ON UNBIASED ESTIMATION OF DENSITY FUNCTIONS

by

Allan Henry Seheult

Institute of Statistics
Mimeograph Series No. 649
Ph.D. Thesis - January 1970
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>1.</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>REVIEW OF THE LITERATURE</td>
<td>6</td>
</tr>
<tr>
<td>3.</td>
<td>GENERAL THEORY</td>
<td>13</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>13</td>
</tr>
<tr>
<td>3.2</td>
<td>The Model</td>
<td>13</td>
</tr>
<tr>
<td>3.3</td>
<td>The General Problem and Definitions</td>
<td>15</td>
</tr>
<tr>
<td>3.4</td>
<td>Existence Theorems</td>
<td>20</td>
</tr>
<tr>
<td>4.</td>
<td>APPLICATIONS</td>
<td>26</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>26</td>
</tr>
<tr>
<td>4.2</td>
<td>Methods</td>
<td>27</td>
</tr>
<tr>
<td>4.3</td>
<td>Discrete Distributions</td>
<td>34</td>
</tr>
<tr>
<td>4.4</td>
<td>Absolutely Continuous Distributions</td>
<td>38</td>
</tr>
<tr>
<td>5.</td>
<td>SUMMARY</td>
<td>49</td>
</tr>
<tr>
<td>6.</td>
<td>LIST OF REFERENCES</td>
<td>53</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

4.1 The U.M.V.U.E. of $\frac{x}{\theta}I_{(0,\theta)}$ ........................................ 46
1. INTRODUCTION

The problem of estimating probability density functions is of fundamental importance to statistical theory and its applications. Parzen (1962) states; "The problem of estimation of a probability density function \( f(x) \) is interesting for many reasons. As one possible application, we mention the problem of estimating the hazard, or conditional rate of failure, function \( f(x)/(1-F(x)) \), where \( F(x) \) is the distribution function corresponding to \( f(x) \)." Hájek and Šidák (1967) in the introduction to chapter one of their book make the statement; "We have tried to organize the multitude of rank tests into a compact system. However, we need to have some knowledge of the form of the unknown density in order to make a rational selection from this system. The fact that no workable procedure for the estimation of the form of the density is yet available is a serious gap in the theory of rank tests." It is implicitly understood in both of the statements cited above that the unknown density function corresponds to a member of the class of all probability distributions on the real line that are absolutely continuous with respect to Lebesgue measure. Moreover, the obvious further implication is that the problem falls within the domain of non-parametric statistics. It is not surprising, therefore, that most of the research effort on this problem in the last fifteen years or so has been directed towards developing estimation procedures that are to some extent non-parametric in character. However, like many non-parametric procedures those that have been developed for estimating density functions are somewhat heuristic. Any reasonable heuristic procedure would be expected to have good large sample properties and this is about all that could be hoped for in this
particular non-parametric setting. The problem is that any attempt to search for estimators which in some sense optimize a predetermined criterion invariably leads one to the untenable situation of considering estimators that are functions of the density function to be estimated. Hence within this non-parametric framework the efforts, for example, of Hájek and Šidák; "... to organize the multitude of rank tests into a compact system ...;" is, indeed, a difficult task. The work in this thesis is partly an attempt to resolve some of the difficulties alluded to above. The approach adopted is classical in that attention is restricted to unbiased estimators of density functions. However, the question of the existence of unbiased estimators is of fundamental importance and is one of primary consideration of the research herein.

The general setting is that of a family of probability measures $\mathcal{P}$, dominated by a $\sigma$-finite measure $\mu$ on a measurable space $(\mathcal{X}, \mathcal{G})$, with $\mathcal{P}$ the family of densities (with respect to $\mu$) that correspond to $\mathcal{P}$. It is assumed that there is available $n$ independent observations $X^{(n)} = (X_1, X_2, \ldots, X_n)$ on a random variable $X$ with unknown distribution $P$, a member of $\mathcal{P}$, and density $p \in \mathcal{P}$. The problem is, therefore, to find an estimator $\hat{p}(x) = \hat{p}(x; X^{(n)})$ such that for each $x \in \mathcal{X}$

$$E_p[\hat{p}(x)] = p(x), \quad \text{for all } P \in \mathcal{P}. \quad (1.1)$$

The symbol $E_p$ denotes mathematical expectation under the assumption that $P$ is the true distribution of $X$. Thus, with the exception of one simple but important difference, the problem resembles the usual parametric estimation problem. In the usual context an estimator is required of a
function \( \varphi(P) \) of the distribution \( P \). Here, \( \varphi \) depends on \( x \) as well as \( p \).

Thus a certain global notion has been introduced which is not usually present within the classical framework of point estimation; one might use the phrase 'uniformly unbiased' for estimators \( \hat{p} \) satisfying (1.1).

At this point it is pertinent to make a few remarks about the classical estimation problem. Very often the statistician claims he wants to estimate a function \( \varphi(P) \) using the sample \( x^{(n)} \). However, it is not always clear for what purpose the estimate will be used. If all that is required is to have an estimate of \( \varphi \) then the classical procedures, within their own limitations, are adequate for this purpose. It is often the case, however, that \( P \) (and hence \( p \)) is an explicit function of \( \varphi \) alone; that is, for each \( A \in \mathcal{G} \)

\[
P(A) = P(A; \varphi), \quad \text{for all } P \in \mathcal{P}. \tag{1.2}
\]

For example, if \( \mathcal{P} \) is the family of normal distributions on the real line with unit variance and unknown mean \( \mu \), it is common to identify an unknown member of this family by the symbol \( P_{\mu} \), and a typical function \( \varphi \) is

\[
\varphi(P_{\mu}) = \mu, \quad \text{for all } \mu \in (-\infty, \infty). \tag{1.3}
\]

Once an estimate \( \hat{\varphi} \) of \( \varphi \) has been obtained an estimate \( \hat{P} \) of \( P \) is arrived at by the 'method of substitution'; that is, \( \hat{P} \) is given by

\[
\hat{P} = P(\hat{\varphi}). \tag{1.4}
\]
This procedure implies that the ultimate objective is to estimate \( P \) (or \( p \)) rather than \( \varphi \), and, therefore, desirable criteria should refer to the estimation of \( P \) (or \( p \)) rather than \( \varphi \). This substitution method is perfectly valid if maximum likelihood estimators are required provided that very simple regularity conditions are satisfied. In the above example the mean \( \bar{X} \) is a unique minimum variance unbiased estimator of \( \mu \), but it is not the case that \( P_{\bar{X}} \) and \( p_{\bar{X}} \) are the unique minimum variance unbiased estimators of \( P_\mu \) and \( p_\mu \). It is shown in chapter four that the unique minimum variance unbiased estimator \( \hat{P}_\mu \) of \( P_\mu \) is a normal distribution with mean \( \bar{X} \) but with variance \( (n-1)/n \). This should be compared with the maximum likelihood estimator \( P_{\bar{X}} \).

It transpires that in the search for estimators satisfying (1.1) it is necessary to consider unbiased estimators \( \hat{P}(A) = \hat{P}(A; X^{(n)}) \) of the corresponding probability measure \( P \); that is, \( \hat{P} \) is an unbiased estimator of \( P \) if for each \( A \in G \)

\[
E_{\hat{P}}[\hat{P}(A)] = P(A), \quad \text{for all } P \in \mathcal{P}. \tag{1.5}
\]

It is easy to see that if \( \hat{P} \) is an unbiased estimator of \( p \) then \( \int_A \hat{P} \, d\mu \) is an unbiased estimator of \( P \), but in general the converse is not true. In chapter three of this thesis conditions for the converse to hold are established. It turns out that if an unbiased estimator \( \hat{P} \) of \( p \) exists, then it can be found immediately. The existence or non-existence of the unbiased estimator \( \hat{P} \) depends on whether or not there exists an unbiased estimator \( \hat{P} \) of the corresponding probability measure that is absolutely continuous with respect of the original dominating measure \( \mu \).
If a \( \mu \)-continuous \( \hat{\mu} \) exists, an unbiased estimator of \( \mu \) is given by the Radon-Nikodym derivative, \( \frac{d\hat{\mu}}{d\mu} \). Thus if \( \mu \) is a sub-family of Lebesgue-continuous distributions on the real line and an unbiased density estimator exists, it is obtained by differentiation of the distribution function estimator corresponding to \( \hat{\mu} \).

Chapter three also includes a more precise description of the general theoretical framework and analogous theorems on the existence of unique minimum variance unbiased estimators of density functions. The interesting result here is that if \( \mu \) admits a complete and sufficient statistic then a unique minimum variance unbiased estimator \( \hat{\mu} \) of \( \mu \) exists. Moreover, if this \( \hat{\mu} \) is absolutely continuous with respect to \( \mu \) then \( \frac{d\hat{\mu}}{d\mu} \) is the unique minimum variance unbiased estimator of \( \mu \).

The application of the general theory developed in chapter three to particular families of distributions on the real line forms the main content of chapter four. The dominating measure \( \mu \) is taken to be either ordinary Lebesgue measure or counting measure on some sequence of real numbers. In particular, the result of Rosenblatt (1956), that there exists no unbiased estimator of \( \mu \) if \( \mu \) is the family of all Lebesgue-continuous distributions on the real line, is re-established as a simple corollary to the theorems of chapter three. On the other hand, it is also true that the theorems of chapter three are generalizations of Rosenblatt's result.

