Neyman-Pearson Theory of Testing

$X \sim P_\theta$, $\theta \in \Theta$. $\mathcal{X}$ = sample space = set of all values that X can take. It is of interest to test

$$H_0 : \theta \in \Theta_0 \text{ versus } H_1 : \theta \in \Theta_1,$$

where $\Theta_i \subset \Theta$ and $\Theta_0 \cap \Theta_1 = \phi$.
Simple hypothesis: $\Theta_0 = \{\theta_0\}$, i.e., $H_0 : \theta = \theta_0$.
Composite hypothesis: $\Theta_0 = (-\infty, \theta_0]$, i.e., $H_0 : \theta \leq \theta_0$.

Nonrandomized test. Find a subset S of $\mathcal{X}$, and reject H_0 if the observed value $x \in S$. $S \subset \mathcal{X}$ is called the critical region or the rejection region of the test. One defines a test function for nonrandomized tests ϕ as

$$\phi(x) = \begin{cases} 1 & \text{if } x \in S; \\ 0 & \text{if } x \in S^c. \end{cases}$$

Note that $\phi(x)$ is also the probability of rejecting H_0 upon observing x.

For a level α test, one must have,

$$\sup_{\theta \in \Theta_0} P_\theta(X \in S) \leq \alpha.$$

Note that, if $\Theta_0 = \{\theta_0\}$, then we need $P_{\theta_0}(X \in S) \leq \alpha$.
For nonrandomized tests, it may happen that $\sup_{\theta \in \Theta_0} P_\theta(X \in S) < \alpha$.

Power(θ) = Power of test = $P_\theta(X \in S) = E_\theta[\phi(X)]$ for $\theta \in \Theta_1$ is the power function of the test associated with S.

Randomized test. Any ϕ such that $0 \leq \phi(x) \leq 1$ for all $x \in \mathcal{X}$, and at any x, $\phi(x)$ is the probability of rejecting H_0 if x is observed. Nonrandomized tests form a subset of randomized tests. The power function for randomized tests is given by

$$P_\theta(\text{Reject} H_0) = E_\theta\left[P(\text{Reject} H_0 | X)\right] = E_\theta \phi(X) = \int_{\mathcal{X}} \phi(x)\, dP_\theta(x).$$

Problem. Find ϕ such that $E_\theta \phi$ is maximized when $\theta \in \Theta_1$ and subject to $\sup_{\theta \in \Theta_0} E_\theta \phi(X) \leq \alpha$. Such a test, if it exists, is called a Uniformly Most Powerful (UMP) test.

Consider the two kinds of the errors defined earlier. It turns out that in general if one tries to reduce one error probability the other error probability goes up, so one cannot reduce both simultaneously. Because probability of error of first kind is more important, one first makes it small (by fixing it at a small value such as 0.01 or 0.05). Among all tests satisfying this, one then tries to minimize the probability of committing error of second kind or equivalently, to maximize the power uniformly for all θ in H_1.

N-P Lemma. Suppose $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$. i.e., simple null versus simple alternative. Also, let p_{θ_0} and p_{θ_1} denote the respective densities (pdf or pmf). Then
(a) there exists a test ϕ and a constant $k \geq 0$ such that

$$E_{\theta_0}\phi(X) = \alpha \tag{1}$$

and

$$\phi(x) = \begin{cases} 1 & \text{when } p_{\theta_1}(x) > kp_{\theta_0}(x); \\ 0 & \text{when } p_{\theta_1}(x) < kp_{\theta_0}(x). \end{cases} \tag{2}$$

(b) If a test satisfies (1) and (2) for some k, then it is most powerful for testing $H_0 : P_\theta = P_{\theta_0}$ versus $H_1 : P_\theta = P_{\theta_1}$ at level α.
(c) If ϕ is a most powerful test at level α for testing $H_0 : P_\theta = P_{\theta_0}$ versus $H_1 : P_\theta = P_{\theta_1}$, then it satisfies (1) and (2) for some k. (This answers whether MP test can have some other form.)

