Session 9

9.1

Recall...

The Neyman-Pearson Lemma

Neyman-Pearson Lemma: Let $\Theta = \{\underline{\theta}_0, \underline{\theta}_1\}$, and let $F_{\underline{\theta}_0}(\underline{x})$ be the cdf of the random vector $\underline{X}$ under hypothesis H_0 and $F_{\underline{\theta}_1}(\underline{x})$ be its cdf under hypothesis H_1. Assume that the cdfs $F_{\underline{\theta}_i}(\underline{x})$ have corresponding pdfs or pmfs $f_{\underline{\theta}_i}(\underline{x})$, $i = 0, 1$. Then a test of the form

$$\phi(\underline{x}) = \begin{cases} 1, & \text{for } f_{\underline{\theta}_1}(\underline{x}) > k f_{\underline{\theta}_0}(\underline{x}), \\ \gamma, & \text{for } f_{\underline{\theta}_1}(\underline{x}) = k f_{\underline{\theta}_0}(\underline{x}), \\ 0, & \text{for } f_{\underline{\theta}_1}(\underline{x}) < k f_{\underline{\theta}_0}(\underline{x}), \end{cases}$$

for some $k \geq 0$ and some $0 \leq \gamma \leq 1$ is the most powerful test of size α for testing hypothesis H_0: $\underline{\theta} = \underline{\theta}_0$ versus H_1: $\underline{\theta} = \underline{\theta}_1$.

Recall... 9.2

Interpretation of γ

$$\phi(\underline{x}) = \begin{cases} 1, & \text{for } f_{\underline{\theta}_1}(\underline{x}) > k f_{\underline{\theta}_0}(\underline{x}), \\ \gamma, & \text{for } f_{\underline{\theta}_1}(\underline{x}) = k f_{\underline{\theta}_0}(\underline{x}), \\ 0, & \text{for } f_{\underline{\theta}_1}(\underline{x}) < k f_{\underline{\theta}_0}(\underline{x}), \end{cases}$$

When $\phi(\underline{X}) = \gamma$ we randomly select between H_1 and H_0 such that

$$P(\{\text{Decide } H_1\}|\{\phi(\underline{X}) = \gamma\}) = \gamma$$
$$P(\{\text{Decide } H_0\}|\{\phi(\underline{X}) = \gamma\}) = 1 - \gamma$$

(*i.e.*, flip a biased coin having probability γ of comming up "Heads". If "Heads" occurs, decide H_1. If "Tails" occurs, decide H_0.)

9.3

Proof of the Neyman-Pearson Lemma:

We must show there does not exist a test $\phi'(\underline{x})$ with size $\alpha' \le \alpha$ having power $\beta' > \beta$.

Let $\phi'(\underline{x})$ be a test such that $0 \le \phi'(\underline{x}) \le 1$, $\forall \underline{x} \in \mathbb{R}^N$ having size $\alpha' \le \alpha$.

Note that because of the way $\phi(\underline{x})$ is defined

$$\underbrace{[\phi(\underline{x}) - \phi'(\underline{x})]}_{\ge 0} \underbrace{[f_{\theta_1}(\underline{x}) - k f_{\underline{\theta}_0}(\underline{x})]}_{+} \ge 0, \quad \forall \underline{x} \in \mathbb{R}^n \quad \ge 0$$

≤ 0 $\quad$ $-$ $\quad$ ≥ 0

$(\gamma - \phi'(x)) \cdot 0 = 0$

Thus it follows that

9.4

$$\int_{\mathbb{R}^N} \underbrace{[\phi(\underline{x}) - \phi'(\underline{x})][f_{\underline{\theta}_1}(\underline{x}) - k f_{\underline{\theta}_0}(\underline{x})]}_{\geq 0} \, d\underline{x} \geq 0$$