Finally, the thesis is concluded with a summary and suggestions for further research.
Most of the research on density estimation that appears in the literature is concerned with the following problem. Let 
\[ X^{(n)} = (X_1, X_2, \ldots, X_n) \] 
be \( n \) independent observations on a real-valued random variable \( X \). The only knowledge of the distribution of \( X \) is that its distribution function \( F \) is absolutely continuous. \( F \) and, hence, the corresponding probability density function \( f \) are otherwise assumed to be completely unknown. By making use of the observation \( X^{(n)} \) the objective is to develop a procedure for estimating \( f(x) \) at each point \( x \). For any such estimation procedure let \( f_n(x) = f_n(x;X^{(n)}) \) denote the estimator of \( f \) at the point \( x \). Some researchers have also considered the case where \( X \) takes on values in higher dimensional Euclidean spaces. However, for the purpose of this review, no separate notation will be developed for these multivariate extensions.

At the outset it should be remarked that, within the context outlined above, a result of Rosenblatt (1956) has a direct bearing on the work of this thesis and the relevance of certain sections of the existing literature on the subject of density estimation. He showed that there exists no unbiased estimator of \( f \). As a result much of the emphasis has been on finding estimators with good large sample properties, such as consistency and asymptotic normality. Since these methods and procedures do not have a direct bearing on the present work, only a brief review of the literature on these methods and procedures will be presented here.

Very basic approaches have been adopted by Fix and Hodges (1951), Loftsgaarden and Quesenberry (1965), and Elkins (1968). In connection
with non-parametric discrimination, Fix and Hodges estimate a k-dimensional density by counting the number $N$ of sample points in k-dimensional Borel sets $\Lambda_n$. They showed that if $f$ is continuous at $x$,

$$\lim_{n \to \infty} \sup_{d \in \Lambda_n} |x-d| = 0, \quad \lim_{n \to \infty} n \lambda(\Lambda_n) = 0,$$

then $N/n \lambda(\Lambda_n)$ is a consistent estimator of $f(x)$. Here, $\lambda$ is k-dimensional Lebesgue measure and $\rho(x, y) = |x-y|$ is the usual Euclidean metric. Loftsgaarden and Quesenberry obtain the same consistency result with $\Lambda_n$ hyperspheres about $x$, and by posing the number of points and then finding the radius of $\Lambda_n$ which contains this number of points. In the two-dimensional case Elkin compares the effects of choosing the $\Lambda_n$ to be spheres or squares (centered about $x$) in the Fix and Hodges type of estimator; the criterion of comparison being mean integrated square error.

Apart from the simple estimators discussed above, two other essentially different approaches have been adopted in this complete non-parametric setting; namely, weighting function type estimators, and series expansion type estimators.

Kronmal and Tarter (1968), referring to papers on density estimation by Rosenblatt (1956), Whittle (1958) and Parzen (1962) state;

"The density estimation problem is considered in these papers to be that of finding the focusing function $\delta_m(*)$ such that using the criteria of Mean Integrated Square Error, $M. I. S. E., f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_n(x-X_i)$ would be the best estimator of the density $f$." It was shown by Watson and Leadbetter (1963) that the solution to this problem could be obtained by inverting an expression involving the characteristic function of the density $f$. Papers by Bartlett (1963), Murthy (1965, 1966),
Nadaraya (1965), Cacoullos (1966), and Woodroofe (1967) are all concerned with the properties and extensions to higher dimensions of estimators of the type described above. Anderson (1969) compares some of the above estimators in terms of the M. I. S. E. criterion.

A different approach has been exploited by Čencov (1962), Schwartz (1962) and Kronmal and Tarter (1968). The basic idea on the one-dimensional case is to use a density estimator

\[ f_n(x) = \sum_{j=1}^{q} a_{jn} \varphi_j(x) \], \quad (2.1)

where

\[ a_{jn} = \frac{1}{n} \sum_{i=1}^{n} r(x_i) \varphi_j(x_i) \]. \quad (2.2)

Here, \( q \) is an integer depending on \( n \), and the \( \varphi_j \) form an orthonormal system of functions with respect to a weight function \( r \); that is,

\[ \int_{-\infty}^{\infty} r(x) \varphi_1(x) \varphi_j(x) dx = \delta_{1j}, \quad (2.3) \]

where \( \delta_{ij} \) is the Kronecker delta function. Čencov studied the general case and its \( k \)-dimensional extension, Schwartz specialized to Hermite functions with \( r(x) = 1 \), and Kronmal and Tarter specialized to trigonometric functions with \( r(x) = 1 \).

Brunk (1965), Robertson (1966), and Wegman (1968) restrict their attention to estimating unimodal densities. The maximum likelihood
method is used and the solution can be represented as a conditional expectation given a $\sigma$-lattice. Brunk (1965) discusses such conditional expectations and their application to various optimization problems.

The research which turns out to be most relevant to the present work is to be found in papers concerned with the estimation of a distribution function $F_\theta$ at a point $x$ on the real line. Here, $\theta$ is a parameter which varies over some index set $\Theta$, which is usually a Borel subset of a Euclidean space. With one exception, all of the writers of these papers consider particular families of distributions on the real line such that $\theta$ admits a complete and sufficient statistic $T$, say. Their objective is then to find unique minimum variance unbiased estimators of $F_\theta(x)$ for these particular families. Their approach is the usual Rao-Blackwell-Lehmann-Scheffé method of conditioning a simple estimator of $F_\theta(x)$ by the complete and sufficient statistic $T$. In fact, the usual estimator is given by

$$\hat{F}(x) = P[X_1 \leq x | T]. \quad (2.4)$$

Barton (1961) obtained the estimators for the binomial, the Poisson and the normal distribution functions. Pugh (1963) obtained the estimator for the one-parameter exponential distribution function. Laurent (1963) and Tate (1959) have considered the gamma and two-parameter exponential distribution functions. Folks, et al (1965) find the estimator for the normal distribution function with known mean but unknown variance. Basu (1964) has given a summary of all these methods. Sathe
and Varde (1969) consider the estimation of the so-called reliability function, \( 1 - F_\theta(x) \). Their method is to find a statistic which is stochastically independent of the complete and sufficient statistic and whose distribution can be easily obtained. The unique minimum variance unbiased estimator is based on this distribution. For example, if \( F_\theta(x) \) is the normal distribution function with mean \( \theta \) and variance one, \( X_1 \sim \bar{X} \), say, is independent of the complete and sufficient statistic \( \bar{X} \), and has a normal distribution with mean zero and variance \( (n-1)/n \). They give as their estimator of \( 1 - F_\theta(t) \),

\[
G[[n/(n-1)]^{1/2}],
\]

(2.5)

where

\[
G(w) = \int_w^\infty \exp[-x^2/2]dx.
\]

(2.6)

They also apply their method to most of the distributions considered by the previous authors mentioned above.

As an example of more general functions of single location and scale parameters, Tate considered the problem of estimation \( P_\theta[X \in A] \); that is, \( \theta \) is the parameter in distributions given by densities,

\[
p_\theta(x) = \theta p(\theta x) \quad \text{(scale parameter)}
\]

(2.7)

and
\[ p_\theta(x) = p(x - \theta) \quad \text{(location parameter),} \quad (2.8) \]

where \( p(x) \) is a density function on the real line, and \( \theta > 0 \) in (2.7) and \( -\infty < \theta < \infty \) in (2.8). Tate relies heavily on integral transform theory. For example, he obtains the following result for the family (2.7), under the assumption that \( p(x) \) vanishes for negative \( x \). Let \( H(X^{(n)}) \) be a non-negative homogeneous statistic of degree \( \alpha \neq 0 \) with density \( \theta^\alpha g(\theta^\alpha x) \), and suppose \( g(x) \) and \( \theta p(\theta z) \) (for some fixed positive number \( z \)) both have Mellin transforms. Then, if there exists an unbiased estimator \( \hat{p} \) of \( \theta p(\theta z) \) with a Mellin transform it will be given by

\[ \hat{p}(z; H) = \frac{\alpha}{H} M^{-1}\left[ \frac{M[p(x); \alpha(s-1)+1]}{z^\alpha(s-1)+1} \frac{1}{M[g(x); s]} \right], \quad (2.9) \]

where \( M \) is the Mellin transform defined for a function \( \zeta(x) \), when it exists, by

\[ M[\zeta(x); s] = \int_0^\infty x^{s-1} \zeta(x) dx, \quad s_0 < \text{Re}(s) < s_1; \quad (2.10) \]

and \( M^{-1} \) denotes the inverse function. Tate also obtained many of the previously mentioned results. Washio, et al (1956), again using integral transform methods, study the problem of estimating functions of the parameter in the Koopman-Pitman family of densities (cf. Koopman (1936) and Pitman (1936)). To illustrate their method they consider, as one example, the problem of estimating \( P_\theta[X \in A] \) for the normal
distribution with unknown mean and variance. In an earlier paper Kolmogorov (1950) derived an estimator of $P_\theta[X \in A]$ for the normal distribution with unknown mean and known variance. His approach was to first obtain a unique minimum variance unbiased estimator of the density function (which he assumed existed) and then to integrate the solution on the set $A$. In his derivation he used the 'source solution' of the heat equation

$$
\psi(z, t) = (2\pi t)^{1/2} \exp[-z^2/4t], \ 0 < t < \infty, \ -\infty < z < \infty, \quad (2.11)
$$

to solve the integral equation that resulted directly from the estimation problem.
3. GENERAL THEORY

3.1 Introduction

In this chapter the general set-up of the basic statistical model is described, definitions of unbiasedness for probability measures and densities are given, and two basic theorems are established. The first of these two theorems gives necessary and sufficient conditions for the existence of an unbiased estimator of a density function, and under the additional assumptions of sufficiency and completeness the second theorem gives similar necessary and sufficient conditions for the existence of a unique minimum variance unbiased estimator of a density function. Moreover, these theorems also show that if, in any particular example, an unbiased estimator of the density function exists, then the estimator can be computed immediately.

