

INSEGAMIENTO ESTRICTO Y ERRORES DE DECISI3N
EN EL PROBLEMA DE PRUEBA DE HIP3TESIS

MARÍA DEL CARMEN RUIZ MORENO

TESIS

Presentada como Requisito Parcial para
Obtener el Grado de:

MAESTRA EN
ESTADÍSTICA APLICADA

UNIVERSIDAD AUT3NOMA AGRARIA
“ANTONIO NARRO”

Buenavista, Saltillo, Coahuila, M3xico
Junio de 2013

Universidad Autónoma Agraria Antonio Narro
Subdirección de Postgrado

INSEGAMIEN TO ESTRICTO Y ERRORES DE DECISIÓ N
EN EL PROBLEMA DE PRUEBA DE HIPÓTESIS

TESIS

Por:

MARÍA DEL CARMEN RUIZ MORENO

Elaborada bajo la supervisión del comité particular de asesoría
y aprobada como requisito parcial para optar al grado de

MAESTRA EN
ESTADÍSTICA APLICADA

Comité Particular

Asesor principal:



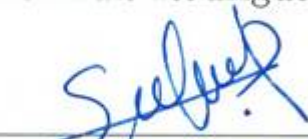
Dr. Rolando Cavazos Cadena

Asesor:



M. C. Luis Rodríguez Gutiérrez

Asesor:



M. C. Félix de Jesús Sánchez Pérez



Dr. Fernando Ruiz Zárate
Subdirector de Postgrado
Buena vista, Saltillo, Coahuila, Junio de 2013

Acknowledgement

Es un placer agradecer a la Universidad Autónoma Agraria Antonio Narro, por la oportunidad que me brindó para estudiar Estadística en su programa de Graduados.

COMPENDIO

INSEGAMIEN TO ESTRICTO Y ERRORES DE DECISIÓN EN EL PROBLEMA DE PRUEBA DE HIPÓTESIS

Por

MARÍA DEL CARMEN RUIZ MORENO

MAESTRÍA EN

ESTADÍSTICA APLICADA

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, Junio de 2013

Dr. Rolando Cavazos Cadena –Asesor–

Palabras clave: Modelos estadísticos paramétricos, Pruebas uniformemente más potentes, Pruebas de Neyman-Pearson, Propiedad de insegmento, Errores de decisión, Cota inferior para la suma de probabilidades de decisiones incorrectas.

Este trabajo trata sobre dos aspectos fundamentales en la teoría de *prueba de hipótesis* para modelos estadísticos paramétricos, a saber, (i) la posibilidad de cometer errores al decidir sobre la validez de una hipótesis dada o su hipótesis complementaria, y (ii) la noción de insegmento de una prueba. Los principales objetivos son (i) establecer una cota inferior para las probabilidades de decidir incorrectamente entre dos hipótesis complementarias dadas, y (ii) demostrar que una prueba de Neyman-Pearson es insegmenta en el sentido estricto. La exposición inicia con una descripción breve del problema de prueba de hipótesis en modelos paramétricos, para continuar con una discusión de los posibles errores en que se puede incurrir al tratar de determinar la validez de una hipótesis o de su complementaria basándose en una observación aleatoria; el resultado principal que se obtiene es una cota inferior para la suma de los posibles errores de decisión, mostrando que dicha cota es, generalmente, positiva y que, bajo condiciones menores, es igual a uno. Adicionalmente, se discute la construcción de pruebas de Neyman-Pearson para decidir entre dos hipótesis simples y se establece el insegmento estricto de dichas pruebas.

ABSTRACT

STRICT UNBIASEDNESS AND DECISION ERRORS
IN HYPOTHESIS TESTING

BY

MARÍA DEL CARMEN RUIZ MORENO

MASTER IN

APPLIED STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, June, 2013

Dr. Rolando Cavazos Cadena –Advisor–

Key Words: Parametric statistical models, Uniformly most powerful tests, Neyman-Pearson tests, Unbiasedness property, Decision Errors, Lower bound for the sum of probabilities of incorrect decisions.

This work is concerned with two basic aspects in the theory of *hypothesis testing* in the context of parametric statistical models, namely, (i) the possibility of taking incorrect decisions when the validity of one of two complementary hypothesis is assessed, and (ii) the idea of *unbiased test*. The main objectives of the thesis are (i) to establish a lower bound for the probabilities of deciding incorrectly between to given complementary hypothesis, and (ii) to prove that a Neyman-Pearson test is strictly unbiased. La exposition begins with a brief description of the testing problem for parametric models, and continues discussing the potential errors that can occur when a random observation vector is used to decide between to complementary hypothesis; the main result established in this direction is that the sum of the probabilities of the possible errors is bounded below by a number that, generally, is positive and, moreover, under mild conditions it is shown that such a bound is equal to 1. Additionally, the existence and construction of a Neyman-Pearson test is discussed, and it is proved that such a test is strictly unbiased.

Contents

1. A General Perspective of This Work	1
1.1 Introduction	1
1.2 Testing a Parametric Hypothesis	2
1.3 Main Objectives and Contributions	3
1.4 The Origin of This Work	4
1.5 The Organization	5
2. Test of Hypothesis	7
2.1 Statistical Models	7
2.2 Statistical Hypothesis	9
2.3 Simple, Composite and Complementary Hypothesis	10
2.4 Testing Procedures	12
2.5 Randomized Tests	14
2.6 Examples	16
3. Decision Errors	19
3.1 Errors and a Lower Bound	19
3.2 Positivity of the Lower Bound	20
3.3 An Improved Result	22
3.4 Null Hypothesis	24
3.5 Examples	27
4. Neyman-Pearson Theory	39
4.1 Likelihood Ratio Tests	39
4.2 Existence of Neyman-Pearson tests	41
4.3 Construction of Neyman-Pearson Tests	43
4.4 Examples	45

5. Unbiasedness Property	51
5.1 Unbiased Tests	51
5.2 Dependence on a Sufficient Statistic	52
5.3 Additional Examples	54
5.4 Strict Unbiasedness	60
References	66

Chapter 1

A General Perspective of This Work

This chapter presents a brief outline of this work, establishing the main objectives and describing the organization of the subsequent material. The main contributions are clearly stated.

1.1. Introduction

This work deals with the problem of *hypothesis testing* in the context of parametric statistical models, which lays in the core of the classical statistical methodology. For instance, the most basic statistical technique used in agriculture is the so called *analysis of variance*, which is widely used to compare the effects of diverse treatments applied to experimental units; when an analysis of variance is performed, the first objective is to decide whether or not the treatments produce different effects on the experimental units, that is, *to test* the claim of absence of treatments effects. In the area of *quality control*, an important problem is *to test* that the real characteristic \mathbf{X} of a product has the value δ stipulated by the design, and that such a characteristic does not change substantially among the different products. Two graphical devices are used in industry—the $\bar{\mathbf{X}}$ and R charts—to decide if the real characteristic coincides in the average with δ and to check if the variability of the relevant characteristic is not larger than the admissible tolerance. Every time that those charts are used, a test of a statistical hypothesis is performed (Montgomery, 2011).

On the other hand, due to its practical and theoretical relevance, every treatise on Statistics dedicates a good amount of space to study the problem of hypothesis testing, analyzing the construction of statistical tests with optimality properties (Dudewicz and Mishra 1998, Wackerly *et al.* 2009, Lehmann and Romano 1999, or Graybill 2000).

The topics studied in the following chapters are mainly concentrated on three aspects of the theory of hypothesis testing:

- (i) The analysis of the errors that can occur when a statistical test is used;
- (ii) The existence and construction of optimal tests for a given problem of hypothesis testing, and
- (iii) The dependence of an optimal statistical test on sufficient statistics.

Each one of the results presented below are illustrated by detailed examples.

1.2. Testing a Parametric Hypothesis

The purpose of a statistical analysis is to use the observed data *to gain knowledge* about some unknown aspect of the process generating the observations. The observable data $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is thought of as a random vector whose distribution is not completely known. Rather, theoretical or modeling considerations lead to assume that the distribution of \mathbf{X} , say $P_{\mathbf{X}}$, belongs to a certain family \mathcal{F} of probability measures defined on (the Borel class of) \mathbb{R}^n :

$$P_{\mathbf{X}} \in \mathcal{F}. \quad (1.2.1)$$

This is a statistical model, and in any practical instance it is necessary to include a precise definition of the family \mathcal{F} . In this work, the main interest concentrates on *parametric models*, for which the family \mathcal{F} can be indexed by a k -dimensional vector θ whose components are real numbers; in such a case the set of possible values of θ , which is referred to as the parameter space, will be denoted by Θ and \mathcal{F} can be written as

$$\mathcal{F} = \{P_{\theta} \mid \theta \in \Theta\}.$$

In this context the model (1.2.1) ensures that *there exists some parameter* $\theta^* \in \Theta$ such that $P_{\mathbf{X}} = P_{\theta^*}$, that is, for every (Borel) subset A of \mathbb{R}^n

$$P[X \in A] = P_{\mathbf{X}}[A] = P_{\theta^*}[A]. \quad (1.2.2)$$

The parameter θ^* satisfying this relation for every (Borel) subset of \mathbb{R}^n is *the true parameter value*. Notice that the model prescribes the existence of $\theta^* \in \Theta$ such that the above equality always holds, but does not specify which is the parameter θ^* ; it is only supposed that θ^* belongs to the parameter space Θ . Indeed, the lack of exact knowledge of θ^* represents ‘the aspects that are not completely known’ to the analyst about the real process generating the observation vector \mathbf{X} . One way of getting knowledge about the unknown true parameter value θ^* can be described as follows: Suppose that the researcher has reasons to think that θ^* belongs to a certain subset Θ_0 of the parameter space Θ , *i.e.*, that

$$\mathcal{H}: \theta^* \in \Theta_0; \quad (1.2.3)$$

such an statement is a (parametric) hypothesis and the objective of the analyst is to use the observed values of the data vector \mathbf{X} to decide whether or not \mathcal{H} occurs. If based on the observed value of \mathbf{X} it is decided that \mathcal{H} is a correct statement, that is, that $\theta^* \in \Theta_0$, then knowledge is gained about θ^* : originally, it was known that $\theta \in \Theta$, but after the observation of \mathbf{X} it is declared that θ belongs to the smaller subset Θ_0 , so that *the statement $\theta \in \Theta_0$ represents less uncertainty than $\theta \in \Theta$* . On the other hand, if the observation \mathbf{X} leads the analyst to conclude that \mathcal{H} is false, then it will be declared that θ^* does not belong to Θ_0 ; since θ is assumed to be a member of Θ , the conclusion is that the statement

$$\mathcal{K}: \theta \in \Theta \setminus \Theta_0$$

holds; since the set $\Theta \setminus \Theta_0$ is smaller than Θ_0 , the above statement also represents less uncertainty than the original inclusion $\theta \in \Theta_0$. In short, assuming that after the observation of X the analyst can declare that \mathcal{H} is a correct or an incorrect statement, knowledge about Θ will be improved after the experiment. However, in practice a definitive conclusion about the validity of a hypothesis like \mathcal{H} above can not be established, that is, it is possible to take an incorrect decision, and two errors are possible:

- To declare that $\mathcal{H}: \theta \in \Theta_0$ is false when it is true, and
- To declare that $\mathcal{K}: \theta \in \Theta \setminus \Theta_0$ is false when it is true.

The (largest) probabilities of incurring in these errors are denoted by $\mathcal{E}(\mathcal{H})$ and $\mathcal{E}(\mathcal{K})$, respectively, and it will be shown explicitly that both error probabilities $\mathcal{E}(\mathcal{H})$ and $\mathcal{E}(\mathcal{K})$ can not be made simultaneously small. Indeed, a positive lower bound for the sum $\mathcal{E}(\mathcal{H}) + \mathcal{E}(\mathcal{K})$ will be obtained, so that if $\mathcal{E}(\mathcal{H})$ is ‘small’, then $\mathcal{E}(\mathcal{K})$ is necessarily ‘large’. It follows that designing a rule to decide which of \mathcal{H} or \mathcal{K} is valid is a complex and interesting problem, a fact that provides the motivation for this work.

1.3. Main Objectives and Contributions

As already mentioned, this note is concerned with the hypothesis testing problem which can be roughly described as follows: Given a statistical model for an observation vector \mathbf{X} whose distribution depends on an unknown parameter θ , it is desired to decide whether or not a statement of the form $\mathcal{H}: \theta \in \Theta_0$ is valid. Any rule based on the observed value of \mathbf{X} that is used to take a decision about the validity of \mathcal{H} is called a *test*, and *the general objective of this work is to study the construction and properties of a ‘good’ test*.

The main specific goals of the thesis are as follows:

- (i) To establish a lower bound for the (largest) probabilities of taking an incorrect decision when the validity of a hypothesis $\mathcal{H}: \theta \in \Theta_0$ or the complementary statement $\mathcal{K}: \theta \in \Theta \setminus \Theta_0$ are being assessed. The lower bound established below is explicit and holds for every possible decision rule (test) that is used by the analyst.

- (ii) To prove that, for the testing problem $\mathcal{H}: \theta = \theta_0$ versus $\mathcal{K}: \theta = \theta_1$, the optimal test is strictly unbiased; the unbiasedness property is a desirable property of any test, and the result in this direction is an improvement of the conclusions usually stated in classical treatises.
- (iii) To illustrate the diverse ideas studied below using carefully analyzed examples.

The *main contributions* presented below can be briefly described as follows: As for the first objective, it is established that the largest error probabilities $\mathcal{E}(\mathcal{H})$ and $\mathcal{E}(\mathcal{K})$ satisfy

$$\mathcal{E}(\mathcal{H}) + \mathcal{E}(\mathcal{K}) \geq \ell(\mathcal{H}, \mathcal{K})$$

where $\ell(\mathcal{H}, \mathcal{K})$ is a generally positive number depending on the model, but not on the test used to decide whether or not \mathcal{H} is a valid claim; moreover, it is shown that under the mild condition that the sets Θ_0 and $\Theta \setminus \Theta_0$ share a boundary point, such a lower bound is equal to 1. Concerning the second objective, the main conclusion establishes that, for the problem of testing $\mathcal{H}: \theta = \theta_0$ versus $\mathcal{K}: \theta = \theta_1$, the relation $\mathcal{E}(\mathcal{H}) < 1 - \mathcal{E}(\mathcal{K})$ occurs, improving the inequality $\mathcal{E}(\mathcal{H}) \leq 1 - \mathcal{E}(\mathcal{K})$ that is usually presented in the literature. Finally, it is worth mentioning that the examples presented below are analyzed in full detail, paying attention to all relevant technical arguments, a feature that makes the presentation self-contained.

1.4. The Origin of This Work

This work may be considered as a product of the activities developed in the project *Mathematical Statistics: Elements of Theory and Examples*, started on July 2011 by the Graduate Program in Statistics at the Universidad Autónoma Agraria Antonio Narro; the founder students were Mary Carmen Ruiz Moreno and Alfonso Soto Almaguer, and other participants include Ana Paula Isais, Alberto Aguilar, María Elena Berlanga, Julieta Bautista, and María de Jesús Pinales Chávez. The basic aims of the project are:

- (i) To be a framework where statistical problems can be freely and fruitfully discussed;
- (ii) To promote the *understanding* of basic statistical and analytical tools through the analysis and detailed solution of exercises.
- (iii) To develop the *writing skills* of the participants, generating an organized set of neatly solved examples, which can be used by other members of the program, as well as by the statistical communities in other institutions and countries.
- (iv) To develop the *communication skills* of the students and faculty through the regular participation in seminars, where the results of their activities are discussed with the members of the program.

The activities of the project are concerned with fundamental statistical theory at an intermediate (non-measure theoretical) level, as in the book *Mathematical Statistics* by

Dudewicz and Mishra (1998). When necessary, other more advanced references that have been useful are Lehmann and Casella (1998), Borobkov (1999) and Shao (2010), whereas deeper probabilistic aspects have been studied in the classical text by Loève (1984). On the other hand, statistical analysis requires algebraic and analytical tools, and in these directions the basic references in the project are Apostol (1980), Fulks (1980), Khuri (2002) and Royden (2003), which concern mathematical analysis, whereas the algebraic aspects are covered in Graybill (2001) and Harville (2008). Initially, the project was concerned with the theory of Point Estimation and, starting on July 2012, the Theory of Hypothesis Testing has been studied; these work, as well as the one in Pinales Chávez (2013), may be considered as a product of all the activities of the project during the last year, and reflect the discussions and different perspectives of analysis of all the participants. In particular, it is a real pleasure to thank to my classmate, Alfonso Soto Almaguer, by his generous help and clever suggestions.

1.5. The Organization

The remainder of this work has been organized as follows:

In Chapter 2 the basic ideas and terminology involved in the study of hypothesis testing are introduced. The presentation begins with a brief description of a parametric statistical model, and then the notion of statistical hypothesis is introduced. Next, after presenting the ideas of simple, composite and complementary hypothesis, the notion of (deterministic, non-randomized) *test* is formally defined, and the exposition concludes studying the concept of randomized test.

Next, in Chapter 3 the possibility of taking an incorrect decision when testing a hypothesis \mathcal{H} is studied. The main conclusion to be established below is that, regardless of the test γ used by the analyst, the possibility of an erroneous decision can not be generally avoided. The main contributions of this work are stated in Theorem 3.1, and Lemmas 3.2.1 and 3.3.1.

Then, Chapter 4 is concerned with parametric statistical models whose parameter space consists of two points θ_0 and θ_1 , and the problem of testing the hypothesis $\mathcal{H}_0: \theta = \theta_0$ versus $\mathcal{H}_1: \theta = \theta_1$ is studied. For this problem, the existence of an optimal (most powerful) test is rigorously established and it is proved that such a test is unbiased.

Finally, the exposition concludes in Chapter 5 analyzing two basic properties that are satisfied by a most powerful test. The first one concerns a natural and intuitive condition: Under a reasonable test, the probability of rejecting the null hypothesis when it is false must be at least equal to the probability of rejection when \mathcal{H}_0 is true; a test with this property is called *unbiased*. The second property refers to the information that must be gathered in order to apply a most powerful test. The main result in this direction establishes that,

when a simple hypothesis is being tested versus another simple hypothesis, a most powerful test can be always specified in terms of a sufficient statistic. The main contribution of this chapter is Theorem 5.4.1, establishing that a Neyman-Pearson test is strictly unbiased.

Chapter 2

Test of Hypothesis

In this chapter the basic ideas and terminology involved in the study of hypothesis testing are introduced. The presentation begins with a brief description of a parametric statistical model, and then the notion of statistical hypothesis is introduced. Next, after presenting the ideas of simple, composite and complementary hypothesis, the notion of (deterministic, non-randomized) *test* is formally defined, and the exposition concludes studying the concept of randomized test.

2.1. Statistical Models

A statistical model for a random vector postulates that the distribution of the vector belongs to a certain family of probability measures. As it was already mentioned, in this work it will be assumed that the class of potential distributions of the observation vector \mathbf{X} is indexed by a parameter θ whose components are real numbers, and P_θ stands for the distribution associated with θ . Thus, a statistical model may be written as

$$\mathbf{X} \sim P_\theta, \quad \theta \in \Theta, \tag{2.1.1}$$

where $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is the observation vector and, for each θ in the parameter space $\Theta \subset \mathbb{R}^k$, the probability measure P_θ is defined on (the Borel sets of) \mathbb{R}^n . This model prescribes that the distribution $P_{\mathbf{X}}$ of the observation vector \mathbf{X} is P_θ for some $\theta \in \Theta$, that is,

there exists $\theta \in \Theta$ such that for ‘any set $A \subset \mathbb{R}^n$ ’ the probability that \mathbf{X} belongs to A is $P_{\mathbf{X}}[A] = P_\theta[A]$.

More explicitly, there exists $\theta \in \Theta$ such that

(i) If P_θ is a continuous distribution with density $f(\mathbf{x}; \theta)$, then

$$P_X[A] = P[\mathbf{X} \in A] = \int_A f(\mathbf{x}; \theta) d\mathbf{x},$$

and

(ii) If P_θ is a discrete distribution with probability function $f(\mathbf{x}; \theta)$, then

$$P_X[A] = P[\mathbf{X} \in A] = \sum_{\mathbf{x} \in A} f(\mathbf{x}; \theta).$$

The following basic *identifiability* condition is enforced throughout the remainder.

Assumption 2.1.1. For each $\theta_0, \theta_1 \in \Theta$ with $\theta_0 \neq \theta_1$, the corresponding distributions are different. More explicitly, if $\theta_0 \neq \theta_1$, then there exists a set A such that

$$P_{\theta_0}[A] \neq P_{\theta_1}[A].$$

In words, this condition establishes that, for the model (2.1.1), different parameters correspond to different distributions.

Remark 2.1.1. Usually, a statistical model is specified by giving a condition that allows to determine a family of probability measures containing the distribution of the observation vector \mathbf{X} . A common situation is as follows:

- (i) It is supposed that the components of \mathbf{X} are independent and identically distributed with common density or probability function $f(\cdot; \theta)$, and
- (ii) The unknown parameter θ is a member of a given set Θ ,

requirements that are summarized as

$$X_1, X_2, \dots, X_n \sim i.i.d. \ f(x; \theta), \quad \theta \in \Theta. \quad (2.1.2)$$

When the common density or probability function of the variables X_i is $f(x; \theta)$, the independence property in condition (i) yields that the distribution of \mathbf{X} is the probability measure P_θ specifying that, for each (Borel) set $A \subset \mathbb{R}^n$,

$$P_\theta[A] = \begin{cases} \int_A f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) d\mathbf{x}, & \text{if } f(\cdot; \theta) \text{ is a density,} \\ \sum_{\mathbf{x} \in A} f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta), & \text{if } f(\cdot; \theta) \text{ is a probability function,} \end{cases}$$

where

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \quad \text{and} \quad d\mathbf{x} = dx_1 dx_2 \cdots dx_n;$$

since condition (ii) establishes that θ is an arbitrary element of parameter space Θ , the statement (2.1.2) is equivalent to (2.1.1). On the other hand, the above conditions (i) and (ii) are frequently summarized as

$$\mathbf{X} \text{ is a sample of } f(x; \theta), \quad \theta \in \Theta,$$

so that this statement and (2.1.2) are equivalent; if the distribution corresponding to $f(\cdot; \theta)$ is denoted by $\mathcal{D}(\theta)$, then the above display and

$$\mathbf{X} \text{ is a sample of } \mathcal{D}(\theta), \quad \theta \in \Theta$$

are equivalent statements. □

2.2. Statistical Hypothesis

Notice that the specific parameter value θ that corresponds to the distribution of \mathbf{X} is *unknown*, because the model prescribes only that such a parameter belongs to the space Θ . The objective of performing the experiment rendering the vector \mathbf{X} is to use the observed value $\mathbf{X} = \mathbf{x}$ to obtain ‘information’ about the *precise* or *true* value of θ such that $P_{\mathbf{X}} = P_{\theta}$; the statistical procedures using the observed values of \mathbf{X} to obtain ‘conclusions’ about the true parameter value constitute that is called *statistical inference*. This chapter studies a particular form of statistical inference called *hypothesis testing* which can be described as follows:

Originally, the model prescribes that the distribution $P_{\mathbf{X}}$ of \mathbf{X} is a member of the family $\{P_{\theta}, \theta \in \Theta\}$, and the problem consists in deciding whether or not the observed values of \mathbf{X} support the assertion that the distribution of \mathbf{X} belongs to the smaller family $\{P_{\theta}, \theta \in \Theta_0\}$, where Θ_0 is a *subset* of Θ .

