1965

Exponential families in testing statistical hypotheses

Anton Kraft

The University of Montana

Follow this and additional works at: https://scholarworks.umt.edu/etd

Let us know how access to this document benefits you.

Recommended Citation

Kraft, Anton, "Exponential families in testing statistical hypotheses" (1965). Graduate Student Theses, Dissertations, & Professional Papers. 8323.
https://scholarworks.umt.edu/etd/8323

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
EXPONENTIAL FAMILIES IN
TESTING STATISTICAL HYPOTHESES

By
ANTON KRAFT
B.A. Montana State University, 1963
Presented in partial fulfillment of the requirements
for the degree of
Master of Arts

UNIVERSITY OF MONTANA

1965

Approved by:

Howard E. Reiland
Chairman, Board of Examiners

Dean, Graduate School

JUL 20 1965
Date
ACKNOWLEDGEMENTS

I would like to express my appreciation to Professor Howard E. Reinhardt for his guidance throughout the preparation of this thesis. Also, I would like to thank Professors William M. Myers and Robert Banaugh for reading of the thesis.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>INTRODUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 1. PRELIMINARIES</td>
</tr>
<tr>
<td>1. Probability and measure................. 1</td>
</tr>
<tr>
<td>2. Conditional probability and expectation 2</td>
</tr>
<tr>
<td>3. Discrete and continuous probability spaces.......................... 3</td>
</tr>
<tr>
<td>4. Parameter spaces and sufficiency........ 8</td>
</tr>
<tr>
<td>CHAPTER 2. UNIFORMLY MOST POWERFUL TESTS</td>
</tr>
<tr>
<td>1. Exponential families..................... 10</td>
</tr>
<tr>
<td>2. Stating the problem..................... 19</td>
</tr>
<tr>
<td>3. Existence of tests...................... 22</td>
</tr>
<tr>
<td>CHAPTER 3. UNIFORMLY MOST POWERFUL UNBIASED TESTS</td>
</tr>
<tr>
<td>1. Unbiasedness, similarity and completeness............................. 31</td>
</tr>
<tr>
<td>2. One-parameter exponential families...... 35</td>
</tr>
<tr>
<td>3. Multiparameter exponential families.... 39</td>
</tr>
<tr>
<td>4. Applications............................. 42</td>
</tr>
<tr>
<td>REFERENCES</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
INTRODUCTION

In this paper the theory of testing statistical hypotheses is restricted to the class of distributions called the exponential family of distributions. The material closely follows that given by Lehmann in [3] with an attempt to exclude many general measure theoretic concepts. The proofs of all theorems regarding distributions are given in either the case of discrete distributions or the case of continuous distributions and the case given is specified. The necessary fundamentals of probability and measure theory are outlined in chapter 1. The exponential family of distributions and concepts of hypothesis testing are given in chapter 2. The theorems in this chapter develop the existence of uniformly most powerful tests of hypotheses. Chapter 3 introduces the idea of unbiasedness and develops the existence of uniformly most powerful unbiased tests within the framework of exponential families of distributions.
1. PROBABILITY AND MEASURE

Consider an outcome space $\mathcal{X}$ and a $\sigma$-algebra $\mathcal{A}$ of subsets of $\mathcal{X}$. We will call $(\mathcal{X}, \mathcal{A})$ a measurable space and the elements of $\mathcal{A}$ the measurable sets. A countably additive nonnegative set function $\mu$ defined over $\mathcal{A}$ is called a measure. If $\mu(\mathcal{X}) = 1$ it is a probability measure. The probabilities over $(\mathcal{X}, \mathcal{A})$ refer to the points $x \in \mathcal{X}$, which are the possible outcomes for a random experiment. Let $X$ denote these observations and let the probability of $X$ falling in a set $A$ be $P\{X \in A\} = P(A)$. We will call $X$ a random variable over the space $(\mathcal{X}, \mathcal{A})$ and refer to the probability measure $P$ or $P^X$ as the probability distribution of $X$.

For a function $T : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{T}, \mathcal{B})$ the probability of the event $T \in B$ is defined by $P$ if and only if $B \in \mathcal{B}$ and $T^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$. Then $T$ is said to be a measurable function or a statistic. A function $T$ of $\mathcal{X}$ taking on values in some space $\mathcal{T}$ generates in $\mathcal{T}$ the $\sigma$-algebra $\mathcal{B}'$ of sets $B$ whose inverse image $A = T^{-1}(B) = \{x : x \in \mathcal{X}, T(x) \in B\}$ is in $\mathcal{A}$. $T = T(x)$ is a random variable over the space $(\mathcal{T}, \mathcal{B}')$ and since $X \in T^{-1}(B)$ if and only if $T(X) \in B$, the probability distribution of $T$ over $(\mathcal{T}, \mathcal{B}')$ is given by $P^T(B) = P\{T(x) \in B\}$.
\[ P \{ X \in T^{-1}(B) \} = P^X(T^{-1}(B)) \]. A statistic induces in the original sample space, \((X, \mathcal{A})\), the subalgebra \( \mathcal{A}_o = T^{-1}(\mathcal{B}) = \{ T^{-1}(B) : B \in \mathcal{B} \} \).

The following theorem is proved in [2].

**Theorem 1:** Let the statistic \( T \) from \((X, \mathcal{A})\) into \((T, \mathcal{B})\) induce the subfield \( \mathcal{A}_o \). Then a real-valued \( \mathcal{A} \)-measurable function \( f \) is \( \mathcal{A}_o \)-measurable if and only if there exists a real-valued \( \mathcal{B} \)-measurable function \( g \) such that \( f(x) = g[T(x)] \) for all \( x \).

2. **Conditional Expectation and Probability**

In this section we will give a general outline of conditional expectation and probability and omit the proofs of the theorems evident from the discussion.

Let \( P \) be a probability measure over \((X, \mathcal{A})\), \( T \) a statistic with range space \((T, \mathcal{B})\), and \( \mathcal{A}_o \) the subalgebra it induces. Consider a nonnegative function \( f \) which is \( \mathcal{A} \)-measurable and \( P \)-integrable. Then \( \int_A f \text{d}P \) is defined for all \( A \in \mathcal{A} \) and hence for all \( A_o \in \mathcal{A}_o \). It follows from the Radon-Nikodym theorem that there exists a function \( f_o \), which is \( \mathcal{A}_o \)-measurable and \( P \)-integrable, and such that

\[ (1) \quad \int_{A_o} f \text{d}P = \int_{A_o} f_o \text{d}P \quad \text{for all } A_o \in \mathcal{A}_o, \]

and that \( f_o \) is unique up to a set of \( P \)-measure zero. Now \( f_o(x) \) is taken as the general definition of the conditional expectation \( E[f(X)|T(x)] \). We see by Theorem 1, that \( f_o \)
depends on $x$ only through $T(x)$ so that $f_0(x) = g[T(x)]$.
We can write $E[f(X)|T = t] = g(t)$ so that $E[f(X)|t]$ is a $\mathcal{B}$-measurable function.

In the case of discrete and continuous distributions in the next section we will see that the function $g$ can be defined directly in terms of $f$ through the relationship between integrals of the functions $f$ and $g$ of Theorem 1, where $P^T$ is the probability measure on $(\mathcal{T}, \mathcal{B})$:

$$\int_{T^{-1}(B)} f(x) \, dP^X(x) = \int_B g(t) \, dP^T(t) \quad \text{for all } B \in \mathcal{B}. 
$$

Since $P\{X \in A\} = E[I_A(X)] = 1 \cdot P(X \in A) + 0 \cdot P(X \notin A)$, where $I_A$ denotes the indicator of the set $A$, it seems natural to define the conditional probability of $A$ given $T = t$ by

$$P(A|t) = E[I_A(X)|t]. 
$$

This definition has the pleasing property that it agrees with the notion of conditional probability used in discrete spaces. In view of (2) we have

$$\int_{T^{-1}(B)} I_A(x) \, dP^X(x) = \int_B E[I_A(x)|t] \, dP^T(t),
$$

and hence (3) may be written as

$$P^X(A \cap T^{-1}(B)) = \int_{A \cap T^{-1}(B)} dP^X(x) = \int_{B} P(A|t) \, dP^T(t) \quad \text{for all } B \in \mathcal{B}. 
$$

3. DISCRETE AND CONTINUOUS PROBABILITY SPACES

**Definition 1:** A space $(\mathcal{X}, \mathcal{A})$ is called **discrete** if $\mathcal{X}$ is countable and $\mathcal{A}$ contains all single outcome sets
If we define a probability measure on the discrete space, it can be shown that \( P(\mathcal{X}) = 1 \) and \( P(B) = \sum_{x \in B} P(x) \).

**DEFINITION 2**: A space \((\mathcal{X}, \mathcal{A})\) is called continuous if \(\mathcal{X}\) is an n-dimensional Euclidean space and \(\mathcal{A}\) the \(\sigma\)-algebra generated by the open and closed subsets of \(\mathcal{X}\).

A probability measure on the continuous space is given by \( P(A) = \int_A f(x) \, dx \) for every \( A \in \mathcal{A} \) where \( f(x) \) is a nonnegative function such that \( \int_{\mathcal{X}} f(x) \, dx = 1 \).

We will let \((\mathcal{X}, \mathcal{A}, P^X)\) denote a probability space \((\mathcal{X}, \mathcal{A})\), where \( P^X \) is the probability measure defined on it. If \( T : (\mathcal{X}, \mathcal{A}, P^X) \to (\mathcal{T}, \mathcal{B}, P^T) \) is a statistic, then the joint probability measure \( P^{X,T} \) defined on the measurable space \((\mathcal{X} \times \mathcal{T}, \mathcal{A} \times \mathcal{B})\) is given by \( P^{X,T}(A \times B) = P^X(A \cap T^{-1}(B)) \) for all \( A \in \mathcal{A} \) and \( B \in \mathcal{B} \).

The following is a discussion of conditional expectation and probability in the particular cases of discrete and continuous spaces. Details and proofs can be found in [2].

**DEFINITION 3**: (1) If \((\mathcal{X}, \mathcal{A})\) is discrete and \( f \) is a real-valued function defined on \(\mathcal{X}\), then the expected value of \( f(X) \) is defined as \( \mathbb{E}[f(X)] = \sum_\mathcal{X} f(x) P^X(x) \) provided this series converges absolutely.

(2) If \((\mathcal{X}, \mathcal{A})\) is continuous and \( f \) is a real-valued measurable function defined on \(\mathcal{X}\), then the
expected value of \( f(X) \) is defined as
\[
E[f(X)] = \int \! f(x)p^X(x) \, dx,
\]
where \( p^X(x) \) is the probability density.

Suppose \((\mathcal{X}, \mathcal{A})\) is discrete and \( T : (\mathcal{X}, \mathcal{A}, p^X) \rightarrow (\mathcal{T}, \mathcal{B}, p^T) \) is a statistic. Denote the set of all points in \( \mathcal{X} \) with positive probability by \( \{x_n\} \) and let \( T \{x_n\} = \{t_m\} \). Then for \( t \in \{t_m\} \) define the conditional probability distribution of \( X \) given \( T = t \) by
\[
p^X | t(A) = \frac{p^X(A \cap T^{-1}(t))}{p^T(t)}.
\]

\( p^X | t \) is a probability measure over \((\mathcal{X}, \mathcal{A})\) for each \( t \in \{t_m\} \). Now suppose \((\mathcal{X}, \mathcal{A})\) is continuous and \( T : (\mathcal{X}, \mathcal{A}, p^X) \rightarrow (\mathcal{T}, \mathcal{B}, p^T) \) is a statistic. Then the conditional probability distribution of \( X \) given \( T = t \) has density \( p^X | t \) with respect to Lebesgue measure and is given by
\[
p^X | t(x) = \frac{p^X, T(x, t)}{\int_p^X, T(x, t) \, dx} = \frac{p^X, T(x, t)}{p^T(t)},
\]
where \( p^T(t) \) is the density of \( p^T \).

