \to +\infty} n^\alpha m_j(K_n, t) = \ell_j(t)$$

3. Define

$$M_2(K_n, t) := \sum_{k=0}^{\infty} |K_n(t, v_{n,k})|(v_{n,k} - t)^2$$

For the above $\alpha > 0$ and for every $t \in I$ there exists a positive constant $H(t)$ and a $\overline{n} \in \mathbb{N}$ such that

$$n^\alpha M_2(K_n, t) \leq H(t)$$

for every $n \geq \overline{n}$ and for every $\delta > 0$ and $t \in I$

$$\sum_{|v_{n,k} - t| > \delta} |K_n(t, v_{n,k})|(v_{n,k} - t)^2 = o(n^{-\alpha}), \quad n \to +\infty$$

The following is a theorem from Bardaro and Mantellini (2009):

**Theorem 7.** (Bardaro and Mantellini) Let $f \in \text{Dom}S \cap L^\infty(I)$ be a function such that $f''(t)$ exists at a point $t \in I$. Under the above assumptions there holds

$$\lim_{n \to +\infty} n^\alpha [(L_n f)(t) - f(t)] = f'(t)\ell_1(t) + \frac{f''(t)}{2}\ell_2(t)$$

In the language of probability,

1. If $K_n(t, v_{n,k}) \geq 0$ then $K_n(t, v_{n,k})$ is a density

2. If $K_n(t, v_{n,k})$ is a density, $\mathbb{E}[(v_{n,k} - t)^j]$ is finite for $j = 1, 2$. Furthermore, there are $\alpha > 0$ and real number $\ell_j$ such that

$$\lim_{n \to +\infty} n^\alpha \mathbb{E}[(v_{n,k} - t)^j] = \ell_j(t)$$
3. If \( K_n(t, v_{n,k}) \) is not a density, then \( \sum_{k=0}^{\infty} K_n(t, v_{n,k})(v_{n,k} - t)^2 \) must be absolutely convergent. If \( K_n(t, v_{n,k}) \) is a density, then \( M_2(K_{n,k}, t) = m_2(K_{n,k}, t) \) and

\[
\sum_{|v_{n,k} - t| > \delta} |K_n(t, v_{n,k})|(v_{n,k} - t)^2 = o(n^{-\alpha}), \quad n \to +\infty
\]

gives a bound on the tail of the second moment of \( (v_{n,k} - t) \).

In our context, \( t = \theta \), and

\[
K_n(t, v_{n,k}) = P_\theta(S_n = k) = A(n, k) \frac{\theta^k}{(G(\theta))^n} = \frac{((G(\theta))^n)^{(k)}}{k!} \frac{\theta^k}{(G(\theta))^n}
\]

\[
v_{n,k} = A(n, k - m) \frac{A(n, k)}{A(n, k - m)} = \frac{kP_{k-m}}{(k-m)!} \frac{((G(\theta))^n)^{(k-m)}}{((G(\theta))^n)^{(k)}}
\]

where \( kP_{k-m} = \frac{k!}{(k-m)!} \).

Condition 2) states that there exists a positive constant \( \alpha \) and positive real functions \( \ell_1(t) \) and \( \ell_2(t) \) such that

\[
\lim_{n \to +\infty} n^\alpha \mathbb{E}[(h_m(S_n) - \theta)] = \ell_1(\theta)
\]

and

\[
\lim_{n \to +\infty} n^\alpha \mathbb{E}[(h_m(S_n) - \theta)^2] = \ell_2(\theta)
\]

and condition 3) there is a positive real function \( H(t) \) and a \( \overline{n} \in \mathbb{N} \) such that

\[
n^\alpha \text{Var}(h_m(S_n) - \theta) \leq H(\theta)
\]

for all \( n \geq \overline{n} \) and \( t \in I \). That is, the tail of the sequence \( (n^\alpha \text{Var}(h(S_n) - \theta))_{n=1}^{\infty} \) is bounded by a positive real function. For \( PSD(\theta) \), all we need is condition is \( \alpha = 1 \) and condition 3) and we have \( \ell_1(t) = 0 \) and \( \ell_2 = \frac{m^2 \theta^2}{\sigma^2} \). and the theorem gives exactly what we expect when \( f(x) = x^2 \).
CHAPTER 5