Definition 2.2.1. Consider the statistical model (2.1.1) for the observation vector \mathbf{X} . A hypothesis \mathcal{H} is a statement of the form

$$\mathcal{H} : \theta \in \Theta_0, \tag{2.2.1}$$

where Θ_0 is a subset of the parameter space Θ .

The hypothesis (2.2.1) is interpreted as the assertion that the distribution of \mathbf{X} belongs to the reduced family $\{P_{\theta}, \theta \in \Theta_0\}$, and represents conditions whose validity is of interest for the analyst

Example 2.2.1. To study the labor security in a local factory, the number of accidents X_1, X_2, \dots, X_n occurring in n weeks is recorded. Suppose that the variables X_i are independent and identically distributed random variables with Poisson distribution:

$$X_1, X_2, \dots, X_n \text{ is a sample of } \mathcal{P}(\theta), \quad \theta > 0.$$

In this case the parameter space is $\Theta = (0, \infty)$ and the (unknown) parameter θ is the expected number of accidents in a week. Several hypothesis might be of interest. For instance, suppose that in other countries the expected number of accidents in similar factories is 3.

(i) The condition that the security in the national factory is higher than the one in foreign factories is expressed by the hypothesis $\mathcal{H} : \theta \in (0, 3)$, that is,

$$\mathcal{H} : \theta < 3.$$

(ii) The hypothesis

$$\mathcal{H} : \theta = 3$$

is interpreted as the assertion that the security in the local factory is the same as in the foreign ones, whereas

(iii) The assertion that the security in the local factory is worse than the one in foreign factories is expressed as

$$\mathcal{H} : \theta > 3$$

In each one of these cases the problem is to see if the observed value of the variables X_1, X_2, \dots, X_n support the validity of the corresponding hypothesis. \square

2.3. Simple, Composite and Complementary Hypothesis

In this section a simple classification of hypothesis is formulated and the idea of complementary hypothesis is introduced.

Definition 2.3.1. A hypothesis $\mathcal{H} : \theta \in \Theta_0$ is *simple* if $\Theta_0 = \{\theta_0\}$ is a singleton, so that \mathcal{H} can be expressed as

$$\mathcal{H} : \theta = \theta_0,$$

whereas \mathcal{H} is *composite* when Θ_0 has two or more elements.

In the context of Example 2.2.1 the Hypothesis in parts (i) and (iii) are composite, whereas the Hypothesis $\mathcal{H} : \theta = 3$ in part (ii) is simple. The following examples discuss the ideas in Definition 2.3.1.

Exercise 2.3.1. Suppose that X_1, X_2, \dots, X_n is a sample of $\mathcal{N}(\mu, \sigma^2)$.

(i) If σ^2 is *known*, determine whether each hypothesis is simple or composite, where μ_0 is a known constant.

(a) $\mathcal{H} : \mu = \mu_0$;

(b) $\mathcal{H} : \mu > \mu_0$;

(c) $\mathcal{H} : \mu \neq \mu_0$;

(d) $\mathcal{H} : \mu \in I$, where I is a specified interval with positive length.

(ii) If σ^2 is *unknown*, determine whether each of the four hypothesis formulated above is simple or composite.

Solution. (i) Notice that the parameter is μ , and the parameter space is $(-\infty, \infty)$. In case (a) the value of the parameter is completely determined by the hypothesis, which can be written as $\mathcal{H} : \mu \in \{\mu_0\}$, and then the hypothesis is simple. However, in cases (b), (c) and (d), the hypothesis is composite, since it does not determine exactly the value of the parameter. For instance, \mathcal{H} can be written as

$$\mathcal{H} : \mu \in (\mu_0, \infty)$$

in case (b), and as

$$\mathcal{H} : \mu \in (-\infty, \mu_0) \cup (\mu_0, \infty)$$

in case (c).

(ii) In this context, the parameter is $\theta = (\mu, \sigma^2)$ and

$$\Theta = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\} = \mathbb{R} \times (0, \infty).$$

is the parameter space. Neither of the hypothesis determines the value of $\theta = (\mu, \sigma^2)$ exactly, and then all of the hypothesis are composite. For instance, \mathcal{H} can be written as $\mathcal{H} : \theta = (\mu, \sigma^2) \in \Theta_0 = \{0\} \times (0, \infty)$ in case (a), and then, since Θ_0 contains more than one element, it follows that \mathcal{H} is composite. \square

Remark 2.3.1. Other common formulation of the idea of simple hypothesis is the following:

A hypothesis \mathcal{H} is simple if, under \mathcal{H} , the distribution of the observation vector is uniquely determined.

Notice that if \mathcal{H} can be expressed as $\mathcal{H} : \theta = \theta_0$ (where θ_0 is known), then under \mathcal{H} the distribution of \mathbf{X} is uniquely determined and is equal to P_{θ_0} . \square

Exercise 2.3.2. Are the following hypothesis simple or composite?

(a) $\mathcal{H} : \mathbf{X}$ is a sample of an exponential distribution with some parameter θ ;

(b) If $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a sample of size n of $\mathcal{N}(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$,

$$\mathcal{H} : \frac{E_{\theta}[X_1]}{\text{Var}_{\theta}[X_1]} = c_0$$

where c_0 is a known constant.

(c) If $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a sample of size n of *Bernoulli*(θ) where $\theta \in (0, 1)$,

$$\mathcal{H} : \frac{E_\theta[X_1]}{\text{Var}_\theta[X_1]} = c_0$$

where c_0 is a known constant.

Solution. (a) Under \mathcal{H} the distribution of the variables X_i is exponential, but the parameter of the distribution is unknown; thus, under \mathcal{H} the distribution of the observation vector \mathbf{X} is not uniquely determined, and then \mathcal{H} is composite.

(b) Since $E_\theta[X_1] = \mu$ and $\text{Var}_\theta[X_1] = \sigma^2$, the hypothesis can be explicitly written as

$$\mathcal{H} : \frac{\mu}{\sigma^2} = c_0, \quad \text{or} \quad \mathcal{H} : \mu = c_0 \sigma^2.$$

Therefore, under \mathcal{H} the parameter $\theta = (\mu, \sigma^2)$ is not uniquely determined, and then \mathcal{H} is composite.

(c) Using that $E_\theta[X_1] = \theta$ and $\text{Var}_\theta[X_1] = \theta(1 - \theta)$, in the present case \mathcal{H} can be stated as

$$\mathcal{H} : \frac{\theta}{\theta(1 - \theta)} = c_0,$$

that is,

$$\mathcal{H} : \theta = 1 - \frac{1}{c_0}.$$

Consequently, under \mathcal{H} the distribution of \mathbf{X} is uniquely determined, so that \mathcal{H} is a simple hypothesis. \square

A natural companion to a hypothesis \mathcal{H} is the statement that the property stipulated by \mathcal{H} does not occur.

Definition 2.3.2. Let $X \sim P_\theta$, $\theta \in \Theta$ be a statistical model. Given a hypothesis $\mathcal{H} : \theta \in \Theta_0$, the corresponding *complementary* hypothesis is

$$\mathcal{K} : \theta \in \Theta \setminus \Theta_0.$$

Example 2.3.1. The following examples show a hypothesis \mathcal{H} and the corresponding complementary hypothesis \mathcal{K} .

(i) $\mathcal{H} : \mu > \mu_0$, $\mathcal{K} : \mu \leq \mu_0$.

(ii) $\mathcal{H} : \frac{\mu}{\sigma^2} = c_0$, $\mathcal{K} : \frac{\mu}{\sigma^2} \neq c_0$.

(iii) $\mathcal{H} : \mu + 1.96\sigma \leq 3$, $\mathcal{K} : \mu + 1.96\sigma > 3$. \square

2.4. Testing Procedures

As already mentioned, given a statistical hypothesis \mathcal{H} , the main objective of the analyst is to decide whether or not the observed data support the properties of the parameter stipulated by \mathcal{H} . The classical theory faces this problem as a decision between the validity of the original hypothesis \mathcal{H} and the complementary hypothesis \mathcal{K} , that is, if it is decided that the observed data support the validity of \mathcal{H} , then \mathcal{H} is ‘accepted’ and, consequently, \mathcal{K} is ‘rejected’; of course, when according to the criterion of the analyst the available data do not support \mathcal{H} , then \mathcal{H} is ‘rejected’ and \mathcal{K} is ‘accepted’.

Definition 2.4.1. Given a statistical model $\mathbf{X} \sim P_\theta$, $\theta \in \Theta$, consider a hypothesis $\mathcal{H} : \theta \in \Theta_0$ and the corresponding complementary hypothesis $\mathcal{K} : \theta \in \Theta \setminus \Theta_0$. Let \mathcal{X} be the set of possible values of the observation vector \mathbf{X} , that is,

$$\mathcal{X} = \{\mathbf{x} \mid f(\mathbf{x}; \theta) > 0 \text{ for some } \theta \in \Theta\}, \quad (2.4.1)$$

and notice that

$$P_\theta[\mathbf{X} \in \mathcal{X}] = 1, \quad \theta \in \Theta. \quad (2.4.2)$$

A test γ of \mathcal{H} versus \mathcal{K} is a partition of \mathcal{X} into two sets \mathcal{A} and $\mathcal{R} = \mathcal{X} \setminus \mathcal{A} = \mathcal{A}^c$, called the *acceptance and rejection* regions, respectively. When using γ , the decision taken after observing $\mathbf{X} = \mathbf{x}$ is denoted by $\gamma(\mathbf{x})$ and is determined as follows:

$$\gamma(\mathbf{x}) = \begin{cases} \text{Accept } \mathcal{H} \text{ (equivalently, Reject } \mathcal{K}), & \text{if } \mathbf{x} \in \mathcal{A} \\ \text{Reject } \mathcal{H} \text{ (equivalently, Accept } \mathcal{K}), & \text{if } \mathbf{x} \in \mathcal{R} \end{cases}$$

Thus, after observing \mathbf{X} , the analyst always ends up with a decision about the validity of \mathcal{H} ; the properties stipulated by \mathcal{H} are ‘accepted’ or ‘rejected’.

Example 2.4.1. Let $\mathbf{X} = (X_1, X_2, \dots, X_{16})$ be a sample of the *Poisson*(θ) distribution, where $\theta \in (0, \infty)$, and consider the hypothesis

$$\mathcal{H} : \theta \leq 12,$$

as well as the companion complementary hypothesis $\mathcal{K} : \theta > 12$. Observing that θ is the mean of the variables X_i , it is natural to estimate the unknown value of θ by the sample average $\bar{\mathbf{X}} = (X_1 + X_2 + \dots + X_{16})/16$, which is taken as an approximation of the true parameter value. Since according to \mathcal{H} the parameter does not exceed 12, it is natural to reject \mathcal{H} if $\bar{\mathbf{X}}$ is ‘substantially’ larger than 12, for instance if $\bar{\mathbf{X}} > 15$. These considerations lead to formulate the following procedure to reject or accept \mathcal{H} :

$$\gamma(\mathbf{X}) = \begin{cases} \text{Accept } \mathcal{H} : \theta \leq 12 \text{ (equivalently, Reject } \mathcal{K} : \theta > 12), & \text{if } \bar{\mathbf{X}} \leq 15 \\ \text{Reject } \mathcal{H} : \theta \leq 12 \text{ (equivalently, Accept } \mathcal{K} : \theta > 12), & \text{if } \bar{\mathbf{X}} > 15. \end{cases} \quad (2.4.3)$$

Notice that in the present case the set of possible values of the observation vector \mathbf{X} is

$$\mathcal{X} = \{\mathbf{x} = (x_1, x_2, \dots, x_{16}) \mid x_i \text{ is a nonnegative integer, } i = 1, 2, \dots, 16\},$$

and that the acceptance and rejection regions corresponding to the test γ are given by

$$\mathcal{A} = \{\mathbf{x} \in \mathcal{X} \mid \bar{x} \leq 15\}, \quad \mathcal{R} = \{\mathbf{x} \in \mathcal{X} \mid \bar{x} > 15\},$$

where $\bar{x} = (x_1 + x_2 + \dots + x_{16})/16$. □

Example 2.4.2. Suppose now that $\mathbf{X} = (X_1, X_2, \dots, X_{16})$ is a sample of the $\mathcal{N}(\theta, 1)$ distribution, where $\theta \in \mathbb{R}$ and, consider the hypothesis

$$\mathcal{H} : 10 \leq \theta \leq 12$$

as well as the complementary hypothesis

$$\mathcal{K} : \theta < 10 \quad \text{or} \quad \theta > 12.$$

Since θ is the mean of the variables X_i , the sample average $\bar{\mathbf{X}} = (X_1 + X_2 + \dots + X_{16})/16$ is a natural estimator of θ . On the other hand, because according to \mathcal{H} the true parameter lays between 10 and 12, it is natural to reject \mathcal{H} if $\bar{\mathbf{X}}$ is ‘substantially’ far from the interval $[10, 12]$, say if $\bar{\mathbf{X}} > 15$ or $\bar{\mathbf{X}} < 7$, and this leads to formulate the following testing procedure to reject or accept \mathcal{H} :

$$\gamma(\mathbf{X}) = \begin{cases} \text{Accept } \mathcal{H} : 10 \leq \theta \leq 12, & \text{if } 7 \leq \bar{\mathbf{X}} \leq 15 \\ \text{Reject } \mathcal{H} : 10 \leq \theta \leq 12, & \text{if } \bar{\mathbf{X}} < 7 \quad \text{or} \quad \bar{\mathbf{X}} > 15. \end{cases} \quad (2.4.4)$$

In this example the set of possible values of the observation vector \mathbf{X} is

$$\mathcal{X} = \{\mathbf{x} = (x_1, x_2, \dots, x_{16}) \mid x_i \in \mathbb{R} \text{ for all } i = 1, 2, \dots, 16\} = \mathbb{R}^{16},$$

whereas that the acceptance and rejection regions corresponding to the test γ are given by

$$\mathcal{A} = \{\mathbf{x} \in \mathcal{X} \mid 7 \leq \bar{x} \leq 15\}, \quad \mathcal{R} = \{\mathbf{x} \in \mathcal{X} \mid \bar{x} < 7 \text{ or } \bar{x} > 15\}$$

where, as before, \bar{x} is the average of the components of \mathbf{x} . □

2.5. Randomized Tests

In this section the class of tests in Definition 2.4.1 is generalized to include the so called *randomized tests*.

Definition 2.5.1. Let $\mathbf{X} \sim P_\theta$, $\theta \in \Theta$ be a statistical model, and let \mathcal{X} be the set of possible values of the observation \mathbf{X} ; see (2.4.2).

(i) A mapping $\varphi: \mathcal{X} \rightarrow [0, 1]$ is referred to as a *critical function*.

(ii) Consider a hypothesis $\mathcal{H} : \theta \in \Theta_0$ and the complementary hypothesis $\mathcal{K} : \theta \in \Theta \setminus \Theta_0$. The randomized test $\gamma_\varphi \equiv \gamma$ corresponding to a critical function φ is the random variable $\gamma(\mathbf{X})$ whose conditional distribution given \mathbf{X} is determined as follows: $\gamma(\mathbf{X})$ takes on the two values ‘Accept \mathcal{H} ’ and ‘Reject \mathcal{H} ’ with probabilities

$$P[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H} | \mathbf{X}] = \varphi(\mathbf{X}),$$

whereas

$$P[\gamma(\mathbf{X}) = \text{Accept } \mathcal{H} | \mathbf{X}] = 1 - \varphi(\mathbf{X}).$$

For a given critical function φ , the value $\varphi(\mathbf{X})$ may be thought of as the strength of the evidence to reject \mathcal{H} that, according to the analyst’s criterion, is provided by the observation vector \mathbf{X} .

Remark 2.5.1. (i) An observer using the randomized test γ_φ corresponding to a critical function φ operates as follows:

If $\mathbf{X} = \mathbf{x}$ is observed, then

- \mathcal{H} is rejected when $\varphi(\mathbf{x}) = 1$,
- \mathcal{H} is accepted if $\varphi(\mathbf{x}) = 0$,
- When $0 < \varphi(\mathbf{x}) < 1$, a Bernoulli experiment with success probability $\varphi(\mathbf{x})$ is performed, and if the result of the experiment is ‘success’, the hypothesis \mathcal{H} is rejected, whereas \mathcal{H} is accepted when the result is ‘failure’.

As it will be shown later, randomized tests have a great theoretical importance.

(ii) It is interesting to observe that the (non-randomized) tests in Definition 2.4.1 are a particular class of randomized tests. Indeed, let γ be a test in the sense of Definition 2.4.1, and let \mathcal{R} be the corresponding rejection region. Define $\varphi: \mathcal{X} \rightarrow [0, 1]$ as the indicator function of \mathcal{R} , that is,

$$\varphi(\mathbf{x}) = I_{\mathcal{R}}(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathcal{R} \\ 0, & \text{if } \mathbf{x} \in \mathcal{X} \setminus \mathcal{R} = \mathcal{A}. \end{cases}$$

Now, consider the randomized test γ_φ corresponding to this critical function φ , so that

$$P[\gamma_\varphi(\mathbf{X}) = \text{Reject } \mathcal{H} | \mathbf{X}] = \begin{cases} 1, & \text{if } \varphi(\mathbf{X}) = 1, \text{ that is, if } \mathbf{X} \in \mathcal{R}, \\ 0, & \text{if } \varphi(\mathbf{X}) = 0, \text{ i.e., if } \mathbf{X} \in \mathcal{A}. \end{cases}$$

It follows that the decision $\gamma_\varphi(\mathbf{X})$ is ‘Reject \mathcal{H} ’ when $\mathbf{X} \in \mathcal{R}$, and $\gamma_\varphi(\mathbf{X})$ is ‘Accept \mathcal{H} ’ if $\mathbf{X} \in \mathcal{A} = \mathcal{X} \setminus \mathcal{R}$; thus, γ_φ is the test γ with rejection region \mathcal{R} and acceptance region \mathcal{A} . This argument shows that the tests described in Definition 2.4.1 correspond to critical functions attaining exclusively the values 0 and 1. \square

2.6. Examples

In this section the notions of randomized and nonrandomized test are illustrated in the context of specific examples.

Example 2.6.1. (i) As in Example 2.4.1, let $\mathbf{X} = (X_1, X_2, \dots, X_{16})$ be a sample of the *Poisson*(θ) distribution, where $\theta \in (0, \infty)$, and consider the hypothesis $\mathcal{H} : \theta \leq 12$. The test in (2.4.3) corresponds to the critical function

$$\varphi(\mathbf{X}) = \begin{cases} 1, & \text{if } \bar{\mathbf{X}} > 15 \\ 0, & \text{if } \bar{\mathbf{X}} \leq 15; \end{cases}$$

using the notation of indicator functions, $\varphi(\mathbf{X}) = I_{(15, \infty)}(\bar{\mathbf{X}})$.

(ii) Suppose again that $\mathbf{X} = (X_1, X_2, \dots, X_{16})$ is a sample of the *Poisson*(θ) distribution with $\theta > 0$. An example of a randomized test for $\mathcal{H} : \theta \leq 12$ has the following critical function:

$$\varphi(\mathbf{X}) = \begin{cases} 1, & \text{if } \bar{\mathbf{X}} > 15 \\ 1/2 & \text{if } \bar{\mathbf{X}} = 15 \\ 0, & \text{if } \bar{\mathbf{X}} < 15. \end{cases}$$

The test γ_φ corresponding to this critical function rejects \mathcal{H} when $\bar{\mathbf{X}} > 15$ and accepts \mathcal{H} if $\bar{\mathbf{X}} < 15$; however, when $\bar{\mathbf{X}} = 15$ a Bernoulli experiment with success probability 1/2 is performed, and \mathcal{H} is rejected if a success is obtained, whereas a failure leads to acceptance of \mathcal{H} . With the notation of indicator function φ can be specified by

$$\varphi(\mathbf{X}) = I_{(15, \infty)}(\bar{\mathbf{X}}) + \frac{1}{2}I_{\{15\}}(\bar{\mathbf{X}}).$$

(iii) Continuing with Example 2.4.2, suppose that $\mathbf{X} = (X_1, X_2, \dots, X_{16})$ is a sample of the $\mathcal{N}(\theta, 1)$ distribution, where $\theta \in \mathbb{R}$ and, consider the hypothesis

$$\mathcal{H} : 10 \leq \theta \leq 12$$

and the complementary

$$\mathcal{K} : \theta < 10 \quad \text{or} \quad \theta > 12.$$

The test γ in (2.4.4) is associated to the critical function

$$\varphi(\mathbf{X}) = \begin{cases} 0, & \text{if } 7 \leq \bar{\mathbf{X}} \leq 15 \\ 1, & \text{if } \bar{\mathbf{X}} < 7 \quad \text{or} \quad \bar{\mathbf{X}} > 15; \end{cases}$$

notice that the test γ is non-randomized, and for this reason the critical function takes only the values 0 and 1. With the notation of indicator functions,

$$\varphi(\bar{\mathbf{X}}) = I_{(-\infty, 7) \cup (15, \infty)}(\bar{\mathbf{X}}).$$

(iv) In the context of the previous part, a ‘reasonable’ randomized test will be now described. Recall that $\bar{\mathbf{X}}$ is a ‘good’ estimator of the mean θ of the $\mathcal{N}(\theta, 1)$ distribution, and suppose that the analyst considers the distance between $\bar{\mathbf{X}}$ and $[7, 10]$ as a measure of the evidence provided by that data against \mathcal{H} ; as the distance gets larger, the evidence against \mathcal{H} increases. Moreover, assume that the analyst considers a distance of three or less units as an event due to chance, but a distance of size five or more units is regarded as a definitive evidence against \mathcal{H} . For distances from 3 to 5, the evidence against \mathcal{H} increases linearly from 0 to 1. These considerations lead to formulate the critical function of the analyst as follows:

$$\varphi(\mathbf{X}) = \begin{cases} 0, & \text{if } 7 \leq \bar{\mathbf{X}} \leq 15 \\ 1, & \text{if } \bar{\mathbf{X}} < 5 \text{ or } \bar{\mathbf{X}} > 17 \\ (7 - \bar{\mathbf{X}})/2, & \text{if } 5 \leq \bar{\mathbf{X}} < 7 \\ (\bar{\mathbf{X}} - 15)/2, & \text{if } 15 < \bar{\mathbf{X}} \leq 17. \end{cases}$$

This function φ attains values in $[0, 1]$ and the is a critical function determining a genuine test. \square

Remark 2.6.1. A test γ and the corresponding critical function determine each other, and they will be identified in the subsequent development. \square

To conclude this section, it will be shown that the (unconditional) probability of rejecting a hypothesis \mathcal{H} when using a test γ , can be evaluated by taking the expectation of the corresponding critical function.