3.2 The Model

In this section the set-theoretical details of the general statistical model are described. Let $(\mathcal{X}, \mathcal{G}, \mu)$ be a $\sigma$-finite Euclidean measure space; that is, $\mathcal{X}$, with generic element $x$, is a Borel set in a Euclidean space, $\mathcal{G}$ is the class of Borel subsets of $\mathcal{X}$, and $\mu$ is a $\sigma$-finite measure on $\mathcal{G}$. Furthermore, suppose that there is given a family $\mathcal{P}$ of probability measures $P$ on $\mathcal{G}$ which is dominated by $\mu$; that is, each $P \in \mathcal{P}$ is absolutely continuous with respect to $\mu$. Then by the Radon-Nikodym theorem there exists, for each $P$, a $\mu$-unique, finite valued, non-negative, $\mathcal{G}$-measurable function $p$, such that
The family $p$, of functions $p$, will be referred to as the family of density functions (or densities) corresponding to $P$.

Let $X$ be a random variable over the space $(\mathcal{X}, \mathcal{G})$. It will be assumed that the probability distribution of $X$ identifies with an unknown member $P$ of $\mathcal{F}$. Let $X^{(n)} = (X_1, X_2, \ldots, X_n)$ be $n$ independent random variables that are identically distributed as $X$, and denote by $x^{(n)} = (x_1, x_2, \ldots, x_n)$ an observation on $X^{(n)}$. Denote by $\mathcal{X}^{(n)}$ the sample space of observations $x^{(n)}$, and $\mathcal{G}^{(n)}$ the product $\sigma$-algebra of subsets of $\mathcal{X}^{(n)}$ that is determined by $\mathcal{G}$ in the usual way. For any measure $Q$ on $\mathcal{G}$ the corresponding product measure on $\mathcal{G}^{(n)}$ will be denoted by $Q^{(n)}$.

All statistics to be considered are $\mathcal{G}^{(n)}$-measurable functions on $\mathcal{X}^{(n)}$ to a Euclidean measurable space $(\mathcal{J}, \mathcal{B})$. If $T = T(x^{(n)})$ is such a statistic, then denote by $P_T$ that family of probability measures $P_T$ on $\mathcal{B}$ induced by $T$; that is, for each $B \in \mathcal{B}$

$$P_T(B) = P^{(n)}[x^{(n)} : T(x^{(n)}) \in B], \quad \text{for all } P \in \mathcal{F}. \quad (3.2)$$

Since both of the spaces $(\mathcal{X}, \mathcal{G})$ and $(\mathcal{J}, \mathcal{B})$ are Euclidean it will be assumed that all conditional probability distributions are regular; that is, if $T$ is a statistic and $Q$ an arbitrary probability measure on $\mathcal{G}^{(n)}$, then a conditional probability function $Q^T$ defined, for each $A \in \mathcal{G}^{(n)}$ and $T = t$, by
\[ \int_B \mathbb{Q}_T^t(dQ_T(t)) = \mathbb{Q}[A \cap (T \in B)], \quad \text{for all } B \in \mathfrak{A}, \quad (3.3) \]

is, for each fixed \( t \), a probability measure on \( \mathfrak{A}(n) \). Theorems 4 and 7 in chapter two of Lehmann(1959), for example, validate this assumption. Actually, the assumption of Euclidean spaces is unnecessarily restrictive. As long as there exist regular conditional probability measures all the results of this chapter will still be valid. Of course, from a practical viewpoint, spaces other than Euclidean spaces are of little interest. This will be true in the next chapter where \( \mathcal{Y} \) will be taken to be a Borel subset of the real line.

Finally, it will be assumed that the definitions of completeness and sufficiency for a family of probability measures are known, and hence no separate definitions of these notions will be given.

3.3 The General Problem and Definitions

In this section the general problem of density estimation is discussed within the framework of decision theory, and definitions of unbiased estimators for probability measures and probability density functions are given to prepare the way for the theorems of the next section.

The problem to be considered here is typical of a large class of statistical problems. A set of observations \( x^{(n)} \) are available that are assumed to have been generated from a distribution \( P \) with corresponding density \( p \). All that is assumed about \( P \) (or) is that it is a member of some particular class of probability measures (probability
density functions) $p(x)$. Hence, given the observations $x^{(n)}$, the problem here is to estimate $p$ in some optimal way. Viewed from the standpoint of decision theory an estimator $\hat{p} = \hat{p}(x;X^{(n)})$ of $p$ at a point $x \in X$ is required that minimizes the risk

$$R_{\hat{p}}(x)(P) = E_P[L(\hat{p}(x;X^{(n)}), p(x))], \quad \text{for all } P \in P. \quad (3.4)$$

The symbol $E_P$ represents mathematical expectation under the assumption that $p^{(n)}$ is the true distribution of $X^{(n)}$, and $L(.,.)$ is a loss function which expresses the loss incurred (monetary or otherwise) when estimating the density as $p(x)$ when, in fact, $p(x)$ is the true density at $x$. Rather than think in terms of estimating $p$ at some fixed point $x$ it is often desirable to think in terms of estimating the entire density function. Hence the problem becomes one of selecting a $\hat{p}$ which not only minimizes the risk in $(3.4)$ for all $P \in P$, but also, simultaneously, minimizes it for all $x \in X$. Minimum risk estimators are rarely obtainable so that the additional complication introduced by this global notion increases the difficulty in general. This problem can be circumvented in the following way. Let $w$ be a measure on $G$ which in some way expresses the intensity of interest of the statistician in the points of $X$ with reference to the estimation of $p$. For example, if $X$ is the real line and $w$ is Lebesgue measure, the intensity of interest is uniform over the points of $X$. The problem becomes one of finding a $\hat{p}$ which minimizes
\[ R_P(P) = \int \mathcal{L} R_P(x)(P) dw(x), \quad \text{for all } P \in \mathcal{P}. \quad (3.5) \]

To overcome the difficulty of finding estimators which minimize, for all \( P \in \mathcal{P} \), the risk function in (3.4) and the integrated risk function in (3.5) the notions of minimax estimators and Bayes estimators have been introduced. The Bayes procedure in the context of integrated risk, for example, introduces yet another averaging process; a prior distribution \( \pi \) on \( \mathcal{P} \). In this case the Bayes estimator would be the one that minimizes the Bayes integrated risk

\[ R_\mathcal{P}^\pi = \int \mathcal{P} R_P(P) d\pi(P). \quad (3.6) \]

The approach to this problem of density estimation adopted in this thesis is classical. The loss function is squared error loss; that is

\[ L(\hat{p}(x), p(x)) = (\hat{p}(x) - p(x))^2. \quad (3.7) \]

The object will be to find unbiased estimators \( \hat{p} = \hat{p}(x; X^{(n)}) \), if they exist, that minimize the risk function in (3.4) for all \( (x, P) \in \mathcal{X} \times \mathcal{P} \). Note that, with the loss function in (3.7) and the restriction to unbiased estimators, the risk function in (3.4) becomes the variance

\[ \text{Var}_P[\hat{p}(x)] = E_P[\hat{p}(x) - p(x)]^2, \quad (3.8) \]
of \( \hat{p} \) at the point \( x \). It is convenient at this point to give a
definition of an unbiased estimator of a density function. In general
the estimator will depend on a statistic \( T = T(x^{(n)}) \); for example,
\( T = x^{(n)} \). The notation developed in the previous section concerning
such statistics will be adhered to in the definition below and the
subsequent ones in the remainder of this section.

**Definition 3.1** A real-valued function \( \hat{p} = \hat{p}(x; t) \) on \( \mathcal{X} \times \mathcal{F} \) is an
unbiased estimator of a density \( p \in \mathcal{P} \) if it is \( \mathcal{G} \times \mathcal{B} \) measurable and
satisfies

\[
E_T[\hat{p}(x; t)] = p(x), \quad \text{a.e. } \mu, \quad \text{for all } p \in \mathcal{P}. \tag{3.9}
\]

Note, \( E_T \) is an abbreviation for \( E_{p_T} \).

Notice that this definition does not require that an unbiased
estimator have the properties of a density, such as, being non-
negative, and integrating to unity. Notice also that the definition
allows for the estimator to be biased on a subset of \( \mathcal{X} \) with \( \mu \)-measure
zero. This is a technical necessity the reason for which will be made
clear in the proofs of the theorems in the next section. Apart from
this technical detail the above definition has, therefore, a certain
global content; an unbiased estimator of the entire function is
required.