**DEFINITION 4:** (i) If \((\mathcal{X}, \mathcal{A})\) is discrete and \( f \) is a real-valued function defined on \( \mathcal{X} \), then the conditional expectation of \( f(X) \) given \( T(x) = t \) is defined as
\[
E[f(X)|t] = \sum_x f(x)p^X | t(x),
\]
provided the series converges absolutely.

(ii) If \((\mathcal{X}, \mathcal{A})\) is continuous and \( f \) is a real-valued measurable function defined on \( \mathcal{X} \), then the conditional expectation of \( f(X) \) given \( T(x) = t \) is defined as
\[ E[f(X) \mid t] = \int_{\mathcal{X}} f(x) p^X_t(x) \, dx, \text{ provided the integral exists.} \]

If \( E[f(X)] \) exists, then \( E[f(X) \mid t] \) is \( \mathcal{B} \) - measurable and defined a.e. \( P^T \) on \( \mathcal{T} \) (a.e. \( P^T \) means it is defined everywhere except on a set of \( P^T \)-measure zero).

In the case of discrete distributions we again let \( \{t_m\} \) be the image of the set of all points in \( \mathcal{X} \) with positive probability and let \( g(t) = E[f(X) \mid t] \). Then the integral representation of conditional expectation in (2) is given by \( \sum_{x \in \mathcal{T}^{-1}(B)} f(x) p^X_t(x) = \sum_{t \in \mathcal{T}} g(t) p^T_t(t) \) for all \( B \subseteq \mathcal{B} \).

Also, in terms of continuous distributions we have
\[
\int_{\mathcal{T}^{-1}(B)} f(x) p^X_t(x) \, dx = \int_B g(t) p^T_t(t) \, dt \text{ for all } B \subseteq \mathcal{B}.
\]

We will prove the following lemma to be used in section 1, chapter 2.

**Lemma 1:** Let \((\mathcal{T}, \mathcal{B})\) and \((\mathcal{Y}, \mathcal{C})\) be Euclidean spaces and let \( P^{T,Y}_{\mathcal{O}} \) be a distribution over the product space \((\mathcal{X}, \mathcal{A}) = (\mathcal{T} \times \mathcal{Y}, \mathcal{B} \times \mathcal{C})\). Suppose that another distribution \( P^{T,Y}_{\mathcal{I}} \) over \((\mathcal{X}, \mathcal{A})\) is such that \( dP^{T,Y}_{\mathcal{I}}(t,y) = a(y)b(t)dP^{T,Y}_{\mathcal{O}}(t,y) \) with \( a(y) > 0 \) for all \( y \). Then under \( P_{\mathcal{I}} \) the marginal distribution of \( T \) and a version of the conditional distribution of \( Y \) given \( t \) are given by
\[
dP^{T}_{\mathcal{I}}(t) = b(t)[\int a(y)dP^{Y\mid t}_{\mathcal{O}}(y)]dP^{T}_{\mathcal{O}}(t)
\]
and
\[
dP^{Y\mid t}_{\mathcal{I}}(y) = \frac{a(y)dP^{Y\mid t}_{\mathcal{O}}(y)}{\int a(y)dP^{Y\mid t}_{\mathcal{O}}(y)}.
\]
PROOF: Let us consider the particular case of continuous distributions and note that the case of discrete distributions is treated similarly. Under both \( P_0 \) and \( P_1 \) the conditional probability distribution of \( Y \) given \( T = t \) has density with respect to Lebesgue measure given by

\[
p^Y|_T(y) = \frac{p^{T,Y}(t,y)}{p^T(t)} \quad \text{a.e.m (where } m \text{ is Lebesgue measure).}
\]

Now under \( P_1 \) the density for the marginal distribution of \( Y \) given \( T = t \) is given by

\[
p^1(t) = \int a(y)p^Y|_t(y)dy = b(t)\int a(y)p^T,Y(t,y)dy
\]

\[
= b(t)[\int a(y)p^Y|_t(y)dy]p^T_0(t).
\]

Using this result we see that the conditional distribution of \( Y \) given \( T = t \) is given by

\[
p^Y|_1(t,y) = \frac{p^T,Y(t,y)}{p^1(t)} = \frac{a(y)b(t)p^T,Y(t,y)}{b(t)[\int a(y)p^Y|_t(y)dy]p^T_0(t)}
\]

\[
= \frac{a(y)b(t)p^Y|_t(y)p^T_0(t)}{b(t)[\int a(y)p^Y|_t(y)dy]p^T_0(t)} = \frac{a(y)p^Y|_t(y)}{\int a(y)p^Y|_t(y)dy}
\]

We should note that the use of an expression like \( dP^X \) will be taken to mean that \( P^X \) has a density \( p^X \) with respect to counting measure in the case of discrete distributions or with respect to Lebesgue measure in the case of continuous distributions and that \( dP^X \) is the density with respect to counting measure in the discrete case and in the
continuous case \( d\mu_x = \mu^x \, dx \) where \( \mu^x \) is the density.

4. PARAMETER SPACES AND SUFFICIENCY

Consider a random variable \( X \) with probability distribution \( P_\theta \). The parameter \( \theta \) specifies the distribution of \( X \) and lies in a set \( \Omega \), called the parameter space. In general the distribution of \( X \) is unknown or partly unknown and the problem is to use the observational material to obtain information about the distribution of \( X \) or the parameter \( \theta \) with which it is labeled. We state a statistical hypothesis pertaining to the parameter of a distribution and make some decision on the basis of a particular test of the hypothesis. The methods used in the solution of a statistical problem depend in part on the class of probability distributions to which the distribution of \( X \) is assumed to belong. This class of distributions is written as \( \mathcal{P} = \{ P_\theta, \theta \in \Omega \} \). Some common classes are the binomial, Poisson and normal distributions.

It is often desirable to discard that part of the observational material which contains no information regarding the unknown distribution \( P_\theta \). This leads to the idea of a sufficient statistic the theory of which is developed in [2].

**DEFINITION 5**: A statistic \( T \) is said to be sufficient for the family \( \mathcal{P} = \{ P_\theta, \theta \in \Omega \} \) defined over a common sample space \( (\mathcal{X}, \mathcal{A}) \), if the conditional distribution of the random variable \( X \) given \( T = t \) is independent of \( \theta \).
The following theorem is a factorization criterion for sufficiency. The proof of this theorem is given in [2].

**THEOREM 2:** If the distributions $P_\theta$ of $\mathcal{P}$ have probability densities $p_\theta = dP_\theta/dy$ with respect to a $\sigma$-finite measure $\mu$, then the statistic $T$ with range space $(T, \mathcal{B})$ is sufficient for $\mathcal{P}$ if and only if there exist nonnegative $\mathcal{B}$-measurable functions $g_\theta$ on $T$ and a non-negative $\mathcal{A}$-measurable function $h$ on $\mathcal{X}$ such that $p_\theta(x) = g_\theta[T(x)]h(x)$.
1. EXPONENTIAL FAMILIES

**DEFINITION 6:** The exponential family of distributions is defined by probability densities of the form

\[ p_\theta(x) = C(\theta) \exp \left[ \sum_{j=1}^{k} \theta_j T_j(x) \right] h(x) \]

with respect to a σ-finite measure \( \mu \) over a Euclidean sample space \( (\mathcal{X}, \mathcal{A}) \).

Particular cases are the distributions of a sample \( X = (X_1, \ldots, X_n) \) from a binomial, Poisson, or normal distribution. For example, in the binomial case the density with respect to counting measure is \( p^n (1-p)^{n-x} \). This can be written as \( (1-p)^n \exp[x \log \frac{p}{1-p}] \). Letting \( \theta = \log \frac{p}{1-p} \) we can solve for \( p \) in order to obtain an expression for \( (1-p)^n \) in terms of \( \theta \). Hence \( \theta = \log \frac{p}{1-p} \Rightarrow p = \frac{e^\theta}{1+e^\theta} \Rightarrow (1-p)^n = \left( \frac{1}{1+e^{-\theta}} \right)^n \). So we can write

\[ \left( \frac{\bar{x}}{\theta} \right) p^n (1-p)^{n-x} = \left( \frac{1}{1+e^{-\theta}} \right)^n \exp(\theta x) \]

where in comparison with (5), \( k = 1, \theta = \log \frac{p}{1-p} \),

\[ C(\theta) = \left( \frac{1}{1+e^{-\theta}} \right)^n, T(x) = x \text{ and } h(x) = \left( \frac{\bar{x}}{\theta} \right). \]

**THEOREM 3:** If \( X_1, X_2, \ldots, X_n \) is a sample from a distribution with density (5), the joint distribution
of the X's constitute an exponential family.

**PROOF:** Consider the case of a discrete distribution. 

\[ P(X_1 = x_1, \ldots, X_n = x_n) = p_\theta(x_1) \cdots p_\theta(x_n) \]

\[ = C_1(\theta) \exp \left[ \sum_{j=1}^{k} \theta_j T_j(x) \right] h(x_1) \cdots C_n(\theta) \exp \left[ \sum_{j=1}^{k} \theta_j T_j(x) \right] h(x_n) \]

\[ = C^*(\theta) \exp \left[ \sum_{j=1}^{k} \theta_j \sum_{i=1}^{n} T_j(x_i) \right] h^*(x). \]

In this case, \( \sum_{i=1}^{n} T_j(x_i), j = 1, \ldots, k \) are sufficient statistics for the exponential family. Thus there exists a k-dimensional sufficient statistic for \((X_1, \ldots, X_n)\) regardless of the sample size.

We shall often write an exponential family in the form 

\[ dP_\theta(x) = p_\theta(x) d\mu(x), \]

absorbing \( h(x) \) into \( \mu \) where 

\[ d\mu = h \, dx \]

and so that

\[ (6) \quad \frac{p_\theta(x)}{C(\theta)} = \exp \left[ \sum_{j=1}^{k} \theta_j T_j(x) \right]. \]

**DEFINITION 7:** The set \( \Theta \) of parameter points \( \theta = (\theta_1, \ldots, \theta_k) \) for which (6) is a probability density is called the natural parameter space of the exponential family (6).

Following are properties of exponential families required for the discussion in the remainder of the paper.

**LEMMA 2:** The natural parameter space of an exponential family is convex.

**PROOF:** Let \( \theta = (\theta_1, \ldots, \theta_k) \) and \( \theta' = (\theta_1', \ldots, \theta_k') \) be two parameter points in \( \Theta \) for which the integral of (6) is finite. For any \( 0 < \alpha < 1 \) we wish to show that
\[ \bar{\theta} = \alpha \theta + (1-\alpha)\theta' \] is a parameter point in \( \Omega \), that is the integral of (6) for \( \bar{\theta} \) is finite where the integrand is nonnegative. Consider the case of a continuous distribution where \( \mu \) is Lebesgue measure.