Conclusions

We have established some asymptotic properties for the UMVU estimator for a general power series. We have shown that the limiting distribution of the UMVU estimator for $\theta^m$ is asymptotically normal. We have given the limiting mean deviation and the convergence rate for the limiting mean deviation. We have shown that the UMVU estimator is uniformly integrable for $1 \leq r < 2$ and have given the limiting variance of the estimator under certain conditions. Lastly, we have shown how UMVU estimation for $\theta^m$ in the area of general powers series distributions has links to estimation in left-truncated and censored power series distributions and links to some classical operators from approximation theory.

The main results are as follows.

Theorem 1. When $\sigma^2 = Var_\theta(X_1) < \infty$, the UMVU estimator $h_m(S_n) = \frac{A(n, S_n - m)}{A(n, S_n)}$ is asymptotically normal and

$$\sqrt{n}\left(\frac{A(n, S_n - m)}{A(n, S_n)} - \theta^m\right) \xrightarrow{dist} N\left(0, \frac{m^2 \theta^{2m}}{\sigma^2}\right).$$

Theorem 2. For any $\theta \in (0, R)$, we have

$$\mathbb{E}_\theta |h_m(S_n) - \theta^m| = O(n^{-1/2}).$$

Theorem 3. For any $\theta \in (0, R)$, we have

$$\lim_{n \to \infty} \sqrt{n} \mathbb{E}_\theta |h_m(S_n) - \theta^m| = \frac{m \theta^m}{\sigma} \sqrt{\frac{2}{\pi}}.$$
Theorem 4. Let $G(\theta) = a_0 + a_1 \theta + a_2 \theta^2 + \ldots$ be a function with radius of convergence $R > 0$ and $a_k > 0 \ \forall k \geq \kappa \in \mathbb{N}_0$ and $a_k = 0, \ \forall k < \kappa$. Let $X_1, X_2, \ldots \overset{iid}{\sim} PSD(\theta)$. When

$$\sigma^2 = \text{Var}_\theta(X_1) < \infty,$$

if $\text{Var} \left( \frac{aX - m}{aX} \right)$ exists for $\theta \in (0, R)$ and

$$\sqrt{n} \frac{\mathbb{P}(S_n = k - 1) - \mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k)}$$

is bounded by a measurable function on $(0, R)$ for all $n \in \mathbb{N}$ and $k \geq \kappa n$, then for any $\theta \in (0, R),$

$$\lim_{n \to \infty} n \mathbb{E}_\theta (h(S_n) - \theta)^2 = \frac{\theta^2}{\sigma^2}.$$

Theorem 5. If $h(S_n)$ is the UMVU estimator of $\theta$ and if $Y = \frac{a(X-1)}{a(X)}$ has finite variance, then for any $1 \leq r < 2$, we have

$$\lim_{n \to \infty} \mathbb{E} (\sigma \sqrt{n} |h(S_n) - \theta|)^r = \frac{\theta^r}{\sigma^r} \int_{-\infty}^{\infty} |t|^r \phi(t) \, dt.$$

Therefore, $(\sigma \sqrt{n} |h(S_n) - \theta|)^r$ with $n \geq 1$ is uniformly integrable.
APPENDIX A

Appendix - Petrov Expansions

This appendix contains some background information concerning the Petrov expansions cited in this paper. Valentin V. Petrov’s book *Sums of Independent Random Variables* (published 1972, translated from the Russian in 1975 by A. A. Brown) contains many useful facts for dealing with sums of independent random variable. Instead of the usual *independent identically distributed* requirement, Petrov starts with a few basic assumptions, and uses only independence. In a manner like Euclid’s *Elements*, the book builds on these basic principles. This makes the book a joy to read, and it makes it rather long to explain a certain result, as one must use several previous results to explain the one result.