Hence, with minor modifications, the problem is not new. The
search is for minimum variance unbiased estimators of density functions.
The solution to the problem has also not changed; if \( \mathcal{P} \) admits a complete
and sufficient statistic then unbiased estimators based on this statistic will be unique minimum variance unbiased. However, the question of existence is the fundamental problem that arises. In this connection it is convenient to consider unbiased estimators of the corresponding probability measure $P$. Therefore, the following definition will formalize this notion.

**Definition 3.2** A real-valued function $\hat{P} = \hat{P}(A; t)$ on $G \times \mathcal{F}$ is an unbiased estimator of $P \in \mathcal{P}$ if for each $A \in \mathcal{G}$ it is a $\mathcal{G}$-measurable function and satisfies

$$E_T[\hat{P}(A; T)] = P(A), \quad \text{for all } P \in \mathcal{P}.$$  \hspace{1cm} (3.10)

Notice that this definition does not require $\hat{P}$ to be a probability measure or, in fact, a measure on $G$ for each fixed $t$. However, in what follows, it will be assumed that, for each $t \in \mathcal{F}$, such estimators are, in fact, measures on $G$; and only such estimators will be considered in the sequel. In the following development it will become clear that it is desirable to restrict attention to estimators $\hat{P}$ which are measures.

**Definition 3.3** An estimator $\hat{P} = \hat{P}(A; T)$ of $P \in \mathcal{P}$, which is a measure on $G$ for each $t \in \mathcal{F}$, is said to be absolutely continuous with respect to a $\sigma$-finite measure $\mu$ on $G$ if, for every $A \in \mathcal{G}$ for which $\mu(A) = 0$, $\hat{P}(A; t) = 0$, a.e. $P_T$; that is, except for a set of $t$-values with $P_T$-measure zero, for all $P_T \in \mathcal{P}_T$.

A consequence for this definition is that if $\hat{P}$ is such an estimator
then, by the Radon-Nikodym Theorem, there exists, almost everywhere \( P_n \), an \( \mathcal{G} \)-measurable, \( \mu \)-unique, non-negative function \( f(x; t) \), such that

\[
\hat{P}(A; t) = \int_A f(x; t) \, d\mu(x), \quad \text{for every } A \in \mathcal{G}. \tag{3.11}
\]

Although it is \( \mathcal{G} \)-measurable, \( f(x; t) \), in general, will not be \( \mathcal{G} \times \mathcal{G} \)-measurable. The \( \mathcal{G} \times \mathcal{G} \)-measurability of such functions \( f(x; t) \) is necessary for the application of Fubini's theorem in the proofs of the theorems in the next section. Hence, it will be assumed, in the sequel, that all such \( \mu \)-continuous estimators \( \hat{P} \) of \( P \) have \( \mathcal{G} \times \mathcal{G} \)-measurable Radon-Nikodym derivatives.

### 3.4 Existence Theorems

In this section the problems concerning the existence of an unbiased estimator and a unique minimum variance unbiased estimator of a density function are solved. A desirable property of these solutions is that they also provide a method for computing such an estimator, should it exist.

The terminology, notations, and conventions, introduced in the previous sections of this chapter, will be adhered to throughout the remainder of the present section.

Theorem 3.1, below, is concerned with the existence of an unbiased estimator of a density \( p \) which is assumed to be a member of a family of densities \( p \) corresponding (with respect to a \( \sigma \)-finite dominating measure \( \mu \)) to a family \( \mathcal{P} \) of probability measures \( P \).
Theorem 3.1 An unbiased estimator of $p$ exists if and only if there exists an unbiased estimator $\hat{P}$ of $P$ that is absolutely continuous with respect to the original dominating measure $\mu$. Moreover, when such a $\hat{P}$ exists, an unbiased estimator of $p$ is given by the Radon-Nikodym derivative, $\frac{d\hat{P}}{d\mu}$.

Proof: Let $\hat{P} = \hat{P}(\cdot;T)$ be an unbiased estimator of $P$ that is absolutely continuous with respect to $\mu$. Then, by the unbiasedness property, for each $A \in G$

$$E_T[\hat{P}(A;T)] = P(A) = \int_A p(x)\mu(x), \text{ for all } P \in \mathcal{P}. \hspace{1cm} (3.12)$$

Also, by the Radon-Nikodym theorem, there exists a non-negative function $\hat{p} = \hat{p}(x;t)$ such that for each $A \in G$,

$$\hat{P}(A;t) = \int_A \hat{p}(x;t)\mu(x), \text{ a.e. } P_T. \hspace{1cm} (3.13)$$

It follows, on taking expectations (with respect to $P_T$) in (3.13), and applying Fubini's theorem to the third member of (3.14) below, that, for each $A \in G$,

$$P(A) = E_T[\hat{P}(A;T)] = E_T[\int_A \hat{p}(x;t)\mu(x)] = \int_A E_T[\hat{p}(x;T)]\mu(x), \text{ for all } P \in \mathcal{P}. \hspace{1cm} (3.14)$$

Then, by the uniqueness part of the Radon-Nikodym theorem, the desired
result follows from (3.12) and (3.14); that is,

\[ E_T[\hat{p}(x;T)] = p(x), \quad \text{a.e. } \mu, \quad \text{for all } p \in \mathcal{P}. \quad (3.15) \]

Now suppose that \( \hat{p} = \hat{p}(x;T) \) is an unbiased estimator of \( p \); that is, (3.15) is true. It then follows, on applying Fubini's theorem to the second member of (3.16) below, that, for each \( A \in \mathcal{G} \),

\[ P(A) = \int_A E_T[\hat{p}(x;T)] \, d\mu(x) = E_T[\int_A \hat{p}(x;T) \, d\mu(x)], \quad \text{for all } P \in \mathcal{P}. \quad (3.16) \]

Hence, from (3.16), \( \int_A \hat{p} \, d\mu \) is an unbiased estimator of \( P(A) \) for every \( A \in \mathcal{G} \), and is clearly absolutely continuous with respect to \( \mu \). This completes the proof.

Thus, if an unbiased estimator of \( P \) is available that happens to be absolutely continuous with respect to \( \mu \), it can be immediately concluded that an unbiased estimator of \( p \) exists, and such an estimator is given by the \( \mu \)-derivative of the estimator of \( P \). Furthermore, if there exists a complete and sufficient statistic for \( \mathcal{P} \), the usual methods of the Rao-Blackwell-Lehmann-Scheffé theory can be used to obtain a unique minimum variance unbiased estimator of \( p \) at every point \( x \in \mathcal{X} \). However, the following lemma and theorem give another method of achieving the same end.

Denote by \( \mathcal{C} \) the class of subsets \( E \) of the form \( A \times \mathcal{X}^{(n-1)} \) with \( A \) an element of \( \mathcal{G} \). Thus, \( \mathcal{C} \) is a class of cylinder sets with bases in the first coordinate space of \( \mathcal{X}^{(n)} \). Clearly, \( \mathcal{C} \) is a sub-\( \sigma \)-algebra
of \( G^{(n)} \). If \( T \) is a complete and sufficient statistic for \( \mathcal{P} \), the conditional probability measure \( P^T \) on \( G^{(n)} \) is, therefore, constant in \( P \in \mathcal{P} \). Denote by \( P^T_\mathcal{E} \) the restriction of \( P^T \) to the class of sets \( \mathcal{E} \), and define an estimator \( \hat{P} \) of \( P \) by

\[
\hat{P}(A;T) = P^T_\mathcal{E}(E), \quad \text{for all } E \in \mathcal{E}, \tag{3.17}
\]

where \( E = A \times X^{(n-1)} \).

For each \( t \in J \), \( P^t_\mathcal{E} \) is a probability measure on \( \mathcal{E} \) and, therefore, \( \hat{P} \), defined in (3.17), is a probability measure on \( G \) for each such \( t \).

The following lemma is now easy to prove.

**Lemma 3.1** If \( T \) is a complete sufficient statistic for a family \( \mathcal{P} \) of probability measures \( P \), then \( \hat{P} = P^T_\mathcal{E} \) is the unique minimum variance unbiased estimator of \( P \).

**Proof:** Put \( B = J \) and identify \( P^{(n)} \) with \( Q \) in (3.3). Then, for every \( E = A \times X^{(n-1)} \in \mathcal{E} \) it follows that

\[
E[\hat{P}(A;T)] = P^{(n)}(E) = P(A), \quad \text{for all } P \in \mathcal{P}. \tag{3.18}
\]

Hence \( \hat{P} = P^T_\mathcal{E} \) is an unbiased estimator of \( P \) and, being a function of \( T \) alone, is, moreover, the unique minimum variance unbiased estimator of \( P \); and this completes the proof of the lemma.

It should be noted that the \( \hat{P} \) defined above is, in fact, a probability measure on \( G \) for each \( t \in J \). This property has an important
implication in the next theorem which is concerned with both the existence and computation of unique minimum variance unbiased estimators of density functions. Note also that $\mathcal{E}$ could have been defined with the bases of its sets in any one of the coordinate spaces of $X^{(n)}$. The notation of the previous lemma will be used without further comment.

**Theorem 3.2** Let $T$ be a complete and sufficient statistic for $\mathcal{P}$. A unique minimum variance unbiased estimator $\hat{p}$ of $p \in \mathcal{P}$ exists if and only if $\hat{P} = P^T_E$ is absolutely continuous with respect to the original dominating measure $\mu$. Moreover, if such a $\hat{p}$ exists it is given by the Radon-Nikodym derivative, $\frac{dP_E^T}{d\mu}$.

