Neyman-Pearson Lemma

For two parameter values θ_0 and θ_1 consider the likelihood ratio

$$\mathrm{LR}(\mathbf{x}) = \frac{f(\mathbf{x};\theta_0)}{f(\mathbf{x};\theta_1)} \tag{1}$$

The rejection region based on the likelihood ratio (1) is of the form

$$R = \{\mathbf{x} : \mathrm{LR}(\mathbf{x}) < c\} \tag{2}$$

We use the notation $P(\mathbf{X} \in R|\theta)$ in place of the notation $P_\theta(\mathbf{X} \in R)$. Note that it is not intended to mean conditional probability, so please do not misread it as such. However the notation makes many such statements easier to read so we do not have to pay attention to the subscript of a subscript, provided one has a reasonable interpretation of the notation.

Recall for a decision rule, say $d(\mathbf{x})$ which is a statistic

- Size
$$\alpha = P_{\theta_0}\left(d(\mathbf{X}) \in R\right)$$
is the probability of rejecting H_0 when the null hypothesis is true, the is probability of type I error
- Power
$$1 - \beta = P_{\theta_1}\left(d(\mathbf{X}) \in R\right)$$
is the probability of rejecting H_0 when the alternative is true, the is 1 minus the probability of type II error, where the probability of type II error is
$$\beta = P_{\theta_1}\left(d(\mathbf{X}) \in R^c\right)$$

Decreasing the size also decreases the power, and increasing the power increases the size. It is best to keep the size small and to increase the power or equivalently keep the probability of type II error small. In order to find a *best* test procedure Neyman and Pearson tried to find a test with highest power but constrained to have size less than or equal to some specified value, say α. The Theorem below, called the Neyman-Pearson Lemma, does this in the case of a simple null hypothesis versus simple alternative. The conclusion is that the likelihood ratio test or decision rule is the best.

Notice that we can also match up a decision rule with an indicator function of $\mathbf{x}$ being in the rejection region. Thus a decision rule corresponds to an indicator of a rejection region. In the Theorem below this is even slightly generalized so that one really just needs $d(\mathbf{x})$ to take values between 0 and 1 inclusive.

Neyman-Pearson Lemma : Consider the simple null hypothesis $H_0 : \theta = \theta_0$ versus the simple alternative $H_A : \theta = \theta_1$. Consider the rejection region R given by (2) of size $\alpha = P(\mathbf{X} \in R|\theta_0)$. Let R^* be any other rejection region of size $\alpha^* = P(\mathbf{X} \in R^*|\theta_0) \leq \alpha$. Then the likelihood ratio test is more powerful than this other test, that is

$$P(\mathbf{X} \in R^*|\theta_1) \leq P(\mathbf{x} \in R|\theta_1) \ .$$

Proof:

This is given in the case of X having a density.

For a set A let I_A be the indicator function of this set. Thus we can make a 1 to 1 correspondence between a rejection region and its indicator function

$$\begin{aligned} R &\leftrightarrow \mathrm{I}_R \\ R^* &\leftrightarrow \mathrm{I}_{R^*} \end{aligned}$$

Next notice that

$$\begin{aligned} cf(\mathbf{x};\theta_1) - f(\mathbf{x};\theta_0) > 0 \quad &\text{if} \quad \mathrm{I}_R(\mathbf{x}) = 1 \\ cf(\mathbf{x};\theta_1) - f(\mathbf{x};\theta_0) \leq 0 \quad &\text{if} \quad \mathrm{I}_R(\mathbf{x}) = 0 \end{aligned}$$

Thus for every possible value of $\mathbf{x}$ we obtain

$$\mathrm{I}_{R^*}(\mathbf{x})\,(cf(\mathbf{x};\theta_1) - f(\mathbf{x};\theta_0)) \leq \mathrm{I}_R(\mathbf{x})\,(cf(\mathbf{x};\theta_1) - f(\mathbf{x};\theta_0)) \ .$$

Remark: If we integrate the left hand side with respect to (w.r.t.) $\mathbf{x}$ we get

$$\int \mathrm{I}_{R^*}(\mathbf{x})\,(cf(\mathbf{x};\theta_1) - f(\mathbf{x};\theta_0))\,d\mathbf{x} = cP(R^*|\theta_1) - P(R^*|\theta_0)$$

Thus we get c times the power of R^* minus the size of R^*. The other side will give a similar term.

Integrate both sides w.r.t. $\mathbf{x}$. This gives

$$cP(R^*|\theta_1) - P(R^*|\theta_0) \leq cP(R|\theta_1) - P(R|\theta_0)$$

Rearranging gives

$$P(R|\theta_0) - P(R^*|\theta_0) \leq c\,\{P(R|\theta_1) - P(R^*|\theta_1)\}$$

Thus if R^* is a test (or rejection region) of size $\leq \alpha = P(R|\theta_0)$, the LHS is ≥ 0, and hence the same is true for the RHS (recall $c > 0$) giving

$$P(R|\theta_1) - P(R^*|\theta_1) \geq 0 \ .$$

This later piece says that the power with rejection region R is greater than or equal to the power with rejection region R^*. Thus the test of hypothesis, of a simple null versus simple alternative hypothesis, based on the likelihood ratio (1) is more powerful than any other test of the same or smaller size.