Lemma 2.6.1. Given a statistical model $\mathbf{X} \sim P_\theta$, $\theta \in \Theta$, consider a hypothesis $\mathcal{H} : \theta \in \Theta_0$ and let γ be a test with corresponding critical function φ , so that

$$P[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H} | \mathbf{X}] = \varphi(\mathbf{X}).$$

In this case, for each $\theta \in \Theta$

$$P_\theta[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}] = E_\theta[\varphi(\mathbf{X})].$$

Proof. Using that the probability of an event A equals the expectation of the corresponding indicator function, the double expectation theorem yields that

$$\begin{aligned} P_\theta[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}] &= E_\theta[I[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}]] \\ &= E_\theta[E_\theta[[I[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}] | \mathbf{X}]]] \\ &= E_\theta[[P_\theta[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}] | \mathbf{X}]]. \end{aligned}$$

On the other hand, under the test γ , the conditional probability of rejecting \mathcal{H} given \mathbf{X} is $\varphi(\mathbf{X})$ *regardless of the parameter value*, so that

$$P_{\theta}[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H} \mid \mathbf{X}] = P[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H} \mid \mathbf{X}] = \varphi(\mathbf{X});$$

combining this fact with the previous display it follows that

$$P_{\theta}[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}] = E_{\theta}[\varphi(\mathbf{X})],$$

concluding the argument. □

Chapter 3

Decision Errors

In this chapter the possibility of taking an incorrect decision when testing a hypothesis \mathcal{H} is studied. The main conclusion to be established below is that, regardless of the test γ used by the analyst, the possibility of an erroneous decision can not be generally avoided.

3.1. Errors and a Lower Bound

To begin with, consider a hypothesis

$$\mathcal{H} : \theta \in \Theta_0$$

and its complement

$$\mathcal{K} : \theta \in \Theta \setminus \Theta_0.$$

Recall now that after observing the vector \mathbf{X} a test γ attains two possible values, namely, ‘Reject \mathcal{H} ’, and ‘Accept \mathcal{H} ’, and let φ be the critical function associated to γ .

- The decision $\gamma(\mathbf{X}) = \text{‘Reject } \mathcal{H}\text{’}$ is an error when $\theta \in \Theta_0$, and the probability of incurring in this error is $E_\theta[\varphi(\mathbf{X})]$, by Lemma 2.6.1. The largest probability of rejecting \mathcal{H} incorrectly is

$$\mathcal{E}_\gamma[\mathcal{H}] = \max_{\theta_0 \in \Theta_0} E_{\theta_0}[\varphi(\mathbf{X})] \tag{3.1.1}$$

- The decision $\gamma(\mathbf{X}) = \text{‘Accept } \mathcal{H}\text{’}$ —which is equivalent to ‘Reject \mathcal{K} ’—is incorrect when $\theta \in \Theta \setminus \Theta_0$. By Lemma 2.6.1 the probability of accepting \mathcal{H} is $1 - E_\theta[\varphi(\mathbf{X})] = E_\theta[1 - \varphi(\mathbf{X})]$, so that the largest probability of rejecting \mathcal{K} incorrectly is

$$\mathcal{E}_\gamma[\mathcal{K}] = \max_{\theta_1 \in \Theta \setminus \Theta_0} E_{\theta_1}[1 - \varphi(\mathbf{X})] \tag{3.1.2}$$

In the remainder of the section a lower bound for the sum $\mathcal{E}_\gamma[\mathcal{H}] + \mathcal{E}_\gamma[\mathcal{K}]$ will be determined. The starting point is the following idea.

Definition 3.1.1. Consider a statistical model $\mathbf{X} \sim P_\theta$, $\theta \in \Theta$ where the observation vector \mathbf{X} has dimension n , and let \mathcal{H} and \mathcal{K} be the complementary hypothesis given by

$$\mathcal{H}: \theta \in \Theta_0, \quad \mathcal{K}: \theta \in \Theta \setminus \Theta_0,$$

where Θ_0 is a subset of Θ . The number $\ell(\mathcal{H}, \mathcal{K})$ corresponding to the pair of hypothesis \mathcal{H} and \mathcal{K} is defined as follows:

(i) If P_θ is a continuous distribution with density $f(\mathbf{x}; \theta)$ for each $\theta \in \Theta$

$$\ell(\mathcal{H}, \mathcal{K}) := \max_{\theta_0 \in \Theta_0, \theta_1 \in \Theta \setminus \Theta_0} \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_0), f(\mathbf{x}; \theta_1)\} d\mathbf{x},$$

whereas

(ii) If P_θ is a discrete distribution with probability function $f(\mathbf{x}; \theta)$ for each $\theta \in \Theta$

$$\ell(\mathcal{H}, \mathcal{K}) := \max_{\theta_0 \in \Theta_0, \theta_1 \in \Theta \setminus \Theta_0} \sum_{\mathbf{x} \in \mathbb{R}^n} \min\{f(\mathbf{x}; \theta_0), f(\mathbf{x}; \theta_1)\}.$$

The following simple result plays a central role in the subsequent development.

Theorem 3.1.1. In the context of Definition 3.1.1, for any test γ the largest probabilities $\mathcal{E}_\gamma[\mathcal{H}]$ and $\mathcal{E}_\gamma[\mathcal{K}]$ of an incorrect rejection of \mathcal{H} and \mathcal{K} , respectively, satisfy that

$$\mathcal{E}_\gamma[\mathcal{H}] + \mathcal{E}_\gamma[\mathcal{K}] \geq \ell(\mathcal{H}, \mathcal{K}).$$

Proof. Let γ be a test with critical function φ . Suppose that P_θ is a continuous distribution with density $f(\mathbf{x}; \theta)$ and select $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta \setminus \Theta_0$. Notice now that

$$\begin{aligned} E_{\theta_0}[\varphi(\mathbf{X})] + E_{\theta_1}[1 - \varphi(\mathbf{X})] &= \int_{\mathbb{R}^n} \varphi(\mathbf{x})f(\mathbf{x}; \theta_0) d\mathbf{x} + \int_{\mathbb{R}^n} (1 - \varphi(\mathbf{x}))f(\mathbf{x}; \theta_1) d\mathbf{x} \\ &\geq \int_{\mathbb{R}^n} \varphi(\mathbf{x}) \min\{f(\mathbf{x}; \theta_0), f(\mathbf{x}; \theta_1)\} d\mathbf{x} \\ &\quad + \int_{\mathbb{R}^n} (1 - \varphi(\mathbf{x})) \min\{f(\mathbf{x}; \theta_0), f(\mathbf{x}; \theta_1)\} d\mathbf{x} \\ &= \int_{\mathbb{R}^n} [\varphi(\mathbf{x}) + (1 - \varphi(\mathbf{x})) \min\{f(\mathbf{x}; \theta_0), f(\mathbf{x}; \theta_1)\}] d\mathbf{x} \end{aligned}$$

and then

$$E_{\theta_0}[\varphi(\mathbf{X})] + E_{\theta_1}[1 - \varphi(\mathbf{X})] \geq \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_0), f(\mathbf{x}; \theta_1)\} d\mathbf{x};$$

taking the maximum over $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta \setminus \Theta_0$, from (3.1.1), (3.1.2) and Definition 3.1.1 it follows that $\mathcal{E}_\gamma[\mathcal{H}] + \mathcal{E}_\gamma[\mathcal{K}] \geq \ell(\mathcal{H}, \mathcal{K})$. The argument in the discrete case is similar. \square

3.2. Positivity of the Lower Bound

As it is shown below, in general $\ell(\mathcal{H}, \mathcal{K})$ is a positive number.

Lemma 3.2.1. In the context of Definition 3.1.1 the following assertions (i) and (ii) hold.

(i) Suppose that P_θ is a continuous distribution with density $f(\mathbf{x}; \theta)$, and that there exist parameters $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta \setminus \Theta_0$ as well as a point $\mathbf{x} \in \mathbb{R}^n$ such that

$$\begin{aligned} &\text{The densities } f(\cdot; \theta_0) \text{ and } f(\cdot; \theta_1) \text{ are continuous at } \mathbf{x}, \text{ and} \\ &f(\mathbf{x}; \theta_0) > 0 \quad \text{and} \quad f(\mathbf{x}; \theta_1) > 0. \end{aligned} \tag{3.2.1}$$

In this case, $\ell(\mathcal{H}, \mathcal{K}) > 0$.

(ii) Suppose that P_θ is a discrete distribution with probability function $f(\mathbf{x}; \theta)$, and that for some parameters $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta \setminus \Theta_0$ and a point $\mathbf{x} \in \mathbb{R}^n$

$$f(\mathbf{x}; \theta_0) > 0 \quad \text{and} \quad f(\mathbf{x}; \theta_1) > 0.$$

In this framework,

$$\ell(\mathcal{H}, \mathcal{K}) > 0.$$

On the other hand,

(iii) The relation

$$\ell(\mathcal{H}, \mathcal{K}) \leq 1,$$

is always valid.

Proof. (i) Let the parameters $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta \setminus \Theta_0$ as well as $\mathbf{x} \in \mathbb{R}^n$ be as in (3.2.1). Since the densities $f(\cdot; \theta_0)$ and $f(\cdot; \theta_1)$ are continuous at \mathbf{x} , there exist a $\varepsilon > 0$ such that $f(\mathbf{y}; \theta_0)$ and $f(\mathbf{y}; \theta_1)$ are positive numbers when $\|\mathbf{y} - \mathbf{x}\| \leq \varepsilon$, and then

$$\int_{\mathbb{R}^n} \min\{f(\mathbf{y}; \theta_0), f(\mathbf{y}; \theta_1)\} d\mathbf{y} \geq \int_{\mathbf{y}: \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon} \min\{f(\mathbf{y}; \theta_0), f(\mathbf{y}; \theta_1)\} d\mathbf{y} > 0.$$

It follows that

$$\ell(\mathcal{H}, \mathcal{K}) = \max_{\theta_0 \in \Theta_0; \theta_1 \in \Theta \setminus \Theta_0} \int_{\mathbb{R}^n} \min\{f(\mathbf{y}; \theta_0), f(\mathbf{y}; \theta_1)\} d\mathbf{y} > 0;$$

this completes the proof of part (i), whereas the second assertion can be established along similar lines. To establish part (iii), just observe that the relation

$$\min\{f(\mathbf{y}; \theta_0), f(\mathbf{y}; \theta_1)\} \leq f(\mathbf{y}; \theta_0)$$

is always valid, so that

$$\int_{\mathbb{R}^n} \min\{f(\mathbf{y}; \theta_0), f(\mathbf{y}; \theta_1)\} d\mathbf{y} \leq \int_{\mathbb{R}^n} f(\mathbf{y}; \theta_0) d\mathbf{y} = 1;$$

taking the maximum with respect to $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta \setminus \Theta_0$, the conclusion $\ell(\mathcal{H}, \mathcal{K}) \leq 1$ follows. \square

Remark 3.2.1. The essential condition in Lemma 3.2.1 is that the supports of same distributions P_{θ_0} and P_{θ_1} with $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta \setminus \Theta_0$ have nonempty intersection. This requirement is very weak and is practically satisfied in *all* models considered in applications. Thus, the conclusion of Lemma 3.2.1 allows to establish the following assertion:

Regardless of the test γ being used to decide which of the complementary hypothesis \mathcal{H} and \mathcal{K} holds, *in practically all interesting applications the largest error probabilities $\mathcal{E}_\gamma[\mathcal{H}]$ and $\mathcal{E}_\gamma[\mathcal{K}]$ satisfy the inequality*

$$\mathcal{E}_\gamma[\mathcal{H}] + \mathcal{E}_\gamma[\mathcal{K}] \geq \ell(\mathcal{H}, \mathcal{K}) > 0.$$

This inequality shows that, given an observation vector \mathbf{X} , the maximum error probabilities $\mathcal{E}_\gamma[\mathcal{H}]$ and $\mathcal{E}_\gamma[\mathcal{K}]$ can not be made *simultaneously* ‘small’. \square

3.3. An Improved Result

The following result is a refinement of the previous lemma, and provides conditions under which the number $\ell(\mathcal{H}, \mathcal{K})$ is equal to 1.

Lemma 3.3.1. In the context of Definition 3.1.1, let $f(\mathbf{x}; \theta)$ be the density or probability function corresponding to P_θ , and suppose that there exists $\theta^* \in \Theta$ satisfying that

- (i) $\lim_{\theta \rightarrow \theta^*} \int_{\mathbb{R}^n} |f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta^*)| d\mathbf{x} = 0$ if P_θ is a continuous distribution, or
- $$\lim_{\theta \rightarrow \theta^*} \sum_{x \in \mathbb{R}^n} |f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta^*)| = 0$$
- if the distribution
- P_θ
- is discrete.

- (ii) There exist sequences $\{\theta_{0n}\} \subset \Theta_0$ and $\{\theta_{1n}\} \subset \Theta \setminus \Theta_0$ such that

$$\lim_{n \rightarrow \infty} \theta_{0,n} = \theta^* = \lim_{n \rightarrow \infty} \theta_{1,n}.$$

In this case, $\ell(\mathcal{H}, \mathcal{K}) = 1$.

Proof. The argument will be presented for the case in which P_θ is a continuous distribution. Suppose that $\theta^* \in \Theta \setminus \Theta_0$, and select $\{\theta_{0,n}\} \subset \Theta_0$ such that $\lim_{n \rightarrow \infty} \theta_{0,n} = \theta^*$. In this case condition (i) yields that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^n} |f(\mathbf{x}; \theta_{0,n}) - f(\mathbf{x}; \theta^*)| d\mathbf{x} = 0. \quad (3.3.1)$$

Notice now that

$$\begin{aligned} \max\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} + \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} &= f(\mathbf{x}; \theta_{0,n}) + f(\mathbf{x}; \theta^*) \\ \max\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} - \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} &= |f(\mathbf{x}; \theta_{0,n}) - f(\mathbf{x}; \theta^*)|, \end{aligned}$$

so that

$$\int_{\mathbb{R}^n} \max\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x} + \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x} = 2, \quad (3.3.2)$$

and

$$\int_{\mathbb{R}^n} \max\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x} - \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

where the convergence follows from (3.3.1). Equation (3.3.2) allows to obtain the equality $\int_{\mathbb{R}^n} \max\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x} = 2 - \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x}$, and replacing this expression in the above display it follows that

$$\left[2 - \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x}\right] - \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x} \rightarrow 0,$$

which is equivalent to

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x} = 1.$$

Notice now that the inclusions $\theta_{0,n} \in \Theta_0$ and $\theta^* \in \Theta \setminus \Theta_0$ yield that the inequality

$$\ell(\mathcal{H}, \mathcal{K}) \geq \int_{\mathbb{R}^n} \min\{f(\mathbf{x}; \theta_{0,n}), f(\mathbf{x}; \theta^*)\} d\mathbf{x}$$

occurs, by Definition 3.1.1, and after taking the limit as $n \rightarrow \infty$, it follows that $\ell(\mathcal{H}, \mathcal{K}) \geq 1$; combining this inequality with the relation $\ell(\mathcal{H}, \mathcal{K}) \leq 1$ established in Lemma 3.2.1(iii), it follows that $\ell(\mathcal{H}, \mathcal{K}) = 1$. The case $\theta^* \in \Theta_0$ can be analyzed along similar lines. \square

Remark 3.3.1. Assume that each distribution P_θ is continuous, and that the $\{P_\theta\}_{\theta \in \Theta}$ is a location-scale family, that is, P_θ is the distribution of $A\mathbf{Z} + \mu$ where \mathbf{Z} has a fixed (known) continuous distribution, whereas A is an invertible matrix and μ is a vector. In this case, condition (i) in Lemma 3.3.1 holds; this fact can be verified by using the following analytical results: If $g(\cdot)$ is an integrable function, then

$$\int_{\mathbb{R}^n} |g(\mathbf{x} + \mu) - g(\mathbf{x})| d\mathbf{x} \rightarrow 0 \quad \text{as } \mu \rightarrow 0,$$

and

$$\int_{\mathbb{R}^n} |g(A\mathbf{x}) - g(\mathbf{x})| d\mathbf{x} \rightarrow 0 \quad \text{as } A \rightarrow I,$$

where I is the identity matrix. On the other hand, the second condition in Lemma 3.3.1 is satisfied when the sets Θ_0 and its complement $\Theta \setminus \Theta_0$ share a common boundary point. For instance, suppose that $\Theta_0 = [0, 4]$ and $\Theta \setminus \Theta_0 = (4, \infty)$. In this case $\theta^* = 4$ can be approximated using points in Θ_0 and points in $\Theta \setminus \Theta_0$. In discrete models, the two above displayed relations can be verified using the dominated convergence theorem, or Weierstrass M -test for the convergence of series. In short, in common applications the conditions in Lemma 3.3.1 are valid, so that $\ell(\mathcal{H}, \mathcal{K}) = 1$, and then $\mathcal{E}_\gamma[\mathcal{H}] + \mathcal{E}_\gamma[\mathcal{K}] \geq 1$. \square

3.4. Null Hypothesis

When a test γ is used to decide which of the complementary hypothesis \mathcal{H} or \mathcal{K} occurs, the objective of the analyst is to select the correct hypothesis. However, the results in the previous section showed that, regardless of the test γ , the largest probabilities of rejecting incorrectly \mathcal{H} or \mathcal{K} —which are given by $\mathcal{E}_\gamma[\mathcal{H}]$ and $\mathcal{E}_\gamma[\mathcal{K}]$, respectively—can not be made arbitrarily small *simultaneously*. The theory handles this problem by choosing the hypothesis whose incorrect rejection is ‘more serious’, and selecting a test that makes the probability of that serious error ‘small’, while trying to keep the probability of the other error as small as possible.

Definition 3.4.1. Given two complementary hypothesis \mathcal{H} and \mathcal{K} ,

- (i) The *null hypothesis* is the one whose incorrect rejection represents the more serious error; the null hypothesis is denoted by \mathcal{H}_0 .
- (ii) The hypothesis that is complementary to \mathcal{H}_0 is referred to as the *alternative hypothesis* and is denoted by \mathcal{H}_1 .

Example 3.4.1. A disease is prevented by a vaccine in 95% of the cases. A new vaccine has been produced and it is claimed that it is effective in at least 98% of the cases, an assertion

that must be verified by the Health Department before proceeding to massive application of the new formula. The following hypothesis are considered:

$$\mathcal{H}: \theta \in [0.98, 1], \quad \text{and} \quad \mathcal{K}: \theta \in [0, 0.98),$$

where $\theta \in \Theta = [0, 1]$ is the proportion of cases in which the new vaccine is effective. When using a test, two errors may occur:

- \mathcal{H} is rejected when \mathcal{H} is true.

In this case the new vaccine is effective in 98% or more of the cases, but the test fails in detecting the improvement with respect to the vaccine that is currently in use. Thus, the Health Department keeps on using the old vaccine, missing the opportunity to obtain at less 3% more protection.

- \mathcal{K} is rejected when \mathcal{K} is true.

In this case the new vaccine is effective in less than 98% of the cases—in particular, it may be effective in less than 95% of the cases—but according to the test it is declared that it is 3% or more effective than the vaccine presently used. Thus, the Health Department eagerly recommends the massive application of the new vaccine, but this decision may diminish the protection of the population against the disease.

In short:

- If \mathcal{H} is incorrectly rejected, an improvement of 3% in the protection rate is missed;
- If \mathcal{K} is incorrectly rejected, the protection of the population against the disease may decrease below the present level.

Between these errors, the second one (rejecting \mathcal{K} incorrectly) seems to have the most serious consequences, and then the null hypothesis is \mathcal{K} , that is,

$$\mathcal{H}_0: \theta \in [0, 0.98)$$

and the alternative hypothesis is $\mathcal{H}_1: \theta \in [0.98, 1]$. □

After selecting the null and alternative hypothesis, the performance of a test is measured by the power function, an idea that is now introduced.

Definition 3.4.2. Let $\mathbf{X} \sim P_\theta$, $\theta \in \Theta$ be a statistical model and consider a null hypothesis

$$\mathcal{H}_0: \theta \in \Theta_0$$

as well as the corresponding alternative hypothesis.

$$\mathcal{H}_1: \theta \in \Theta \setminus \Theta_0.$$

Let γ be a given test.

