Comparing Variations of the Neyman-Pearson Lemma

William Jake Johnson

Department of Mathematical Sciences
Montana State University

April 29, 2016

A writing project submitted in partial fulfillment
of the requirements for the degree

Master of Science in Statistics

APPROVAL

of a writing project submitted by

William Jake Johnson

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date	Steve Cherry Writing Project Advisor
Date	Mark C. Greenwood Writing Projects Coordinator

Abstract:
Hypothesis testing is a key component in statistical analyses and many of the processes that are utilized in hypothesis testing can be summarized in various theorems and lemmas. Of these results, few are as prominent as the Neyman-Pearson Lemma. This paper hosts a collection of various forms of the Neyman-Pearson Lemma from ten different Mathematical Statistics text books and provides comparisons between the lemma statements and the provided proofs. Additionally a couple of examples regarding the lemma are provided and, given the fame the lemma has acquired, a brief look into the lemma nomenclature is explored.

Introduction

For any test, there are four possibilities between what the test concludes and what the truth is regarding the associated hypotheses. On the positive spectrum, the test can reject the null hypothesis when the null is in fact false or the test can fail to reject the null hypothesis when the null is actually true. On the other hand, the test can make an incorrect decision. It can reject the null hypothesis when the null is actually true or it can fail to reject the null hypothesis when, in fact, the null is false. These two errors are given the respective names of Type I error and Type II error. Table 1 summarizes these results.

		Decision	
		Reject H_0	Fail to Reject H_0
Truth	H_0	Type I Error	Correct Decision
	H_1	Correct Decision	Type II Error

Table 1: Errors in Hypothesis Testing

If R is the rejection region for a particular test, then under the null hypothesis, Type I error is defined as $P_\theta(X \in R | H_0 \text{ is true})$. On a similar note, Type II error is defined as $P_\theta(X \in R^c | H_0 \text{ is false})$ where R^c is the compliment of R. Often, mathematical statistics books will denote the probability of a Type II error as $1 - P_\theta(X \in R | H_0 \text{ is false})$ where this probability is defined as the probability of making a correct decision given H_0 is false with respect to a parameter θ. At this point, it seems clear that both Type I and Type II error can be written in terms of a common probability where the difference is only attributed to the truth of the null hypothesis. So why does this matter? The probability that X is in the rejection region R is denoted as the power of a test. It is common to denote the power function as $\beta(\theta) = P_\theta(X \in R)$ although there are other widely used forms. It is through this power function that we can asses the usefulness of a test. Ideally, we would like to create a test that has an arbitrarily small probability of making a Type I or Type II error. Unfortunately, it is impossible to make both of these probabilities very small. Because of this, it is common to restrict the probability of making a Type I error by specifying a particular probability of making such an error, typically denoted as α. From the set of all tests that make a Type I error with probability α (where $\alpha \in [0,1]$), the test that has the smallest probability of making a Type II error is typically used. A test that makes a Type I error with probability α is defined to be a *size α test.* That is, under the null hypothesis, the largest value the power function takes on with respect to θ is α. Related to this idea is a *level α test* which satisfies the condition the largest value the power function takes on with respect to θ under the null hypothesis is at most α. Mathematically, these two concepts are expressed as follows:

$$\text{Size } \alpha \text{ Test: } \sup_{\theta \in \Theta_0} \beta(\theta) = \alpha \qquad\qquad \text{Level } \alpha \text{ Test: } \sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

where Θ_0 is the set of all possible values for θ under the null hypothesis. Regarding the use of this restriction on a Type I error, in practice the most common used values are $\alpha = 0.1$, 0.05, and 0.01.
The above use of size α and level α tests are important to the statement of the Neyman-Pearson Lemma. However, the Neyman-Pearson Lemma allows us to determine a test that is *Uniformly Most Powerful.* A definition for this is provided below by Casella and Berger (2002).

> Let C be a class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1^c$. A test in class C, with power function $\beta(\theta)$, is a *uniformly most powerful (UMP) class C test* if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class C.

Essentially, the above definition can be interpreted as follows: from the set of all tests that are defined to be of size α, then the test whose power function is as large as or larger than any power function of any other test for all values of θ under the alternative hypothesis is said to be uniformly most powerful. At this point, we have everything needed to proceed with the comparisons of the Neyman-Pearson Lemma variants. However, a few notes about the notation used in these variations will first be discussed.

Due to the nature of obtaining a collection of lemmas from various texts, it is worth noting that each of the variations of the Neyman-Pearson Lemma utilize various forms in mathematical notation. Most notation is quite similar among all the lemmas and their proofs but there are a few differences. Regarding the hypotheses statements, the majority follow the form H_0 for the null hypothesis and H_1 for the alternative hypothesis. Other variations which occur are H and A, H_0 and H_A, and H and K where each pair of letters correspond to the null hypothesis and the alternative hypothesis respectively. Critical regions are most commonly labeled as C but some lemma statements and proofs make use of C^*, A, and R. Some variations in the lemma statements and proofs will make use of explicit notation to denote the power functions and test functions that are utilized while others do not provide a name. For those variations that include such definitions, letters utilized for power are β, Q, and π and letters used for test functions are ϕ, ψ, and δ (sometimes written as $\phi(x)$, $\psi(x)$, and $\delta(x)$).

In the following section, Each lemma variant will be provided with the name given by the author(s) from the text it appeared and the proof will follow immediately after. Comments will be made in between the lemmas such that most of the general comments about a particular lemma variant will follow in the paragraphs immediately following the proof.

Lemma Statements & Proofs

> **Neyman-Pearson Fundamental Lemma** (Roussas, 1997)
> Let $X_1, \ldots, X_n$ be iid random variables with pdf $f(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Omega} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$. We are interested in testing the hypothesis $H : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative $A : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ at level $\alpha (0 \leq \alpha \leq 1)$. Let ϕ be the test defined as follows:

$$\phi(x_1,\ldots,x_n)=\begin{cases}1, & \text{if } f(x_1;\boldsymbol{\theta}_1)\cdots f(x_n;\boldsymbol{\theta}_1)>C\,f(x_1;\boldsymbol{\theta}_0)\cdots f(x_n;\boldsymbol{\theta}_0)\\ \gamma, & \text{if } f(x_1;\boldsymbol{\theta}_1)\cdots f(x_n;\boldsymbol{\theta}_1)=C\,f(x_1;\boldsymbol{\theta}_0)\cdots f(x_n;\boldsymbol{\theta}_0)\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

where the constants $\gamma(0\le\gamma\le 1)$ and $C(>0)$ are determined so that

$$E_{\boldsymbol{\theta}_0}\phi(X_1,\ldots,X_n)=\alpha. \tag{2}$$

Then, for testing H against A at level α, the test defined by (1) and (2) is MP within the class of all tests whose level is $\le\alpha$.

The proof is presented for the case that the X's are of the continuous type, since the discrete case is dealt with similarly by replacing integrals by summation signs.

Proof For convenient writing, we set

$$\boldsymbol{z}=(x_1,\ldots,x_n)',\, d\,\boldsymbol{z}=d\,x_1\cdots d\,x_n,\, \boldsymbol{Z}=(X_1,\ldots,X_n)'$$

and $f(\boldsymbol{z};\boldsymbol{\theta}), f(\boldsymbol{Z};\boldsymbol{\theta})$ for $f(x_1;\boldsymbol{\theta})\cdots f(x_n;\boldsymbol{\theta}), f(X_1;\boldsymbol{\theta})\cdots f(X_n;\boldsymbol{\theta})$ respectively. Next, let T be the set of points $\boldsymbol{z}\in\mathbb{R}^n$ such that $f_0(\boldsymbol{z})>0$ and let $D^c=\boldsymbol{Z}^{-1}(T^c)$. Then

$$P_{\boldsymbol{\theta}_0}(D^c)=P_{\boldsymbol{\theta}_0}(\boldsymbol{Z}\in T^c)=\int_{T^c} f_0(\boldsymbol{z})\,d\,\boldsymbol{z}=0,$$

and therefore in calculating $P_{\boldsymbol{\theta}_0}-$ probabilities we may redefine and modify random variables on the set D^c. Thus, we have, in particular,

$$\begin{aligned}E_{\boldsymbol{\theta}_0}\phi(\boldsymbol{Z}) &= P_{\boldsymbol{\theta}_0}\left[f_1(\boldsymbol{Z})>C\,f_0(\boldsymbol{Z})\right]+\gamma\,P_{\boldsymbol{\theta}_0}\left[f_1(\boldsymbol{Z})=C\,f_0(\boldsymbol{Z})\right]\\ &= P_{\boldsymbol{\theta}_0}\left\{\left[f_1(\boldsymbol{Z})>c\,F_0(\boldsymbol{Z})\right]\cap D\right\}+\gamma\,P_{\boldsymbol{\theta}_0}\left\{\left[f_1(\boldsymbol{Z})=C\,f_0(\boldsymbol{Z})\right]\cap D\right\}\\ &= P_{\boldsymbol{\theta}_0}\left\{\left[\frac{f_1(\boldsymbol{Z})}{f_0(\boldsymbol{Z})}>C\right]\cap D\right\}+\gamma\,P_{\boldsymbol{\theta}_0}\left\{\left[\frac{f_1(\boldsymbol{Z})}{f_0(\boldsymbol{Z})}=C\right]\cap D\right\}\\ &= P_{\boldsymbol{\theta}_0}\left[(Y>C)\cap D\right]+\gamma\,P_{\boldsymbol{\theta}_0}\left[(Y=C)\cap D\right]\\ &= P_{\boldsymbol{\theta}_0}(Y>C)+\gamma\,P_{\boldsymbol{\theta}_0}(Y=C)\end{aligned} \tag{3}$$

Where $Y=f_1(\boldsymbol{Z})/f_0(\boldsymbol{Z})$ on D and let Y be arbitrary (but measurable) on D^c. Now let $a(C)=P_{\boldsymbol{\theta}_0}(Y>C)$, so that $G(C)=1-a(C)=P_{\boldsymbol{\theta}_0}(Y\le C)$ is the distribution function of the random variable Y. Since G is a distribution function, we have $G(-\infty)=0$, $G(\infty)=1$, G is nondecreasing and continuous from the right. These properties of G imply that the function a is such that $a(-\infty)=1$,, $a(\infty)=0$,, a is nonincreasing and continuous from the right. Furthermore,

$$P_{\boldsymbol{\theta}_0}(Y=C)=G(C)-G(C-)=[1-a(C)]-[1-a(C-)]=a(C-)-a(C).$$

and $a(C)=1$ for $C<0$, since $P_{\boldsymbol{\theta}_0}(Y\geq 0)=1$.

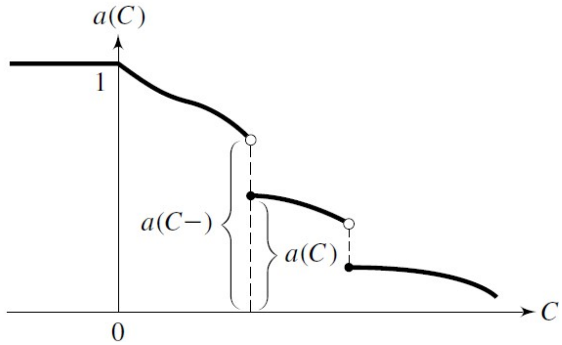

Figure 1

Figure 1 represents the graph of a typical function a. Now for any $\alpha\,(0<\alpha<1)$ there exists $C_0\,(\geq 0)$ such that $a(C_0)\leq\alpha\leq a(C_0-)$ (See Fig. 1). At this point, there are two cases to consider. First, $a(C_0)=a(C_0-)$; that is, C_0 is a continuity point of the function a. Then $\alpha=a(C_0)$ and if in (1) C is replaced by C_0 and $\gamma=0$, the resulting test is of level α. In fact, in this case (3) becomes

$$E_{\boldsymbol{\theta}_0}\phi(\boldsymbol{Z})=P_{\boldsymbol{\theta}_0}(Y>C_0)=a(C_0)=\alpha,$$

as was to be seen.

