國立交通大學

統計學研究所

碩士論文

高密度顯著性檢定以複合式假設為例

Highest Density Significance Test for Composite Hypothesis

研 究 生：曾義家

指導教授：陳鄰安　教授

中華民國九十五年六月

高密度顯著性檢定以複合式假設為例

Highest Density Significance Test for Composite Hypothesis

研 究 生：曾義家　　Student : Yi-Chia Tseng
指導教授：陳鄰安　　Advisor : Dr. Lin-An Chen

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis

Submitted to Institute of Statistics

College of Science

Nation Chiao-Tung University

in partial Fulfillment of the Requirements

for the degree of Master

in

Statistics

June 2006

Hsinchu, Taiwan

中華民國九十五年六月

高密度顯著性檢定以複合式假設爲例

學　生：曾義家　　　　　　　　　　　　　　指　導　教　授：陳鄰安

國立交通大學統計學研究所　碩士班

摘　　要

延伸 Chen(2005)的想法，我們提出針對複合式虛無假設的高密度顯著性檢定。且推導出這個檢定的存在性以及它的最佳化的性質，並且針對常態分配的參數，將此方法應用於其上。對於一些較爲複雜的高密度顯著性檢定問題，我們利用近似性的高密度顯著性檢定加以解決。

Highest Density Significance Test for Composite Hypothesis

student : Yi-Chia Tseng   Advisors : Dr. Lin An Chen

Institute of Statistics

National Chiao Tung University

ABSTRACT

Extending from the idea of Chen(2005), we proposed highest density significance (HDS) test for composite null hypothesis. Existence and optimality of this test are derived. Examples of HDS test for normal parameters are provided. For problems that HDS tests are complicated to derived, we propose the approximate HDS tests.

誌　　　　謝

時光飛似，轉眼間研究所求學階段即將結束，謝謝師長們的教導以及同學們的陪伴，讓我經歷了兩年愉快且充實的碩士生涯。

首先，我要感謝我的指導老師 陳鄰安教授，謝謝老師在忙碌之餘還能撥冗教導我，不厭其煩的指導我在研究時所遇到的問題，使我能如期完成論文。還要感謝博士班　陳弘家學長，謝謝學長在我有困惑時為我解惑，使我釐清學長論文中的觀念及問題。還要謝謝口試委員對我論文的指導與建議。當然還要感謝研究室的同學給予我協助，以及閒暇之餘一起運動娛樂，感謝你們豐富了這兩年的生活，留下美好的回憶。

最後，要感激父母及家人這麼多年來的栽培與支持，你們的關心與照顧，我才能在毫無顧慮下順利完成學業，在此，將以本篇論文獻給曾經給我鼓勵協助的家人、師長、朋友以及同學們，並致上我最誠摯的謝意。

義家　　　謹誌于

國立交通大學統計學研究所
中華民國九十五年六月

Contents

Highest Density Significance Test for Composite Hypothesis

Abstract

Extending from the idea of Chen (2005), we proposed highest density significance (HDS) test for composite null hypothesis. Existence and optimality of this test are derived. Examples of HDS test for normal parameters are provided. For problems that HDS tests are complicated to derive, we propose the approximate HDS tests.

Key words: Fisherian significance test; Hypothesis testing; significance test;

1. Introduction

In the hypothesis testing, there are two important categories of hypothesis specification, the siginificance test and the Neyman-Pearson formulation. The Neyman-Pearson formulation considers a decision problem that we want to choose one from the null hypothesis H_0 and an alternative hypothesis H_1. On the other hand, the significance test considers only one hypothesis, the null hypothesis H_0. The significance test may occurs that H_0 is drawn from a scientific guess and we are vague about the alternative, and cannot easily parameterize them. Another case is that the model when H_0 is true is developed by a selection process on a subset and is to be checked with new data. Then the problem for significance test is more general than the Neyman-Pearson formulation in that when H_0 is not true there are many possibilities for the true alternative.

Over 200 years of development of significance test, it has been used in many branches of applied sicences. Some earliest use of significance test include that, for examples, Armitage (1983) claims to have found the germ of the idea in a medical discussion from 1662 and Arbuthnot (1710) observed that the male births exceeded female births in Lonton for each of the past 82 years that violates the assumption of equal chance of male birth. Some important significance tests latter developed include the Karl Pearson's (1900) chi-squares test and W. S. Gosset's (1908) student paper that proposed the first solution to the problem of small-sample tests. Significance tests were given their modern justification and then popularized by Fisher that he derived most of the test statistics that we now use, in a series of papers and books during 1920s and 1930s. Traditionally the siginificance test is to examine how to decide whether or not a given set of data is consistent with H_0 it is conducted through several key steps:

1. stimulating a suitable null hypothesis H_0, 2. choosing a test statistic T to rank possible experimental outcomes, 3. determining the p-value which is the probability of the set of values of T at least as extreme as the value observed when H_0 is true. 4. H_0 being accepted (rejected) if the p-value is large (small) enough. This classical way that the user have to select a test statistic, although often been recommended a sufficient statistic, to formulate the test is generally called the Fisherian significance test since R. A. Fisher made the contribution on significance test the most.