Then we have

\[
\int \exp\left[ \sum_{j=1}^{k} \bar{\theta}_j T_j(x) \right] dx = \int \exp\left[ \sum_{j=1}^{k} \left[ \alpha \theta_j + (1-\alpha)\theta_j' \right] T_j(x) \right] dx \\
= \int \exp\left[ \alpha \sum_{j=1}^{k} \theta_j T_j(x) + (1-\alpha) \sum_{j=1}^{k} \theta_j' T_j(x) \right] dx \\
= \int \exp\left[ \alpha \sum_{j=1}^{k} \theta_j T_j(x) \right] \exp\left[ (1-\alpha) \sum_{j=1}^{k} \theta_j' T_j(x) \right] dx \\
\leq \left[ \int \exp\left[ \sum_{j=1}^{k} \theta_j T_j(x) \right] dx \right]^\alpha \left[ \int \exp\left[ \sum_{j=1}^{k} \theta_j' T_j(x) \right] dx \right]^{(1-\alpha)} < \infty
\]

where the inequality follows from Hölder's Inequality,

\[
\int |f| \ dx \leq \left[ \int |f|^r \ dx \right]^{\frac{1}{r}} \left[ \int |g|^s \ dx \right]^{\frac{1}{s}} \text{ for } r > 1 \text{ and } \frac{1}{r} + \frac{1}{s} = 1,
\]

by letting \( f = \exp\left[ \sum_{j=1}^{k} \theta_j T_j(x) \right] \), \( g = \exp\left[ (1-\alpha) \sum_{j=1}^{k} \theta_j' T_j(x) \right] \), \( \alpha = \frac{1}{r} \) and \( 1 - \alpha = \frac{1}{s} \).

The assumption is made that \( \Omega \) is \( k \)-dimensional since if it lies in a linear space of dimension less than \( k \), (6) can be written in a form involving fewer than \( k \) components of \( T \). It also follows from the factorization theorem for sufficient statistics, Theorem 2, that \( T(x) = (T_1(x), \ldots, T_K(x)) \) is sufficient for \( \mathcal{P} = \{ P_{\theta}, \theta \in \Omega \} \).

**Lemma 3:** Let \( X \) be distributed according to the exponential family

\[
dP_{\theta, \alpha}^X (x) = C(\theta, \alpha) \exp\left[ \sum_{i=1}^{r} \theta_i U_i(x) + \sum_{j=1}^{s} \alpha_j T_j(x) \right] d\mu(x).
\]
Then there exist measures $\lambda_\theta$ and probability measures $v_t$ over $s$ and $r$-dimensional Euclidean space respectively such that

1. the distribution of $T = (T_1, \ldots, T_s)$ is an exponential family of the form
   \[ dP_{\theta, \nu}^T(t) = C(\theta, \nu) \exp\left( \sum_{j=1}^{s} \nu_j t_j \right) d\lambda_\theta(t) \]

2. the conditional distribution of $U = (U_1, \ldots, U_r)$ given $T = t$ is an exponential family of the form
   \[ dP_{\theta, \nu}^{U|T}(u) = C_{t}(\theta) \exp\left( \sum_{i=1}^{r} \theta_i u_i \right) dv_t(u) \]

and hence in particular is independent of $\nu$.

**Proof:** Express the distribution of $X$ in the form

\[ dP_{\theta, \nu}^{T,U}(t,u) = C(\theta, \nu) \exp\left( \sum_{i=1}^{r} \theta_i u_i + \sum_{j=1}^{s} \nu_j t_j \right) du(t,u) \]

Let $(\theta^0, \nu^0)$ be a point of the natural parameter space.

Then $dP_{\theta, \nu}^{T,U}(t,u) = \frac{C(\theta, \nu)}{C(\theta^0, \nu^0)} \exp\left( \sum_{i=1}^{r} (\theta_i - \theta^0_i) u_i + \sum_{j=1}^{s} (\nu_j - \nu^0_j) t_j \right) dP_{\theta^0, \nu^0}^{T,U}.$

Applying Lemma 1 we have two distributions over $(X, A)$ such that $dP_{\theta, \nu}^{T,U}(t,u) = a(u)b(t) dP_{\theta^0, \nu^0}^{T,U}(t,u)$ where

\[ a(u) = C(\theta, \nu) \exp\left( \sum_{i=1}^{r} (\theta_i - \theta^0_i) u_i \right) \]

\[ b(t) = \frac{1}{C(\theta^0, \nu^0)} \exp\left( \sum_{j=1}^{s} (\nu_j - \nu^0_j) t_j \right). \]

So by Lemma 1, under $P_{\theta, \nu}^{T,U}$ the marginal distribution of $T$ is $dP_{\theta, \nu}^{T}(t) = b(t) \left[ \int a(u) dP_{\theta^0, \nu^0}^{U|T}(u) \right] dP_{\theta^0, \nu^0}(t)$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[ \frac{1}{C(\theta^0, \omega^0)} \exp[\sum_{j=1}^{s}(\omega_j - \omega_j^0)t_j] \]

\[ \left\{ \frac{\int C(\theta, \omega) \exp[\sum_{j=1}^{s}(\theta_j - \omega_j^0)u_j]dP_{\omega}^U|T=t \right\} dP_T^{\theta^0, \omega^0}(u) \]

\[ = \frac{C(\theta, \omega) \exp[\sum_{j=1}^{s}\omega_j]}{C(\theta^0, \omega^0) \exp[\sum_{j=1}^{s}\omega_j^0]t_j} \left\{ \frac{\int \exp[\sum_{j=1}^{s}(\theta_j - \omega_j^0)u_j]dP_{\omega}^U|T=t \right\} dP_T^{\theta^0, \omega^0}(u) \]

\[ = C(\theta, \omega) \exp[\sum_{j=1}^{s}\omega_j] \left\{ \frac{\int \exp[\sum_{j=1}^{s}(\theta_j - \omega_j^0)u_j]dP_{\omega}^U|T=t \right\} dP_T^{\theta^0, \omega^0}(u) \]

Absorbing \( C(\theta^0, \omega^0) \) into \( C(\theta, \omega) \) and letting

\[ d\lambda_\theta(t) = \exp[\sum_{j=1}^{s}\omega_j] \left\{ \frac{\int \exp[\sum_{j=1}^{s}(\theta_j - \omega_j^0)u_j]dP_{\omega}^U|T=t \right\} dP_T^{\theta^0, \omega^0}(u) \]

we get

\[ dP_{\theta^0, \omega^0}(t) = C(\theta, \omega) \exp[\sum_{j=1}^{s}\omega_j]d\lambda_\theta(t). \]

By the second part of Lemma 1 we have the conditional distribution of \( U \) given \( T = t \) given by

\[ dP_{\theta, \omega}^U|T=t(u) = \frac{a(u)dP_{\omega}^U|T=t}{\int_{U} a(u')dP_{\omega}^U|T=t} \]

\[ = \frac{\int \exp[\sum_{j=1}^{s}(\theta_j - \omega_j^0)u_j]dP_{\omega}^U|T=t}{\int \exp[\sum_{j=1}^{s}(\theta_j - \omega_j^0)u_j]dP_{\omega}^U|T=t} \]

\[ = \frac{\exp[\sum_{j=1}^{s}\theta_ju_j]}{\exp[\sum_{j=1}^{s}\theta_ju_j]} \cdot \frac{\int \exp[\sum_{j=1}^{s}(\theta_j - \omega_j^0)u_j]dP_{\omega}^U|T=t}{\int \exp[\sum_{j=1}^{s}(\theta_j - \omega_j^0)u_j]dP_{\omega}^U|T=t} \]
Now the denominator is seen to be a function of $\theta$ and is in fact independent of $\gamma_j$. Denote this function by $C_t'(\theta)$ and let $C_t(\theta) = 1/C_t'(\theta)$ and $d_v(t) = \frac{dP^u_t}{\theta^0, \gamma^0(u)}$ so that $d_P^u_t(u) = C_t(\theta) \exp[\sum_{i=1}^r \theta_i u_i] d_v(t)$.

The following lemma is used in the proof of Theorem 4.

**Lemma 4:**

\[ \left| \frac{\exp(az) - 1}{z} \right| \leq \exp(\frac{\delta |a|}{\delta}) \text{ for } |z| \leq \delta. \]

**Proof:**

\[
\left| \frac{\exp(az) - 1}{z} \right| = \left| \sum_{n=0}^{\infty} \frac{(az)^n}{n!} - 1 \right| = \left| \sum_{n=0}^{\infty} \frac{a^n z^{n-1}}{n!} \right| \\
\leq \sum_{n=1}^{\infty} \left| a \right| \frac{|z|^{n-1}}{n!} \\
\leq \sum_{n=1}^{\infty} \frac{|a|^n |z|^{n-1}}{n!} \leq \frac{1}{\delta} + \sum_{n=1}^{\infty} \frac{|a|^n \delta^{n-1}}{n!} = \sum_{n=0}^{\infty} \frac{|a|^n \delta^n}{\delta^n n!} = \frac{\exp(\frac{\delta |a|}{\delta})}{\delta}. 
\]

**Theorem 4:** Let $\phi$ be any bounded measurable function on $(\mathcal{X}, \mathcal{A})$. Then

1. the integral 

\[ \int \phi(x) \exp[\sum_{j=1}^k \theta_j T_j(x)] d\mu(x) \]

considered as a function of the complex variables $\theta_j = \delta_j + i \eta_j$ ($j = 1, \ldots, k$) is an analytic function in each of these variables in the region $R$ of parameter
points for which \((\xi^0_1, \ldots, \xi^0_k)\) is an interior point of the natural parameter space \(\Omega\);

(ii) the derivatives of all orders with respect to the \(\theta'\)'s of the integral (7) can be computed under the integral sign.

PROOF: Consider the case of a continuous distribution and suppose \(\mu\) has density \(f(x)\) with respect to Lebesgue measure.

If \(|\phi| \leq M\), then
\[
|\phi(x)\exp\left[\sum_{j=1}^{k} \theta_j T_j(x)\right]| = |\phi(x)||\exp\left[\sum_{j=1}^{k} \theta_j T_j(x)\right]|
\]
\[
\leq M \exp\left[\sum_{j=1}^{k} \theta_j T_j(x)\right] \text{ since } |e^z| = e^x \text{ for } z = x + iy \text{ where } x \text{ and } y \text{ are real.}
\]

Then
\[
|\int \phi(x) \exp\left[\sum_{j=1}^{k} \theta_j T_j(x)\right] f(x) dx|
\]
\[
\leq \int |\phi(x)| \exp\left[\sum_{j=1}^{k} \theta_j T_j(x)\right] f(x) dx
\]
\[
\leq M \int \exp\left[\sum_{j=1}^{k} \theta_j T_j(x)\right] f(x) dx < +\infty \text{ since } (\xi^0_1, \ldots, \xi^0_k)
\]
\[
\text{is an interior point of } \Omega. \text{ So the integral (7) exists and is finite for all points } (\xi^0_1, \ldots, \xi^0_k) \text{ of } \Omega.
\]

Let \((\xi^0_1, \ldots, \xi^0_k)\) be any fixed point in the interior of \(\Omega\), and consider one of the variables in question, say \(\theta_1\). Express the integrand of (7) as

\[
\exp[\theta_1 T_1(x)] \phi(x) \exp[\theta_2 T_2(x) + \ldots + \theta_k T_k(x)] f(x).
\]

For the point \((\xi^0_1, \ldots, \xi^0_k)\) break the second factor into its real and imaginary parts.
(8) \( \phi(x) \exp[(\frac{i}{2} + i\eta \frac{0_2}{2})T_2(x) + ...]