**Proof:** If $\hat{P} = P^T_E$ is absolutely continuous with respect to $\mu$, then, by the above lemma and theorem 3.1 an unbiased estimator of $p$ is given by $\frac{d\hat{P}}{d\mu}$. But, $\frac{d\hat{P}}{d\mu}$ is a function of $T$ alone and hence is the unique minimum variance unbiased estimator of $p$.

On the other hand, if $\hat{p}$ is the unique minimum variance unbiased estimator of $p$, it must be a function, $\hat{p}(x;T)$, of $T$ alone. Hence, for each $A \in G$,

$$P(A) = \int_A E_p^{\hat{p}}[\hat{p}(x;T)]d\mu(x) = E_T^{\hat{P}}[\int_A \hat{p}(x;T)d\mu(x)], \text{ for all } P \in \mathcal{P}. \quad (3.19)$$

But, by the lemma $\hat{P} = P^T_E$ is an unbiased estimator of $P$ and, therefore, for each $A \in G$,

$$E_T^{\hat{P}}[\hat{P}(A;T) - \int_A \hat{p}(x;T)d\mu(x)] = 0, \text{ for all } P \in \mathcal{P}. \quad (3.20)$$
But, $T$ is complete, and hence

$$
\hat{P}(A; t) = \int_A \hat{p}(x; t) d\mu(x), \quad \text{a.e. } \nu_T
$$

(3.21)

Hence, $\hat{P} = P^T$ is absolutely continuous with respect to $\mu$. Finally, it should be noted that if $P^T_T$ is absolutely continuous with respect to $\mu$,

$$
dP_T^T
$$

then $\frac{dP_T^T}{d\mu}$ is, in fact, a proper density function. This completes the proof.

Hence, once $P^T_T$ has been computed, it is only necessary to check whether or not it is absolutely continuous with respect to $\mu$; and if it is, the unique minimum variance unbiased estimator of $P$ is obtained by differentiating $P^T_T$ with respect to $\mu$. For example, if $\mathcal{P}$ is a family of probability measures on the real line with $\mathcal{F}$ the corresponding family of distribution functions, and $\mu$ is ordinary Lebesgue measure, then one needs only to check if $\hat{F}$, the distribution function corresponding to $\hat{P}$, is an absolutely continuous function of $x$. If it is, the unbiased estimator, at $x$, is given by the ordinary derivative $\frac{d\hat{F}(x)}{dx}$. Such families of distributions on the real line will be the object of study in the next chapter.
4. APPLICATIONS

4.1 Introduction

In this chapter applications of the general theory presented in the previous chapter are discussed for families of distributions on the real line; that is, the basic space is a Borel set of the real line together with its Borel subsets, and the families of densities and corresponding families of distribution functions are taken to be Borel-measurable functions. Without exception, the dominating measure will be either ordinary Lebesgue measure or counting measure.

In the search for an unbiased estimator of a density function the general method will be to first find an unbiased estimator of the corresponding distribution function and then test it for absolute continuity. If, then, the distribution function estimator is absolutely continuous an unbiased estimator of the density function is obtained by differentiating this distribution function estimator. Most of the families of distributions that are considered have the feature that they admit a complete and sufficient statistic. In these cases the search for estimators will be restricted to those which are functions of the complete and sufficient statistic, and then the unbiased estimators, if they exist, will also be unique and have uniform minimum variance. Before proceeding to particular examples, three different methods of finding unbiased estimators of distribution functions will be discussed. The choice of method is sometimes dictated by the structure of the particular family of distributions under consideration, whereas in some situations computational ease is
the only criterion.

Essentially, the remainder of the chapter will be divided into two main parts; one part dealing with discrete distributions and the other with families of absolutely continuous distributions. In the discrete case a solution is given for a one-parameter, power series type family which includes several of the well-known discrete distributions with possible values on the set \( \{0, 1, 2, \ldots\} \). In the absolutely continuous case many of the well-known families of distributions are discussed including a general truncation-parameter family. Rosenblatt's result, on the non-existence of an unbiased estimator of a density function that corresponds to an unknown absolutely continuous distribution function on the real line, is also re-established as a simple corollary to Theorem 3.2 of the previous chapter.

4.2 Methods

The literature, to date, contains essentially three different methods for finding unbiased estimators of distribution functions. Actually, all three methods can be used to find unbiased estimators of more general parametric functions. These methods are given below. However, before giving a formal statement of each method, a short discussion will first be given comparing them on the basis of their scope of applicability and computational ease.

The first method, due to Tate (1959) and Washio, et al (1956), makes use of integral transform theory. In general, the application of this theory does not depend on the existence of sufficient
statistics, although Washio, et al do restrict their study to the Koopman-Pitman family of densities (cf. Koopman (1936) and Pitman (1936)). Recall that this family, by definition, admits a sufficient statistic and has range independent of the parameter. On the other hand, Tate restricts his attention to finding estimators which are functions of homogeneous statistics and statistics with the translation property for the scale and location parameter families, respectively.

In what follows interest will be confined to finding unbiased estimators of distribution functions and in most of the examples treated these integral transform methods would be somewhat heavy-handed and the other two methods, described below, are usually easier to apply.

The second and third methods depend, for their application, on the existence of sufficient or complete and sufficient statistics. The second method is a straightforward application of the well-known Rao-Blackwell and Lehmann-Scheffé theory, and the theorem due to Basu (1955, 1958) is the connecting link between these two methods. Whereas the third method is generally the easiest to apply, its application depends, somewhat artificially, on one being able to find a statistic with distribution independent of the index parameter of the family of distributions under consideration. On the other hand, the distribution problems encountered in applying the Rao-Blackwell-Lehmann-Scheffé theory are often difficult and laborious.

Throughout the remainder of this chapter the following conventions,
notation, and terminology will be adopted. \( \mathcal{X} \) will denote a Borel set of the real line, \( \mathbb{R} \); \( \mathcal{G} \), the Borel subsets of \( \mathcal{X} \); \( \lambda \), ordinary Lebesgue measure; \( \nu \), counting measure when \( \mathcal{X} \) is a finite or infinite sequence of points, \( \{a_n : n \geq 1\} \); \( p_\theta \), \( F_\theta \), and \( p_\theta \), the unknown probability measure, distribution function, and density function, respectively; with \( \theta \) an element of a parameter space \( \Theta \). In each example it will be assumed that the estimation procedure will be based on a statistic \( T \) which is a function of \( n \) independent observations, \( x^{(n)} = (X_1, X_2, \ldots, X_n) \). In all the examples considered, \( T \) will be a complete and sufficient statistic. Each example will be characterized by a vector, \( \{p_\theta, x, \Theta; T\} \).

Finally, \( \hat{F}(x) = \hat{F}(x; T) \) and \( \hat{p}(x) = \hat{p}(x; T) \) will denote, respectively, the estimators of \( F_\theta \) and \( p_\theta \); and \( \hat{F}(A) = \hat{F}(A; T) \), the estimator of \( p_\theta \), will be given by

\[
\hat{F}(A) = \int_A d\hat{F}(x), \quad \text{for all } A \in \mathcal{G}.
\] (4.1)

The integral in (4.1) is, for each \( t \in \mathcal{T} \), a Lebesgue-Stieltjes integral but in all examples can be evaluated as a Riemann-Stieltjes integral.

**Method 1.** The transform method considered here will not be discussed in any generality but will be illustrated by the work of Tate in section four of his paper where he applies the Mellin transform to find unbiased estimators of arbitrary functions of scale parameters. The Mellin transform of a function \( \mathfrak{f}(x) \), when it exists, is defined by
\[ M[J(x);s] = \int_0^\infty x^{s-1} J(x)dx, \quad s_0 < \text{Re}(s) < s_1. \quad (4.2) \]

If \( \delta(s) \) is such a transform the inverse transform will be denoted by

\[ M^{-1}[\delta(s);x]. \quad (4.3) \]

Let \( p(x) \) be a known, piecewise continuous density that vanishes for negative \( x \). Consider the following scale parameter family, defined by

\[ p_{\theta}(x) = \theta p(\theta x), \quad 0 < \theta < \infty. \quad (4.4) \]

Tate considers the problem of estimating \( \theta p(\theta z) \) for a fixed value \( z \).

He considers only those estimators that are functions of non-negative homogeneous statistics; that is, \( H \) is such a homogeneous statistic of degree \( \alpha \neq 0 \) if,

(i) \( H: [0,\infty)^{(n)} \rightarrow [0,\infty) \),

(ii) \( H(\theta x^{(n)}) = \theta^\alpha H(x^{(n)}), \quad 0 < \theta < \infty. \quad (4.5) \)

When \( X^{(n)} \) is a simple random sample from a member of the family defined in (4.4), \( H(X^{(n)}) \) has a density of the form \( \hat{p}(z;H) \). Thus an unbiased estimator \( \hat{p}(z;H) \), if it exists, is a solution of the integral equation
\[ \int_{0}^{\infty} \hat{p}(z; h) e^{\alpha h} dh = \Theta p(\Theta z). \] (4.6)

Assume that both \( g(x) \) and \( \Theta p(\Theta z) \) have Mellin transforms. Then, if there exists an unbiased estimator of \( \Theta p(\Theta z) \) with a Mellin transform, it will be given by

\[ p(z; H) = \frac{\alpha}{H} M^{-1} \left\{ \frac{M[p(x); \alpha(s-1)+1]}{z^{\alpha(s-1)+1} M[g(x); s]} ; \frac{1}{H} \right\}. \] (4.7)

Tate uses the Laplace and Bilateral Laplace transforms to find unbiased estimators of functions of location parameter families.