Remark. The most powerful test is of the form: reject H_0 if $p_{\theta_1}(x)/p_{\theta_0}(x) > k$. i.e., when the likelihood ratio exceeds a threshold. This is intuitively meaningful because, on the one hand we want $\int_S p_{\theta_0}(x)\,dx \leq \alpha$ and on the other $\int_S p_{\theta_1}(x)\,dx =$ maximum. To achieve this one must put all x that give very large values of $p_{\theta_1}(x)/p_{\theta_0}(x)$ in S.

Proof. Let $0 < \alpha < 1$ and define

$$\alpha(c) = P_{\theta_0}\left(p_{\theta_1}(X) > cp_{\theta_0}(X)\right) = P_{\theta_0}\left(\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} > c\right).$$

(The second equality is because $P_{\theta_0}(p_{\theta_0}(X) = 0) = 0$.) Then,

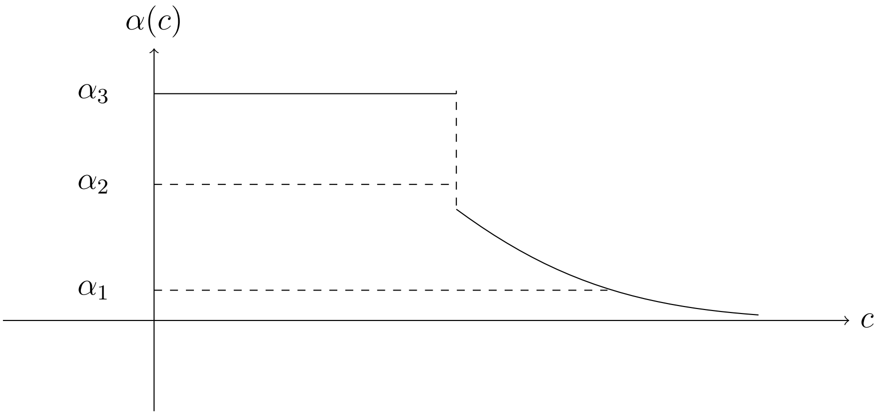
$$1 - \alpha(c) = P_{\theta_0}(r(X) \leq c), \text{ where } r(X) = \tfrac{p_{\theta_1}(X)}{p_{\theta_0}(X)}.$$

Therefore, $1 - \alpha(c)$ is the cdf of $r(X)$. Thus, $1 - \alpha(c)$ is nondecreasing and right continuous. Therefore, $\alpha(c)$ is nonincreasing and right continuous. i.e., $1 - \alpha(-\infty) = 0$, $1 - \alpha(\infty) = 1$, $1 - \alpha(c) \nearrow$ and $1 - \alpha(c+) = 1 - \alpha(c)$. Therefore, $\alpha(-\infty) = 1$, $\alpha(\infty) = 0$, $\alpha(c) \searrow$ and $\alpha(c+) = \alpha(c)$. We would like to find c_0, if possible, such that $\alpha(c_0) = \alpha$. Note that

$$\alpha(c-) - \alpha(c) = P_{\theta_0}(r(X) = c) = P_{\theta_0}\left(\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} = c\right).$$

For fixed $0 < \alpha < 1$, find c_0 such that $\alpha(c_0) \leq \alpha \leq \alpha(c_0-)$ and define

$$\phi(x) = \begin{cases} 1 & \text{if } p_{\theta_1}(x) > c_0 p_{\theta_0}(x); \\ 0 & \text{if } p_{\theta_1}(x) < c_0 p_{\theta_0}(x); \\ \frac{\alpha - \alpha(c_0)}{\alpha(c_0-) - \alpha(c_0)} & \text{if } p_{\theta_1}(x) = c_0 p_{\theta_0}(x). \end{cases}$$



If $\alpha(c_0) = \alpha(c_0-)$, (i.e., α_1 or α_3 in the figure) then

$$P_{\theta_0}\left(\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} = c_0\right) = 0,$$

i.e., $P_{\theta_0}(p_{\theta_1}(X) = c_0 p_{\theta_0}(X)) = 0$. Therefore, $\phi(.)$ is defined a.e. w.r.t. P_{θ_0}.

