國 立 交 通 大 學

統計學研究所

博 士 論 文

最高密度顯著性檢定

Highest Density Significance Test

研究生：陳弘家

指導教授：陳鄰安教授

中 華 民 國 九 十 六 年 六 月

最高密度顯著性檢定
Highest Density Significance Test

研　究　生：陳弘家　　　　Student：Hung-Chia Chen

指導教授：陳鄰安　　　　Advisor：Lin-An Chen

國 立 交 通 大 學

統 計 學 研 究 所

博 士 論 文

A Thesis

Submitted to Institute of Statistics

College of Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Ph. D.

in

Statistics

June 2007

Hsinchu, Taiwan

中華民國九十六年六月

最高密度顯著性檢定

學生：陳弘家　　　　　　　　　　　　　　　　　指導教授：陳鄰安

國立交通大學統計學研究所 博士班

中文摘要

顯著性檢定是一種藉由計算p值的方法，用於衡量是否違反虛無假設的統計證據，傳統的顯著性檢定選擇一個統計量$T = t(X)$，同時決定一個極端的集合，此集合包含比觀測值 $t(x_0)$極端或相當的所有點。但是，這個方法有可能無法找到一個合適的統計量，或者不存在一個普遍性的最佳性質來支持既存的顯著性檢定。因此，我們提出一個新的顯著性檢定，設定極端集合包含所有發生機率比觀測值x_0機率小的點，稱為最高密度顯著性(HDS)檢定。此方法應用到較小的機率顯示存在更強的證據否定虛無假設的概念，且將一個樣本X藉由機率比分為極端與非極端的兩個集合。在相同的p值檢定中，我們發現HDS檢定的非極端集合體積最小，此為一最佳性質。我們更進一步延伸HDS檢定來建立控制圖，同時監控所有的參數，並且能精準的達到第一類誤差的機率。藉由監控樣本點的機率來辨識是否受到控制，不像傳統的控制圖是依據樣本平均和全距來監控。

關鍵詞：控制圖，最高密定顯著性檢定，p 值，最小體積

Highest Density Significance Test

student：Hung-Chia Chen Advisor：Dr. Lin-An Chen

Institute of Staistics
National Chiao Tung University

Abstract

The significance test is a method for measuring statistical evidence against null hypothesis H_0 by computing *p*-value. The classical significance test chooses a test statistic $T = t(X)$ and determines the extreme set representing the sample set with values $t(x)$ greater than or equal to $t(x_0)$, where x_0 is the observed sample. It may be difficult to choose a suitable test statistic for the test, or there is no generally accepted optimal theory to support the existed significance tests. Now, we propose a new significance test, called the highest density significance (HDS) test, setting extreme set including those sample points with probabilities less than or equal to it of x_0. It applies the concept that the smaller probability of an observation $X = x_0$ reveals stronger evidence of departure from H_0. This test virtually classifies the sample space of random sample X into extreme set and the non-extreme set through a concept of probability ratio. We also show that this test shares an optimal property for that it has smallest volume among the class of non-extreme sets of significance tests with the same *p*-value. Further, we extend HDS test to set up a control chart which can monitor all the parameters simultaneously and the probability of type I error is precisely attained. Unlike the classical control charts that track statistics such as sample mean $\overline{X}$ or sample range R, it is tracking the density value of the sample point to classify if it is in control.

Key words: Control chart; highest density significance test; p-value; smallest volume

Contents

List of Figures

List of Tables

1 Introduction

With great application in numerous different fields such as natural science, social science, or some others, statistics is frequently used to interpret characteristics of some phenomena involving randomness and variability. It plays a key role in information extraction and drawing conclusions for characteristics of the phenomena when a data set is collected from experiments or field studies. Information extraction includes data reduction and summarizing statistics from the raw data. Statistics not only analyze the observed data from experiments, but also assists designing the experiments. Prior to analysis of statistics, the practitioners or scientists should establish a statistical model that describes the data generation mechanism, and this is for developing statistical methods to condense the information for further making statistical inferences.

The simplest generic form of a statistical model, which was formulated by Fisher and may be seen in Azzalini(1996) and Spanos(1999), contains a probability model and a sampling model, and it purports to describe the population and its variation for data reduction. The probability model represents a probability space assigning a probability distribution P_θ on S where $\theta \in \Theta$ is parameter point and S is the sample space. The sampling model is viewed as the realization of the probability model or events drawn from the population. A generalized statement of statistical model including regression related models may be seen in McCullagh(2002) which is extended from Spanos(1999). Generally, the true parameter θ of the probability model is usually unknown, but the parameters or the distributional characteristics may be drawn conclusions from the observed data via statistical approaches.

The efficiency of a statistical approach relies on the amount and appropriateness of the information being captured from the data. For specific, in-appropriateness of choosing statistical model will never collect full information for statistical inferences. On the other hand, with a specified statistical model, various principles of information extraction resulted in various information sets. Then the usefulness of the information will depend on the preferences of the practitioners or scientists for model and principle selection. The fact of statistical inferences in the literature, point estimation and Neyman-Pearson framework of hypothesis testing are solved respectively with solutions supporting with some desirable optimal properties. However, as interpreted in Birnbaum(1962) *"method such as significance tests and* *interval estimates are in wide standard use for the purposes of reporting and interpreting the essential features of statistical evidence."* These two important statistical inference problems are not solved with solutions supported with desirable properties(Detailed description of these points will be stated latter).

Eventually, the practitioners or scientists would pay attention on what they should do, what they should believe, or what the data does tell, and all of these problems depend on the assumed statistical models and the ways of the data collection. Each specific inference problem require a specific amount of information, i.e., there is no principle that provides amount of information enough for every inference problem. For significance test problem, we will study what an amount of information is appropriate and how it does to reveal a desirable optimal property. Before to do this, we will review statistical inference problems. The following sections will contain three categories which are the most parts of statistical inference, point estimation, interval estimation and hypothesis test in statistical inference.

1.1 Point Estimation

Point estimation tries to present a value based on the observation to estimate the unknown parameter (or distributional characteristics) which often represents a location point or scale for the distribution. In point estimation with a mass data set, data reduction helps in information condensation that help practitioners summarizing useful statistics. When a summarized statistic is used for estimation of the unknown parameter, it is called an estimator of θ. Although computational easiness and naivety often be factors for data reduction to the practitioners, however, estimators developed from these considerations may not be efficient for statistical inferences.

With a great effort for efficient estimation technique, there are many approaches proposed for point estimation, for examples, the method of moments, maximum likelihood estimator, Bayes estimators, or EM algorithm, etc. For selection of estimators, criterions for evaluating estimators have also been proposed in literature. Some important ones are mean squared error, best unbiased, equivariance, average risk optimality, minimaxity, admissibility and asymptotic optimality (see Lehmann and Casella(1999) or Casella and Berger(2001)). In the literature, it has been done in deriving estimators supported with optimal properties such as the uniformly minimum variance unbiased estimator, uniformly most powerful test, Bayes estimator etc. With this achievement, we will not spend any effort on point estimation to seek for any other optimality. However, the maximum likelihood estimator does provide interesting information in the statistical model that is helpful for other statistical inference problems.

For technique with asymptotic optimality, a widely adopted estimator in the parametric model is maximum likelihood estimator which can be traced back to Fisher(1912, 1922a, 1922b, 1925a, 1925b, 1935). The maximum likelihood estimator is derived from solving a parameter point with highest likelihood. It represents the most likely value in the parameter space for which a sample point has already occurred. The use of most likely value leads the maximum likelihood estimator to be not only consistent but also efficient. Besides from estimation for estimators, the Fisher information revealed in the efficiency of the maximum likelihood estimator has been applied in construction of other inference techniques such as the score test and Wald's test. This concept of most likely values may also be interesting for some unsolved statistical inference problems.

1.2 Hypothesis Testing

In the topic of hypothesis testing, there are two important categories for hypothesis specification, the significance test and the Neyman-Pearson formulation. The Neyman-Pearson formulation considers a decision problem which is composed by two components, null hypothesis and alternative hypothesis. 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 scientific guess and we have no idea on assumption about alternative hypothesis when the null hypothesis is not true. Another case is that the model is developed to be checked with new data by a selection process on a subset when H_0 is true. 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.

1.2.1 Neyman-Pearson Formulation

The hypothesis testing for Neyman-Pearson formulation considering a null hypothesis H_0 and an alternative hypothesis H_1 is different from significance test which considers only the null hypothesis. The theory of Neyman-Pearson lemma applies the ratio of the likelihoods with one that H_0 is true and one that H_1 is true. This leads to the result of most powerful test when H_0 and H_1 are both simple and uniformly most powerful test for some composite hypotheses and some specific distributions. Hacking(1965) interpreted the the law of likelihood in the following:

> *If one hypothesis, H_1, implies that a random variable X taking the value x with probability is $f_1(x)$, while another hypothesis, H_2, implies that the probability is $f_2(x)$, then the observation X=x is evidence supporting H_1 over H_2 if $f_1(x) > f_2(x)$, and the likelihood ratio, $f_1(x)/f_2(x)$, measures the strength of the evidence.*

The law of likelihood gathers the information of likelihoods for that H_0 and H_1 are true and use the likelihood ratio to help statisticians in drawing conclusion of acceptance or rejection of null hypothesis.

Besides that it has been with derived optimality, the Neyman-Pearson hypothesis testing can be solved with sample size determination to have a desired power. The significance test is not available in justifying the sample size for the fact that anything is possible when H_0 is not true. There is one other advantage for the Neyman-Pearson hypothesis testing. The use of likelihood ratio will automatically derive the test statistic. This desired property is not shared with most other hypothsis testing problems. With the interesting results including the optimal property, samples size determination and test statistic derivation, the theory of Neyman-Pearson lemma provides us a desired test when an alternative hypothesis may be specified. However, in many practical situations, to specify an alternative hypothesis is not appropriate. In this case, what can we do?

1.2.2 Significance Hypothesis Test

For developing more than 200 years, the significance test has been popularly used in many branches of applied science. Some earlier applications of significance test include Armitage's

(1983) claim finding the germ of the idea in a medical discussion from 1662 and Arbuthnot's(1710) observation that the male births exceeds the female births. Some important significance tests developed latter include the Karl Pearson's(1900) χ^2 test and W.S. Gosset's (1908) student test , the first small-sample test. Significance tests were given their modern justification and then popularized by Fisher who derived most of the test statistics that were broadly adopted in a series of papers and books during 1920s and 1930s. Traditionally, a significance test is to examine whether a given data is in concordance with H_0. The practitioners generally formulate a null hypothesis of interest and specify a test statistic to interpret if the observation provides evidence against H_0. Then, the p-value is determined as the probability of the sample set for that the test statistic is at least as extreme as its observed value when H_0 is true:

$$p = Pr(\text{T at least as extreme as the value observed } |H_0).,$$

where T represents the test statistic.

A significance test always drawn conclusion in terms of p-value. It interprets the p-value as evidence for that the data is consistent with the null hypothesis by concluding that the hypothesis is significant or not to be true. This is different from the Neyman-Pearson framework which always draws conclusion of acceptance or rejection of null hypothesis. The significance test is often being criticized for that it is hard to provide acceptable reason in supporting the chosen test statistic although the sufficient statistic is usually recommended. There is other way in the interpretation of the p-value by saying that it represents the strength of evidence against null hypothesis. From this point, the extreme set has to contain sample points which are at least as large as the observed value or absolute value of the test statistic (see this point in Schervish(1996) and Sackrowitz and Samuel-Chan (1999))

Schervish(1996), Royall(1997) and Donahue(1999) argued that the typical p-value couldn't completely interpret the statistical evidence. With this concern, the p-value has been redefined as the probability of the extreme points determined by the joint density function. In the empirical studies, Hung *et al.*(1997), and Donahue(1999) proposed modifications for significance test based on the Neyman-Pearson approach where two hypotheses are assumed. They discussed p-value in the class of Neyman-Pearson formulation regardless of no alternative assumption. In other words, they connect the Neyman-Pearson formulation to the significance test. This argument is different from the approach that we will introduce.

It is known that the likelihood function has been recognized as a mathematical representation of the evidence (Birnbaum 1962). However, without consistent technique to defining evidence, the classical approaches for the discipline of the significance test really make users confused for that the hypothesis may be significant for one test statistic and insignificant for the other one. Thus, there needs one approach to interpret the statistical evidence that is more convincing than the existing approaches. Hopefully there are interesting properties for this new approach.

1.3 Interval Estimation

Point estimation is always criticized for reporting only a single value for the unknown parameter to the practitioners without flexibility presenting the description of variability. There are two techniques to overcome this drawback. One is a range or interval containing the parameter that contains a specified level of probability or confidence. The other one is hypothesis testing which provides a tool with decision, either determining to reject or accept H_0 or reporting a p-value.

The techniques of interval estimation are always with close relation to them for point estimation and hypothesis test. The interval with confidence level $1-\alpha$ may be derived from the inversion of a level $1-\alpha$ acceptance region of hypothesis testing. Interval estimation may also be derived from which Bayesian approaches which one is criticized for the reason of its determination of subjectively prior distribution. Mood (1974) also categorized the methods of parametric interval estimation into two main techniques. One is based on the pivotal-quantity, and the other is based on the statistical method. Statistical method constructs the confidence interval from the distribution of a statistic. The pivotal-quantity method is the most popular one to derive the confidence interval. But, it requires the existence of a pivotal quantity. However, the pivotal quantity is frequently absent. When this happens, the alternative methods would be the statistical method or the inversion of hypothesis test. Most of the methods stated above are connected with point estimation. They consider the interval as a set composed by two points involving parameters, and they estimate the two points via point estimation.

Without using point estimation, Hyndman (1996) proposed the highest density regions consisting of parameter points of relatively high density due as the confidence set for the parameters. Formulated from likelihood function, the highest density region has the property of smallest volume in sample space which will be discussed in latter section.

1.4 Control Chart

The control charts in Statistical Process Control (SPC) are useful tools to monitor whether various production processes in industry are in statistical control or not. A process is in statistical control when the probability distribution of a random variable which represents the characteristic of the process is unchanged in time. It is often that the distribution involves several parameters. The general way to treat this process is to construct a control chart for each parameter,for example, the $\bar{X}$ and R charts. As we have done for a process with several characteristics(values), the most popular way to monitor the process is to draw conclusion through the results from several control charts.

According to the fact that conclusion from a combination of control charts for variables, it may lead to incorrect control (probability) limits. This drawback has been received extensive attention in literature (Mason, et al. (1997) and Wierda (1994)). Incorrect control limits happen not only in a multivariate variable case but also in the case of univariate variable with distribution involving several parameters. Why not we construct a control chart to assure that the control limits are correct for this circumstance? This problem is similar to test several hypotheses simultaneously. Hence we could consider all of them in a single hypothesis which could interpret if a process is in statistical control.

1.5 Summary

Started from introducing of statistical model, we reviewed the statistical inference approaches in previous sections. The approach with techniques based on likelihood function is especially interesting. The use of maximum likelihood estimation leads to the desired property of asymptotically attaining Cramer-Rao lower bound. On the other hand, the use of likelihood ratio results the most power test. In the following sections, we will study a point in statistical inference that is somehow different from the sufficiency and conditionality. We will investigate the information that is contained in the statistical model and introduce two statistical principles. Then we will follow these principles to define a new extreme set and conduct a new significance hypothesis test, highest density significance test, in section 3 and 4. Further, we will develop several different control charts for different distributions in section 5.

2 Probability and Plausibility

2.1 Concepts of Probability and Plausibility

Birnbaum(1962) proposed that *"report of experimental results in scientific journals should in principle be descriptions of likelihood functions, when adequate mathematical-statistical models can be assumed, rather than reports of significance levels or interval estimates."* The likelihood function formulated from statistical model has been broadly applied with a long history in statistical inference. The simplest form of the statistical model is the case that there existed a random sample $X_1, ..., X_n$ drawn from a distribution with pdf $f(x, \theta), x \in R_x, \theta \in \Theta$ where R_x is sample space of variable X and Θ is the parameter space.

Most statistical inference problems in parametric model arise from the fact that we do not know the true parameter θ as we have the observations. According to the occurrence of the observation, we wish to infer something requested from problems concerning about the true parameter. For dealing with the raised statistical problem, we generally want to summarize important and necessary information to develop a statistical procedure with desired property. Among many available procedures, we expect a desired one to achieve the followings:

(a) Generally accepted criterion of optimal procedures.

(b) Demonstrate that the best procedure is possible, i.e. the best one is obtainable.

Various statistical problems need various amount of information to accomplish the achievement. For a general study, it is interesting to observe that the total amount information is contained in the statistical model postulated by Fisher as

$$\Gamma_x = \{f(x, \theta)x \in R_x, \theta \in \Theta, X = (X_1, ..., X_n) \text{ is a random sample}\}. \quad (1)$$

For aiming in solving various statistical problems, there are several techniques implemented for data reduction, which are proposed trying to gather possible information for inference of θ. With disciplines of statistical problem, it has not been known if there is a technique of data reduction that may accomplish our goal for dealing with various statistical inference problems. The reason raising this question is related to the concept of the sufficiency which was introduced by Fisher(1922). This concept is now played a central role in frequency based inferences and then many frequentist approaches are recommended to rely on the sufficient statistic. It is general in literature that we define a statistic $S(X)$ to be sufficient for the family $\{f(x,\theta),\theta\in\Theta\}$ if the conditional distribution of the random sample X given a value of $S(X)$ does not depend on θ (see, for example Spanos (1999, p627)). Suppose that the sufficient statistic provides enough information for dealing with all statistical inference problems. We must see that there is a statistical technique involving sufficient statistic for every inference problem is shown to have some desirable properties. However, this is not true in some important statistical inference problems.

Sufficient statistic perhaps plays the most important part in developing minimum variance unbiased estimator. Let's see its role in some other statistical inference problems. Mainly due to Barnard (1949,1980), the pivotal quantity, an elegant technique, has been very popularly applied for constructing the confidence interval. On the other hand, a significance test, formulated by Fisher (1915), is a method for measuring statistical evidence against a null hypothesis H_0 and is done by selecting a test statistic T and computing the probability of the tail area of the distribution of T beyond its observed value which is called the p-value. The pivotal quantity for confidence interval and the test statistic for significance test are generally recommended to be constructed involving the sufficient statistic without any careful justification. Unfortunately, the statistical procedures based on them are not justified with any desired optimal properties. It may be that the sufficient statistic, especially the minimal one, condenses the information in the statistical model Γ_x so much that it is not appropriate to be applied to all statistical problems.

Vapnik (1998, p12) argued the techniques for problems in statistical inferences that a restricted amount of information only can solve some special problems and it can never solve all different statistical problems with effective procedures. In interval estimation and significance test, the lack of optimal properties reveals that there must have interesting evidence that can't be discovered in a sufficient statistic. Silvey (1975) and Lindsey (1996) both pointed out that many frequentist theory and techniques for confidence sets appear ad hoc because they are not wholly model-based, relying on the likelihood function, and other single unified principle. It is lack a technique that can capture useful information embedded in the data and the model for construction of inference methods.

Without applying the likelihood function, the techniques for confidence interval in literature are not convincing in terms of plausibility which has the information represented by the size of the likelihood function $L(\theta, x)$ when $X = x$ is observed. This indicates that we should be careful in using sufficient statistic to construct confidence interval. The problem raised by Silvey(1975) and Lindsey(1996) also occurs in the significance test problem. With assuming that H_0 is true, the existed significance tests do not restrict the set of probable non-extreme points. This results that the p-values computed from these significant tests not appropriate as evidence against H_0.

Let's examine the concern of Silvey (1975) and Lindsey (1996) about plausibility and probability of sufficient-statistic based on confidence interval and significance test. When vector $X = x$ is observed , we say that θ_1 is more plausible than θ_2 if $L(x, \theta_1) > L(x, \theta_2)$, where $L(\theta, x)$ is the likelihood function for random sample X. Regarding to the null hypothesis $H_0 : \theta = \theta_0$, we may say that sample point x_1 is more probable than another point x_2 if $L(x_1, \theta) > L(x_2, \theta)$. Suppose that $C(X)$ is a $100(1-\alpha)\%$ confidence interval for θ and $A(x_0)$ is the non-extreme set for a significance test. The likelihood sets corresponding with confidence interval $C(X)$ and significance test when $X = x_0$ is observed are, respectively,

$$LS_C = \{L(\theta, x) : \theta \in C(x), x \in R_x^n\} \text{ for confidence interval, and}$$
$$LS_{A(x_0)} = \{L(\theta_0, x) : x \in A(x_0)\} \text{ for significance test.}$$

For interval estimation, the likelihood set is set of plausibilities values for a confidence interval. As we have discussed above, a set to be more plausibility may be more suitable to play as a confidence interval or significance test. Thus we have to choose a confidence set whose corresponding likelihood set stay away from zero. On the other words, we want to construct a confidence interval which includes the most plausible points. The likelihood sets of some typical confidence intervals will be shown in the following examples.

Example 2.1. Let $X_1, ..., X_n$ be a random sample drawn from the normal distribution $N(\mu, \sigma^2)$. First we consider the confidence interval for mean μ with known variance $\sigma = 1$ for convenience, and then the sample space is $R_x = R$. The popularly used $100(1-\alpha)\%$ confidence interval based on sufficient statistic $\bar{X}$ for μ is

$$(\bar{X} - z_{\alpha/2}\frac{1}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{1}{\sqrt{n}}).$$

Three facts are employed in deriving the likelihood set. (1). $\bar{x} - z_{\alpha/2}\frac{1}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{1}{\sqrt{n}}$ if and only if $n(\bar{x}-\mu)^2 \leq z^2_{\alpha/2}$. (2). $\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$ which has sample space

$(0, \infty)$. (3). $\bar{X}$ and $\sum\limits_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$ are independent. The following we derive its corresponding likelihood set.

$$
\begin{aligned}
LS_C &= \{\tfrac{1}{(2\pi)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}} : \mu \in (\bar{x} - z_{\alpha/2}\tfrac{1}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\tfrac{1}{\sqrt{n}}), \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n\} \\
&= \{\tfrac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2)} : \mu \in (\bar{x} - z_{\alpha/2}\tfrac{1}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\tfrac{1}{\sqrt{n}}), \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n\} \\
&= \{\tfrac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(y_1+y_2)} : 0 \le y_2 \le z^2_{\alpha/2}, 0 < y_1 < \infty\} \\
&= \{\tfrac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}y} : 0 \le y < \infty\} \\
&= \tfrac{1}{(2\pi)^{n/2}}(0, 1].
\end{aligned}
$$

In this example, it has been shown that the likelihood set includes zero which indicates that this confidence interval is implausible.