Yes, over 200 years development, however, the concept and theory of Fisherian significance test are now scarcely introduced in modern texts of statistical inferences. There are some reasons reflecting this fact. (a) The Fisherian significance tests are questioned about how to use test statistics and which test statistic to use under which circumstances. For one specific hypothesis, there may have several test statistics for use such as the Pearson's (1900) chi-square test and the normal approximation method. With this difficulty, people may be bothered for making decision when it is happened that the null hypothesis H_0 is rejected by using one test statistic inducing strong evidence against H_0 and accepted by using another test statistic inducing no real evidence against H_0. (b) On the other hand, there is also the problem of choosing the one sided Fisherian significance test or the two sided Fisherian significance test. In practice, it often depends only on practicer's convenience. It is quite common that if the distribution of the test statistic is available and is symmetric then a two sided Fisherian significance test is often implemented and a one sided Fisherian significance test is implemented when it has an asymmetric distribution. However, if approximation for the distribution of the test statistic such as the normal aprroximation and the Pearsone's chi-square test has been used then they usually implement a two sided one. For example, for testing $H_0 : p = p_0$ with X obeying a binomial distribution, the Fisherian significance test may be conducted by the two sided Pearson's chi-square and normal approximation tests and by choosing X as a test statistic for a one sided test (see the latter one in Garthwaite et al. (2002)). For one hypothesis problem, a two sided Fisherian significance test may have p-value approximated twice as it for an one sided one. Without a fair justification in deciding a one sided or two sided test, the conclusion based on p-value may be misleading. (c) There is absence of a direct indication of any departure from H_0. This happens mainly because that Fisherian significance tests generall are constructed based on the sufficient statistic of the parameter assumed in null hypothesis. (d) There is lack of desired optimal property as a justification to support the calssically used Fisherian significance tests. In the late 1920s, E. S. Pearson, son of Karl Pearson, approached Jerzy Neyman with a question that bother him. If you test whether data fit a particular probability distribution and the test statistic is not large enough to reject the distribution, how do you know that this is the best that you could done? How do you know that some other test statistics might not have rejected that probability distribution? The resulting collaboration (see this in a series of papers collected in Neyman and Pearson (1967)) between them produced the Neyman-Pearson formulation that requires tests selecting one between the null hypothesis H_0 and a well specified alternative hypothesis H_1. With Neyman-Pearson formulation, a mathematical theory of hypothesis testing in which tests are derived as solutions of clearly stated optimum problems has been developed. The optimal property does made the Neyman-Pearson theory popular in statistical inferences.

In this paper, we consider the hypothesis $H_0 : \theta = \theta_0$. By letting $X = x_0$ be the observed sample, our approach of HDS test sets all x's satisfying $L(x, \theta_0) \leq L(x_0, \theta_0)$ as set of extreme points for computing p-value. This setting of determining the extreme points not only gets rid off the difficulty of deciding a one sided or two sided test but also automatically determine the test statistic. The defects (a) and (b) ocuured for the Fisherian significance test are then solved. Traditionally the chosen test statistic used in Fisherian significance test involves information only related to the parameter considered in H_0 which often is either a location or scale parameter. But, the likelihood based HDS test classifies a point x if it is an extreme by measuring its corresponding likelihood using all information involving the size of its likelihood. This often result in a test statistic that involve information with amount bigger than that contained in the traditionally used test statistic. Therefore, we may expect that likelihood function for constructing a significance test definitely will gain some more advantages. First, we will show that this new significance test has a desired property of optimality which then does provide a justification for the use of this new test. The defect (d) for Fisherian significance test is also overcomed. Second, often using extra information often provides indication of departure from H_0. Furthermore, examples of HDS test for both continuous distribution and discrete distribution will be provided associated with discussion of comparison with the Fisherian significance tests.

2. Motivation for Highest Density Significance Test

Let $X_1, ..., X_n$ be a random sample drawn from a distribution having a probability density function (pdf) $f(x, \theta)$ with parameter space Ω. Consider the simple hypothesis $H_0 : \theta = \theta_0$ for some $\theta_0 \in \Omega$. By letting vector X with $X' = (X_1, ..., X_n)$ and sample space Λ, let's denote the join pdf of X as $L(x, \theta)$, also called it the likelihood function.

What is generally done in classical approach for significance test when $X = x_0$ is observed, particularly influenced by R. A. Fisher and being called the Fisherian significance test, is to determine the extreme set based on the distribution (or approximated distribution) of a test statistic. With a test statistic $T = t(X)$, it then define the p-value as

$$p_{x_0} = P_{\theta_0}(T \text{ at least as extreme as the value observed}). \qquad (2.1)$$

While this approach is applicable in certain practical problems, it is scarcely of sufficient generality to warrant trying to find necessary and sufficient conditions for its applicability. There are several reasons for this point. First, as questioned from E. P. Perason that, for given one observation x_0, we may reject H_0 for having small p-value that provides evidence against H_0 with one test statistic but accept H_0 for having large p-value that provides no real evidence against H_0 with another test statistic. How can we decide to choose the test statistic? Second, the extreme set E in (2.1) varies in choosing the one sided or two sided Fisherian significance test. However, in application, it often is decided based on practicer's convenience where we should know that the two sided Fisherian significance test may have p-value as large as twice the p-value of the one sided Fisherian significance test. This increases the difficulty in understanding the $p$value. Third, so far, there is no justification of desired optimal property for this test-statistic based Fisherian significance test.

Suppose that we set out to order points in the sample space Λ according to the amount of evidence they provide for $H_0 : \theta = \theta_0$. We should naturally order them according to the value of the probability $L(x, \theta_0)$; any x with small $L(x, \theta_0)$ revealing evidence against H_0. Then, when $X = x_0$ is observed and if we must choose subset of possible observations which indicates that H_0 is true, then it seems sensible to put into this subset those x's for which the probability $L(x, \theta_0)$ is large - in other words to choose a subset of the form $\{x : L(x, \theta_0) > L(x_0, \theta_0)\}$. On the other hand, the subset which indicates that H_0 is not true seems sensible to be of the form $\{x : L(x, \theta_0) \leq L(x_0, \theta_0)\}$.

With expectation that each extreme point has to be at least as extreme as the observed value x_0, this leads the following definition for defining a new type of significance test.