Notation: Let

$$\sigma_j^2 = \mathbb{E}X_j^2, \quad \varphi_j(t) = \mathbb{E}e^{itX_j} = \int_{-\infty}^{\infty} e^{itz} dV_j(x), \quad V_j(x) = \mathbb{P}(X_j < x)$$

$$B_n = \sum_{j=1}^{n} \sigma_j^2, \quad F_n(x) = \mathbb{P}(B_n^{-1/2} \sum_{j=1}^{n} X_j < x)$$

Let $\lambda_{\nu n} = \frac{n^{\nu/2} \gamma_{\nu j}}{(B_n)^{\nu/2}}$, where $\gamma_{\nu m}$ is the cumulant of order $\nu$. Let $H_m(x)$ be the $m^{th}$ Chebyshev-Hermite polynomial. Lastly, let $f_n(t)$ be the characteristic function of the random variable $B_n^{-1/2} \sum_{j=1}^{n} X_j$.

We start with a fact: a distribution with the characteristic function $\varphi(t)$ is a lattice distribution if and only if there exists a $t_0 \neq 0$ such that $|\varphi(t_0)| = 1$. The span $h$ is maximal if and only if $|\varphi\left(\frac{2\pi}{h}\right)| = 1$ and $|\varphi(t)| < 1$ in the interval $0 < t < \frac{2\pi}{h}$. Consequently, if $\varphi(t)$ is the characteristic function of a lattice distribution with the maximal span $h$, then given $\varepsilon > 0$, there exists a $c > 0$ such that $|\varphi(t)| \leq e^{-c}$ for every $t$ in the interval $\varepsilon \leq |t| \leq \frac{2\pi}{h} - \varepsilon$. 

80
Lemma 1. (Petrov Lemma 1, Chapter V, pp. 109-111)

Let $X_1, \ldots, X_n$ be independent random variables with $E X_j = 0$, $|E X_j|^3 < \infty$ for $j = 1, \ldots, n$, and let $L_n = B_n^{-3/2} \sum_{j=1}^{n} E|X_j|^3$. Then $|f_n(t) - e^{-t^2/2}| \leq 16L_n|t|^3 e^{-t^2/3}$ for $|t| \leq \frac{1}{4L_n}$.

Proof. First, if $|t| \geq \frac{1}{2} L_n^{-1/3}$, then $8L_n|t|^3 \geq 1$. Plugging this into 1 gives

$$|f_n(t) - e^{-t^2}| \leq |f_n(t)| + e^{-t^2/2} \leq 2e^{-t^2/3}.$$ 

Since $\varphi_j(t) = E e^{itX_j}$ $(j = 1, \ldots, n)$, a random variable $\bar{X}_j = X_j - Y_j$, where $Y_j$ has the same distribution as $X_j$ but does not depend on $X_j$ has the characteristic function $|\varphi_j(t)|^2$ and variance $2\sigma_j^2$. Furthermore, $E|\bar{X}_j|^3 \leq 8E|X_j|^3$, and

$$|\varphi(t)|^2 \leq 1 - \sigma_j^2 t^2 + \frac{4}{3} |t|^3 E|X_j|^3 \leq \exp \left\{ -\sigma_j^2 t^2 + \frac{4}{3} |t|^3 E|X_j|^3 \right\}.$$ 

So in the interval $|t| \leq \frac{1}{4L_n}$ we have the estimate

$$|f_n(t)|^2 = \prod_{j=1}^{n} \left| \phi_j \left( \frac{t}{\sqrt{B_n}} \right) \right|^2 \leq \exp \left\{ -t^2 + \frac{4}{3} L_n |t|^3 \right\} \leq \exp \left\{ -\frac{2}{3} t^2 \right\}$$ 

and 1 holds.

If $|t| \leq \frac{1}{4L_n}$ and $|t| < \frac{1}{2} L_n^{-1/3}$, then for $j = 1, \ldots, n$,

$$\frac{\sigma_j}{\sqrt{B_n}} |t| \leq \left( \frac{E|X_j|^3}{\sqrt{B_n}} \right)^{1/3} |t| < L_n^{1/3} |t| < \frac{1}{2}$$ 

and

$$\varphi_j \left( \frac{t}{\sqrt{B_n}} \right) = 1 - r_j$$ 

where

$$r_j = \frac{\sigma_j^2 t^2}{2B_n} + \eta_j \frac{E|X_j|^3}{6B_n^{3/2}} |t|^3, \quad |\eta_j| \leq 1,$$ 