Similarly, the popularly used significance test based on sufficient statistic $\sum\limits_{i=1}^{n} X_i$ when $X = x_0$ is to compute p-value $P(\sqrt{n}|\bar{X}| \ge \sqrt{n}|\bar{x}_0| | \mu = 0)$ where $\mu = 0$ representing that H_0 is true. By denoting $\bar{x}_0$ as the sample mean of vector x_0, the likelihood set of the significance test derived in the following

$$
\begin{aligned}
LS_{A(x_0)} &= \{\tfrac{1}{(2\pi)^{n/2}} e^{-\frac{\sum_{i=1}^{n} x_i^2}{2}} : |\bar{x}| \le |\bar{x}_0|, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n\} \\
&= \{\tfrac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(\sum_{i=1}^{n}(x_i-\bar{x})^2 + n\bar{x}^2)} : |\bar{x}| \le |\bar{x}_0|, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n\} \\
&= LS_C.
\end{aligned}
$$

is identical to it for the confidence interval for mean.

Next, we consider the confidence interval for variance σ^2 where $\mu = \mu_0$ is also assumed to be known. With sufficient statistic $\sum_{i=1}^{n}(x_i = \mu_0)^2$, the widely adopted $100(1-\alpha)\%$ confidence interval for σ^2 is

$$
\left(\frac{\sum\limits_{i=1}^{n}(x_i - \mu_0)^2}{\chi^2_{1-\alpha/2}}, \frac{\sum\limits_{i=1}^{n}(x_i - \mu_0)^2}{\chi^2_{\alpha/2}}\right).
$$

With σ^2 restricting to the inequality, $\chi^2_{\alpha/2} \leq \frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}$ and each x_i may take any value in R, the likelihood set is

$$LS_C = \{\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2\sigma^2}} : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n, \frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{1-\alpha/2}} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{\alpha/2}}\}$$

$$= (0, \infty).$$

The result is also shown that the likelihood set includes zero and its neighbors. The confidence interval for σ is not a desired confidence interval in sense of plausibility.

Example 2.2. (Likelihood set for confidence interval based on sufficient statistic) Let $X_1, ..., X_n$ be a random sample drawn from the negative exponential distribution with pdf $f(x, \theta) = e^{-(x-\theta)} I(\theta < x < \infty)$. The typical confidence interval for θ is constructed by the pivotal quantity $Y_1 = X_{(1)} - \theta$ which has distribution $Y_1 = Gamma(1, \frac{1}{n})$ which uses the the sufficient statistic $X_{(1)}$. Let a and b be two positive constants satisfying $1 - \alpha = P(a < Y_1 < b)$, and then a $100(1-\alpha)\%$ confidence interval for θ is

$$(X_{(1)} - b, X_{(1)} - a). \tag{2}$$

Since $\frac{\sum_{i=1}^{n}(X_i - X_{(1)})}{n} \sim Gamma(n-1, \frac{1}{n})$ and it is independent of $X_{(1)} - \theta$, we may derive the likelihood set for confidence interval (2) as

$$LS_C = \{e^{-\sum_{i=1}^{n}(x_i-\theta)} \pi_{i=1}^{n} I(\theta < x_i < \infty) : \theta \in (x_{(1)} - b, x_{(1)} - a), \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n\}$$

$$= \{e^{-[\sum_{i=1}^{n}(x_i - x_{(1)}) + n(x_{(1)} - \theta)]} I(\theta \leq x_{(1)}) : a < x_{(1)} - \theta < b, 0 < \sum_{i=1}^{n}(x_i - x_{(1)}) < \infty\}$$

$$= (0, 1).$$

Similarly, it also happens in the case that using only the sufficient statistic makes able to stay from zero. It also happens for a confidence interval which has likelihood set including less plausible points.

In the case of null hypothesis $H_0 : \theta = \theta_0$, suppose that the observation of the random sample is $x_0 = (x_{10}, ..., x_{n0})'$ with $x_{0(1)}$ value of the first order statistic $x_{(1)}$. Then the significance test defines the p-value as $P(X_{(1)} \geq x_{0(1)} | \theta = \theta_0)$. The significance test has the same likelihood set for the confidence interval as shown in the following:

$$LS_{A(x_0)} = \{e^{-\sum_{i=1}^{n}(x_i - \theta_0)} \pi_{i=1}^{n} I(\theta_0 < x_1 < \infty) : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n, x_{(1)} < x_{0(1)}\}$$

$$= \{e^{-[\sum_{i=1}^{n}(x_i - x_{(1)}) + n(x_{(1)} - \theta_0)]} I(\theta_0 < x_1 < \infty) : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n, x_{(1)} < x_{0(1)}\}$$

$$= LS_C.$$

We have examined the likelihood sets for the confidence interval and significance test that are constructed by the sufficient statistic. With these results, we have several conclusions and comments:

(a) Classically, the confidence interval may not contain the most plausible point, the maximum likelihood estimate, and a significance test may not contain the most probable point, the point x^* achieving $\max_{x \in R_x^n} L(x, \theta_0)$. However, the examples mentioned above do contain corresponding most plausible points and most probable points in the corresponding intervals or significance tests. This indicates that the sufficient-statistic based statistical procedures seems to be efficient in catching the information in the statistical model (1) for problems searching for most plausible and most probable points.

(b) From the analyzed examples, the likelihood sets for confidence intervals includes all displausible points (those as closer to zero), and the likelihood sets for significance tests include all disprobable points. This provides evidence to support the concern of Silvey (1975) and Lindsey (1996) about displausibility for existed confidence intervals, where we also have the analogous result for significance test. We then may conclude that the sufficient statistic does not contain all plausibility information and probability information, and then it is inappropriate to say that it is sufficient for the family $\{f(x, \theta), \theta \in \Theta\}$.

(c) As argued by Vapnik (1998), there requires information much more than it provided by a sufficient statistic to deal with a more general statistical problems. Then the information desired to construct the confidence interval is much more than what that the sufficient statistic has provided.

2.2 Probability and Plausibility Principles

In the interval estimation and hypothesis testing, we are looking for either subsets of parameter space or subsets of sample space respectively to estimate and test the parameter. However, it is lack of a justification of optimal property for interval estimation or techniques of some hypothesis testing problems. We expect to propose refined statistical principles which may help in construction of desired and improved techniques.

In point estimation, the Roa-Blackwell theorem plays the most important key for developing the minimum variance unbiased estimator. The theorem mainly provides contribution of sufficiency on point estimation. Birnbaum(1962) has shown that the sufficiency principle and conditionality principle imply the likelihood principle. These principles are defined as follow.

Sufficiency principle: *Consider an experiment* $E = (X, \theta, \{f(x|\theta)\})$ *and suppose* $S(X)$ *is a sufficient statistic for* θ*. If* x *and* y *are sample points satisfying* $S(x) = S(y)$*, then the conclusions drawn from* x *and* y *should be identical.*

Conditionality principle: *Suppose that* $E_1 = (X_1, \theta, \{f_1(x_1|\theta)\})$ *and* $E_2 = (X_2, \theta, \{f_1(x_1|\theta)\})$ *are two experiments, where only the unknown parameter* θ *need to be common. Consider the mixed experiment in which the random variable* J *is observed, where* $P(J = 1) = P(J = 2) = \frac{1}{2}$ *(independent of* θ*,* X_1*, or* X_2*), and then experiment* E_J *is performed. Formally,* $f^*(x^*|\theta) = f^*((j, x_j)|\theta) = \frac{1}{2}f_j(x_j|\theta)$*. Then* $Ev(E^*, (j, x_j)) = Ev(E_j, x_j)$*.*

Likelihood principle: *Suppose that* $E_1 = (X, \theta, \{f_1(x|\theta\})$ *and* $E_2 = (Y, \theta, \{f_2(y|\theta\})$ *are two experiments, where the unknown parameter* θ *is the same in both experiments. Suppose* x^* *and* y^* *are sample points from* E_1 *and* E_2*, respectively, such that* $L(\theta|x^*) = CL(\theta|y^*)$ *for all* θ *and for some constant* C *that may depend on* x^* *and* y^* *but not* θ*. Then the conclusions drawn from* x^* *and* y^* *should be identical.*

We have introduced the concepts of plausibility and probability, and we will develop some principles based on them. These principles will lead us to construct interval estimation and significance test with some optimal properties.

Plausibility principle: *For given x, if $L(\theta, x) = L(\theta', x)$, then any conclusion based on θ and θ' should be identical.*

Probability principle: *For given θ, if $L(\theta, x) = L(\theta, x')$, then any conclusion based on x and x' should be identical.*

These two principles may solve Lindsey's problem to avoid the *ad hoc* appearance of the frequentist theory for that they are lack of unified optimality properties for interval estimation and significance test.

Further, we want to investigate that the application of the probability and plausibility principles do preserve information of probability and plausibility contained in likelihood function through studying some procedures for some general statistical problems. We consider a significance test with null hypothesis $H_0 : \theta = \theta_0$ to decide an acceptance region in R_x^n to support H_0. On the other hand, we need to decide a region in parameter space Θ to support x for interval estimation problem when $X = x$ is observed. The following definition specifies one rule in deciding statistical procedures for many general statistical problems.

Definition 2.1. (i) A likelihood based supporting subregion of sample space for parameter value θ_0 is $\{x : L(\theta_0, x) \geq a\}$. (ii) A likelihood based supporting subregion of parameter space for observation x is $\{\theta : L(\theta, x) \geq a\}$.

In the following examples, we propose procedures based on supporting subregions for some statistical problems to investigate their properties of probability and plausibility.

Example 2.3. (Upper likelihood set) (a) Let $X_1, ..., X_n$ be a random sample drawn from the normal distribution $N(\mu, 1)$, and the sample space $R_x = R$. Let's consider a level α test which has the acceptance region containing the set of values μ on the top of the likelihood function for significance test with null hypothesis $H_0 : \mu = \mu_0$. By the fact that $\frac{1}{(2\pi)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{2}} \geq a$ if and only if $\sum_{i=1}^{n}(X_i - \mu_0)^2 \leq b$ for some b, and we choose $b = \chi^2_{n,\alpha}$ where $\alpha = P(\chi^2_n \geq \chi^2_\alpha)$. With acceptance region

$$\{x : \sum_{i=1}^{n}(x_i - \mu_0)^2 \leq \chi^2_\alpha\},$$

we derive its corresponding likelihood set

$$LS = \{\frac{1}{(2\pi)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2}} : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n, \text{ subjected to } \sum_{i=1}^{n}(x_i-\mu_0)^2 \le \chi^2_\alpha\}$$

$$= \{\frac{1}{(2\pi)^{n/2}} e^{-\frac{y}{2}} : 0 \le y \le \chi^2_\alpha\}$$

$$= \frac{1}{(2\pi)^{n/2}} [e^{-\chi^2_\alpha/2}, 1].$$

(b) Let $X_1, ..., X_n$ be a random sample drawn from the negative exponential distribution with pdf $f(x,\theta) = e^{-(x-\theta)} I(\theta < x < \infty)$. In the case of significance test for the null hypothesis $H_0 : \theta = \theta_0$, a level α test is considering the highest area of the likelihood function as the acceptance region. According to that $L(\theta, x_1, ..., x_n) \ge a$ if and only if $\sum_{i=1}^{n}(X_i - \theta) \le b$, and similarly we have $b = \frac{\chi^2_{2n,\alpha}}{2}$ for significance test. The acceptance region of a significance level α test is $\{x : \sum_{i=1}^{n}(x_i - \theta_0) \le \frac{\chi^2_{2n,\alpha}}{2}\}$. We have its corresponding likelihood set as following.

$$LS = \{e^{-\sum_{i=1}^{n}(x_i-\theta_0)} : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n, \text{ subjected to } 0 \le \sum_{i=1}^{n}(x_i-\theta_0) \le \frac{\chi^2_{2n,\alpha}}{2}\}$$

$$= \{e^{-y} : 0 \le y \le \frac{\chi^2_{2n,\alpha}}{2}\}$$

$$= [e^{\chi^2_{2n,\alpha}/2}, 1].$$

The confidence intervals and acceptance regions are with likelihood sets away from zero. It leads to the following conclusions:

(i) The confidence intervals constructed by likelihood function accomplish the desirability of Lindsey (1996) in sense of containing plausible points of θ. Similarly, the acceptance regions for significance test constructed in the same way contain only probable points of x. This shows us that the likelihood function contains information of probability and plausibility.

(ii) Since the likelihood supporting regions are constructed based on two proposed principles, information of probability and plausibility does been retained in the inference techniques for interval estimation and significance problems.

(iii) The decomposition of the normal likelihood function as

$$\frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}[\sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2]}$$

is involved with the sufficient statistic $\sum_{i=1}^{n} X_i$ and the ancillary statistic $\sum_{i=1}^{n}(X_i - \bar{X})^2$. For the negative exponential case, it also involves sufficient statistic $X_{(1)}$ and ancillary statistic $\sum_{i=1}^{n}(X_i - X_{(1)})$ because

$$e^{-\sum_{i=1}^{n}(x_i-\theta)} = e^{-[n(x_{(1)}-\theta)+\sum_{i=1}^{n}(x_i-x_{(1)})]}.$$

This indicates that probability and plausibility information is partially contained in the ancillary statistic which provides an evidence with significant contribution of ancillary statistic to these two statistical inference problems.

(iv) Generally, the interval estimation and significance test require more information than the sufficiency information to obtain desired inference procedures.

When we deal with confidence interval or significance test for μ in (a) of Example 2.3, the procedures involve $\sum_{i=1}^{n}(X_i - \bar{X})^2$ which is an ancillary statistic for μ. This also exists in (b) of Example 2.3 where the procedures involve $\sum_{i=1}^{n}(X_i - X_{(1)})$ which is also ancillary since it has distribution independent of θ. We will apply these principles to construct a highest density significance test and discuss its properties latter.

3 Highest Density Significance Test

We have early reviewed the idea of significance testing. Classically p-value for a significance test is the probability of an extreme determined from a pivotal quantity which is generally recommended to be constructed with a sufficient statistic. There is no unified theory for developing significance test so that this test doesn't deserve any desirable optimal property and then it is questionable for its interpretation. Hence people are still confused to interpret the p-value since the p-values computed from two test statistics may be dramatically different. People also argue the appropriateness in using the Neyman-Pearson formulation. First, in many situations, we do not really know which assumption for H_1 is appropriate. Second, in some situations, there do have appropriate alternative hypotheses, however the best ones (UMP tests) may not exist. Therefore, it is desired to develop a unified theory for significance test that may automatically and appropriately decide the extreme sets for some computing its$p$-value. Here we proposed a technique using the joint pdf to decide the extreme set that guarantees to include only the less probable sample points in it.

A significance test will be called the highest density significance test when its corresponding extreme set for computing p-value includes only sample points that are less probable than the observed sample point. It may be interpreted as a test with smallest volume non-extreme set. There is an interesting connection between this new significance test and the Neyman-Pearson formulation. The latter one is appealing for being interpreted as a most powerful test and uniformly most powerful test. This appealing actually is contributed with setting an alternative hypothesis so that a likelihood ratio may be applied to decide its corresponding extreme set or the critical region. The appealing for using the likelihood function has also been interpreted by Tsou and Royall (1995), and Hacking(1965) that the likelihood function is a proper tool to capture the evidence of the statistical data for statistical inferences.

The Neyman-Pearson lemma considers the ratio of two likelihoods when observation x is given to detect if this observation x is an extreme point. However, the highest density significance (HDS) test considers, given a θ_0, the probability ratio of a sample point x and observation x_0 to detect if x is an extreme point. Gibbons and Pratt (1975) criticized that p-value of the two-sided minimum likelihood method could lead to absurdities when the underlying distribution is the non-unimodal. They also recommended p-value for one tailed technique reporting that it could retain clear interpretation for the test. However, we argue that the extremity for a sample point should be determined by its probability size, not its corresponding value of a specified test statistic. We consider the HDS test motivated from intuitive implication in probability and plausibility. We will later define the HDS test and discuss it detail.

3.1 Definition and Properties

3.1.1 Likelihood-Based 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. By letting vector X with $X' = (X_1, ..., X_n)$ and sample space Λ, we denote the join pdf of X as $L(x, \theta)$ and also called it the likelihood function. The interest of hypothesis testing is the simple one $H_0 : \theta = \theta_0$, for some $\theta_0 \in \Omega$.

What is generally done in classical approach for significance test with observation $X = x_0$, particularly influenced by R. A. Fisher and being called the Fisherian significance test, is to choose a test statistic and to determine a sample set consistent in terms of the distribution of the test statistic to that H_0 is not true. This sample set is called the extreme set. With a test statistic $T = t(X)$, it then defined the p-value as

$$p_{x_0} = P_{\theta_0}(T \text{ at least as extreme as the observed } t(x_0))$$

Although it is applicable in certain practical problems, however, its dependence on a specified test statistic and the choice of one sided or two sided critical region are often questionable. Moreover, there is generally no suitable justification of optimality even through the procedure has been involved with sufficient statistic. Therefore, this classical significance test is often arguable for dealing with null hypothesis H_0.

In the way of classifying extreme set done in Neyman-Pearson framework with hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, the theorem specifies the extreme set which is often called the critical region or rejection region as

$$\{x : \frac{L(x, \theta_0)}{L(x, \theta_1)} \leq \frac{L(x_0, \theta_0)}{L(x_0, \theta_1)}\}$$

where $k > 0$ is chosen to achieve the restriction of significance level. With the help of using the likelihood function to determine the extreme points, this test has a nice justification for that it is a most powerful (MP) test. It has often been argued that the limitation of the significance tests could not attain any desired optimal property because of being unable to apply the technique of likelihood ratio for measuring the evidence against the null hypothesis since there is no specified alternative hypothesis.

There is reason for that the maximum likelihood estimators is asymptotically efficient. It is that the maximum likelihood chooses the most plausible parameter value when an observation x is observed as the estimate of parameter θ. That is, it is the parameter that greatest probability to x. For the significance test problem that is no specification of alternative hypothesis, we may determine whether an observation x is an extreme point through the probability of $X = x$ when H_0 is true. The joint pdf $L(x, \theta_0)$ expresses the relative "probability" of sample value x when H_0 is assumed to be true. Then, for two observation points $x_a, x_b \in \Lambda$, $X = x_a$ is more or equally probable than $X = x_b$ if the probability ratio $\frac{L(x_a,\theta_0)}{L(x_b,\theta_0)}$ is greater than or equal to 1. A sample point is claimed as an point only if it is at least as extreme as the observed value x_0. This leads to an application of the probability ratio between x and x_0 for defining a new significance test.

Definition 3.1. Consider the null hypothesis $H_0 : \theta = \theta_0$. The HDS test defines the p-value as

$$p_{hd} = \int_{\{x: \frac{L(x,\theta_0)}{L(x_0,\theta_0)} \leq 1\}} L(x, \theta_0) dx.$$

There are some facts following with the method of highest density for significance test. Based on some statistic $T = t(X)$, Fisherian significance test considers the ratio of observations $t(x)$ and $t(x_0)$, $\frac{t(x)}{t(x_0)}$ to determine the extreme set. Its extreme set varies with chosen test statistic. On the other hand, the HDS test consistently applies the probability ratio to determine its extreme set. Therefore, the extreme points included in Fisherian significance test may be excluded from the extreme points of HDS test. This is somehow weird, and the two approaches are discordant. One other interest result is that the random probability ratio $\frac{L(X,\theta_0)}{L(x_0,\theta_0)}$ will automatically determine the test statistic and one sided or two sided test will be simultaneously determined from the probability ratio, which always confuse the practitioners when the classical significance test is applied.

3.1.2 Smallest Volume Non-Extreme Set significance Test

There exist desired properties in statistical inference for point estimation and hypothesis testing, such as uniformly minimum variance of unbiased estimation and uniformly most power tests. However, there is lack of any desired optimal property for the Fisherian significance tests. This makes it difficult in selecting a test statistic. We will define an evaluation technique for significance test.

Definition 3.1. If a test with p-value p_0 has smallest volume of non-extreme set among the class of tests with the same p-value (p_0), we call this significance test the p_0 smallest volume non-extreme set(SVNES) significance test.

Theorem 3.2. Suppose that the observation of the sample is x_0. The HDS test with p-value p_{hd} is the p_{hd} SVNES significance test.

proof. Consider a significance test with p-value p_{hd} that has set of no-extreme points $B(x_0)$. Then,

$$1 - p_{hd} = \int_{L(x,\theta_0)\geq L(x_0,\theta_0)} L(x,\theta_0)dx = \int_{B(x_0)} L(x,\theta_0)dx. \tag{3}$$

Deleting the common subset of $\{x : L(x,\theta_0) \geq L(x_0,\theta_0)\}$ and $B(x_0)$ yields

$$\int_{\{x:L(x,\theta_0)\geq L(x_0,\theta_0)\}\bigcap B(x_0)^c} L(x,\theta_0)dx = \int_{B(x_0)\bigcap\{x:L(x,\theta_0)\geq L(x_0,\theta_0)\}^c} L(x,\theta_0)dx. \tag{4}$$

Now, for $x_a \in \{x : L(x,\theta_0) \geq L(x_0,\theta_0)\}\bigcap B(x_0)^c$ and $x_b \in B(x_0)\bigcap\{x : L(x,\theta_0) \geq L(x_0,\theta_0)\}$, we have $L(x_a,\theta_0) > L(x_b,\theta_0)$. Thus,

$$\begin{aligned} &\text{volume}(\{x : L(x,\theta_0) \geq L(x_0,\theta_0) \text{ and } x \in B(x_0)^c\}) \\ &\leq \text{volume}(\{x : L(x,\theta_0) < L(x_0,\theta_0) \text{ and } x \in B(x_0)\}). \end{aligned} \tag{5}$$

So, adding the volume $\{x : L(x,\theta_0) \geq L(x_0,\theta_0)\}\bigcap B(x_0)$ to both sides of (3),

$$\text{volume}(\{x : L(x,\theta_0) \geq L(x_0,\theta_0)\}) \leq \text{volume}(B(x_0)).\square \tag{6}$$

When $X = x_0$ is observed, there are many alternative Fisherian significance tests dealing with the hypothesis $H_0 : \theta = \theta_0$. The likelihood based significance test, HDS test, is appealing to the justification of smallest volume of non-extreme set in the class of tests with the same p-value.