Next, we assume that C_0 is a discontinuity point of a.. In this case, take again $C=C_0$ in (1) and also set

$$\gamma=\frac{\alpha-a(C_0)}{a(C_0-)-a(C_0)}$$

$(\text{so that } 0\leq\gamma\leq 1)$. Again we assert that the resulting test is of level α. In the present case, (3) becomes as follows:

$$\begin{aligned} E_{\boldsymbol{\theta}_0}\phi(\boldsymbol{Z}) &= P_{\boldsymbol{\theta}_0}(Y>C_0)+\gamma P_{\boldsymbol{\theta}_0}(Y=C_0) \\ &= a(C_0)+\frac{\alpha-a(C_0)}{a(C_0-)-a(C_0)}[a(C_0-)-a(C_0)]=\alpha. \end{aligned}$$

Summarizing what we have done so far, we have that with $C=C_0$, as defined above, and

$$\gamma=\frac{\alpha-a(C_0)}{a(C_0-)-a(C_0)}$$

(which it is to be interpreted as 0 whenever it is of the form 0/0), the test defined by (1) is of level α .That is, (2) is satisfied.

Now it remains for us to show that the test so defined is MP, as described in the theorem. To see this, let ϕ^* be any test of level $\leq\alpha$ and set

$$B^{+}=\{z\in\mathbb{R}^n;\phi(z)-\phi^{*}(z)>0\}=(\phi-\phi^{*}>0),$$
$$B^{-}=\{z\in\mathbb{R}^n;\phi(z)-\phi^{*}(z)<0\}=(\phi-\phi^{*}<0).$$

Then $B^{+}\cap B^{-}=\emptyset$ and, clearly,

$$\left.\begin{aligned} B^{+}&=(\phi>\phi^{*})\subseteq(\phi=1)\cup(\phi=\gamma)=(f_1\geq C f_0)\\ B^{-}&=(\phi<\phi^{*})\subseteq(\phi=1)\cup(\phi=\gamma)=(f_1\leq C f_0)\end{aligned}\right\} \tag{4}$$

Therefore

$$\begin{aligned}&\int_{\mathbb{R}^n}[\phi(z)-\phi^{*}(z)][f_1(z)-C f_0(z)]\,dz\\ &=\int_{B^{+}}[\phi(z)-\phi^{*}(z)][f_1(z)-C f_0(z)]\,dz\\ &\quad+\int_{B^{-}}[\phi(z)-\phi^{*}(z)][f_1(z)-C f_0(z)]\,dz\end{aligned}$$

and this is ≥ 0 on account of (4). That is

$$\int_{\mathbb{R}^n}[\phi(z)-\phi^{*}(z)][f_1(z)-C f_0(z)]\,dz\geq 0$$

which is equivalent to

$$\int_{\mathbb{R}^n}[\phi(z)-\phi^{*}(z)]f_1(z)\,dz\geq C\int_{\mathbb{R}^n}[\phi(z)-\phi^{*}(z)]f_0(z)\,dz \tag{5}$$

But

$$\begin{aligned}\int_{\mathbb{R}^n}[\phi(z)-\phi^{*}(z)]f_0(z)\,dz &=\int_{\mathbb{R}^n}\phi(z)f_0(z)\,dz-\int_{\mathbb{R}^n}\phi^{*}(z)f_0(z)\,dz\\ &=E_{\boldsymbol{\theta}_0}\phi(\boldsymbol{Z})-E_{\boldsymbol{\theta}_0}\phi^{*}(\boldsymbol{Z})=\alpha-E_{\boldsymbol{\theta}_0}\phi^{*}(\boldsymbol{Z})\geq 0,\end{aligned} \tag{6}$$

And similarly,

$$\int_{\mathbb{R}^n}[\phi(z)-\phi^{*}(z)]f_1(z)\,dz=E_{\boldsymbol{\theta}_1}\phi(\boldsymbol{Z})-E_{\boldsymbol{\theta}_1}\phi^{*}(\boldsymbol{Z})=\beta_{\phi}(\boldsymbol{\theta}_1)-\beta_{\phi^{*}}(\boldsymbol{\theta}_1). \tag{7}$$

Relations (5), (6), and (7) yield $\beta_{\phi}(\boldsymbol{\theta}_1)-\beta_{\phi^{*}}(\boldsymbol{\theta}_1)\geq 0$, or $\beta_{\phi}(\boldsymbol{\theta}_1)\geq\beta_{\phi^{*}}(\boldsymbol{\theta}_1)$. This completes the proof of the theorem. ■

At this point, with only one lemma provided, not much can be said in terms of a comparison. However, some notes can be made. Regarding Figure 1, the author's use of the term “typical function” seems misleading. As stated, $a(C)$ is the probability that $Y=f_1(\boldsymbol{Z})/f_0(\boldsymbol{Z})$ is greater than some constant C with respect to $\boldsymbol{\theta}_0$ which implies $G(C)=1-a(C)$ is a distribution function of Y. Because this proof assumes continuous random variables, these jumps exhibited in the figure do not occur in a “typical” continuous random variable distribution function. These jumps are indicative that the probability density function for these random variables have point masses at the points where the jumps occur. Despite this, the figure is used to illustrate the two cases for showing the test is of level α.

Neyman-Pearson Lemma (Casella and Berger, 2002)
Consider testing $H_0:\theta=\theta_0$ versus $H_1:\theta=\theta_1$, where the pdf or pmf corresponding to θ_i is $f(\boldsymbol{x}|\theta_i), i=0,1$, using a test with rejection region R that satisfies

$$\boldsymbol{x}\in R \text{ if } f(\boldsymbol{x}|\theta_1)>k\, f(\boldsymbol{x}|\theta)$$

and

$$\boldsymbol{x}\in R^c \text{ if } f(\boldsymbol{x}|\theta_1)<k\, f(\boldsymbol{x}|\theta_0) \tag{8}$$

for some $k\geq 0$, and

$$\alpha=P_{\theta_0}(\boldsymbol{X}\in R). \tag{9}$$

Then
a. (Sufficiency) Any test that satisfies (8) and (9) is a uniformly most powerful (UMP) level α test.
b. (Necessity) If there exists a test satisfying (8) and (9) with $k>0$, then every UMP level α test is a size α test (satisfies (9)) and every UMP level α test satisfies (8) except perhaps on a set A satisfying $P_{\theta_0}(\boldsymbol{X}\in A)=P_{\theta_1}(\boldsymbol{X}\in A)=0$.

Proof We will prove the theorem for the case that $f(\boldsymbol{x}|\theta_0)$ and $f(\boldsymbol{x}|\theta_1)$ are pdfs of continuous random variables. The proof for discrete random variables can be accomplished by replacing integrals with sums.
Note first that any test satisfying (9) is a size α and, hence, a level α test because $sup_{\theta\in\Theta_0} P_\theta(\boldsymbol{X}\in R) = P_{\theta_0}(\boldsymbol{X}\in R)=\alpha$, since Θ_0 has only one point.
To ease notation, we define a test function, a function on the sample space that is 1 if $\boldsymbol{x}\in R$ and 0 if $\boldsymbol{x}\in R^c$. That is, it is the indicator function of the rejection region. Let $\phi(\boldsymbol{x})$ be the test function of a test satisfying (8) and (9). Let $\phi'(\boldsymbol{x})$ be the test function of any other level α test, and let $\beta(\theta)$ and $\beta'(\theta)$ be the power functions corresponding to the tests ϕ and ϕ' respectively. Because $0\leq\phi'(\boldsymbol{x})\leq 1$, (8) implies that $(\phi(\boldsymbol{x})-\phi'(\boldsymbol{x}))(f(\boldsymbol{x}|\theta_1)-k\, f(\boldsymbol{x}|\theta_0))\geq 0$ for every $\boldsymbol{x}$ (since $\phi=1$ if $f(\boldsymbol{x}|\theta_1)>k\, f(\boldsymbol{x}|\theta_0)$ and $\phi=0$ if $f(\boldsymbol{x}|\theta_1)<k\, f(\boldsymbol{x}|\theta_0)$). Thus

$$\begin{aligned} 0 &\leq \int\left[\phi(\boldsymbol{x})-\phi'(\boldsymbol{x})\right]\left[f(\boldsymbol{x}|\theta_1)-k\, f(\boldsymbol{x}|\theta_0)\right]d\,\boldsymbol{x} \\ &= \beta(\theta_1)-\beta'(\theta_1)-k\left(\beta(\theta_0)-\beta'(\theta_0)\right). \end{aligned} \tag{10}$$

Statement (a) is proved by noting that, since ϕ' is a level α test and ϕ is a size α test, $\beta(\theta_0)-\beta'(\theta_0)=\alpha-\beta'(\theta_0)\geq 0$. Thus (10) and $k\geq 0$ imply that

$$0\leq\beta(\theta_1)-\beta'(\theta_1)-k\left(\beta(\theta_0)-\beta'(\theta_0)\right)\leq\beta(\theta_1)-\beta'(\theta_1),$$

showing that $\beta(\theta_1) \geq \beta'(\theta_1)$ and hence ϕ has greater power than ϕ'. Since ϕ' was an arbitrary level α test and θ_1 is the only point in Θ_0^c, ϕ is a UMP level α test.

To prove statement (b), let ϕ' now be the test function for any UMP level α test. By part (a), ϕ', the test satisfying (8) and (9), is also UMP level α test, thus $\beta(\theta_1)=\beta'(\theta_1)$. This fact, (10), and $k \geq 0$ imply

$$\alpha - \beta'(\theta_0) = \beta(\theta_0) - \beta'(\theta_0) \leq 0.$$

Now since ϕ' is a level α test, $\beta'(\theta_0) \leq \alpha$. Thus $\beta'(\theta_0)=\alpha$, that is, ϕ' is a size α test, and this also implies that (10) is an equality in this case. But the nonnegative integrand $\left(\phi(\boldsymbol{x}) - \phi'(\boldsymbol{x})\right)\left(f(\boldsymbol{x}|\theta_1) - k f(\boldsymbol{x}|\theta_0)\right)$ will have a zero integral only if ϕ' satisfies (8), except perhaps on a set A with $\int_A f(\boldsymbol{x}|\theta_i) d\boldsymbol{x} = 0$.. This implies that the last assertion in statement (b) is true. ■

With this second lemma variant supplied, there is direct mention that this lemma applies to both continuous and discrete random variables where the lemma statement by Roussas only mentions continuous random variables. In terms of the parameter θ, Casella and Berger imply the parameter is one-dimensional in contrast to the multidimensional parameter inferred by Roussas. In light of these differences, the one-dimensional parameter is commonly used as examples used to demonstrate the Neyman-Pearson Lemma typically involve one parameter of interest. Should there be more than one parameter (e.g. Gamma or Binomial distributions) the parameters not of interest are typically assumed to be known and fixed. Of course, simple hypotheses can be in terms of more than one parameter such as a Normal distribution. From the lemma statement by Roussas, possible hypotheses can be $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs. $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ where $\boldsymbol{\theta} = (\mu, \sigma^2)^T$. The Casella and Berger lemma statement also explicitly addresses the components of sufficiency and necessity. These statements are provided in Roussas' statement of the lemma but they are merged into one sentence and not directly pointed out.