Note that

$$\begin{aligned} E_{\theta_0}\phi(X) &= P_{\theta_0}\left(\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} > c_0\right) + \frac{\alpha - \alpha(c_0)}{\alpha(c_0-) - \alpha(c_0)} P_{\theta_0}\left(\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} = c_0\right) \\ &= \alpha(c_0) + \frac{\alpha - \alpha(c_0)}{\alpha(c_0-) - \alpha(c_0)} (\alpha(c_0-) - \alpha(c_0)) = \alpha \end{aligned}$$

Choose $k = c_0$ to satify (1) and (2).

(b) Suppose ϕ is a test which satisfies (1) and (2). We want to show ϕ is MP at level α. Let ϕ^* be any other test such that $E_{\theta_0}\phi^*(X) \leq \alpha$. (Note that $0 \leq \phi(x) \leq 1$ and $0 \leq \phi^*(x) \leq 1$.) Let

$$S^+ = \{x : \phi(x) - \phi^*(x) > 0\} \text{ and } S^- = \{x : \phi(x) - \phi^*(x) < 0\}.$$

Then, if $x \in S^+$, $\phi(x) > \phi^*(x) \geq 0$. Therefore,

$$\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} \geq k, \text{ or } p_{\theta_1}(x) \geq k p_{\theta_0}(x).$$

If $x \in S^-$, $\phi(x) < \phi^*(x) \le 1$, so that $p_{\theta_1}(x) \le kp_{\theta_0}(x)$. Therefore,

$$\begin{aligned}
&\int_{\mathcal{X}} (\phi(x) - \phi^*(x))(p_{\theta_1}(x) - kp_{\theta_0}(x))\,dx \\
&\quad = \int_{S^+ \cup S^-} (\phi(x) - \phi^*(x))(p_{\theta_1}(x) - kp_{\theta_0}(x))\,dx \ge 0.
\end{aligned}$$

Hence,

$$\begin{aligned}
\int_{\mathcal{X}} (\phi(x) - \phi^*(x))p_{\theta_1}(x)\,dx &\ge k \int_{\mathcal{X}} (\phi(x) - \phi^*(x))p_{\theta_0}(x)\,dx \\
&= k \left[\int_{\mathcal{X}} \phi(x)p_{\theta_0}(x)\,dx - \int_{\mathcal{X}} \phi^*(x)p_{\theta_0}(x)\,dx \right] \\
&= k\left[\alpha - E_{\theta_0}\phi^*(X)\right] \ge 0.
\end{aligned}$$

i.e., $E_{\theta_1}\phi(X) \ge E_{\theta_1}\phi^*(X)$.

(c) Suppose ϕ^* is the most powerful test and ϕ satisfies (1) and (2). Let

$$S = \left(S^+ \cup S^-\right) \cap \{x : p_{\theta_1}(x) \neq kp_{\theta_0}(x)\}.$$

Suppose S has positive probability (or positive Lebesgue measure in the continuous case). Then, since $(\phi(x) - \phi^*(x))(p_{\theta_1}(x) - kp_{\theta_0}(x)) > 0$ on S,

$$\begin{aligned}
&\int_{S^+ \cup S^-} (\phi(x) - \phi^*(x))(p_{\theta_1}(x) - kp_{\theta_0}(x))\,dx \\
&\quad = \int_{S} (\phi(x) - \phi^*(x))(p_{\theta_1}(x) - kp_{\theta_0}(x))\,dx > 0.
\end{aligned}$$

Therefore,

$$\int_{\mathcal{X}} (\phi(x) - \phi^*(x))p_{\theta_1}(x)\,dx > k \int_{\mathcal{X}} (\phi(x) - \phi^*(x))p_{\theta_0}(x)\,dx \ge 0,$$

and hence,

$$\int_{\mathcal{X}} \phi(x)p_{\theta_1}(x)\,dx > \int_{\mathcal{X}} \phi^*(x)p_{\theta_1}(x)\,dx.$$

i.e., $E_{\theta_1}\phi(X) > E_{\theta_1}\phi^*(X)$, which contradicts the assumption that ϕ^* is MP. Therefore, S must have probability 0, which implies that ϕ and ϕ^* can differ only on the set $\{x : p_{\theta_1}(x) = kp_{\theta_0}(x)\}$. (So, the MP tests may differ by picking different subsets of this set to satisfy the level condition.)