Theorem 3.3. Let $X = (X_1, ..., X_n)$ be a random sample from $f(x, \theta)$ with a observation $X = x$. Consider the hypothesis $H_0 : \theta = \theta_0$ 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 $t(x)$, then the test with p-value

$$p_{hd} = P_{\theta_0}(t(X) \leq t(x))$$

is a SVNES significance test.

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

$$p_{hd} = P_{\theta_0}(t(X) \geq t(x))$$

is a SVNES significance test.

The proof is omitted because it is resulted from its monotone property. The theorem also shows us that the HDS test is an one-sided Fisherian significance test. We suggest letting the likelihood function to decide if it is an one-sided or two-sided significance test.

Theorem 3.4 The HDS test is invariant in linear transformation.

proof. Suppose X_i's are i.i.d. as $f_\theta(x)$ and x_i's are the observations. Let $Y_i = aX_i + b$ for $a \neq 0$. Then $g_\theta(y)$,the density function of Y_i, is $\frac{1}{|a|}f(\frac{y-b}{a})$.

$$\begin{aligned} p_{hd} &= P(L(Y, \theta) \leq L(y, \theta)) \\ &= P(\prod_{i=1}^{n} g_\theta(Y_i) \leq \prod_{i=1}^{n} g_\theta(y_i)) \\ &= P(\prod_{i=1}^{n} \tfrac{1}{|a|} f_\theta(\tfrac{Y_i-b}{a}) \leq \prod_{i=1}^{n} \tfrac{1}{|a|} f_\theta(\tfrac{y_i-b}{a})) \\ &= P(\prod_{i=1}^{n} f_\theta(X_i) \leq \prod_{i=1}^{n} f_\theta(x_i)).\Box \end{aligned}$$

Obviously, not every nonlinear transformation of HDS test is still a HDS test because the Jacobian depends on the transformation.

Sharing with one optimal property, it is interesting to compare the HDS test with the Neyman-Pearson's MP test in every aspect. We list some in a table that will show that the HDS test share some other interesting properties.

A comparison of MP test and HDS test

	MP test	HDS test
Hypothesis	$H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$	$H_0 : \theta = \theta_0$
Classes of Tests	Test with size $p \leq \alpha$	Test with p-value $= p_{hd}$
Test Derivation	Likelihood Ratio	Probability Ratio
Optimality	Most Power when H_1 is true	Smallest volume for non-extreme set

3.2 HDS Test for Continuous Distributions

We will illustrate the differences from two aspects between the classical Fisherian significance and HDS test through a study of these two tests under several continuous distributions. First, one is to see how much information contained in the model has been involved to measure the evidence again H_0. We will present a study of the HDS test with several examples where we will see evidence against H_0 drawn from sources from the statistical model. Second, it will be seen that the HDS test has the advantage of observing a distributional shift. For this, we will design some special situations assisting to explain these advantages. As recommended by R.A. Fisher, the classical Fisherian significance tests should consider involving only the sufficient statistics. Here we first use the normal distribution to interpret these two aspects.

Example 3.1. Let $X_1, ..., X_n$ be a random sample drawn from a normal distribution $N(\mu, \sigma)$ and consider the null hypothesis $H_0 : \mu = \mu_0, \sigma = \sigma_0$. An appropriate way to interpret this hypothesis is to say that X is drawn from $N(\mu_0, \sigma_0)$ when H_0 is true and anything is possible when H_0 is not true including non-normal distribution. With the fact that $L(x_a, \mu_0, \sigma_0) \geq L(x_b, \mu_0, \sigma_0)$ if and only if $\Sigma_{i=1}^n (x_{ia} - \mu_0)^2 \leq \Sigma_{i=1}^n (x_{ib} - \mu_0)^2$ for $x_a^{'} = (x_{1a}, ..., x_{na})$ and $x_b^{'} = (x_{1b}, ...x_{nb})$, the p-value of the HDS test with observed $X_1 = x_{10}, ..., X_{n0} = x_n$ is

$$p_{hd} = P_{\mu_0}\left(\sum_{i=1}^n (X_i - \mu_0)^2 \geq \sum_{i=1}^n (x_{i0} - \mu_0)^2\right) = P\left(\chi_n^2 \geq \sum_{i=1}^n \frac{(x_{i0} - \mu_0)^2}{\sigma_0^2}\right),$$

where χ_n^2 is the random variable distributed as χ^2 with degrees of freedom n.

In the hypothesis involving assumption of both μ and σ, the HDS test gives an exact p-value which provides the evidence against the assumption that $N(\mu_0, \sigma_0^2)$ is true. The extreme set based on this HDS when observation $x_0 = (x_{10}, ..., x_{n0})'$ is observed is

$$E_{hd}(x_0) = \left\{ x \mid \sum_{i=1}^n (x_i - \overline{x})^2 + n(\overline{x} - \mu_0)^2 \geq \sum_{i=1}^n (x_{i0} - \mu_0)^2, x \in R^n \right\},$$

where the determination of extreme sample point x relies on two variations $(\overline{x} - \mu_0)^2$ and $\Sigma_{i=1}^n (x_i - \overline{x})^2$. One measures the departure of the sample mean from location parameter and the other measures the dispersion of the sample point.

Let's further consider the hypothesis $H_0^* : \mu = 0$ and we assume that $\sigma = 1$ is known. The classical significance test of Fisher based sufficient statistic $\bar{X}$ gives p-value

$$p_x = P_{\mu_0}(\sqrt{n}|\overline{X}| \geq \sqrt{n}|\overline{x}|) = P(Z \geq \sqrt{n}|\bar{x}|),$$

where Z has the standard normal distribution $N(0, 1)$. In fact, the HDS test for H_0^* is exactly the same as it for the hypothesis H_0 and then it also has p-value. Let's interpret the difference between the HDS test and the classical significance test. Suppose that we have drawn a sample of size even number n and the observation is as follows:

$$x_{i0} = \begin{cases} i \times 1000 & \text{if } i = 1, 3, 5, 7, ..., 2n - 1 \\ -(i-1) \times 1000 & \text{if } i = 2, 4, 6, 8, ..., 2n \end{cases}$$

Here $\Sigma_{i=1}^n x_{i0}^2$ is a huge value but $\bar{x}_0 = 0$ such that the Fisherian p-value $p_x = P(|Z| \geq 0) = 1$ and HDS test p-value p_{hd} is approximately 0. There are completely opposite ways indicated from the p-values of two significance test. The unsignificant p-value for Fisherian significance test provides no evidence against H_0^*, but the HDS test provides very strong evidence against H_0^*. It is interesting that the results of these two tests are completely different. Without specified alternative hypothesis, the HDS test gives the p-value indicating that the pre-assumption $\sigma = 1$ may be wrong although $H_0^* : \mu = 0$ is probably valid. In Fisherian significance test, unsignificance p-value leads us to accept H_0^* and do nothing further for this wild observation. In fact, the strong evidence provided by the HDS test indicates that we will not blindly believe that the population mean μ has been changed, but we will probably suspect that σ or the distribution is no longer true.

Let's see the use of HDS test on multivariate data. We consider that $X_1, ..., X_n$ is a random sample drawn from a multivariate normal distribution $N_k(\mu, \Sigma)$. Suppose that the null hypothesis is $H_0 : \mu = \mu_0, \Sigma = \Sigma_0$, where μ_0 and Σ_0 are a known k-vector and a $k \times k$ positive definite matrix, and we have observation $(x_{10}, ..., x_{n0})$. It is seen that $L(X, \mu_0, \Sigma_0) \leq L(x_0, \mu_0, \Sigma_0)$ if and only if $\sum_{i=1}^n (X_i - \mu_0)' \Sigma_0^{-1} (X_i - \mu_0) \geq \sum_{i=1}^n (x_{i0} - \mu_0)' \Sigma_0^{-1} (x_{i0} - \mu_0)$. Then, the p-value of HDS test for this multivariate normal distribution is

$$\begin{aligned} p_{hd} &= P_{\mu_0,\Sigma_0}(\sum_{i=1}^{n}(X_i-\mu_0)'\Sigma_0^{-1}(X_i-\mu_0) \geq \sum_{i=1}^{n}(x_{i0}-\mu_0)'\Sigma_0^{-1}(x_{i0}-\mu_0)) \\ &= P(\chi^2(nk) \geq \sum_{i=1}^{n}(x_{i0}-\mu_0)'\Sigma_0^{-1}(x_{i0}-\mu_0))\Box \end{aligned}$$

Example 3.2. Let $X_1, ..., X_n$ be a random sample drawn from the negative exponential distribution with density function $f(x,\theta) = e^{-(x-\theta)}I(\theta < x < \infty)$. Consider significance tests for the null hypothesis $H_0 : \theta = \theta_0$. Let $X_{(1)}$ represent the first order statistic of this random sample and we denote as random variable with gamma distribution $\Gamma(a, b)$ where a and b are its corresponding parameters. Given an observation $(x_{10}, ..., X_{n0})$, the Fisherian significance test generally chooses sufficient statistic $X_{(1)}$ as the test statistic that yields p-value as follows.

$$\begin{aligned} p_x &= P_{\theta_0}(X_{(1)} - \theta_0 \geq x_{(1)0} - \theta_0) \\ &= P(\Gamma(1, \tfrac{1}{n}) \geq x_{(1)0} - \theta_0) \end{aligned}$$

since $X_{(1)} - \theta$ has gamma distribution $\Gamma(1, \frac{1}{n})$, where $x_{(1)0}$ is the observed value of $X_{(1)}$.

Now, let's consider the HDS test for this hypothesis testing problem. Since $L(x_a, \theta) \geq L(x_b, \theta)$ is equivalent to $\sum_{i=1}^{n}(x_{ia} - \theta) \leq \sum_{i=1}^{n}(x_{ia} - \theta)$, the p-value yielded for the HDS test is

$$\begin{aligned} p_{hd} &= P_{\theta_0}(\tfrac{1}{n}\sum_{i=1}^{n}(X_i - \theta_0) \geq \tfrac{1}{n}\sum_{i=1}^{n}(x_{i0} - \theta_0)) \\ &= P_{\theta_0}(\tfrac{1}{n}\sum_{i=1}^{n}(X_i - X_{(1)}) + (X_{(1)} - \theta_0) \geq \tfrac{1}{n}\sum_{i=1}^{n}(x_{i0} - x_{(1)0}) + (x_{(1)0} - \theta_0)) \\ &= P(\Gamma(n, \tfrac{1}{n}) \geq \tfrac{1}{n}\sum_{i=1}^{n}(x_{i0} - x_{(1)0}) + (x_{(1)0} - \theta_0)), \end{aligned}$$

where $\frac{1}{n}\sum_{i=1}^{n}(X_i - X_{(1)})$ is distributed as $\Gamma(n-1, \frac{1}{n})$ and is independent of $X_{(1)} - \theta$ which has distribution $\Gamma(n, \frac{1}{n})$.

The Fisherian significance test traditionally uses only the sufficient statistic, $X_{(1)}$, to compute p-value and it will claims to have strong evidence of departure from H_0 if $x_{(1)} - \theta_0$ is large enough to yield a small p-value. However, the HDS test computes the p-value based on both $X_{(1)}$ and $\Sigma_{i=1}^{n}(X_i - X_{(1)})$, where the latter measures the sum of distances between each observation X_i and the first order statistic $X_{(1)}$. Thus, it uses information more than it provided by the sufficient statistic to compute p-value.

In consideration of an extreme case, let's assume that $\theta_0 = 0$ and the observation is $x_{(1)0} = 0.01$ and $x_{(i)0} = 100, i = 2, ..., n$. In this situation, the Fisherian significance test will claim that there is no enough evidence against the null hypothesis. However, in this rare case for an exponential distribution produced an observation like this, the small p-value for HDS test leads to strong evidence against null hypothesis. There is no specified alternative hypothesis in a significance test, leads us to suspect if the distribution is no longer an exponential one. This needs a further investigation. □

We have two conclusions drawn from the above two examples:

(I) The results showing in the examples indicate us that the HDS test is more sensitive than Fisherian significance test significance to the rare events.

(II) This sensitivity makes the HDS test to have a desired property in sense of smallest volume of non-extreme set that doesn't happen in classical Fisherian significance tests.

Occasionally, HDS test leads to a test statistic exactly the same as it for a classical Fisherian significance test.

Example 3.3. Let $X_1, ..., X_n$ be a random sample drawn from a distribution with pdf

$$f(x,\theta) = \theta x^{\theta-1}, 0 < x < 1, \theta \in R \text{ and } \theta \neq 0.$$

Suppose that we consider the null hypothesis $H_0 : \theta = \theta_0$. The joint pdf of $X_1, ..., X_n$ under H_0 is $L(x, \theta_0) = \theta_0^n (\prod_{i=1}^{n} x_i)^{\theta_0 - 1}$. Since $L(x_a, \theta_0) \geq L(x_a, \theta_0)$ if and only if $0 < \prod_{i=1}^{n} x_{ib} \leq \prod_{i=1}^{n} x_{ia} < 1$when $\theta_0 > 1$ and $0 < \prod_{i=1}^{n} x_{ia} \leq \prod_{i=1}^{n} x_{ib} < 1$ when $\theta_0 < 1$ and the fact that $-\sum_{i=1}^{n} lnX_i \sim Gamma(n, \frac{1}{\theta_0})$ when H_0 is true, the p-value of the HDS test is

$$p_{hd} = \begin{cases} P(\Gamma(n, \frac{1}{\theta_0}) < -\sum_{i=1}^{n} \ln x_{i0}) & \text{if } \theta_0 < 1, \text{ and } \sum_{i=1}^{n} -\ln x_{i0} < \infty \\ P(-\sum_{i=1}^{n} \ln x_{i0} < \Gamma(n, \frac{1}{\theta_0})) & \text{if } \theta_0 > 1, \text{ and } \sum_{i=1}^{n} -\ln x_{i0} > 0 \end{cases}$$

where $x_0 = (x_{10}, ..., x_{n0})$ is the observation.

The HDS test uses the distribution shape under H_0 automatically to classify the extreme sets producing $E_{hd} = \{(x_1, ..., x_n)' : 0 < -\sum_{i=1}^{n} \ln x_{i0} < -\sum_{i=1}^{n} \ln x_{i0}\}$ if $\theta_0 < 1$ and $E_{hd} = \{(x_1, ..., x_n)' : 0 < -\sum_{i=1}^{n} \ln x_{i0} < -\sum_{i=1}^{n} \ln x_i\}$ if $\theta_0 > 1$. However, this advantage is not shared with the Fisherian significance test.

It is known that $X_i^{\theta_0}$, $i = 1, ..., n$ are iid uniform distribution $U(0, 1)$ when H_0 is true. For the Fishserian significance test, it traditionally uses the best statistic $\prod_{i=1}^{n} x_i^{\theta_0}$. One of the one-sided tests with extreme sets such as $\{(x_1,...,x_n)' : \prod_{i=1}^{n} x_i^{\theta_0} > \prod_{i=1}^{n} x_{i0}^{\theta_0}\}$ or $\{(x_1,...,x_n)' : \prod_{i=1}^{n} x_i^{\theta_0} < \prod_{i=1}^{n} x_{i0}^{\theta_0}\}$ and a two sided test is an alternative choice, where Mood et al.(1974) choose a two-sided version.

In this example, the HDS test and the classical test both are constructed via a same statistic. However, the HDS test has the advantage of automatically determining the extreme set. □

From these examples, we summarize some other conclusions in support of the HDS test. Firstly, the HDS tests automatically determine the extreme set to compute p-values. On the other hand, the Fisherian significance test may be struggling in determining a test statistic or deciding if it is a one sided or two sided test. Secondly, the HDS test constructs the extreme set E_{hd} containing sample points more weirder than those in the non-extreme set such that $L(x_1,\theta_0) < L(x_2,\theta_0)$ for $x_1 \in E_{hd}$ and $x_2 \notin E_{hd}$. This property holds in general only for the HDS test. Thirdly, the HDS test usually use a statistic containing information in the data related to both location and scale parameters to determine the extreme set. This statistic, hence, often combines sufficient and ancillary statistics. However, the Fisherian significance test only involves information contained in the data related to the parameter which is discussed in null hypothesis. With using rich information, the HDS test seems to be quite satisfactory in detecting distributional shift when the observation gives a small p-value. In this situation, a further investigation is needed to detect what happen for a small p-value. It may be thatH_0 is not true or anothter distributional shift. Actually, these possible conclusions are resulted from the use of likelihood function.

3.3 HDS Test for Discrete Distributions

Hypothesis testing for discrete distributions has seldom been discussed in literature. The difficulty of constructing a test for hypotheses of parameters for discrete distribution is that it could be rarely possible to construct a test with an exactly specified significance level when the sample size is finite (see for examples, Welsh (1996, page 146) and Kotz and Johnson (1982)).

The most popular use of level α test for discrete random variables is the normal approximation. However, it is an approximate level α test which in general is not a level α test in the case of finite sample, and its exact significance level is not known without further calculation. One other possibility for the discrete case is the Fisher's fiducial approach. If we consider hypothesis $H_0 : p = p_0$ for a binomial random variable, we may reject H_0 then if and only if $P_{p_0}(X \leq x) < \alpha/2$ or $P_{p_0}(X \geq x) < \alpha/2$ (see Garthwaite et al. (2002, p103)) where x is observed. We prefer to use the p-value to interpret a hypothesis testing problem in replace of the significance level α test since we are dealing with significance test. We re-state the HDS test in an appropriate way for dealing with discrete distribution.

Definition 3.5. Consider the hypothesis $H_0 : \theta = \theta_0$. The HDS test for discrete distribution defines the p-value with observation $X = x_0$ as

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

Generally there are many versions for constructing Fisherian significance test. The choice of a test statistic may affects the conclusion of a test as it behaves in continuous case. The following example presents the results from our concern.

Example 3.4. Let X be a random variable with binomial distribution $b(n, p)$. Consider the hypothesis $H_0 : p = p_0$ and assume that an observation x_0 is available. There are two popularly used Fisherian significance tests defining the p values. First, the normal approximation defines p-value as

$$p_{x_0} = P(|Z| \geq \frac{|x_0 - np_0|}{\sqrt{np_0(1 - p_0)}})$$

where Z has the standard normal distribution $N(0, 1)$. On the other hand, taking X as a test statistic, it has a binomial distribution $b(n, p_0)$ when H_0 is true. Typically, the test statistic is applied to construct a one-sided Fisherian significance test (see Garthwaite et al. (2002)) with p value as follows

$$p_{x_0} = \begin{cases} \sum_{x=x_0}^{n} \binom{n}{x} p_0^x (1 - p_0)^{n-x} & \text{if } x_0 \geq np_0 \\ \sum_{x=0}^{x_0} \binom{n}{x} p_0^x (1 - p_0)^{n-x} & \text{if } x_0 < np_0 \end{cases}.$$

Suppose that the observation is $x_0 = 18$ with sample size n=20. We consider hypothesis $H_0 : p = p_0 = 0.7$. In this situation, p-value p_{x_0} for the normal approximated Fisherian significance test is 0.051 and it for the one-sided Fisherian significance test is 0.0355. Therefore, it often provides various conclusions when we use different Fisherian significance tests. Otherwise, the HDS test defines the p-value as

$$p_{hd} = \sum_{x=0}^{n} \binom{n}{x} p_0^x (1-p_0)^{n-x} I(\binom{n}{x} p_0^x (1-p_0)^{n-x} \leq \binom{n}{x_0} p_0^{x_0} (1-p_0)^{n-x_0})$$

where $I(a \leq b)$ is the indicator function of the event if $a \leq b$ which generates the p-value as 0.0526.□

We present the results in Example 3.4 only for reflecting the fact that in the discrete distribution case there may also have several versions of test statistic for Fisherian significance test and various conclusions may be drawn from these tests. Only a consistent way in defining the significance test makes it sense in the interpretation of a p-value for statisticians. However, as long as we decide to apply the HDS test to determine the bound of p-value so that we may classify a test as a significant one or a non-significant one, this classification should be further studied.

Whatever a Fisherian significance test can interpret, this test is not justified with any desired optimal property. However, it is generally not true that tests for parameters of discrete distributions are with the same p-value (see this point in Welsh (1996)). We slightly revise the definition to enlarge the class of tests for comparison that results the following theorem which may be analogously derived as we did for Theorem 3.2.

Theorem 3.6. Consider that the underlying distribution is discrete and the hypothesis is $H_0 : \theta = \theta_0$. For given observation $X = x_0$, suppose that the HDS test has p-value, p_{hd}, and the number of elements x in the non-extreme set E_{hd}^c is denoted by n_{hd}. For any significance test with p-value smaller than or equal to p_{hd} and its element number in its corresponding non-extreme set is denoted by n_0. Then, $n_0 \geq n_{hd}$.

This optimality of smaller element number for the HDS test does provide a justification for its application to the discrete distributions.

Let's consider the case that $p_0 = 0.5$ and $n = 50$ to verify the result presented in Theorem 3.6, and we list the numbers of non-extreme sets for both HDS test and the one-sided Fisherian significance test that they have p-values of some interest ones less than or equal to 0.05.

Table 1. Numbers of non-extreme points for HDS test and one-sided Fisherian significance test with approximated equal p-value.

p-value	HDS test	One-sided test
0.001	$23 \sim 25$	$37 \sim 38$
0.005	$21 \sim 23$	$35 \sim 36$
0.01	$19 \sim 21$	$34 \sim 35$
0.02	$17 \sim 19$	$33 \sim 34$
0.03	$17 \sim 19$	$33 \sim 34$
0.04	$15 \sim 17$	$32 \sim 33$
0.05	$15 \sim 17$	$32 \sim 33$

For interpretation, suppose that we are interesting in the tests with p-values around 0.02. Then, the HDS test takes about number 17 to 19 of x's in sample space in the non-extreme set for having p-value $p_{hds} \approx 0.02$. However, one-sided Fisherian significance test takes about number 33 to 34 of x's in sample space in the non-extreme set to have the same p-value which is about twice the number for HDS test.

Explicit formulations of p-value for Fisherian significance test and HDS test are generally not obtainable in the case of discrete distribution. For considering the HDS test, we show two results related to the computation of p-value when variable X follows a binomial distribution.

Theorem 3.7. Consider that random variable X has a binomial distribution $b(n,p)$. Then the p-value of the HDS test for hypothesis $H_0 : p = p_0$ with observation $X = x_0$ and it for hypothesis $H_0 : p = 1 - p_0$ with observation $X = n - x_0$ are identical.