In the above proof, like in Roussas' proof, the proof assumes continuous random variables. At this point, it is worth mentioning that all remaining proofs for the following lemma variants will be assuming continuous random variables. It is also the case that the authors for the remaining variants make mention that the proof holds for the discrete case by merely replacing integration with summation.

Neyman-Pearson Lemma (Bain, Engelhardt, 1990)

Suppose that $X_1, \ldots, X_n$ have joint pdf $f(x_1, \ldots, x_n; \theta)$. Let

$$\lambda(x_1, \ldots, x_n; \theta_0, \theta_1) = \frac{f(x_1, \ldots, x_n; \theta_0)}{f(x_1, \ldots, x_n; \theta_1)} \tag{11}$$

and let C^* be the set

$$C^* = \{(x_1, \ldots, x_n) | \lambda(x_1, \ldots, x_n; \theta_0, \theta_1) \leq k\} \tag{12}$$

Where k is a constant such that

$$P\left[(X_1, \ldots, X_n) \in C^* \middle| \theta_0\right] = \alpha \tag{13}$$

Then C^* is a most powerful critical region of size α for testing $H_0: \theta = \theta_0$ versus $H_A: \theta = \theta_1$.

Proof For convenience, we will adopt vector notation $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{x} = (x_1, \ldots, x_n)$. Also, if A is an n – dimensional event, let

$$P\left[\boldsymbol{X} \in A \middle| \theta\right] = \int_A f(\boldsymbol{x}; \theta)\, d\boldsymbol{x} = \int \cdot \underset{A}{\cdot} \cdot \int f(x_1, \ldots, x_n; \theta)\, dx_1, \ldots, dx_n \tag{14}$$

for the continuous case. The discrete case would be similar, with integrals replaced by summations. We also will denote the complement of a set C by $\bar{C}$. Note that if A is a subset of C^*, then

$$P\left[\boldsymbol{X} \in A \middle| \theta_0\right] \le k\, P\left[\boldsymbol{X} \in A \middle| \theta_1\right] \tag{15}$$

Because $\int_A f(\boldsymbol{x}; \theta_0)\, d\boldsymbol{x} \le \int_A k\, f(\boldsymbol{x}; \theta_1)\, d\boldsymbol{x}$. Similarly, if A is a subset of $\bar{C}^*$, then

$$P\left[\boldsymbol{X} \in A \middle| \theta_0\right] \ge k\, P\left[\boldsymbol{X} \in A \middle| \theta_1\right] \tag{16}$$

Notice that for any critical region C, we have

$$C^* = (C^* \cap C) \cup (C^* \cap \bar{C}) \qquad \text{and} \qquad C = (C \cap C^*) \cup (C \cap \bar{C}^*)$$

Thus,

$$\pi_{C^*}(\theta) = P\left[\boldsymbol{X} \in C^* \cap C \middle| \theta\right] + P\left[\boldsymbol{X} \in C^* \cap \bar{C} \middle| \theta\right]$$

and

$$\pi_C(\theta) = P\left[\boldsymbol{X} \in C^* \cap C \middle| \theta\right] + P\left[\boldsymbol{X} \in C \cap \bar{C}^* \middle| \theta\right]$$

and the difference is

$$\pi_{C^*}(\theta) - \pi_C(\theta) = P\left[\boldsymbol{X} \in C^* \cap \bar{C} \middle| \theta\right] - P\left[\boldsymbol{X} \in C \cap \bar{C}^* \middle| \theta\right] \tag{17}$$

Combining equation (17) with $\theta = \theta_1$ and inequalities (15) and (16), we have

$$\pi_{C^*}(\theta_1) - \pi_C(\theta_1) \ge (1/k)\left\{P\left[\boldsymbol{x} \in C^* \cap \bar{C} \middle| \theta_0\right] - P\left[\boldsymbol{X} \in C \cap \bar{C}^* \middle| \theta_0\right]\right\}$$

Again, using (17) with $\theta = \theta_0$ and the right side of this inequality, we obtain

$$\pi_{C^*}(\theta_1) - \pi_C(\theta_1) \ge (1/k)\left[\pi_{C^*}(\theta_0) - \pi_C(\theta_0)\right]$$

If C is a critical region of size α, then $\pi_{C^*}(\theta_0)-\pi_C(\theta_0)=\alpha-\alpha=0$, and the right side of the last inequality is 0, and thus $\pi_{C^*}(\theta_1)\geq\pi_C(\theta_1)$. ■

Similar to the lemma statement and proof provided by Casella and Berger, Bain and Engelhardt's version of the lemma implies a one-dimensional parameter (lack of boldface or vector notation on θ). This version also makes extensive use of the power function (denoted as π) defined in terms of probabilities. Compared to the two proofs from earlier, the version from Bain and Engelhardt is quite short and simple. This stems from the lemma statement given by Bain and Engelhardt as this lemma version only mentions the "sufficiency" component like in the Casella and Berger lemma statement. This version by Bain and Engelhardt assumes the critical region is of size α as (13) has a strict equality. The idea of a level α critical region is still utilized as the arbitrary critical region C is first assumed to be of level α. The last statement in the proof considers what happens if the arbitrary critical region is of size α. If both C and C^* are of size α, the difference in power between the two regions will simply be zero. As mentioned earlier, all the lemmas make use of the continuous case and the authors point out that the discrete case is handled by simply replacing integration with summation. This version by Bain and Engelhardt make use of probability statements which are only implied to reference the continuous case through (14). Thus, this proof is very well adapted to cover both the discrete and continuous case almost simultaneously.

The Fundamental Lemma of Neyman and Pearson (Lehmann, 1991)
Let P_0 and P_1 be probability distributions possessing densities p_0 and p_1 respectively with respect to a measure μ.

(i) Existence. For testing $H: p_0$ against the alternative $K: p_1$ there exists a test ϕ and a constant k such that

$$E_0\phi(X)=\alpha \tag{18}$$

and

$$\phi(x)=\begin{cases}1 & \text{when} \quad p_1(x)>k\,p_0(x)\\ 0 & \text{when} \quad p_1(x)<k\,p_0(x)\end{cases} \tag{19}$$

(ii) Sufficient condition for a most powerful test. If a test satisfies (18) and (19) for some k, then it is most powerful for testing p_0 against p_1 at level α.

(iii) Necessary condition for a most powerful test. If ϕ is most powerful at level α for testing p_0 against p_1, then for some k it satisfies (19) a.e. μ. It also satisfies (18) unless there exists a test of size $<\alpha$ and with power 1.

Proof: For $\alpha=0$ and $\alpha=1$ the theorem is easily seen to be true provided the value $k=+\infty$ is admitted in (19) and $0\cdot\infty$ is interpreted as 0. Through the proof we shall therefore assume $0<\alpha<1$.

(i): Let $\alpha(c)=P_0\{p_1(x)>c\,p_0(X)\}$. Since the probability is computed under P_0, the inequality need be considered only for the set where $p_0(x)>0$, so that $\alpha(c)$ is the probability that the random variable $p_1(X)/p_0(X)$ exceeds c. Thus $1-\alpha(c)$ is a cumulative distribution function, and $\alpha(c)$ is nonincreasing and continuous in the right, $\alpha(c-0)-\alpha(c)=P_0\{p_1(X)/p_0(X)=c\}$, $\alpha(-\infty)=1$, and $\alpha(\infty)=0$. Given any $0<\alpha<1$, let c_0 be such that $\alpha(c_0)\le\alpha\le\alpha(c_0-0)$, and consider the test ϕ defined by

$$\phi(x)=\begin{cases} 1 & \text{when} \quad p_1(x)>c_0 p_0(x), \\ \dfrac{\alpha-\alpha(c_0)}{\alpha(c_0-0)-\alpha(c_0)} & \text{when} \quad p_1(x)=c_0 p_0(x), \\ 0 & \text{when} \quad p_1(x)<c_0 p_0(x). \end{cases}$$

Here the middle expression is meaningful unless $\alpha(c_0)=\alpha(c_0-0)$; since then $P_0\{p_1(X)=c_0 p_0(X)\}=0$, ϕ is defined a.e. The size of ϕ is

$$E_0\phi(X)=P_0\left\{\frac{p_1(X)}{p_0(X)}>c_0\right\}+\frac{\alpha-\alpha(c_0)}{\alpha(c_0-0)-\alpha(c_0)}P_0\left\{\frac{p_1(X)}{p_0(X)}=c_0\right\}=\alpha$$

so that c_0 can be taken as the k of the theorem.

It is of interest to note that c_0 is essentially unique. The only exception is the case than an interval of c's exists for which $\alpha(c)=\alpha$. If (c', c'') is such an interval, and

$$C=\left\{x: p_0(x)>0 \,\text{and}\, c'<\frac{p_1(x)}{p_0(x)}<c''\right\}$$

then $P_0(C)=\alpha(c')-\alpha(c''-0)=0$. By Problem 3 of Chapter 2, this implies $\mu(C)=0$ and hence $P_1(C)=0$. Thus the sets corresponding to two different values of c differ only in a set of points which has probability 0 under both distributions, that is, points that could be excluded from the sample space.

(ii): Suppose that ϕ is a test satisfying (18) and (19) and that ϕ^* is any other test with $E_0\phi^*(X)\le\alpha$. Denote by S^+ and S^- the sets in the sample space where $\phi(x)-\phi^*(x)>0$ and <0 respectively. If x is in S^+, $\phi(x)$ must be >0 and $p_1(x)\ge k\,p_0(x)$. In the same way $p_1(x)\le k\,p_0(x)$ for all x in S^-, and hence

$$\int(\phi-\phi^*)(p_1-k\,p_0)\,d\mu=\int_{S^+\cup S^-}(\phi-\phi^*)(p_1-k\,p_0)\,d\mu\ge 0\,.$$

The difference in power between ϕ and ϕ^* therefore satisfies

$$\int(\phi-\phi^*)\,p_1\,d\mu\ge k\int(\phi-\phi^*)\,p_0\,d\mu\ge 0\,,$$

as was to be proved.

(iii): Let ϕ^* be most powerful at level α for testing p_0 against p_1, and let ϕ satisfy (18) and (19). Let S be the intersection of the set $S^+\cup S^-$,, on which ϕ and ϕ^* differ, with the set $\{x: p_1(x)\ne k\,p_0(x)\}$, and suppose that $\mu(S)>0$. Since $(\phi-\phi^*)(p_1-k\,p_0)$ is positive on S, it follows from Problem 3 of Chapter 2 that

$$\int_{S^+\cup S^-} (\phi-\phi^*)(p_1-k\,p_0)\,d\mu=\int_S (\phi-\phi^*)(p_1-k\,p_0)\,d\mu>0$$

and hence that ϕ is more powerful against p_1 than ϕ^*. This is a contradiction, and therefore $\mu(S)=0$,, as was to be proved. ■

The notation that Lehmann uses (as well as some other authors from above) is sloppy in the sense that the lemma statement and proof appear to be done in the univariate case as there is a lack of vector notation or bold-facing. However, Lehmann is working under a multidimensional case and it is understood in the context of his text. The proof for the Lehmann variation begins by stating the theorem is easily seen to be true if α is either 0 or 1, a condition that was not mentioned in the proofs for the previous three lemma statements. This variation also breaks the lemma up into three nice results: the existence of a most powerful test, the sufficient condition for most powerful tests, and the necessary condition for most powerful tests. This form is very similar to the Casella and Berger variation but the existence of a most powerful test is not directly mentioned. In the proof of the lemma, the test function $\phi(x)$ is defined almost identically to the test function in Roussas' statement of the lemma. The only difference between them is where Roussas says $\phi(x_1,\ldots,x_n)=\gamma$ if $\prod f(x_i,\boldsymbol{\theta}_i)=C\prod f(x_i,\boldsymbol{\theta}_0)$ for $i=1,\ldots,n$ (from (1)) and defines γ later as $\frac{\alpha-a(C_0)}{a(C_0-)-a(C_0)}$, Lehmann's version directly places this value for γ in the definition of the test function. With that being said, it is worth mentioning that Lehmann defines the notation of $\alpha(c_0-0)$ the same as Roussas defines $a(C_0-)$. Going back to Figure 1 in the proof provided by Roussas, $a(C_0-)$ is defined as the limit of $a(C)$ as $C\to C_0$ from the left. Which implies $\alpha(c_0-0)$ is defined as the limit of $\alpha(c)$ as $c\to c_0$ from the left. In fact, Lehmann states that $1-\alpha(c)$ is a cumulative distribution function, similar to Roussas' statement about $a(C)$.

