# 7: THE NEYMAN-PEARSON LEMMA

Suppose we are testing a simple null hypothesis $H_0 : \theta = \theta'$ against a simple alternative $H_1 : \theta = \theta''$, where $\theta$ is the parameter of interest, and $\theta'$, $\theta''$ are particular values of $\theta$. We are given a random sample $(X_1, \ldots, X_n)$ which are *iid*, each with the p.d.f. $f(x;\theta)$.

A p.d.f. for a random variable $X$, as defined by Hogg and Craig, p. 39, is either the probability density function (if $X$ is a continuous random variable) or the probability mass function $f(x) = Pr(X = x)$ (if $X$ is a discrete random variable). This definition is not the standard one, however, as the term p.d.f.

is usually reserved for the density of a continuous random variable. Also note that Hogg and Craig are assuming that $X$ is either discrete or continuous, even though there are other possibilities.

We are going to reject $H_0$ if $(X_1, . . . , X_n) \in C$, where $C$ is a region of the $n$-dimensional sample space called the **critical region.** This specifies a test. We say that the critical region $C$ has **size** $\alpha$ if the probability of a Type I error is $\alpha$:

$$Pr[(X_1, . . . , X_n) \in C ; H_0] = \alpha \ .$$

We call $C$ a **best critical region** of size $\alpha$ if it has size $\alpha$, and

$$Pr[(X_1, . . . , X_n) \in C ; H_1] \geq Pr[(X_1, . . . , X_n) \in A ; H_1]$$

for every subset $A$ of the sample space for which $Pr[(X_1, \ldots, X_n) \in A; H_0] = \alpha$. Thus, the power of the test associated with the best critical region $C$ is at least as great as the power of the test associated with any other critical region $A$ of size $\alpha$.

- The Neyman-Pearson Lemma provides us with a way of finding a best critical region.

The joint p.d.f. of $X_1, \ldots, X_n$, evaluated at the observed values $x_1, \ldots, x_n$ is called the **likelihood function,**

$$L(\theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) .$$

We often think of $L(\theta)$ as a function of $\theta$ alone, although it clearly depends on the data as well.

Define the **likelihood ratio** as $L(\theta')/L(\theta'')$. Informally, we can think of this as measuring the plausibility of $H_0$ relative to $H_1$. Therefore, if the likelihood ratio is sufficiently small, we might be inclined to reject $H_0$. Example 1, p. 396 of Hogg and Craig shows that for a binomial random variable with $n = 5$, the best critical region for testing a simple null versus a simple alternative involving the probability $\theta$ of success is the one for which $L(\theta')/L(\theta'') \leq k$, where $k$ is some constant chosen to ensure that the test has level $\alpha$. The Neyman-Pearson Lemma asserts that, in general a best critical region can be found by finding the $n$-dimensional points in the sample space for which the likelihood ratio is smaller than some constant.

**The Neyman-Pearson Lemma:** If $k > 0$ and $C$ is a subset of the sample space such that

$$L(\theta')/L(\theta'') \leq k \quad \text{for } all \ (x_1, \ldots, x_n) \in C \qquad \text{(a)}$$

$$L(\theta')/L(\theta'') \geq k \quad \text{for } all \ (x_1, \ldots, x_n) \in C^* \qquad \text{(b)}$$

$$\alpha = Pr[(X_1, X_2, \ldots, X_n) \in C; H_0] \qquad \text{(c)}$$

where $C^*$ is the complement of $C$, then $C$ is a best critical region of size $\alpha$ for testing the simple hypothesis $H_0 : \theta = \theta'$ against the alternative simple hypothesis $H_1 : \theta = \theta''$.