Expanding the integrand and integrating term by term yields

$$\int_{\mathbb{R}^N} \phi(\underline{x}) f_{\underline{\theta}_1}(\underline{x})\, dx - \int_{\mathbb{R}^N} \phi'(\underline{x}) f_{\underline{\theta}_1}(\underline{x})\, d\underline{x}$$
$$- k \int_{\mathbb{R}^N} \phi(\underline{x}) f_{\underline{\theta}_0}(\underline{x})\, dx + k \int_{\mathbb{R}^N} \phi'(\underline{x}) f_{\underline{\theta}_0}(x)\, dx \geq 0$$

or equivalently

9.5

$$\underbrace{\int_{\mathbb{R}^N} \phi(\underline{x}) f_{\underline{\theta}_1}(\underline{x})\, d\underline{x}}_{\beta} - \underbrace{\int_{\mathbb{R}^N} \phi'(\underline{x}) f_{\underline{\theta}_1}(\underline{x})\, d\underline{x}}_{\beta'}$$
$$\geq k \underbrace{\int_{\mathbb{R}^N} \phi(\underline{x}) f_{\underline{\theta}_0}(\underline{x})\, d\underline{x}}_{\alpha} - k \underbrace{\int_{\mathbb{R}^N} \phi'(\underline{x}) f_{\underline{\theta}_0}(\underline{x})\, d\underline{x}}_{\alpha'}$$

which simplifies to

$$(\beta - \beta') \geq \underbrace{k}_{\geq 0} (\alpha - \alpha')$$

$$\Rightarrow \quad \beta \geq \beta'$$

because $k \geq 0$ and $\alpha - \alpha' \geq 0$ by hypothesis. ■

Choosing the Threshold for the Neyman-Pearson Test

To choose a threshold k and parameter γ to produce a N-P test of the form

$$\phi(\underline{x}) = \begin{cases} 1, & \text{for } f_{\underline{\theta}_1}(\underline{x}) > k f_{\underline{\theta}_0}(\underline{x}), \\ \gamma, & \text{for } f_{\underline{\theta}_1}(\underline{x}) = k f_{\underline{\theta}_0}(\underline{x}), \\ 0, & \text{for } f_{\underline{\theta}_1}(\underline{x}) < k f_{\underline{\theta}_0}(\underline{x}), \end{cases}$$

with the desired size α, we note that

$P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) \geq k f_{\underline{\theta}_0}(\underline{X})\})$

$$\begin{aligned} \alpha &= \mathrm{E}_{\underline{\theta}_0}[\phi(\underline{X})] \\ &= P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) > k f_{\underline{\theta}_0}(\underline{X})\}) + \gamma P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) = k f_{\underline{\theta}_0}(\underline{X})\}) \\ &= 1 - P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) \leq k f_{\underline{\theta}_0}(\underline{X})\}) + \gamma P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) = k f_{\underline{\theta}_0}(\underline{X})\}) \end{aligned}$$

$$\alpha = 1 - P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) \leq k f_{\underline{\theta}_0}(\underline{X})\}) + \gamma P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) = k f_{\underline{\theta}_0}(\underline{X})\}).$$

This term can be made equal to zero (e.g., $\gamma = 0$)

If there exists a threshold k_0 such that

$$P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) \leq k_0 f_{\underline{\theta}_0}(\underline{X})\}) = 1 - \alpha$$

then we can take

$$k = k_0$$

$$\gamma = 0$$

and achieve a test of size α.

$$P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) < k_0 f_{\underline{\theta}_0}(\underline{X})\}) < 1-\alpha < P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) \le k_0 f_{\underline{\theta}_0}(\underline{X})\})$$

Inclusion of k_0 results in strict bracketing of $1-\alpha$

This can only occur when

$$P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) = k_0 f_{\underline{\theta}_0}(\underline{X})\}) \neq 0.$$

In this case, we select $k = k_0$ such that the bracketing occurs and then solve for γ to achieve a size α test.

The resulting value of γ is

$$\gamma = \frac{P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) \le k_0 f_{\underline{\theta}_0}(\underline{X})\}) - (1-\alpha)}{P_{\underline{\theta}_0}(\{f_{\underline{\theta}_1}(\underline{X}) = k_0 f_{\underline{\theta}_0}(\underline{X})\})}.$$

Recall ... The Neyman-Pearson Lemma

Neyman-Pearson Lemma: Let $\Theta = \{\underline{\theta}_0, \underline{\theta}_1\}$, and let $F_{\underline{\theta}_0}(\underline{x})$ be the cdf of the random vector $\underline{X}$ under hypothesis H_0 and $F_{\underline{\theta}_1}(\underline{x})$ be its cdf under hypothesis H_1. Assume that the cdfs $F_{\underline{\theta}_i}(\underline{x})$ have corresponding pdfs or pmfs $f_{\underline{\theta}_i}(\underline{x})$, $i = 0, 1$. Then a test of the form

$$\phi(\underline{x}) = \begin{cases} 1, & \text{for } f_{\underline{\theta}_1}(\underline{x}) > k f_{\underline{\theta}_0}(\underline{x}), \\ \gamma, & \text{for } f_{\underline{\theta}_1}(\underline{x}) = k f_{\underline{\theta}_0}(\underline{x}), \\ 0, & \text{for } f_{\underline{\theta}_1}(\underline{x}) < k f_{\underline{\theta}_0}(\underline{x}), \end{cases}$$

for some $k \ge 0$ and some $0 \le \gamma \le 1$ is the most powerful test of size α for testing hypothesis H_0: $\underline{\theta} = \underline{\theta}_0$ versus H_1: $\underline{\theta} = \underline{\theta}_1$.