so that $|r_j| < \frac{1}{6}$ and

$$|r_j|^2 \leq 2 \left( \frac{\sigma_j^2 t^2}{2B_n} \right)^2 + 2 \left( \frac{E|X_j|^3}{6B_n^{3/2}} |t|^3 \right)^2 \leq \frac{E|X_j|^3}{3B_n^{3/2}} |t|^3.$$
This gives
\[
\log \varphi_j \left( \frac{t}{\sqrt{B_n}} \right) = -\frac{\sigma_j^2 t^2}{2B_n} + \eta_j \frac{\mathbb{E}|X_j|^3}{2B_n^{3/2}}|t|^3
\]
and
\[
\log f_n(t) = -\frac{t^2}{2} + \eta \frac{L_n}{2}|t|^3, \quad |\eta| \leq 1.
\]
Since \(L_n|t|^3 < \frac{1}{8}\) implies \(\exp \left\{ \frac{1}{2} L_n |t|^3 \right\} < 2\), we have
\[
|f_n(t) - e^{-t^2/2}| \leq e^{-t^2/2} \left| e^{\frac{3}{2} L_n |t|^3} - 1 \right|
\]
\[
\leq \frac{L_n}{2} |t|^3 \exp \left\{ -\frac{t^2}{2} + \frac{L_n}{2} |t|^3 \right\}
\]
\[
\leq L_n |t|^3 e^{-t^2/2}.
\]
The result follows. \(\square\)

**Lemma 2.** (Petrov Lemma 12, Chapter VI, p. 179, Corollary to Lemma 1, Chapter V, pp. 109-111)

Let \(X_1, \ldots, X_n\) be independent random variables with \(\mathbb{E}X_j = 0, |\mathbb{E}X_j|^3 < \infty\) for \(j = 1, \ldots, n\). Let \(B_n = \sum_{j=1}^{n} \mathbb{E}X_j^2 > 0\) and let \(L_n = B_n^{-3/2} \sum_{j=1}^{n} \mathbb{E}|X_j|^3\). Let \(b\) be an arbitrary positive constant. If \(|t| \leq \frac{b}{L_n}\), then
\[
|f_n(t)| \leq \exp \left\{ -\frac{3 - 4b^2}{6} t^2 \right\}.
\]
**Proof.** The result follows from the previous lemma. \(\square\)

**Definition 1.** A sequence \(\{X_n\}\) belongs to the class \(S(k, l, \alpha)\) if it satisfies the following conditions:

(i) \(\liminf \frac{B_n}{n} > 0, \limsup \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}|X_i|^k < \infty,\)

(ii) \(\frac{1}{n} \sum_{j=1}^{n} \int_{|x|>n^\tau} |x|^k dV_j(x) \to 0, \) for some \(0 < \tau < \frac{1}{2},\)

(iii) \(n^\alpha \int_{|t|>\epsilon} |t|^{-1} \prod_{j=1}^{n} |\varphi_j(t)| dt \to 0 \) for every fixed \(\epsilon > 0.\)
Theorem 8. (Petrov theorem 7, Ch. V, pp. 175-179)

If the sequence \( \{X_n\} \) of independent random variables with means zero belongs to the class \( S \left( k, l, \frac{k + l - 2}{2} \right) \), then for all \( x \) and sufficiently large \( n \) there exists a continuous derivative \( \frac{d^l}{dx^l} F_n(x) \) and moreover

\[
\frac{d^l}{dx^l} F_n(x) = \frac{d^l}{dx^l} \left( \Phi(x) + \sum_{\nu=1}^{k-2} \frac{Q_{\nu n}}{n^{\nu/2}} \right) + o \left( \frac{1}{n^{(k-2)/2}} \right)
\]

uniformly in \( x \) \((-\infty < x < \infty)\). Here

\[
Q_{\nu n}(x) = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \sum H_{\nu + 2s - 1}(x) \prod_{m=1}^{\nu} \frac{1}{k!} \left( \frac{\lambda_{m+2,n}}{(m+2)!} \right)^{k_m},
\]

where the sum is carried out over all non-negative integer solutions of the equation \( k_1 + 2k_2 + \cdots + \nu k_{\nu} = \nu \) and \( s = k_1 + k_2 + \cdots k_{\nu} \).