**Method 2.** This method, as noted in the discussion above, is a straightforward application of the Rao-Blackwell-Lehmann-Scheffé theory. To find the unique minimum variance unbiased estimator of \( F_\Theta \) at a fixed point \( x \) a simple unbiased estimator is first found which when conditioned on the complete and sufficient statistic yields the required result. Usually, the simple estimator is

\[ I_A(X_1), \] (4.8)

where \( I_A \) denotes the indicator function of the set \( A = (-\infty, x] \) and \( X_1 \) is the first observation from a sample of \( n \). Thus, if \( T \) is a complete and sufficient statistic for \( \Theta \) the minimum variance unbiased estimator of \( F_\Theta(x) \) is given uniquely by
\[ \hat{F}(x; T) = P[X_\perp \leq x | T]. \quad (4.9) \]

Clearly, in the discrete case the minimum variance unbiased estimator of the probability mass function can be obtained directly; absolute continuity of the distribution function estimator with respect to counting measure is guaranteed. Hence, the required density estimator in the discrete case is given by

\[ \hat{p}(x; T) = P[X_\perp = x | T], \quad (4.10) \]

at the point \( x \).

Method 3. This method, due to Sathe and Varde (1969), depends, for its application, on a theorem of Basu (1955, 1958). The original result will be specialized as follows. Denote by \( U \) the simple estimator given in (4.8) and suppose \( T \) is a complete and sufficient statistic for \( \theta \). If there exists a function \( V = V(X_\perp, T) \) such that,

1. it is stochastically independent of \( T \),
2. it is a strictly increasing function of \( X_\perp \) for fixed \( T \),

then the minimum variance unbiased estimator of \( F_\theta \) at \( x \) is given uniquely by
\[
\hat{F}(x; T) = \begin{cases} 
1 & \text{if } V(x, T) > b, \\
\int_a^{V(x, T)} dH(u) & \text{if } a < V(x, t) \leq b, \\
0 & \text{if } V(x, T) \leq a,
\end{cases} \tag{4.11}
\]

where \( H \) is the distribution function of \( V \) and satisfies
\[
H(x) = \begin{cases} 
0 & \text{for } x \leq a, \\
1 & \text{for } x > b. 
\end{cases} \tag{4.12}
\]

The proof is immediate by noting
\[
\hat{F}(x; T) = P[X_1 \leq x | T] = P[V \leq V(x, T) | T] \text{ by (ii)} = P[V \leq V(x, T)] \text{ by (i)}.
\]

The important step in the application of this theorem is the selection of a suitable statistic \( V \). The following theorem, due to Basu (1955, 1958), gives sufficient conditions for \( V \) to satisfy (i).

**Theorem 4.1** If \( T \) is a boundedly complete and sufficient statistic for \( \theta \) and \( V \) is a statistic (which is not a function of \( T \) alone) that has distribution independent of \( \theta \) then \( V \) is stochastically independent of \( T \).
Therefore, once $V$ is found and its distribution has been derived, the required estimator is given by (4.11).

### 4.3 Discrete Distributions

In this section method 2 will be used to find unique minimum variance unbiased estimators of discrete-type density functions. Included is a general expression for an estimator for a wide class of discrete distributions. These distributions, first introduced by Noak (1950) and later studied in detail by Roy and Mitra (1957), include many of the well-known discrete distributions.

**Example 4.1: The Binomial Distribution.** This class of distributions, in the notation introduced in section 4.2, is characterized by the vector

$$\{\binom{k}{x} \theta^x (1-\theta)^{k-x}, (0, 1, \ldots, k); \sum_{i=1}^{n} x_i\}$$

where $k$ is a known positive integer.

The distribution of $T$ is given by

$$\binom{nk}{t} \theta^t (1-\theta)^{nk-t}, \quad t \in \{0, 1, \ldots, nk\};$$

and the conditional distribution of $X^{(n)}$ given $T = t$ is

$$\prod_{i=1}^{n} \binom{k}{x_i} \binom{nk}{t}, \quad x^{(n)} \in S_n(t),$$

where $S_n(t) = \{x^{(n)} \in \mathcal{X}^{(n)}: \sum_{i=1}^{n} x_i = t\}$. 

Hence,

\[ \hat{p}(x;t) = \sum_{i=1}^{n} \frac{k}{t} \binom{\binom{k}{x}}{t} \cdot x \in \{0, 1, \ldots, \min(t, k)\}, \quad (4.16) \]

where the summation in (4.16) is over the set \( \{x(n): x^{(n-1)} \in S_{n-1}(t-x), x_n = x\} \). This expression simplifies to give the required estimator

\[ \hat{p}(x;T) = \frac{k}{x} \binom{nk-k}{T-x} \binom{nk}{T} \cdot x \in \{0, 1, \ldots, \min(T,k)\}. \quad (4.17) \]

It is interesting to note that the binomial distribution is estimated by a hypergeometric distribution with a range that has random end-point, \( \min(T,k) \).

**Example 4.2: The Noak Distributions.** These distributions are characterized by the vector

\[ \{a(x)\theta^x/f(\theta), (0, 1, \ldots), (0, c); \sum_{i=1}^{n} X_i\}, \quad (4.18) \]

where \( c \) is finite or infinite, \( a(x) > 0, a(0) = 1 \) and

\[ f(\theta) = \sum_{x=0}^{\infty} a(x)\theta^x. \quad (4.19) \]

The Poisson, negative binomial and logarithmic distributions occur as special cases of the above. They will be considered separately to
illustrate Theorem 4.2 below.

Roy and Mitra proved that \( T = \sum_{i=1}^{n} X_i \) is a complete and sufficient statistic for \( \theta \) and has distribution

\[
C(t, n) \theta^t / f^n(\theta), \quad t \in \{0, 1, \ldots\};
\]

where

\[
C(t, n) = \sum_{i=1}^{n} a(x_i).
\]

The summation in (4.21) is over the set \( S_n(t) = \{x^{(n)} \in X^{(n)}: \sum_{i=1}^{n} x_i = t\} \).

The above results and notation will be used in the statement and proof of the following theorem.

**Theorem 4.2** The minimum variance unbiased estimator of a Poisson density function is given by

\[
\hat{p}(x; T) = \frac{a(x) C(T-x, n-1)}{C(m, n)}, \quad x \in \{0, 1, \ldots, T\}.
\]

**Proof:** The conditional distribution of \( X^{(n)} \) given \( T = t \) is given by

\[
\prod_{i=1}^{n} \frac{a(x_i)}{C(t, n)}, \quad x^{(n)} \in S_n(t).
\]
Hence,

\[ \hat{p}(x,t) = \sum_{i=1}^{n} a(x_i)/C(t,n), \quad x \in \{0, 1, \ldots, t\}, \quad (4.24) \]

where the summation is over the set \( \{x^{(n)}: x^{(n-1)} \in S_{n-1}(t-x), x_n = x\} \)
the expression in (4.24) simplifies to give the required estimator.

The above result will now be illustrated for the Poisson, negative binomial and logarithmic distributions.

**The Poisson Distribution.** For this distribution,

\[ a(x) = 1/x!; \quad f(\theta) = e^\theta; \quad c = +\infty; \quad C(t,n) = n^t/t!. \]

Therefore,

\[ \hat{p}(x,T) = \binom{T}{x} \left(\frac{1}{n}\right)^x \left(\frac{1-n}{n}\right)^{T-x}, \quad x \in \{0, 1, \ldots, T\}. \quad (4.26) \]

Hence, the Poisson distribution is estimated by a binomial distribution with parameters \( 1/n \) and \( T \).

**The Negative Binomial Distribution.** For this distribution,

\[ a(x) = \binom{k+x-1}{x}; \quad f(\theta) = (1-\theta)^{-k}; \quad c = 1; \quad C(t,n) = \binom{nk+t-1}{t}, \quad (4.27) \]

where \( k \) is a positive integer. Therefore,
\[ \hat{p}(x; T) = \binom{k+1}{x} \binom{nk-k+T-x-1}{T-x} / \binom{nk+T-1}{T}, \quad x \in \{0, 1, \ldots, T\}. \quad (4.28) \]

Note that the minimum variance unbiased estimator for the geometric distribution is obtained by putting \( k = 1 \) in (4.28). For fixed \( T = t \), \( \hat{p}(x; t) \) has the following interpretation. Suppose there are \( nk \) cells into which \( t \) indistinguishable balls are placed at random. If it is assumed that all of the \( \binom{nk+t-1}{t} \) distinguishable arrangements have equal probabilities, then \( \hat{p}(x; t) \) is the probability that a group of \( k \) prescribed cells contain a total of exactly \( x \) balls.