9.10

The Likelihood Ratio Test

We can rewrite the Neyman-Pearson decision rule in terms of the *Likelihood Ratio*

$$L(\underline{x}) = \frac{f_{\underline{\theta}_1}(\underline{x})}{f_{\underline{\theta}_0}(\underline{x})}.$$

The Neyman-Pearson test can be rewritten as

$$\phi(\underline{X}) = \begin{cases} 1, & \text{for } L(\underline{x}) > k, \\ \gamma, & \text{for for } L(\underline{x}) = k, \\ 0, & \text{for } L(\underline{x}) < k. \end{cases}$$

9.11

$$\boxed{\phi(\underline{X}) = \begin{cases} 1, & \text{for } L(\underline{x}) > k, \\ \gamma, & \text{for for } L(\underline{x}) = k, \\ 0, & \text{for } L(\underline{x}) < k. \end{cases}}$$

If there is a k_0 such that

$$P_{\underline{\theta}_0}(\{L(\underline{X}) \leq k_0(\underline{X})\}) = 1 - \alpha$$

take $k = k_0$.

If not, then find a k_0 such that

$$P_{\underline{\theta}_0}(\{L(\underline{X}) < k_0\}) < 1 - \alpha < P_{\underline{\theta}_0}(\{L(\underline{X}) \leq k_0)\})$$

and take $k = k_0$ and

$$\gamma = \frac{P_{\underline{\theta}_0}(\{L(\underline{X}) \leq k_0\}) - (1 - \alpha)}{P_{\underline{\theta}_0}(\{L(\underline{X}) = k_0\})}$$

Because $L(\underline{X})$ is a function of a random vector $\underline{X}$, it is itself a scalar random variable, and it takes on only nonnegative values.

If $P_{\underline{\theta}_0}(\{L(\underline{X}) = k\}) = 0$, then the threshold k achieving false alarm probability α can be found by solving

$$\alpha = P_{\underline{\theta}_0}(\{L(\underline{X}) > k\}) = \int_k^\infty f_{L,\underline{\theta}_0}(l)\, dl,$$

for k, where $f_{L,\underline{\theta}_0}(l)$ is the density function of $L(\underline{X})$ under H_0

We will find it convenient to use the *log-likelihood ratio*

$$\ell(\underline{X}) = \log\left(L(\underline{X})\right).$$

Because $\log(\cdot)$ is a monotonically increasing function on $(0, \infty)$, the most powerful test of size α equivalent to the likelihood ratio test will take the form

$$\phi(\underline{X}) = \begin{cases} 1, & \text{for } \ell(\underline{X}) > \ell_0, \\ \gamma, & \text{for } \ell(\underline{X}) = \ell_0, \\ 0, & \text{for } \ell(\underline{X}) < \ell_0, \end{cases}$$

where the threshold $\ell_0 = \log k$.

Working with $\ell(\underline{X})$ often yields simpler results than $L(\underline{X})$.

Notation:

$$L(\underline{X}) \underset{H_0}{\overset{H_1}{\gtrless}} k \qquad \text{or} \qquad \ell(\underline{X}) \underset{H_0}{\overset{H_1}{\gtrless}} \ell_0,$$

Example 1

Let X be a Gaussian random variable.

Under H_0: $X \sim \mathcal{N}[0, \sigma^2]$

Under H_1: $X \sim \mathcal{N}[\mu, \sigma^2]$

Find the most powerful test of size α, and determine an expression for the power β as a function of α, μ, and σ.