Consider the null hypothesis $H_0 : \theta = \theta_0$. The HDS test proposed by Chen defines the p-value as

$$p_{hd} = \int_{L(x,\theta_0)\leq L(x_0,\theta_0)} L(x,\theta_0)dx.$$

The method of highest density for significance test is obviously appealing for the follwoing facts:

(1) Fisherian significance test chooses a test statistic $T = t(X)$ that gives an ordering of the sample points as evidence against H_0: $t(x_1) > t(x_2)$ means that x_1 is stronger than x_2 as evidence against H_0. This evidence may varies in the chosen test statistic. On the other hand, the ordering based on HDS test means that x_1 is stronger than x_2 as evidence against H_0 when it occurs $L(x_1,\theta_0) < L(x_2,\theta_0)$.

(2) The inequality $L(x,\theta_0) \leq L(x_0,\theta_0)$ automatically decide two desirabilities: the test statistic involving in the test and setting if it is a one sided or two sided test. This solves the defects (a) and (b) occurred in Fisherian significance test.

(3) For this test, we set the non-extreme set including the points x's in sample space the highest part of the joint density function $L(x,\theta_0)$; such a set includes relatively more probable points when H_0 is true. On the other hand, this extreme set $E_{hd} = \{x : L(x,\theta_0) < L(x_0,\theta_0)\}$ of the HDS test is weirder than its corresponding non-extreme set $\tilde{E}_{hd} = \{x : L(x,\theta_0) \geq L(x_0,\theta_0))\}$ in the sense that $L(x_1,\theta_0) < L(x_2,\theta_0)$ for $x_1 \in E_{hd}$ and $x_2 \in \tilde{E}_{hd}$. None of the traditional Fisherian significance tests has this appealing in determining the extreme set for computing p-value of given observation x.

Definition 2.1. Consider the null hypothesis $H_0 : \theta \in \Theta_0$. The HDS test defines the p-value as

$$p_{hd} = \sup_{\theta\in\Theta_0} \int_{L(x,\theta)\leq L(x_0,\theta)} L(x,\theta)dx. \tag{2.2}$$

We classify the hypothesis testing problems whose level α MP, UMP, or UMPU tests are proposed in the literature into the following three categories:

(A) $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ and $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$

(B) $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ versus $H_1 : \theta_1 < \theta < \theta_2$, where both θ_1 and θ_2 are known real-valued constants with $\theta_1 < \theta_2$

(C) $H_0 : \theta_1 \leq \theta \leq \theta_2$ versus $H_1 : \theta < \theta_1$ or $\theta > \theta_2$, where both θ_1 and θ_2 are known real-valued constants with $\theta_1 \leq \theta_2$

(1) From any family of distributions with the monotone likelihood ratio property, there exists a level α UMP test for any hypothesis testing problem belonging to Category A. See, e.g., Lehmann (1986, Theorem 2 in Chapter 3).

(2) From any one-parameter exponential family, there exists a level α UMP test for any hypothesis testing problem belonging to Category B. See, e.g., Lehmann (1986, Theorem 6 in Chapter 3).

(3) From any one-parameter exponential family, there exists a level α UMPU test for any hypothesis testing problem belonging to Category C. See, e.g., Lehmann (1986, Section 4.2).

Theorem 2.2. Let $X = (X_1, ..., X_n)$ be a random sample from $f(x, \theta)$ with an observation $X = x$. Consider the hypothesis $H_0 : \theta = \theta \in \Theta_0$ and we assume that the family of densities $\{f(x, \theta) : \theta \in \Theta\}$ has a monotone likelihood in the statistic $T = t(X)$:
(a) If the monotone likelihood is nondecreasing in a function $t(x, \theta)$, then the test with p-value

$$p_{hd} = \sup_{\theta \in \Theta_0} P_\theta(t(X, \theta) \leq t(x_0, \theta))$$

is a HDS test.
(b) If the monotone likelihood is nonincreasing in $t(x)$, then the test with p-value

$$p_{hd} = \sup_{\theta \in \Theta_0} P_\theta(t(X, \theta) \geq t(x_0, , \theta))$$

is a HDS test.

3. Existence and Optimality

Theorem 3.1. Let $X_1, ..., X_n$ be a random sample drawn from a distribution with pdf f. Suppose that there exists a partition, $A_0, A_1, ..., A_k$, of the sample space of random variable X such that $P_{\theta_0}(X \in A_0) = 0$ and f is continuous on each A_i. Further, suppose there exist functions $f_1, ..., f_k$, defined on $A_1, ..., A_k$, respectively, satisfying
(i) $f(x) = f_i(x)$, for $x \in A_i$,
(ii) f_i is monotone on A_i,
(iii) the set $B = \{y : y = f_i(x)$ for some $x \in A_i\}$ is the same for each $i = 1, ..., k$, and

(iv) f_i^{-1} has a continuous derivative on B, for each $i = 1, ..., k$.
Then, HDS test exists.
Proof. Conditions (i)-(iv) are set for variable transformation of continuous random variable which provides a continuous pdf of new variable $f(X_i, \theta_0)$ (see, for example Cassella and Berger (p53)). As the fact that $L(X_1, ..., X_n, \theta_0)$ is a product of $f(X_i, \theta_0), i = 1, ..., n$, then $P(L(X_1, ..., X_n, \theta_0) \leq a)$ is a continuous and monotone increasing function of a. Then, from the intermediate theorem,

$$P(L(X, \theta) \leq L(x_0, \theta)) \tag{3.1}$$

exists for $\theta \in \theta_0$. The result is followed from the fact that the set of values of (3.1) with $\theta \in \theta_0$ is bounded which indicating the existence of infimum. □

There are three remarks for this theorem:
(a) Most continuous type distributions appeared in literature fulfill the conditions (i)-(iv) in the theorem and then the HDS tests exist for any level α. Example that this theorem does not hold includes the uniform distribution.
(b) (i)-(iv) provide only a sufficien conditions and this set is definitely not necessary conditions and then the HDS tests for any level α exist in a wider famimly of distributions.
(c) We can ignore the exceptional set A_0 since $P_{\theta_0}(X \in A_0) = 0$. It is a technical device that is used to handle endpoints of intervals. Example for this case includes the double exponential distribution.