\quad + (\frac{i}{2} + i\eta \frac{0_2}{2})T_k(x)]f(x)

= \phi(x) \exp[\sum \frac{k}{j} T_j(x)] \exp[\sum \eta \frac{0_j}{j}T_j(x)]f(x)

= \phi(x) \exp[\sum \frac{k}{j} T_j(x)] \left\{ \cos[\sum \eta \frac{0_j}{j}T_j(x)]

\quad + i \sin[\sum \eta \frac{0_j}{j}T_j(x)] \right\} f(x)

= \phi(x) \exp[\sum \frac{k}{j} T_j(x)] \cos[\sum \eta \frac{0_j}{j}T_j(x)]f(x)

\quad + i \phi(x) \exp[\sum \frac{k}{j} T_j(x)] \sin[\sum \eta \frac{0_j}{j}T_j(x)]f(x)

Except for the factor \( f(x) \) denote the real and imaginary parts by \( u(x) \) and \( v(x) \) respectively and break each of these functions into their positive and negative parts. Then (8) can be written as

\[ [u^+(x) - u^-(x) + iv^+(x) - iv^-(x)]f(x) \]

Hence the integral of (7) is

\[ \int \exp[\Theta_1 T_1(x)]u^+(x)f(x)dx - \int \exp[\Theta_1 T_1(x)]u^-(x)f(x)dx \]

\[ + i \int \exp[\Theta_1 T_1(x)]v^+(x)f(x)dx \]

\[ - i \int \exp[\Theta_1 T_1(x)]v^-(x)f(x)dx. \]

Now absorb each of the four factors obtained above into the density \( f(x) \) and the integral of (7) becomes

\[ \int \exp[\Theta_1 T_1(x)]f_1(x)dx - \int \exp[\Theta_1 T_1(x)]f_2(x)dx \]

\[ + i \int \exp[\Theta_1 T_1(x)]f_3(x)dx - i \int \exp[\Theta_1 T_1(x)]f_4(x)dx. \]
Hence it is sufficient to prove the result for integrals of the form $\Psi(\theta_1) = \int \exp[\theta_1 T_1(x)] f(x) dx$.

Since $(\xi_1^0, \ldots, \xi_k^0)$ is in the interior of $\Omega$, there exists $\delta > 0$ such that $\Psi(\theta_1)$ exists and is finite for all $\theta_1$, with $|{\xi}_1 - \xi_1^0| \leq \delta$ by definition of the natural parameter space. Consider the difference quotient

$$\frac{\Psi(\theta_1) - \Psi(\theta_1^0)}{\theta_1 - \theta_1^0} = \int \frac{\exp[\theta_1 T_1(x)] - \exp[\theta_1^0 T_1(x)]}{\theta_1 - \theta_1^0} f(x) dx.$$

The integrand can be written as

$$\exp[\theta_1^0 T_1(x)] \left[ \frac{\exp(\theta_1 - \theta_1^0) T_1(x)] - 1}{\theta_1 - \theta_1^0} \right] f(x).$$

Applying Lemma 4 to the second factor we see that

$$\left| \exp[\theta_1^0 T_1(x)] \left[ \frac{\exp(\theta_1 - \theta_1^0) T_1(x)] - 1}{\theta_1 - \theta_1^0} \right] f(x) \right|$$

$$\leq \left| \exp[\theta_1^0 T_1(x)] \right| \frac{\exp(\delta |T_1(x)|)}{\delta} f(x) = \frac{1}{\delta} \left| \exp[\theta_1^0 T_1(x)] + \delta |T_1(x)| \right| f(x)$$

$$\leq \frac{1}{\delta} \left| \exp[(\theta_1^0 + \delta) T_1] + \exp[(\theta_1^0 - \delta) T_1] \right| f(x) \text{ for } |\theta_1 - \theta_1^0| \leq \delta.$$

The last inequality follows from the fact that the exponential on the left of the inequality is exactly one of the exponentials on the right.

Now for any sequence of points $\theta_1^{(n)}$ tending to $\theta_1^0$ let

$$\Psi_1(\theta_1^{(n)}) = \frac{\Psi(\theta_1^{(n)}) - \Psi(\theta_1^0)}{\theta_1^{(n)} - \theta_1^0}.$$
19

Now
\[
\lim_{\theta_1^{(n)} \to \theta_1^0} \psi_1(\theta_1^{(n)}) = \lim_{\theta_1^{(n)} \to \theta_1^0} \int \frac{\exp[\theta_1^{(n)} T_1(x)] - \exp[\theta_1^0 T_1(x)]}{\theta_1^{(n)} - \theta_1^0} f(x) dx.
\]

But
\[
\lim_{\theta_1^{(n)} \to \theta_1^0} \frac{\exp[\theta_1^{(n)} T_1(x)] - \exp[\theta_1^0 T_1(x)]}{\theta_1^{(n)} - \theta_1^0} = T_1(x) \exp[\theta_1^0 T_1(x)]
\]

and
\[
\left| \frac{\exp[\theta_1^{(n)} T_1(x)] - \exp[\theta_1^0 T_1(x)]}{\theta_1^{(n)} - \theta_1^0} \right| \leq \frac{1}{\delta} \left| \exp(\theta_1^0 + \delta) T_1(x) \right|
\]

\[+ \exp(\theta_1^0 - \delta) T_1(x) \right|\]

which is integrable, hence by the Lebesgue bounded convergence theorem
\[
\lim_{\theta_1^{(n)} \to \theta_1^0} \psi_1(\theta_1^{(n)}) = \int T_1(x) \exp[\theta_1^0 T_1(x)] f(x) dx.
\]

This completes the proof of (i) and proves (ii) for the first derivative. The proof for the higher derivatives follows by induction.

3. STATING THE PROBLEM

In testing a statistical hypothesis we have two basic decisions: accept or reject the hypothesis. A decision procedure is called a test of the hypothesis in question. The decision is to be based on the value of a certain random variable \( X \), the distribution \( P_\theta \) of which is known to belong to a class \( \mathcal{P} = \{ P_\theta \mid \theta \in \Omega \} \). We let \( H \) denote the distributions of \( \mathcal{P} \) for which the hypothesis is true and \( K \) denote the distributions of \( \mathcal{P} \) for which the
hypothesis is false. We also denote the corresponding subsets of the parameter space \( \Omega \) by \( \Omega_H \) and \( \Omega_K \). Then we have that \( H \cup K = \emptyset \) and \( \Omega_H \cup \Omega_K = \Omega \). Now the hypothesis is equivalent to the statement that \( P_\theta \in H \) or \( \theta \in \Omega_H \), so we let \( H \) denote the hypothesis. The distributions in \( K \) are the alternatives to \( H \), so \( K \) is called the class of alternatives.

We make the following definitions:

**DEFINITION 8**: A nonrandomized test procedure assigns to each value \( x \) of \( X \) one of the two decisions, reject or accept the hypothesis.

In this case the sample space is divided into two complementary regions \( S_0 \) called the region of acceptance and \( S_1 \) called the region of rejection or critical region.

**DEFINITION 9**: When performing a test, rejecting the hypothesis when it is true is called an error of the first kind and accepting it when it is false is called an error of the second kind.

In carrying out the test it is desirable to minimize the probabilities of the two types of error. Since they cannot both be minimized at the same time, we assign a bound to the probability of rejecting \( H \) when it is true and attempt to minimize the probability of accepting \( H \) when it is false. The bound is a number \( \alpha \) between 0 and 1 and is called the level of significance. Classical
hypotheses testing considers minimizing the probability of accepting a false hypothesis subject to the condition that $P_\theta \{X \in S_1\} \leq \alpha$ for all $\theta \in \Omega_H$.

**DEFINITION 10:** The size of the test or critical region $S_1$ is defined as $\sup P_\theta \{X \in S_1\} = \alpha$ over $\Omega_H$.

**DEFINITION 11:** $P_\theta \{X \in S_1\}$ for a given $\theta \in \Omega_K$ is called the power of the test against the alternative $\theta$. As a function of $\theta$ for all $\theta \in \Omega$ this probability is called the power function of the test and is denoted by $\beta$.

Consider now the structure of a randomized test. For any value $x$ such a test rejects or accepts the hypothesis with probabilities depending on $x$ and denoted by $\phi(x)$ and $1-\phi(x)$ respectively. A randomized test is then characterized by a critical function $\phi$ with $0 \leq \phi(x) \leq 1$ for all $x$. If $\phi$ takes on only the values 1 and 0, then it is a nonrandomized test. The classical approach is to consider $\phi(x)$ as the conditional probability of rejection given $x$. Then it is reasonable to make the following definition:

**DEFINITION 12:** If the distribution of $X$ is $P_\theta$ with critical function $\phi(x)$, then the probability of rejection is defined as

$$E_\theta \phi(X) = \int \phi(x) dP_\theta(x),$$

which is the conditional probability $\phi(x)$ of rejection given $x$, integrated with respect to the probability distribution of $X$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The problem then is to select \( \phi \) so as to maximize the power

\[
\beta_{\phi}(\theta) = E_{\theta} \phi(X) \quad \text{for all } \theta \in \Omega_K
\]

subject to the condition

\[
E_{\theta} \phi(X) \leq \alpha \quad \text{for all } \theta \in \Omega_H.
\]

**DEFINITION 13:** A uniformly most powerful (UMP) test function \( \phi \) of size \( \alpha \) is one having maximum power for each \( \theta \in \Omega_K \).

Usually, the test that maximizes the power against a particular alternative in \( K \) depends on this alternative, which may indicate the difficulty in finding uniformly most powerful tests. We will find in the next section that if \( K \) contains only one distribution, then the problem is completely specified by (9) and (10). We will also see that under certain conditions the same test maximizes the power for alternatives in \( K \) even when there is more than one.

4. **EXISTENCE OF TESTS**

We now state the fundamental lemma of Neyman and Pearson for testing a simple hypothesis against a simple alternative, that is, when \( H \) and \( K \) each contain only a single distribution. The proof of this theorem will be omitted but may be found in [3].

**THEOREM 5:** Let \( P_0 \) and \( P_1 \) be probability distributions possessing densities \( p_0 \) and \( p_1 \) respectively with
respect to a measure \( \mu \).

(1) Existence: For testing \( H : \mu \) against the alternative \( K : \mu_{1} \) there exists a test \( \phi \) and a constant \( k \) such that

\[ (11) \quad E_{\mu} \phi(x) = \alpha \]

and

\[ (12) \quad \phi(x) = \begin{cases} 
1 \text{ when } \mu_{1}(x) > k \mu(x) \\
0 \text{ when } \mu_{1}(x) < k \mu(x)
\end{cases} \]

(ii) Sufficient condition for a most powerful test: If a test satisfies (11) and (12) for some \( k \), then it is most powerful for testing \( \mu \) against \( \mu_{1} \) at level \( \alpha \).

(iii) Necessary condition for a most powerful test: If \( \phi \) is most powerful at level \( \alpha \) for testing \( \mu \) against \( \mu_{1} \), then for some \( k \) it satisfies (12) a.e.\( \mu \). It also satisfies (11) unless there exists a test of size \( \alpha \) and with power 1.

A useful result now follows from Theorem 5.