(i) The *power function* of the test γ is denoted by π_γ and is defined by

$$\pi_\gamma(\theta) = P_\theta[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}_0], \quad \theta \in \Theta.$$

(ii) An *error of type I* occurs if \mathcal{H}_0 is rejected when \mathcal{H}_0 is true and, when $\theta \in \Theta_0$, the probability of incurring in such an error is denoted by $\alpha_\gamma(\theta)$, that is,

$$\alpha_\gamma(\theta) = P_\theta[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}_0], \quad \theta \in \Theta_0.$$

(iii) An *error of type II* occurs if \mathcal{H}_1 is rejected when \mathcal{H}_1 holds—equivalently, if \mathcal{H}_0 is accepted when \mathcal{H}_0 is false—and, for each $\theta \in \Theta \setminus \Theta_0$, the probability of incurring in an error of type II is denoted by $\beta_\gamma(\theta)$, *i.e.*,

$$\beta_\gamma(\theta) = P_\theta[\gamma(\mathbf{X}) = \text{Accept } \mathcal{H}_0], \quad \theta \in \Theta \setminus \Theta_0;$$

equivalently,

$$\beta_\gamma(\theta) = 1 - \pi_\gamma(\theta), \quad \theta \in \Theta \setminus \Theta_0; \tag{3.4.1}$$

(iv) The power of the test γ against the alternative value $\theta \in \Theta \setminus \Theta_0$ is

$$\pi_\gamma(\theta) = 1 - \beta_\gamma(\theta), \quad \theta \in \Theta \setminus \Theta_0. \tag{3.4.2}$$

(v) The *size of the test* γ is denoted by α^γ and is defined by

$$\alpha^\gamma = \max_{\theta \in \Theta_0} \alpha_\gamma(\theta) = \max_{\theta \in \Theta_0} P_\theta[\gamma(\mathbf{X}) = \text{Reject } \mathcal{H}_0]. \tag{3.4.3}$$

The size of a test measures the risk of incurring in an error of type I—the most serious error—and the first objective of the analyst is to use a test that has a ‘small’ size, say less than or equal to a specified number $\alpha \in (0, 1)$. After this property is granted, the next goal is to have a probability of incurring in an error of type II as small as possible. A test that satisfies these properties is formally described below:

Definition 3.4.3. Consider the null hypothesis $\mathcal{H}_0: \theta \in \Theta_0$ and the corresponding alternative $\mathcal{H}_1: \theta \in \Theta \setminus \Theta_0$. A test γ is *uniformly most powerful* of size (less than or equal to) α if the following conditions (i) and (ii) hold:

(i) The size of γ does not exceed α , that is,

$$\alpha^\gamma = \max_{\theta \in \Theta_0} \pi_\gamma(\theta) \leq \alpha;$$

(ii) If $\tilde{\gamma}$ is an arbitrary test with size less than or equal to α , then the probability of incurring in an error of type II is smaller under γ than under $\tilde{\gamma}$:

$$\beta_{\gamma}(\theta) \leq \beta_{\tilde{\gamma}}(\theta), \quad \theta \in \Theta \setminus \Theta_0,$$

that is,

$$\pi_{\gamma}(\theta) \geq \pi_{\tilde{\gamma}}(\theta), \quad \theta \in \Theta \setminus \Theta_0;$$

see (3.4.2).

The construction of uniformly most powerful tests will be illustrated in the following sections.

Remark 3.4.1. When a test γ is used to test the hypothesis \mathcal{H}_0 versus \mathcal{H}_1 , the probability of an error of type I is less than or equal to the size α^{γ} , that is, rejecting \mathcal{H}_0 when it is true has a probability at most α^{γ} . Thus, when the null hypothesis is rejected by a test γ with a small size, then the analyst will be confident that the rejection is correct. This comments signals a useful rule to select which of two complementary conditions on the parameter should be selected as the null hypothesis: If it is desired to declare confidently that a condition \mathcal{C} on the parameter holds, select the null hypothesis as the opposite of the condition \mathcal{C} :

$$\mathcal{H}_0: \text{Condition } \mathcal{C} \text{ does not occur.}$$

With this selection, once \mathcal{H}_0 is rejected by a test with ‘small’ size, the alternative \mathcal{H}_1 —which asserts that Condition \mathcal{C} occurs—will be established on ‘firm’ grounds. \square

3.5. Examples

In this section the diverse concepts introduced above will be illustrated.

Exercise 3.5.1. An urn contains 10 balls, of which θ balls are blue (the rest being red and white). To test the null hypothesis $\mathcal{H}_0: \theta = 3$ versus $\mathcal{H}_1: \theta = 4$, a sample of size 3 balls is taken and the color of the balls in the sample is recorded. The test γ used in this problem rejects \mathcal{H}_0 if all three balls drawn are blue. Compute the probabilities α and β of incurring in an error of type I and II, respectively, if

- (i) Sampling is done without replacement.
- (ii) Sampling is done with replacement.

Solution. Let X be the number of blue balls contained in the sample of size 3.

(i) When the sampling is without replacement, X has *Hipergeometric* $(10, \theta, 3)$ distribution, where $\theta \in \Theta = \{3, 4\}$, that is,

$$P_{\theta}[X = x] = \frac{\binom{\theta}{x} \binom{10 - \theta}{3 - x}}{\binom{10}{3}}.$$

The power function π corresponding to γ is given by

$$\begin{aligned} \pi(\theta) &= P_{\theta}[\gamma(X) = \text{'Reject } \mathcal{H}_0\text{'}] \\ &= P_{\theta}[X = 3] \\ &= \frac{\binom{\theta}{3} \binom{10 - \theta}{3 - 3}}{\binom{10}{3}} = \frac{\binom{\theta}{3}}{\binom{10}{3}} = \frac{\theta(\theta - 1)(\theta - 2)}{10(9)(8)}, \quad \theta \in \{3, 4\} \end{aligned}$$

Therefore,

$$\alpha = \pi(3) = \frac{1}{120}, \quad \text{and} \quad \beta = 1 - \pi(4) = 1 - \frac{1}{30} = \frac{29}{30}.$$

(ii) Under sampling with replacement, X has *Binomial* $\left(3, \frac{\theta}{10}\right)$ distribution, where $\theta \in \Theta = \{3, 4\}$, that is,

$$P_{\theta}[X = x] = \binom{3}{x} \left(\frac{\theta}{10}\right)^x \left(1 - \frac{\theta}{10}\right)^{3-x}.$$

The power function π corresponding to γ is given by

$$\begin{aligned} \pi(\theta) &= P_{\theta}[\gamma(X) = \text{'Reject } \mathcal{H}_0\text{'}] \\ &= P_{\theta}[X = 3] \\ &= \binom{3}{3} \left(\frac{\theta}{10}\right)^3 \left(1 - \frac{\theta}{10}\right)^{3-3} = \frac{\theta^3}{1000}, \quad \theta \in \{3, 4\}. \end{aligned}$$

so that $\alpha = \pi(3) = 27/1000$ and $\beta = 1 - \pi(4) = 1 - 64/1000 = 936/1000$. \square

Exercise 3.5.2. The manufacturer of an imported automobile claims that the average miles per gallon (mpg) of this type of car is at least 30. Suppose that the mpg of a randomly selected car is a normally distributed random variable with mean μ and variance $\sigma^2 = 25$. To check the claim, consider the testing problem:

$$\mathcal{H}_0: \mu < 30 \quad \text{versus} \quad \mathcal{H}_1: \mu \geq 30.$$

A sample of 36 cars of the manufacturer's output is taken and the the average mpg \bar{X} for the sampled cars is determined. Suppose it is decided to reject \mathcal{H}_0 if, and only if, $\bar{X} \geq 32$.

- (i) Compute the size of the critical region (*i.e.*, α).
- (ii) Compute the power $\pi(\mu)$ of the test against each alternative value $\mu \geq 30$, and then find $\beta(\mu)$, the probability of an error of type II.

Solution. Since $\bar{\mathbf{X}}$ is the average of 36 independent and normally distributed random variables with mean μ and variance $\sigma^2 = 25$, it follows that $\bar{\mathbf{X}} \sim \mathcal{N}(\mu, 25/36)$. On the other hand, the test rejects \mathcal{H}_0 when $\bar{\mathbf{X}} \geq 32$, and then the power function is given by

$$\pi(\mu) = P_\mu[\bar{\mathbf{X}} \geq 32] = P_\mu \left[\frac{\bar{\mathbf{X}} - \mu}{5/6} \geq \frac{32 - \mu}{5/6} \right] = P \left[Z \geq \frac{32 - \mu}{5/6} \right],$$

where Z is a random variable with the standard normal distribution. Thus,

$$\pi(\mu) = 1 - \Phi \left(\frac{32 - \mu}{5/6} \right) \quad \text{for every } \mu, \quad (3.5.1)$$

where, as usual, Φ is the cumulative distribution function of the standard normal distribution; notice that $\pi(\cdot)$ is an *increasing* function.

- (i) The null hypothesis is rejected when $\bar{\mathbf{X}} \geq 32$. Thus, the *rejection or critical region* is $\mathcal{R} = \{\mathbf{x} \mid \bar{\mathbf{x}} \in [32, \infty)\}$. Observing that the expressions ‘size of the critical region’ and ‘size of the test’ are synonymous, it follows that

$$\alpha = \max_{\mu < 30} \pi(\mu) = \max_{\mu < 30} \left[1 - \Phi \left(\frac{32 - \mu}{5/6} \right) \right] = 1 - \Phi \left(\frac{32 - 30}{5/6} \right) = 1 - \Phi(2.4)$$

- (ii) Combining (3.4.1) and (3.5.1), it follows that the probability of an error of type II at each alternative parameter μ is given by

$$\beta(\mu) = \Phi \left(\frac{32 - \mu}{5/6} \right), \quad \mu \geq 30;$$

notice that $\beta(\mu)$ is decreasing, and then the largest value of $\beta(\mu)$ is attained at $\mu = 30$: $\beta(30) = \Phi(2.4)$. It follows that the sum of the largest probability of an error of type I—the size of the test, which is given by $1 - \Phi(2.4)$ —and the largest possible probability of an error of type II is equal to 1, as it is typical in most interesting testing problems, where the sets specified by the null and alternative hypothesis share a boundary point; in the present context such a point is $\mu = 30$. \square

Exercise 3.5.3. It is desired to test $\mathcal{H}_0: \mu \leq 10$ versus $\mathcal{H}_1: \mu > 10$ on the basis of a random sample \mathbf{X} of size 25 from a normal population with unknown mean μ and variance $\sigma^2 = 4$. If the (largest) probability of an error of type I is to be $\alpha = 0.025$ and the test is determined by the critical function $\varphi(\mathbf{X}) = 1$ if $\bar{\mathbf{X}} > c$ and $\varphi(\mathbf{X}) = 0$ otherwise,

- (i) Find c (and hence the critical region), and

(ii) Find $\beta(\mu)$.

Solution. The sample mean $\bar{\mathbf{X}}$ has the normal distribution $\mathcal{N}(\mu, 4/25)$. The power function of the test is

$$\begin{aligned}\pi(\mu) &= E_{\mu}[\varphi(\mathbf{X})] \\ &= P_{\mu}[\bar{\mathbf{X}} > c] \\ &= P_{\mu}\left[\frac{\bar{\mathbf{X}} - \mu}{\sqrt{4/25}} > \frac{c - \mu}{\sqrt{4/25}}\right] \\ &= P\left[Z > \frac{5}{2}(c - \mu)\right] = 1 - \Phi\left(\frac{5}{2}(c - \mu)\right), \quad \mu \in \mathbb{R}.\end{aligned}$$

(i) Observe that (regardless of the value of c) the power function is an increasing function of μ . Since the null hypothesis states that $\mu \leq 10$, the size of the test is

$$\alpha = \max_{\mu \leq 10} \pi(\mu) = \max_{\mu \leq 10} \left[1 - \Phi\left(\frac{5}{2}(c - \mu)\right)\right] = 1 - \Phi\left(\frac{5}{2}(c - 10)\right).$$

Therefore, the condition $\alpha = 0.025$ leads to

$$1 - \Phi\left(\frac{5}{2}(c - 10)\right) = 0.025, \quad \text{i.e.,} \quad \Phi\left(\frac{5}{2}(c - 10)\right) = 0.975.$$

It follows that

$$\frac{5}{2}(c - 10) = \Phi^{-1}(0.975) = 1.96,$$

and then $c = 10 + 3.92/5 = 10.7805$. This specifies completely the power function:

$$\pi(\mu) = 1 - \Phi\left(\frac{5}{2}(10.7805 - \mu)\right), \quad \mu \in \mathbb{R}.$$

(ii) The probability of an error of type II is

$$\beta(\mu) = 1 - \pi(\mu) = \Phi\left(\frac{5}{2}(10.7805 - \mu)\right)$$

for each $\mu \in (10, \infty)$. □

Exercise 3.5.4. (i) Consider testing the simple null hypothesis $\mathcal{H}_0: \theta = \theta_0$ versus the simple alternative $\mathcal{H}_1: \theta = \theta_1$ where θ is the mean of a normal distribution with known variance σ^2 . Suppose that a random sample of size n (to be determined) is taken and that the test with the following critical function is used: $\varphi(\mathbf{X}) = 1$ if $\bar{\mathbf{X}} > c$ (and 0 otherwise). How large n is needed to obtain a probability of Type I error α and power of $1 - \beta$?

(ii) Suppose $\theta_0 = 5$ and $\theta_1 = 8$. If $\sigma^2 = 4$, find the values of c and n such that $\alpha = 0.05 = \beta$.

Solution. Since \mathbf{X} is a sample of size n from the $\mathcal{N}(\theta, \sigma^2)$ distribution, it follows that $\bar{\mathbf{X}} \sim \mathcal{N}(\theta, \sigma^2/n)$. Thus, the power function of the test described above is given by

$$\begin{aligned}\pi(\theta) &= E_{\theta}[\varphi(\mathbf{x})] \\ &= P_{\theta}[\bar{\mathbf{X}} > c] \\ &= P_{\theta} \left[\frac{\bar{\mathbf{X}} - \theta}{\sigma/\sqrt{n}} > \frac{c - \theta}{\sigma/\sqrt{n}} \right] \\ &= P \left[Z > \frac{c - \theta}{\sigma/\sqrt{n}} \right] = 1 - \Phi \left(\frac{c - \theta}{\sigma/\sqrt{n}} \right), \quad \theta \in \{\theta_0, \theta_1\},\end{aligned}$$

where Z has the standard normal distribution.

(i) The size α of the test is

$$\alpha = \pi(\theta_0) = 1 - \Phi \left(\frac{c - \theta_0}{\sigma/\sqrt{n}} \right), \quad \text{that is,} \quad 1 - \alpha = \Phi \left(\frac{c - \theta_0}{\sigma/\sqrt{n}} \right),$$

whereas the probability of an error of type II is given by $\beta = 1 - \pi(\theta_1)$, that is,

$$\beta = \Phi \left(\frac{c - \theta_1}{\sigma/\sqrt{n}} \right).$$

These two last displays together yield that

$$\begin{aligned}\frac{c - \theta_0}{\sigma/\sqrt{n}} &= z_{1-\alpha} \\ \frac{c - \theta_1}{\sigma/\sqrt{n}} &= z_{\beta},\end{aligned}\tag{3.5.2}$$

where $z_{\delta} = \Phi^{-1}(\delta)$ for each $\delta \in (0, 1)$, that is, z_{δ} is the quantile of order δ corresponding to the standard normal distribution. Taking the difference of these equations it follows that

$$\frac{\theta_1 - \theta_0}{\sigma/\sqrt{n}} = z_{1-\alpha} - z_{\beta}, \quad \text{that is,} \quad \frac{\sigma}{\sqrt{n}} = \frac{\theta_1 - \theta_0}{z_{1-\alpha} - z_{\beta}}.\tag{3.5.3}$$

It follows that

$$n = \sigma^2 \left(\frac{z_{1-\alpha} - z_{\beta}}{\theta_1 - \theta_0} \right)^2,\tag{3.5.4}$$

whereas the second equation in (3.5.3) and the first equality in (3.5.2) together lead to

$$c = \theta_0 + z_{1-\alpha} \frac{\theta_1 - \theta_0}{z_{1-\alpha} - z_{\beta}}.\tag{3.5.5}$$

(ii) When α and β take the value 0.05, it follows that

$$z_{1-\alpha} = z_{0.95} = 1.645$$

and

$$z_{0.05} = -1.645.$$

Thus, if $\theta_1 = 8$, $\theta_0 = 5$ and $\sigma^2 = 4$, the expressions (3.5.4) and (3.5.5) yield that

$$n = 4 \left(\frac{2(1.645)}{3} \right)^2 \quad \text{and} \quad c = 6.5 .$$

□

Exercise 3.5.5. Consider the *Poisson*(λ) distribution, where $\lambda \in \{2, 3\}$. The null hypothesis $\mathcal{H}_0: \lambda = 3$ versus the alternative $\mathcal{H}_1: \lambda = 2$ will be tested based on a sample \mathbf{X} of size 5. The test γ calls for rejection of \mathcal{H}_0 if $\bar{\mathbf{X}} < c$, and accepts \mathcal{H}_0 otherwise. If α is to be the number as closest to 0.05 as possible, find the critical region to use.

Solution. The power function of γ is $\pi(\lambda) = E_\lambda[\bar{\mathbf{X}} > c]$, $\lambda \in \{2, 3\}$. The exercise calls for the number c such that the size of the test γ —which is $\pi(2) = P_2[\bar{\mathbf{X}} > c]$ —provides the better approximation to 0.05. Notice that, under \mathcal{H}_0 ,

$$Y = 5\bar{\mathbf{X}} \sim \text{Poisson}(10)$$

and that

$$P_2[\bar{\mathbf{X}} > c] = P[Y > 5c].$$

The table of the *Poisson*(10) distribution, which is reproduced below, shows that the best approximation from below to 0.05 by $P[Y > 5c]$ is attained when $5c = 15$, that is, when $c = 3$, and the critical (or rejection) region is $\mathcal{R} = \{\mathbf{x} \in \mathcal{X} \mid \bar{\mathbf{x}} > 3\}$, where

$\mathcal{X} = \{\mathbf{x} = (x_1, x_2, x_3, x_4, x_5) \mid x_i \text{ is a nonnegative integer}\}$.

n	$P[Y = n]$	$P[Y \leq n]$	$P[Y > n]$
0	0.0000	0.0000	1.0000
1	0.0005	0.0005	0.9995
2	0.0023	0.0028	0.9972
3	0.0076	0.0103	0.9897
4	0.0189	0.0293	0.9707
5	0.0378	0.0671	0.9329
6	0.0631	0.1301	0.8699
7	0.0901	0.2202	0.7798
8	0.1126	0.3328	0.6672
9	0.1251	0.4579	0.5421
10	0.1251	0.5830	0.4170
11	0.1137	0.6968	0.3032
12	0.0948	0.7916	0.2084
13	0.0729	0.8645	0.1355
14	0.0521	0.9165	0.0835
15	0.0347	0.9513	0.0487
16	0.0217	0.9730	0.0270
17	0.0128	0.9857	0.0143
18	0.0071	0.9928	0.0072
19	0.0037	0.9965	0.0035
20	0.0019	0.9984	0.0016
21	0.0009	0.9993	0.0007
22	0.0004	0.9997	0.0003
23	0.0002	0.9999	0.0001
24	0.0001	1.0000	0.0000
25	0.0000	1.0000	0.0000

□

Exercise 3.5.6. It is desired to test $\mathcal{H}_0: \mu = 10$ versus $\mathcal{H}_1: \mu \neq 10$, where μ is the population mean of a normal distribution with standard deviation $\sigma = 3$. A random sample of size $n = 9$ is taken, and the test γ_c with critical function $\varphi(\mathbf{X}) = 1$ if $|\bar{\mathbf{X}} - 10| > c$ (and $\varphi(\mathbf{X}) = 0$ otherwise) is used.

- (i) Find the values of c to use for each size of the test $\alpha = 0.01, 0.02, 0.05, 0.10$.
- (ii) Find the power function of the test γ_c .

Solution. Since \mathbf{X} is a random sample of size $n = 9$ from the $\mathcal{N}(\mu, 9)$, it follows that

$$\bar{\mathbf{X}} \sim \mathcal{N}(\mu, 1). \quad (3.5.6)$$

- (i) The test γ_c that rejects $\mathcal{H}_0: \mu = 10$ when $|\bar{\mathbf{X}} - 10| > c$ has size

$$\alpha_c = P_{10}[|\bar{\mathbf{X}} - 10| > c] = 2(1 - \Phi(c)),$$

where it was used that, under $\mathcal{H}_0: \mu = 10$, the random variable $\bar{\mathbf{X}} - 10$ has the standard normal distribution. Therefore,

$$c = \Phi^{-1}\left(1 - \frac{\alpha_c}{2}\right).$$

The values of c for the different values of α are shown below:

α	0.01	0.02	0.05	0.1
c	2.576	2.326	1.960	1.645

(ii) Using (3.5.6), it follows that the power function of the test π_{γ_c} is given by

$$\begin{aligned} \pi_{\gamma_c}(\mu) &= P_\mu[|\bar{\mathbf{X}} - 10| > c] \\ &= 1 - P_\mu[|\bar{\mathbf{X}} - 10| \leq c] \\ &= 1 - P_\mu[10 - c \leq \bar{\mathbf{X}} \leq 10 + c] \\ &= 1 - P_\mu[10 - \mu - c \leq \bar{\mathbf{X}} - \mu \leq 10 - \mu + c] \\ &= 1 - [\Phi(10 - \mu + c) - \Phi(10 - \mu - c)] \\ &= 1 - \Phi(10 - \mu + c) + \Phi(10 - \mu - c) \\ &= \Phi(\mu - 10 - c) + \Phi(10 - \mu - c). \end{aligned}$$

Notice that

$$\pi_{\gamma_c}(10 + \mu) = \Phi(\mu - c) + \Phi(-\mu - c) = \pi_{\gamma_c}(10 - \mu),$$

so that π_{γ_c} is symmetric about 10. □

Exercise 3.5.7. It is desired to test $\mathcal{H}_0: p = 0.2$ versus $\mathcal{H}_1: p = 0.4$ for a binomial distribution with $n = 10$. For the test γ with critical function $\varphi(X) = 1$ if $X \leq 3$ (and $\varphi(X) = 0$ otherwise), find α and β . Can you find a better test (*i.e.*, with improved probabilities of type I and II errors) with the same $n = 10$?

Solution. The power function of γ is

$$\pi_\gamma(p) = P_p[X \leq 3] = (1 - p)^{10} + 10(1 - p)^9 p + 45(1 - p)^8 p^2 + 120(1 - p)^7 p^3.$$

Therefore,

$$\alpha = \pi_\gamma(0.2) = 0.8791 \quad \text{and} \quad \beta = 1 - \pi_\gamma(0.4) = 1 - 0.3823 = 0.6177.$$

Notice that γ rejects \mathcal{H}_0 when $X \leq 3$. Intuitively, it is more reasonable to reject H_0 when X is ‘large’. Consider now the test $\tilde{\gamma}$ with critical function $\tilde{\varphi}(X) = 1$ if $X \geq 4$ and zero otherwise; notice that $\tilde{\varphi} = 1 - \varphi$, so that

$$\pi_{\tilde{\gamma}}(p) = 1 - \pi_\gamma(p),$$

and then

$$\tilde{\alpha} = \pi_{\tilde{\gamma}}(0.2) = 0.1209 \quad \text{and} \quad \tilde{\beta} = 1 - \pi_{\tilde{\gamma}}(0.4) = 1 - 0.6177 = 0.3823;$$

since $\tilde{\alpha} < \alpha$ and $\tilde{\beta} < \beta$, the test $\tilde{\gamma}$ is better than γ . \square

Exercise 3.5.8. Let $\mathbf{X} = (X_1, X_2, X_3)$ be a sample of the *Exponential*(λ) distribution, where $\lambda \in \{1/2, 1/3\}$. Consider testing $\mathcal{H}_0: \lambda = 1/2$ versus $\mathcal{H}_1: \lambda = 1/3$, and let γ be the test with critical function

$$\varphi(\mathbf{x}) = 1 \quad \text{if} \quad \frac{\prod_{i=1}^3 f(x_i; 1/3)}{\prod_{i=1}^3 f(x_i; 1/2)} \geq 1$$

(and $\varphi(\mathbf{x}) = 0$ otherwise), where $f(x; \lambda) = \lambda e^{-\lambda x} I_{[0, \infty)}(x)$ is the exponential density with parameter λ . Find α and β for this test.