Most distributions of continuous type we have seen in literature are having pdf's of monotone cases or unimodels and they fullfill the conditions (i)-(iv) in Theorem 3.1 such that the HDS tests exist where the unimodels have ranges of the form $(0, x_{mod})$ where x_{mod} is the mode of the distribution.

From Theorem 3.1, the level α HDS test has acceptance region $\{(x_1, ..., x_n) : L(x_1, ..., x_n, \theta_0) \geq a_\alpha\}$ where a_α satisfies $1 - \alpha = P_{\theta_0}(L(X_1, ..., X_n, \theta_0) \geq a_\alpha))$. The key of deriving the HDS test based on this theorem is that we need to know the distribution of the random form of the likelihood function under the assumption that $H_0 : \theta = \theta_0$ is true.

Theorem 3.2. Consider the hypothesis $H_0 : \theta \in \Theta_0$ where θ may be a vector parameters and suppose that the observation of the sample be x_0. For any significance test with set of non-extreme points $B(\theta)$, then for any $\theta \in \Theta_0$ such that

$$\int_{L(x,\theta)\geq L(x_0,\theta)} L(x,\theta)dx = \int_{B(\theta)} L(x,\theta)dx, \tag{3.2}$$

then we have

$$\text{volume}(\{x : L(x,\theta) \geq L(x_0,\theta)\}) \leq \text{volume}(B(\theta)). \tag{3.3}$$

Proof. Suppose taht (3.2) holds. Deleting the subset common to $\{x : L(x,\theta) \geq L(x_0,\theta)\}$ and $B(\theta)$ yields

$$\int_{\{x:L(x,\theta)\geq L(x_0,\theta)\}\cap B(\theta)^c} L(x,\theta)dx = \int_{B(\theta)\cap\{x:L(x,\theta)\geq L(x_0,\theta)\}^c} L(x,\theta)dx.$$

Now, for $x_a \in \{x : L(x,\theta) \geq L(x_0,\theta)\} \cap B(\theta)^c$ and $x_b \in B(\theta) \cap \{x : L(x,\theta) \geq L(x_0,\theta)\}$, we have $L(x_a,\theta) > L(x_b,\theta)$. Thus,

$$\begin{aligned}&\text{volume}(\{x : L(x,\theta) \geq L(x_0,\theta) \text{ and } x \in B(\theta)^c\})\\ &\leq \text{volume}(\{x : L(x,\theta) < L(x_0,\theta) \text{ and } x \in B(\theta)\}).\end{aligned} \tag{3.4}$$

So, adding the volume of $\{x : L(x,\theta) \geq L(x_0,\theta)\} \cap B(\theta)$ to both sides of (3.4), we have the theorem. □

4. Testing Hypothesis for Normal Parameters

Suppose that we are willing to accept as a fact that the outcome of the variable of a random experiment has a normal distribution with mean μ and variance σ^2. Now, let $X_1, ..., X_n$ be a random sample drawn from normal distribution $N(\mu, \sigma^2)$. We first consider the hypothesis about mean μ with assuming that σ is known.

Case 1: $\sigma^2 = \sigma_0$ is known constant. Consider the null hypothesis $H_0 : \mu \in \Theta_\mu$. The likelihood function is

$$L(x,\mu) = (2\pi\sigma_0^2)^{-n/2} e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma_0^2}}.$$

This is a monotone decreasing function in $\sum_{i=1}^n (x_i-\mu)^2$. With the fact that $L(X,\mu) \leq L(x_0,\mu)$ if and only if $\sum_{i=1}^n (X_i-\mu)^2 \geq \sum_{i=1}^n (x_i-\mu)^2$ and $\sum_{i=1}^n (x_i-\mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x}-\mu)^2$, the p-value for H_0 is

$$p_{hd} = \sup_{\Theta_\mu} P(\chi^2(n) \geq \frac{1}{\sigma_0^2}[\sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2]).$$

This p-value varies in observation $x_0 = (x_1, ..., x_n)$ and assumption on μ. We list the corresponding p-values associated with several hypotheses in Tables 1 and 2.

Table 1. p-values for some hypotheses H_0 about normal mean

Hypothesis	p-value p_{hd}
$H_0 : \mu \leq \mu_0$	$\begin{cases} P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2}) & \text{if } \bar{x} \leq \mu_0 \\ P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma_0^2}) & \text{if } \bar{x} > \mu_0 \end{cases}$
$H_0 : \mu \geq \mu_0$	$\begin{cases} P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma_0^2}) & \text{if } \bar{x} < \mu_0 \\ P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2}) & \text{if } \bar{x} \geq \mu_0 \end{cases}$
$H_0 : \mu_0 \leq \mu \leq \mu_1$	$\begin{cases} P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma_0^2}) & \text{if } \bar{x} < \mu_0 \\ P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} & \text{if } \mu_0 \leq \mu \leq \mu_1 \\ P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sigma_0^2}) & \text{if } \bar{x} > \mu_1 \end{cases}$

Example 1. In a semiconductor manufacturering process CVD metal thickness was measured on 30 wafers obtained over approximately 2 weeks. A data of this experiment has been given in Montgomery, Runger and Hubele (2004) where $\sigma = 1$ is assumed to be known and we have $\bar{x} = 15.99$ and $\sum_{i=1}^n (x_i - \bar{x})^2 = 29.107$. We have the following tests:

(a) If we test $H_0 : 15 \leq \mu \leq 17$, the p-value is

$$p_{hd} = \sup_{(13,19)} P(\chi^2(30) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}) = 0.5119.$$

(b) If we test $H_0 : 17 \leq \mu \leq 19$, the p-value is

$$p_{hd} = \sup_{(17,19)} P(\chi^2(30) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} + \frac{30(\bar{x} - 17)^2}{\sigma^2}) = 0.00099.$$