**Corollary 1:** Let \( \beta \) denote the power of the most powerful level \( \alpha \) test for testing \( \mu \) against \( \mu_{1} \). Then \( \alpha < \beta \) unless \( \mu = \mu_{1} \).

**Proof:** We have that \( \beta = E_{\mu} \phi(x) \) and \( \alpha = E_{\mu} \phi(x) \). Suppose \( \phi(x) = \alpha \). Then the level \( \alpha \) test given by \( \phi(x) = \alpha \) has power

\[ E_{\mu} \phi(x) = \alpha P_{1}[\mu_{1}(x) > k \mu(x)] + \alpha P_{1}[\mu_{1}(x) = k \mu(x)] + \alpha P_{1}[\mu_{1}(x) < k \mu(x)] \]

= \( \alpha \). But the most powerful level \( \alpha \) test has maximum power \( \beta \), so \( \alpha \leq \beta \). If \( \alpha = \beta < 1 \), then \( \phi(x) = \alpha \) is most powerful.
and by Theorem 5 (iii) satisfies (12). That is,
\[ \phi = \begin{cases} 1 & \text{if } \frac{p_1}{p_0} > k \\ 0 & \text{if } \frac{p_1}{p_0} < k. \end{cases} \]

But since \( \phi(x) = \alpha \) we have that \( p_1 = p_0 k \) a.e.

In case \( \mu \) is Lebesgue measure we have
\[ 1 = \int p_1(x)dx = \int kp_0(x)dx = k \cdot 1 \] which implies that \( k = 1 \).
So \( p_0 = p_1 \) a.e. and \( P_0 = P_1 \).

More generally hypothesis testing problems involve a parametric family of distributions depending on one or more parameters, such as the case when the distributions depend on a single real-valued parameter \( \theta \) and the hypothesis is one-sided, say \( H : \theta \leq \theta_0 \). We should also state that in general, the most powerful test of a hypothesis \( H \) against an alternative \( \theta_1 > \theta_0 \) depends on \( \theta_1 \) and it is clear from Theorem 5 that it is then not UMP.

The following theorem gives a UMP test under an additional assumption.

**DEFINITION 14:** The real-parameter family of densities \( p_\theta(x) \) is said to have **monotone likelihood ratio** if there exists a real-valued function \( T(x) \) such that for any \( \theta < \theta' \) the distributions \( P_\theta \) and \( P_{\theta'} \), are distinct, and the ratio \( p_{\theta'}(x)/p_\theta(x) \) is a nondecreasing function of \( T(x) \).

That is, \( T(x_1) \geq T(x_2) \) implies that \( \frac{p_{\theta'}(x_1)}{p_{\theta}(x_1)} \geq \frac{p_{\theta'}(x_2)}{p_{\theta}(x_2)} \).
**THEOREM 6:** Let $\theta$ be a real parameter, and let the random variable $X$ have probability density $p_\theta(x)$ with monotone likelihood ratio in $T(x)$.

(i) For testing $H : \theta \leq \theta_0$ against $K : \theta > \theta_0$, there exists a UMP test, which is given by

$$\phi(x) = \begin{cases} 1 & \text{when } T(x) > C \\ \gamma & \text{when } T(x) = C \\ 0 & \text{when } T(x) < C \end{cases}$$

(13) where $C$ and $\gamma$ are determined by

$$(14) \quad \mathbb{E}_{\theta_0} \phi(X) = \alpha.$$ 

(ii) The power function $\beta(\theta) = \mathbb{E}_\theta \phi(X)$ of this test is strictly increasing for all points $\theta$ for which $\beta(\theta) < 1$.

(iii) For all $\theta'$, the test determined by (13) and (14) is UMP for testing $H' : \theta \leq \theta'$ against $K' : \theta > \theta'$ at level $\alpha' = \beta(\theta')$.

(iv) For any $\theta < \theta_0$ the test minimizes $\beta(\theta)$ (the probability of an error of the first kind) among all tests satisfying (14).

**PROOF:** (i) and (ii). Consider the hypothesis $H_0 : \theta = \theta_0$ and a simple alternative $\theta_1 > \theta_0$. By Theorem 5, the most powerful test rejects when $\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} > C$. Since $X$ has probability density $p_\theta(x)$ with monotone likelihood ratio in $T(x)$, it follows from Definition 14 that the most powerful test rejects when $T(x) > C$. So by Theorem 5 (1),
there exist $C$ and $\delta$ such that (13) and (14) hold.

By Theorem 5 (11), the resulting test is most powerful for testing $P_{\Theta}$ against $P_{\Theta''}$ at level $\alpha' = \beta(\theta')$ for $\theta' < \theta''$. Now $\beta(\theta'')$ is the power of this most powerful level $\alpha'$ test, so by Corollary 1, $\beta(\theta') < \beta(\theta'')$ unless $P_{\Theta'} = P_{\Theta''}$. But by Definition 14, $P_{\Theta'}$ and $P_{\Theta''}$ are distinct. Hence the power function is strictly increasing for all $\theta$ such that $\beta(\theta) < 1$.

Since $\beta(\theta)$ is nondecreasing, $E_{\Theta} \phi(X) \leq E_{\Theta_{o}} \phi(X) = \alpha$ for $\theta \leq \theta_{o}$ and we have

$$E_{\Theta} \phi(X) \leq \alpha \text{ for } \theta \leq \theta_{o}.$$ (15)

If $\phi$ satisfies $E_{\Theta} \phi(X) \leq \alpha$, then it satisfies $E_{\Theta} \phi(X) \leq \alpha$ for $\theta \leq \theta_{o}$. So the class of tests satisfying (15) is contained in the class satisfying $E_{\Theta_{o}} \phi(X) \leq \alpha$. The given test maximizes $\beta(\theta_{1})$ within the class satisfying $E_{\Theta_{o}} \phi(X) \leq \alpha$, hence it also maximizes $\beta(\theta_{1})$ subject to (15). The test is also independent of the particular alternative $\theta_{1} > \theta_{o}$ chosen and therefore, it is UMP against $K$.

Part (iii) is proved analogously. When we applied Theorem 5 we found a test of $H_{0}: \theta = \theta_{o}$ against $K: \theta = \theta_{1} > \theta_{o}$ which maximized the power. Reversing all inequalities produces a test which minimizes the power for testing $H_{0}: \theta = \theta_{o}$ against $K: \theta = \theta_{1} < \theta_{o}$. Hence part (iv) follows by applying Theorem 5 with all inequalities reversed.

The following is a corollary to Theorem 6 and holds for the one-parameter exponential families of distributions.
by letting \( \phi(\theta) = \theta \).

**COROLLARY 2:** Let \( \theta \) be a real parameter, and let 
\( X \) have probability density with respect to some measure \( \mu \) 
\[ p_\theta(x) = C(\theta) \exp[\phi(\theta)T(x)]h(x) \]
where \( \phi \) is strictly monotone. Then there exists a UMP 
test \( \phi \) for testing \( H : \theta \leq \theta_0 \) against \( K : \theta > \theta_0 \). If \( \phi \)
is increasing, the test is given by (13) and (14). If \( \phi \)
is decreasing the inequalities are reversed.

**PROOF:** It is only necessary to show that the 
family of densities \( p_\theta(x) \) have monotone likelihood ratio 
in \( T(x) \). If \( \phi \) is increasing, then \( \phi(\theta') - \phi(\theta) > 0 \) for 
\( \theta < \theta' \). Suppose \( T(x) \geq T(x') \). Then 
\[ (\phi(\theta') - \phi(\theta))T(x) \geq (\phi(\theta') - \phi(\theta))T(x') \]
\[ \Rightarrow \frac{C(\theta')}{C(\theta)} \exp[(\phi(\theta') - \phi(\theta))T(x)] \geq \frac{C(\theta')}{C(\theta)} \exp[(\phi(\theta') - \phi(\theta))T(x')] \]
for positive \( C(\theta') \) and \( C(\theta) \). Therefore, we have that
\[ \frac{p_{\theta'}(x)}{p_\theta(x)} = \frac{C(\theta')}{C(\theta)} \exp[\phi(\theta')T(x)] \geq \frac{C(\theta')}{C(\theta)} \exp[\phi(\theta')T(x')] \]
\[ = \frac{p_{\theta'}(x)}{p_\theta(x')} \]
and \( p_\theta(x) \) has monotone likelihood ratio in \( T(x) \). The re­
sult follows from Theorem 6. The results are similar for 
\( \phi \) decreasing.

The following theorem is a generalization of 
Theorem 5.

**THEOREM 7:** Let \( f_1, \ldots, f_{m+1} \) be real-valued
functions defined on a Euclidean space \( \mathcal{X} \) and integrable \( \mu \), and suppose that for given constants \( C_1, \ldots, C_m \) there exists a critical function \( \phi \) satisfying
\[
\int \phi f_1 d\mu = C_i, \quad i = 1, \ldots, m.
\]
Let \( \mathcal{C} \) be the class of critical functions \( \phi \) for which (17) holds.

(i) Among all members of \( \mathcal{C} \) there exists one that maximizes
\[
\int \phi f_{m+1} d\mu.
\]

(ii) A sufficient condition for a member of \( \mathcal{C} \) to maximize (18) is the existence of constants \( k_1, \ldots, k_m \) such that
\[
\begin{cases}
1 & \text{when } f_{m+1}(x) > \sum_{i=1}^{m} k_i f_i(x) \\
0 & \text{when } f_{m+1}(x) < \sum_{i=1}^{m} k_i f_i(x).
\end{cases}
\]

(iii) If a member of \( \mathcal{C} \) satisfies (19) with \( k_1, \ldots, k_m \geq 0 \), then it maximizes (18) among all critical functions satisfying
\[
\int \phi f_1 d\mu \leq C_i, \quad i = 1, \ldots, m.
\]

(iv) The set \( M \) of points in \( m \)-dimensional space whose coordinates are
\[
(\int \phi f_1 d\mu, \ldots, \int \phi f_m d\mu)
\]
for some critical function \( \phi \) is convex and closed. If \( (C_1, \ldots, C_m) \) is an inner point of \( M \), then there exist constants \( k_1, \ldots, k_m \) and a test \( \phi \) satisfying (17) and (19), and a necessary condition for a member of \( \mathcal{C} \) to
maximize (18) is that (19) holds a.e.u.

Existence of a UMP test for two-sided hypotheses for exponential families is given by the following theorem.

**THEOREM 8:**

(i) For testing the hypothesis $H : \theta \leq \theta_1$ or $\theta \geq \theta_2$ $(\theta_1 < \theta_2)$ against the alternatives $K : \theta_1 < \theta < \theta_2$ in a one-parameter exponential family there exists a UMP test given by

\[
\phi(x) = \begin{cases} 
1 & \text{when } C_1 < T(x) < C_2 \quad (C_1 < C_2) \\
\gamma_i & \text{when } T(x) = C_i, \quad i = 1, 2 \\
0 & \text{otherwise}
\end{cases}
\]

where the $C$'s and $\gamma$'s are determined by

\[
E_{\theta_1} \phi(X) = E_{\theta_2} \phi(X) = \alpha.
\]

(ii) This test minimizes $E_{\theta} \phi(X)$ subject to (22) for each $\theta < \theta_1$ or $> \theta_2$.

(iii) For $0 < \alpha < 1$ the power function of this test has a maximum at a point $\theta_0$ between $\theta_1$ and $\theta_2$ and decreases strictly as $\theta$ tends away from $\theta_0$ in either direction, unless there exist two values $t_1, t_2$ such that $P_{\theta} \{T(X) = t_1\} + P_{\theta} \{T(X) = t_2\} = 1$ for all $\theta$.