The likelihood ratio is

$$L(X) = \frac{f_1(X)}{f_0(X)} = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-(X-\mu)^2}{2\sigma^2}\right\}}{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-X^2}{2\sigma^2}\right\}}$$

$$= \exp\left\{\frac{2X\mu - \mu^2}{2\sigma^2}\right\} = \boxed{\exp\left\{\frac{\mu X}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\} \underset{H_0}{\overset{H_1}{\gtrless}} k.}$$

The log-likelihood ratio is

$$\ell(X) = \frac{2\mu X - \mu^2}{2\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ell_0$$

which simplifies to

$$X \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma^2}{\mu}\left(\frac{\mu^2}{2\sigma^2} + \ell_0\right) = \frac{\mu}{2} + \frac{\sigma^2 \ell_0}{\mu}$$

So our test is of the form

$$\phi(X) = \begin{cases} 1, & \text{for } X > \lambda, \\ 0, & \text{for } X \leq \lambda, \end{cases}$$

where

$$\lambda = \frac{\mu}{2} + \frac{\sigma^2 \ell_0}{\mu}$$

$$\boxed{\phi(X) = \begin{cases} 1, & \text{for } X > \lambda, \\ 0, & \text{for } X \le \lambda, \end{cases}}$$

$X \sim \mathcal{N}[0, \sigma^2]$ under H_0

$$\alpha = \mathrm{E}_{H_0}[\phi(X)] = \int_{\lambda}^{\infty} f_0(x)\, dx = 1 - \Phi\left(\frac{\lambda}{\sigma}\right)$$

Solving for λ achieving a size α test:

$$\boxed{\lambda_\alpha = \sigma \Phi^{-1}(1-\alpha)}$$

$X \sim \mathcal{N}[\mu, \sigma^2]$ under H_1

$$\beta = \mathrm{E}_1[\phi(X)] = \int_{\lambda_\alpha}^{\infty} f_1(x)\, dx$$

$$= 1 - \Phi\left(\frac{\lambda_\alpha - \mu}{\sigma}\right) = \boxed{1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\mu}{\sigma}\right).}$$

Example 2

The photon count N observed by a laser radar is a Poisson random variable:

Under H_0: $N \sim \text{Poisson}(\lambda_0)$

Under H_1: $N \sim \text{Poisson}(\lambda_1)$

We assume $\lambda_1 > \lambda_0$.

We have probability mass functions (pmfs):

$$p_0(n) = \frac{\lambda_0^n e^{-\lambda_0}}{n!}, \quad n = 0, 1, 2, \dots\,,$$

$$p_1(n) = \frac{\lambda_1^n e^{-\lambda_1}}{n!}, \quad n = 0, 1, 2, \dots\,,$$

9.18

The likelihood ratio is

$$L(N) = \frac{p_1(N)}{p_0(N)} = \frac{\lambda_1^N e^{-\lambda_1}/N!}{\lambda_0^N e^{-\lambda_0}/N!} = \left(\frac{\lambda_1}{\lambda_0}\right)^N e^{-(\lambda_1-\lambda_0)} \underset{H_0}{\overset{H_1}{\gtrless}} k$$

The log-likelihood ratio is

$$\ell(N) = \ln L(N) = N \ln\left(\frac{\lambda_1}{\lambda_0}\right) - (\lambda_1 - \lambda_0) \underset{H_0}{\overset{H_1}{\gtrless}} \ell_0 = \ln k.$$

Hence we can express the test in the form

$$N \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\ell_0 + (\lambda_1 - \lambda_0)}{\ln(\lambda_1/\lambda_0)} = \eta.$$

9.19

$$\boxed{N \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\ell_0 + (\lambda_1 - \lambda_0)}{\ln(\lambda_1/\lambda_0)} = \eta.}$$

The most powerful test of size α is

$$\phi(N) = \begin{cases} 1, & \text{for } N > \eta, \\ \gamma, & \text{for } N = \eta, \\ 0, & \text{for } N < \eta, \end{cases}$$

The power β of this size α test is

$$\beta = \mathrm{E}_1[\phi(N)]$$

$$= P_1(\{N > \eta_\alpha\}) + \gamma_\alpha \cdot P_1(\{N = \eta_\alpha\})$$

$$= 1 - P_1(\{N \le \eta_\alpha\}) + \gamma_\alpha \cdot P_1(\{N = \eta_\alpha\})$$

$$= 1 - \sum_{n=0}^{\eta_\alpha} \frac{\lambda_1^n e^{-\lambda_1}}{n!} + \gamma_\alpha \frac{\lambda_1^{\eta_\alpha} e^{-\lambda_1}}{\eta_\alpha!}$$

The Receiver Operating Characteristic (ROC)