(c) If we test $H_0 : 13 \leq \mu \leq 15$, the p-value is

$$p_{hd} = \sup_{(13,15)} P(\chi^2(30) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} + \frac{30(\bar{x} - 15)^2}{\sigma^2}) = 0.00139.$$

Table 2. p-values for some hypotheses H_0 about normal mean

Hypothesis	p-value p_{hd}
$H_0 : \mu \in (\mu_0, \mu_1) \cup (\mu_2, \mu_3)$	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} + \frac{n(\bar{x}-\mu_0)^2}{\sigma_0^2})$ if $\bar{x} \leq \mu_0$ $P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2})$ if $\mu_0 \leq \bar{x} \leq \mu_1$ or $\mu_2 \leq \bar{x} \leq \mu_3$ $P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} + \frac{n(\bar{x}-\mu_1)^2}{\sigma_0^2})$ if $\mu_1 \leq \bar{x} \leq \mu_2$ with $\lvert\bar{x} - \mu_1\rvert \leq \lvert\bar{x} - \mu_2\rvert$ $P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} + \frac{n(\bar{x}-\mu_2)^2}{\sigma_0^2})$ if $\mu_1 \leq \bar{x} \leq \mu_2$ with $\lvert\bar{x} - \mu_1\rvert > \lvert\bar{x} - \mu_2\rvert$ $P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} + \frac{n(\bar{x}-\mu_3)^2}{\sigma_0^2})$ if $\bar{x} \geq \mu_3$

Table 3. p-values for some hypotheses H_0 about normal mean

Hypothesis	p-value p_{hd}
$H_0 : \mu \leq \mu_0$ or $\mu \geq \mu_1$	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} + \frac{n(\bar{x}-\mu_0)^2}{\sigma_0^2})$ if $\bar{x} \leq \mu_0$ $P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} + \frac{n(\bar{x}-\mu_0)^2}{\sigma_0^2})$ if $\mu_0 \leq \bar{x} \leq \mu_1$ with $\lvert\bar{x} - \mu_0\rvert \leq \lvert\bar{x} - \mu_1\rvert$ $P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} + \frac{n(\bar{x}-\mu_1)^2}{\sigma_0^2})$ if $\mu_0 \leq \bar{x} \leq \mu_1$ with $\lvert\bar{x} - \mu_0\rvert \geq \lvert\bar{x} - \mu_1\rvert$ $P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} + \frac{n(\bar{x}-\mu_1)^2}{\sigma_0^2})$ if $\bar{x} \geq \mu_1$

5. Testing Hypothesis For Standard Deviation and For Both Mean and Standard Deviation

Case 2: $\mu = \mu_0$ is known.

Table 4. p-values for some hypotheses H_0 about normal variance

Hypothesis	p-value p_{hd}
$H_0 : \sigma^2 \leq \sigma_0^2$	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma_0^2})$
$H_0 : \sigma_0^2 \leq \sigma^2 \leq \sigma_1^2$	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma_1^2})$

Table 5. p-values for some hypotheses H_0 about normal mean and variance

Hypothesis	p-value p_{hd}
$H_0 : \mu \leq \mu_0, \sigma_0^2 \leq \sigma^2 \leq \sigma_1^2$	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_1^2})$ if $\bar{x} < \mu_0$
	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma_1^2})$ if $\bar{x} \geq \mu_0$
$H_0 : \mu_0 \leq \mu \leq \mu_1, \sigma_0^2 \leq \sigma^2 \leq \sigma_1^2$	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma_1^2})$ if $\bar{x} < \mu_0$
	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_1^2})$ if $\mu_0 \leq \bar{x} \leq \mu_1$
	$P(\chi^2(n) \geq \frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sigma_1^2})$ if $\bar{x} > \mu_1$

6. Some Other Distributions

Let's consider the testing hypothesis about the parameter in the following exponetial distribution

$$f(x, \theta) = \theta e^{-\theta x}, x > 0.$$

Table 5. p-values for some hypotheses H_0 about exponential parameter

Hypothesis	p-value p_{hd}
$H_0 : \theta \leq \theta_0$	$P(Gamma(n, 1) \geq \frac{n\bar{x}}{\theta_0})$
$H_0 : \theta \in (\theta_0, \theta_1) \cup (\theta_2, \theta_3)$	$P(Gamma(n, 1) \geq \frac{n\bar{x}}{\theta_3})$

7. Approximate Highest Density Significance Test

In statistical inferences for some distributions, approximate techniques are often desirable. Sometimes there are due to that direct evaluation of exact statistical inference is overwhemingly difficult and sometimes the approximations are cheaper and quiker. For example, consider that we have a random sample $X_1, ..., X_n$ drawn from the Gamma distribution with pdf

$$f(x, \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, x > 0$$

where $\alpha, \beta > 0$ are parameters. We further assume that α is known constant and we want to test the composite null hypothesis $H_0 : \beta \in \Theta_\beta$. With this case the p-value for the HDS test is

$$p_{hd} = \sup_{\beta \in \Theta_\beta} P_\beta \{(\pi_{i=1}^n X_i)^{\alpha-1} e^{-\frac{\sum_{i=1}^n X_i}{\beta}} \leq (\pi_{i=1}^n x_i)^{\alpha-1} e^{-\frac{\sum_{i=1}^n x_i}{\beta}}\}. \tag{6.1}$$

The p-value of (6.1) may be derived only if we have an explicit distribution of the statistic $\pi_{i=1}^n X_i)^{\alpha-1} e^{-\frac{\sum_{i=1}^n X_i}{\beta}}$ under the distribution P_β. Howewver, this is complicated to derive it. Then, an approximation technique for this topic of highest density significance tests is needed.