**Proof:** Suppose for simplicity that the random variables $X_1, \ldots, X_n$ are continuous. (If they were discrete, the proof would be the same, except that integrals would be replaced by sums). Let $X = (X_1, \ldots, X_n)$. For any region $R$ of $n$-dimensional space, we will denote the probability that $X \in R$ by $\int_R L(\theta)$, where theta is the true value of the parameter. The full notation, omitted to save space, would be

$$Pr[X \in R; \theta] = \int \cdots_R \int L(\theta; x_1, \ldots, x_n)\, dx_1 \cdots dx_n \quad .$$

We need to prove that if $A$ is another critical region of size $\alpha$, then the power of the test associated with $C$ is at least as great as the power of the test associated with $A$, or in the present notation, that

$$\int_A L(\theta'') \le \int_C L(\theta'') \ . \tag{1}$$

Suppose $X \in A^* \cap C$. Then $X \in C$, so by (a),

$$\int_{A^* \cap C} L(\theta'') \ge \frac{1}{k} \int_{A^* \cap C} L(\theta') \ . \tag{2}$$

Next, suppose $X \in A \cap C^*$. Then $X \in C^*$, so by (b),

$$\int_{A \cap C^*} L(\theta'') \le \frac{1}{k} \int_{A \cap C^*} L(\theta') \ . \tag{3}$$

We now establish (1), thereby completing the proof.

$$\int_A L(\theta'') = \left[ \int_{A \cap C} L(\theta'') \right] + \int_{A \cap C^*} L(\theta'')$$

$$= \left[ \int_C L(\theta'') - \int_{A^* \cap C} L(\theta'') \right] + \int_{A \cap C^*} L(\theta'')$$

$$\leq \int_C L(\theta'') - \frac{1}{k} \int_{A^* \cap C} L(\theta') + \frac{1}{k} \int_{A \cap C^*} L(\theta') \quad (\textit{See}\,(2),(3))$$

$$\left[ -\frac{1}{k} \int_{A \cap C} L(\theta') + \frac{1}{k} \int_{A \cap C} L(\theta') \right] \qquad (\textit{Add Zero})$$

$$= \int_C L(\theta'') - \frac{1}{k} \int_C L(\theta') + \frac{1}{k} \int_A L(\theta') \qquad (\textit{Collect Terms})$$

$$= \int_C L(\theta'') - \frac{\alpha}{k} + \frac{\alpha}{k}$$

(Since both $C$ and $A$ have size $\alpha$)

$$= \int_C L(\theta'') \ .$$

**Eg:** Suppose $X_1, \ldots, X_n$ are *iid* $N(\theta, 1)$, and we want to test $H_0: \theta = \theta'$ versus $H_1: \theta = \theta''$, where $\theta'' > \theta'$. According to the $z$-test, we should reject $H_0$ if $Z = \sqrt{n}\,(\bar{X} - \theta')$ is large, or equivalently if $\bar{X}$ is large. We can now use the Neyman-Pearson Lemma to show that the $z$-test is best. The likelihood function is

$$L(\theta) = (2\pi)^{-n/2} \exp\{-\sum_{i=1}^{n} (x_i - \theta)^2/2\} \;.$$

According to the Neyman-Pearson Lemma, a best critical region is given by the set of $(x_1, \ldots, x_n)$ such that $L(\theta')/L(\theta'') \le k_1$, or equivalently, such that $\frac{1}{n} \log[L(\theta'')/L(\theta')] \ge k_2$. But

$$\frac{1}{n}\log[L(\theta'')/L(\theta')] = \frac{1}{n}\sum_{i=1}^{n}[(x_i-\theta')^2/2-(x_i-\theta'')^2/2]$$

$$= \frac{1}{2n}\sum_{i=1}^{n}[(x_i^2-2\theta' x_i+\theta'^2)-(x_i^2-2\theta'' x_i+\theta''^2)]$$

$$= \frac{1}{2n}\sum_{i=1}^{n}[2(\theta''-\theta')x_i+\theta'^2-\theta''^2]$$

$$= (\theta''-\theta')\bar{x}+\frac{1}{2}[\theta'^2-\theta''^2] \ .$$

So the best test rejects $H_0$ when $\bar{x} \geq k$, where $k$ is a constant. But this is exactly the form of the rejection region for the $z$-test. Therefore, the $z$-test is best.