Proof. It is followed from the fact that $L(x_0, p = p_0) = L(n - x_0, p = 1 - p_0)$. $\square$

The above theorem indicates that when the p-values of the HDS test for hypothesis $H_0 : p = p_0$ with $p_0 \leq 0.5$ are available, then those for cases $p_0 > 0.5$ are automatically implied. It is very often that the hypothesis about binomial p is the case that $p_0 = 0.5$. We list some results of p-value for HDS test in the following theorem.

Theorem 3.8. Consider that random variable X has a binomial distribution $b(n, p)$ and hypothesis $H_0 : p = 0.5$. We also redenote the p-value for HDS test at $X = x_0$ by p_{hd,x_0}. Then, for $n = 2k$, the p-value of the HDS test is

(a) $p_{hd,x_0} = p_{hd,2k-x_0}$ if $x_0 = k+1, ..., 2k$

(b) $p_{hd,x_0} = \begin{cases} 2\sum_{x=0}^{x_0} L(x, p=0.5) & \text{if } x_0 = 0, 1, ..., k-1 \\ 1 & \text{if } x_0 = k \end{cases}$.

For $n = 2k+1$, we have

(c) $p_{hd,x_0} = p_{hd,2k+1-x_0}$ if $x_0 = k+1, ..., 2k$

(d) $p_{hd,x_0} = \begin{cases} 2\sum_{x=0}^{x_0} L(x, p=0.5) & \text{if } x_0 = 0, 1, ..., k \\ 1 & \text{if } x_0 = k, k+1 \end{cases}$.

Proof. For $n = 2k$, we see that

$$\frac{L(x, p=0.5)}{L(x+1, p=0.5)} = \frac{x+1}{2k-x} \begin{cases} < 1 & \text{if } x = 0, 1, ..., k-1 \\ > 1 & \text{if } x = k, k+1, ..., 2k \end{cases}$$

indicating that $L(x, p=0.5)$ is monotone increasing on $\{0, 1, ..., k\}$ and monotone decreasing on $\{k, k+1, ..., 2k\}$. This indicates the results in (a) and (b).

On the other hand, for $n = 2k+1$, we have

$$\frac{L(x, p=0.5)}{L(x+1, p=0.5)} = \frac{x+1}{2k+1-x} \begin{cases} < 1 & \text{if } x = 0, 1, ..., k-1 \\ = 1 & \text{if } x = k \\ > 1 & \text{if } x = k+1, ..., 2k+1 \end{cases}$$

indicating that $L(x, p=0.5)$ is monotone increasing on $\{0, 1, ..., k\}$ and monotone decreasing on $\{k+1, ..., 2k+1\}$ and the results in (c) and (d) are followed.□

For application, it is desired to have table of p-value for all cases of binomial distributions. It is not difficult in computation of it for every situation of p_0. For considering only $p_0 = 0.5$, we here display p-values in Appendix Tables A.1-A.6 for cases with sample size n is less than or equal to 30.

3.4 The Best Significance Tests

It is known that the Fisherian significance test is not generally accepted since tests of this type do not automatically fulfill any desirable optimal property. With the fact that all HDS tests are best significance tests in some sense, it then raises the question that if there is a guide that some Fisherian significance tests are also best significance tests. The following theorem provides this guide, simply for a special case.

Theorem 3.9. Let $X = (X_1, ..., X_n)$ be a random sample from $f(x, \theta)$ with an observation $X = x$. Consider the hypothesis $H_0 : \theta = \theta_0$ and we assume that the family of densities $\{f(x, \theta) : \theta \in \Theta\}$ has a monotone likelihood in a sufficient statistic $T = t(X)$:

(a) If the monotone likelihood is nondecreasing in $t(x)$, then the left hand Fisherian significance test with p-value

$$p_x = P_{\theta_0}(t(X) \leq t(x))$$

is a best significance test.

(b) If the monotone likelihood is nonincreasing in $t(x)$, then the right hand Fisherian significance test with p-value

$$p_x = P_{\theta_0}(t(X) \geq t(x))$$

is a best significance test.

Example 3.5. (a) Let X be random sample from the exponential distribution with pdf $f(x, \theta) = \theta e^{-\theta x} I(0 < x < \infty)$. The family of this exponential distribution has a monotone likelihood nonincreasing in sufficient statistic $\sum_{i=1}^{n} X_i$. Then the Fisherian significance test with p-value

$$p_x = P_{\theta_0}(\sum_{i=1}^{n} X_i \geq \sum_{i=1}^{n} x_i) = P(\Gamma(n, \frac{1}{\theta_0}) \geq \sum_{i=1}^{n} x_i)$$

is a best significance test.

(b) Let $X = (X_1, ..., X_n)$ be a random sample from the uniform distribution $U(0, \theta), \theta > 0$. Considering the null hypothesis $H_0 : \theta = \theta_0$, the likelihood function is

$$L(\theta, x) = \frac{1}{\theta^n} I(0 < x_{(n)} < \theta)$$

where $X_{(n)}$ is the largest order statistic which is sufficient for parameter θ. Since the space of $x_{(n)}$ is $(0, \infty)$, the family of unform distribution under H_0 has monotone likelihood non-increasing in the sufficient statistic $Y_{(n)}$. The right hand Fisherian test with p-value

$$p_x = P_{\theta_0}(X_{(n)} \geq x_{(n)}) = \begin{cases} \int_{x_{(n)}}^{\theta_0} \frac{n}{\theta_0^n} x^{n-1} dx & \text{if } x_{(n)} < \theta_0 \\ 0 & \text{if } \geq \theta_0 \end{cases}$$

$$= \begin{cases} 1 - (\frac{x_{(n)}}{\theta_0})^n & \text{if } x_{(n)} < \theta_0 \\ 0 & \text{if } x_{(n)} \geq \theta_0 \end{cases}$$

is a best significance test.

From the theory we have developed for best significance test and Fisherian significance tests, we may draw several conclusions from the fact that we apply likelihood function for the HDS test:

(a) If the likelihood function $L(x, \theta)$ involves a univariate parameter θ and has a monotone likelihood in a sufficient statistic, then a one sided Fisherian significance test based on the sufficient statistic is a best significance test. This provides a guide in selecting tests statistic and the choice of one sided or two sided test so that it shares an optimal property.

(b) If the likelihood function $L(x, \theta)$ involves a univariate parameter θ and statistics including a sufficient one and some other ancillary statistics, then the sufficient statistic based Fisherian significance test doesn't share the optimal property. For examples, the best significance test for hypothesis $H_0 : \theta = \theta_0$ in the negative exponential distribution of Example 3.2 involves $\sum_{i=1}^{n}(X_i - \theta_0)$ which is an ancillary statistic.

(c) If the likelihood function $L(x, \theta_1, ..., \theta_k)$ involves k parameters and we testing hypothesis $H_0 : \theta_1 = \theta_{10}, ..., \theta_k = \theta_{k0}$, the Fisherian significance test is classically constructed by Bonferroni technique combining tests, using separate sufficient statistics, for hypotheses $H_0 : \theta_j = \theta_{j0}$. This doesn't share the optimal property. We have an example to explain this point. The HDS test for hypothesis $H_0 : \mu = \mu_0, \sigma = \sigma_0$ where sample is drawn from normal distribution is based on statistic $\sum_{i=1}^{n}(X_i - \mu_0)^2$. This test statistic is the sum of $\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $n(\bar{X} - \mu_0)^2$, one is a function of sample variance S^2 and the other is a function of sample mean where S^2 and $\bar{X}$ are separately sufficient for σ^2 and μ.

(d) If the likelihood function $L(x, \theta_1, ..., \theta_k)$ involves k parameters but we preassume that

$\theta_2 = \theta_{20}, ..., \theta_k = \theta_{k0}$ are known, the interest is to test $H_0 : \theta_1 = \theta_{10}$, the traditional Fisherian significance test is to construct the test statistic based on the sufficient statistic for parameter θ_1. This test definitely doesn't share the optimal property. In this situation, the likelihood function is of the form $L(x, \theta_{10}, ..., \theta_{k0})$ that generally involve extra sufficient statistics for $\theta_2, ..., \theta_k$. However, these extra sufficient statistics are ancillary since their corresponding parameters are preassumed to be known and then are not involved in the Fisherian significance test.

(e) The HDS tests always employ all information of statistics involving in the likelihood function. This is the reason that they are always optimal in sense of smallest volume.

In the rest of this section, we will evaluate the performance of HDS and Fisherian significance tests for cases of several hypothesis problems. For significance test with only null hypothesis, we are allowed to have data departure not only from the null hypothesis, but also from any pre-assumption set on the statistical model including independence, identical distribution, or some pre-assumed parameter values. We expect that a significance test may provide p-value with strong evidence for any of these departures.

A model that the data will be drawn for simulation is called the true model. A model that is assumed by the statistician before the execution of hypothesis testing is called the pre-assumed statistical model. This pre-assumed model may coincides or not coincides with the true statistical model for this situation of significance test. We design the following situations for simulation study:

(a) The true model and the pre-assumed model are identical.

(b) The pre-assumed model has parameter values varies with the true model. This inconsistency could be that the null hypothesis isn't correct or other parameters in the model aren't correctly specified.

(c) The true model is with correlated sample and the pre-assumed model is with random sample assumption.

Example 3.6. We consider two cases for simulation study where one is that the model assumption and the null hypothesis are all correct and the other one is that the model assumption is incorrect. In every simulation for various model assumption, we choose several sample sizes $n = 5, 10, 20, 40, 100$ and replications $m = 10000$. Suppose that $p_j, j = 1, ..., m$ represents the computed p-values of all replication for one test, we compute average p-value $\bar{p} = \frac{1}{m}\sum_{j=1}^{m} p_j$ and standard error $\hat{\sigma}^p = \frac{1}{m}\sum_{j=1}^{m}(p_j - \bar{p})^2$.

Case A: In this simulation, we draw random sample $X = (X_1, ..., X_n)$ from normal distribution $N(0, 1)$.

In first case, we assume that we have random sample drawn from $N(\mu, 1)$ and the considered assumption is $H_0 : \mu = 0$. This is the situation that H_0 is correct and all preassumed assumptions such as random sample, normality and σ and null hypothesis regarding with μ are all correct. Significance tests for this case are expected to have large p-values. The HDS test has p-value, with $x_i, i = 1, ..., n$,

$$p_{hd} = P_{\mu=0,\sigma=1}(\sum_{i=1}^{n} X_i^2 \geq \sum_{i=1}^{n} x_i^2) = P(\chi_n^2 \geq \sum_{i=1}^{n} x_i^2). \tag{7}$$

The Fisherian significance two sided test is with p-value

$$p_x = P_{\mu=0,\sigma=1}(\sqrt{n}|\bar{X}| \geq \sqrt{n}|\bar{x}|) = P(|Z| \geq \sqrt{n}|\bar{x}|) \tag{8}$$

We display the average p-values and their standard errors for the two significance tests.

Table 2. Average p-value for two significance tests under $H_0 : \mu = 0$ when $N(0, 1)$ is true.

n	$\bar{p}_{hds}$	$\hat{\sigma}^p_{hds}$	$\bar{p}_z$	$\hat{\sigma}^p_z$
5	0.4988	0.0829	0.5028	0.0838
10	0.5005	0.0820	0.4992	0.0823
20	0.5004	0.0830	0.4960	0.0832
40	0.5033	0.0849	0.5007	0.0829
100	0.4996	0.0829	0.4995	0.0828

We have two conclusions drawn from the results in Table 2:

(a) The average p-values and their corresponding standard errors for the HDS and Fisherian significance tests are with values very close. So, these two tests perform quite similar when the assumed statistical model is identical with the true statistical model.

(b) With p-value in average nearly 0.5, these two significance tests both provide no real evidence against the assumed statistical model.

Case B: In this simulation, we draw random sample $X = (X_1, ..., X_n)$ from normal distribution $N(0, 4)$.

In first case, we assume that we have random sample drawn from $N(\mu, 1)$ and the considered assumption is $H_0 : \mu = 0$. This is the situation that H_0 is correct but the preassumed $\sigma = 1$ is not true. The HDS test has p-value exactly the same as it in (7) and the Fisherian significance two sided test is with p-value exactly the same as it in (8). We display the average p-values and their standard errors for the two significance tests.

Table 3. Average p-value for two significance tests under $H_0 : \mu = 0$ when $N(0, 4)$ is true.

n	$\bar{p}_{hds}$	$\hat{\sigma}^p_{hds}$	$\bar{p}_z$	$\hat{\sigma}^p_z$
5	0.0780	0.0282	0.2939	0.0930
10	0.0200	0.0063	0.2992	0.0951
20	0.0015	0.0002	0.2952	0.0932
40	$8.4316e-07$	$2.0733e-07$	0.3008	0.0956
100	$1.7347e-15$	$1.1108e-26$	0.2967	0.0929

In the second case, we have samples drawn from the same normal distribution $N(0, 4)$. We assume that we draw them from $N(\mu, 1)$ and the considered assumption is $H_0 : \mu = 1$. This is a situation that H_0 and the preassumed assumption on σ are all incorrect. Then smaller p-values are again desired. The HDS test defines p-value as

$$p_{hd} = P_{\mu=1,\sigma=1}(\sum_{i=1}^{n}(X_i - 1)^2 \geq \sum_{i=1}^{n}(x_i - 1)^2) = P(\chi^2_n \geq \sum_{i=1}^{n}(x_i - 1)^2).$$

The Fisherian significance two sided test is with p-value

$$p_x = P_{\mu=1,\sigma=1}(\sqrt{n}|\bar{X} - 1| \geq \sqrt{n}|\bar{x} - 1|) = P(|Z| \geq \sqrt{n}|\bar{x} - 1|).$$

The following table displays the average p-values and MSE's for two significance tests.

Table 4. Average p-value for two significance tests under $H_0 : \mu = 1$ when $N(0, 4)$ is true.

n	$\bar{p}_{hds}$	$\hat{\sigma}^p_{hds}$	$\bar{p}_z$	$\hat{\sigma}^p_z$
5	0.0500	0.0177	0.1719	0.0704
10	0.0086	0.0022	0.1016	0.0459
20	0.0003	$4.9857e-05$	0.0341	0.0156
40	$1.8765e-07$	$8.706e-11$	0.0039	0.0012
100	0	0	$9.3025e-07$	$2.3016e-09$

We have conclusions drawn from Tables 3 and 4:

(a) The HDS and Fisherian significance tests are with average p-values also decreasing when the sample size decreases. This show that these two significance tests are more efficient in detecting the evidence against the model change when the sample size is large.

(b) The average p-value in each corresponding sample size is smaller than it of the Fisherian significance test. This shows that the former one is more efficient than the latter one in this detection of evidence for a model shift.

(c) The average standard errors of the p-values for the HDS test are relatively smaller than those of the Fisherian significance test. This shows the stability of using the HDS test.

Case C: In this simulation, we draw random sample $X = (X_1, ..., X_n)$ from the following $AR(1)$ model

$$X_i = \eta_i, i = 1, ..., n$$
$$\eta_i = \rho\eta_{i-1} + \epsilon_i$$

where ϵ_i's are i.i.d. with normal distribution $N(0, 1)$.

In this case, we assume that we have a random sample from $N(\mu, 1)$ and the considered assumption is $H_0 : \mu = 0$. This is the situation that H_0 is correct, however, the error variables are not iid and are with $AR(1)$ structure. It is expected to have small p-values for significance tests. The p-values for HDS and Fisherian significance tests are with the same forms, respectively, of (7) and (8). Suppose we will reject H_0 when p value is less than 0.05.

Table 5. Average p-value for two significance tests under $H_0 : \mu = 0$ when $AR(1)$ is true.

ρ	$\bar{p}_{hds}$	n_{rej}	$\bar{p}_z$	n_{rej}
0.1	0.4819	624	0.2330	1398
0.2	0.4245	1001	0.2155	1866
0.3	0.3345	1873	0.1957	2423
0.4	0.2163	3675	0.1713	3282
0.5	0.1131	6228	0.1477	4061
0.6	0.0432	8337	0.1226	5013
0.7	0.0098	9608	0.0947	6106
0.8	0.0012	9945	0.0647	7345
0.9	$6.8908e-05$	9996	0.0335	8600
1.0	$6.4842e-07$	10000	0.0051	9777

We have several conclusions drawn from Table 5:

(a) In this situation that the data drawn from an $AR(1)$ model, however, we compute two significance tests based on a model of iid random variables. The p-values of two considered significance tests are both decreasing when ρ is increasing from zero. It is reasonable that when ρ is close to zero the true statistical and the assumed statistical model are similar.

(b) In cases that ρ's are smaller than 0.4, the Fisherian significance test seems to be better than the HDS test. When ρ's are larger than 0.5, our new test seems to be better.

(c) When we set a significance level as 0.05 these two significance tests are with numbers of rejection increase in ρ in a reasonable trend. One interesting fact is that for case that $\rho = 0.4$ the HDS test has larger average p value but with larger number of rejection.

3.5 Some Further Developments of HDS Test

3.5.1 Approximate HDS Test

Approximation techniques are important for some statistical inference problems. Sometimes the desirability comes from the mathematics difficult for deriving an inference technique, and sometimes it is due to cheaper and quicker technique derivation. We then also consider approximation technique for HDS tests. THe p-value for the HDS test may be written as

$$\begin{aligned} p_{hd} &= P(L(\theta_0, X) \le L(\theta_0, x)) \\ &= P(\prod_{i=1}^{n} f(X_i, \theta_0) \le \prod_{i=1}^{n} f(x_i, \theta_0)) \\ &= P(\sum_{i=1}^{n} lnf(X_i, \theta_0) \le \sum_{i=1}^{n} lnf(x_i, \theta_0)) \\ &= P(\sum_{i=1}^{n}(lnf(X_i, \theta_0) - Elnf(X, \theta_0)) \le \sum_{i=1}^{n}(lnf(x_i, \theta_0) - Elnf(X, \theta_0))) \end{aligned}$$

Suppose that the variance of the log likelihood is finite. By the central limit theorem, we have the following asymptotic p-value of HDS test.

$$\begin{aligned} p_{hd} &\approx P\left(Z \le \frac{\frac{1}{n}\sum_{i=1}^{n} lnf(X_i,\theta_0) - Elnf(X,\theta_0)}{\sqrt{\frac{1}{n}Varlnf(X,\theta_0)}}\right) \\ &= \Phi\left(\frac{\frac{1}{n}\sum_{i=1}^{n} lnf(x_i,\theta_0) - Elnf(X,\theta_0)}{\sqrt{\frac{1}{n}Varlnf(X,\theta_0)}}\right) \end{aligned} \tag{9}$$

Where Z represent the standard normal random variable, and Φ is its distribution function.

For application of approximate HDS test, suppose that now we have a random sample $X_1, ..., X_n$ from binomial distribution $b(k, p)$ with pdf

$$f(x, p) = \binom{k}{x} p^x (1-p)^{k-x}, x = 0, 1, .., k,$$

and we consider hypothesis $H_0 : p = p_0$.

Let's denote

$$\begin{aligned} \mu_p &= E[ln\ f(X, p)] \\ &= \sum_{i=0}^{k} ln\{\binom{k}{i} p^i (1-p)^{k-i}\} \binom{k}{i} p^i (1-p)^{k-i}, \end{aligned} \tag{10}$$

and

$$\begin{aligned} \sigma_p^2 &= E[ln\ f(X, p) - \mu_p]^2 \\ &= \sum_{i=0}^{k} [ln\{\binom{k}{i} p^i (1-p)^{k-i}\} - \mu_p]^2 \binom{k}{i} p^i (1-p)^{k-i}. \end{aligned} \tag{11}$$

Then the p-value of an approximate HDS test is

$$p_{hd.app}^{bin} = \Phi(\frac{\frac{1}{n}\sum_{i=1}^{k}[ln\ \binom{k}{x_i} + x_i\ ln\ (p_0(1-p_0)^{-1}) + k\ ln\ (1-p_0)] - \mu_{p_0}}{n^{-1/2}\sigma_{p_0}}). \tag{12}$$

There is a simulation result of $b(5, p)$ example in Table 6. The true p's are 0.1, 0.2, 0.3, 0.4, and 0.5. n's are 10, 20, 30, 50, and 100.

Table 6. p-value of approximate HDS test for binomial distribution under $H_0 : p = 0.5$.

true p	0.1	0.2	0.3	0.4	0.5
$n = 10$	0.0001	0.0212	0.1714	0.4088	0.5174
$n = 20$	$5.7196e-08$	0.0019	0.0864	0.3612	0.5117
$n = 30$	$6.2980e-11$	0.0002	0.0463	0.3275	0.5092
$n = 50$	$2.1231e-18$	$2.9588e-06$	0.0143	0.2808	0.5079
$n = 100$	$7.1441e-52$	$3.3474e-11$	0.0009	0.1998	0.5036

Consider another example that we have a random sample from a Cauchy distribution with $\frac{1}{\pi\sigma}\frac{1}{1+(\frac{x-\mu}{\sigma})^2}$. For this distribution, we have

$$\begin{aligned} Elnf_{\mu,\sigma}(X) &= \int_{-\infty}^{\infty} ln(\tfrac{1}{\pi\sigma}\tfrac{1}{1+(\frac{(x-\mu)}{\sigma})^2})\tfrac{1}{\pi\sigma}\tfrac{1}{1+(\frac{(x-\mu)}{\sigma})^2}dx \\ &= -ln(\pi\sigma) + 2\int_{-\pi/2}^{\pi/2}\tfrac{ln(cos(\theta))}{\pi}d\theta \\ &= -ln(\pi\sigma) - 2ln2, \end{aligned}$$

$$\begin{aligned} E(lnf_{\mu,\sigma}(X))^2 &= \int_{-\infty}^{\infty}(ln(\tfrac{1}{\pi\sigma}\tfrac{1}{1+(\frac{(x-\mu)}{\sigma})^2}))^2\tfrac{1}{\pi\sigma}\tfrac{1}{1+(\frac{(x-\mu)}{\sigma})^2}dx \\ &= (ln(\pi\sigma))^2 - \tfrac{4ln(\pi\sigma)}{\pi}\int_{-\pi/2}^{\pi/2}ln(cos\theta)d\theta + \tfrac{4}{\pi}\int_{-\pi/2}^{\pi/2}(ln(cos\theta))^2 d\theta \\ &= (ln(\pi\sigma))^2 + 4ln(\pi\sigma)ln2 + \tfrac{4}{\pi}\cdot 4.0932, \end{aligned}$$

and

$$\begin{aligned} Varlnf_{\mu,\sigma}(X) &= E(lnf_{\mu,\sigma}(X))^2 - (Elnf_{\mu,\sigma}(X))^2 \\ &= \tfrac{4}{\pi}\cdot 4.0932 - (2ln2)^2. \end{aligned}$$

Then the p-value of the approximate HDS test $H_0 : \mu = \mu_0$ for Cauchy(μ, σ) is

$$p_{hd,app} = \Phi\left(\frac{\frac{1}{n}\sum_{i=1}^{n} lnf(x_i, \mu_0, \sigma) - Elnf(X, \mu_0, \sigma)}{\sqrt{\frac{1}{n}Varlnf(X, \mu_0, \sigma)}}\right)$$

where σ is known.