The p-value of (2.2) for a highest density significance test may be reformulated as

$$\begin{aligned} p_{hd} &= \sup_{\theta \in \Theta_0} P\{\pi_{i=1}^n f(X_i, \theta) \leq \pi_{i=1}^n f(x_i, \theta)\} \\ &= \sup_{\theta \in \Theta_0} P\{\sum_{i=1}^n lnf(X_i, \theta) \leq \sum_{i=1}^n lnf(x_i, \theta)\}. \end{aligned} \tag{6.2}$$

The asymptotic theory may be applied on the statistic $\sum_{i=1}^n lnf(X_i, \theta)$ of (6.2) when θ is true for this test due the fact that $lnf(X_i, \theta), i = 1, ..., n$ are independent and identically distributed with further assumptions that its mean $E_\theta[ln\ f(X, \theta)]$ and variance $Var_\theta[ln\ f(X, \theta)]$ exist. With the central limit theorem, the p-value of an approximate highest density significance test is

$$p_{hd.app} = \sup_{\theta \in \Theta_0} \Phi(\frac{n^{-1} \sum_{i=1}^n ln\ f(x_i, \theta) - E_\theta[ln\ f(X, \theta)]}{\sqrt{n^{-1} Var_\theta[ln\ f(X, \theta)]}}) \tag{6.3}$$

where Φ is the distribution function of the standard normal distribution.

It is interesting to see if the approximate highest density significance test is appropriate to use when the exact one is not available. We consider the appropriateness based on the efficiencies of the approximation technique. Let's consider a simulation when the underlying distribution is normal as an example. Suppose that now we have a random sample $X_1, ..., X_n$ from normal distribution $N(\mu, \sigma^2)$ and we consider hypothesis $H_0 : (\mu, \sigma) \in \Theta_0$. To formulate the test in (6.3), the logarithm of normal pdf is

$$ln\ f(x, \mu, \sigma) = -\frac{1}{2} ln\ (2\pi\sigma^2) - \frac{(x-\mu)^2}{2\pi\sigma^2}$$

which implies that $E[ln\ f(X, \theta)] = -\frac{1}{2} ln(2\pi\sigma^2) - \frac{1}{2}$ and $Var[ln\ f(X, \theta)] = \frac{1}{2}$. With some arrangements, we have that the approximate p-value under the normal distribution is

$$p_{hd.app}^{N(\mu,\sigma^2)} = \sup_{(\mu,\sigma)\in\Theta_0} \Phi((2n)^{-1/2} \sum_{i=1}^{n} [1 - \frac{(x_i-\mu)^2}{\sigma^2}]). \quad (6.4)$$

For studying the efficiencies of the approximate p-value of formula (6.4), we further assume that σ is known to be value 1. Then the approximate p-value is reduced to

$$p_{hd.app}^{N(\mu,\sigma^2)} = \sup_{\mu\in\Theta_{mu}} \Phi((2n)^{-1/2} \sum_{i=1}^{n} [1 - \frac{(x_i-\mu)^2}{\sigma_0^2}]). \quad (6.5)$$

We generate random sample of size n from normal distribution $N(\mu, 1)$ and, from this sample, we compute p-values from Table 1 and (6.5). This simulation is done with replication 100,000 and we compute the average p-values of the exact one and approximate one. The following table display these results.

Table 6. p-values for exact and approximate highest density significance tests when H_0 is not true

Sample size	Exact test	Appro. test	Exact test	Appro. test
	$H_0 : \mu \in (-2, -1)$		$H_0 : \mu \in (-1, 1)$	
$n = 10$	$2.47E-09$	$1.35E-17$	0.1216	0.1243
$n = 20$	$2.62E-18$	$7.39E-50$	0.0492	0.0467
$n = 30$	$6.60E-28$	$3.48E-92$	0.0204	0.0179
$n = 50$	$1.44E-42$	$1.70E-149$	0.0040	0.0031
$n = 100$	$3.29E-112$	0.0000	0.0000	0.0000
	$H_0 : \mu \in (3, 4)$		$H_0 : \mu \in (5, 6)$	
$n = 10$	0.1202	0.1227	$4.76E-09$	$2.52E-15$
$n = 20$	0.0479	0.0453	$1.29E-18$	$6.60E-53$
$n = 30$	0.0208	0.0182	$5.03E-30$	$4.33E-105$
$n = 50$	0.0044	0.0034	$3.23E-51$	$2.29E-206$
$n = 100$	0.0001	0.0000	$2.71E-110$	0.0000

Table 7. p-values for exact and approximate highest density significance tests when H_0 is true

Sample size	Exact test	Appro. test	Exact test	Appro. test
	$H_0 : \mu \in (1,3)$		$H_0 : \mu \in (0,4)$	
$n = 10$	0.5688	0.5864	0.5666	0.5833
$n = 20$	0.5473	0.5606	0.5477	0.5609
$n = 30$	0.5347	0.5460	0.5381	0.5492
$n = 50$	0.5305	0.5398	0.5246	0.5337
$n = 100$	0.5157	0.5224	0.5214	0.5280

In the application of approximate highest density significance test, we first consider the approximate p-value of (6.1) with random sample from Gamma distribution.