Solution. Notice that

$$\frac{\prod_{i=1}^3 f(x_i; 1/3)}{\prod_{i=1}^3 f(x_i; 1/2)} = \frac{(1/3)^3 e^{-(x_1+x_2+x_3)/3}}{(1/2)^3 e^{-(x_1+x_2+x_3)/2}} = \frac{8}{27} e^{(x_1+x_2+x_3)/6}.$$

Therefore, the critical function φ can be explicitly written as

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{if } x_1 + x_2 + x_3 \geq 6 \log(27/8) \\ 0, & \text{otherwise,} \end{cases}$$

and the corresponding critical function π is given by

$$\pi(\lambda) = P_\lambda[X_1 + X_2 + X_3 \geq 6 \log(27/8)].$$

To compute the above probability, notice that

- If $Y \sim \text{Exponential}(\lambda)$, then $2\lambda Y \sim \text{Exponential}(1/2) = \Gamma(1, 1/2) = \chi_2^2$.

Since X_1, X_2, X_3 are independent with the *Exponential*(λ) distribution, it follows that $2\lambda X_1, 2\lambda X_2, 2\lambda X_3$ is a sample of the χ_2^2 distribution, and then

$$2\lambda X_1 + 2\lambda X_2 + 2\lambda X_3 \sim \chi_6^2.$$

Consequently,

$$\pi(\lambda) = P_\lambda[2\lambda X_1 + 2\lambda X_2 + 2\lambda X_3 \geq 12\lambda \log(27/8)] = P[W \geq 12\lambda \log(27/8)],$$

where

$$W \sim \chi_6^2.$$

Therefore,

$$\alpha = \pi(1/2) = P[W \geq 6 \log(27/8)] = 0.2941,$$

and

$$\beta = 1 - \pi(1/3) = 1 - P[W \geq 4 \log(27/8)] = 1 - 0.5612 = 0.4388 .$$

Notice that $\alpha + \beta < 1$, a relation that can be traced back to the fact that the null subset $\{1/2\}$ and the alternative subset $\{1/3\}$ do not share a boundary point. \square

Exercise 3.5.9. Let φ be a randomized test. Show that $\bar{\varphi} = 1 - \varphi$ is also a randomized test. Compute the relationship between the power functions of tests φ and $\bar{\varphi}$.

Solution. Recall that a critical function is a map from \mathcal{X} into $[0, 1]$, where \mathcal{X} is the set of possible values of the observation vector \mathbf{X} , and that the critical function is identified with the corresponding (randomized) test γ . Thus, starting with a critical function, it follows that the function $\bar{\varphi} = 1 - \varphi$, also maps \mathcal{X} into $[0, 1]$, and consequently is a critical function. The power functions associated to φ and $\bar{\varphi}$ satisfy

$$\pi_{\bar{\varphi}}(\theta) = E_{\theta}[\bar{\varphi}] = E_{\theta}[1 - \varphi] = 1 - E_{\theta}[\varphi] = 1 - \pi_{\varphi}(\theta),$$

so that $\pi_{\varphi} + \pi_{\bar{\varphi}} = 1$. \square

Exercise 3.5.10. Consider a random sample of size n from a distribution with mean μ and variance σ^2 , with n large enough that it is reasonable to assume that $\bar{\mathbf{X}}$ has a normal distribution. To test $\mathcal{H}_0: \mu = \mu_0$ versus $\mathcal{H}_1: \mu \neq \mu_0$, a suggested test is $\varphi(\mathbf{X}) = 1$ if $|\bar{\mathbf{X}} - \mu_0| \geq c$, and $\varphi(\mathbf{X}) = 0$ otherwise.

- (i) Find an approximate α -level critical region for this test if σ^2 is known.
- (ii) Find an approximation for the power of the test in part (i).

Solution. Given $c > 0$, the power function π corresponding to φ is

$$\begin{aligned} \pi(\mu) &= P_{\mu} [|\bar{\mathbf{X}} - \mu_0| \geq c] \\ &= 1 - P_{\mu} [|\bar{\mathbf{X}} - \mu_0| \leq c] \\ &= 1 - P_{\mu} [\mu_0 - c \leq \bar{\mathbf{X}} \leq \mu_0 + c] \\ &= 1 - P_{\mu} [\mu_0 - \mu - c \leq \bar{\mathbf{X}} - \mu \leq \mu_0 - \mu + c] \\ &= 1 - P_{\mu} \left[\sqrt{n} \frac{\mu_0 - \mu - c}{\sigma} \leq \frac{\bar{\mathbf{X}} - \mu}{\sigma/\sqrt{n}} \leq \sqrt{n} \frac{\mu_0 - \mu + c}{\sigma} \right]; \end{aligned}$$

from this point, the central limit theorem allows to establish the following approximation, where Z is a random variable with the standard normal distribution:

$$\begin{aligned} \pi(\mu) &\approx 1 - P \left[\sqrt{n} \frac{\mu_0 - \mu - c}{\sigma} \leq Z \leq \sqrt{n} \frac{\mu_0 - \mu + c}{\sigma} \right]; \\ &= 1 - \left[\Phi \left(\sqrt{n} \frac{\mu_0 - \mu + c}{\sigma} \right) - \Phi \left(\sqrt{n} \frac{\mu_0 - \mu - c}{\sigma} \right) \right] \\ &= 1 - \Phi \left(\sqrt{n} \frac{\mu_0 - \mu + c}{\sigma} \right) + \Phi \left(\sqrt{n} \frac{\mu_0 - \mu - c}{\sigma} \right) \end{aligned}$$

so that, using the relation $1 - \Phi(x) = \Phi(-x)$ for every $x \in \mathbb{R}$,

$$\pi(\mu) \approx \Phi\left(\sqrt{n} \frac{\mu - \mu_0 - c}{\sigma}\right) + \Phi\left(\sqrt{n} \frac{\mu_0 - \mu - c}{\sigma}\right). \quad (3.5.7)$$

The approximate size of the test is

$$\pi(\mu_0) \approx \Phi\left(\frac{-\sqrt{n} c}{\sigma}\right) + \Phi\left(\frac{-\sqrt{n} c}{\sigma}\right) = 2\Phi\left(\frac{-\sqrt{n} c}{\sigma}\right).$$

(i) If $\alpha = \pi(\mu_0)$, then

$$\alpha \approx 2\Phi\left(\frac{-\sqrt{n} c}{\sigma}\right)$$

so that

$$c \approx -\frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{\alpha}{2}\right) = \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}. \quad (3.5.8)$$

Thus, the test with critical region

$$\mathcal{R} = \{\mathbf{x} \mid |\bar{\mathbf{X}} - \mu_0| \geq \sigma z_{1-\alpha/2} / \sqrt{n}\}$$

has size approximately α .

(ii) Replacing the value of c obtained in (3.5.8) into (3.5.7) it follows that the power function of the test with the critical region in the above display satisfies

$$\pi(\mu) \approx \Phi\left(\sqrt{n} \frac{\mu - \mu_0}{\sigma} - z_{1-\alpha/2}\right) + \Phi\left(\sqrt{n} \frac{\mu_0 - \mu}{\sigma} - z_{1-\alpha/2}\right),$$

completing the analysis. \square

Exercise 3.5.11. As in Exercise 3.5.10, consider a random sample of size n from a distribution with mean μ and variance σ^2 , with n large enough that it is reasonable to assume that $\bar{\mathbf{X}}$ has a normal distribution. To test $\mathcal{H}_0: \mu = \mu_0$ versus $\mathcal{H}_1: \mu \neq \mu_0$, a suggested test is $\varphi(\mathbf{X}) = 1$ if $|\bar{\mathbf{X}} - \mu_0|/S \geq c$, and $\varphi(\mathbf{X}) = 0$ otherwise, where $S = \sqrt{\sum_{i=1}^n (X_i - \bar{\mathbf{X}})^2 / (n-1)}$ is the sample standard deviation.

(i) Find an approximate α -level critical region for this test if σ^2 is known.

(ii) Find an ‘approximation’ for the power of the test in part (i).

Solution. Proceeding as in Exercise 3.5.11, given $c > 0$, the power function π corresponding to φ is

$$\begin{aligned} \pi(\mu) &= P_\mu[|\bar{\mathbf{X}} - \mu_0|/S \geq c] \\ &= 1 - P_\mu[|\bar{\mathbf{X}} - \mu_0|/S \leq c] \\ &= 1 - P_\mu[\mu_0 - cS \leq \bar{\mathbf{X}} \leq \mu_0 + cS] \\ &= 1 - P_\mu[\mu_0 - \mu - cS \leq \bar{\mathbf{X}} - \mu \leq \mu_0 - \mu + cS] \\ &= 1 - P_\mu\left[\sqrt{n} \frac{\mu_0 - \mu - cS}{S} \leq \frac{\bar{\mathbf{X}} - \mu}{S/\sqrt{n}} \leq \sqrt{n} \frac{\mu_0 - \mu + cS}{S}\right] \\ &= 1 - P_\mu\left[\sqrt{n} \left(\frac{\mu_0 - \mu}{S} - c\right) \leq \frac{\bar{\mathbf{X}} - \mu}{S/\sqrt{n}} \leq \sqrt{n} \left(\frac{\mu_0 - \mu}{S} + c\right)\right]. \end{aligned} \quad (3.5.9)$$

When n increases, ‘large’, $S \rightarrow \sigma$ and $(\bar{\mathbf{X}} - \mu)/(\sigma/\sqrt{n}) \xrightarrow{d} \mathcal{N}(0, 1)$, so that Slutsky’s theorem implies that

$$\frac{\bar{\mathbf{X}} - \mu}{S/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

On the other hand, when n is ‘large’, S is ‘near’ σ with high probability, and treating S as a constant, the last probability in (3.5.9) can be *estimated* by

$$\pi(\mu) \approx 1 - \left[\Phi \left(\sqrt{n} \left[\frac{\mu_0 - \mu}{S} + c \right] \right) - \Phi \left(\sqrt{n} \left[\frac{\mu_0 - \mu}{S} - c \right] \right) \right].$$

Thus,

$$\begin{aligned} \pi(\mu) &\approx 1 - \Phi \left(\sqrt{n} \left[\frac{\mu_0 - \mu}{S} + c \right] \right) + \Phi \left(\sqrt{n} \left[\frac{\mu_0 - \mu}{S} - c \right] \right) \\ &= \Phi \left(\sqrt{n} \left[\frac{\mu - \mu_0}{S} - c \right] \right) + \Phi \left(\sqrt{n} \left[\frac{\mu_0 - \mu}{S} - c \right] \right). \end{aligned} \quad (3.5.10)$$

The approximate size of the test is

$$\pi(\mu_0) \approx \Phi(-\sqrt{n}c) + \Phi(-\sqrt{n}c) = 2\Phi(-\sqrt{n}c).$$

(i) If $\alpha = \pi(\mu_0)$, then

$$\alpha \approx 2\Phi(-\sqrt{n}c)$$

so that

$$c \approx -\frac{1}{\sqrt{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right) = \frac{1}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \frac{z_{1-\alpha/2}}{\sqrt{n}}. \quad (3.5.11)$$

Thus, the test with critical region

$$\mathcal{R} = \{\mathbf{x} \mid |\bar{\mathbf{x}} - \mu_0|/s \geq z_{1-\alpha/2}/\sqrt{n}\} = \{\mathbf{x} \mid \sqrt{n}|\bar{\mathbf{x}} - \mu_0|/s \geq z_{1-\alpha/2}\}$$

has size approximately α .

(ii) Replacing the value of c obtained in (3.5.11) into (3.5.10) it follows that, for a given value of μ , the power function of the test with the critical region in the above display can be *estimated* by

$$\pi(\mu) \approx \Phi \left(\sqrt{n} \frac{\mu - \mu_0}{S} - z_{1-\alpha/2} \right) + \Phi \left(\sqrt{n} \frac{\mu_0 - \mu}{S} - z_{1-\alpha/2} \right),$$

concluding the argument. □

Chapter 4

Neyman-Pearson Theory

Throughout this chapter a statistical model $\mathbf{X} \sim P_\theta$ is considered, where the parameter space consists of two points θ_0 and θ_1 , *i.e.*, $\Theta = \{\theta_0, \theta_1\}$, and the problem of testing the hypothesis $\mathcal{H}_0: \theta = \theta_0$ versus $\mathcal{H}_1: \theta = \theta_1$ will be considered. A most powerful test for this problem will be constructed and it will be proved that such a test is strictly unbiased.

4.1. Likelihood Ratio Tests

Consider the problem of testing the following simple null and alternative hypothesis:

$$\mathcal{H}_0: \theta = \theta_0, \quad \mathcal{H}_1: \theta = \theta_1. \quad (4.1.1)$$

It is supposed that *both distributions* P_{θ_0} and P_{θ_1} are (i) *continuous* or (ii) *discrete*, and $f(\mathbf{x}; \theta_i)$ stands for the density or probability function of P_{θ_i} , $i = 1, 2$. Recall that

$f(\mathbf{x}; \theta_i)$ is the likelihood of θ_i given $\mathbf{X} = \mathbf{x}$.

Hence, if $f(\mathbf{x}; \theta_1)$ is ‘substantially’ larger than $f(\mathbf{x}; \theta_0)$ it is natural to decide that θ_1 is the true parameter value, *i.e.*, that \mathcal{H}_1 holds. A class of tests based on this idea is introduced below.

Definition 4.1.1. Let $k \in [0, \infty)$ and $\psi: \mathcal{X} \rightarrow [0, 1]$ be arbitrary but fixed. The Neyman-Pearson test for $\mathcal{H}_0: \theta = \theta_0$ versus $\mathcal{H}_1: \theta = \theta_1$ determined by k and ψ is the function $\varphi_{k, \psi}: \mathcal{X} \rightarrow [0, 1]$ given by

$$\varphi_{k, \psi}(\mathbf{x}) \equiv \varphi(\mathbf{x}) = \begin{cases} 1, & \text{if } f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0) \\ 0, & \text{if } f(\mathbf{x}; \theta_1) < kf(\mathbf{x}; \theta_0) \\ \psi(\mathbf{x}), & \text{if } f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0) \end{cases} \quad (4.1.2)$$

Remark 4.1.1. An analyst using the test $\varphi_{k,\psi}$ operates as follows: After observing $\mathbf{X} = \mathbf{x}$,

- (i) If $f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)$ the null hypothesis is rejected;
- (ii) If $f(\mathbf{x}; \theta_1) < kf(\mathbf{x}; \theta_0)$ the null hypothesis is accepted;
- (iii) If $f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)$ then a Bernoulli experiment with success probability $\psi(\mathbf{x})$ is performed. If the outcome of the experiment is ‘success’, then the null hypothesis is rejected, and if the outcome is ‘faliure’, then \mathcal{H}_0 is accepted. \square

The following theorem (known as the Neyman-Pearson lemma) shows that a Neyman-Pearson test is most powerful.

Theorem 4.1.1. Given $k \in [0, \infty)$ and $\psi: \mathcal{X} \rightarrow [0, 1]$, let $\varphi_{k,\psi}$ be the critical function specified in (4.1.2). If $\tilde{\varphi}$ is other critical function satisfying

$$E_{\theta_0}[\tilde{\varphi}(\mathbf{X})] \leq E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})],$$

then

$$E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] \leq E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})].$$

In words, if

$$E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] = \alpha,$$

then

$$\varphi_{k,\psi} \text{ is a most powerful test of size } \alpha$$

for the testing problem (4.1.1).

Proof. It will be shown that

$$[\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})][f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)] \geq 0. \quad (4.1.3)$$

To verify this assertion, consider the following three exhaustive cases:

(i) $f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0) > 0$. In this context, (4.1.2) yields that $\varphi_{k,\psi}(\mathbf{x}) = 1$, so that $\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x}) = 1 - \tilde{\varphi}(\mathbf{x}) \geq 0$, since the critical function $\tilde{\varphi}$ takes values in $[0, 1]$. It follows that (4.1.3) occurs in this case.

(ii) $f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0) < 0$. Under this condition, (4.1.2) implies that $\varphi_{k,\psi}(\mathbf{x}) = 0$, and then $\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x}) = 0 - \tilde{\varphi}(\mathbf{x}) \leq 0$ and, again, (4.1.3) follows.

(iii) $f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0) = 0$. In this context, the left-hand side of (4.1.3) is null.

From (4.1.3) it follows that

$$\varphi_{k,\psi}(\mathbf{x})f(\mathbf{x}; \theta_1) - \tilde{\varphi}(\mathbf{x})f(\mathbf{x}; \theta_1) \geq k[\varphi_{k,\psi}(\mathbf{x})f(\mathbf{x}; \theta_0) - \tilde{\varphi}(\mathbf{x})f(\mathbf{x}; \theta_0)];$$

taking the summation or the integral with respect to \mathbf{x} in the discrete or continuous case, respectively, it follows that

$$E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] \geq k[E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_0}[\tilde{\varphi}(\mathbf{X})]]; \quad (4.1.4)$$

since k is nonnegative, this relation shows that

$$E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] \geq E_{\theta_0}[\tilde{\varphi}(\mathbf{X})] \Rightarrow E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})] \geq E_{\theta_1}[\tilde{\varphi}(\mathbf{X})],$$

concluding the argument. \square

4.2. Existence of Neyman-Pearson tests

The above theorem leads naturally to the following question: If $\alpha \in (0, 1)$ is arbitrary, are there $k \in [0, \infty)$ and a function $\psi: \mathcal{X} \rightarrow [0, 1]$ such that the Neyman-Pearson test $\varphi_{k,\psi}$ has size α ? The following result shows that the answer is positive.

Theorem 4.2.1. Given $\alpha \in (0, 1)$, there exist $k \in [0, \infty)$ and a function $\psi: \mathcal{X} \rightarrow [0, 1]$ such that the Neyman-Pearson $\varphi_{k,\psi}$ has size α , *i.e.*, $E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] = \alpha$.

The proof of this theorem relies on the following preliminary result.

Lemma 4.2.1. Given $\alpha \in (0, 1)$ there exists $k \in (0, \infty)$ such that

$$P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\}] \leq \alpha \quad (4.2.1)$$

and

$$P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) \geq kf(\mathbf{x}; \theta_0)\}] \geq \alpha. \quad (4.2.2)$$

Proof. For each constant c , define

$$S_c = \{\mathbf{x} \mid f(\mathbf{x}; \theta_1) > cf(\mathbf{x}; \theta_0)\} \quad (4.2.3)$$

and let

$$S = \{\mathbf{x} \mid f(\mathbf{x}; \theta_0) > 0\}$$

be the support of $f(\mathbf{x}; \theta_0)$, so that

$$P_{\theta_0}[S] = 1. \quad (4.2.4)$$

Observe that $S_{c_1} \subset S_c$ when $c_1 \geq c$, that is, the sets S_c *decrease* as c increases. The monotonicity and continuity properties of a probability measure imply that

$$P_{\theta_0}[S_{c_1}] \leq P_{\theta_0}[S_c] \quad \text{if } c_1 \geq c, \quad (4.2.5)$$

and

$$\lim_{c \nearrow \infty} P_{\theta_0}[S_c] = P_{\theta_0}[S^*] = P_{\theta_0}[S^* \cap S], \quad \text{where } S^* = \bigcap_{c \geq 0} S_c. \quad (4.2.6)$$

Notice now that

$$\begin{aligned} S \cap S^* &= \{\mathbf{x} \mid f(\mathbf{x}; \theta_0) > 0 \text{ and } f(\mathbf{x}; \theta_1) > cf(\mathbf{x}; \theta_0) \text{ for all } c > 0\} \\ &= \{\mathbf{x} \mid f(\mathbf{x}; \theta_0) > 0 \text{ and } f(\mathbf{x}; \theta_1) = \infty\} = \emptyset, \end{aligned}$$

where the last equality is due to the fact that a density or probability function attains only finite values. Therefore, by (4.2.6), $P_{\theta_0}[S^* \cap S] = 0$ and then

$$\lim_{c \nearrow \infty} P_{\theta_0}[S_c] = 0. \quad (4.2.7)$$

Next, given $\alpha \in (0, 1)$, define

$$\mathcal{A} = \{c > 0 \mid P_{\theta_0}[S_c] \leq \alpha\} \quad (4.2.8)$$

and observe that (4.2.5) and (4.2.7) yield that \mathcal{A} is a nonempty ray pointing to the right. Let k be the origin of the ray \mathcal{A} , so that

$$(k, \infty) \subset \mathcal{A} \subset [k, \infty), \quad (4.2.9)$$

where

$$k = \inf \mathcal{A}; \quad (4.2.10)$$

consequently,

$$c > k \Rightarrow c \in \mathcal{A} \Rightarrow P_{\theta_0}[S_c] \leq \alpha. \quad (4.2.11)$$

On the other hand, the specification of the sets S_c yields that

$$S_c \cap S \nearrow S_k \cap S \quad \text{as } c \searrow k,$$

and, *via* (4.2.4), the continuity property of a probability measure leads to

$$\lim_{c \searrow k} P_{\theta_0}[S_c] = \lim_{c \searrow k} P_{\theta_0}[S_c \cap S] = P_{\theta_0}[S_k \cap S] = P_{\theta_0}[S_k].$$

Combining this fact with (4.2.11) it follows that

$$P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\}] = P_{\theta_0}[S_k] \leq \alpha,$$

i.e., (4.2.1) holds. Next, it will be shown that (4.2.2) occurs by analyzing the following two exhaustive cases about the value of k :

(i) If $k = 0$, then

$$\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) \geq 0f(\mathbf{x}; \theta_0)\} = \{\mathbf{x} \mid f(\mathbf{x}; \theta_1) \geq 0\} = \mathbb{R}^k,$$

so that (4.2.2) is equivalent to the inequality $1 \geq \alpha$, a relation that is true.