Parts (ii) and (iii) of Theorem 8 are proved in the same way as part (ii) of Theorem 5 and the proof of parts (i) and (iv) makes use of a weak compactness theorem for critical functions and a theorem on dominated families of distributions. Since it is not the purpose of this
paper to go into such theory, the proof is omitted but can be found in [3].

The proof of Theorem 8 depends strongly on Theorem 7. In the proof attention is restricted to the sufficient statistic $T = T(X)$ whose distribution is represented by $dP_\theta(t) = C(\theta) \exp[Q(\theta)t]dv(t)$, which in particular for $Q(\theta) = \theta$ is of the exponential type. For $\theta_1 < \theta < \theta_2$ the first problem is that of maximizing $E_\theta \Psi(T)$ subject to (22) with $\phi(x) = \Psi(T(x))$. The idea is to let $M$ denote the set of all points $(E_\theta \Psi(T), E_\theta \Psi(T))$ as $\Psi$ ranges over the totality of critical functions and then verify that the point $(a,a)$ is an inner point of $M$. Then part (iv) of Theorem 7 is applied to obtain constants $k_1$, $k_2$ and a test $\Psi_0(t)$ satisfying (22) such that $\Psi_0(t) = 1$ when $k_1 C(\theta_1) \exp[Q(\theta_1)t] + k_2 C(\theta_2) \exp[Q(\theta_2)t]$

$$< C(\theta') \exp[Q(\theta')t],$$

which may be written as $a_1 \exp(b_1 t) + a_2 \exp(b_2 t) < 1$ for $b_1 < 0 < b_2$, and such that $\Psi_0(t) = 0$ when the above expression on the left hand side is $> 1$. This test is then shown to satisfy (21). To prove that the test is UMP and also to prove part (ii) of Theorem 8, part (iv) of Theorem 7 is used again. For the many details the reader is again referred to [3].
1. UNBIASEDNESS, SIMILARITY AND COMPLETENESS

**DEFINITION 15:** A test \( \phi \) is said to be unbiased if the power function \( \beta_\phi(\theta) = E_\theta \phi(X) \) satisfies

\[
\begin{align*}
\beta_\phi(\theta) &\leq \alpha \text{ if } \theta \in \Omega_H \\
\beta_\phi(\theta) &\geq \alpha \text{ if } \theta \in \Omega_K
\end{align*}
\]

A UMP test is unbiased since by Corollary 1 its power cannot fall below that of the test \( \phi(x) = \alpha \). For a large class of problems for which a UMP test does not exist, there does exist a UMP unbiased test.

**THEOREM 9:** Let \( \mathcal{T} \) be a collection of subsets of \( \Omega \) such that \( \Omega \) and the empty set are in \( \mathcal{T} \), arbitrary unions of sets in \( \mathcal{T} \) are in \( \mathcal{T} \) and finite intersections of sets in \( \mathcal{T} \) are in \( \mathcal{T} \). If \( \beta_\phi(\theta) \) is a continuous function of \( \theta \) on \( \mathcal{T} \), then unbiasedness implies

\[
\beta_\phi(\theta) = \alpha \quad \text{for all } \theta \in \omega
\]

where \( \omega \) is the common boundary of \( \Omega_H \) and \( \Omega_K \).

**PROOF:** Let \( \theta \in \omega \) and suppose \( \beta(\theta) < \alpha \). Then there exists an \( \epsilon > 0 \) such that \( \beta(\theta) = \alpha - \epsilon \). Consider the open set \( (\alpha - 2\epsilon, \alpha) \). Let \( U = \beta^{-1}(\alpha - 2\epsilon, \alpha) \), then \( U \) is open since \( \beta \) is continuous. Now \( \beta(\theta) = \alpha - \epsilon \) is in \( (\alpha - 2\epsilon, \alpha) \) so \( \theta \in U \). But \( \theta \in \omega \) implies that \( U - \{\theta\} \cap \Omega_K \) is not empty. So there exists a \( p \in U \) such that \( p \neq \theta \) and \( p \in \Omega_K \). Then \( \beta(p) \geq \alpha \) but \( p \in U \) implies that \( \beta(p) \in (\alpha - 2\epsilon, \alpha) \) which in turn implies that \( \beta(p) < \alpha \), a contradiction. Hence
32

By a similar argument it can be shown that $\beta(\theta) \geq \alpha$. LEt a similar argument it can be shown that $\beta(\theta) \leq \alpha$ so that (24) holds.

**Lemma 5:** If the distributions $P_\theta$ are such that the power function of every test is continuous, and if $\phi_o$ is UMP among all tests satisfying (24) and is a level $\alpha$ test of $H$, then $\phi_o$ is UMP unbiased.

**Proof:** Let $A$ be the set of all tests $\phi$ satisfying (24) and let $B$ be the set of all tests $\phi$ satisfying (23). Then $B \subseteq A$. Now $E_\theta \phi_o(X) \leq \alpha$ for $\theta \in \Omega_H$ and if we let $\phi^*$ be in $A$, then $E_\theta \phi^*(X) \leq E_\theta \phi_o(X)$ for $\theta \in \Omega_K$. But $B \subseteq A$ so for every $\phi' \in B$, $E_\theta \phi'(X) \leq E_\theta \phi_o(X)$ for $\theta \in \Omega_K$ and $\phi_o$ is uniformly at least as powerful as any unbiased test.

For $\phi(x) = \alpha$, $\beta_\phi(\theta) = \alpha$ for $\theta \in \Omega_H$ and $\theta \in \Omega_K$, hence $E_\theta \phi(X) \leq E_\theta \phi(X)$ for $\theta \in \Omega_K$ since $\phi \in A$. So $\phi_o$ uniformly at least as powerful as $\phi(x) = \alpha$. Therefore, $\beta_\phi(\theta) \leq \alpha$ for $\theta \in \Omega_H$ and $\beta_\phi(\theta) \geq \alpha$ for $\theta \in \Omega_K$ and $\phi_o$ is UMP unbiased.

**Definition 16:** A test $\phi$ is said to be similar with respect to $\mathcal{P}^X$ or $\omega$ if $E_\theta \phi(X) = \alpha$ for all distributions of $X$ belonging to $\mathcal{P}^X = \{P_\theta, \theta \in \omega\}$.

Let $T$ be a sufficient statistic for $\mathcal{P}^X$, and let $\mathcal{P}^T = \{P_\theta^T, \theta \in \omega\}$. Then any test satisfying (25)

$$E[\phi(X)|T] = \alpha \quad \text{a.e. } \mathcal{P}^T$$

is similar with respect to $\mathcal{P}^X$ since

$$E_\theta[\phi(X)] = E_\theta \{E[\phi(X)|T]\} = \alpha \quad \text{for all } \theta \in \omega \quad .$$

Such a test is said to have Neyman structure with respect...
to $T$.

**Definition 17**: A family $\mathcal{P}$ of probability distributions $P$ is complete if

\[(26) \quad E_P[f(X)] = 0 \quad \text{for all } P \in \mathcal{P} \]

implies

\[(27) \quad f(x) = 0 \quad \text{a.e. } \mathcal{P}. \]

**Theorem 10**: Let $X$ be a random vector with probability distribution

\[dP_\theta(x) = C(\theta) \exp\left[ \sum_{j=1}^{k} \theta_j T_j(x) \right] d\mu(x) \]

and let $\mathcal{P}^T$ be the family of distributions of $T = (T_1(X), \ldots, T_k(X))$ as $\theta$ ranges over the set $\omega$. Then $\mathcal{P}^T$ is complete provided $\omega$ contains a $k$-dimensional rectangle.

**Proof**: By a suitable change of coordinates suppose $\omega$ contains the rectangle $I = \{ (\theta_1, \ldots, \theta_k) : -a < \theta_j < a, \ j = 1, \ldots, k \}$. Let $f(t) = f^+(t) - f^-(t)$ be such that $E_\theta f(T) = 0$ for all $\theta \in \omega$. We will give the proof only in the case that $\mu$ and $\nu$, the measure induced in $T$-space by $\mu$, are both absolutely continuous with respect to Lebesgue measure. Then $dv(t) = h(t)dt$ and in the following expressions we absorb $h(t)$ into $f(t)$. Then for all $\theta \in I$.

\[ \int f(t)C(\theta) \exp\left[ \sum_{j=1}^{k} \theta_j t_j \right] dt = 0 \]

\[ \Rightarrow \int [f^+(t) - f^-(t)] \exp\left[ \sum_{j=1}^{k} \theta_j t_j \right] dt = 0 \]

\[ \Rightarrow \int \exp\left[ \sum_{j=1}^{k} \theta_j t_j \right] f^+(t) dt = \int \exp\left[ \sum_{j=1}^{k} \theta_j t_j \right] f^-(t) dt \]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[
\Rightarrow \int f^+(t)dt = \int f^-(t)dt, \text{ since in particular we can let } \\
\theta_j = 0 \text{ for } j = 1, \ldots, k. \text{ Dividing } f \text{ by a constant we can take the common value of these two integrals to be 1, so that } P^+(t) \text{ and } P^-(t) \text{ are probability measures with densities } f^+(t) \text{ and } f^-(t) \text{ respectively. Then we have }
\]

\[
(28) \int \exp[\sum_{j=1}^k \theta_j t_j]f^+(t)dt = \int \exp[\sum_{j=1}^k \theta_j t_j]f^-(t)dt
\]

for all \( \Theta \in \mathfrak{I} \). Consider these integrals as functions of the complex variables \( \theta_j = x_j + iy_j, j = 1, \ldots, k \).

For any fixed \( \theta_1, \ldots, \theta_j-1, \theta_{j+1}, \ldots, \theta_k \) with real parts strictly between \(-a\) and \(+a\), these integrals are by Theorem 4 analytic functions of \( \theta_j \) in the strip

\( \Re_j: -a < x_j < a, -\infty < y_j < \infty \) of the complex plane.

For \( \theta_2, \ldots, \theta_k \) fixed, real, and between \(-a\) and \(+a\), (28) holds on the line segment \( \{(x_1, y_1) : -a < x_1 < a, y_1 = 0\} \) and can be extended to the strip \( R_1 \), in which the integrals are finite. By induction (28) holds in the complex region \( \{(\theta_1, \ldots, \theta_k) : (x_j, y_j) \in R_j \text{ for } j = 1, \ldots, k\} \). Then for all real \( (\eta_1, \ldots, \eta_k) \)

\[
\int \exp[\sum_{j=1}^k \eta_j t_j]f^+(t)dt = \int \exp[\sum_{j=1}^k \eta_j t_j]f^-(t)dt.
\]

The left and right hand sides of this equation represent the characteristic functions of the distributions \( P^+ \) and \( P^- \) respectively. By the uniqueness theorem for characteristic functions, \( P^+ \) and \( P^- \) are equal and by their definition, \( f^+(t) = f^-(t) \) a.e. Therefore, \( f(t) = 0 \) a.e. \( \mathcal{P}^T \) and \( \mathcal{P}^T \) is complete.
2. ONE-PARAMETER EXPONENTIAL FAMILIES

Let $\theta$ be a real parameter and $X = (X_1, \ldots, X_n)$ a random vector with probability density

$$p_\theta(x) = C(\theta) \exp[\theta T(x)]h(x),$$

where $x = (x_1, \ldots, x_n)$.