Theorem 6.1. The p-value of an approximate highest density significance test for the Gamma distribution is

$$p^{Gamma}_{hd,app} = \sup_{\beta \in \Theta_\beta} \Phi\Big(\frac{(\alpha-1)[n^{-1}\sum_{i=1}^n lnx_i - (ln(\beta) + PG[0,\alpha])] - (\bar{x}/\beta - \alpha)}{\sqrt{n^{-1}Var[ln\ f(X,\beta)]}}\Big)$$

where $PG[n,z]$ is the n^{th} derivative of the digamma function $\phi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ and

$$\begin{aligned} Var[ln\ f(X,\beta)] &= \alpha[3(ln\ \beta)^2 + 6ln\ \beta PG[0,\alpha] + 3(PG[0,\alpha])^2 + PG[1,\alpha]] \\ &- 2(PG[0,\alpha])^2 - 2(ln\ \beta)^2 - PG[1,\alpha] - 2\alpha(\alpha-1)PG[0,\alpha+1] \\ &- 2\alpha(\alpha-1)(ln\ \beta - 1)PG[0,\alpha] + \alpha(\alpha+1)\beta^{\alpha-1} - (\alpha ln\ \beta)^2 + 2\alpha^2 PG[0,\alpha]. \end{aligned}$$

Let's consider a case for simulation. Suppose that we have known $\alpha = 2$. In this situation, we see that $PG[0,2] = 1-\gamma, PG[1,2] = \frac{\pi^2}{6} - 1, PG[0,3] = \frac{3}{2} - \gamma$ indicating that $E[ln\ f(X,\beta)] = -1-\gamma - ln\ \beta$ and $Var[ln\ f(X,\beta)] = \frac{\pi^2}{6} - 1$. Then, the p-value of the approximate highest density significance test is

$$p^{Gamma}_{hd,app} = \sup_{\beta \in \Theta_\beta} \Phi\Big(\frac{n^{-1}\sum_{i=1}^n lnx_i - \bar{x}/\beta - ln\ \beta + 1 + \gamma}{\sqrt{(\pi^2/6 - 1)/n}}\Big). \qquad (6.6)$$

For comparison, we also consider the following two Fisherian significance tests with p-values

$$p_{cla,A} = \sup_{\beta \in \Theta_\beta} P\{Gamma(2n,1) \geq \frac{\sum_{i=1}^n x_i}{\beta}\}$$

$$\text{and } _{cla,B} = \sup_{\beta \in \Theta_\beta} P\{Gamma(2n,1) \leq \frac{\sum_{i=1}^n x_i}{\beta}\}.$$

The following table displays the simulation results.

Table 7. p-values for exact and approximate highest density significance tests for $H_0 : 1.5 \leq \beta \leq 2.5$

Sample size	$p_{hd,app}^{Gamma}$	$p_{cla,A}$	$p_{cla,B}$
True $\beta = 2$			
$n = 10$	0.7255	0.7547	0.8297
$n = 20$	0.7912	0.8362	0.9065
$n = 30$	0.8372	0.8870	0.9406
$n = 50$	0.8973	0.9451	0.9784
$n = 100$	0.9611	0.9873	0.9976
True $\beta = 3$			
$n = 10$	0.3218	0.2906	0.9833
$n = 20$	0.2338	0.2080	0.9989
$n = 30$	0.1846	0.1573	0.9999
$n = 50$	0.1188	0.0989	1
$n = 100$	0.0472	0.0335	1
True $\beta = 5$			
$n = 10$	0.0160	0.0151	0.9999
$n = 20$	0.0017	0.0012	1
$n = 30$	$8.260e-05$	$4.657e-05$	1
$n = 50$	$4.231e-08$	$4.199e-08$	1
$n = 100$	$3.248e-24$	$2.201e-14$	1

In the next, we consider the Weibull distribution which has been very useful in monitoring the lifetime data. Consider that we have a random sample $X_1, ..., X_n$ from the one parameter Weibull distribution with pdf of the form

$$f(x, \beta) = \beta x^{\beta-1} e^{-x^{\beta}}, x > 0. \tag{6.4}$$

Theorem 6.2. The p-value of an approximate highest density significance test for the Weibull distribution of (6.4) is

$$p_{hd.app}^{Weibull} = \sup_{\beta \in \Theta_\beta} \Phi\left(\frac{\sqrt{n}\{(\beta-1)[n^{-1}\sum_{i=1}^n ln\ x_i + (\gamma/\beta)] - [n^{-1}\sum_{i=1}^n x_i^{\beta} - 1]\}}{((\beta-1)\pi)^2/(6\beta^2) - 2(\beta-1)/\beta + 1}\right).$$

where $\gamma = \int_0^\infty ln\ te^{-t} dt$ is the Euler's constant approximately equaled 0.57722.

Table 8. p-values for approximate highest density significance tests for Weibull distribution when true β is 2

Sample size	$H_0 : \beta \in (2.5, 3)$	$H_0 : \beta \in (1.5, 2.5)$
$n = 10$	0.7180	0.9673
$n = 20$	0.8156	0.4696
$n = 30$	$1.4339E - 07$	0.4596
$n = 50$	0.0001	0.3519
$n = 100$	$8.1764E - 07$	0.7988

In the next example we consider the extreme value distribution with pdf

$$f(x, \mu, \beta) = \frac{1}{\beta} e^{\frac{x-\mu}{\beta}} e^{-e^{\frac{x-\mu}{\beta}}}, x \in R \tag{6.5}$$

where parameters $\mu \in R$ and $\beta > 0$. We assume that we have a random sample $X_1, ..., X_n$ drawn from this distribution. We have p-value for an approximate highest density significance test.

Theorem 6.3. The p-value for an approximate highest density significance test for the extreme value distribution of (6.5) is

$$p_{hd.app}^{Extreme} = \sup_{(\mu,\beta)\in\Theta_{\mu,\beta}} \Phi\left(\frac{(\bar{x} - \mu)/\beta - n^{-1}\sum_{i=1}^{n} e^{(x_i-\mu)/\beta} + 1.57722}{\sqrt{n^{-1}(\pi^2/6 - 1)}}\right)$$

7. Appendix

Proof of Theorem 6.1. With density $f(x, \beta)$, we have

$$ln\ f(x, \beta) = (\alpha - 1)ln\ x - \frac{x}{\beta} - \alpha ln\ \beta - ln\ \Gamma(\alpha)$$

$$(ln\ f(x, \beta))^2 = (\alpha - 1)[ln\ x]^2 - 2(\alpha - 1)\frac{xln\ x}{\beta} - 2(\alpha - 1)(\alpha ln\ \beta + ln\ \Gamma(\alpha))ln\ x$$
$$+ \frac{x^2}{\beta} + 2(\frac{\alpha ln\ \beta + ln\ \Gamma(\alpha)}{\beta})x + (\alpha ln\ \beta)^2 + 2ln\ \Gamma(\alpha)\alpha ln\ \beta + ln\ \Gamma^2(\alpha).$$