Neyman-Pearson Lemma (Dudewicz & Mishra, 1988)
Suppose that X is a random variable with distribution function $F(x|\theta)$ where θ is unknown, $\theta\in\Theta$ and $x\in\mathcal{X}$. Assume $F(x|\theta)$ is either absolutely continuous [in which case let $f(x|\theta)$ denote its density at x] or discrete [in which case let $f(x|\theta)$ denote $P_\theta[X=x]$]. Suppose that $H_0=\{\theta_0\}$ and $H_1=\{\theta_1\}$ [often written as $H_0:\theta=\theta_0$ and $H_1:\theta=\theta_1$ in this case] are simple null and alternative hypotheses, respectively.
For testing H_0 versus H_1 after observing X, if a test with critical function $\phi(\cdot)$ satisfies (for some k)

$$E_{\theta_0}\phi(X)=\alpha \tag{20}$$

and

$$\phi(x)=\begin{cases}1, & \text{when} \quad f(x|\theta_1)>k\,f(x|\theta_0)\\ 0, & \text{when} \quad f(x|\theta_1)<k\,f(x|\theta_0),\end{cases} \tag{21}$$

then it is most powerful for testing H_0 versus H_1 at level α.
Proof: [We give the proof for the absolutely continuous case, the discrete case being analogous with summations replacing integrals.] We write, for example $\int_{-\infty}^{\infty} g(x)\,dx$ as $\int_{x\in\mathcal{R}} g(x)\,dx$ as a shorthand (since the sets over which we are

integrating are notationally complex). For this use, define

$$\begin{cases} c_1 = \{x : f(x|\theta_1) > k\, f(x|\theta_0)\}, \\ c_2 = \{x : f(x|\theta_1) < k\, f(x|\theta_0)\}, \\ c_3 = \{x : f(x|\theta_1) = k\, f(x|\theta_0)\}. \end{cases} \tag{22}$$

Suppose $\phi(\cdot)$ satisfies equations (20) and (21). Let $\psi(\cdot)$ be the critical function of any other test with level α. Then the difference of the powers of the two tests at H_1 is

$$\begin{aligned} E_{\theta_1}\phi(X) - E_{\theta_1}\psi(X) &= E_{\theta_1}\left[\phi(X) - \psi(X)\right] \\ &= \int_{x\in\mathcal{X}} \left[\phi(x) - \psi(x)\right] f(x|\theta_1)\,dx \\ &= \int_{x\in c_1} \left[\phi(x) - \psi(x)\right] f(x|\theta_1)\,dx \\ &\quad + \int_{x\in c_2} \left[\phi(x) - \psi(x)\right] f(x|\theta_1)\,dx \\ &\quad + \int_{x\in c_3} \left[\phi(x) - \psi(x)\right] f(x|\theta_1)\,dx \end{aligned} \tag{23}$$

Using the definitions of c_1, c_2, and c_3, we know that: $\phi(x)=1$ for $x\in c_1$; and $\phi(x)=0$ for $x\in c_2$. Substituting these in equation (23) and then using the facts $f(x|\theta_1) > k\, f(x|\theta_0)$ for $x\in c_1$, $f(x|\theta_1) < k\, f(x|\theta_0)$ for $x\in c_2$, and $f(x|\theta_1) = k\, f(x|\theta_0)$ for $x\in c_3$, we find

$$\begin{aligned} E_{\theta_1}\phi(X) - E_{\theta_1}\psi(X) &= \int_{x\in c_1} \left[1 - \psi(x)\right] f(x|\theta_1)\,dx \\ &\quad + \int_{x\in c_2} \left[-\psi(x)\right] f(x|\theta_1)\,dx \\ &\quad + \int_{x\in c_3} \left[\phi(x) - \psi(x)\right] f(x|\theta_1)\,dx \\ &\geq \int_{x\in c_1} \left[1 - \psi(x)\right] k\, f(x|\theta_0)\,dx \\ &\quad + \int_{x\in c_2} \left[-\psi(x)\right] k\, f(x|\theta_0)\,dx \\ &\quad + \int_{x\in c_3} \left[\phi(x) - \psi(x)\right] k\, f(x|\theta_0)\,dx \end{aligned}$$

$$
\begin{aligned}
&= \int_{x \in c_1} [\phi(x) - \psi(x)] k\, f(x|\theta_1)\, dx \\
&\quad + \int_{x \in c_2} [\phi(x) - \psi(x)] k\, f(x|\theta_1)\, dx \\
&\quad + \int_{x \in c_3} [\phi(x) - \psi(x)] k\, f(x|\theta_1)\, dx \\
&= k \int_{x \in X} [\phi(x) - \psi(x)] f(x|\theta_0)\, dx \\
&= k\, E_{\theta_0}[\phi(X) - \psi(X)] = k(\alpha - \alpha) = 0
\end{aligned} \tag{24}
$$

Hence $E_{\theta_1}\phi(X) \geq E_{\theta_1}(X)$, as was to be shown. ■

Dudewicz and Mishra's version is similar to Lehmann's version as both define the test function in the same way. That is, both versions have the test function stated in terms of the conditions $f(x|\theta_1) > k\, f(x|\theta_0)$ and $f(x|\theta_1) < k\, f(x|\theta_0)$, yet both proofs look at a third condition where equality holds similar to Roussas' proof. Dudewicz and Mishra's proof is rather simple as the proof relies on simply showing the difference in power between the test function $\phi(\cdot)$ which satisfies (20) and (21) and any arbitrary test function $\psi(\cdot)$. Under the alternative hypothesis, this difference is larger than the same difference under the null hypothesis. Working with the expectations, they are first written as their corresponding integrals. Making used of the stipulations indicated in equations (21) and (22), the integrals, which are using $\theta = \theta_1$ are shown greater than their counterparts with $\theta = \theta_0$ which matches the form of the null hypothesis.

Neyman-Pearson Theorem (Hogg, McKean, & Craig, 2005)
Let $X_1, X_2, \ldots, X_n$, where n is a fixed positive integer, denote a random sample from a distribution that has pdf or pmf $f(x;\theta)$. Then the likelihood of $X_1, X_2, \ldots, X_n$ is

$$L(\theta, \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i;\theta), \text{ for } \boldsymbol{x}' = (x_1, \ldots, x_n)$$

Let θ' and θ'' be distinct fixed values of θ so that $\Omega = \{\theta : \theta = \theta', \theta''\}$, and let k be a positive number. Let C be a subset of the sample space such that:

(a) $\dfrac{L(\theta';\boldsymbol{x})}{L(\theta'';\boldsymbol{x})} \leq k$, for each point $\boldsymbol{x} \in C$

(b) $\dfrac{L(\theta';\boldsymbol{x})}{L(\theta'';\boldsymbol{x})} \geq k$, for each point $\boldsymbol{x} \in C^c$

(c) $\alpha = P_{H_0}[\boldsymbol{X} \in C]$

Then C is a best critical region of size α for testing the simple hypothesis $H_0 : \theta = \theta'$ against the alternative simple hypothesis $H_1 : \theta = \theta''$.

Proof: We shall give the proof when the random variables are of the continuous type. If C is the only critical region of size α, the theorem is proved. If there is another critical region of size α, denote it by A. For convenience, we shall let $\int \underset{R}{\cdots} \int L(\theta; x_1, \ldots, x_n)\, dx_1 \ldots d\, x_n$ be denoted by $\int_R L(\theta)$. In this notation we wish to show that

$$\int_C L(\theta'') - \int_A L(\theta'') \geq 0 .$$

Since C is the union of the disjoint sets $C \cap A$ and $C \cap A^c$ and A is the union of the disjoint sets $A \cap C$ and $A \cap C^c$, we have

$$\begin{aligned} \int_C L(\theta'') - \int_A L(\theta'') &= \int_{C \cap A} L(\theta'') + \int_{C \cap A^c} L(\theta'') - \int_{A \cap C} L(\theta'') - \int_{A \cap C^c} L(\theta'') \\ &= \int_{C \cap A^c} L(\theta'') - \int_{A \cap C^c} L(\theta'') \end{aligned} \tag{25}$$

However, by the hypothesis of the theorem, $L(\theta'') \geq (1/k) L(\theta')$ at each point of C, and hence at each point of $C \cap A^c$; thus

$$\int_{C \cap A^c} L(\theta'') \geq \frac{1}{k} \int_{C \cap A^c} L(\theta') .$$

But $L(\theta'') \leq (1/k) L(\theta')$ at each point of C^c, and hence at each point of $A \cap C^c$; accordingly,

$$\int_{A \cap C^c} L(\theta'') \leq \frac{1}{k} \int_{C \cap A^c} L(\theta') .$$

These inequalities imply that

$$\int_{C \cap A^c} L(\theta'') - \int_{A \cap C^c} L(\theta'') \geq \frac{1}{k} \int_{C \cap A^c} L(\theta') - \frac{1}{k} \int_{A \cap C^c} L(\theta') ;$$

and, from equation (25), we obtain

$$\int_C L(\theta'') - \int_A L(\theta'') \geq \frac{1}{k} \left[\int_{C \cap A^c} L(\theta') - \int_{A \cap C^c} L(\theta') \right] . \tag{26}$$

However,

$$\begin{aligned}\int_{C\cap A^c} L(\theta') - \int_{A\cap C^c} L(\theta') &= \int_{C\cap A^c} L(\theta') + \int_{C\cap A} L(\theta') \\ &\quad - \int_{A\cap C} L(\theta') - \int_{A\cap C^c} L(\theta') \\ &= \int_C L(\theta') - \int_A L(\theta') = \alpha - \alpha = 0.\end{aligned}$$

If this result is substituted in inequality (26), we obtain the desired result,

$$\int_C L(\theta'') - \int_A L(\theta'') \geq 0.$$

If the random variables are of the discrete type, the proof is the same with integration replaced by summation. ■

Hogg, McKean, and Craig's version of the lemma makes a strong connection to the likelihood function and likelihood ratio, but does so without the use of a test statistic. Of course, the previous versions of the lemma have made use of likelihoods and likelihood ratios but in a more indirect way. In fact, Bain and Engelhardt's version makes use of the likelihood ratio in its typical form (although joint pdfs are used instead of likelihoods). Hogg, McKean, and Craig's proof also follows similarly to Bain and Engelhardt's proof as both make use of disjoint sets. Differences between these two proofs include the direct usage of integrals by Hogg, McKean, and Craig where Bain and Engelhardt's use of integrals is indirect through the use of probabilities. Additionally, Bain and Engelhardt express the power function using π where Hogg, McKean, and Craig do not and make strict usage of integrals. Aside from these differences, the proof provided by Hogg, McKean, and Craig is otherwise the same as the proof from Bain and Engelhardt.

