

Type I and Type II Errors and the Neyman-Pearson: Lecture XXIII

Charles B. Moss

October 26, 2010

- 1 Introduction
- 2 Type I and Type II Errors
- 3 Neyman-Pearson Lemma

Introduction

- In general there are two kinds of hypotheses: one type concerns the form of the probability distribution (i.e. is the random variable normally distributed) and the second concerns parameters of a distribution function (i.e. what is the mean of a distribution).
- The second kind of distribution is the traditional stuff of econometrics. We may be interested in testing whether the effect of income on consumption is greater than one, or whether the effect of price on the level consumed is equal to zero.
 - ▶ The second kind of hypothesis is termed a simple hypothesis. Under this scenario, we test the value of a parameter against a single alternative.
 - ▶ The first kind of hypothesis (whether the effect of income on consumption is greater than one) is termed a composite hypothesis. Implicit in this test is several alternative values.

- Hypothesis testing involves the comparison between two competing hypothesis, or conjectures.
 - ▶ The null hypothesis, denoted H_0 , is sometimes referred to as the maintained hypothesis.
 - ▶ The competing hypothesis to be accepted if the null hypothesis is rejected is called the alternative hypothesis.

- The general notion of the hypothesis test is that we collect a sample of data $X_1, \dots, X_n$. This sample is a multivariate random variable, E_n (the text refers to this as an element of a Euclidean space).
 - ▶ If the multivariate random variable is contained in space R , we reject the null hypothesis.
 - ▶ Alternatively, if the random variable is in the complement of the space R , we fail to reject the null hypothesis.
 - ▶ Mathematically,

$$\text{if } X \in R \text{ then reject } H_0 \quad (1)$$

$$\text{if } X \notin R \text{ then fail to reject } H_0$$

The set R is called the region of rejection or the critical region of the test.

- Continued

- ▶ In order to determine whether the sample is in a critical region, we construct a test statistics $T(X)$. Note that like any other statistic, $T(X)$ is a random variable. The hypothesis test given this statistic can then be written as

$$T(X) \in R \Rightarrow \text{reject } H_0 \quad (2)$$

$$T(X) \in \bar{R} \Rightarrow \text{fail to reject } H_0$$

- ▶ A statistic used to test hypotheses is called a test statistic.

- **Definition 9.1.1.** A hypothesis is called simple if it specifies the values of all the parameters of a probability distribution. Otherwise, it is called composite.'

Type I and Type II Errors

- **Definition 9.2.1.** A Type I error is the error of rejecting H_0 when it is true. A Type II error is the error of accepting H_0 when it is false (that is when H_1 is true).
- We denote the probability of Type I error of α and the probability of Type II error as β . Mathematically,

$$\begin{aligned}\alpha &= P[X \in R | H_0] \\ \beta &= P[X \in \bar{R} | h_1]\end{aligned}\tag{3}$$

- ▶ The probability of Type I error is also called the size of a test.

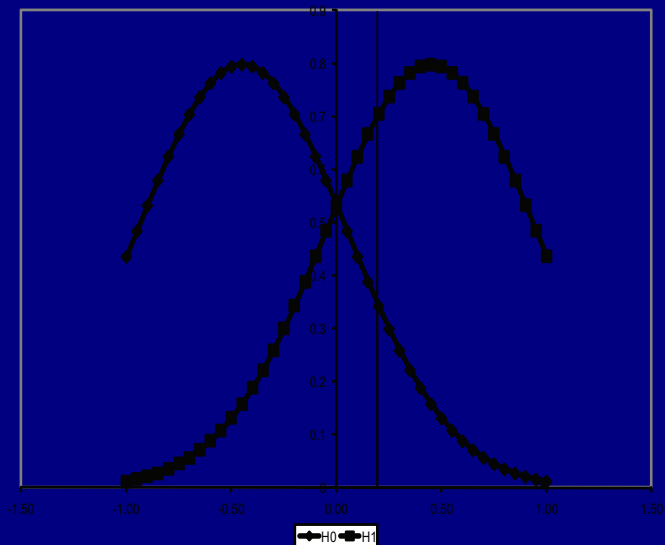


Figure: Type I and Type II Error

- Continued

- ▶ Assume that we want to compare two critical regions R_1 and R_2 . Assume that we choose either confidence region R_1 or R_2 randomly with probabilities δ and $1 - \delta$, respectively. This is called a randomized test.
- ▶ If the probabilities of the two types of error for R_1 and R_2 are (α_1, β_1) and (α_2, β_2) respectively. The probability of each type of error becomes

$$\begin{aligned}\alpha &= \delta \alpha_1 + (1 - \delta) \alpha_2 \\ \beta &= \delta \beta_1 + (1 - \delta) \beta_2\end{aligned}\tag{4}$$

The values (α, β) are called the characteristics of the test.