(ii) Suppose now that k is positive. In this case, for any sufficiently large integer m the inclusion $k - 1/m \in (0, k)$ holds, and then $k - 1/m \notin \mathcal{A}$, by (4.2.9). In this case the specification of \mathcal{A} and (4.2.4) yield that

$$P_{\theta_0}[S \cap S_{k-1/m}] = P_{\theta_0}[S_{k-1/m}] \geq \alpha. \quad (4.2.12)$$

On the other hand, notice that

$$S \cap S_{k-1/m} = \{\mathbf{x} \mid f(\mathbf{x}; \theta_0) > 0, f(\mathbf{x}; \theta_1) > (k - 1/m)f(\mathbf{x}; \theta_0)\},$$

an expression that immediately implies that, as $m \rightarrow \infty$,

$$\begin{aligned} S \cap S_{k-1/m} &\searrow \{\mathbf{x} \mid f(\mathbf{x}; \theta_0) > 0, f(\mathbf{x}; \theta_1) \geq kf(\mathbf{x}; \theta_0)\} \\ &= S \cap \{\mathbf{x} \mid f(\mathbf{x}; \theta_1) \geq kf(\mathbf{x}; \theta_0)\}. \end{aligned}$$

Using the continuity property of a probability measure, this last display and (4.2.12) lead to

$$\begin{aligned} P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) \geq kf(\mathbf{x}; \theta_0)\}] &= P_{\theta_0}[S \cap \{\mathbf{x} \mid f(\mathbf{x}; \theta_1) \geq kf(\mathbf{x}; \theta_0)\}] \\ &= \lim_{m \rightarrow \infty} P_{\theta_0}[S \cap S_{k-1/m}] \geq \alpha, \end{aligned}$$

completing the verification of (4.2.2). \square

Proof of Theorem 4.2.1. Let k be as in Lemma 4.2.1, so that

$$P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) \geq kf(\mathbf{x}; \theta_0)\}] \geq \alpha \geq P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\}],$$

and let $\psi(\cdot) \equiv \psi$ be the constant function defined as follows:

$$\psi = \begin{cases} 1, & \text{if } P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\}] = 0 \\ \frac{\alpha - P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\}]}{P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\}]} & \text{if } P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\}] \neq 0. \end{cases} \quad (4.2.13)$$

Let $\varphi_{k,\psi}$ be the Neyman-Pearson critical function associated to this pair (k, ψ) as in (4.1.2) and observe that

$$\begin{aligned} E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] &= P_{\theta_0}[\varphi_{k,\psi}(\mathbf{X}) = 1] + \psi P_{\theta_0}[\varphi_{k,\psi}(\mathbf{X}) = \psi] \\ &= P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\}] \\ &\quad + \left(\frac{\alpha - P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\}]}{P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\}]} \right) P_{\theta_0}[\{\mathbf{x} \mid f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\}] \\ &= \alpha, \end{aligned}$$

that is, $\varphi_{k,\psi}$ has size α , completing the proof. \square

4.3. Construction of Neyman-Pearson Tests

The results in Theorems 4.1.1 and 4.2.1 can be summarized as follows: For the testing problem (4.1.1), given $\alpha \in (0, 1)$, there exists a Neyman-Pearson test with size α , and such a test is most powerful. Also, the proof of Lemma 4.2.1 and Theorem 4.2.1 allow to state the following procedure to construct a most powerful test of size $\alpha \in (0, 1)$ for the testing problem (4.1.1). Let \mathcal{X} be the set of possible values of the observation vector \mathbf{X} under the condition that θ is θ_0 or θ_1 , that is,

$$\mathcal{X} = \{\mathbf{x} \mid f(\mathbf{x}; \theta_0) > 0 \text{ or } f(\mathbf{x}; \theta_1) > 0\}.$$

- **Step 1:** For each $k \geq 0$, find the region $\mathcal{R}_k = \{\mathbf{x} \in \mathcal{X} \mid f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\}$.
- **Step 2:** Determine the size of each region \mathcal{R}_k , which is given by $P_{\theta_0}[\mathbf{X} \in \mathcal{R}_k]$.
- **Step 3:** Find the region \mathcal{R}_{k^*} such that

$$P_{\theta_0}[\mathbf{X} \in \mathcal{R}_{k^*}] \leq \alpha \quad \text{and} \quad P_{\theta_0}[\mathbf{X} \in \mathcal{R}_k] \geq \alpha \quad \text{for} \quad k < k^*;$$

notice that

$$k^* \text{ is the } \textit{smallest} \text{ nonnegative number } k \text{ such that } P_{\theta_0}[\mathbf{X} \in \mathcal{R}_k] \leq \alpha, \quad (4.3.1)$$

and that, if \mathcal{R}_k is expressed in terms of a statistic with continuous distribution, in common models the region \mathcal{R}_{k^*} is determined by solving the equation

$$P_{\theta_0}[\mathbf{X} \in \mathcal{R}_{k^*}] = \alpha. \quad (4.3.2)$$

- **Step 4:** Define the constant ψ^* as in (4.2.13) with k^* instead of k , and
- **Step 5:** Finally, perform the test using the Neyman-Pearson critical function φ_{k^*,ψ^*} given by

$$\varphi_{k^*,\psi^*}(\mathbf{x}) = \begin{cases} 1, & \text{if } f(\mathbf{X}; \theta_1) > k^* f(\mathbf{X}; \theta_0) \\ \psi^*, & \text{if } f(\mathbf{X}; \theta_1) = k^* f(\mathbf{X}; \theta_0) \\ 0, & \text{if } f(\mathbf{X}; \theta_1) < k^* f(\mathbf{X}; \theta_0). \end{cases}$$

The application of this test leads to the following operation rule:

- ▷ Reject \mathcal{H}_0 if $f(\mathbf{X}; \theta_1) > k^* f(\mathbf{X}; \theta_0)$;
- ▷ Accept \mathcal{H}_0 if $f(\mathbf{X}; \theta_1) < k^* f(\mathbf{X}; \theta_0)$;
- ▷ Randomize the decision when $f(\mathbf{X}; \theta_1) = k^* f(\mathbf{X}; \theta_0)$;

- ▷ ▷ Reject \mathcal{H}_0 with probability ψ^* ;
- ▷ ▷ Accept \mathcal{H}_0 with probability $1 - \psi^*$.

When $P_{\theta_0}[\{\mathbf{x} \in \mathcal{X} | f(\mathbf{x}; \theta_1) > k^* f(\mathbf{x}; \theta_0)\}] = 0$, the constant ψ^* does not have any influence in the size of the test and, usually, ψ^* is set equal to 1; with this assignment the above test has critical function

$$\varphi_{k^*, \psi^*}(\mathbf{x}) = \begin{cases} 1, & \text{if } f(\mathbf{X}; \theta_1) \geq k^* f(\mathbf{X}; \theta_0) \\ 0, & \text{if } f(\mathbf{X}; \theta_1) < k^* f(\mathbf{X}; \theta_0). \end{cases}$$

This procedure to construct a most powerful test will be illustrated in the following section.

Remark 4.3.1. Define

$$\frac{a}{0} = \infty \quad \text{when } a > 0$$

and, as before, let \mathcal{X} be the set where at least one of the densities (or probability functions) $f(\mathbf{x}; \theta_1)$ and $f(\mathbf{x}; \theta_0)$ is positive. In this case, for every $\mathbf{x} \in \mathcal{X}$, the relation $f(\mathbf{x}; \theta_1) > k f(\mathbf{x}; \theta_0)$ is equivalent to

$$\frac{f(\mathbf{x}; \theta_1)}{f(\mathbf{x}; \theta_0)} > k.$$

4.4. Examples

In this section the construction of a Neyman-Pearson test will be illustrated for some common models.

Example 4.4.1. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a sample of the $\mathcal{N}(\mu, \sigma_0^2)$ distribution, where σ_0 is a known positive number, $\mu \in \{\mu_0, \mu_1\}$ and $\mu_1 > \mu_0$. Consider the the testing problem

$$\mathcal{H}_0: \mu = \mu_0 \quad \text{versus} \quad \mathcal{H}_1: \mu = \mu_1.$$

To determine the Neyman-Pearson test with level α , first notice that the density of the sample vector \mathbf{X} when μ is the parameter value is given by

$$\begin{aligned} f(\mathbf{x}; \mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x_i - \mu)^2 / 2\sigma_0^2} \\ &= \frac{1}{(2\pi\sigma_0^2)^{n/2}} e^{-[\sum_{i=1}^n x_i^2 + n\mu^2] / (2\sigma_0^2)} e^{\mu \sum_{i=1}^n x_i / \sigma_0^2} \end{aligned}$$

Next, the desired test will be obtained by an application of the procedure outlined in Remark 4.3.

Step 1: Notice that

$$\begin{aligned}
f(\mathbf{x}; \mu_1) > kf(\mathbf{x}; \mu_0) &\iff \frac{f(\mathbf{x}; \mu_1)}{f(\mathbf{x}; \mu_0)} > k \\
&\iff \frac{\frac{1}{(2\pi\sigma_0^2)^{n/2}} e^{-[\sum_{i=1}^n x_i^2 + n\mu_1^2]/(2\sigma_0^2)} e^{\mu_1 \sum_{i=1}^n x_i/\sigma_0^2}}{\frac{1}{(2\pi\sigma_0^2)^{n/2}} e^{-[\sum_{i=1}^n x_i^2 + n\mu_0^2]/(2\sigma_0^2)} e^{\mu_0 \sum_{i=1}^n x_i/\sigma_0^2}} > k \\
&\iff e^{n(\mu_0^2 - \mu_1^2)/(2\sigma_0^2)} e^{(\mu_1 - \mu_0) \sum_{i=1}^n x_i/\sigma_0^2} > k \\
&\iff e^{(\mu_1 - \mu_0) \sum_{i=1}^n x_i/\sigma_0^2} > k e^{n(\mu_1^2 - \mu_0^2)/(2\sigma_0^2)} \\
&\iff \frac{(\mu_1 - \mu_0) \sum_{i=1}^n x_i}{\sigma_0^2} > \log \left(k e^{n(\mu_1^2 - \mu_0^2)/(2\sigma_0^2)} \right)
\end{aligned}$$

and then, recalling that $\mu_1 > \mu_0$,

$$f(\mathbf{x}; \mu_1) > kf(\mathbf{x}; \mu_0) \iff \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n x_i > \tilde{k},$$

where

$$\tilde{k} = \frac{\sigma_0^2}{n(\mu_1 - \mu_0)} \log \left(k e^{n(\mu_1^2 - \mu_0^2)/(2\sigma_0^2)} \right).$$

Thus, $\mathcal{R}_k = \{\mathbf{x} \mid \bar{\mathbf{x}} > \tilde{k}\}$.

Step 2: The size of the critical region \mathcal{R}_k when θ_0 is the true parameter value is given by

$$P_{\theta_0}[\mathbf{X} \in \mathcal{R}_k] = P_{\theta_0}[\bar{\mathbf{X}} > \tilde{k}] = P_{\theta_0} \left[\frac{\bar{\mathbf{X}} - \theta_0}{\sqrt{\sigma_0^2/n}} > \frac{\tilde{k} - \theta_0}{\sqrt{\sigma_0^2/n}} \right] = 1 - \Phi \left(\frac{\tilde{k} - \theta_0}{\sqrt{\sigma_0^2/n}} \right),$$

where it was used that, under \mathcal{H}_0 , the standardized sample mean $(\bar{\mathbf{X}} - \theta_0)/\sqrt{\sigma_0^2/n}$ has the standard normal distribution.

Step 3: The critical region \mathcal{R}_k is expressed in terms of $\bar{\mathbf{X}}$, whose distribution is continuous.

To find k^* —or equivalently, \tilde{k} —the following equation must be solved:

$$1 - \Phi \left(\frac{\tilde{k} - \theta_0}{\sqrt{\sigma_0^2/n}} \right) = \alpha;$$

see (4.3.2). It follows that

$$\frac{\tilde{k} - \theta_0}{\sqrt{\sigma_0^2/n}} = \Phi^{-1}(1 - \alpha) = z_{1-\alpha},$$

and then

$$\tilde{k} = \theta_0 + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha}.$$

Step 4: Since $P_{\theta_0}[\{\mathbf{X} \in \mathcal{X} | f(\mathbf{X}; \theta_1) = k^* f(\mathbf{x}; \theta_0)\}] = P_{\theta_0}[\bar{\mathbf{X}} = \tilde{k}] = 0$, the constant ψ^* is set equal to 1; see Remark 4.3(ii).

Step 5: The most powerful test with level α is determined by the following critical function:

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{if } \bar{\mathbf{x}} \geq \theta_0 + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha} \\ 0, & \text{otherwise,} \end{cases}$$

concluding the construction. □

Example 4.4.2. Let $\mathbf{X} = (X_1, X_2, X_3, X_4)$ be a sample of the *Poisson*(λ) distribution where $\lambda \in \{1, 2.5\}$. To construct a most powerful test of level $\alpha = 0.05$ for the problem

$$\mathcal{H}_0: \lambda = 2.5 \quad \text{versus} \quad \mathcal{H}_1: \lambda = 1,$$

first notice that \mathbf{X} takes values in

$$\mathcal{X} = \{\mathbf{x} = (x_1, x_2, x_3, x_4) | x_i \text{ is a nonnegative integer}\},$$

and that for $\mathbf{x} \in \mathcal{X}$,

$$f(\mathbf{x}; \lambda) = \prod_{i=1}^4 \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \lambda^{\sum_{i=1}^4 x_i} \frac{e^{-4\lambda}}{x_1! x_2! x_3! x_4!}.$$

Therefore,

$$\frac{f(\mathbf{x}; \lambda_1)}{f(\mathbf{x}; \lambda_0)} = \frac{\lambda_1^{\sum_{i=1}^4 x_i} \frac{e^{-4\lambda_1}}{x_1! x_2! x_3! x_4!}}{\lambda_0^{\sum_{i=1}^4 x_i} \frac{e^{-4\lambda_0}}{x_1! x_2! x_3! x_4!}} = \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^4 x_i} e^{-4(\lambda_1 - \lambda_0)}.$$

Using this expression, the Neyman-Pearson test is constructed as follows:

Step 1: Since $\lambda_1 = 1$ and $\lambda_0 = 2.5$, it follows that for $\mathbf{x} \in \mathcal{X}$

$$\begin{aligned} f(\mathbf{x}; \lambda_1) > k f(\mathbf{x}; \lambda_0) &\iff \frac{f(\mathbf{x}; \lambda_1)}{f(\mathbf{x}; \lambda_0)} > k \\ &\iff \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^4 x_i} e^{-4(\lambda_1 - \lambda_0)} > k \\ &\iff \left(\frac{1}{2.5}\right)^{\sum_{i=1}^4 x_i} e^{-6} > k \\ &\iff (2.5)^{\sum_{i=1}^4 x_i} < \frac{1}{k e^6} \\ &\iff \log(2.5) \sum_{i=1}^4 x_i < \log\left(\frac{1}{k e^6}\right) \end{aligned}$$

and then

$$f(\mathbf{x}; \lambda_1) > kf(\mathbf{x}; \lambda_0) \iff \bar{\mathbf{x}} < \tilde{k}$$

where

$$\tilde{k} = \frac{1}{\log(2.5)} \log\left(\frac{1}{ke^6}\right).$$

Thus, $\mathcal{R}_k = \{\mathbf{x} \in \mathcal{X} \mid f(\mathbf{x}; \lambda_1) > kf(\mathbf{x}; \lambda_0)\} = \{\mathbf{x} \in \mathcal{X} \mid \bar{\mathbf{x}} < \tilde{k}\}$; notice that \tilde{k} is a decreasing function of k .

Step 2: Under \mathcal{H}_0 , the random variables X_i are independent with common *Poisson*(2.5) distribution, so that $\sum_{i=1}^4 X_i \sim \text{Poisson}(10)$, and then the size of the critical region \mathcal{R}_k when $\lambda = 2.5$ is given by

$$P_{\lambda_0}[\mathbf{X} \in \mathcal{R}_k] = P_{\lambda_0}[X_1 + X_2 + X_3 + X_4 < \tilde{k}] = \sum_{0 \leq t < \tilde{k}} e^{-10} \frac{10^t}{t!}.$$

Step 3: If \tilde{k}^* corresponds to k^* in (4.3.1), then \tilde{k}^* is the largest value \tilde{k} such that

$$P_{\lambda_0} \left[\sum_{i=1}^4 X_i < \tilde{k} \right] \leq 0.05,$$

since \tilde{k} is a decreasing function of k . A glance to portion of the table of the *Poisson*(10) distribution displayed below shows that

$$P_{\lambda_0} \left[\sum_{i=1}^4 X_i < 5 \right] = 0.0293, \quad \text{and} \quad P_{\lambda_0} \left[\sum_{i=1}^4 X_i < 6 \right] = .0671.$$

so that $\tilde{k}^* = 5$.

n	$P[Y = n]$	$P[Y \leq n]$	$P[Y > n]$
0	0.0000	0.0000	1.0000
1	0.0005	0.0005	0.9995
2	0.0023	0.0028	0.9972
3	0.0076	0.0103	0.9897
4	0.0189	0.0293	0.9707
5	0.0378	0.0671	0.9329
6	0.0631	0.1301	0.8699
7	0.0901	0.2202	0.7798

Step 4: Since $P_{\theta_0}[\{\mathbf{X} \in \mathcal{X} \mid f(\mathbf{X}; \theta_1) = k^* f(\mathbf{x}; \theta_0)\}] = P_{\theta_0}[\bar{\mathbf{X}} = \tilde{k}^*] = P_{\theta_0}[\bar{\mathbf{X}} = 5] = 0.0378$, it follows that

$$\psi^* = \frac{.05 - P_{\lambda_0}[\bar{X} < 5]}{P_{\lambda_0}[\bar{X} = 5]} = \frac{.05 - .0293}{.0378} = .5476.$$

Step 5: The most powerful test with level $\alpha = 0.5$ is determined by the following critical function:

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{if } \bar{\mathbf{x}} < 5 \\ .5476, & \text{if } \bar{\mathbf{x}} = 5, \\ 0, & \text{if } \bar{\mathbf{x}} > 5. \end{cases}$$

The application of this test is as follows: If $\bar{\mathbf{X}} < 5$ (respectively, $\bar{\mathbf{X}} > 5$) is observed, then the null hypothesis is rejected (respectively, accepted), whereas if $\bar{\mathbf{X}} = 5$ occurs, then a Bernoulli experiment with success probability is performed, and the null hypothesis is rejected if the outcome of the Bernoulli experiment is ‘success’, while \mathcal{H}_0 is ‘accepted’ when a ‘failure’ occurs. \square

In Example 4.4.1, when \mathcal{H}_0 is valid, the decision maker will not use any randomized procedure to decide which of the null or the alternative hypothesis is true. On the other hand, in Example 4.4.2, the analyst will take a randomized decision when $\bar{\mathbf{X}} = 5$, an event that has probability 0.0378 if \mathcal{H}_0 is true. In the following example, the decision maker *always* uses randomization to decide which of the null or alternative hypothesis holds.

Example 4.4.3. Let X be a random variable with the *Uniform*(0, θ) distribution, where $\theta \in \{1, 2\}$, and consider the problem of testing $\mathcal{H}_0: \theta = 1$ versus $\mathcal{H}_1: \theta = 2$. A test with level $\alpha = 0.05$ is constructed as follows:

Notice that, since the parameter space is $\{1, 2\}$, then X takes values in

$$\mathcal{X} = (0, 2)$$

and that for $x \in \mathcal{X}$,

$$f(x; \theta) = \frac{1}{\theta} I_{(0, \theta)}(x).$$

Therefore,

$$\frac{f(x; \theta_1)}{f(x; \theta_0)} = \frac{\frac{1}{\theta_1} I_{(0, \theta_1)}(x)}{\frac{1}{\theta_0} I_{(0, \theta_0)}(x)} = \frac{\theta_0 I_{(0, \theta_1)}(x)}{\theta_1 I_{(0, \theta_0)}(x)} = \begin{cases} \infty, & \text{if } x \in [1, 2) \\ 1/2, & \text{if } x \in (0, 1), \end{cases} \quad (4.4.1)$$

where the last equality used that $\theta_0 = 1$ and $\theta_1 = 2$. Using this expression, the Neyman-Pearson test is constructed as follows:

Step 1: Observe that for each $x \in \mathcal{X}$ and $k \in [0, \infty)$

$$f(x; \theta_1) > kf(x; \theta_0) \iff \frac{f(x; \theta_1)}{f(x; \theta_0)} > k$$

and then

$$\mathcal{R}_k = \{x \in \mathcal{X} \mid f(x; \theta_1) > kf(x; \theta_0)\} = \begin{cases} [1, 2), & \text{if } k > 1/2 \\ (0, 2), & \text{if } k \in [0, 1/2]. \end{cases}$$

Step 2: Under \mathcal{H}_0 , the random variable X has the *Uniform*(0, 1) distribution, so that the size of the region \mathcal{R}_k is given by

$$P_{\theta_0}[X \in \mathcal{R}_k] = \begin{cases} P_{\theta_0}[X \in [1, 2)] = 0, & \text{if } k > 1/2 \\ P_{\theta_0}[X \in (0, 2)] = 1, & \text{if } k \leq 1/2. \end{cases}$$

Step 3: According to the previous display, the minimum (infimum) k^* of all the values k such that $P_{\theta_0}[X \in \mathcal{R}_k] \leq \alpha = 0.05$ is

$$k^* = 1/2.$$

Step 4: By (4.4.1),

$$\begin{aligned} P_{\theta_0}[f(X; \theta_1) = k^* f(X; \theta_0)] &= P_{\theta_0} \left[\frac{f(X; \theta_1)}{f(X; \theta_0)} = k^* \right] \\ &= P_{\theta_0} \left[\frac{f(X; \theta_1)}{f(X; \theta_0)} = \frac{1}{2} \right] \\ &= P_{\theta_0}[X \in (0, 1)] = 1. \end{aligned}$$

Observing that $P_{\theta_0}[f(X; \theta_1) > k^* f(X; \theta_0)] = P_{\theta_0}[X \in [1, 2)] = 0$ it follows that

$$\psi^* = \frac{0.05 - P_{\lambda_0}[X \in [1, 2)]}{P_{\lambda_0}[X \in (0, 1)]} = \frac{0.05 - 0}{1} = 0.05.$$

Step 5: Observing that $\{f(X; \theta_1) < k^* f(X; \theta_0)\} = \emptyset$, the most powerful test with level $\alpha = 0.05$ is determined by the following critical function:

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{if } x \in [1, 2) \\ 0.05, & \text{if } x \in (0, 1). \end{cases}$$

Notice that under \mathcal{H}_0 the variable X takes values in $(0, 1)$ with probability 1, so that, when the null hypothesis is true, the analyst will always decide which of the null or alternative hypothesis holds using a random mechanism. \square

Chapter 5

Unbiasedness Property

This chapter concerns two basic properties that are satisfied by a most powerful test. The first one concerns a natural and intuitive condition: Under a reasonable test, the probability of rejecting the null hypothesis when it is false must be at least equal to the probability of rejection when \mathcal{H}_0 is true; a test with this property is called *unbiased*. The second property refers to the information that must be gathered in order to apply a most powerful test. The main result in this direction establishes that, when a simple hypothesis is being tested versus another simple hypothesis, a most powerful test can be always specified in terms of a sufficient statistic.