Neyman-Pearson Lemma (Bickel & Doksum, 1977)
Let δ_k be the critical function of the test which rejects $H:\theta=\theta_0$ if, and only if, the likelihood ratio is at least k, where $0 \leq k \leq \infty$. Let δ be the critical function of any test whose size is no greater than the size of δ_k, that is

$$\beta(\theta_0, \delta) \leq \beta(\theta_0, \delta_k). \tag{27}$$

Then we must have

$$\beta(\theta_1, \delta) \leq \beta(\theta_1, \delta_k). \tag{28}$$

Proof: We shall prove a little more. Let ψ be any function on $\mathbb{R}^n$ such that,

$$\begin{aligned}&(\mathrm{a})\, 0 \leq \psi(\boldsymbol{x}) \leq 1 \text{ for all } \boldsymbol{x}. \\ &(\mathrm{b})\, E_{\theta_0}(\psi(\boldsymbol{X})) \leq E_{\theta_0}(\delta_k(\boldsymbol{X})).\end{aligned} \tag{29}$$

Then we shall show that

$$E_{\theta_1}(\psi(\boldsymbol{X})) \leq E_{\theta_1}(\delta_k(\boldsymbol{X})). \tag{30}$$

If we recall that for any critical function δ,

$$\beta(\theta, \delta) = E_\theta(\delta(\boldsymbol{X})),$$

we see that the theorem follows from (30) by putting $\psi = \delta$, the critical function of any specified level α test.
To prove (30) suppose that $\boldsymbol{X}$ is continuous. The proof is the same for $\boldsymbol{X}$ discrete except that integrals are replaced by sums. The key idea is to note that since $0 \leq \psi(\boldsymbol{x}) \leq 1$, the definition of δ_k implies that for all $\boldsymbol{x}$

$$\psi(\boldsymbol{x})(p(\boldsymbol{x}, \theta_1) - k\, p(\boldsymbol{x}, \theta_0)) \leq \delta_k(\boldsymbol{x})(p(\boldsymbol{x}, \theta_1) - k\, p(\boldsymbol{x}, \theta_0)). \tag{31}$$

This is because $p(\boldsymbol{x}, \theta_1) - kp(\boldsymbol{x}, \theta_0)$ is < 0 or ≥ 0 accordingly as $\delta_k(\boldsymbol{x})$ is 0 or 1. If we integrate both sides of (31) with respect to $x_1, \ldots, x_n$, we get

$$\begin{aligned} &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi(\boldsymbol{x})\, p(\boldsymbol{x}, \theta_1)\, d\boldsymbol{x} - k \int_{-\infty}^{\infty} \cdots \int_{\infty}^{\infty} \psi(\boldsymbol{x})\, p(\boldsymbol{x}, \theta_0)\, d\boldsymbol{x} \\ &\leq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta_k(\boldsymbol{x})\, p(\boldsymbol{x}, \theta_1)\, d\boldsymbol{x} - k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta_k(\boldsymbol{x})\, p(\boldsymbol{x}, \theta_0)\, d\boldsymbol{x}. \end{aligned} \tag{32}$$

If $k < \infty$, (32) is equivalent to,

$$E_{\theta_1}(\psi(\boldsymbol{X})) - E_{\theta_1}(\delta_k(\boldsymbol{X})) \leq k\{E_{\theta_0}(\psi(\boldsymbol{X})) - E_{\theta_0}(\delta_k(\boldsymbol{X}))\}. \tag{33}$$

Since k is nonnegative (29)(b) and (33) imply (32). The theorem follows, if k is finite. The case $k = \infty$ is left to the reader. ■

This lemma variant provided by Bickel and Doksum is interesting in how they define the necessary condition for the uniformly most powerful test as the power function is used to relate the relative sizes of the critical functions δ_k and δ. Notationally, (31) makes use of p which the authors used to denote the density or frequency function for the x's. This is mentioned in the paragraphs directly preceding the lemma statement. The proof only encompasses the instance that the constant k is finite. As stated in the final sentence of the proof, the case where k is infinite is left to the reader. Because looking at comparisons between the Neyman-Pearson Lemmas is the only topic of interest, further detail on this matter will not be discussed. The authors also make mention that there is a simple generalization of this which is also known as the Neyman-Pearson Lemma, but they do not go into detail about this generalization.

Neyman-Pearson Lemma (Mukhopadhyay, 2000)
Consider a test of H_0 versus H_1 stated as $H_0: \theta=\theta_0$ versus $H_1: \theta=\theta_1$ with the rejection region and acceptance region for the null hypothesis H_0 defined as follows:

$$\boldsymbol{x} \in R \text{ if } L(\boldsymbol{x};\theta_1) > k\, L(\boldsymbol{x};\theta_0)$$
$$\boldsymbol{x} \in R^c \text{ if } L(\boldsymbol{x};\theta_1) < k\, L(\boldsymbol{x};\theta_0)$$

or equivalently, suppose that the test function has the form

$$\psi(\boldsymbol{x}) = \begin{cases} 1 \text{ if } L(\boldsymbol{x};\theta_1) > k\, L(\boldsymbol{x};\theta_0) \\ 0 \text{ if } L(\boldsymbol{x};\theta_1) < k\, L(\boldsymbol{x};\theta_0) \end{cases} \tag{34}$$

where the constant $k(\geq 0)$ is so determined that

$$E_{\theta_0}\{\psi(\boldsymbol{X})\} = \alpha. \tag{35}$$

Any test satisfying (34)-(35) is a MP level α test.

Proof: We give a proof assuming that the X's are continuous random variables. The discrete case can be disposed off by replacing the integrals with the corresponding sums. First note that any test which satisfies (35) has size α and hence it is level α too. We already have a level α test function $\psi(\boldsymbol{x})$ defined by (34)-(35). Let $\psi^*(\boldsymbol{x})$ be the test function of any other level α test. Suppose that $Q(\theta), Q^*(\theta)$ are respectively the power functions associated with the test functions ψ, ψ^*. Now, let us first verify that

$$\{\psi(\boldsymbol{x}) - \psi^*(\boldsymbol{x})\}\{L(\boldsymbol{x},\theta_1) - k\, L(\boldsymbol{x};\theta_0)\} \geq 0 \text{ for all } x \in \mathcal{X}^n. \tag{36}$$

Suppose that $\boldsymbol{x} \in \mathcal{X}^n$ is such that $\psi(\boldsymbol{x})=1$ which implies $L(\boldsymbol{x};\theta_1) - k\, L(\boldsymbol{x};\theta_0) > 0$, by the definition of ψ in (34). Also for such $\boldsymbol{x}$, one obviously has $\psi(\boldsymbol{x}) - \psi^*(\boldsymbol{x}) \geq 0$ since $\psi^*(\boldsymbol{x}) \in (0,1)$. That is, if $\boldsymbol{x} \in \mathcal{X}^n$ is such that $\psi(\boldsymbol{x})=1$, we have verified (36). Next, suppose that $\boldsymbol{x} \in \mathcal{X}^n$ is such that $\psi(\boldsymbol{x})=0$ which implies $L(\boldsymbol{x};\theta_1) - kL(\boldsymbol{x};\theta_0) < 0$, by definition of ψ in (34). Also for such $\boldsymbol{x}$ one obviously has $\psi(\boldsymbol{x}) - \psi^*(\boldsymbol{x}) = 0$ since $\psi^*(\boldsymbol{x}) \in (0,1)$. Again (36) is validated. Now, if $\boldsymbol{x} \in \mathcal{X}^n$ is such that $0 < \psi(\boldsymbol{x}) < 1$, then from (34) we must have $L(\boldsymbol{x};\theta_1) - k\, L(\boldsymbol{x};\theta_0) = 0$, and again (36) is validated. That is, (36) surely holds for all $\boldsymbol{x} \in \mathcal{X}^n$. Hence we have

$$
\begin{aligned}
0 \quad &\leq \int \cdots \int_{\mathcal{X}^n} \{\psi(\boldsymbol{x}) - \psi^*(\boldsymbol{x})\}\{L(\boldsymbol{x};\theta_1) - k\, L(\boldsymbol{x};\theta_0)\} \prod_{i=1}^{n} d\, x_i \\
&= \int \cdots \int_{\mathcal{X}^n} \psi(\boldsymbol{x})\{L(\boldsymbol{x};\theta_1) - k\, L(\boldsymbol{x};\theta_0)\} \prod_{i=1}^{n} dx_i \\
&\quad - \int \cdots \int_{\mathcal{X}^n} \psi^*(\boldsymbol{x})\{(\boldsymbol{x};\theta_1) - k\, L(\boldsymbol{x};\theta_0)\} \prod_{i=1}^{n} dx_i \\
&= \{E_{\theta_1}[\psi(\boldsymbol{X})] - k\, E_{\theta_0}[\psi(\boldsymbol{X})]\} - \{E_{\theta_0}[\psi^*(\boldsymbol{X})] - k\, E_{\theta_0} E[\psi^*(\boldsymbol{X})]\} \\
&= \{Q(\theta_1) - Q^*(\theta_1)\} - k\{Q(\theta_0) - Q^*(\theta_0)\}.
\end{aligned} \tag{37}
$$

Now recall that $Q(\theta_0)$ is the Type I error probability associated with the test ψ defined in (34) and thus $Q(\theta_0)=\alpha$ from (35). Also, $Q^*(\theta_0)$ is the similar entity associated with the test ψ^* which is assumed to have the level α, that is $Q^*(\theta_0)\leq\alpha$. Thus, $Q(\theta_0)-Q^*(\theta_0)\geq 0$ and hence we can rewrite (37) as

$$Q(\theta_1)-Q^*(\theta_1)\geq k\{Q(\theta_0)-Q^*(\theta_0)\}\geq 0,$$

which shows that $Q(\theta_1)\geq Q^*(\theta_1)$. Hence, the test associated with ψ is at least as powerful as the one associated with ψ^*. But, ψ^* is any arbitrary level α test to begin with. The proof is now complete. ■

This variant provided by Mukhopadhyay has a statement that closely matches the version provided by Casella and Berger regarding conditions to meet. Mukhopadhyay's version, however, does not provide insight on the necessity condition like that in Casella and Berger's. Mukhopadhyay's proof begins similarly to several we have seen thus far with perhaps slight differences in notation. The second paragraph of this proof concerns the relationship between (34) and (36). Given how the test function is stated in (34), one might argue that it is intuitively obvious that (36) follows when assuming $\psi(\boldsymbol{x})$ is defined as 0 or 1. Even if this isn't so, it helps in the coherence of the proof. This section of the proof is more interesting when considering $\phi(\boldsymbol{x})\in(0,1)$ as that is not depicted in (34). It's not hard to see that $\psi(\boldsymbol{x})\in(0,1)$ must lead to $L(\boldsymbol{x},\theta_1)=k\, L(\boldsymbol{x},\theta_0)$ as that is the only option left considering the possible values of $\psi(\boldsymbol{x})$ (unless Does Not Exist is a possibility which is a whole other matter). The rest of this proof simply relates ψ and ψ^* to their corresponding power function Q and Q^*, wherein Mukhopadhyay shows in a quite simple fashion that $\psi(\boldsymbol{x})$ is most powerful.