We present simulated results in Table 7 and 8 where the sample is drawn from Cauchy distribution with sample size n and known $\sigma = 1$. In Table 7, the data is further drawn when the null hypothesis $H_0 : \mu = 0$ is true. The proportion of rejection is closer to the 0.05 when the sample size is increasing. It indicates that the approximation HSD test is suitable for testing μ when null hypothesis is true. In Table 8, the samples are drawn from Cauchy(1,1). In this case, the null hypothesis is not true, like some other approximate hypothesis tests, and it is difficult to detect the false hypothesis since the proportion of rejection is 0.089 as $n = 5$. Otherwise, the proportion of rejection increases with the sample size increasing although all of the average p-values are larger than 0.1. For testing $H_0 : \mu = 0$, both examples show that the approximate HDS test for Cauchy sample is easier to make a right decision when the sample size is larger than 100. This agrees with the suggestion of Hass, Bain, and Antle (1970).

Table 7. Approximate HDS test for Cauchy sample when null hypothesis is true.

$H_0 : \mu_0 = 0$

n	$\bar{p}_{hds}$	$\hat{\sigma}^p_{hds}$	n_{rej}
5	0.5255051	0.0792775	676
10	0.5211978	0.08016141	631
20	0.5123106	0.08279323	618
40	0.508165	0.08168721	584
100	0.506423	0.08280831	533
200	0.5054276	0.08325652	524

significance level $\alpha = 0.05$ with 10000 replicates

Table 8. Approximate HDS test for Cauchy sample when $\mu = 1$.

$H_0 : \mu_0 = 0$

n	$\bar{p}_{hds}$	$\hat{\sigma}^p_{hds}$	n_{rej}
5	0.4419071	0.06946822	890
10	0.4024538	0.06854606	1031
20	0.3570021	0.06623502	1302
40	0.2878691	0.05730798	1819
100	0.1891455	0.04020064	3234
200	0.1029821	0.02026250	5333

significance level $\alpha = 0.05$ with 10000 replicates

The approximate HDS test will assist the practitioners to compute the p-value that may automatically derive the test statistics. Hence, we do not need to worry about seeking for the pivotal quantity. The practitioners only need to check the existence of 2nd moment of the logarithmic likelihood function.

3.6 Power of a Test

In Neyman-Pearson formulation, we evaluate a the test statistics based on its power which is the probability of rejecting H_0 when the H_0 is false. Now, we consider the two-sided test where there is no uniformly most powerful test. Suppose the sample is drawn from $N(\mu, \sigma^2)$ and the null hypothesis is $H_0 : \mu = \mu_0$ with known $\sigma = \sigma_0$ and the alternative hypothesis is $H_1 : \mu \neq \mu_0$. The typical level α test using the sufficient statistic $\bar{X}$ sets the rejection region $|\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}| > z_{1-\alpha/2}$ where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal. This is not an SVNES significance test. Thus its power is

$$\begin{aligned} \beta_{NP}(\mu, \sigma) &= P_{\mu,\sigma}(|\bar{X} - \mu_0| > \tfrac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}) \\ &= 1 - \Phi(\tfrac{\mu_0-\mu}{\sigma/\sqrt{n}} + z_{1-\alpha/2}\tfrac{\sigma_0}{\sigma}) + \Phi(\tfrac{\mu_0-\mu}{\sigma/\sqrt{n}} - z_{1-\alpha/2}\tfrac{\sigma_0}{\sigma}) \end{aligned} \tag{13}$$

when $N(\mu, \sigma)$ is the true distribution.

For the HSD test, the test statistic is $\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma_0^2}$ and the power is

$$\begin{aligned} \beta_{HDS}(\mu, \sigma) &= P_{\mu,\sigma}(\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\sigma_0^2} > q_{\chi^2_{n,1-\alpha}}) \\ &= P_{\mu,\sigma}(\chi^2_{n,ncp=\frac{n(\mu-\mu_0)^2}{\sigma^2}} > \tfrac{\sigma_0^2}{\sigma^2} q_{\chi^2_{n,1-\alpha}}), \end{aligned} \tag{14}$$

where $\chi^2_{n,ncp=\frac{n(\mu-\mu_0)^2}{\sigma^2}}$ is a noncentral χ^2 random variable with degree of freedom n and non-centrality $ncp = \frac{n(\mu-\mu_0)^2}{\sigma^2}$ and $q_{\chi^2_{n,1-\alpha}}$ is the $1-\alpha$ quantile of central χ^2_n random variable.

The points of power function of these two tests for $H_0 : \mu = 0$ with known variances $\sigma = 1$ and sample size $n = 5$ are shown in Figure 1 and 2. When the assumed variance is true in Figure 3, the power of HDS test is equal or smaller than one of sufficient statistic based test, which is denoted as NP in the figures. As the true variance is equal to 2 in Figure 4, the comparison of the two tests shows that the HDS test has larger power when the true means is close to it the null hypothesis is true but not in the other ways. On the other hand, in Figure 5, the power of HDS test is larger than it based on the other one when the true variance is equal to 3. Hence the HDS test is more sensitive to detect the distributional change when the true variance is larger than the known value. When the true variance is equal to 0.5 which is smaller than it in null hypothesis in Figure 6, the power of HDS test is always smaller. Thus, in this situation, the HDS test can not be easy to reject the null hypothesis.

Figure 1: Power of HDS test and n=5

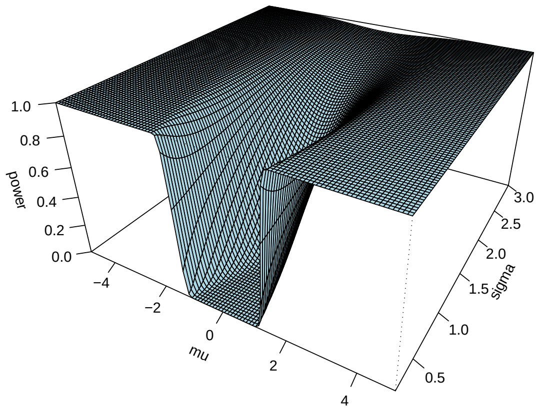


Figure 2: Power of Neyman-Pearson test and n=5

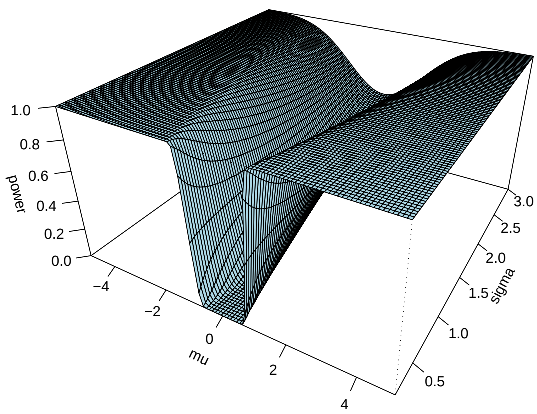


Figure 3: Powers of the two tests as $\sigma^2 = 1$ and n=5

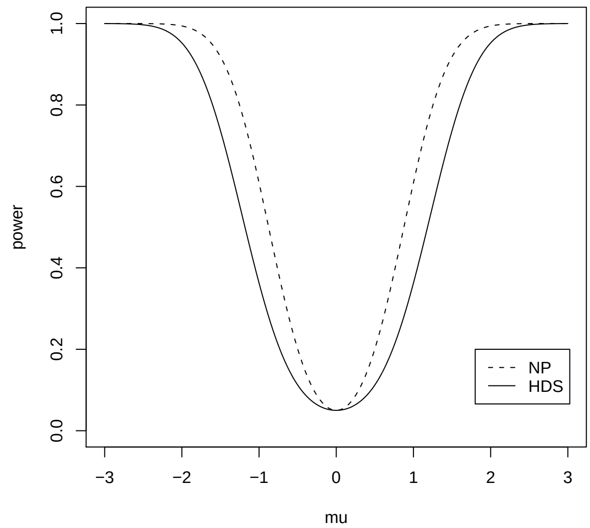


Figure 4: Powers of the two tests as $\sigma^2 = 2$ and n=5

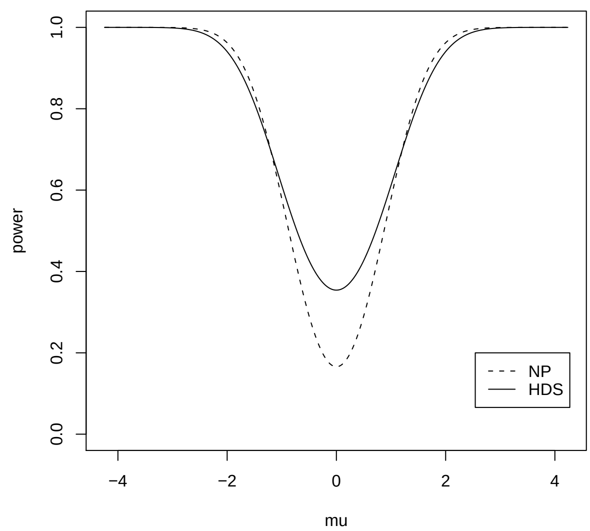


Figure 5: Powers of the two tests as $\sigma^2 = 3$ and n=5

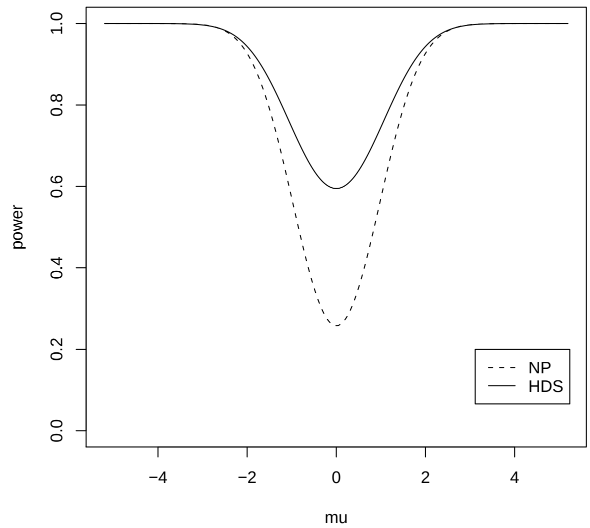


Figure 6: Powers of the two tests as $\sigma^2 = 0.5$ and n=5

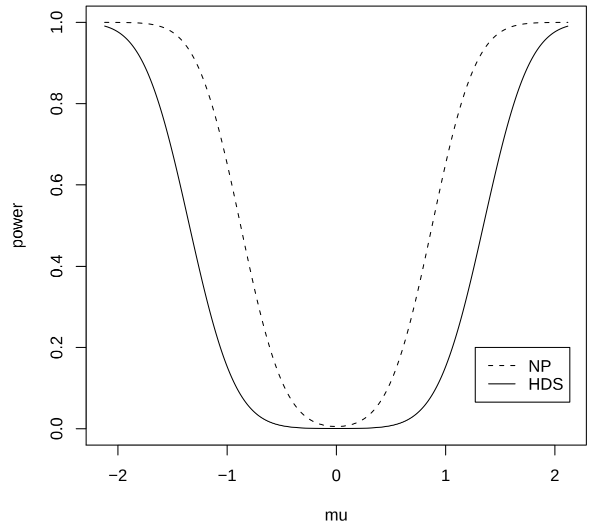


Figure 7: Powers of the two tests as $\mu = 0$ and n=5

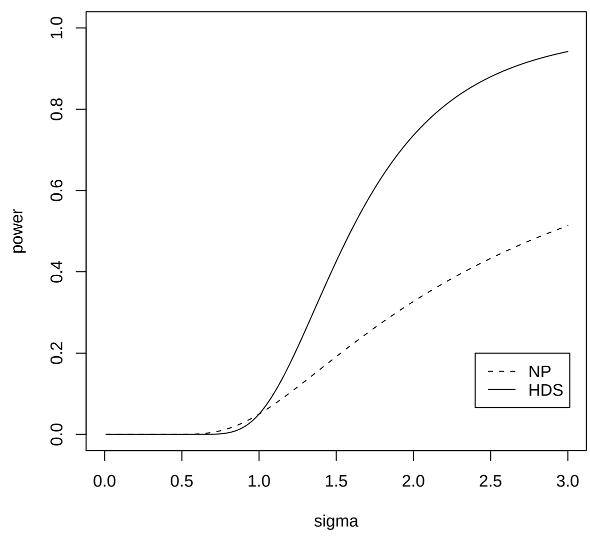
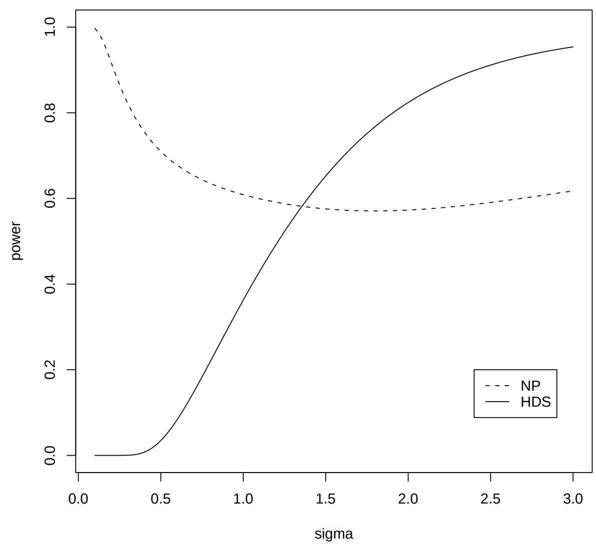


Figure 8: Powers of the two tests as $\mu = 1$ and n=5



Further, the powers of the HDS test are smaller than the other's when variances happen to be smaller in the condition that the mean is fixed to be 0 and 1 in Figure 7 and 8, respectively. With evaluation of the power, we have shown that HDS test for hypothesis of a normal distribution can be easy to detect the false null hypothesis with larger true variance. Although the HDS test has smaller power when the assumed variance is larger than true one, it is not the main consideration of control chart which focuses on the larger dispersion. The extension of the HDS test to control chart will be discussed in the following section.

4 Control Charts

A process statistical control is to see if the process is stable. A stable process indicates that the distribution of the characteristic is unchanged. Hence a process statistical control is to test if there is a distribution shift. Usually, a distribution involves several parameters. Suppose that the density function f with p parameters is denoted $f(x, \theta_1, ..., \theta_p)$. The common method to deal with this problem is to design a statistical process control scheme for each of p parameters separately. It can also be interpreted that a test for each hypothesis is with probability of type I error α. For example, when a variable of a characteristic obeys a normal distribution with mean μ and variance σ^2, the popular technique is to construct $\bar{X}$-chart to monitor the shift of mean μ and R-chart to monitor the change of standard deviation σ.

Suppose that the control charts are constructed by statistics $\hat{\theta}_1, ..., \hat{\theta}_p$ separately for the p parameters and these statistics are independent. Then we will reject the hypothesis to interpret that the process is statistical out of control if some of these charts lead to a verdict of rejection. However, the overall probability of a type I error then becomes $1 - (1 - \alpha)^p$. A second method is to reject the hypothesis whenever all schemes lead to rejection. Then, the overall probability of a type I error becomes α^p. Since the different assertions for testing hypothesis lead to various probability of type I error, it may confuse the user with controlling the error probability. This is one deficit for the classical control charts.

There are another two deficits often occurring in the classical technique to develop control charts. One is the ignorance of possible correlation among the p test statistics. For example, the test statistics involved in $\bar{X}$-chart and R-chart are actually correlated. It leads to incorrect (probability) control charts as we have mentioned. The other one is that an occurrence of a shift in distribution it may not be detectable through a shift in one or more parameters involving in f. In explaining this point, Hoerl and Palm (1992) and Woodall (2000) argued that the control charts are aimed to detect deviations from the model, including the distribution assumption itself.

As considering these deficits, Grimshaw and Alt (1997) argued that the traditional $\bar{X}$ and R charts are efficient in detecting changes of distributional mean and variation. Besides, their efficiencies can be remarkably reduced due to departures from the shape of the density function. Thus they proposed a quantile control chart which has control limits estimated by a confidence band for a quantile vector. They showed that these charts are quite effective in detecting changes in the distributional shape which are undetected through the $\bar{X}$ and R charts. However, we still need to concern that such a nonparameteric control chart generally has less efficiency than appropriate parametric one when the distribution is known.

Unlike the classical Shwhart control chart tracking a sample point through its mean or range, we track the value of its density. Any sample point x to be classified to have either a chance cause or an assignable cause will be only determined by the size of its density. Thus we need only to construct a lower limit for the sample density. This allows us to use only one chart for monitoring the process no matter how many parameters be involved in the distribution and the probability of type I error can be generally controlled with a specified value.

4.1 Density Control Charts

In the hypothesis testing problem, the density function f can be anything other than f_0 such as f_1 when H_0 is false. f_1 can be different from f_0 in its mean, variation or the shape of the density. In control charting, the practitioners concern whether the process is in statistical control interpreted by a distribution. Thus we can extend the optimality of significance test to control charting although there is a debate over the relation between hypothesis testing and control charting (Woodall, 2000). This property is also introduced to tolerance interval based on coverage interval of highest density values by Huang, Chen, and Welsh (2006).

Now, we can establish a control chart which has the same idea of HDS test according to the relative values of a density function. Let $X_1, ..., X_n$ be a random sample drawn from a distribution with pdf f. By setting a HDS test with level α, it leads to region $\{x : L(x, f_0) \leq \ell(f_0)\}$ where $\ell(f_0)$ satisfies

$$\alpha = P_{f_0}(L(X, f_0) \leq \ell(f_0)). \tag{15}$$

We now introduce the framework of a new control chart including a lower density limit and the tracking variable.

Definition 5.1. Let x be the sample point and $L(x, f_0)$ be its joint density. The density

Shewhart control chart specifies the control limit for tracking variable $L(x, f_0)$ as

$$LCL = \ell(f_0)$$
$$\text{Tracking variable: } L(x, f_0)$$

where constant $\ell(f_0)$ satisfies (15).

Although there are many other applications, a control chart is very useful in online process monitoring. To interpret the use of density control chart for detection of assignable causes, let's motivate it from the use of classical control chart. The classical Shwhart control chart monitors a parameter θ with tracking an estimator $\hat{\theta}$ by setting control chart,

$$\begin{aligned} UCL &= \mu_{\hat{\theta}} + 3\sigma_{\hat{\theta}} \\ LCL &= \mu_{\hat{\theta}} - 3\sigma_{\hat{\theta}} \end{aligned} \tag{16}$$

where $\mu_{\hat{\theta}}$ and $\sigma_{\hat{\theta}}$ are, respectively, the mean and standard deviation of $\hat{\theta}$. For control chart in (16), if the sample values of $\hat{\theta}$ fall in the control limits, i.e., $\hat{\theta} \in (LCL, UCL)$, and do not exhibit any systematic pattern, we say that the process is in statistical control at the level indicated by the chart, generally this level is not known. In the case of density control chart, if the process density f remains at the function f_0, then values of $L(x, f_0)$ should be larger than $\ell(f_0)$. Here $\ell(f_0)$ is the lower α percentage point of the distribution of joint density $L(X, f_0)$ when H_0 is true. We then have a rule for online process monitoring through the density control chart as:

If the sample values of $L(x, f_0)$ lies above the lower limit $LCL = \ell(f_0)$ and do not exhibit any systematic pattern, we say that the process is in statistical control at the level $1 - \alpha$.

In this setting, we may set any proper level such as $1 - \alpha = 0.9, 0.95$ or 0.9973. This is in general with difficulty in the classical control chart.

Usually, the pdf f may often be represented in the form $f(x, \theta_1, ..., \theta_p)$ so that the joint pdf $L(x, f)$ and $\ell(f_0)$ may be formulated as $L(x, \theta_1, ..., \theta_p) = \prod_{i=1}^n f(x_i, \theta_1, ..., \theta_p)$ and $\ell(\theta_1, ..., \theta_p)$, respectively. We also generally not know $\theta_1, ..., \theta_p$. Therefore, the unknown parameters must be estimated from preliminary samples taken when the process is thought to be in statistical control. In fact, the answer of the question whether the process is in statistical control may not be known. Thus we have to consider another way to deal with the p parameters such as substitutes for them. Suppose that m samples are available, each containing n observations on the quality characteristic of interest. In practice, Shewhart(1931) substituted $\frac{1}{m}\sum_{j=1}^{m}\bar{x}_j$ and $\frac{1}{m}\sum_{j=1}^{m}s_j$ for μ and σ in the control charts where the $\bar{x}_j$ and s_j are j-th sample mean and standard deviation. This method was also adopted by Chao and Cheng (1996) and Spiring and Cheng (1998). Thus we extend it to all of the parameters of a distribution. Let $\hat{\theta}_{i1}, \hat{\theta}_{i2}, ..., \hat{\theta}_{ip}, i = 1, ..., m$ be the estimated values, respectively, of $\theta_1, ..., \theta_p$ of the m groups of the sample. Then the grand averages of these estimates are

$$\bar{\theta}_j = \frac{1}{m}\sum_{i=1}^{m}\hat{\theta}_{ij}, j = 1, ..., p.$$

Woodall (2000) argued that a control chart is a test concerning the hypothesis that the in-control parameter values are true. From this point, the hypothesis of distribution f for a control chart is appropriate as

$$H_0 : f(x, \theta_1, ..., \theta_p) = f(x, \bar{\theta}_1, ..., \bar{\theta}_p).$$

Then, we have a control chart agreeing with Woodall's point when the underlying distribution involves parameters $\theta_1, ..., \theta_p$.