5.1. Unbiased Tests

To begin with, the most fundamental property of a test is introduced below.

Definition 5.1.1. A test γ for the problem $\mathcal{H}_0: \theta \in \Theta_0$ versus $\mathcal{H}_1: \theta \in \Theta \setminus \Theta_0$ is *unbiased*, if the following condition holds:

$$\max_{\theta \in \Theta_0} \pi_\gamma(\theta) \leq \pi_\gamma(\tilde{\theta}), \quad \tilde{\theta} \in \Theta \setminus \Theta_0.$$

This unbiasedness property for a test can be phrased as follows: The test γ is unbiased if the probability of rejecting \mathcal{H}_0 when it is false is always larger than (or equal to) the probability of rejection when \mathcal{H}_0 is true. As the following lemma shows, this intuitively appealing property is satisfied by a uniformly most powerful test.

Lemma 5.1.1. Let γ be a uniformly most powerful test of size α for the problem $\mathcal{H}_0: \theta \in \Theta_0$ versus $\mathcal{H}_1: \theta \in \Theta \setminus \Theta_0$. In this case γ is unbiased.

Proof. Let γ be a most powerful test of size α for the problem $\mathcal{H}_0: \theta \in \Theta_0$ versus $\mathcal{H}_1: \theta \in \Theta \setminus \Theta_0$. In this case

$$\max_{\theta \in \Theta_0} \pi_\gamma(\theta) = \alpha$$

and, for any other test $\tilde{\gamma}$ of size α ,

$$\pi_{\tilde{\gamma}}(\tilde{\theta}) \leq \pi_\gamma(\tilde{\theta}) \quad \text{for every } \tilde{\theta} \in \Theta \setminus \Theta_0. \quad (5.1.1)$$

Consider now the test $\tilde{\gamma}$ determined by the critical function $\tilde{\varphi}(\mathbf{x}) = \alpha$ for every possible value \mathbf{x} of the observation vector \mathbf{X} . This test satisfies that

$$\pi_{\tilde{\gamma}}(\theta) = E_\theta[\tilde{\varphi}(\mathbf{X})] = E_\theta[\alpha] = \alpha, \quad \theta \in \Theta;$$

in particular, $\tilde{\gamma}$ has size α , and using the relation (5.1.1) with this test $\tilde{\gamma}$ it follows that, for every $\tilde{\theta} \in \Theta \setminus \Theta_0$, $\alpha = E_{\tilde{\theta}}[\tilde{\varphi}(\mathbf{X})] = \pi_{\tilde{\gamma}}(\tilde{\theta}) \leq \pi_\gamma(\tilde{\theta})$, and then

$$\max_{\theta \in \Theta_0} \pi_\gamma(\theta) = \alpha \leq \pi_\gamma(\tilde{\theta}),$$

showing that γ is an unbiased test. □

As an immediate corollary of the previous result, the unbiasedness property of a Neyman-Pearson test is obtained.

Corollary 5.1.1. Consider the testing problem

$$\mathcal{H}_0: \theta = \theta_0 \quad \text{versus} \quad \mathcal{H}_1: \theta = \theta_1, \quad (5.1.2)$$

and let $\varphi_{k, \psi}$ be (the critical function of a) Neyman-Pearson test for this problem. In this case $\varphi_{k, \psi}$ is an unbiased test, that is,

$$\pi_{\varphi_{k, \psi}}(\theta_0) \leq \pi_{\varphi_{k, \psi}}(\theta_1). \quad (5.1.3)$$

Proof. Theorem 4.1.1 establishes that the test $\varphi_{k, \psi}$ is most powerful for the problem (5.1.2) with size $\pi_{\varphi_{k, \psi}}(\theta_0)$, and consequently it is unbiased, by Lemma 5.1.1, that is, the probability of rejecting \mathcal{H}_0 when it is true—which is given by $\pi_{\varphi_{k, \psi}}(\theta_0)$ —is less than or equal to the probability $\pi_{\varphi_{k, \psi}}(\theta_1)$ of rejecting the null hypothesis when it is false, which is the desired conclusion. □

In a later section it will be proved that, under mild conditions, the inequality in (5.1.3) is strict.

5.2. Dependence on a Sufficient Statistic

In this section it will be shown that a Neyman-Pearson test depends on a sufficient statistic; see Dudewicz and Mishra (1998) or Wackerly *et al.* (2009).

Lemma 5.2.1. Let \mathbf{X} be a random vector with density or probability function $f(\mathbf{x}; \theta)$. Consider the testing problem $\mathcal{H}_0: \theta = \theta_0$ versus $\mathcal{H}_1: \theta = \theta_1$, and let $S = S(\mathbf{X})$ be a sufficient statistic for the underlying model. In this context, any Neyman-Pearson test $\varphi_{k, \psi}$ with ψ a constant, is a function of the sufficient statistic S .

Recalling that if a statistic $S = S(\mathbf{X})$ is sufficient, then the analyst may record S and *nothing more* without missing any useful information provided by the observed data \mathbf{X} , it follows from the previous lemma that an observer retaining only the value of $S(\mathbf{X})$ will be always able to construct a most powerful test.

Proof. Since $S = S(\mathbf{X})$ is sufficient, it follows that the density (or probability function) $f(\mathbf{x}; \theta)$ can be expressed as

$$f(\mathbf{x}; \theta) = h(\mathbf{x})g(\theta; S(\mathbf{x}))$$

for certain nonnegative functions h and g . Set

$$A = \{s \mid g(\theta_1; s) > kg(\theta_0; s)\},$$

$$B = \{s \mid g(\theta_1; s) < kg(\theta_0; s)\},$$

$$C = \{s \mid g(\theta_1; s) = kg(\theta_0; s)\},$$

and recall that the set of possible values of \mathbf{X} is given by $\mathcal{X} = \{\mathbf{x} \mid f(\mathbf{x}; \theta_0) > 0\} \cup \{\mathbf{x} \mid f(\mathbf{x}; \theta_1) > 0\}$, so that $h(\mathbf{x}) > 0$ for every $\mathbf{x} \in \mathcal{X}$. Therefore, for every $k \geq 0$ and $\mathbf{x} \in \mathcal{X}$

$$\begin{aligned} f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0) &\iff h(\mathbf{x})g(\theta_1; S(\mathbf{x})) > kh(\mathbf{x})g(\theta_0; S(\mathbf{x})) \\ &\iff g(\theta_1; S(\mathbf{x})) > kg(\theta_0; S(\mathbf{x})) \end{aligned}$$

where the positivity of $h(\mathbf{x})$ was used to set the second equivalence. Thus,

$$\mathbf{x} \in \{f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\} \iff S(\mathbf{x}) \in \{s \mid g(\theta_1; s) > kg(\theta_0; s)\} = A.$$

Similarly,

$$\mathbf{x} \in \{f(\mathbf{x}; \theta_1) < kf(\mathbf{x}; \theta_0)\} \iff S(\mathbf{x}) \in \{s \mid g(\theta_1; s) < kg(\theta_0; s)\} = B$$

and

$$\mathbf{x} \in \{f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\} \iff S(\mathbf{x}) \in \{s \mid g(\theta_1; s) = kg(\theta_0; s)\} = C.$$

Combining these three last displays with the formula (4.1.2) for $\varphi_{k,\psi}$ it follows that

$$\varphi_{k,\psi}(\mathbf{x}) = \begin{cases} 1, & \text{if } S(\mathbf{x}) \in A, \\ 0, & \text{if } S(\mathbf{x}) \in B, \\ \psi, & \text{if } S(\mathbf{x}) \in C \end{cases}$$

or, in a more compact way,

$$\varphi_{k,\psi}(\mathbf{x}) = I_A(S(\mathbf{x})) + \psi I_C(S(\mathbf{x}));$$

keeping in mind that ψ is constant, either of these expressions show that $\varphi_{k,\psi}$ is a function of $S(\mathbf{x})$. \square

5.3. Additional Examples

The following examples illustrate the construction of Neyman-Pearson tests in some specific cases, and show explicitly that the tests are unbiased and depend on a sufficient statistic. First, some technical facts are stated in the remarks before each example; the proofs of the results can be found in Dudewicz and Mishra (1998) or Wackerly *et al.* (2009).

Remark 5.3.1. Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a sample of the density

$$f(x; \theta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} I_{(0,\infty)}(x),$$

which corresponds to the *Gamma*(α, β) distribution. In this case

- (i) $X_1 + X_2 + \dots + X_n \sim \text{Gamma}(n\alpha, \beta)$,
- (ii) $(X_1 + X_2 + \dots + X_n)/\beta \sim \text{Gamma}(n\alpha, 1)$, and
- (iii) $2(X_1 + X_2 + \dots + X_n)/\beta \sim \text{Gamma}(n\alpha, 1/2)$.

Also,

- (iv) The *Gamma*($k/2, 1/2$) distribution is referred to as the *chi squared distribution with k degrees of freedom*, and is denoted as χ_k^2 . \square

Exercise 5.3.1. Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a sample of the *Gamma*(α, β) distribution ($\alpha, \beta > 0$), which has density

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I_{(0,\infty)}(x), \quad x \in \mathbb{R},$$

and assume that α is a *known* positive number.

- (a) For the testing problem $\mathcal{H}_0: \beta = 1$ versus $\mathcal{H}_1: \beta = 2$
- (i) Show explicitly that the best (most powerful) test at a specified level α_0 is based on the sufficient statistic $S = X_1 + X_2 + \cdots + X_n$,
- (ii) Find a most powerful test at level α_0 , and
- (iii) Show directly that this test is unbiased.
- (b) Repeat part (a) for $\mathcal{H}_0: \beta = \beta_0$ versus $\mathcal{H}_1: \beta = \beta_1$, where $\beta_1 > \beta_0$, and hence generalize part (a) to find a most powerful test, and
- (c) Show that if φ is a Neyman-Pearson test of the form

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{if } X_1 + X_2 + \cdots + X_n > k \\ \psi, & \text{if } X_1 + X_2 + \cdots + X_n = k \\ 0, & \text{otherwise,} \end{cases} \quad (5.3.1)$$

then the corresponding power function is a nondecreasing function of β .

Solution. The density $f_{\mathbf{X}}(\mathbf{x}; \alpha, \beta)$ of \mathbf{X} is supported on the set of all n -dimensional vectors \mathbf{x} with positive components. If $x_i > 0$ for $i = 1, 2, \dots, n$,

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}; \alpha, \beta) &= \prod_{i=1}^n \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I_{(0,\infty)}(x) \\ &= \frac{1}{\Gamma(\alpha)^n} \left(\prod_{i=1}^n x_i \right)^{\alpha-1} \frac{1}{\beta^{n\alpha}} e^{-\sum_{i=1}^n x_i/\beta} \\ &= h(\mathbf{x}) \frac{1}{\beta^{n\alpha}} e^{-\sum_{i=1}^n x_i/\beta} \end{aligned} \quad (5.3.2)$$

where

$$h(\mathbf{x}) = \frac{1}{\Gamma(\alpha)^n} \left(\prod_{i=1}^n x_i \right)^{\alpha-1}.$$

- (a) If \mathbf{x} has positive components,

$$\begin{aligned} f(\mathbf{x}; \alpha, 2) > \tilde{k} f(\mathbf{x}; \alpha, 1) &\iff \frac{f(\mathbf{x}; \alpha, 2)}{f(\mathbf{x}; \alpha, 1)} > \tilde{k} \\ &\iff \frac{h(\mathbf{x}) \frac{1}{2^{n\alpha}} e^{-\sum_{i=1}^n x_i/2}}{h(\mathbf{x}) e^{-\sum_{i=1}^n x_i}} > \tilde{k} \\ &\iff e^{\sum_{i=1}^n x_i/2} > \tilde{k} 2^{n\alpha} \\ &\iff \sum_{i=1}^n x_i > 2 \log(\tilde{k} 2^{n\alpha}) \equiv k. \end{aligned}$$

and it follows that a Neyman-Pearson can be described as

$$\varphi(\mathbf{X}) = \begin{cases} 1, & \text{if } X_1 + X_2 + \cdots + X_n \geq k \\ 0, & \text{if } X_1 + X_2 + \cdots + X_n < k; \end{cases} \quad (5.3.3)$$

see (4.1.2). Notice that the event $[f(\mathbf{X}; \alpha, 2) = \tilde{k}f(\mathbf{X}; \alpha, 1)] = [X_1 + X_2 + \cdots + X_n = k]$ has been incorporated to the rejection error, a choice that does not affect the size of the test, since $X_1 + X_2 + \cdots + X_n$ has $Gamma(n\alpha, \beta)$ distribution, which is continuous, and then the event has probability zero. The above expression shows that the Neyman-Pearson test depends only on the sufficient statistic $S = X_1 + X_2 + \cdots + X_n$. The power function of the test is $\pi_\varphi(\beta) = P_\beta[X_1 + X_2 + \cdots + X_n \geq k]$, and the level of this test for the problem $\mathcal{H}_0: \beta = 1$ versus $\mathcal{H}_1: \beta = 2$ is $\pi_\varphi(1)$. To obtain the level (size) a_0 , the constant k must be selected in such a way that

$$a_0 = \pi(1) = P_1[X_1 + X_2 + \cdots + X_n > k]. \quad (5.3.4)$$

Notice that when $\beta = 1$, $S = X_1 + X_2 + \cdots + X_n \sim \Gamma(n\alpha, 1)$, so that the density of S is given by

$$f_S(y) = \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} I_{(0,\infty)}(x).$$

Therefore, the number k in (5.3.4) must be selected as the (right-hand side) percentil p_{a_0} of size a_0 for this density:

$$\int_{p_{a_0}}^{\infty} \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} I_{(0,\infty)}(x) dx = a_0,$$

and the test of level a_0 is obtained by replacing k by p_{a_0} in the above expression for the critical function φ ; the power function is

$$\pi(\beta) = P_\beta[X_1 + X_2 + \cdots + X_n > p_{a_0}].$$

Of course, $\pi(1) = a_0$, and it will be shown that $\pi(2) > a_0$. To achieve this goal, notice that if X_1, X_2, \dots, X_n is a sample of the $Gamma(\alpha, \beta)$ distribution, then X_i/β , $i = 1, 2, \dots, n$ are *i.i.d.* with common distribution $Gamma(\alpha, 1)$, so that $(X_1 + X_2 + \cdots + X_n)/\beta \sim Gamma(n\alpha, 1)$, that is,

$$\frac{X_1 + X_2 + \cdots + X_n}{\beta} \quad \text{has density} \quad \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} I_{(0,\infty)}(x). \quad (5.3.5)$$

Therefore,

$$\pi(\beta) = P_\beta \left[\frac{X_1 + X_2 + \cdots + X_n}{\beta} > \frac{p_{a_0}}{\beta} \right] = \int_{p_{a_0}/\beta}^{\infty} \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} dx.$$

In particular,

$$\pi(2) = \int_{p_{a_0}/2}^{\infty} \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} dx > \int_{p_{a_0}}^{\infty} \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} dx = a_0 = \pi(1),$$

showing that the test (5.3.3) with $k = p_{a_0}$ is unbiased for the testing problem $\mathcal{H}_0: \beta = 1$ versus $\mathcal{H}_1: \beta = 2$.

(b) Consider the testing problem $\mathcal{H}_0: \beta = \beta_0$ versus $\mathcal{H}_1: \beta = \beta_1$, where $\beta_1 > \beta_0$. Using (5.3.2), it follows that if \mathbf{x} has positive components,

$$\begin{aligned} f(\mathbf{x}; \alpha, \beta_1) > \tilde{k} f(\mathbf{x}; \alpha, \beta_0) &\iff \frac{f(\mathbf{x}; \alpha, \beta_1)}{f(\mathbf{x}; \alpha, \beta_0)} > \tilde{k} \\ &\iff \frac{h(\mathbf{x}) \frac{1}{\beta_1^{n\alpha}} e^{-\sum_{i=1}^n x_i/\beta_1}}{h(\mathbf{x}) \frac{1}{\beta_0^{n\alpha}} e^{-\sum_{i=1}^n x_i/\beta_0}} > \tilde{k} \\ &\iff e^{(1/\beta_0 - 1/\beta_1) \sum_{i=1}^n x_i} > \tilde{k} \left(\frac{\beta_1}{\beta_0} \right)^{n\alpha} \\ &\iff \sum_{i=1}^n x_i > \frac{\beta_0 \beta_1}{\beta_1 - \beta_0} \log \left(\tilde{k} \left(\frac{\beta_1}{\beta_0} \right)^{n\alpha} \right) \equiv k. \end{aligned}$$

Thus, incorporating the set $\{\mathbf{x} \mid f_{\mathbf{X}}(\mathbf{x}; \alpha, \beta_1) = \tilde{k} f_{\mathbf{X}}(\mathbf{x}; \alpha, \beta_0)\}$ to the rejection region, the Neyman-Pearson test for the present testing problem can be written as in (5.3.3), just in terms of the sufficient statistic $S = X_1 + X_2 + \cdots + X_n$. To determine the value of k to obtain a test of size a_0 , notice that the power function of the test is given by $\pi(\beta) = P_{\beta}[X_1 + X_2 + \cdots + X_n \geq k]$, and then k must be selected in such a way that

$$a_0 = \pi(\beta_0) = P_{\beta_0}[X_1 + X_2 + \cdots + X_n > k]. \quad (5.3.6)$$

Recalling that $(X_1 + X_2 + \cdots + X_n)/\beta$ has the *Gamma*($n\alpha, 1$) distribution when β is the true parameter value, it follows that

$$\pi(\beta) = P_{\beta} \left[\frac{X_1 + X_2 + \cdots + X_n}{\beta} > \frac{k}{\beta} \right] = \int_{k/\beta}^{\infty} \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} dx,$$

so that (5.3.6) is equivalent to

$$a_0 = \int_{k/\beta_0}^{\infty} \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} dx,$$

that is,

$$\frac{k}{\beta_0} = p_{a_0},$$

the right percentil of order a_0 of the *Gamma*($\alpha, 1$) distribution. Thus, $k = \beta_0 p_{a_0}$, and the Neyman-Person test is given by

$$\varphi(\mathbf{X}) = \begin{cases} 1, & \text{if } X_1 + X_2 + \cdots + X_n \geq \beta_0 p_{a_0} \\ 0, & \text{if } X_1 + X_2 + \cdots + X_n < \beta_0 p_{a_0}. \end{cases}$$

Now, it will be shown that this is an unbiased test. Indeed,

$$\begin{aligned} \pi(\beta_1) &= P_{\beta_1}[X_1 + X_2 + \cdots + X_n > \beta_0 p_{a_0}] \\ &= P_{\beta_1} \left[\frac{X_1 + X_2 + \cdots + X_n}{\beta_1} > \frac{\beta_0 p_{a_0}}{\beta_1} \right] \\ &= \int_{\beta_0 p_{a_0}/\beta_1}^{\infty} \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} I_{(0, \infty)}(x) dx, \end{aligned}$$

where (5.3.5) was used to set the second equality; since $\beta_0 < \beta_1$, it follows that $\beta_0 p_{a_0} / \beta_1 < p_{a_0}$, and then

$$\pi(\beta_1) > \int_{p_{a_0}}^{\infty} \frac{1}{\Gamma(n\alpha)} x^{n\alpha-1} e^{-x} dx = a_0 = \pi(\beta_0),$$

showing that the test is unbiased.

(iii) As already mentioned, $P_\beta[X_1 + X_2 + \cdots + X_n = k] = 0$, so that the power function of the test (5.3.1) is given by

$$\begin{aligned} \pi(\beta) &= P_\beta[X_1 + X_2 + \cdots + X_n > k] + \psi P_\beta[X_1 + X_2 + \cdots + X_n = k] \\ &= P_\beta[X_1 + X_2 + \cdots + X_n \geq k]. \end{aligned}$$

Let $\beta_0 < \beta_1$ be two positive numbers and set $a_0 = \pi(\beta_0)$, so that

$$a_0 = P_{\beta_0}[X_1 + X_2 + \cdots + X_n \geq k].$$

In this case, as it was established in part (b), $k = \beta_0 p_{a_0}$ and

$$\pi(\beta_1) = P_{\beta_1}[X_1 + X_2 + \cdots + X_n \geq \beta_0 p_{a_0}] > a_0 = \pi(\beta_0),$$

showing that π is increasing. □

Remark 5.3.2. The following fact will be useful in the analysis of the example below. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a sample of the *Exponential*(λ) distribution, so that

$$P_\lambda[X_i > x] = e^{-\lambda x}, \quad x \geq 0.$$

Observing that for every nonnegative x

$$[\min(X_i) > x] = \bigcap_{i=1}^n [X_i > x],$$

the independence of the variables X_i yields that

(i) $P_\lambda[\min(X_i) > x] = \prod_{i=1}^n P_\lambda[X_i > x] = e^{-n\lambda x}$ for all $x \geq 0$, that is,

$$\min(X_i) \sim \text{Exponential}(n\lambda).$$

Consequently, an application of the change of variable formula, leads to the following result:

(ii) $n\lambda \min(X_i) \sim \text{Exponential}(1)$. □

Exercise 5.3.2. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a random sample of the displaced exponential density

$$f(x; \theta) = e^{-(x-\theta)} I_{(\theta, \infty)}(x).$$

Use the Neyman-Pearson Lemma to determine an α -level most powerful test for $\mathcal{H}_0: \theta = \theta_0$ versus $\mathcal{H}_1: \theta = \theta_1$, where $\theta_1 > \theta_0$.