Neyman-Pearson Lemma (Mood, Graybill, & Boes, 1973)
Let $X_1,\ldots,X_n$ be a random sample from $f(x;\theta)$, where θ is one of the two known values θ_0 or θ_1 and let $0<\alpha<1$ be fixed.
Let k^* be a positive constant and C^* be a subset of $\mathcal{X}$ which satisfy:

(i) $P_{\theta_0}[(X_1,\ldots,X_n)\in C^*]=\alpha.$ (38)

(ii) $\lambda=\dfrac{L(\theta_0;x_1,\ldots,x_n)}{L(\theta_1;x_1,\ldots,x_n)}=\dfrac{L_0}{L_1}\leq k^*$ if $(x_1,\ldots,x_n)\in C^*$ and $\lambda\geq k^*\in(x_1,\ldots,x_n)\in\bar{C}^*$ (39)

Then the test $\mathcal{Y}^*$ corresponding to the critical region C^* is a most powerful test of size α of $H_0:\theta=\theta_0$ versus $H_1:\theta=\theta_1$. [Recall that $L_j=L(\theta_j;x_1,\ldots,x_n)=\prod_{i=1}^n f(x_i;\theta_j)$ for $j=0$ or 1 and $\bar{C}^*$ is the compliment of C^*; that is, $\bar{C}^*=\mathcal{X}-C^*$.]

Proof: Suppose that k^* and C^* satisfying conditions (i) and (ii) exist. If there is no other test of size α or less, then $\mathcal{Y}^*$ is automatically most powerful. Let $\mathcal{Y}$ be another test of size α or less, and let C be its corresponding critical region. We have $P_{\theta_0}[(X_1,\ldots,X_n)\in C]\leq\alpha$. We must show that $\pi_{\mathcal{Y}^*}(\theta_1)\geq\pi_{\mathcal{Y}}(\theta_1)$ to complete the proof. {For any subset R of $\mathcal{X}$, let us abbreviate $\underset{R}{\int\cdots\int}\left[\prod_{i=1}^n f(x_i;\theta_j)dx_i\right]$ as $\int_R L_j$ for $j=0,1$. Our notation indicates that $f_0(\cdot)$ and $f_1(\cdot)$ are probability density functions. The same proof holds for discrete density functions.} Showing that $\pi_{\mathcal{Y}^*}(\theta_1)\geq\pi_{\mathcal{Y}}(\theta_1)$ is equivalent to showing that $\int_{C^*}L_1\geq\int_C L_1$. See Figure 2.

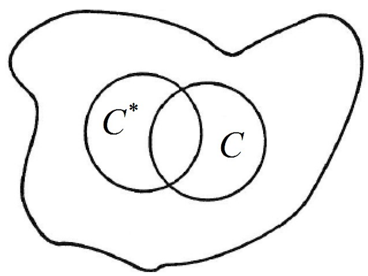

Figure 2

Now $\int_{C^*}L_1-\int_C L_1=\int_{C^*\bar{C}}L_1-\int_{C\bar{C}^*}L_1\geq(1/k^*)\int_{C^*\bar{C}}L_0-(1/k^*)\int_{C\bar{C}^*}L_0$ since $L_1\geq L_0/k^*$ on C^* (hence also on $C^*\bar{C}$) and $L_1\leq L_0/k^*$, or $-L_1\geq -L_0/k^*$,, on $\bar{C}^*$ (hence also on $C\bar{C}^*$). But $(1/k^*)\left(\int_{C^*\bar{C}}L_0-\int_{C\bar{C}^*}L_0\right)=(1/k^*)\left(\int_{C^*\bar{C}}L_0+\int_{C^*C}L_0-\int_{C^*C}L_0-\int_{C\bar{C}^*}L_0\right)$ $=(1/k^*)\left(\int_{C^*}L_0-\int_C L_0\right)=(1/k^*)(\alpha-\text{ size of test }\mathcal{Y})\geq 0$; so $\int_{C^*}L_1-\int_C L_1>0$, as was to be shown. ■

Bain and Engelhardt's version of the lemma is quite similar to this version by Mood, Greybill, and Boes as both directly define a most powerful critical region through the use of the likelihood ratio. In fact, Hogg, McKean, and Craig's version is similar in this approach as well. Among these three variations, there are no significant differences in the lemma statement. As for the proof, it can be argued that Mood, Greybill, and Boes have the most elegant proof compared to Bain and Engelhardt's version and Hogg, McKean, and Craig's as the proof provided by Mood, Greybill and Boes is the shortest (although not by much). It also helps in comparing the various critical regions and their respective complements with the provided figure for the proof (Figure 2). This helps to make clear the potential relationship between the critical region of interest (C^*) and an arbitrary one (C). With the exception of the included figure, this proof provided by Mood, Greybill, and Boes is quite similar to the proof by Hogg, McKean, and Craig as both provide very similar variations to the lemma.

The Neyman-Pearson Theorem (Young & Smith, 2005)

(a) (Optimality). For any K and $\gamma(x)$, the test ϕ_0 has maximum power among all tests whose sizes are no greater than the size of ϕ_0.

(b) (Existence). Given $\alpha\in(0,1)$, there exists constants K and γ_0 such that the LRT defined by this K and $\gamma(x)=\gamma_0$ for all x has size exactly α.

(c) (Uniqueness). If the test ϕ has size α, and is of maximum power amongst all possible tests of size α, then ϕ is necessarily a likelihood ratio test, except possibly on a set of values of x which has probability 0 under H_0 and H_1.

Proof: assuming absolute continuity

(a) Let ϕ be any test for which $E_{\theta_0}\phi(X)\le E_{\theta_0}\phi_0(X)$. Define $U(x)=\{\phi_0(x)-\phi(x)\}\{f_1(x)-Kf_0(x)\}$. When $f_1(x)-Kf_0(x)>0$ we have $\phi_0(x)=1$, so $U(x)\ge 0$. When $f_1(x)-Kf_0(x)<0$ we have $\phi_0(x)=0$, so $U(x)\ge 0$. For $f_1(x)-Kf_0(x)=0$, of course $U(x)=0$. thus $U(x)\ge 0$ for all x. Hence

$$\begin{aligned}0 &\le \int \{\phi_0(x)-\phi(x)\}\{f_1(x)-Kf_0(x)\}dx\\ &=\int \phi_0(x)f_1(x)dx-\int \phi(x)f_1(x)dx+K\left\{\int \phi(x)f_0(x)dx-\int \phi_0(x)f_0(x)dx\right\}\\ &=E_{\theta_1}\phi_0(X)-E_{\theta_1}\phi(X)+K\{E_{\theta_0}\phi(X)-E_{\theta_0}\phi_0(X)\}\end{aligned}$$

However, the expression in curly brackets is ≤ 0, because of the assumption that the size of ϕ is no greater than the size of ϕ_0. Thus

$$\int \phi_0(x)f_1(x)dx-\int \phi(x)f_1(x)dx\ge 0,$$

which establishes that the power of ϕ cannot be greater than the power of ϕ_0, as claimed.

(b) The probability distribution function $G(K)=P_{\theta_0}\{\Lambda(X)\le K\}$ is non-decreasing as K increases; it is also right-continuous (so that $G(K)=\lim_{y\to K}G(y)$ for each K). Try to find a value K_0 for which $G(K_0)=1-\alpha$. As can be seen from Figure 3, there are two possibilities: (i) such K_0 exists, or (ii) we cannot exactly solve the equation $G(K_0)=1-\alpha$ but we can find a K_0 for which $G_{\_}(K_0)=P_{\theta_0}\{\lambda(X)<K_0\}\le 1-\alpha<G(K_0)$. In Case (i), we are done, (set $\gamma_0=0$). In Case (ii), set

$$\gamma_0=\frac{G(K_0)-(1-\alpha)}{G(K_0)-G_{\_}(K_0)}$$

Then it is an easy exercise to demonstrate that the test has size exactly α, as required.

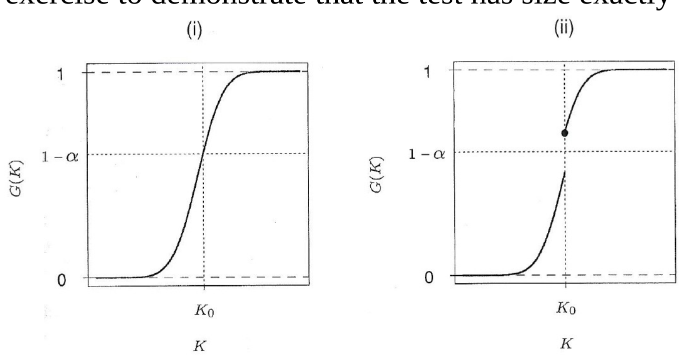

Figure 3

(c) Let ϕ_0 be the LRT defined by the constant K and function $\gamma(x)$, and suppose ϕ is another test of the same size α and the same power as ϕ_0. Define $U(x)$ as in (a). Then $U(x) \geq 0$ for all x, but because ϕ and ϕ_0 have the same size and power, $\int U(x)dx = 0$. So the function $U(x)$ is non-negative and integrates to 0: hence $U(x) = 0$ for all x, except possibly on a set, S say, of values of x, which has probability zero under both H_0 and H_1. This in turn means that, except on the set S, $\phi(x) = \phi_0(x)$ or $f_1(x) = K f_0(x)$, so that $\phi(x)$ has the form of a LRT. This established the uniqueness result, and so completes the proof of the theorem. ■

In the above lemma statement, regarding the first component, Young and Smith define ϕ_0 in terms of K and $\gamma(x)$ as follows:

$$\phi_0(x) = \begin{Bmatrix} 1 & \text{if } f_1(x) > K f_0(x) \\ \gamma(x) & \text{if } f_1(x) = K f_0(x) \\ 0 & \text{if } f_1(x) < K f_0(x) \end{Bmatrix}$$

where $K \geq 0$ and $\gamma(x)$ is an arbitrary function such that $\gamma(x) \in [0,1]$. Additionally, $\Lambda(x) = \frac{f_1(x)}{f_0(x)}$, as referred to in part (b) of the proof. That being said, this statement of the lemma provided by Young and Smith mimics the versions from Roussas, Lehmann, Dudewicz and Mishra, and Mukhopadhyay as they all present the test function in more or less the same fashion (although Young and Smith presented their definition of the test function before the lemma statement). This version of the statement is unique from the rest as it has the three components similar to Lehmann's but no preliminary statement is addressed before the three components in the theorem. This layout of the lemma stems from the approach Young and Smith took by defining several expressions beforehand. With regard to the three components, the Uniqueness component corresponds to the Necessity component from Casella and Berger's variant as both infer the same result. The proof by Young and Smith resembles their lemma statement as the proof is broken up into matching corresponding parts, which makes following through the proof straight forward. Not much needs to be said for part (a) of the proof as this same argument has been covered several times up to this point. With part (b), the argument is quite similar to those presented by Roussas and Lehmann as these three versions make use of a function that has properties of a cumulative distribution function. Roussas made use of Figure 1 to aid in the discussion on this as do Young and Smith with Figure 3. Notationally, Young and Smith make use of a subscript minus sign to indicate the left-handed limit of G, similar to how Roussas made use of $a(C-)$.

Examples

At this point, ten different versions of the Neyman-Pearson Lemma have been compared and despite some of their different statements and proofs, they all provide the same conclusion. This brings up the next question: how is the Neyman-Pearson Lemma used? This section looks at two examples for use of the Neyman-Pearson Lemma. One example considered is a typical example and is 'well behaved' in conclusion. The other example looks at potential problems that can occur with the lemma, particularly where the test function needs to be utilized in an intermediate state according to several lemma statements that involve the test function.

Exponential Example

This first example comes from Intermediate Mathematical Statistics (Beaumont, 1980) and shows how the Neyman-Pearson Lemma is typically used.