Theorem 9. (Petrov Theorem 13, pp. 205-6)

Let \( \{X_n\} \) be a sequence of integer-valued random variables having a common distribution. Suppose that \( \text{Var}(X_1) = \sigma^2 > 0 \), \( E|X_1|^k < \infty \) for some integer \( k \geq 3 \) and that the maximal span of the distribution of \( X_1 \) is equal to 1. Then

\[
\sigma \sqrt{n} \mathbb{P}_n(N) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + \sum_{\nu=1}^{k-2} \frac{q_{\nu n}(x)}{n^{\nu/2}} + o \left( \frac{1}{n^{(k-2)/2}} \right)
\]

uniformly in \( N \) \((-\infty < N < \infty)\). Here \( x = \frac{N - n\mathbb{E}X_1}{\sigma \sqrt{n}} \), \( \mathbb{P}_n(N) = \mathbb{P} \left( \sum_{j=1}^{n} X_j = N \right) \) and the \( q_{\nu n}(x) = \frac{d}{dx} Q_{\nu n}(x) \).

Proof. The result follows from the previous theorem. \( \square \)

Theorem 10. Theorem (Petrov Theorem 16, pp. 207-9) Let \( \{X_n\} \) be a sequence of integer valued random variables having a common distribution. If \( \text{Var}(X_1) = \sigma^2 > 0 \), \( E|X_1|^k < \infty \) for some integer \( k \geq 3 \), and the maximal span of distribution of \( X_1 \) is equal to 1, then

\[
(1 + |x|^k) \left( \sigma \sqrt{n} \mathbb{P}_n(N) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} - \sum_{\nu=1}^{k-2} \frac{q_{\nu n}(x)}{n^{\nu/2}} \right) = o \left( \frac{1}{n^{(k-2)/2}} \right)
\]

uniformly in \( x \). Where \( x \), \( \mathbb{P}_n(N) \) and the \( q(t) \) are defined above.
Proof. Let $\mu = \mathbb{E}X_1$, and let the characteristic function of the random variable $X_1$ and the characteristic function of the normalized sum $Z_n = \frac{1}{\sigma \sqrt{n}} \sum_{j=1}^{n} (X_j - \mu)$ be denoted by $\varphi(t)$ and $f_n(t)$ respectively. It is clear that $f_t = \sum_{N=-\infty}^{\infty} e^{itx} \mathbb{P}_n(N)$. Multiplying both sides of this by $e^{-itx}$ for fixed $n$ and integrating over the interval $|t| < \pi \sigma \sqrt{n}$ gives

$$2\pi \sigma \sqrt{n} \mathbb{P}_n(N) = \int_{|t|<\pi \sigma \sqrt{n}} e^{-itx} f_n(t) dt.$$ 

The condition that $\mathbb{E}|X_1| < \infty$ implies that the derivative $\frac{d^k}{dt^k} f_n(t)$ exists and that

$$\frac{d^k}{dt^k} f_n(t) = \sum_{N=-\infty}^{\infty} (ix)^k e^{itx} \mathbb{P}_n(N).$$

The previous two equations give

$$2\pi \sigma (ix)^k \sqrt{n} \mathbb{P}_n(N) = \int_{|t|<\pi \sigma \sqrt{n}} e^{-itx} \frac{d^k}{dt^k} f_n(t) dt.$$ 

Let

$$\varphi_k(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + \sum_{v=1}^{k-2} \frac{q_v(x)}{n^{v/2}}$$

and

$$u_k = \int_{-\infty}^{\infty} e^{itx} \varphi_k(x) dx$$

so that

$$u_k(t) = e^{-t^2/2} \left( 1 + \sum_{v=1}^{k-2} \frac{\mathbb{P}_v(it)}{n^{v/2}} \right)$$

where

$$e^{-t^2/2} \mathbb{P}_v(it) = \int_{-\infty}^{\infty} e^{-itx} q_v(x) dx.$$ 

Differentiating $A k$ times and applying the inversion formula gives

$$x^k \varphi_k(x) = \frac{1}{2\pi i^k} \int_{-\infty}^{\infty} e^{-itx} \frac{d^k}{dt^k} u_k(t) dt.$$ 