Solution. In the context of this example the parameter space is $\{\theta_0, \theta_1\}$, and the set \mathcal{X} of possible values of \mathbf{X} consists of all n -dimensional vectors \mathbf{x} whose components are all larger than θ_0 . For all $\mathbf{x} \in \mathcal{X}$,

$$f_{\mathbf{X}}(\mathbf{x}; \theta) = \prod_{i=1}^n e^{-(x_i-\theta)} I_{(\theta, \infty)}(x_i) = e^{n\theta} e^{-\sum_{i=1}^n x_i} I_{(\theta, \infty)}(\min x_i).$$

From this point, the ratio $f(\mathbf{x}; \theta_1)/f(\mathbf{x}; \theta_0)$ can be evaluated as

$$\frac{f(\mathbf{x}; \theta_1)}{f(\mathbf{x}; \theta_0)} = \frac{e^{n\theta_1} e^{-\sum_{i=1}^n x_i} I_{(\theta_1, \infty)}(\min x_i)}{e^{n\theta_0} e^{-\sum_{i=1}^n x_i} I_{(\theta_0, \infty)}(\min x_i)} = e^{n(\theta_1-\theta_0)} \frac{I_{(\theta_1, \infty)}(\min x_i)}{I_{(\theta_0, \infty)}(\min x_i)},$$

and recalling that $\theta_1 > \theta_0$,

$$\frac{f(\mathbf{x}; \theta_1)}{f(\mathbf{x}; \theta_0)} = \begin{cases} 0, & \text{if } \theta_0 < \min x_i \leq \theta_1 \\ e^{n(\theta_1-\theta_0)}, & \text{if } \min x_i > \theta_1. \end{cases}$$

This expression shows that $f(\mathbf{x}; \theta_1)/f(\mathbf{x}; \theta_0)$ is an increasing function of $\min x_i$; indeed,

$$\frac{f(\mathbf{x}; \theta_1)}{f(\mathbf{x}; \theta_0)} = g(\min x_i),$$

where the increasing function $g(\cdot)$ is given by

$$g(t) = \begin{cases} 0, & \text{if } t \leq \theta_1 \\ e^{n(\theta_1-\theta_0)}, & \text{if } t > \theta_1. \end{cases}$$

Consequently, a most powerful test for $\mathcal{H}_0: \theta = \theta_0$ versus $\mathcal{H}_1: \theta = \theta_1$ is given by

$$\varphi(\mathbf{X}) = \begin{cases} 1, & \text{if } \min X_i > k \\ \psi & \text{if } \min X_i = k \\ 0, & \text{if } \min X_i < k, \end{cases}$$

where k and $\psi \in [0, 1]$ are determined to achieve the desired size $\alpha = E_{\theta_0}[\varphi(\mathbf{X})]$. To complete the specification of the test, notice that if θ_0 is the true parameter value, then $X_i - \theta_0$ has the exponential density $e^x I_{(0, \infty)}(x)$, that is, $X_i - \theta_0 \sim \text{Exponential}(1)$. Consequently,

$$(\min X_i) - \theta_0 = \min(X_i - \theta_0) \sim \text{Exponential}(n)$$

(see Remark 5.3.2) and than

$$n[(\min X_i) - \theta_0] \sim \text{Exponential}(1).$$

It follows that $\min X_i$ has a continuous distribution, and then ψ can be selected arbitrarily in $[0, 1]$, say $\psi = 1$. To find the value of k , notice that

$$\begin{aligned} E_{\theta_0}[\min X_i > k] &= P_{\theta_0}[\min X_i > k] \\ &= P_{\theta_0}[n[\min X_i - \theta_0] > n[k - \theta_0]] \\ &= e^{-n(k-\theta_0)}. \end{aligned}$$

Thus, $E_{\theta_0}[\min X_i > k] = \alpha$ if and only if $\alpha = e^{-n(k-\theta_0)}$, and then

$$k = \theta_0 - \frac{1}{n} \log(\alpha) = \theta_0 + \frac{1}{n} \log\left(\frac{1}{\alpha}\right).$$

In short: A most powerful test with level α is obtained by rejecting \mathcal{H}_0 if

$$\min X_i \geq \theta_0 - \log(\alpha)/n,$$

and accepting \mathcal{H}_0 otherwise. □

5.4. Strict Unbiasedness

In Corollary 5.1.1 it was show that any Neyman-Pearson test $\varphi_{k, \psi}$ for the problem $\mathcal{H}_0: \theta = \theta_0$ versus $\mathcal{H}_1: \theta = \theta_1$ is unbiased, that is,

$$E_{\theta_0}[\varphi_{k, \psi}(\mathbf{X})] \leq E_{\theta_1}[\varphi_{k, \psi}(\mathbf{X})].$$

The main objective of this section is to show that, in general, the inequality in this relation is strict, a result that is formally stated as follows.

Theorem 5.4.1. Let \mathbf{X} be a random vector with density or probability function $f(\mathbf{x}; \theta)$, where $\theta \in \{\theta_0, \theta_1\}$ and, given $\alpha \in (0, 1)$, let $\phi_{k, \psi}$ be a Neyman-Pearson test of level α for the problem

$$\mathcal{H}_0: \theta = \theta_0 \quad \text{versus} \quad \mathcal{H}_1: \theta = \theta_1.$$

In this case the test $\phi_{k, \psi}$ is *strictly unbiased*, that is,

$$E_{\theta_0}[\varphi_{k, \psi}(\mathbf{X})] < E_{\theta_1}[\varphi_{k, \psi}(\mathbf{X})]. \tag{5.4.1}$$

The proof of this result relies on the following theorem analyzing the relation between two most powerful tests.

Theorem 5.4.2. Consider the testing problem () and, given $\alpha \in (0, 1)$, let $\varphi_{k, \psi}$ be the Neyman-Pearson test of size α , where $k \geq 0$ and $\psi: \mathcal{X} \rightarrow [0, 1]$ and define \mathcal{D} as the region where $f(\mathbf{x}; \theta_1)$ and $kf(\mathbf{x}; \theta_0)$ are different, that is,

$$\mathcal{D} = \{\mathbf{x} \in \mathcal{X} \mid f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_0)\} \cup \{\mathbf{x} \in \mathcal{X} \mid f(\mathbf{x}; \theta_1) < kf(\mathbf{x}; \theta_0)\}. \tag{5.4.2}$$

With this notation, if $\tilde{\varphi}$ is other a most powerful test of size α , then

$\tilde{\varphi}$ and $\varphi_{k,\psi}$ coincide on the region \mathcal{D} .

More precisely,

$$P_\theta[[\mathbf{X} \in \mathcal{D}] \cap [\varphi_{k,\psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})]] = P_\theta[\mathbf{X} \in \mathcal{D}], \quad \theta = \theta_0, \theta_1. \quad (5.4.3)$$

Remark 5.4.1. (a) Suppose that $P_\theta[\mathcal{D}] \neq 0$. In this case, notice that the equality in (5.4.3) is equivalent to

$$P_\theta[\varphi_{k,\psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X}) | \mathbf{X} \in \mathcal{D}] = \frac{P_\theta[[\mathbf{X} \in \mathcal{D}] \cap [\varphi_{k,\psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})]]}{P_\theta[\mathbf{X} \in \mathcal{D}]} = 1.$$

In words, this equality means that, under the condition that $\mathbf{X} \in \mathcal{D}$ has been observed, with probability 1 the critical functions $\tilde{\varphi}$ and $\varphi_{k,\psi}$ coincide.

(b) The following simple analytical facts will be useful:

(i) If G is a nonnegative function defined on a region $\mathcal{X} \subset \mathbb{R}^n$, then

$$\int_{\mathcal{X}} G(\mathbf{x}) d\mathbf{x} = 0 \Rightarrow \int_{\mathcal{A}} d\mathbf{x} = 0, \quad \text{where } \mathcal{A} = \{\mathbf{x} \in \mathcal{X} | G(\mathbf{x}) \neq 0\}$$

(ii) For $\mathcal{A} \subset \mathbb{R}^n$

$$\int_{\mathcal{A}} d\mathbf{x} = 0 \Rightarrow \int_{\mathcal{A}} f(\mathbf{x}) d\mathbf{x} = 0 \quad \text{for any } f: \mathcal{A} \rightarrow \mathbb{R}.$$

□

The proof of Theorem 5.4.2 is particularly simple when the observation vector \mathbf{X} is discrete, and for this reason the argument is presented in two parts.

Proof of Theorem 5.4.2. The backbone of the argumentation is the relation

$$0 \leq [\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})][f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)], \quad \mathbf{x} \in \mathcal{X}; \quad (5.4.4)$$

established in the proof of Theorem 4.1.1. To begin with, recall that the Neyman-Pearson test $\varphi_{k,\psi}$ is most powerful for the problem () at level α . Since $\tilde{\varphi}$ is also most powerful at the significance level α , it follows that the power functions of these tests coincide at $\theta = \theta_1$, that is,

$$E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] = \pi_{\tilde{\varphi}}(\theta_1) = \pi_{\varphi_{k,\psi}}(\theta_1) = E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})]$$

i.e.,

$$E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] = 0.$$

On the other hand, by construction, the size of the test $\varphi_{k,\psi}$ is α , that is, $E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] = \alpha$ whereas, since $\tilde{\varphi}$ is most powerful at size α , it follows that $E_{\theta_0}[\tilde{\varphi}(\mathbf{X})] \leq \alpha$; see Definition 3.4.3. Consequently, $E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_0}[\tilde{\varphi}(\mathbf{X})] \geq 0$ and then, recalling that k is nonnegative number,

$$-k[E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_0}[\tilde{\varphi}(\mathbf{X})]] \leq 0.$$

Combining these two last displays it follows that

$$E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] - k[E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_0}[\tilde{\varphi}(\mathbf{X})]] \leq 0. \quad (5.4.5)$$

From this point, it is convenient to analyze the discrete and continuous cases separately.

Case 1: The observation vector \mathbf{X} is discrete.

Taking the summation over $\mathbf{x} \in \mathcal{X}$ in both sides of (5.4.4) it follows that

$$\begin{aligned} 0 &\leq \sum_{\mathbf{x} \in \mathcal{X}} [\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})][f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)] \\ &= \sum_{\mathbf{x} \in \mathcal{X}} [\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})]f(\mathbf{x}; \theta_1) - k \sum_{\mathbf{x} \in \mathcal{X}} [\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})]f(\mathbf{x}; \theta_0) \\ &= \sum_{\mathbf{x} \in \mathcal{X}} \varphi_{k,\psi}(\mathbf{x})f(\mathbf{x}; \theta_1) - \sum_{\mathbf{x} \in \mathcal{X}} \tilde{\varphi}(\mathbf{x})f(\mathbf{x}; \theta_1) \\ &\quad - k \left(\sum_{\mathbf{x} \in \mathcal{X}} \varphi_{k,\psi}(\mathbf{x})f(\mathbf{x}; \theta_0) - \sum_{\mathbf{x} \in \mathcal{X}} \tilde{\varphi}(\mathbf{x})f(\mathbf{x}; \theta_0) \right) \\ &= E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] - k[E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_0}[\tilde{\varphi}(\mathbf{X})]] \end{aligned} \quad (5.4.6)$$

Combining this relation with (5.4.5) it follows that

$$E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] - k[E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_0}[\tilde{\varphi}(\mathbf{X})]] = 0,$$

an equality that, *via* (5.4.6), is equivalent to

$$\sum_{\mathbf{x} \in \mathcal{X}} [\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})][f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)] = 0.$$

Since all the terms in this last summation are nonnegative, by (5.4.4), it follows that

$$[\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})][f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)] = 0, \quad \mathbf{x} \in \mathcal{X}.$$

Since $[f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)] \neq 0$ when $\mathbf{x} \in \mathcal{D}$, it follows that

$$\mathbf{x} \in \mathcal{D} \Rightarrow \varphi_{k,\psi}(\mathbf{x}) = \tilde{\varphi}(\mathbf{x}),$$

so that the inclusion $[\mathbf{X} \in \mathcal{D}] \subset [\varphi_{k,\psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})]$ holds. Consequently,

$$[\mathbf{X} \in \mathcal{D}] \cap [\varphi_{k,\psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})] = [\mathbf{X} \in \mathcal{D}],$$

an equality that immediately leads to (5.4.3).

Case 2: The observation vector \mathbf{X} is continuous.

The argument is similar to the one used in the discrete case. Taking the integral over $\mathbf{x} \in \mathcal{X}$ in both sides of (5.4.4) and paralleling the above argument it follows that

$$\begin{aligned} 0 &\leq \int_{\mathbf{x} \in \mathcal{X}} [\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})][f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)] d\mathbf{x} \\ &= E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] - k[E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_0}[\tilde{\varphi}(\mathbf{X})]], \end{aligned}$$

a relation that together with (5.4.5) yields that

$$E_{\theta_1}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_1}[\tilde{\varphi}(\mathbf{X})] - k[E_{\theta_0}[\varphi_{k,\psi}(\mathbf{X})] - E_{\theta_0}[\tilde{\varphi}(\mathbf{X})]] = 0,$$

and then

$$\int_{\mathbf{x} \in \mathcal{X}} [\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})][f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)] d\mathbf{x} = 0.$$

Since the product in this integral is nonnegative, Remark 5.4.1(b) yields that

$$\int_{\mathbf{x} \in \mathcal{A}} d\mathbf{x} = 0,$$

and then

$$P_{\theta_0}[\mathbf{X} \in \mathcal{A}] = \int_{\mathbf{x} \in \mathcal{A}} f(\mathbf{x}; \theta_0) d\mathbf{x} = 0 = \int_{\mathbf{x} \in \mathcal{A}} f(\mathbf{x}; \theta_0) d\mathbf{x} = P_{\theta_1}[\mathbf{X} \in \mathcal{A}], \quad (5.4.7)$$

where

$$\begin{aligned} \mathcal{A} &= \{\mathbf{x} \in \mathcal{X} \mid [\varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x})][f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0)] \neq 0\} \\ &= \{\mathbf{x} \in \mathcal{X} \mid \varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x}) \neq 0\} \cap \{\mathbf{x} \in \mathcal{X} \mid f(\mathbf{x}; \theta_1) - kf(\mathbf{x}; \theta_0) \neq 0\} \\ &= \{\mathbf{x} \in \mathcal{X} \mid \varphi_{k,\psi}(\mathbf{x}) - \tilde{\varphi}(\mathbf{x}) \neq 0\} \cap \mathcal{D}; \end{aligned}$$

see (5.4.2) for the last equality. It follows that the event $[\mathbf{X} \in \mathcal{A}]$ can be expressed as

$$[\mathbf{X} \in \mathcal{A}] = [\varphi_{k,\psi}(\mathbf{X}) \neq \tilde{\varphi}(\mathbf{X})] \cap [\mathbf{X} \in \mathcal{D}]$$

and then (5.4.7) yields that

$$P_{\theta}[[\varphi_{k,\psi}(\mathbf{X}) \neq \tilde{\varphi}(\mathbf{X})] \cap [\mathbf{X} \in \mathcal{D}]] = 0, \quad \theta = \theta_0, \theta_1.$$

Therefore, for $\theta = \theta_0, \theta_1$,

$$\begin{aligned} P_{\theta}[\mathbf{X} \in \mathcal{D}] &= P_{\theta}[[\varphi_{k,\psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})] \cap [\mathbf{X} \in \mathcal{D}]] + P_{\theta}[[\varphi_{k,\psi}(\mathbf{X}) \neq \tilde{\varphi}(\mathbf{X})] \cap [\mathbf{X} \in \mathcal{D}]] \\ &= P_{\theta}[[\varphi_{k,\psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})] \cap [\mathbf{X} \in \mathcal{D}]] \end{aligned}$$

establishing (5.4.3). □

Proof of Theorem 5.4.1. The argument relies heavily on Assumption 2.1.1. Given $\alpha \in (0, 1)$ let $\varphi_{k, \psi}$ be the Neyman-Pearson test of size α . By (5.1.1), the inequality $E_{\theta_0}[\varphi_{k, \psi}(\mathbf{X})] \leq E_{\theta_1}[\varphi_{k, \psi}(\mathbf{X})]$ holds. Thus, to establish the desired conclusion it is sufficient to show that

$$E_{\theta_0}[\varphi_{k, \psi}(\mathbf{X})] = E_{\theta_1}[\varphi_{k, \psi}(\mathbf{X})] \quad (5.4.8)$$

does not occur. Proceeding by contradiction, *suppose* that this equality holds. Now, defining $\tilde{\varphi}(\mathbf{x}) = \alpha$ for every $\mathbf{x} \in \mathcal{X}$, it follows that

$$\alpha = E_{\theta_0}[\tilde{\varphi}(\mathbf{X})] = E_{\theta_1}[\tilde{\varphi}(\mathbf{X})],$$

so that $\tilde{\varphi}$ is also a most powerful test with size α . Then, Theorem 5.4.2 yields that

$$P_{\theta}[[\mathbf{X} \in \mathcal{D}] \cap [\varphi_{k, \psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})]] = P_{\theta}[\mathbf{X} \in \mathcal{D}], \quad \theta = \theta_0, \theta_1,$$

where \mathcal{D} is given in (5.4.2). Notice now that on the event $[\mathbf{X} \in \mathcal{D}]$, one of the inequalities $f(\mathbf{X}; \theta_1) > kf(\mathbf{X}; \theta_0)$ or $f(\mathbf{X}; \theta_1) < kf(\mathbf{X}; \theta_0)$, and in these cases $\varphi_{k, \psi}(\mathbf{X}) = 1$ or $\varphi_{k, \psi}(\mathbf{X}) = 0$, respectively. Since $\varphi(\mathbf{X})$ attains only the value $\alpha \in (0, 1)$, it follows that the equality $\varphi_{k, \psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})$ can not occur. In short,

$$[\mathbf{X} \in \mathcal{D}] \cap [\varphi_{k, \psi}(\mathbf{X}) = \tilde{\varphi}(\mathbf{X})] = \emptyset,$$

and combining this and the previous display it follows that

$$P_{\theta}[\mathbf{X} \in \mathcal{D}] = 0, \quad \theta = \theta_0, \theta_1,$$

that is,

$$P_{\theta}[f(\mathbf{X}; \theta_1) = kf(\mathbf{X}; \theta_0)] = P_{\theta}[\mathbf{X} \in \mathcal{D}^c] = 1, \quad \theta = \theta_0, \theta_1,$$

an equality that leads to

$$P_{\theta}[A] = P_{\theta}[A \cap [f(\mathbf{X}; \theta_1) = kf(\mathbf{X}; \theta_0)]], \quad A \subset \mathcal{X}, \quad \theta = \theta_0, \theta_1, \quad (5.4.9)$$

To continue, suppose that \mathbf{X} is discrete. and notice that, for every $A \subset \mathbb{R}^n$,

$$\begin{aligned} P_{\theta_1}[A] &= P_{\theta_1}[A \cap [f(\mathbf{X}; \theta_1) = kf(\mathbf{X}; \theta_0)]] \\ &= \sum_{\mathbf{y} \in A \cap \{\mathbf{x} | f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\}} f(\mathbf{y}; \theta_1) \\ &= \sum_{\mathbf{y} \in A \cap \{\mathbf{x} | f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\}} kf(\mathbf{y}; \theta_0) \\ &= k \sum_{\mathbf{y} \in A \cap \{\mathbf{x} | f(\mathbf{x}; \theta_1) = kf(\mathbf{x}; \theta_0)\}} f(\mathbf{y}; \theta_0) \\ &= kP_{\theta_0}[A \cap [f(\mathbf{X}; \theta_1) = kf(\mathbf{X}; \theta_0)]] \\ &= kP_{\theta_0}[A] \end{aligned}$$

Since $1 = P_{\theta_0}[\mathcal{X}] = P_{\theta_1}[\mathcal{X}]$, setting $A = \mathcal{X}$ in the above display it follows that

$$1 = P_{\theta_1}[\mathcal{X}] = kP_{\theta_0}[\mathcal{X}] = k,$$

and then

$$P_{\theta_1}[A] = P_{\theta_0}[A] \text{ for every } A,$$

that is, the distributions P_{θ_1} and P_{θ_0} coincide, contradicting Assumption 2.1.1. Therefore, (5.4.8) does not occur, a fact that, as already mentioned, establishes (5.4.1). When \mathbf{X} is continuous, a similar argument—replacing the summation by integrals—allows to obtain the desired conclusion. \square

The identifiability condition in Assumption 2.1.1 is practically satisfied in all models considered in the literature (with exceptions in the theory of experimental designs). For instance, in Example 4.4.1 the problem of testing $\mathcal{H}_0: \mu = \mu_0$ versus $\mathcal{H}_1: \mu = \mu_1$ based on a sample $\mathbf{X} = (X_1, X_2, \dots, X_n)$ of the $\mathcal{N}(\mu, \sigma_0^2)$ distribution was considered, and it was observed that the most powerful test satisfies that $\pi(\mu_0) < \pi(\mu_1)$; the reason behind the strict inequality is that the model is identifiable.

References

- [1]. T. M. Apostol (1980), *Mathematical Analysis*, Addison Wesley, Reading, Massachusetts.
- [2]. A. A. Borovkov (1999), *Mathematical Statistics*, Gordon and Breach, New York
- [3]. E. Dudewicz y S. Mishra (1998). *Mathematical Statistics*, Wiley, New York.
- [4]. W. Fulks (1980), *Cálculo Avanzado*, Limusa, México, D. F.
- [5]. F. A. Graybill (2000), *Theory and Application of the Linear Model*, Duxbury, New York.
- [6]. F. A. Graybill (2001), *Matrices with Applications in Statistics* Duxbury, New York.
- [7]. D. A. Harville (2008), *Matrix Algebra Form a Statistician's Perspective*, Springer-Verlaf, New York.
- [8]. A. I. Khuri (2002), *Advanced Calculus with Applications in Statistics*, Wiley, New York.
- [9]. E. L. Lehmann and G. B. Casella, (1998), *Theory of Point Estimation*, Springer, New York.
- [10]. E. L. Lehmann and J. P. Romano, (1999), *Testing Statistical Hypothesis*, Springer, New York.
- [11]. M. Loève (1984), *Probability Theory, I*, Springer-Verlag, New York.
- [12]. D. C. Montgomery (2011), *Introduction to Statistical Quality Control*, 6th Edition, Wiley, New York.
- [13]. A. M. Mood, D. C. Boes and F. A. Graybill (1984), *Introduction to the Theory of Statistics*, McGraw-Hill, New York.
- [14]. W. Rudin (1984), *Real and Complex Analysis*, McGraw-Hill, New York.
- [15]. H. L. Royden (2003), *Real Analysis*, MacMillan, London.
- [16]. J. Shao (2010), *Mathematical Statistics*, Springer, New York.
- [17]. D. Wackerly, W. Mendenhall y R. L. Scheaffer (2009), *Mathematical Statistics with Applications*, Prentice-Hall, New York.