It is easy to check that $E[X] = \alpha\beta$ and $E[ln\ X] = ln\ \beta + PG[0, \alpha]$ which indicates that

$$E[ln\ f(X, \beta)] = (\alpha - 1)(ln\ \beta + PG[0, \alpha]) - \alpha(1 + ln\ \beta) - ln\ \Gamma(\alpha). \tag{6.6}$$

Furthermore, we also have

$$E[X^2] = \frac{\Gamma(\alpha + 2)}{\Gamma(\alpha)}\beta^{\alpha}$$
$$E[(ln\ X)^2] = (ln\ \beta + PG[0, \alpha])^2 + PG[1, \alpha]$$
$$E[Xln\ X] = \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)}\beta(ln_{\beta} + PG[0, \alpha + 1]).$$

With the above results, we have

$$E[(ln\ f(X,\beta))^2] = (\alpha-1)\{(ln\ \beta + PG[0,\alpha])^2 + PG[1,\alpha]\} \tag{6.7}$$
$$-\,2(\alpha-1)\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}(ln_\beta + +PG[0,\alpha+1]) - 2(\alpha-1)(\alpha ln\ \beta + ln\ \Gamma(\alpha))$$
$$(ln\ \beta + PG[0,\alpha]) + \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)}\beta^{\alpha-1} + 2\alpha(\alpha ln\ \beta + ln\ \Gamma(\alpha)) + (\alpha ln\ \beta)^2 + ln\ \Gamma(\alpha)^2.$$

Then $Var[ln\ f(X,\beta)]$ is induced from (6.6) and (6.7) and then the theorem is followed. □

Proof of Theorem 6.2. With pdf $f(x,\beta)$ of (6.4), its logarithm is $ln\ f(x,\beta) = ln\ \beta + (\beta-1)ln\ x - x^\beta$. For obtaining the mean and variance of $ln\ f(X,\beta)$, we may derive the followings

$$E[ln(X)] = -\frac{\gamma}{\beta},\ E[ln(X)X^\beta] = \frac{1-\gamma}{\beta},$$
$$E[ln(X)^2] = \frac{6\gamma^2+\pi^2}{6\beta^2},\ E[X^\beta] = 1,\ E[X^{2\beta}] = 2.$$

From the above results, we then have

$$E[ln\ f(X,\beta)] = ln\ \beta - \frac{(\beta-1)\gamma}{\beta} - 1,$$
$$Var[ln\ f(X,\beta)] = -2\frac{\beta-1}{\beta} + (\frac{\beta-1}{\beta})^2\frac{\pi^2}{6} + 1.$$

Imposing these results in (6.1), we then have the theorem.

Proof of Theorem 6.3. The logarithm of the pdf is $ln\ f(x,\mu,\beta) = \frac{x-\mu}{\beta} - ln\ \beta - e^{\frac{x-\mu}{\beta}}$. we also have,

$$E[ln\ f(X,\mu,\beta)] = E[\frac{X-\mu}{\beta}] - ln\ \beta - E[e^{\frac{X-\mu}{\beta}}]$$
$$E[(ln\ f(X,\mu,\beta))^2] = E[\left(\frac{X-\mu}{\beta}\right)^2] - 2ln\ \beta E[\frac{X-\mu}{\beta}] - E[2(\frac{X-\mu}{\beta})e^{\frac{X-\mu}{\beta}}]$$
$$+\,2ln\ \beta E[e^{\frac{X-\mu}{\beta}}] + E[e^{2(\frac{X-\mu}{\beta})}] + (ln\ \beta)^2$$

where $E[\frac{X-\mu}{\beta}] = -0.57722$, $E[\left(\frac{X-\mu}{\beta}\right)^2] = \frac{\pi^2}{6} + (0.57722)^2$, $E[e^{\frac{X-\mu}{\beta}}] = 1$, $E[e^{2(\frac{X-\mu}{\beta})}] = 2$, $E[2(\frac{X-\mu}{\beta})e^{\frac{X-\mu}{\beta}}] = 0.84557$. Henceful, we have

$$E[ln\ f(X,\mu,\beta)] = -0.57722 - ln\ \beta$$
$$Var[ln_f(X,\mu,\beta)] = \frac{\pi^2}{6} - 1.$$

The theorem is followed with implementing the above results in (6.3). □

References

Arbuthnott, J. (1710). An argument for Divine Providence, taken from the constant regularity observ'd in the birth of both sexes. *Philos. Trans.* 27, 186-190. Reprinted in Kendall, M. G. and Plackett, R. L., *Studies in the History of Statistics and Probability, Vol II.* London: Charles Griffin, 1977, 30-34.

Armitage, P. (1983). Trials and errors: the emergence of clinical statistics. *Journal of the Royal Statistical Society A.* 146, 321-334.

Christensen, R. (2005). Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician.* 59, 121-126.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. London A* 222, 309-368.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Edinburg: Oliver and Boyd.

Garthwaite, P. H., Jolliffe, I. T. and Jones, B. (2002). *Statistical Inference.* Oxford University Press: Oxford.

Gossett, W. (1908). The probable error of the mean. *Biometrika* 6, 1-25.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. John Wiley and Sons: New York.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the Theory of Statistics.* McGraw-Hill, Inc.

Neyman, J. and Pearson, E. S. (1967). *Joint Statistical Papers.* Cambridge University Press.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably be supposed to have arisen from random sampling. *Philosophical Magazine* 5, 157-175.

Welsh,A. H. (1996). *Aspects of Statistical Inference.* Wiley: New York.