Suppose $X_{1,} \ldots, X_n$ is a random sample from the exponential distribution with parameter λ. $H_0: \lambda = \lambda_0$; $H_1: \lambda = \lambda_1 > \lambda_0$.

Since the X_i are independent,

$$\begin{aligned} f(x_{1,} \ldots, x_n | \lambda) &= \prod_{i=1}^{n} f_i(x_i | \lambda) \\ &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\ &= \lambda^n e^{-\lambda \sum x_i} \end{aligned}$$

By the Neyman-Pearson Lemma, the most powerful critical region of its size consists of those sample points which satisfy

$$\frac{\lambda_1^n e^{-\lambda_1 \sum x_i}}{\lambda_0^n e^{-\lambda_0 \sum x_i}} > k \Leftrightarrow \underbrace{\left(\frac{\lambda_1}{\lambda_0}\right)^n}_{>0} \exp\left[(\lambda_0 - \lambda_1) \sum x_i\right] > k$$

$$\Leftrightarrow \underbrace{(\lambda_0 - \lambda_1)}_{<0} \sum x_i > \ln(k) + n \ln\left(\frac{\lambda_0}{\lambda_1}\right) \Leftrightarrow \sum x_i < \frac{\ln(k) + n \ln\left(\frac{\lambda_0}{\lambda_1}\right)}{\lambda_0 - \lambda_1}$$

Thus, the Neyman-Pearson Lemma states that for all tests of size α, the most powerful test satisfies the condition that $\sum_{i=1}^{n} x_i < k^*$ where k^* is determined such that $P\left(\sum_{i=1}^{n} X_i < k^* \middle| \lambda = \lambda_0\right) = \alpha$.

Do note that since the X_i's are independent and identically distributed, their sum follows a Gamma distribution. That is, $Y = \sum X_i \sim Gamma(n, \lambda)$ which has pdf $g(y|\lambda) = \left[\lambda^n / \Gamma(n)\right] y^{n-1} e^{-\lambda y}$. Thus the size of this test is defined as

$$\int_0^{k^*} g(y | \lambda = \lambda_0)\, dy = \int_0^{k^*} \frac{\lambda_0^n}{\Gamma(n)} y^{n-1} e^{-\lambda_0 y}\, dy = \alpha \text{ for } y \in [0, \infty).$$

On a similar note, the power of this test is

$$P\left[\sum_{i=1}^{n} X_i < k^* \middle| \lambda = \lambda_1\right] = \int_0^{k^*} g(y | \lambda_1)\, dy = \int_0^{k^*} \frac{\lambda_1^n}{\Gamma(n)} y^{n-1} e^{-\lambda_1 y}\, dy.$$

It can be seen that so long as $\lambda_1 > \lambda_0$, the form of the test does not change for all $\lambda > \lambda_0$ under a fixed α value, the value of k^* is determined solely by λ_0. Thus, this test is uniformly most powerful for testing the hypotheses $H_0: \lambda = \lambda_0$ versus $H_1: \lambda = \lambda_1 > \lambda_0$. ■

This example with the exponential is a nice one because the method in finding the uniformly most powerful test is pretty straight forward. Note that this example expressed the test in terms of a sufficient statistic. This is usually the case when finding tests using the Neyman-Pearson Lemma but not always, as is the case in this second example, which actually looks at two methods at arriving to more or less the same conclusion.

Uniform Example

This second example comes from Modern Mathematical Statistics (Dudewicz and Mishra) and considers a special case where $f(\vec{x}|\theta_0)$ may be zero. It turns out that there is not a unique most powerful test for this case in general and the way randomization occurs when $f(\vec{x}|\theta_1)=k\, f(\vec{x}|\theta_0)$ can be very important.

Suppose $X_1,\dots,X_n$ are independent and identically distributed from a uniform distribution on the interval $(0,\theta)$ with θ unknown. i.e. $X_1,\dots,X_n \overset{iid}{\sim} Unif(0,\theta)$. Suppose the goal is to find the most powerful test of level $\alpha=0.05$ for testing $H_0:\theta=1$ versus $H_1:\theta=2$ in one of two ways. By observing simply the sample (denoting ϕ_1 as the critical function for this first test) or observing the sufficient statistic $Y=max(X_1,\dots,X_n)$ (denoting ϕ_2 as the critical function for this second test).

First Case: Given we have observed $X_1,\dots,X_n$, we want to find a most powerful level $\alpha=0.05$ test of the hypotheses $H_0:\theta=1$ versus $H_1:\theta=2$. By the Neyman-Pearson Lemma, ϕ_1 will be most powerful provided it satisfies

$$\phi_1(\vec{x})=\begin{cases}1 & \text{when } f(\vec{x}|\theta_1)>k\, f(\vec{x}|\theta_0)\\ 0 & \text{when } f(\vec{x}|\theta_1)<k\, f(\vec{x}|\theta_0)\end{cases}$$

where k is chosen such that $E_{\theta_0}\left[\phi_1(\vec{X})\right]=\alpha$.

Now in the case of the uniform distribution, for an arbitrary value of θ, the joint probability density function is

$$f(\vec{x}|\theta)=\begin{cases}\frac{1}{\theta^n} & 0\le x_{(1)}<x_{(n)}\le\theta\\ 0 & \text{otherwise}\end{cases}$$

where $x_{(1)}=min(x_1,\dots,x_n)$ and $x_{(n)}=max(x_1,\dots,x_n)$. This implies that for the hypothesized values of θ, we get the following joint densities:

$$f(\vec{x}|\theta=1)=\begin{cases}1 & 0\le x_{(1)}<x_{(n)}\le 1\\ 0 & \text{otherwise}\end{cases}$$

$$f(\vec{x}|\theta=2)=\begin{cases}\frac{1}{2^n} & 0\le x_{(1)}<x_{(n)}\le 2\\ 0 & \text{otherwise}\end{cases}$$

Because the values of the X_i's depend on the value of θ, the possible values of $f(\vec{x}|\theta=1)$ and $f(\vec{x}|\theta=2)$ can be divided into four cases. These are summarized in Table 2.

Case	$f(\vec{x}\vert\theta=1)$	$f(\vec{x}\vert\theta=2)$
1. $x_{(1)}<0$	0	0
2. $x_{(1)}\geq 0$ and $x_{(n)}\leq 1$	1	$1/2^n$
3. $x_{(1)}\geq 0$ and $1<x_{(n)}\leq 2$	0	$1/2^n$
4. $x_{(1)}\geq 0$ and $x_{(n)}>2$	0	0

Table 2: $f(\vec{x}|\theta=1)$ and $f(\vec{x}|\theta=2)$

From these possible values the joint probability densities can take on, we can summarize the three relationships between $f(\vec{x}|\theta=2)$ and $f(\vec{x}|\theta=1)$. These are shown in Table 3.

Case	$f(\vec{x}\vert\theta=2)>k\,f(\vec{x}\vert\theta=1)$	$f(\vec{x}\vert\theta=2)=k\,f(\vec{x}\vert\theta=1)$	$f(\vec{x}\vert\theta=2)=k\,f(\vec{x}\vert\theta=1)$
1.	Never	Always	Never
2.	If and only if $k<1/2^n$	If and only if $k=1/2^n$	If and only if $k>1/2^n$
3.	Always	Never	Never
4.	Never	Always	Never

Table 3: $f(\vec{x}|\theta=2)$ versus $k\,f(\vec{x}|\theta=1)$

Based on the results in Table 3, regardless of what value of k is chosen, ϕ_1 will always be equal to 1 whenever case 3 occurs (as $1/2^n>k\cdot 0$ for all n). As for cases 1 and 4, any chosen value of k will result in any value for ϕ_1 chosen between 0 and 1 to be acceptable (as $0=k\cdot 0$). Case 2 is the only interesting case as all three possible relationships shown in Table 3 are possible in this instance with $\phi_1(\vec{x})=1$ if $k<1/2^n$ and $\phi(\vec{x})=0$ if $k>1/2^n$. For the instance under case 2 when $k=1/2^n$, $\phi(\vec{x})$ may be chosen as needed.

Considering cases 1, 3, and 4

$$P_{\theta=1}[Case\,1,\ \text{or}\ Case2,\ \text{or}\ Case\,3]=0,$$

so even if the null was rejected based on observing an $\vec{x}$ under cases 1, 3, and 4, the tests would always be level 0. This implies that a level $\alpha=0.05$ test can only be obtained when the observed X_i fall under case 2. Unfortunately, $k>1/2^n$ does not allow for any kind of level $\alpha=0.05$ test and $k<1/2^n$ always results in a level 1 test (always reject is required). Thus it must be true that $k=1/2^n$ in order to obtain a level $\alpha=0.05$ test. On another note, for the observed X_i's under case 1, the null must be false as X_i's under case 1 are impossible under the assumption $\theta=1$. This also holds true for cases 3 and 4 as both involve impossible values for the X_i's when $\theta=1$. Thus the null will always be rejected under cases 1, 3, and 4. Case 2 is the only case where the null hypotheses is possible. Because the value of $\phi_1(\vec{x})$ can be chosen to be anything when under case 2, this value will be denoted as γ where $\gamma\in(0,1)$. All that remains is to choose γ such that $E_{\theta_0}[\phi_1(\vec{X})]=\alpha$. Now

$$E_{\theta=1}\left[\phi_1(\vec{x})\right]=\gamma\, P_{\theta=1}[case\,2]+P_{\theta=1}\left[Case\,1,\text{ or } Case2,\text{ or } Case\,3\right]=\gamma\cdot 1+0=\gamma$$

This implies choosing $\gamma=\alpha$ will satisfy the condition $E_{\theta_0}\left[\phi_1(\vec{X})\right]=\alpha$ and thus, the most powerful test for testing $H_0:\theta=1$ versus $H_1:\theta=2$ is given by

$$\phi(\vec{x})=\begin{cases}0.05 & 0\le x_{(1)}<x_{(n)}\le 1\\ 1 & \text{otherwise}\end{cases}$$

For any arbitrary value of θ such that $\theta>0$, the power is computed as follows. For $0<\theta<1$:

$$\begin{aligned}E_\theta\left[\phi_1(\vec{X})\right] &= 0.05\,P_\theta\left[0\le X_{(1)}<X_{(n)}\le 1\right]+1\,P_\theta\left[X_{(1)}<0\ \text{ or }\ X_{(n)}>1\right]\\ &=(0.05)(1)+(1)(0)=0.05\end{aligned}$$

For $\theta>1$:

$$\begin{aligned}E_\theta\left[\phi_1(\vec{X})\right] &= 0.05\,P_\theta\left[0\le X_{(1)}<X_{(n)}\le 1\right]+1\,P_\theta\left[X_{(1)}<0\ \text{ or }\ X_{(n)}>1\right]\\ &=0.05\left(\frac{1}{\theta^n}\right)+1\left(1-\left(\frac{1}{\theta^n}\right)\right)=1-\frac{0.95}{\theta^n}\end{aligned}$$

Thus, for $H_0:\theta=1$ versus $H_1:\theta=2$, we have

$$E_{\theta=1}\left[\phi_1(\vec{X})\right]=0.05 \text{ and } E_{\theta=2}\left[\phi_1(\vec{X})\right]=1-\frac{0.95}{2^n}.$$

The plot for the power of ϕ_1 is shown in Figure 4 as the solid line assuming $n=3$. It also shows that the power of this test under the alternative hypotheses is close to 0.9.