Therefore

$$|x^k(\sigma \sqrt{n} \mathbb{P}_n(N) - \varphi_k(x))| \leq I_1 + I_2 + I_3 + I_4,$$
where

\[
I_1 = \int_{|t| < n^{1/7}} \left| \frac{d^k}{dt^k} \left( f_n(t) - u_k(t) \right) \right| dt,
\]

\[
I_2 = \int_{|t| > n^{1/7}} \left| \frac{d^k}{dt^k} u_k(t) \right| dt
\]

\[
I_3 = \int_{n^{1/7} < |t| < T_n} \left| \frac{d^k}{dt^k} f_n(t) \right| dt
\]

\[
I_4 = \int_{T_n < |t| < \pi \sqrt{n}} \left| \frac{d^k}{dt^k} f_n(t) \right| dt
\]

where

\[
T_n = \frac{\sigma^3 \sqrt{n}}{4E|X_1 - \mu|^\delta}.
\]

Note that the integrals \(I_1, \ldots, I_4\) do not depend on \(N\). To estimate these integrals, we need the following lemmas:

**Lemma 3.** If \(E|X_1|^k < \infty\) for some integer \(k \geq 3\), we have in the interval \(|t| < n^{1/7}\)

\[
\left| \frac{d^k}{dt^k} \left( f_n(t) - u(t) \right) \right| \leq \frac{\varepsilon(n)}{n^{(k-2)/2}} (1 + |t|^{k(2k-1)}) e^{-t^2/2},
\]

where \(\varepsilon(n)\) is independent of \(t\) and \(\varepsilon(n) \to 0\).

\[
\square
\]

**Lemma 4.** Suppose that a function \(y = y(x)\) has a derivative of order \(\nu \geq 1\). Then

\[
\frac{d^\nu}{dx^\nu} = \nu! \sum_{k=1}^{\min(\nu, n)} \sum' \frac{n!}{(n-k)!} y^{n-k}(x) \prod_{m=1}^{\nu} \frac{1}{k_m!} \left( \frac{d^m}{dx^m} y(x) \right)^{k_m}
\]

where \(\sum'\) denotes summation over all non-negative integer solutions of the system of equations

\[
k_1 + 2k_2 + \cdots + \nu k_\nu = \nu
\]

\[
k_1 + k_2 + \cdots + k_\nu = k,
\]

and we write \(0^0 = 1\).
Proof. The first lemma implies that $I_1 = o(n^{-(k-2)/2})$ and clearly the same estimate holds for $I_2$.

From the second lemma and from the inequalities

$$\left| \frac{d^s}{dt^s} \varphi(t) \right| \leq \mathbb{E}|X_1|^s < \infty (s = 1, \ldots, k)$$

we have

$$I_3 \leq Cn^{2k} \int_{n^{1/7} < |t| < T_n} \left| \varphi \left( \frac{t}{\sigma \sqrt{n}} \right) \right|^{n-k} dt.$$ 

Now, Petrov Lemma 12, Chapter VI above gives

$$I_3 \leq Cn^{2k} \int_{|t| > n^{1/7}} e^{-(n-k)t^2/3n} dt = o(n^{-(k-2)/2}).$$

Finally, if $f(t)$ is the characteristic function of a lattice distribution with maximal span $h$, then for every $\varepsilon > 0$ there exists a $c > 0$ such that $|f(t)| \leq e^{-c}$ for $\varepsilon \leq |t| \leq \frac{2\pi}{h} - \varepsilon$.

Therefore,

$$I_4 \leq Cn^{3k} \int_{b < |t| < \pi} |\varphi(t)|^{n-k} dt = O(e^{-cn^{1/2}})$$

where $b = \frac{\sigma^2}{4} (\mathbb{E}|X_1 - \mu|^3)^{-1}$. The assertion of the theorem follows from these estimates and the inequality from Petrov Theorem 13. \qed
REFERENCES


[57] Lorentz, G. G. (1953), Bernstein Polynomials, Univ. of Toronto Press.