- Continued

- ▶ **Definition 9.2.2.** Let (α_1, β_1) and (α_2, β_2) be the characteristics of two tests. The first test is better (or more powerful) than the second test if $\alpha_1 \leq \alpha_2$, and $\beta_1 \leq \beta_2$ with a strict inequality holding for at least one point.
- ▶ If we cannot determine that one test is better by the definition, we could consider the relative cost of each type of error. Classical statisticians typically do not consider the relative cost of the two errors because of the subjective nature of this comparison.
- ▶ Bayesian statisticians compare the relative cost of the two errors using a loss function.

- **Definition 9.2.3.** A test is inadmissible if there exists another test which is better in the sense of Definition 9.2.2. Otherwise it is called admissible.
- **Definition 9.2.4.** R is the most powerful test of size α if $\alpha(R) = \alpha$ and for any test R_1 of size α , $\beta(R) \leq \beta(R_1)$.
- **Definition 9.2.5.** R is the most powerful test of level α and for any test R_1 of level α (that is, such that $\alpha(R_1) \leq \alpha$), $\beta(R) \leq \beta(R_1)$.
 - ▶ **Example 9.2.2.** Let X have the density

$$\begin{aligned}
 f(x) &= 1 - \theta + x \text{ for } \theta - 1 \leq x \leq \theta \\
 &= 1 + \theta - x \text{ for } \theta \leq x \leq \theta + 1
 \end{aligned}
 \tag{5}$$

This funny looking beast is a triangular probability density function as depicted in Figure 2. Assume that we want to test $H_0 : \theta = 0$ against $H_1 : \theta = 1$ on the basis of a single observation of X .

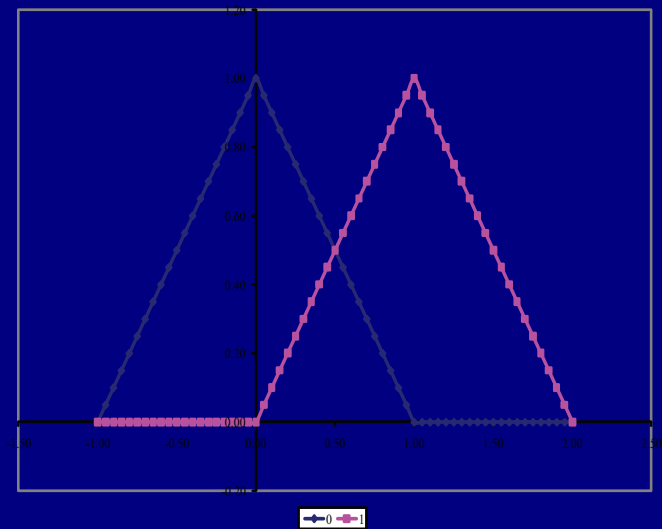


Figure: Hypothesis Test for Triangular Distribution

- Continued

- ▶ Type I and Type II errors are then defined by the choice of t , the cut off region

$$\alpha = \frac{1}{2} (1 - t)^2 \quad (6)$$

$$\beta = \frac{1}{2} t^2$$

Deriving β in terms of α yields

$$\beta = \frac{1}{2} \left(1 - \sqrt{2\alpha} \right)^2 \quad (7)$$

- Continued

- ▶ Note that the choice of any t yields an admissible test. However, any randomized test is inadmissible.
- ▶ **Theorem 9.2.1.** The set of admissible characteristics plotted on the α , β plane is a continuous, monotonically decreasing, convex function which starts at a point with $[0, 1]$ on the β axis and ends at a point within the $[0, 1]$ on the α axis.