Finally, by the Neyman-Pearson Lemma, for any other test ψ such that $E_{\theta=1}\left[\psi(\vec{X})\right]=\alpha$, its power must satisfy $E_{\theta=2}\left[\psi(\vec{X})\right]\le E_{\theta=2}\left[\phi_1(\vec{X})\right]$.

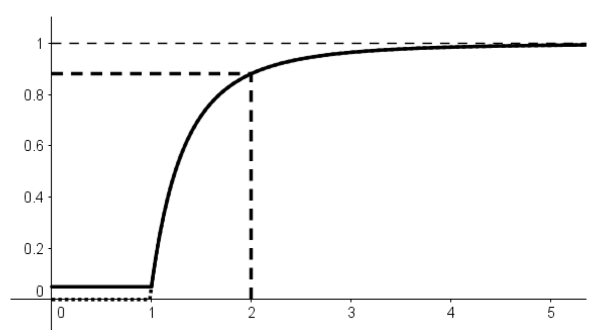

Figure 4

Second Case: Now suppose $Y = max(X_{1,} \ldots, X_n)$ is observed and we want to find the most powerful level $\alpha = 0.05$ test for $H_0: \theta = 1$ versus $H_1: \theta = 2$. Again, the Neyman-Pearson Lemma will be used. Note that $X_1, \ldots, X_n$ are independent and identically distributed. Thus

$$\begin{aligned} P_\theta[Y < y] &= P_\theta[max(X_1, \ldots, X_n) < y] \\ &= P_\theta[X_1 < y, \ldots, X_n < y] \\ &\overset{iid}{=} (P_\theta[X_1 < y])^n \\ &= \begin{cases} 0 & y < 0 \\ y^n/\theta^n & 0 \le y \le \theta \\ 1 & y > \theta \end{cases} \end{aligned}$$

Here, the density for Y is

$$g(y|\theta) = \begin{cases} n\dfrac{y^{n-1}}{\theta^n} & 0 \le y \le \theta \\ 0 & \text{otherwise} \end{cases}$$

Thus

$$g(y|\theta=1) = \begin{cases} ny^{n-1} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$g(y|\theta=2) = \begin{cases} n\dfrac{y^{n-1}}{2^n} & 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

A similar result to that shown in Table 2 is provided in Table 4.

Case	$g(y\|\theta=1)$	$g(y\|\theta=2)$
1. $y < 0$	0	0
2. $0 \le y \le 1$	ny^{n-1}	$n\dfrac{y^{n-1}}{2^n}$
3. $1 < y \le 2$	0	$n\dfrac{y^{n-1}}{2^n}$
4. $y > 2$	0	0

Table 4: $g(y|\theta=1)$ and $g(y|\theta=2)$

Same as before when we were working with the raw sample. $\phi_2(y)$ can be chosen as desired when $y < 0$ or $y > 2$ regardless of the value of k. Also, the value of $\phi_2(y)$ must be set to 1 if $1 < y \le 2$ again regardless of the value of k. When $0 \le y \le 1$, $\phi_2(y)$ is set to 1 only when

$$n\frac{y^{n-1}}{2^n} > kny^{n-1} \Leftrightarrow k < \frac{1}{2^n}$$

As before, $\phi_2(y)$ can be a level $\alpha=0.05$ test only when $k=1/2^n$. In this instance, however, it seems intuitive that large values of Y are more likely to support $H_1:\theta=2$ being true than small values of Y when working in case 2. So for $0\leq y\leq 1$, rather than setting $\phi_2(y)=0.05$, $\phi_2(y)$ will be defined as

$$\phi_2(y)=\begin{cases} 1 & y>c \\ 0 & y<c \end{cases}$$

Where c is defined such that $\phi_2(y)$ is level $\alpha=0.05$. Thus

$$\begin{aligned} 0.05 &= E_{\theta=1}[\phi_2(Y)] = P_{\theta=1}[Y>c] \\ &= 1-P_{\theta=1}[Y\leq c] = 1-c^n \end{aligned}$$

This implies $c=0.95^{1/n}$. Hence, the function $E_\theta[\phi_2(Y)]$ is

$$\begin{aligned} E_\theta[\phi_2(Y)] &= P_\theta[Y>0.95^{1/n}] \\ &= \begin{cases} 0 & \theta<0.95^{1/n} \\ 1-\dfrac{0.95}{\theta^n} & \theta>0.95^{1/n} \end{cases} \end{aligned}$$

The graph is also shown in Figure 4. It is the dotted line toward the bottom left and merges with the power function of ϕ_1 for θ close to 1 and larger than 1. ■

This second example provides a more unusual way of finding the most powerful test as the uniform differs from most other common distributions in the sense that the set of all possible sample points inherently depends on the parameters. In the case of the uniform, a most powerful test was based around the boundary point in the test function and the Neyman-Pearson Lemma doesn't provide any leeway on what to do here. Typically, the boundary (where $f(x|\theta_1)=k\,f(x|\theta_0)$) has probability zero under both the null and alternative hypotheses under the continuous setting (which all proofs above assumed) and thus does not affect the choice of k based on the lemma. However, under a discrete setting, these boundaries often occur with probability greater than zero. It is when this is possible that the test function takes on the form shown in some of the above proofs. That is,

$$\phi(x)=\begin{cases} 1 & f(x|\theta_1)>k\,f(x|\theta_0) \\ \gamma & f(x|\theta_1)=k\,f(x|\theta_0) \\ 0 & f(x|\theta_1)<k\,f(x|\theta_0) \end{cases}$$

If it is the case that the boundary point has a non-zero probability of occurring, then a value of k needs to be found such that

$$P_{\theta_0}[f(X|\theta)>k\,f(X|\theta_0)]<\alpha<P_{\theta_0}[f(X|\theta)\geq k\,f(X|\theta_0)]$$

where the difference between the two sides is the probability of the boundary. It is worth noting that if non of the boundary is put into the rejection region, the test will be less than level α. On the contrary, putting all of the boundary into the rejection region will result in a test larger than level α. Thus a portion of the boundary must be included to make the test level α. Specifically, γ defines this portion which is defined as

$$\gamma = \frac{\alpha - P_{\theta_0}[f(X|\theta_1) > k f(X|\theta_0)]}{P_{\theta_0}[f(X|\theta_1) = k f(X|\theta_0)]}$$

which is the same as

$$\gamma = \frac{\alpha - a(C_0)}{a(C_0 -) - a(C_0)}$$

from Roussas' proof of Neyman-Pearson.

The Lemma Dilemma

As seen from the above ten Neyman-Pearson results, the majority of them are referred to as a lemma with Roussas and Lehmann going as far to specify the lemma as fundamental. On the other end of the spectrum, Hogg, McKean, Craig, and Young and Smith call the result simply a theorem. Based on these differences, it seems not all statisticians view the result under the same filter. Placing some clarification on the matter, below are a few definitions for the term 'lemma'. From the Wolfram Alpha website, we get the following definition:

> Lemma: A subsidiary proposition that is assumed to be true in order to prove another proposition.

Closely related to Wolfram Alpha is Wolfram MathWorld which provides a definition for lemma as:

> Lemma: A short theorem used in proving a larger theorem.

A Third definition for lemma comes from the website Division By Zero:

> Lemma: A minor result whose sole purpose is to help in proving a theorem. It is a stepping stone on the path to proving a theorem. Very occasionally, lemmas can take on a life of their own.

All the above definitions imply the Neyman-Pearson Lemma is a minor result and would otherwise only be in existence to provide structure to some other theorem. Yet, any student who has taken a mathematical statistics course should know that the Neyman-Pearson result is certainly not a minor result. As indicated by the third definition of lemma by Division By Zero, lemmas can occasionally take on a life of their own. This gets at the idea that a lemma may begin as a minor result but progression in the field points to the result being more of a mile stone rather than a stepping one. Referring back to the paragraph before the lemma definitions, some authors called the Neyman-Pearson result a theorem. Below is a definition for theorem from Division By Zero:

> Theorem: A mathematical statement that is proved using rigorous mathematical reasoning. In a mathematical paper, the term theorem is often reserved for the most important result.

The second statement in the definition for theorem certainly matches up with the Neyman-Pearson Lemma as it is a major result for hypothesis testing. So based on the above definitions, it seems it would be more appropriate to refer to the Neyman-Pearson result as a theorem rather than a lemma. So then why do some many people continue to refer to it as a lemma? In hopes of determining this, the original statement of the Neyman-Pearson 'Lemma', which dates back to 1933, was found.

In their paper On the Problem of the most Efficient Tests of Statistical Hypotheses, Neyman and Pearson discuss methods for finding the best critical regions for both simple hypotheses and composite hypotheses. They begin by providing an outline of the general theory which discusses the theory for both simple and composite hypotheses. They spend roughly 5 pages doing so. On an interesting side note, in this section, they make mention that working in the case of discrete variables is analogous to working in the case with continuous variables. They then state that only the continuous case will be considered. As previously mentioned, all the above mentioned proofs for the result assumed a continuous case but nearly all proofs mention that the discrete case is analogous. Perhaps the lack of proofs which assume the discrete case stems from following the footsteps of Neyman and Pearson. In their general overview, Neyman and Pearson define the best critical region in terms of Type I and Type II errors. In short, for a given α level $(0<\alpha<1)$ and a set of all critical regions that satisfy the condition $P(reject\ H_0 | H_0\ is\ true)=\alpha$, the best critical region is defined as the region that has the smallest probability of failing to reject a false null hypothesis. Despite having found the original result by Neyman and Pearson, no mention is given as to where this results may have been used as a lemma. So in conclusion, at least for now, the origins of why Neyman and Pearson's result acquired the name of lemma remains unknown.

References

Bain, Lee J., and Max Engelhardt. *Introduction to Probability and Mathematical Statistics*. 2nd ed. Boston: PWS-KENT Pub., 1992. Print.

Beaumont, G. P. *Intermediate Mathematical Statistics*. London: Chapman and Hall, 1980. Print.

Bickel, Peter J., and Kjell A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day, 1977. Print.

Casella, George, and Roger L. Berger. *Statistical Inference.* 2nd ed. Australia: Duxbury, 2002. Print.

Dudewicz, Edward J., and Satya Narain. Mishra. *Modern Mathematical Statistics*. New York: Wiley, 1988. Print.

Hogg, Robert V., Joseph W. McKean, and Allen T. Craig. *Introduction to Mathematical Statistics*. 6th ed. Upper Saddle River, NJ: Pearson Education, 2005. Print.

Lehmann, E. L. *Testing Statistical Hypotheses*. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced & Software, 1991. Print.

Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes. *Introduction to the Theory of Statistics*. New York: McGraw-Hill, 1973. Print.

Mukhopadhyay, Nitis. *Probability and Statistical Inference*. Vol. 162. New York: Marcel Dekker, 2000. Print.

Neyman, J., and E. S. Pearson. “On the Problem of the Most Efficient Tests of Statistical Hypotheses”. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 231 (1933): 289 – 337.* http://www.jstor.org/stable/91247

Richeson, Dave. "What Is the Difference between a Theorem, a Lemma, and a Corollary?" Division by Zero. 2008. Web. 08 Apr. 2016.

Roussas, George G. *A Course in Mathematical Statistics*. 2nd ed. San Diego, CA: Academic, 1997. Print.

Weisstein, Eric. "Lemma." -- from Wolfram MathWorld. Web. 08 Apr. 2016.

"Wolfram|Alpha: Computational Knowledge Engine." Wolfram|Alpha: Computational Knowledge Engine. Web. 08 Apr. 2016.

Young. *Essentials of Statistical Inference: G.A. Young, R.L. Smith (Cambridge Series in Statistical and Probabilistic Mathematics ;)*. Cambridge UP, 2005. Print.