Definition 5.2. The density Shewhart control chart specifies the framework of density control chart as

$$LCL = \ell(\bar{\theta}_1, ..., \bar{\theta}_p)$$

$$\text{Tracking variable: } L(x, \bar{\theta}_1, ..., \bar{\theta}_p)$$

We notice that the rule for online process monitoring is still valid with the sample density values replacing $L(x, f_0)$ by $L(x, \bar{\theta}_1, ..., \bar{\theta}_p)$.

4.2 Density Control Charts for Some Distributions

4.2.1 Density Control Charts for Normal Distribution

One of the most important uses of a control chart is to improve the process. Consequently, we may also use the the density control chart to evaluate if there are assignable causes. For example, when a sample point $X = x$ falls below the density control limit, this x may reveals to be an out of control sample point and there may exist an assignable cause. If assignable causes can be eliminated from the process, variability will be reduced and the process will be improved. In this section, we introduce density control charts for normal distribution.

Suppose that a quantity characteristic is normally distributed with unknown mean μ and variance σ^2, and then we have density at x as $L(x,\mu,\sigma) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}}$. Since $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$, the inequality $L(X,\mu,\sigma) \geq \ell(\mu,\sigma)$ subjected to $1-\alpha = P_{\mu,\sigma}(L(X,\mu,\sigma) \geq \ell(\mu,\sigma))$ yields $\ell(\mu,\sigma) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\frac{\chi^2_{n,1-\alpha}}{2}}$. The framework of the normal density control chart is

$$LCL = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\frac{\chi^2_{n,1-\alpha}}{2}}$$

$$\text{Tracking variable: } L(x,\mu,\sigma) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

According to Shewhart (1931), we can replace μ and σ by $\bar{\bar{x}} = \frac{1}{m}\sum_{j=1}^{m}\bar{x}_j$ and $\bar{s} = \frac{1}{m}\sum_{j=1}^{m}s_j$ respectively, where $\bar{x}_j$ and s_j are the sample mean and the standard deviation of j-th group respectively. The density control chart turns out to have the framework as

$$LCL = \frac{1}{(2\pi)^{n/2}(\bar{s}^2)^{n/2}} e^{-\frac{\chi^2_{n,1-\alpha}}{2}}$$

$$\text{Tracking variable: } L(x,\bar{\bar{x}},\bar{s}) = \frac{1}{(2\pi)^{n/2}(\bar{s}^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\bar{\bar{x}})^2}{2\bar{S}^2}}$$

Because the value of likelihood function is usually too small to monitor and no obviously difference between in-control and out-of-control, the log-likelihood will be easier to identify the out-of-control points. The framework for log-likelihood control chart is

$$LCL = -\frac{n}{2}ln(2\pi\bar{s}^2) - \frac{\chi^2_{n,1-\alpha}}{2}$$

$$\text{Tracking variable: } ln(L(x,\bar{\bar{x}},\bar{s})) = -\frac{n}{2}ln(2\pi\bar{s}^2) - \frac{\sum_{i=1}^{n}(x_i-\bar{\bar{x}})^2}{2\bar{S}^2}$$

For a given observation $x = (x_1,...,x_n)'$, we compare its probability $L(x,\bar{\bar{x}},\bar{s})$ with the lower control limit LCL and the control chart indicates a sign of out of control if $L(x,\bar{\bar{x}},\bar{s}) < LCL$.

Example 4.1. The process control of vane operating, which is an important functional parameter for a component part for a jet aircraft engine, has been studied in constructing statistical control charts by Montgomery, Runger and Hubele (2004) to assess the statistical stability of this manufacturing process. With preliminary 20 samples of sample size 5, they first constructed $\bar{X}$ chart that indicates the samples, numbered $6, 8, 11, 19$, are departure from the process mean and R chart that indicates the sample, numbered 9, is shift with variation. Removed these samples potentially resulted from assignable causes, they further construct $\bar{X}$ and R charts from the rest of 15 samples for future judgement of statistical stability of the manufacturing process. These procedures are shown in Figure 9, 10, 11, and 12.

Figure 9: $\bar{X}$ chart for Vane Opening

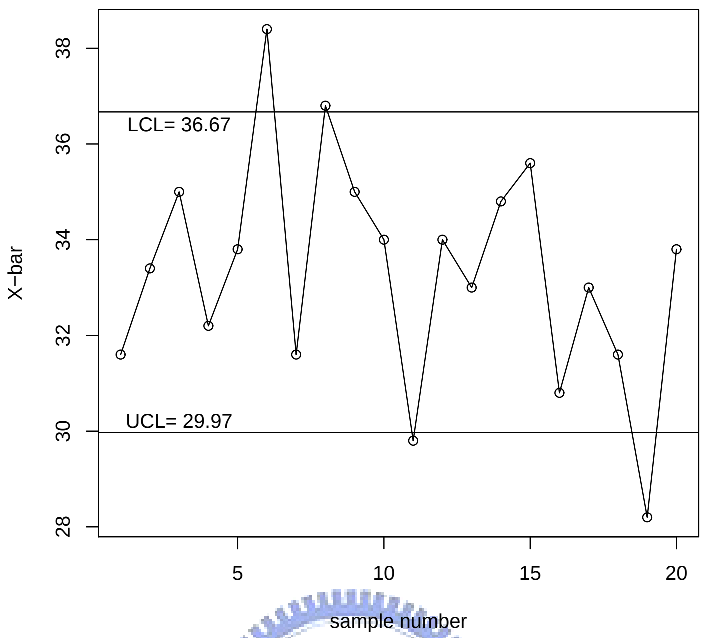


Figure 10: R chart for Vane Opening

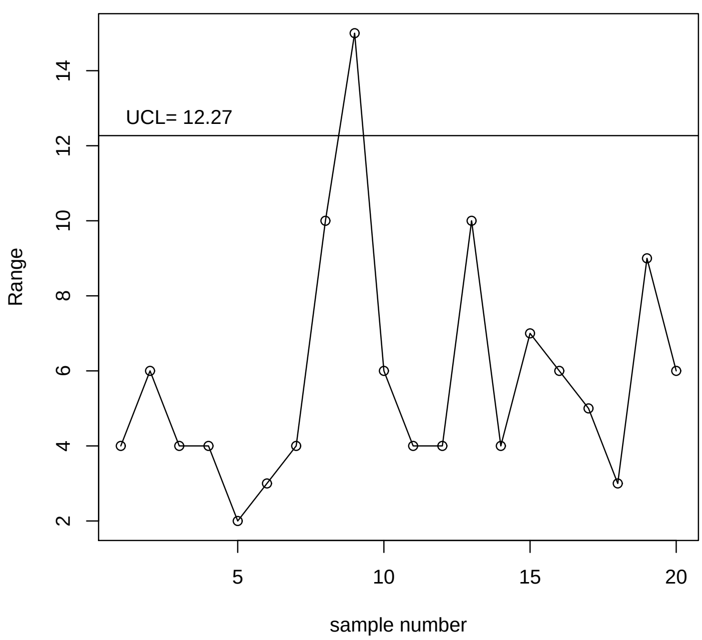


Figure 11: $\bar{X}$ chart for Vane Opening,revised limits

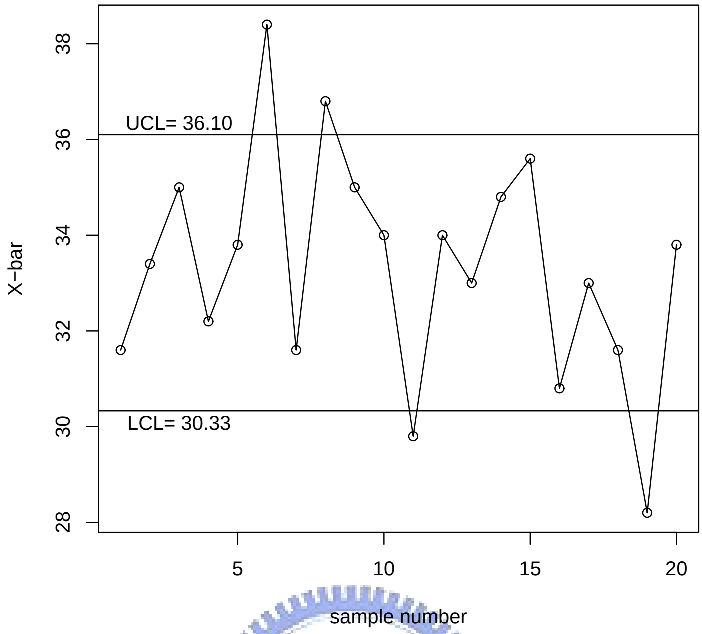


Figure 12: R chart for Vane Opening, revised limit

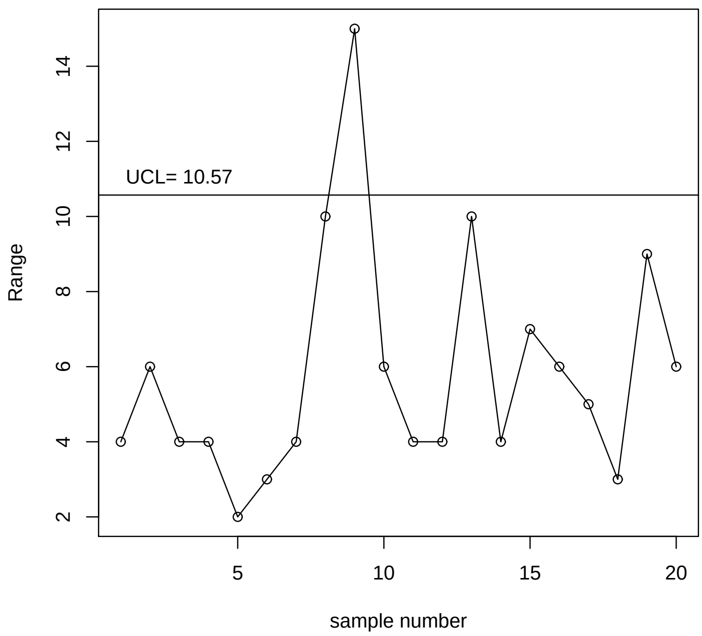
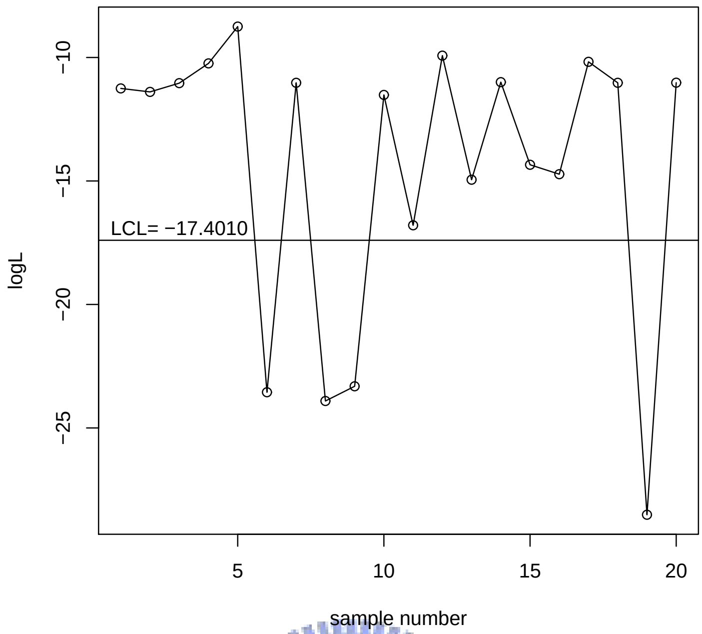


Figure 13: Log-density control chart for Vane Opening



Now, we want to construct density control chart from this preliminary samples. For easy presentation in this study, we construct the log-density control chart as

$$LCL = -\frac{n}{2}ln(2\pi\bar{s}^2) - \frac{\chi^2_{n,1-\alpha}}{2}$$

$$\text{Test statistic function: } lnL(x_1, ..., x_n, \bar{\bar{x}}, \bar{s}) = -\frac{n}{2}ln(2\pi\bar{s}^2) - \frac{\sum_{i=1}^{n}(x_i-\bar{\bar{x}})^2}{2\bar{S}^2} \tag{17}$$

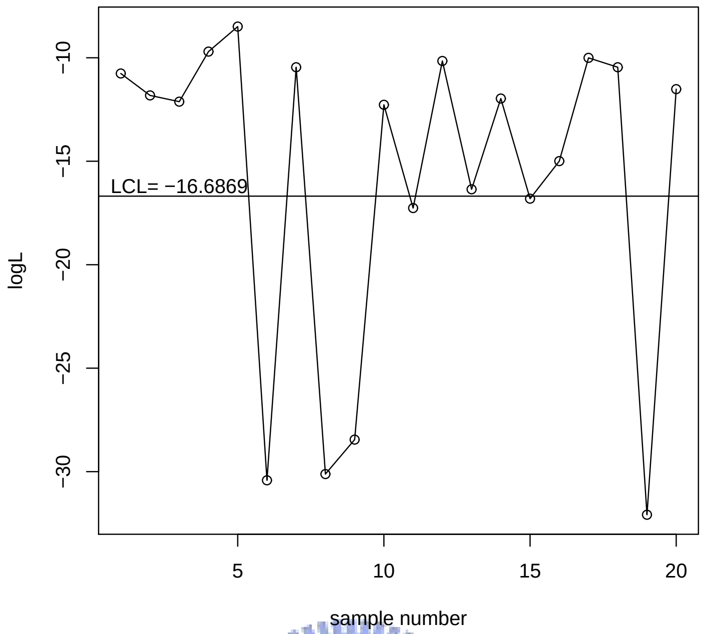
Computed from these 20 preliminary samples, we have $\bar{\bar{x}} = 33.32$ and $\bar{S} = 2.09748$. Hence the lower control limit for the log-density control chart is

$$LCL = -17.4010.$$

Plotting the test statistic function $lnL(x_1, ..., x_5, \bar{\bar{x}} = 33.32, \bar{s} = 2.09748) = -8.298398 - \frac{\sum_{i=1}^{n}(x_i-33.32)^2}{8.7989}$ for these twenty samples associated with the control limit, we have the log-density control chart in Figure 13.

We see that sample numbers 6, 8, 9, 19 are out of control on this log-density control chart. We should discard these four samples, considered as being resulted from assignable causes, and recompute the log-density control limit. Computing from the rest 16 samples, we have

Figure 14: Log-density control chart for Vane Opening, revised limits



$\bar{\bar{x}} = 33$ and $\bar{S} = 1.81833$. Hence the lower control limit for the log-density control chart is

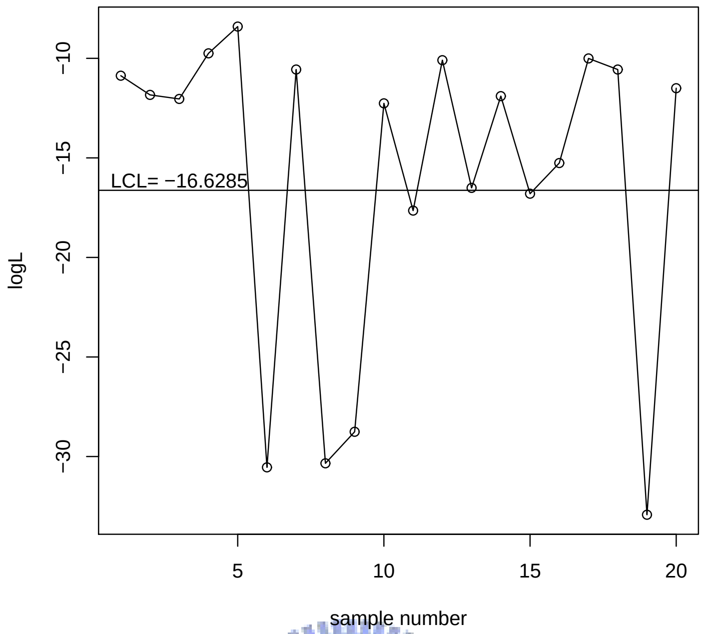
$$LCL = -16.6869.$$

Again, plotting the test statistic function $lnL(x_1, ..., x_5, \bar{\bar{x}} = 33, \bar{s} = 1.81833) = -7.584304 - \frac{\sum_{i=1}^{n}(x_i - 33)^2}{6.6127}$ for the preliminary twenty samples associated with the new control limit in Figure 14, we may see that the samples of possibly resulted from assignable causes is the set numbered $6, 8, 9, 11, 15, 19$. Since there is one more sample, numbered 11, found to be one possibly resulted from assignable cause, we discard these five samples and recompute the log-density control limit that yields $\bar{\bar{x}} = 33.0428$ and $\bar{S} = 1.79721$. Hence we have a new log-density control chart as

$$LCL = -16.6285$$
$$\text{Test statistic function: } lnL(x_1, ..., x_n) = -7.525884 - \frac{\sum_{i=1}^{n}(x_i - 33.0428)^2}{6.4599} \tag{18}$$

We may see that the samples of possibly resulted from assignable causes is still the set numbered $6, 8, 9, 11, 15, 19$. Hence, the log-density control chart of (18) in Figure 15 can now be used to judge the statistical control of the manufacturing process. It is interesting that

Figure 15: Log-density control chart for Vane Opening, twice revised limits



the sample numbered 15 appears out of control in density control chart but not in $\bar{X}$ and R charts. This shows that a sample possibly resulted from assignable causes may not be detected from the $\bar{X}$ and R charts. The control chart of (18) should be revised periodically and when the process has been improved.□

Let $\chi_o^2 = \frac{\sum_{i=1}^n (x_i - \bar{\bar{x}})^2}{\bar{s}^2}$ which has χ^2 distribution with degrees of freedom n, where $\bar{\bar{x}}$ is the in control mean and $\bar{s}$ is the in control standard deviation. From the relation $L(x, \bar{\bar{x}}, \bar{s}) \leq LCL$ if and only if $\chi_o^2 \geq \chi^2_{n,1-\alpha}$, we may see that the density control chart is equivalent to the following chi-square control chart,

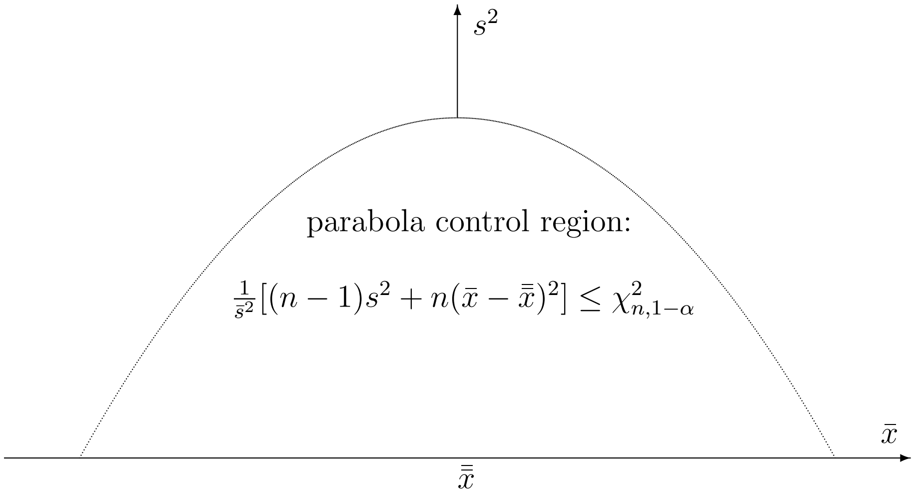
$$UCL = \chi^2_{n,1-\alpha}$$

$$LCL = 0$$

$$\text{Tracking variable: } \chi_o^2 = \frac{\sum_{i=1}^n (x_i - \bar{\bar{x}})^2}{\bar{s}^2}$$

When x is observed, the control chart checks x by comparing the chi-square value χ_o^2 with the lower and upper control limits. The rule for process online monitoring is: If the sample chi-square values χ_o^2 fall within the control limits, LCL and UCL, and do not exhibit any systematic pattern, we say that the process is in statistical control at level $1 - \alpha$. The

Figure 16: A control parabola for normal distribution.



chi-square control chart may be represented graphically.

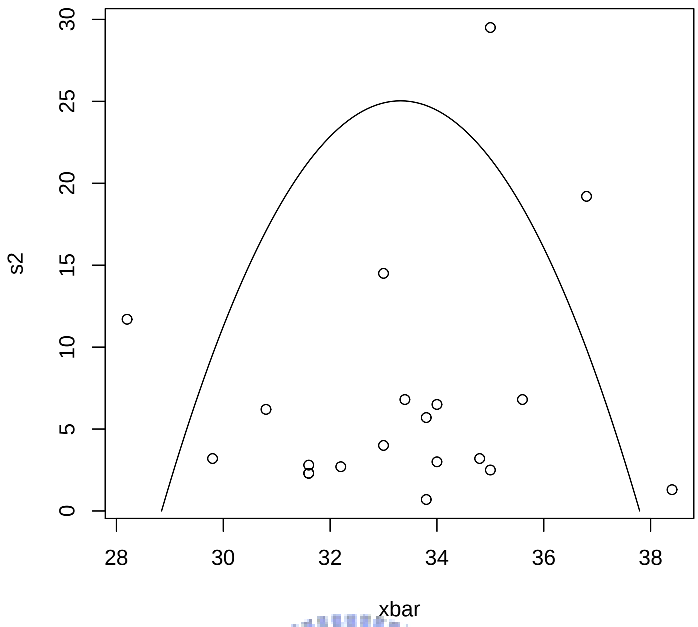
In other words, the monitor value of chi-square control chart is composed by two parts since

$$\chi^2_o = \frac{n(\bar{x} - \bar{\bar{x}})^2}{\bar{s}^2} + \frac{(n-1)s^2}{\bar{s}^2} \tag{19}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the mean and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the variance for the sample point x with $x' = (x_1, ..., x_n)$. Equation (19) defines a parabola centered at $(\bar{\bar{x}}, o)$ with principal axes parallel to the s^2 axis shown in Figure 16. The representation of this control chart is analogous to semicircle chart proposed by Chao and Cheng (1996). Hence the semicircle chart is a density control chart for normal distribution.

Taking χ^2_o in (19) equal to $\chi^2_{n,1-\alpha}$ implies that the sample x with sample mean $\bar{x}$ and sample variance s^2 is on the curve with value χ^2_o. Any observation locating inside the parabola indicates that the process is statistical in control, but otherwise it is statistical out of control. We may call it the control parabola and Figure 17 is the example of Vane Opening. Although it loses tracking time sequence which is also argued in semicircle control by Cheng and Thaga (2006), this problem is popular in a single chart to monitor multiple parameters. The log-likelihood control chart is suitable for monitoring the change of the process with multiple parameters in the time sequence.