We have seen that a UMP test exists for

(i) $H : \theta \leq \theta_0$, $K : \theta > \theta_0$ by Corollary 2,

(ii) $H : \theta \leq \theta_1$ or $\geq \theta_2$, $K : \theta_1 < \theta < \theta_2$ by Theorem 8, but existence of a UMP test has not been given for

(iii) $H : \theta_1 < \theta < \theta_2$, $K : \theta < \theta_1$ or $\theta > \theta_2$,

(iv) $H : \theta = \theta_0$, $K : \theta \neq \theta_0$.

Existence of a UMP unbiased test for (iii) and (iv) is given by the following theorems.

**THEOREM 11:** There exists a UMP unbiased test for (iii) given by

$$\phi(x) = \begin{cases} 1 & \text{when } T(x) < C_1 \text{ or } > C_2 \\ \gamma'_1 & \text{when } T(x) = C_1, \, i = 1, 2 \\ 0 & \text{when } C_1 < T(x) < C_2 \end{cases}$$

where the $C$'s and $\gamma$'s are determined by

$$E_{\theta_1} \phi(X) = E_{\theta_2} \phi(X) = \alpha.$$

**PROOF:** $E_{\theta} \phi(X)$ is continuous by Theorem 4 so that Lemma 5 is applicable, where $\omega$ consists of the two points $\theta_1$ and $\theta_2$. In the case at hand,

$$1 - \phi(x) = \begin{cases} 1 & \text{when } C_1 < T(x) < C_2 \\ \gamma'_1 & \text{when } T(x) = C_1, \, k = 1, 2 \\ 0 & \text{when } T(x) < C_1 \text{ or } > C_2 \end{cases}$$
so by Theorem 8 (ii), $1-\phi(x)$ minimizes $1-E_\theta \phi(X)$ subject
to $1-E_{\theta_1} \phi(X) = 1-E_{\theta_2} \phi(X) = \alpha^* = 1-\alpha$ for all $\theta < \theta_1$ and $> \theta_2$.

This implies that $E_\theta \phi(X)$ is a maximum subject to

\begin{equation}
E_{\theta_1} \phi(X) = E_{\theta_2} \phi(X) = \alpha.
\end{equation}

Then $\phi$ is UMP among all tests satisfying (31) and hence
UMP unbiased by Lemma 5. Also by Theorem 8 (iii) the
power function of the test $\phi(x)$ has a minimum at a point
between $\theta_1$ and $\theta_2$ and is strictly increasing as $\theta$ tends
away from this minimum in either direction.

**THEOREM 12**: There exists a UMP unbiased test
for (iv) given by (29) where the $C$'s and $\mathcal{X}$'s are now
determined by

\begin{equation}
E_{\theta_0}[\phi(X)] = \alpha
\end{equation}

and

\begin{equation}
E_{\theta_0}[T(X)\phi(X)] = E_{\theta_0}[T(X)]\alpha.
\end{equation}

**PROOF**: Let $\theta'$ be any particular alternative and
restrict attention to the sufficient statistic $T$, the dis­
tribution of which by Lemma 3 is of the form

dP_\theta(t) = C(\theta) \exp[\theta t]dv(t). \quad \text{We will give the proof for}
the case that $v$ is absolutely continuous with respect to
Lebesgue measure so that $dv(t) = g(t)dt$. Let $\phi(x) = \Psi[T(x)]$.
Unbiasedness of $\Psi$ implies (32) and for every $\theta \neq \theta_0$,
$\beta_{\phi}(\theta) \geq \alpha$ and so $\beta_{\phi}(\theta) = E_{\theta_0}[\Psi(T)]$ must be a minimum.

By Theorem 4 $\beta(\theta)$ is differentiable and the derivative can
be computed under the expectation sign. So for all tests $\Psi(t)$
\[ \beta'(\theta) = \int \psi(t) [t c(\theta) \exp (\theta t) g(t) + c'(\theta) \exp (\theta t) g(t)] dt \]
\[ = \int t \psi(t) c(\theta) \exp (\theta t) g(t) dt + \frac{c'(\theta)}{c(\theta)} \int \psi(t) c(\theta) \exp (\theta t) g(t) dt \]
\[ = E_\theta [T \psi(T)] + \frac{c'(\theta)}{c(\theta)} E_\theta [\psi(T)]. \]

For \( \psi(t) = \alpha \), \( \beta(\theta) = E_\theta [\psi(T)] = \alpha \) and \( \beta'(\theta) = 0 \). So the above equation becomes
\[ 0 = \alpha E_\theta (T) + \frac{c'(\theta)}{c(\theta)} \alpha \Rightarrow \frac{c'(\theta)}{c(\theta)} = -E_\theta (T). \]

Hence \( \beta'(\theta) = E_\theta [T \psi(T)] - E_\theta (T) E_\theta [\psi(T)]. \)

Now \( \beta(\theta) \) is a minimum at \( \theta_0 \) so \( \beta'(\theta_0) = 0 \) and replacing \( \psi(T) \) by \( \phi(X) \) again,
\[ 0 = E_\theta_0 [T \phi(X)] - E_\theta_0 (T) E_\theta_0 [\phi(X)] \Rightarrow E_\theta_0 [T \phi(X)] = E_\theta_0 (T) \alpha \]
and hence unbiasedness implies (33) as well as (32).

Let \( M \) be the set of points \( (E_\theta_0 [\psi(T)], E_\theta_0 [T \psi(T)]) \)
as \( \psi \) ranges over the totality of critical functions.

Suppose \( (E_\theta_0 [\psi_1(T)], E_\theta_0 [T \psi_1(T)]) \) and \( (E_\theta_0 [\psi_2(T)], E_\theta_0 [T \psi_2(T)]) \)
are points in \( M \) and \( \alpha \) is such that \( 0 < \alpha < 1 \).
Then
\[ \alpha (E_\theta_0 [\psi_1(T)], E_\theta_0 [T \psi_1(T)]) + (1-\alpha) (E_\theta_0 [\psi_2(T)], E_\theta_0 [T \psi_2(T)]) \]
\[ = (E_\theta_0 [\alpha \psi_1(T)], E_\theta_0 [\alpha T \psi_1(T)]) + (E_\theta_0 [(1-\alpha) \psi_2(T)], E_\theta_0 [(1-\alpha) T \psi_2(T)]) \]
\[ = (E_\theta_0 [\alpha \psi_1(T) + (1-\alpha) \psi_2(T)], E_\theta_0 [T(\psi_1(T) + (1-\alpha) \psi_2(T))]) \]
\[ = (E_\theta_0 [\psi(T)], E_\theta_0 [T \psi(T)]) \text{ is in } M \text{ since } \psi(T) = \alpha \psi_1(T) + (1-\alpha) \psi_2(T) \text{ is also a critical function.} \]
Hence M is convex and contains all points \((u, uE_{\Theta_o}(T))\) with \(0 < u < 1\). A result of Corollary 2 is that \(\beta'(\theta) > 0\) and so \(u_2 = E_{\Theta_o}[T\Psi(T)] > E_{\Theta_o}(T)E_{\Theta_o}[\Psi(T)]\). Then M contains points \((\alpha, u_2)\) with \(u_2 > \alpha E_{\Theta_o}(T)\). Similarly it contains points \((\alpha, u_1)\) with \(u_1 < \alpha E_{\Theta_o}(T)\). Hence the point \((\alpha, \alpha E_{\Theta_o}(T))\) is an inner point of the convex set M of points \((E_{\Theta_o}[\Psi(T)], E_{\Theta_o}[T\Psi(T)])\) where

\[
E_{\Theta_o}[\Psi(T)] = \int \Psi(t)C(\Theta_o) \exp [\Theta_o t]g(t)dt \quad \text{and} \quad E_{\Theta_o}(T\Psi(T)) = \int t\Psi(t)C(\Theta_o) \exp [\Theta_o t]g(t)dt.
\]

Let \(f_1 = C(\Theta_o) \exp [\Theta_o t]\) and \(f_2 = t C(\Theta_o) \exp [\Theta_o t]\), then applying Theorem 7 (iv) there exist constants \(k_1\) and \(k_2\) and a test \(\Psi\) satisfying

\[
\int \Psi(t)f_1g(t)dt = E_{\Theta_o}[\Psi(T)] = \alpha \quad \text{and} \quad \int \Psi(t)f_2g(t)dt = E_{\Theta_o}[T\Psi(T)] = \alpha E_{\Theta_o}(T)
\]

such that \(\Psi(t) = 1\) when

\[
k_1C(\Theta_o) \exp [\Theta_o t] + k_2 t C(\Theta_o) \exp [\Theta_o t] < C(\theta') \exp [\theta't]
\]

which may be written as \(a_1 + a_2 t < \exp (bt)\).

Considering the graphical representations of both sides of the inequality it is readily seen that this region is either one-sided or the outside of an interval. But by Theorem 6 (ii) a one-sided test has a strictly monotone power function which implies that \(\beta'(\Theta_o) \neq 0\) contrary to our derivation of (33). Therefore the region is the out-
side of an interval and \( \psi(t) = 1 \) when \( t < C_1 \) or \( > C_2 \). So the most powerful test subject to (32) and (33) is given by (29). This test is unbiased by comparing it with \( \phi(x) \equiv \alpha \), but the class of tests satisfying (32) and (33) includes the class of unbiased tests, so the given test is UMP unbiased.

3. MULTIPARAMETER EXPONENTIAL FAMILIES

Let \( X \) be distributed according to

\[
(34) \quad dP_{\theta, \mathcal{A}}^{X} (x) = C(\theta, \mathcal{A}) \exp[\theta U(x) + \sum_{i=1}^{k} \alpha_i T_i(x)]d\mu(x),
\]

\( (\theta, \mathcal{A}) \in \Omega \) and let \( \mathcal{V} = (\mathcal{V}_1, \ldots, \mathcal{V}_k) \), \( T = (T_1, \ldots, T_k) \) and \( \Omega \) be convex with dimension \( k + 1 \). The sufficient statistics \( (U, T) \) have joint distribution

\[
(35) \quad dP_{\theta, \mathcal{A}}^{U,T}(u,t) = C(\theta, \mathcal{A}) \exp \left[ \theta u + \sum_{i=1}^{k} \alpha_i t_i \right]dv(u,t),
\]

\( (\theta, \mathcal{A}) \in \Omega \). By Lemma 3 the conditional distribution of \( U \) given \( t \) constitutes an exponential family

\[
(36) \quad dP_{\theta}^{U|t}(u) = C_t(\theta) \exp (\theta u)dv_t(u).
\]

In this conditional case the following hypotheses have been considered:

\[
H_1 : \theta \leq \theta_0 \quad K_1 = \theta > \theta_0
\]

\[
H_2 : \theta \leq \theta_1 \text{ or } \theta \geq \theta_2 \quad K_2 : \theta_1 < \theta < \theta_2
\]

\[
H_3 : \theta_1 \leq \theta \leq \theta_2 \quad K_3 : \theta < \theta_1 \text{ or } \theta > \theta_2
\]

\[
H_4 : \theta = \theta_0 \quad K_4 : \theta \neq \theta_0
\]

Recall that we have the following results.