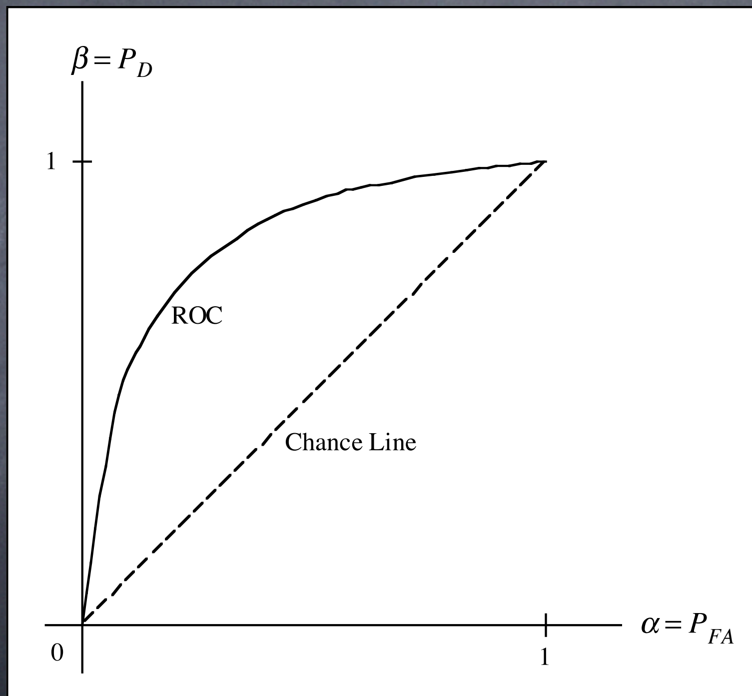
The performance of a binary test is characterized by the pair (α, β)

For a likelihood ratio test, we achieve different pairs $(\alpha(k), \beta(k))$ for different threshold values k

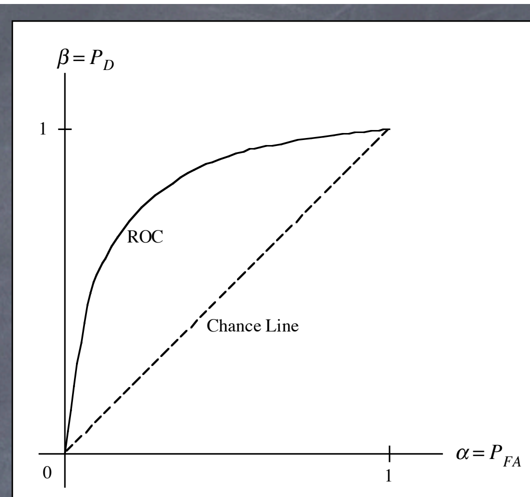
The locus of points $\{(\alpha(k), \beta(k));\ k \in (0, \infty)\}$ specifies all achievable (α, β) that can be obtained by varying the threshold k.

Such a curve is called a *Receiver Operating Characteristic* (ROC)

A typical ROC appears as follows:

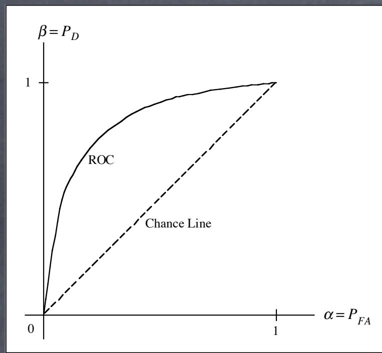


Properties of the ROCs of Likelihood Ratio Tests



1. All continuous likelihood ratio tests have ROCs that are convex downward.

2. Points on the *chance line* ($\beta = \alpha$) can be achieved without observing any data by picking hypothesis H_1 at random with probability α. (i.e., flipping a biased coin with probability α of coming up "heads").



Properties of the ROCs of Likelihood Ratio Tests (Continued)

3. All continuous likelihood ratio tests have ROC's that are above the chance line. This is a consequence of property 1, because the points $(\alpha, \beta) = (0, 0)$ and $(\alpha, \beta) = (1, 1)$ are contained on all ROC's.

4. The slope of the ROC at any particular point is given by the threshold achieving that operating point, i.e.,

$$\frac{d\beta}{d\alpha} = \frac{d\beta/dk}{d\alpha/dk} = \frac{f_{L,\theta_1}(\ell)}{f_{L,\theta_0}(\ell)} = k \geq 0$$