Figure 17: Parabola control chart for vane opening



4.2.2 Density Control Charts for Negative Exponential Distribution

Suppose that a process distribution for a quality characteristic is negative exponentially distributed with probability density function

$$f(x,\theta_1,\theta_2) = \frac{1}{\theta_1}e^{-\frac{x-\theta_2}{\theta_1}}, x \geq \theta_2 \tag{20}$$

where $\theta_1 > 0$ and $\theta_2 \in R$ are unknown parameters. In this situation, we have joint density at $x_1, ..., x_n$ as $L(x_1, ..., x_n, \theta_1, \theta_2) = \frac{1}{\theta_1^n}e^{-\frac{\sum_{i=1}^{n}(x_i-\theta_2)}{\theta_1}}$. Suppose that we have a training sample $x_{ij}, i = 1, ..., n, j = 1, ..., m$ of m groups of size n from an in control distribution. We can then calculate m location estimates $\hat{\theta}_{2j} = \min\{x_{1j}, ..., x_{nj}\}, j = 1, ..., m$ and scale estimates $\hat{\theta}_{1j} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \hat{\theta}_{2j}), j = 1, ..., m$; as well as their average $\bar{\theta}_1 = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_{1j}$ and $\hat{\theta}_2 = \min\{\hat{\theta}_{21}, ..., \hat{\theta}_{2m}\}$. The appropriate hypothesis for this in control process distribution is $H_0 : X \sim f(x, \bar{\theta}_1, \hat{\theta}_2)$. When we consider that $\bar{\theta}_1$ and $\hat{\theta}_2$ as the true θ_1 and θ_2, we have $\frac{\sum_{i=1}^{n}(X_i-\hat{\theta}_2)}{\bar{\theta}_1} \sim \Gamma(n, 1)$. The inequality $L(X_1, ..., X_n, \bar{\theta}_1, \hat{\theta}_2) \geq \ell(\bar{\theta}_1, \hat{\theta}_2)$ subjected to $1 - \alpha = P_{\bar{\theta}_1,\hat{\theta}_2}(L(X_1, ..., X_n, \bar{\theta}_1, \hat{\theta}_2) \geq \ell(\bar{\theta}_1, \hat{\theta}_2))$ yields $\ell(\bar{\theta}_1, \hat{\theta}_2) = \frac{1}{\bar{\theta}_1^n}e^{-\frac{\chi^2_{2n,1-\alpha}}{2}}$. We then have the following new control chart when the process distribution is negative exponential:

$$
\begin{gathered}
LCL = \frac{1}{\bar{\theta}_1^n} e^{-\frac{\chi^2_{2n,1-\alpha}}{2}} \\
\text{Test statistic function: } L(x_1, ..., x_n, \bar{\theta}_1, \hat{\theta}_2) = \frac{1}{\bar{\theta}_1^n} e^{-\frac{\sum_{i=1}^{n}(x_i - \hat{\theta}_2)}{\bar{\theta}_1}}.
\end{gathered} \tag{21}
$$

For a given observation $x_1, ..., x_n$, we compare its density value $L(x_1, ..., x_n, \bar{\theta}_1, \hat{\theta}_2)$ with the lower control limit LCL. The control chart indicates being out of control if $L(x_1, ..., x_n, \bar{\theta}_1, \hat{\theta}_2) < LCL$.

We may choose $t(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\theta}_2)$ as the test statistic. With the fact that $\chi_0^2 = \frac{2\sum_{i=1}^{n}(X_i - \hat{\theta}_2)}{\bar{\theta}_1}$ has distribution χ^2_{2n} when H_0 is true and the relation $L(x_1, ..., x_n, \bar{\theta}_1, \hat{\theta}_2) \leq LCL$ if and only if $\chi_0^2 \geq \chi^2_{2n,1-\alpha}$, we may see that the new control chart is exactly a chi-square control chart as

$$
\begin{gathered}
UCL = \frac{\bar{\theta}_1}{2n} \chi^2_{2n,1-\alpha} \\
LCL = 0 \\
\text{Test function: } t(x_1, ..., x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta}_2)
\end{gathered} \tag{22}
$$

The average run length (ARL) tell us, for a given situation in the distribution, how long on the average we will plot successive control chart points before we detect a point beyond the control chart. The ARL from the chart in (22) for this negative exponential distribution is

$$
ARL = \frac{1}{P(\chi^2_{2n} \geq \frac{\bar{\theta}_1}{\theta_1}\chi^2_{2n,1-\alpha} - \frac{2n(\theta_2 - \hat{\theta}_2)}{\theta_1})}.
$$

We performed a simulation to study the ARL with respect to location (θ_2) shift and scale (θ_1) shift. With sample size $n = 5$ and in-control parameter values $\bar{\theta}_1 = 1.0$ and $\hat{\theta}_2 = 0$, the following table display show the resulting ARL.

Table 9. ARL for the mean and location shifts.

	$\theta_2 = 0$	$\theta_2 = 0.2$	$\theta_2 = 0.5$	$\theta_2 = 1.0$	$\theta_2 = 2.0$	$\theta_2 = 3.0$
$n = 5$						
$\theta_1 = 1.0$	370.37	258.01	151.41	64.03	13.06	3.43
$\theta_1 = 1.2$	76.26	57.57	38.10	19.65	5.91	2.23
$\theta_1 = 1.5$	17.83	14.57	10.84	6.79	2.97	1.57
$\theta_1 = 2.0$	5.01	4.42	3.68	2.77	1.71	1.22
$\theta_1 = 2.5$	2.66	2.44	2.17	1.80	1.34	1.10
$\theta_1 = 3.0$	1.87	1.76	1.62	1.44	1.18	1.05
$n = 10$						
$\theta_1 = 1.0$	370.37	204.90	87.36	23.55	2.97	1.09
$\theta_1 = 1.2$	50.63	32.81	17.69	7.00	1.78	1.03
$\theta_1 = 1.5$	9.25	6.98	4.71	2.67	1.26	1.01
$\theta_1 = 2.0$	2.53	2.20	1.81	1.40	1.06	1.00
$\theta_1 = 2.5$	1.51	1.40	1.27	1.13	1.02	1.00
$\theta_1 = 3.0$	1.21	1.16	1.11	1.05	1.00	1.00

We have several conclusions drawn from the results in Table 9:

1. The ARL is strictly decreasing when either one of θ_1 and θ_2 or both increase. This density control chart is with rapid detection of large shifts in the process level.

2. The rapidity in increases of θ_1 and θ_2 are significantly different. The above results reveals that the detection of process out of control is not very sensitive when there is only location change.

3. When sample size n increases the ARLs are decrease that reflect the efficiency of statistical inferences for large sample sizes.

Example 4.2. Monitoring the semiconductor manufacturing process often involves a characteristic having non-normal distribution. Levinson and Polny (1999) studied a particle counts data generated from an Applied Materials etcher. This is data of size 48. For this data and its generation in detail, see Levinson and Polny (1999). They fitted the data with a χ^2 test and showed that the three-parameter gamma distribution with following probability density function

$$f(x,\theta_1,\theta_2,\theta_3)=\frac{1}{\Gamma(\theta_3)\theta_1^{\theta_3}}(x-\theta_2)^{\theta_3-1}e^{-\frac{x-\theta_2}{\theta_1}},x\geq\theta_2$$

with mean $\mu=\theta_1\theta_3+\theta_2$ is appropriate to explain the this characteristic of particle counts. In their study, the maximum likelihood estimates of distribution parameters are $\hat{\theta}_1=222.32,\hat{\theta}_2=34$ and $\hat{\theta}_3=1.172$. The control limits for the Shewhart control chart of μ with $n=1$ are $[-362.63,951.79]$, and it has a serious problem since negative particles do not exist.

With the fact that the value of $\hat{\theta}_3$ is close to 1, the negative exponential distributions in (20) may provide a simpler family to explain the particle counts. To investigate this conjecture, we set hypothesis H_0 for assuming that X follows the negative exponential distribution and perform the χ^2 test with cell classification identical to Levinson and Polny (1999) that leads to p-value 0.2466 and the K-S test that leads to p-value 0.7889. Both tests support the use of a negative exponential distribution.

We divide the data into 12 samples of size $n=4$ and compute the density control chart of (21) that yields the lower control limit and the sample values of the test statistic function listed in the following tables and the log-density control charts in Figure 18 and 19.

Table 10. Lower control limit and sample values of the test statistic function

Control limit $(\times 10^{-10})$	$L(x_1,..,x_4,\hat{\theta}_1,\hat{\theta}_2)(\times 10^{-10})$			
$LCL=0.000115$	0.158882	0.000014	0.025054	0.657089
	0.003889	0.000107	0.067155	0.22817
	0.005071	0.019951	0.621203	1.898203

This shows that samples numbered 2 and 6 are out of control. Although the density control only uses only one test statistic to monitor the process quality, however, we still can decompose the variability of each sample point into elements (for example, mean shift and variation shift) that may be explained by the classical Shewhart control charts ($\bar{X}$ and R charts). For this purpose, we consider the chi-square chart of (22) for explain.

After revised the control limit, the upper control limit of the chi-square chart of (22) is $UCL=401.83$. By denoting $v=\frac{1}{n}\sum_{i=1}(x_i-x_{(1)})$ and $u=x_{(1)}$ for the chart in Figure 20, variable v measures the variation of the sample and the distance between u and $\hat{\theta}_2$ measures

Figure 18: Log-density control chart for Particle Count

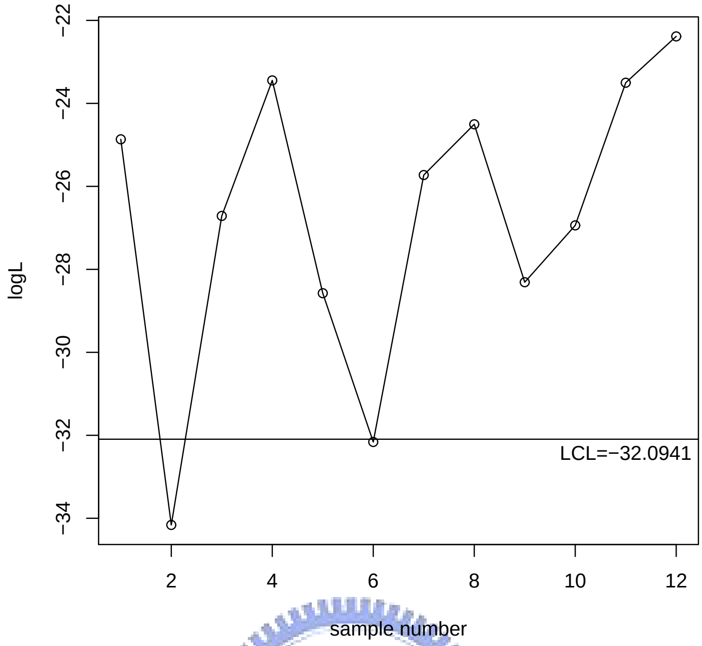


Figure 19: Log-density control chart for Particle Count, revised limit

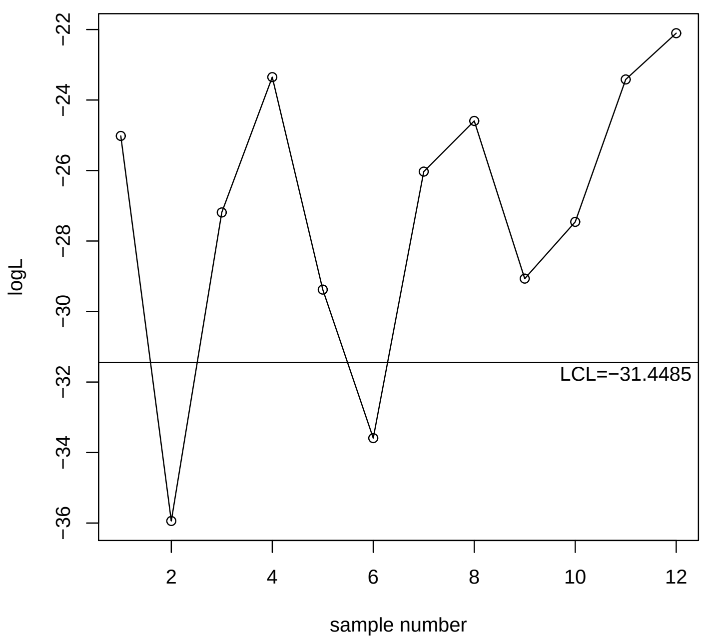
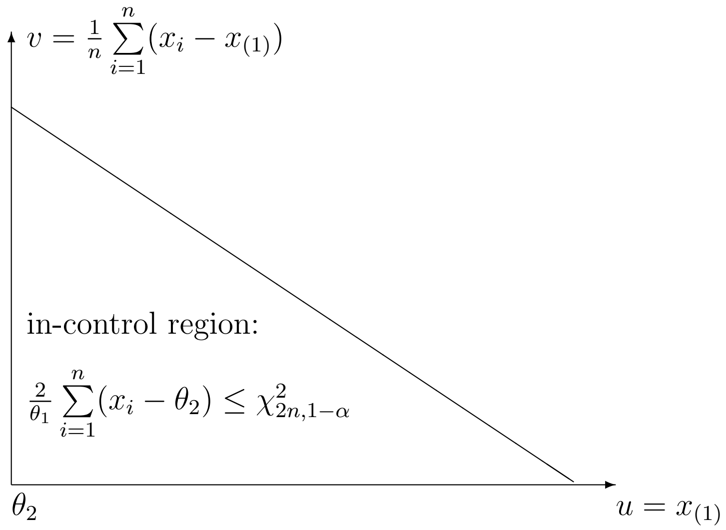


Figure 20: χ^2 control chart for negative exponential distribution.



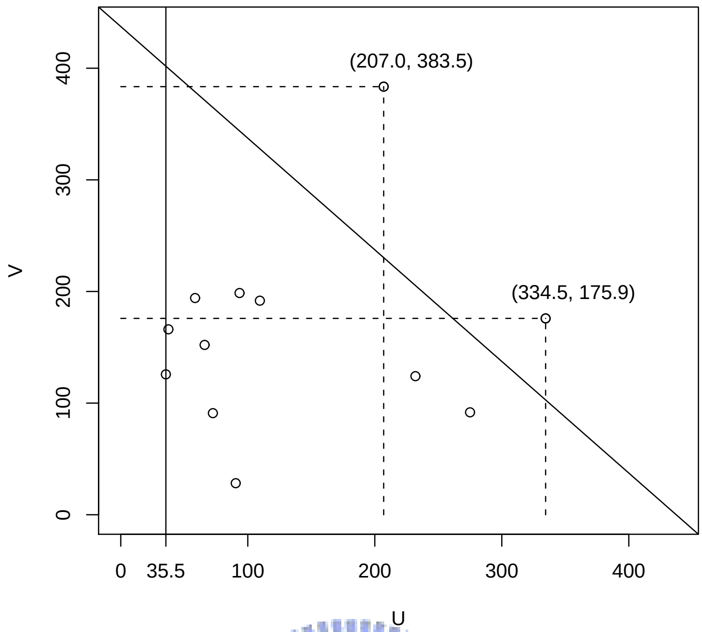
its location shift. Hence the diagonal line with axes u and v defined from the equation

$$v + (u - 35.5) = 401.83$$

represents a control limit to monitor the location and variation. Any sample point $(x_1, ..., x_n)'$ with values u and v lying above the diagonal line is classified as an out of control point and we can tell if it is caused by the location shift and/or the variation shift. We would like a sample point $(x_1, ..., x_n)'$ with value $u = x_{(1)}$ to be close to $\hat{\theta}_2 = 35.5$. Then when this u deviates sufficiently from the line $u = 35.5$, this indicates that this is due to a location shift. We also would like the variability v to be as close to zero as possible. If this sample point is out of the diagonal line and its position deviates sufficiently from the horizontal line, this indicates that is caused by the variation. We plot the sample values of u and v for these 12 samples and the diagonal line in Figure 21.

There are two out of control sample points, numbered 2 and 6 with $(u, v) = (207.0, 383.5)$ and $(334.5, 175.9)$ respectively, indicated both with location shift and variation shift. This figure is interesting for that we may detect the shifts shown in various classical Shewhart control charts.$\square$

Figure 21: χ^2 control chart for Particle Count, revised limit



4.3 Approximate Density Control Charts

We now revise the approximate HDS test to construct control chart. There are some reasons for the need to construct approximate density control chart. First, when we treat the density function $L(X, f_0)$ as a random variable, the quantile point $\ell(f_0)$ of (15) is not easy to be derived for some reasons such as that the distribution of random density $L(X, f_0)$ is not known. This leads to the fact that the exact density control chart is not attainable. Some other lifetime characteristics are the examples of this type. Second, when we deal with random variables having discrete distributions, to construct density control chart is more convenient through the approximation methods.

Let $X_1, ..., X_n$ be a random sample from the distribution that $H_0 : f = f_0$ is true, where $f_0 = f(x, \theta_1, ..., \theta_p)$. We derive a technique of approximate density control chart through (9) in section 4.1. If both $E[ln\ f_0(X)]$ and $Var[ln\ f_0(X)]$ exist, we choose a constant $\ell(f_0)$ such that

$$1 - \alpha = \Phi(\frac{n^{-1} ln\ \ell(f_0) - E[ln\ f_0(X)]}{\sqrt{n^{-1} Var[ln\ f_0(X)]}}), \tag{23}$$

where Φ is the distribution function of the standard normal distribution. From (23), an approximate density control chart is

$$LCL = e^{n\{E[ln\ f_0(X)]+z_\alpha\sqrt{n^{-1}Var[ln\ f_0(X)]}\}}$$
$$\text{Tracking variable: } L(x, f_0)$$

In the same way, let $\bar{\theta}_1, ..., \bar{\theta}_p$ be the estimates of these parameters, respectively, from the m samples. Then an approximate density control chart has the framework as

$$LCL = e^{n\{E[ln\ f(X,\bar{\theta}_1,...,\bar{\theta}_p]+z_\alpha\sqrt{n^{-1}Var[ln\ f(X,\bar{\theta}_1,...,\bar{\theta}_p]}\}}$$
$$\text{Tracking variable: } L(x, \bar{\theta}_1, ..., \bar{\theta}_p) \tag{24}$$

For given observation x, if its density value $L(x, \bar{\theta}_1, ..., \bar{\theta}_p)$ lies below the lower density control limit LCL of (24), then this observation is suspicious to be generated from a process of statistical out of control. We explain this approach with an example.

It is of interest to compare the exact and approximate density control charts in simulation. We compute the average run length (ARL) for this comparison. Assuming that the characteristic has a normal distribution $N(\mu, \sigma^2)$ with in-control mean $\bar{\bar{x}} = \mu_0$ and standard deviation $\bar{s} = \sigma_0$. From the exact chi-square control chart in previous Section 4.2.2, we may see that the ARL when the underlying normal distribution has mean μ and variance σ^2 is

$$ARL_{exact} = \frac{1}{P\left(\chi^2_{n,ncp=\frac{n(\mu-\mu_0)^2}{\sigma^2}} \geq \frac{\sigma_0^2\chi^2_{n,1-\alpha}}{\sigma^2}\right)}.$$

On the other hand, the approximate density control chart for normal distribution may be formulated as

$$LCL = -\tfrac{n}{2}log(2\pi\sigma_0^2) - \tfrac{n}{2} + z_{1-\alpha}\tfrac{n}{\sqrt{2n}}$$
$$\text{Test statistic function: } -\tfrac{n}{2}log(2\pi\sigma_0^2) - \frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2\sigma_0^2}$$

We may derive the ARL for this approximate density control chart as

$$ARL_{Appro} = \frac{1}{P(\chi^2_{n,ncp=\frac{n(\mu-\mu_0)^2}{\sigma^2}} \geq \frac{\sigma_0^2}{\sigma^2}(n - \sqrt{2n}z_{1-\alpha}))}$$

We let $1 - \alpha = 0.9973$ and $\mu_0 = 0$ and $\sigma_0 = 1$ for computing the two ARL's. The results are displayed in the following tables.

Table 11. $\begin{pmatrix} ARL_{exact} \\ ARL_{appro} \end{pmatrix}$ for normal distribution ($n = 5$).

	$\mu = 0$	$\mu = 0.5$	$\mu = 1.0$	$\mu = 2.0$	$\mu = 3.0$
$\sigma = 1.00$	$\begin{pmatrix} 370.37 \\ 59.01 \end{pmatrix}$	$\begin{pmatrix} 91.75 \\ 20.88 \end{pmatrix}$	$\begin{pmatrix} 12.10 \\ 4.61 \end{pmatrix}$	$\begin{pmatrix} 1.34 \\ 1.12 \end{pmatrix}$	$\begin{pmatrix} 1.00 \\ 1.10 \end{pmatrix}$
$\sigma = 1.25$	$\begin{pmatrix} 25.07 \\ 8.62 \end{pmatrix}$	$\begin{pmatrix} 13.75 \\ 5.61 \end{pmatrix}$	$\begin{pmatrix} 4.83 \\ 2.64 \end{pmatrix}$	$\begin{pmatrix} 1.30 \\ 1.12 \end{pmatrix}$	$\begin{pmatrix} 1.01 \\ 1.10 \end{pmatrix}$
$\sigma = 1.50$	$\begin{pmatrix} 6.61 \\ 3.41 \end{pmatrix}$	$\begin{pmatrix} 4.99 \\ 2.81 \end{pmatrix}$	$\begin{pmatrix} 2.85 \\ 1.90 \end{pmatrix}$	$\begin{pmatrix} 1.25 \\ 1.12 \end{pmatrix}$	$\begin{pmatrix} 1.01 \\ 1.00 \end{pmatrix}$
$\sigma = 2.00$	$\begin{pmatrix} 2.11 \\ 1.58 \end{pmatrix}$	$\begin{pmatrix} 1.96 \\ 1.51 \end{pmatrix}$	$\begin{pmatrix} 1.65 \\ 1.35 \end{pmatrix}$	$\begin{pmatrix} 1.17 \\ 1.09 \end{pmatrix}$	$\begin{pmatrix} 1.02 \\ 1.01 \end{pmatrix}$
$\sigma = 2.50$	$\begin{pmatrix} 1.40 \\ 1.22 \end{pmatrix}$	$\begin{pmatrix} 1.37 \\ 1.20 \end{pmatrix}$	$\begin{pmatrix} 1.29 \\ 1.16 \end{pmatrix}$	$\begin{pmatrix} 1.11 \\ 1.06 \end{pmatrix}$	$\begin{pmatrix} 1.03 \\ 1.01 \end{pmatrix}$
$\sigma = 3.00$	$\begin{pmatrix} 1.18 \\ 1.10 \end{pmatrix}$	$\begin{pmatrix} 1.17 \\ 1.09 \end{pmatrix}$	$\begin{pmatrix} 1.14 \\ 1.08 \end{pmatrix}$	$\begin{pmatrix} 1.07 \\ 1.04 \end{pmatrix}$	$\begin{pmatrix} 1.02 \\ 1.01 \end{pmatrix}$

The ARLs for the approximate density control chart are all smaller than the corresponding ARLs of the excat density control chart. Although we choose $1-\alpha = 0.9973$ for constructing both exact and approximate density control charts, their actual probability of type I errors are actually equal to 0.0027 only for the excat density chart where the approximate density control chart has relative larger probability of type I error than 0.0027.