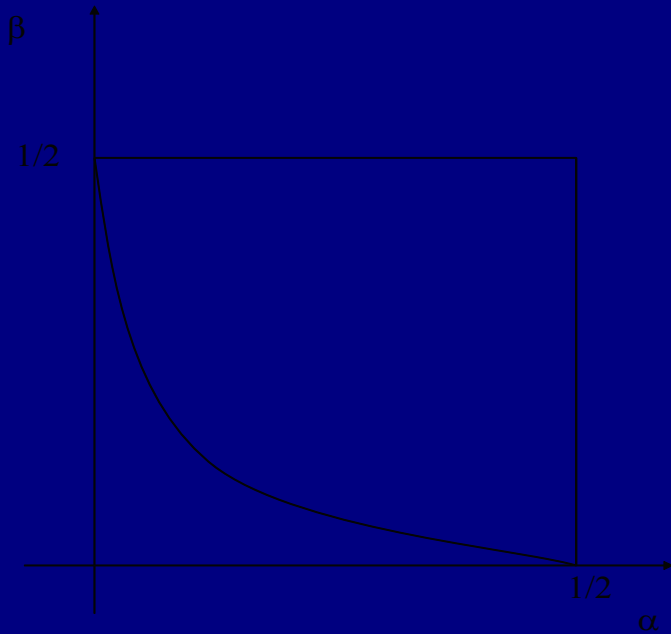


Figure: Tradeoff of the Power of the Test

Neyman-Pearson Lemma

- How does the Bayesian statistician choose between test?
 - ▶ The Bayesian chooses between the test H_0 and H_1 based on the posterior probability of the hypotheses: $P(H_0|X)$ and $P(H_1|X)$.
 - ▶ Table 1 presents the loss matrix for hypothesis testing.

Table: Loss Matrix in Hypothesis Testing

	State of Nature	
	H_0	H_1
Decision H_0	0	γ_1
Decision H_1	γ_2	0

- Continued

- ▶ The Bayesian decision is then based on this loss function

$$\text{Reject } H_0 \text{ if } \gamma_1 P(H_0|X) < \gamma_2 P(H_1|X) \quad (8)$$

The critical region for the test then becomes

$$R_0 = \{ x | \gamma_1 P(H_0|x) < \gamma_2 P(H_1|x) \} \quad (9)$$

- Continued

- ▶ Alternatively, the Bayesian problem can be formulated as that of determining the critical region R in the domain X so as to

$$\begin{aligned}\min \phi(R) = & \gamma_1 P(H_0|X \in R) P(X \in R) + \\ & \gamma_2 P(H_1|X \in \bar{R}) P(X \in \bar{R})\end{aligned}\tag{10}$$

- ▶ We can write this expression as

$$\begin{aligned}\phi(R) = & \gamma_1 P(H_0) P(R|H_0) + \gamma_2 P(H_1) P(R|H_1) \\ = & \eta_0 \alpha(R) + \eta_1 \beta(R) \\ \eta_0 = & \gamma_1 P(H_0) \\ \eta_1 = & \gamma_2 P(H_1)\end{aligned}\tag{11}$$

- Continued

- ▶ Choosing between admissible test statistics in the (α, β) plane then becomes like the choice of a utility maximizing consumption point in utility theory. Specifically, the relative tradeoff between the two characteristics becomes $-\eta_0/\eta_1$.
- ▶ This fact is the basis of the Neyman-Pearson Lemma. Let $L(X)$ be the joint density function of X .

$$\begin{aligned}\gamma_1 P(H_0|x) &< \gamma_2 P(H_1|x) \\ \gamma_1 P(H_0|x) L(x) &< \gamma_2 P(H_1|x) L(x)\end{aligned}\tag{12}$$

- Continued

- ▶ The Bayesian optimal test R_0 can then be written as

$$R_0 = \left\{ x \mid \frac{L(x|H_1)}{L(x|H_0)} > \frac{\eta_0}{\eta_1} \right\} \quad (13)$$

- **Theorem 9.3.1.** (Neyman-Pearson Lemma) If testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, the best critical region is given by

$$R = \left\{ x \mid \frac{L(x|\theta_1)}{L(x|\theta_0)} > c \right\} \quad (14)$$

where L is the likelihood function and c (the critical value) is determined to satisfy

$$P(R|\theta_0) = \alpha \quad (15)$$

provided c exists

- **Theorem 9.3.2** The Bayes test is admissible.