The Logarithmic Distribution. For this distribution,

\[ a(x) = 1/(x+1); \quad f(\theta) = -\log(1-\theta)/\theta; \quad c = 1; \quad C(t, n) = \sum_{i=1}^{n} 1/(x_i+1), \quad (4.29) \]

where the summation is over the set \( S_n(t) \). No closed expression for \( C(t, n) \) is known by the author and, hence, \( \hat{p}(x, t) \) does not have an obvious probabilistic interpretation.

4.4 Absolutely Continuous Distribution

In this section some of the well-known families of Lebesgue-continuous distributions will be discussed. All of these families admit a complete and sufficient statistic so that only unique minimum variance unbiased estimators of the distribution functions will be considered. The estimators of these distributions functions have all previously been derived elsewhere and, therefore, only a few of these will be re-established to illustrate the methods of section two. The remainder of the results will
be stated and checked for absolute continuity to decide whether or not unbiased estimators of the corresponding density functions exist. Also, Rosenblatt's result is re-established as a simple corollary to Theorem 3.2.

**Example 4.3:** The Class of all Absolutely Continuous Distributions. In this example $F(x)$ is an arbitrary absolutely continuous distribution function with corresponding density $p(x)$. It is well known that the order statistic $T = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is complete and sufficient for this family and that the minimum variance unbiased estimator of $F$ is given uniquely by the empirical distribution function

$$
\hat{F}(x; T) = \frac{1}{n} \sum_{i=1}^{n} I_{A_i}(X_i)/n, \quad x \in \mathbb{R}, \quad (4.30)
$$

where $I_{A_i}$ is the indicator function of the set $A_i = [X_i \leq x]$. Clearly, for each fixed $T = t$, $\hat{F}(x; t)$ is not absolutely continuous and hence, by Theorem 3.2, there is no unbiased estimator of $p(x)$ for any sample size $n$.

**Example 4.4:** The Normal Distribution. This distribution is characterized by the vector

$$
\{(2\pi\theta_2)^{-\frac{1}{2}} \exp[-(x-\theta_1)^2/2\theta_2], R, R \times (0, \infty); (\Sigma X_i, \Sigma (X_i - \bar{X})^2)\},
$$

As well as (c), the complete family, two sub-families will be discussed; namely, (a)$\theta_2 = 1$, and (b)$\theta_1 = 0$. For simplicity, write $T/n = (U, W)$;
that is,

\[ U = \sum_{i=1}^{n} \frac{X_i}{n}, \]  

and

\[ W = \sum_{i=1}^{n} \frac{(X_i - U)^2}{n}. \]  

(a) \( \theta_1 \) unknown, \( \theta_2 = 1 \). Here, \( U \) is the complete and sufficient statistic. Moreover, \( V = X_1 - U \) has a normal distribution with \( (\theta_1, \theta_2) = (0, (n-1)/n) \), and the function \( x - u \) is monotonic increasing in \( x \) for fixed \( u \). Hence, method 3 applies and the required minimum variance unbiased estimator of \( F \) is given uniquely by

\[ \hat{F}(x; U) = \int_{-\infty}^{x-U} (2\pi(n-1)/n)^{-1/2} \exp[-n(s^2/2(n-1))]ds. \]  

It should be noted that, for \( n \geq 2 \), \( \hat{F} \) is absolutely continuous and, therefore, the minimum variance unbiased estimator of the density function exists for \( n \geq 2 \) and is given by

\[ \hat{p}(x; U) = \left[2\pi(n-1)/n\right]^{-1/2} \exp[-n(x-U)^2/2(n-1)], \quad x \in \mathbb{R}. \]  

When \( n = 1 \) \( \hat{F} \) is not absolutely continuous and is, in fact, the empirical distribution function. It will become apparent in several of the examples that the existence or non-existence of unbiased estimators
depends on sample size. The smallest sample size \( m \) for which an unbiased estimator exists will be called the **degree of estimability**. In example 4.3, \( m = \infty \) and in the present example, \( m = 2 \). The value \( m = 2 \) in this example could be given the following interpretation: one 'degree of freedom' for estimating the empirical distribution function and one 'degree of freedom' for estimating \( \theta_1 \). Actually, \( \theta_1 \) does have one degree of estimability. Finally, it should be observed that the unbiasedness of \( \hat{p}(x;U) \) in (4.35) can easily be verified directly by integration.

(b) \( \theta_1 = 0, \theta_2 \) unknown. Here, the complete and sufficient statistic is

\[
S = \sum_{i=1}^{n} x_i^2.
\]

(F.36)

Folks, et al, using method two, obtained the following estimator:

\[
\hat{F}(x;S) = \begin{cases} 
0, & x \leq -S_{\frac{1}{2}}^n \\
G_{n-1}\left[\frac{1}{(n-1)^{\frac{1}{2}}}x/(S-x^2)^{\frac{1}{2}}\right], & -S_{\frac{1}{2}}^n < x \leq S_{\frac{1}{2}}^n \\
1, & x > S_{\frac{1}{2}}^n 
\end{cases}
\]

(F.37)

where \( G_{n-1}(u) \) is Student's t-distribution function with \( (n-1) \) degrees of freedom. Clearly, \( \hat{F} \) is absolutely continuous for \( n \geq 2 \) and, hence, the required minimum variance unbiased estimator of the density function is given uniquely by
\[
\hat{p}(x; S) = \begin{cases} 
\frac{S^{-\frac{1}{2}}[1-x^2/S]^{(n-1)/2-1}/B(\frac{1}{2}, (n-1)/2)}{\sqrt{2\pi}} & -S^{\frac{1}{2}} < x < S^{\frac{1}{2}}, \\
0, & \text{elsewhere},
\end{cases}
\]  
(4.38)

where \( B(\alpha, \beta) \) is the complete Beta function.

It is interesting to compare the estimators in cases (a) and (b). The estimator in (4.35) is strictly positive for all values of \( x \) and any fixed \( T = t \), whereas, the estimator in (4.38) is zero outside the interval \([-t^{\frac{1}{2}}, t^{\frac{1}{2}}]\). In both cases, of course, the original density functions are positive everywhere for all values of the unknown parameters. In case (b), \( m = 2 \) and in case (a), \( m = 2 \). The value, \( m = 2 \), in case (b) can be decomposed into one 'degree of freedom' for \( \theta_2 \) and one degree for the empirical distribution function.

(c) \( \theta_1 \) unknown, \( \theta_2 \) unknown. In this case the complete and sufficient statistic is \( T = (nU, nW) \). Using method 3 of section two and the fact that \( V = (X_1 - U)/W^{\frac{1}{2}} \) has distribution

\[
\begin{cases} 
\frac{[1-x^2/(n-1)]^{(n-4)/2}/[B((n-3)/2, (n-2)/2)]}{(n-2)^{\frac{1}{2}}B(\frac{3}{2}, (n-2)/2)} & |x| < (n-1)^{\frac{1}{2}}, \\
0, & \text{elsewhere}.
\end{cases}
\]  
(4.39)

Sathe and Varde obtained, in a slightly different form, the estimator
\[ F(x,T) = \begin{cases} 
0, & x < U - \left[ \frac{(n-1)W}{2} \right], \\
G_{n-2}\left( \frac{x-U}{\left[ \frac{(n-1)W}{2} \right]} \right), & |x-U| \leq \left[ \frac{(n-1)W}{2} \right], (4.40) \\
1, & x > U + \left[ \frac{(n-1)W}{2} \right]. 
\]

Clearly, \( \hat{F} \) is absolutely continuous for \( n \geq 3 \) and the unique minimum variance unbiased estimator \( \hat{p} \) of the density function is obtained by replacing \( x \) with \( (x-U)/W^\frac{1}{2} \) in (4.39). It should be noted that, once again, the estimator \( \hat{p} \) is zero outside a finite interval for any fixed value of \( T \). The degree of estimability \( m = 3 \); two 'degrees of freedom' for \( (\theta_1, \theta_2) \) and one for the empirical distribution function. It appears that the degree of estimability equals the degree of estimability of the usual indexing parameter plus one degree for the empirical distribution function.

**Example 4.5: The Truncation Distributions.** This family of distributions includes two types; and the characterizing vector for each type is given by

- **Type I:** \( \left\{ k_1(\theta)h_1(x), \ (\theta < x < b), \ (a,b); \ X(1) \right\} \) (4.41)

- **Type II:** \( \left\{ k_2(\theta)h_2(x), \ (a < x < \theta), \ (a,b); \ X(n) \right\} \), (4.42)

where the interval \((a,b)\) is either finite, semi-infinite or infinite and, \( k_1 \) and \( h_1 \) and \( k_2 \) and \( h_2 \) have the obvious relations.
\[
\frac{1}{k_1(\theta)} = \int_{\theta}^{b} h_1(x)dx, \quad \text{for all } \theta \in (a, b) \tag{4.43}
\]

\[
\frac{1}{k_2(\theta)} = \int_{a}^{\theta} h_2(x)dx, \quad \text{for all } \theta \in (a, b) \tag{4.44}
\]

The statistics \( X_{(1)} \) and \( X_{(n)} \) are the smallest and largest order statistics, respectively. Tate used essentially method 3 to obtain the following distribution function estimators for the Type I and Type II families defined above:

\[
\hat{F}(x; X_{(1)}) = \begin{cases} 
0, & a < x < X_{(1)} < b, \\
\frac{1}{n} \left( 1 - \frac{1}{n} \right) \frac{h_1(u)du}{X(1)} + \frac{\int_{X(1)}^{x} h_1(u)du}{X(1)}, & a < X_{(1)} < x < b.
\end{cases} \tag{4.45}
\]

\[
\hat{F}(x; X_{(n)}) = \begin{cases} 
\frac{1}{n} \left( 1 - \frac{1}{n} \right) \frac{h_2(u)du}{X(n)} + \frac{\int_{a}^{x} h_2(u)du}{x}, & a < x < X_{(n)} < b, \\
1, & a < X_{(n)} < x < b.
\end{cases} \tag{4.46}
\]

Both of the above estimators are mixed distributions with a jump of \( 1/n \) at \( x = X_{(1)} \) in the first case and at \( x = X_{(n)} \) in the second case.
Hence by Theorem 3.2 no minimum variance unbiased estimator of the density function exists for either type. Two examples of these distributions will now be given to illustrate the above estimators.