4.3.1 Approximate Binomial Density Control Charts

When a quality characteristic is not a numerical variable, we usually classify each item inspected as either defective or nondefective to the specifications on that quality characteristic. Quality control charts for this type of characteristic are called the attributes control charts. Among them, one is dealing with number of defects observed, and it is called the *np* chart. We may analogously develop an approximate density control chart on number of defects.

Suppose that now we have a random sample $X_1, ..., X_n$ from Bernoulli distribution. The joint pdf of the sample is

$$L(x,p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i},$$

and then

$$E[lnf(X,p)] = (1-p)ln(1-p) + pln(p)$$

$$Var[lnf(X,p)] = p(1-p)(lnp - ln(1-p))^2.$$

Suppose that we have a training sample of m groups of sample size n from a statistical in control process. Let $\hat{p}_j$ be the sample mean of the j-th group. We then denote $\bar{p} = \frac{1}{m}\sum_{j=1}^{m}\hat{p}_j$. The framework of an approximate density control chart is

$$LCL = e^{n\{E[lnf(X,\bar{p})]+z_\alpha\sqrt{n^{-1}Var[lnf(x,\bar{p})]}\}}$$

Tracking variable: $L(x,\bar{p})$

4.3.2 Approximate Gamma Density Control Charts

Although the use of an exponential distribution to monitor the lifetime data makes the statistical inferences very simple, however, the applicability of this distribution is fairly limited due to the reason that its hazard function is a constant. There are some other distributions useful for lifetime data analysis, for examples, the Gamma, Weibull and Extreme value distributions. We consider one of them for constructing its density control chart. Suppose that we have a random sample $X_1, ..., X_n$ drawn from a Gamma distribution $\Gamma(\theta_1, \theta_2)$ with pdf

$$f(x,\theta_1,\theta_2) = \frac{1}{\Gamma(\theta_1)\theta_2^{\theta_1}} x^{\theta_1-1}e^{-\frac{x}{\theta_2}}, x > 0.$$

This distribution fits variety of lifetime data adequately. It also arises in some situations involving the exponential distribution. When we are considering construction of density control chart for Gamma random variable, the fact that the distribution of the random density $f(X,\theta_1,\theta_2)$ is not easy to solve leads us to consider an approximate one. The joint pdf of the random sample is

$$L(x,\theta_1,\theta_2) = \frac{1}{(\Gamma(\theta_1))^n\theta_2^{n\theta_1}}(\pi_{i=1}^{n} x_i)^{\theta_1-1}e^{-\frac{\sum_{i=1}^{n} x_i}{\theta_2}}.$$

By letting

$$m_1(\theta_1,\theta_2) = E[lnf(X,\theta_1,\theta_2)]$$

and

$$m_2(\theta_1,\theta_2) = E[(lnf(X,\theta_1,\theta_2))^2],$$

we have

$$m_1(\theta_1,\theta_2) = -ln(\Gamma(\theta_1)) - \theta_1 ln(\theta_2) + (\theta_1 - 1)E(ln(X)) - \theta_1$$

and

$$\begin{aligned} m_2(\theta_1,\theta_2) &= [ln(\Gamma(\theta_1))]^2 + \theta_1^2[ln(\theta_2)]^2 + (\theta_1-1)^2E[(ln(X))^2] + \theta_1 + \theta_1^2 + 2\theta_1 ln(\theta_2)ln(\Gamma(\theta_1)) \\ &\quad -2ln(\Gamma(\theta_1))(\theta_1-1)E(ln(X)) + 2ln(\Gamma(\theta_1))\theta_1 - 2\theta_1 ln(\theta_2)(\theta_1-1)E(ln(X)) \\ &\quad +2\theta_1^2 ln(\theta_2) - 2\frac{(\theta_1-1)}{\theta_2}E[Xln(X)]. \end{aligned}$$

We further let

$$Var(\theta_1,\theta_2) = m_2(\theta_1,\theta_2) - (m_1(\theta_1,\theta_2))^2.$$

In the same way, we assume that we have a training sample of m groups of size n from the statistical in control distribution, and $E(ln(X))$, $E[(ln(X))^2]$, and $E[Xln(X)]$ are not easy to compute. Then we may let

$$\begin{aligned} &\bar{\bar{\theta}}_1 = \frac{1}{m}\sum_{j=1}^m \bar{\theta}_{1j} \\ &\bar{\bar{\theta}}_2 = \frac{1}{m}\sum_{j=1}^m \bar{\theta}_{2j} \\ &\bar{\bar{E}}\,(ln(X)) = \frac{1}{mn}\sum_{j=1}^m\sum_{i=1}^n ln(x_{ij}) \\ &\bar{\bar{E}}\,[(ln(X))^2] = \frac{1}{mn}\sum_{j=1}^m\sum_{i=1}^n (ln(X_{ij}))^2 \\ &\bar{\bar{E}}\,[Xln(X)] = \frac{1}{mn}\sum_{j=1}^m\sum_{i=1}^n x_{ij}ln(x_{ij}). \end{aligned}$$

The framework of an approximate Gamma density control chart may be derived as the following:

$$LCL = e^{n\{m_1(\bar{\bar{\theta}}_1,\bar{\bar{\theta}}_2)+z_{1-\alpha}(n^{-1}Var(\bar{\bar{\theta}}_1,\bar{\bar{\theta}}_2)^{1/2}\}}$$

Tracking variable: $L(x,\bar{\bar{\theta}}_1,\bar{\bar{\theta}}_2)$

5 Future Study

5.1 Nuisance Parameters

Let $\theta = (\theta_1, \theta_2)'$ be a vector of parameters in the statistical model. We assumed that θ_1 is the parameter of interest and θ_2 represents the nuisance parameter. It has received much attention for testing hypothesis about θ_1 with θ_2 leaving unknown. For the examples in section 3.2 and 3.3, the null hypotheses are the cases that one parameter is tested and the other parameters are assumed to be known. In fact, these cases are not very common in the multiple parameter problems. Similarly, the HDS test may be revised to deal with this hypothesis testing with nuisance parameter problems.

By the definition, $p_{hd} = \int_{\left\{x: \frac{L(x,\theta_1,\theta_2)}{L(x_o,\theta_1,\theta_2)} \leq 1\right\}} L(x,\theta_1,\theta_2)dx$, and x_o is the observation. We have some possible solutions as followings:

(i) $\int_{\left\{x: \frac{L(x,\theta_1,\widehat{\theta_2})}{L(x_o,\theta_1,\widehat{\theta_2})} \leq 1\right\}} L(x,\theta_1,*)dx$, where * could be any assigned value.

(ii) $\int_{\left\{x: \frac{L(x,\theta_1,\widehat{\theta_2}(x))}{L(x_o,\theta_1,\widehat{\theta_2}(x))} \leq 1\right\}} L(x,\theta_1,\widehat{\theta_2}(x))dx$.

(iii) $\sup_{\theta_2} \int_{\left\{x: \frac{L(x,\theta_1,\theta_2)}{L(x_o,\theta_1,\theta_2)}\right\}} L(x,\theta_1,\theta_2)dx$.

5.2 Significance Test for Hypothesis of Distribution Function

We have discussed the HDS test for exponential random variables. That is, we have a random sample $Y_1, ..., Y_n$ with the corresponding observations $y_1, ..., y_n$ to test hypothesis $H_0 : Y \sim Exp(1)$. In this case, the p-value of the HDS test is

$$\begin{aligned} p_{hd} &= P(\textstyle\prod_{i=1}^n e^{-Y_i} \leq \prod_{i=1}^n e^{-y_i}) \\ &= P(\textstyle\sum_{i=1}^n Y_i \geq \sum_{i=1}^n y_i) \\ &= P(\Gamma(n,1) \geq \textstyle\sum_{i=1}^n y_i) \end{aligned}$$

where $\Gamma(n,1)$ is a random variable with Gamma distribution Gamma(n,1).

Thus we may extend the test based on joint pdf to product of distribution functions. Suppose we have a continuous random variable with distribution function, $F_\theta(x)$. It is known that the random variable $F_\theta(X)$ is a uniform random variable where θ is true parameter with distribution $U(0,1)$. Then, $-lnF_\theta(X)$ and $-ln(1-F_\theta(X))$ both have exponential distribution $Exp(1)$. Let $X_1, ..., X_n$ be random variables with identical distribution function F_θ, and $x_1, ..., x_n$ represent the sample observation. We may define a new significance test for hypothesis $H_0 : X \sim F_\theta$ with p-value as

$$p_F = P(\prod_{i=1}^{n} e^{-lnF_\theta(X_i)} \le \prod_{i=1}^{n} e^{-lnF_\theta(x_i)})$$

or

$$p_F = P(\prod_{i=1}^{n} e^{-ln(1-F_\theta(X_i))} \le \prod_{i=1}^{n} e^{-ln(1-F_\theta(x_i))})$$

Thus,

$$\begin{aligned} p_F &= P(\sum_{i=1}^{n} -lnF_\theta(X_i) \ge \sum_{i=1}^{n} -lnF_\theta(x_i)) \\ &= P(\Gamma(n,1) \ge \sum_{i=1}^{n} -lnF_\theta(x_i)) \end{aligned}$$

or

$$p_F = P(\Gamma(n,1) \ge \sum_{i=1}^{n} -ln\ (1-F_\theta(x_i))).$$

If we invert the former p_F above to construct confidence interval, the result is analogous to the confidence interval by pivotal-quantity method (Mood, 1974), which could provide another approach to determine the confidence interval.

Obviously, it is doubted that the application of HDS test in goodness-of-test could not work better than Kolmogorov-Smirnov goodness-of-test, but HDS test is an exact result. We need more works on this part in the future. In Kolmogorov-Smirnov Goodness-of-test, the asymptotic p-value is $1 - H(\sqrt{n} \sup_{-\infty<x<\infty} |F_n(x) - F_\theta(x)|)$, where $H(t) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2t^2}$. Thus, we can also compare the K-S test to our new test.

There is also another question to decide one from above two versions of p-value. We have to develop an unified theory and perform a simulation study.

5.3 Incomplete Data

Missing or incomplete data frequently happens in many statistical applications, and the practitioners are usually supported by *ad hot* methods such as case deletion or imputation. Omitting or deletion could lead to discard large amounts of information or bias. For example, the EM algorithm would be the most popular method of point estimation, which is composed with E-step and M-step. The iterative algorithm is frequently applied for simulating multiple imputations of missing values, and then it will lead to hypothesis test or interval estimation.

Here we will propose another approach for incomplete data via the viewpoint of highest density. The simple cases of bivariate normal random variables are considered, and only some observations are incomplete. Let $X_1, X_2, ..., X_n$ be independently drawn from $N_k(\mu, \Sigma)$. Now the p-value for the HDS test in this case is

$$\begin{aligned} p_{hd} &= P(\prod_{i=1}^{n}(2\pi)^{-k/2}|\Sigma|^{-1/2}e^{(X_i-\mu_0)'\Sigma^{-1}(X_i-\mu_0)} \leq \prod_{i=1}^{n}(2\pi)^{-k/2}|\Sigma|^{-1/2}e^{(x_i-\mu_0)'\Sigma^{-1}(x_i-\mu_0)}) \\ &= P(\chi^2_{nk} \geq \sum_{i=1}^{n}(x_i-\mu_0)'\Sigma^{-1}(x_i-\mu_0)) \end{aligned}$$

We may modify the complete data model shown above to the case of incomplete model by replacing the joint pdf with the marginal density.

Besides, the lost follow-up is another case of incomplete data such as censoring in survival analysis. In parametric model of survival analysis, the likelihood function of right censored data is defined as $L(\theta, x) = \prod_{i=1}^{k} f_\theta(x_i) \prod_{j=1}^{c}(1 - F_\theta(x_i))$, where i is in the set of complete data and j is in the set of censored data. According to the definition of HDS test, the p-value of HDS test $H_0 : \theta = \theta_0$ for survival analysis can be expressed as

$$p_{hd} = P(\prod_{i=1}^{k} f_{\theta_0}(X_i) \prod_{j=1}^{c}(1 - F_{\theta_0}(X_j)) \leq \prod_{i=1}^{k} f_{\theta_0}(x_i) \prod_{j=1}^{c}(1 - F_{\theta_0}(x_j))).$$

We need further study this HDS test.

Appendix: p-values for Binomial HDS Test

Table A.1. p-value for binomial distribution under $H_0 : p = 0.5$ $(n = 2, ..., 5)$.

x	$n = 2$	3	4	5
0	0.5	0.25	0.125	0.0625
1	1	1	0.625	0.375
2			1	1

Table A.2. p-value for binomial distribution under $H_0 : p = 0.5$ $(n = 6, ..., 10)$.

x	6	7	8	9	10
0	0.0313	0.0156	0.0078	0.0039	0.0020
1	0.2188	0.1250	0.0703	0.0391	0.0215
2	0.6875	0.4531	0.2891	0.1797	0.1094
3	1	1	0.7266	0.5078	0.3438
4			1	1	0.7539
5					1

Table A.3. p-value for binomial distribution under $H_0 : p = 0.5$ $(n = 11, ..., 15)$.

x	$n = 11$	$n = 12$	13	14	15
0	0.0010	0.0005	0.0002	0.0001	0.0001
1	0.0117	0.0063	0.0034	0.0018	0.0010
2	0.0654	0.0386	0.0225	0.0129	0.0074
3	0.2266	0.1460	0.0923	0.0574	0.0352
4	0.5488	0.3877	0.2668	0.1796	0.1185
5	1	0.7744	0.5811	0.4240	0.3018
6		1	1	0.7905	0.6072
7				1	1

Table A.4. p-value for binomial distribution under $H_0 : p = 0.5$ ($n = 16, ..., 20$).

x	16	17	18	19	20
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0005	0.0003	0.0001	0.0001	0.0000
2	0.0042	0.0023	0.0013	0.0007	0.0004
3	0.0213	0.0127	0.0075	0.0044	0.0026
4	0.0768	0.0490	0.0309	0.0192	0.0118
5	0.2101	0.1435	0.0963	0.0636	0.0414
6	0.4545	0.3323	0.2379	0.1671	0.1153
7	0.8036	0.6291	0.4807	0.3593	0.2632
8	1	1	0.8145	0.6476	0.5034
9			1	1	0.8238
10					1

Table A.5. p-value for binomial distribution under $H_0 : p = 0.5$ ($n = 21, ..., 25$).

x	$n = 21$	$n = 22$	23	24	25
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0002	0.0001	0.0001	0.0000	0.0000
3	0.0015	0.0009	0.0005	0.0003	0.0002
4	0.0072	0.0043	0.0026	0.0015	0.0009
5	0.0266	0.0169	0.0106	0.0066	0.0041
6	0.0784	0.0525	0.0347	0.0227	0.0146
7	0.1892	0.1338	0.0931	0.0639	0.0433
8	0.3833	0.2863	0.2100	0.1516	0.1078
9	0.6636	0.5235	0.4049	0.3075	0.2295
10	1	0.8318	0.6776	0.5413	0.4244
11		1	1	0.8388	0.6900
12				1	1

Table A.6. p-value for binomial distribution under $H_0 : p = 0.5$ ($n = 26, ..., 30$).

x	$n = 26$	$n = 27$	28	29	30
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0005	0.0003	0.0002	0.0001	0.0001
5	0.0025	0.0015	0.0009	0.0005	0.0003
6	0.0094	0.0059	0.0037	0.0023	0.0014
7	0.0290	0.0192	0.0125	0.0081	0.0052
8	0.0755	0.0522	0.0357	0.0241	0.0161
9	0.1686	0.1221	0.0872	0.0614	0.0428
10	0.3269	0.2478	0.1849	0.1360	0.0987
11	0.5572	0.4421	0.3449	0.2649	0.2005
12	0.8450	0.7011	0.5716	0.4583	0.3616
13	1	1	0.8506	0.7111	0.5847
14			1	1	0.8555
15					1

References

[1] 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.

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

[3] Azzalini, A. (1996). *Stastistical Inference: Based on the likelihood*, London: Chapman & Hall.

[4] Barnard, G. A. (1967). The use of the likelihood function in statistical practice, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 1 (eds. L. LeCam and J. Neyman), Berkeley: University of California Press.

[5] Barnard, G. A. (1980). Pivotal inference and the Bayesian controversy (with discussion). *Bayesian Statistics* (J. M. Bemardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.). Valencia: University Press.

[6] Barrnett, V. D. (1966). Evaluation of the maximum-likelihood estimaotr where the likelhihood equation has multiple roots. *Biometrika*, 53, 151-165.

[7] Berger, J. O. and Wolpert, R. L. (1988). *The Likelihood Principle.* Institute of Mathmatical Statistics.

[8] Birnbaum, A. (1962). On the foundation of statistical inference. *Journal of American Statistical Association*, 57, 269-306.

[9] Casella, G. and Berger, R. L. (2001). *Statistical Inference*, 2nd edition. Duxbury Press: Belmont.

[10] Chao, M. T. and Cheng, S.W. (1996) Semicircle control chart for variables data. *Quality Engineering*, 8, 441-446.

[11] Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* Chapman & Hall: London.

[12] Donahue, R. M. J.(1999). A note on information seldom reported via the p value. *The American Statistician*, 53, 303-306.

[13] Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society, series B*, 85, 87-94.

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

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

[16] Gibbsons, J. D. and Pratt, J. W. (1975). P-values: interpretation and methodology. *The American Statistican*, 29, 20-25.

[17] Grimshaw, S. D. and Alt, F. B. (1997). Control charts for quntile function values. *Journal of Quality Technology,* 29, 1-7

[18] Hacking, I. (1965). *Logic of Statistical Inference.* New York: Cambridge University Press.

[19] Hass, G., Bain, L. and Antle, C. (1970). Inferences for the Cauchy distributed based on maximum likelihood estimators. *Biometrika,* 57, 403-408

[20] Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *Annals of Statistics*, 19, 2244-2253.

[21] Hoerl, R. W. and Palm, A. C. (1992). Discussion: intetrating SPC and APC. *Technometrics*, 34, 268-272.

[22] Huang, J.-Y., Chen, L.-A. and Welsh, A. H. (2006). Reference limits from the mode interval. Submitted for publication (In revision).

[23] Hung, H. M. J., O'Neill R. T., Bauer, P. and Köhne, K. (1997). The Behavior of The P-Value When The Aternative Hypothesis Is True. *Biometrics*, 53, 11-22.

[24] Hyndman, R. J. (1996). Computing and Graphing Highest Density Regions. *The American Statistician.* 50, 120-126.

[25] Kotz, S. and Johnson, N. L. (1982). *Encyclopedia of Statistical Sciences*. Wiley: New York.

[26] Lawless, J. F. (1972). Conditional Confidence Interval Procedures for The Location and Scale Parameters of The Cauchy and Logistic Distributions. *Biometrika*, 59, 337-386.

[27] Lehmann, E. L. and Casella, G. (1999). *Theory of Point Estimation* 2nd edition. Springer Verlag: New York.

[28] Levinson, W. A. and Polny, A. (1999). SPC for tool particle counts. *Semiconductor International*, 22, 117-122.

[29] Lindsey, J. K. (1996). *Parametric Statistical Inference*. Oxford Science Publication: New York.

[30] Mason, R. L., Champ, C. W., Tracy, N. D., Wierda, S. J. and Young, J. C. (1997). Assessment of Multivariate Process Control Techniques. *Journal of Quality Technology*, 29, 140-143.

[31] McCullagh, P. (2002). What Is Statistical Model? *The Annals of Statistics*, 30, 1225-1267.

[32] Montgomery, D.C., Runger, G. C. and Hubel, N. F. (2004). *Engineering Statistics*. New York: Wiley.

[33] Mood, A. M., Graybill, F. A., and Boes, D. C. (1974).*Introduction to the Theory of Statistics* 3rd edition. McGraw-Hill: Singapore.

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

[35] Reid, N. (2000). Likelihood. *Journal of the American Statistical Association*, 95, 1335-1340.

[36] Robins, J. and Wasserman, L. (2000). Conditioning, Likelihood, and Coherence: A Review of Some Foundational Concepts. *Journal of the American Statstical Association*, 95, 1340-1346.

[37] Royall, R. (1997). *Satistical Evidence: A Likelihood Paradigm.* Chapman & Hall/CRC.

[38] Royall, R. (2000). On the Probabitily of Observing Misleading Statistical Evidence. *Journal of the American Statistical Association*, 95, 760-768.

[39] Schervish, M. J. (1996). *P* Values: What They Are and What They Are Not. *The American Statistician*, 50, 203-206.

[40] Sackrowitz, H. and Samuel-Cahn, E. (1999). *P* Values as Random Variables—Expected *P* Values. *The American Statistician,* 53, 326-331.

[41] Silvey, S. D. (1975). *Statistical Inference.* Chapman and Hall: London.

[42] Spanos, A. (1999). *Probability Theory and Statistical Inference: Econometric Modeling with Observation Data.* Cambridge University Press.

[43] Spiring, F. A. and Cheng, S. M. (1998). An atlernate variables control chart: the univariate and multivariate case. *Statistica Sinica*, 8, 273-287.

[44] Tsou, T. S. and Royall, R. M. (1995). Robust Likelihoods. *Journal of American Statistical Association*, 90, 316-320.

[45] Vapnik, V. N. (1998). *Statistical Learning Theory.* Wiley: New York.

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

[47] Wierda, S. J. (1994). Multivariate statistical process control - recent results and directions for future research. *Statistica Neerlandica*, 48, 147-168.

[48] Woodall, H. W. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology*, 32, 341-350.