There exists by Corollary 2 a UMP test for \( H_1 \) with critical function \( \phi_1 \) satisfying
\begin{equation}
\phi(u,t) = \begin{cases} 
1 & \text{when } u > C_o(t) \\
\gamma_o(t) & \text{when } u = C_o(t) \\
0 & \text{when } u < C_o(t)
\end{cases}
\end{equation}

where the functions \(C_o\) and \(\gamma_o\) are determined by

\begin{equation}
E_{\Theta_o} [\phi_1(U,T)|t] = \alpha \text{ for all } t.
\end{equation}

There exists by Theorem 8 a UMP test for \(H_2\) with critical function \(\phi_2\) satisfying

\begin{equation}
\phi(u,t) = \begin{cases} 
1 & \text{when } C_1(t) < u < C_2(t) \\
\gamma_1(t) & \text{when } u = C_i(t), \ i = 1, 2 \\
0 & \text{when } u < C_1(t) \text{ or } > C_2(t)
\end{cases}
\end{equation}

where the \(C_i\)'s and \(\gamma_i\)'s are determined by

\begin{equation}
E_{\Theta_1} [\phi_2(U,T)|t] = E_{\Theta_2} [\phi_2(U,T)|t] = \alpha.
\end{equation}

There exists by Theorem 11 a UMP unbiased test for \(H_3\) with critical function \(\phi_3\) satisfying

\begin{equation}
\phi(u,t) = \begin{cases} 
1 & \text{when } u < C_1(t) \text{ or } > C_2(t) \\
\gamma_1(t) & \text{when } u = C_i(t), \ i = 1, 2 \\
0 & \text{when } C_1(t) < u < C_2(t)
\end{cases}
\end{equation}

where the \(C_i\)'s and \(\gamma_i\)'s are determined by

\begin{equation}
E_{\Theta_1} [\phi_3(U,T)|t] = E_{\Theta_2} [\phi_3(U,T)|t] = \alpha.
\end{equation}

There exists by Theorem 12 a UMP unbiased test for \(H_4\) with critical function \(\phi_4\) satisfying (41) with the \(C_i\)'s and \(\gamma_i\)'s determined by

\begin{equation}
E_{\Theta_o} [\phi_4(U,T)|t] = \alpha \text{ and }
\end{equation}

\begin{equation}
E_{\Theta_o} [U\phi_4(U,T)|t] = \alpha E_{\Theta_o} [U|t].
\end{equation}
We would now like to prove the same results with regard to the joint distribution.

**THEOREM 13**: The critical functions \( \phi_1, \phi_2, \phi_3 \) and \( \phi_4 \) defined above constitute UMP unbiased level \( \alpha \) tests for testing the hypotheses \( H_1, \ldots, H_4 \) respectively when the joint distribution of \( U \) and \( T \) is given by (35).

**PROOF**: We will restrict the proof to the case of discrete distributions. For fixed \( \theta \) the statistic \( T \) is sufficient for \( \xi \) and hence is sufficient for each

\[
\omega_j = \{(\theta, \xi) : (\theta, \xi) \in \Omega, \theta = \theta_j^0, j = 0, 1, 2.\}
\]

By Lemma 3 the family of distributions of \( T \) is given by

\[
P_{\theta_j, \xi}(t) = C(\theta_j, \xi) \exp\left[\sum_{i=1}^{k-1} \xi_i t_i\right] n_\theta(t),
\]

\((\theta_j, \xi) \in \omega_j; j = 0, 1, 2.\)

Now \( \Omega \) is convex and of dimension \( k + 1 \). Let us further assume it contains points on both sides of \( \theta = \theta_j \). Then since \( \theta_j \) is fixed, \( \omega_j \) is convex, of dimension \( k \) and contains a \( k \)-dimensional rectangle. Hence by Theorem 10

\[
\mathcal{P}_j = \left\{ P_{\theta_j, \xi}; (\theta, \xi) \in \omega_j \right\}
\]

is complete and similarity of a test \( \phi \) on \( \omega_j \) implies

\[
E_{\theta_j}[\phi(U,T)|t] = \alpha.
\]

Consider \( H_1 \). By Theorem 4 the power function of each test is continuous, so we can apply Lemma 5 and show \( \phi_1 \) to be UMP among all tests similar on \( \omega_0 \). The overall power of a test \( \phi \) against an alternative \((\theta, \xi)\) is

\[
E_{\theta, \xi} \left[ \phi(U,T) \right] = \sum_{t, u} \left[ \sum_{t, u} \phi(u,t) p_{\theta, \xi}^U(t, u) p_{\theta, \xi}^T(t) \right].
\]
where \( p_{\theta}^{U|t}(u) \) and \( p_{\theta,\alpha}^{T}(t) \) are the densities with respect to counting measure of \( p_{\theta}^{U|t}(u) \) and \( p_{\theta,\alpha}^{T}(t) \) respectively. We have already seen that \( \phi_1 \) maximizes the conditional power against any \( \theta > \theta_0 \) subject to (38). Since this conditional power is given inside the brackets in (45) we have that \( \phi_1 \) maximizes the over-all power and the result follows from Lemma 5. The proof for \( H_2 \) and \( H_3 \) is similar.

By Theorem 9 unbiasedness of a test of \( H_4 \) implies similarity on \( \omega_0 \) and

\[
\frac{\partial}{\partial \theta} \left[ E_{\theta,\alpha}^{T} \phi(U, T) \right] = 0 \text{ on } \omega_0.
\]

By Theorem 4 we can differentiate under the expectation sign and our result follows exactly as in the consideration of (33) in Theorem 12. Then the equation is equivalent to

\[
E_{\theta,\alpha}^{T} [U \phi(U, T) - \alpha U] = 0 \text{ on } \omega_0
\]

and since \( \mathcal{D}_\alpha^{T} \) is complete, unbiasedness implies (43) and (44). Now \( \phi_4 \) also satisfies (41) and is UMP among all tests satisfying (43) and (44). By comparison with the test \( \phi(u, t) = \alpha \) we have that \( \phi_4 \) is UMP unbiased.

4. APPLICATIONS

In the present section we will be concerned with discrete distributions only.

Consider the problem of comparing two Poisson distributions. We would like to formulate the problem in such a manner as to use the results of section 3. Suppose
the random variables $X$ and $Y$ are independently Poisson distributed with parameters $\lambda$ and $\mu$ respectively. Their joint distribution

$$P(X = x, Y = y) = \frac{\exp[-(\lambda + \mu)] \lambda^x \mu^y}{x!y!}$$

can be written in the form of an exponential family as

(46) $$P(X = x, Y = y) = \frac{\exp[-(\lambda + \mu)]}{x!y!} \exp[y \log \frac{\mu}{\lambda} + (x+y) \log \lambda].$$

By Theorem 13 there exist UMP unbiased tests of $H_1$, ..., $H_4$ concerning $\Theta = \log(\mu/\lambda)$ or $\rho = \frac{\mu}{\lambda}$ so in particular for testing $H : \mu \leq \lambda$ against $K : \mu > \lambda$, $\mu \leq \lambda$ implies $\log \frac{\mu}{\lambda} \leq 0$. Comparing (46) with (34) we see that $U = Y$ and $T = X + Y$ and the tests are performed conditionally on the integer points of the line segment $X + Y = t$.

It is known that $T$ has a Poisson distribution with parameter $\mu + \lambda$ and the conditional distribution of $Y$ given $T = t$ is

$$P(Y = y \mid X+Y = t) = \binom{t}{y} \left(\frac{\mu}{\lambda + \mu}\right)^y \left(1 - \frac{\mu}{\lambda + \mu}\right)^{t-y}.$$  

So $P(Y = y \mid X+Y = t) = t\binom{t}{y} \left(\frac{\mu}{\lambda + \mu}\right)^y \left(\frac{\Lambda}{\lambda + \mu}\right)^{t-y}$, $y = 0, 1, \ldots, t$

which is the binomial distribution with $t$ trials and probability $p = \frac{\mu}{\lambda + \mu}$ of success. Now $\mu \leq a\lambda \Rightarrow \frac{\mu}{\lambda + \mu} \leq a(1 - \frac{\mu}{\lambda + \mu})$

$\Rightarrow p = \frac{\mu}{\lambda + \mu} \leq \frac{a}{a + 1}$. So the hypothesis $H : \mu \leq a\lambda$ becomes $H : p \leq \frac{a}{a + 1}$ and is rejected when $Y$ is too large.
The constants of the test are determined from tables of the binomial distribution.

Similarly we may compare two binomial distributions. Suppose $X$ and $Y$ are independent binomial variables with joint distribution

$$P(X = x, Y = y) = \binom{m}{x} p_1^x q_1^{m-x} \binom{n}{y} p_2^y q_2^{n-y}$$

which may be written as

$$P(X = x, Y = y) = \binom{m}{x} \binom{n}{y} p_1^x q_1^{m-x} q_2^y \exp[y \log \frac{p_2/q_2}{p_1/q_1} + (x+y) \log \frac{p_1}{q_1}]$$

Comparing (47) with (35) we see that $U = Y$, $T = X + Y$ and the hypotheses can be tested concerning

$$\theta = \log \frac{p_2/q_2}{p_1/q_1} \text{ or } \varphi = \frac{p_2/q_2}{p_1/q_2}.$$

In particular

$$p_2 \leq p_1 \Rightarrow q_1 \leq q_2 \Rightarrow \frac{p_2 q_1}{p_1 q_2} \leq 1 \Rightarrow \log \frac{p_2/q_2}{p_1/q_1} \leq 0$$

so the hypotheses include $H_1': p_2 \leq p_1$ against $p_2 > p_1$ and $H_4': p_2 = p_1$ against $p_2 \neq p_1$. As before we consider

$$P(Y = y | X + Y = t) = \frac{P(X+Y = t | Y = y) P(Y = y)}{P(X+Y = t)}$$

$$= \frac{P(X = t-y) P(Y = y)}{\sum_{y'=0}^t P(Y = y' \text{ and } X+Y = t)}$$

The numerator is

$$\binom{m}{t-y} p_1^{t-y} q_1^{m-(t-y)} \binom{n}{y} p_2^y q_2^{n-y}$$

$$= \binom{m}{t-y} p_1^{t} q_1^{m-(t-y)} \binom{n}{y} \left[ \frac{p_2/q_2}{p_1/q_1} \right]^y$$
and the denominator becomes \( \left( \frac{p_1}{q_1} \right)^{m} \left( \frac{q_2}{p_2} \right)^{n} \sum_{y'=0}^{y} \binom{m}{t-y'} \binom{n}{y'} \left[ \frac{p_2/q_2}{p_1/q_1} \right]^{y'} \).

Hence, \( P(Y=y|X+Y=t) = \frac{\binom{m}{t} \binom{n}{y} \rho^{y}}{\sum_{y'=0}^{t} \binom{m}{t-y'} \binom{n}{y'} \rho^{y'}} \).

For \( H_1 \) and \( H_4 \), the boundary value of (38), (43) and (44) is \( \theta_0 = 0 \) so that \( \rho = 1 \). Since \( \sum_{y=0}^{t} \binom{m}{t-y} \binom{n}{y} = \binom{m+n}{t} \), the conditional distribution becomes

\[
P(Y=y|X+Y=t) = \frac{\binom{m}{t} \binom{n}{y}}{\binom{m+n}{t}}, \quad y = 0, \ldots, t,
\]

which is the hypergeometric distribution. The tests are then carried out by means of appropriate tables. This test is actually R.A. Fisher's exact test for two by two tables which was arrived at largely through intuition. Applying the theory we have developed to this problem we see that his test for two by two tables was quite reasonable. The reader may find the details in section 21.02 of [1].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
REFERENCES