The Pareto Distribution. This distribution has a Type I density with \( k_1(\theta) = \theta, \quad h_1(x) = 1/x^2, \quad a = 0 \) and \( b = +\infty \). The estimator in (4.45) takes the form,

\[
\hat{F}(x; X_{(1)}) = \begin{cases} 
0, & 0 < x < X_{(1)} < \infty, \\
\frac{1}{n} + \left(1 - \frac{1}{n}\right)(1 - X_{(1)}/x), & 0 < X_{(1)} < x < \infty.
\end{cases}
\]  

(4.47)

Note that if \( n = 1 \), \( \hat{F} \) in (4.47) reduces to the empirical distribution function.

The Uniform Distribution. This is a Type II distribution with \( k_2(\theta) = 1/\theta, \quad h_2(x) = 1, \quad a = 0 \) and \( b = +\infty \). The estimator in (4.46) takes the form,

\[
\hat{F}(x, X_{(n)}) = \begin{cases} 
(n-1)x/nX_{(n)}, & 0 < x < X_{(n)} < \infty, \\
1, & 0 < X_{(n)} < x < \infty.
\end{cases}
\]  

(4.48)

The following figure illustrates the estimator in (4.48).
Figure 4.1 The U.M.V.U.E. of \((x/\theta)I(0,\theta)\).

This example serves to illustrate a point that emphasizes the subtle difference between the work in this thesis and the usual point estimation problem. It is simple to show that if \( n \geq 2 \), \((n-1)/nX(n)\) is the unique minimum variance unbiased estimator of \(1/\theta\) and yet there exists no such estimator for the density

\[
p_\theta(x) = \begin{cases} 
1/\theta, & 0 < x < \theta, \\
0, & \text{elsewhere.} 
\end{cases}
\]  

(4.49)

The point is that in the first instance all that is required is an unbiased estimator of the parametric function \(1/\theta\); and in the second case an unbiased estimator is required of the function specified in (4.49). Thus, the notion of a uniform (in \(x\)) unbiased estimator does create a different estimation problem.
Example 4.6: The Two-Parameter Exponential Distribution. The characterizing vector for this family of distributions is

$$\{\theta_2 \exp[-\theta_2(x-\theta_1)], \theta_1, \theta_2, \theta_2(0, \infty); X(1), \sum_{i=1}^{n} X_i\}$$

Tate obtained the following distribution function estimator using method 1:

$$\hat{F}(x; T) = \begin{cases} 
0, & x < X(1), \\
1 - \left(1 - \frac{1}{n}\right) \left(1 - \frac{x - X(1)}{Y}\right)^{n-2}, & X(1) < x < X(1) + Y, \\
1, & x > X(1) + Y,
\end{cases}$$

(4.51)

where $Y = \sum_{i=1}^{n} X_i - nX(1)$.

Clearly, $\hat{F}$ is not absolutely continuous but is a mixed distribution with mass $1/n$ at $x = X(1)$. Hence, no unbiased estimator of the density function exists for this family. However, if $\theta_1$ is known to be equal to zero (say) then the estimator for this sub-family is obtained by putting $X(1) = 0$, and changing the exponent to $n-1$ and $(1 - 1/n)$ to 1 in (4.51). This estimator is absolutely continuous and, therefore, the required density estimator is given by

$$\hat{f}(x; T) = \begin{cases} 
(n-1)(1-x/T)^{n-2}/T, & 0 < x < T, \\
0, & \text{elsewhere},
\end{cases}$$

(4.52)
where $T = \sum_{i=1}^{n} X_i$. The degree of estimability in this case is two; one degree for $\theta_2$ and one degree for the empirical distribution function.
5. SUMMARY

The work in this thesis is concerned with the problem of unbiased estimation of probability density functions. The general model considered is that of a \( \mu \)-dominated family \( P \) of probability measures \( P \) on a \( \sigma \)-finite measure space \( (\mathcal{I}, \mathcal{G}, \mu) \). Given \( n \) observations \( X^{(n)} = (X_1, X_2, \ldots, X_n) \) on a random variable \( X \) (with values in \( \mathcal{I} \)) the object is to find estimators \( \hat{p} \) of \( p \) - the probability density function corresponding to \( P \) - such that

\[
E_p[\hat{p}(x)] = p(x), \quad \text{for all } P \in P \text{ and all } x \in \mathcal{I}.
\]

The uniform in \( x \) property of such estimators \( \hat{p} \) is the feature that distinguishes this problem from the usual problem of point estimation.

Here, an estimator of a function is sought.

In chapter three definitions of unbiasedness for both density function and probability measure estimators are given. After giving a suitable definition of \( \mu \)-continuity for an estimator \( \hat{P} \) of \( P \) two theorems are established that give necessary and sufficient conditions that guarantee the existence of unbiased estimators \( \hat{p} \) of \( p \). It is shown in Theorem 3.1 that such a \( \hat{p} \) exists if and only if there exists an unbiased estimator \( \hat{P} \) of \( P \) that is absolutely continuous with respect to the original dominating measure \( \mu \). Moreover, if such a \( \hat{P} \) exists then it is shown that the Radon-Nikodym derivative \( \frac{d\hat{P}}{d\mu} \) is an unbiased estimator of \( p \). Theorem 3.2 gives a stronger result if there exists a complete and sufficient statistic \( T \) for \( P \); a unique minimum variance unbiased estimator
of \( p \) exists if and only if \( P_T^{\hat{c}} \) is absolutely continuous with respect to \( \mu \). Moreover, if such a \( \hat{p} \) exists, it is given by \( \frac{dP_T^{\hat{c}}}{d\mu} \) which is also a probability density function for almost all \( (P_T) \) t. (Here, \( P_T^{\hat{c}} \) is the conditional probability measure \( P_T \) on \( G(n) \), given \( T \), restricted to the sub-\( \sigma \)-algebra \( \mathcal{E} \) of \( G(n) \) whose sets are cylinders with bases in a fixed but arbitrary coordinate space of \( \mathcal{X}(n) \).) Hence, for any particular family of distributions if it is possible to derive an estimator of \( P \) which is absolutely continuous with respect to \( \mu \) then an unbiased estimator of \( p \) exists and is obtained by differentiating the probability measure estimator with respect to \( \mu \).

In chapter four the general theory is specialized to well-known families of distributions on the real line that are either absolutely continuous with respect to ordinary Lebesgue measure or absolutely continuous with respect to counting measure on a sequence of points. In the Lebesgue-continuous case it is shown, for example, that for the normal distribution with unknown mean and variance that if the sample size exceeds or equals three then a minimum variance unbiased estimator of the normal density function exists. This estimator has some interesting properties when compared with the corresponding maximum likelihood estimator obtained by substituting the maximum likelihood estimators of the parameters for the parameters in the functional form of the normal density. For example, for any fixed value of the sufficient statistic the estimator is zero outside a finite interval (cf. the maximum likelihood estimator). It is also interesting to note that the normal density function is estimated by a Student's t-distribution,
whereas, the likelihood principle demands that the estimator should also be a normal distribution. Examples of minimum variance unbiased estimators of distribution functions which are not absolutely continuous are the family of all Lebesgue-continuous distributions on the real line and the uniform distribution on \((0, \theta)\). Estimators for the binomial distribution and a general power series type distribution are obtained to give examples of the procedures for discrete distributions.

Interesting sample size considerations are observed in many of the examples illustrating the absolutely continuous case. For example, in the normal distribution it has already been noted that at least three observations are required to obtain an unbiased estimator. The maximum likelihood procedure, on the other hand, requires a minimum of only two observations. It is apparent in several of the examples that the degree of estimability can be decomposed into one degree for the empirical distribution function and the degree for the index parameter in the functional form of the density. The degree for the parameter is the smallest number of observations that are required to have an unbiased estimator of that parameter.

As in any area of research some problems still remain to be solved. Some of these are:

1. Find conditions that guarantee the existence of Bayes, minimax and invariant estimators of density functions.

2. Find equivalent conditions on \( \rho \) that guarantee \( P_{\xi}^T \) to be absolutely continuous with respect to \( \mu \).

3. Compare, for example, maximum likelihood estimators with
unbiased estimators of density functions, when the latter exist.

4. The likelihood ratio is often used to partition the sample space in the construction of decision rules. In discrimination procedures, for example, it is sometimes necessary to estimate the likelihood ratio and hence the resultant partition of the sample space. A study of the properties of such procedures when unbiased estimators of density functions are used to estimate these partitions would be interesting.
6. LIST OF REFERENCES


