Outline
1) Hypothesis testing
2) Neyman–Pearson Lemma
Hypothesis Testing

In hypothesis testing, we use data $X$ to infer which of two submodels generated $X$.

Model $\mathcal{S} = \{ P_\theta : \theta \in \Theta \}$

Null hypothesis $H_0 : \theta \in \Theta_0$.

Alternative hyp. $H_1 : \theta \in \Theta \setminus \Theta_0$.

(Whenever $H_1$ unspecified, assume $\Theta_0 = \Theta \setminus \Theta_0$.)

$H_0$ is “default choice”: we either

1. accept $H_0$ (fail to reject, no definite concl.)
2. reject $H_0$ (conclude $\Theta_0$ false, $\Theta_1$ true)

Ex $X \sim N(\theta, 1)$ $H_0 : \theta \leq 0$ vs $H_1 : \theta > 0$

or $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$

Ex $X_1, \ldots, X_n \sim P$ $Y_1, \ldots, Y_m \sim Q$ $H_0 : P = Q$ vs $H_1 : P \neq Q$

Common conceptual objection: we “know” $\theta \neq 0$ or $P \neq Q$ already, why bother?

We will return to this.
Power Function

Can describe a test formally by its critical function (a.k.a. test function)

\[ \phi(X) = \begin{cases} 0 & \text{accept } H_0 \\ \pi \in (0,1) & \text{reject w.p. } \pi \\ 1 & \text{reject } H_0 \end{cases} \]

In practice, randomization rarely used \( (\phi(X) = 0, 1) \)

(In theory, simplifies discussions.)

For non-randomized \( \phi \), the rejection region is \( R = \{ x : \phi(x) = 1 \} \)

\( A = X \setminus R \) called acceptance region

Power function: \( \beta_\phi(\theta) = P_{\theta} [\phi(x)] = P_{\theta} [\text{Reject } H_0] \)

fully summarizes behavior of test.

The significance level of \( \phi \) is \( \sup_{\theta \in \Theta} \beta_\phi(\theta) \)

We say \( \phi \) is a level-\( \alpha \) test if its significance level is \( \leq \alpha \in [0, 1] \) (Sorry)

Ubiquitous choice is \( \alpha = 0.05 \)

["Most influential offhand remark in history of science"]

Goal: Maximize \( \beta_\phi(\theta) \) on \( \Theta_1 \), subject to level-\( \alpha \) constraint
Examples

Ex: \( X \sim N(\theta, 1) \)  \( H_0: \theta = 0 \)  \( H_1: \theta \neq 0 \)
Let \( z_{\alpha} = \Phi^{-1}(1-\alpha) \), \( \Phi = \text{normal cdf} \).

\[
\phi_2(x) = 1 \{ |x| > z_{\alpha/2} \} \quad \text{(2-sided test)}
\]
\[
\phi_1(x) = 1 \{ x > z_{\alpha} \} \quad \text{(1-sided test)}
\]
\[
\phi_3(x) = 1 \{ x < -z_{\alpha/2} \text{ or } x > z_{2\alpha/3} \}
\]

Sometimes a unique best test exists:

Ex: \( X \sim N(\theta, 1) \)  \( H_0: \theta \leq 0 \)  \( H_1: \theta > 0 \)
\( \phi \) is best possible level-\( \alpha \) test.

\[
\beta_\phi(\theta)
\]

\[
\Phi
\]

\[
\phi_3
\]

\[
\phi_1
\]

\[
\phi_2
\]

\[
\Phi
\]

\[
\beta_\phi(\theta)
\]

\[
\Phi
\]

\[
\phi_3
\]

\[
\phi_1
\]

\[
\phi_2
\]
Likelihood Ratio Test

A simple hypothesis is a single distribution:
\( \Theta_0 = \theta_0 \) or \( \Theta_1 = \{ \theta, \gamma \} \)

When null/alt. both simple, there exists a unique best test which rejects for large values of the likelihood ratio:

\[
\Phi^*(x) = \begin{cases} 
1 & \frac{p_1(x)}{p_0(x)} > c \\
\gamma & \frac{p_1(x)}{p_0(x)} = c \\
0 & \frac{p_1(x)}{p_0(x)} < c 
\end{cases}
\]

where \( p_1, p_0 \) are null/alt. densities.

(Note dominating measure always exists, e.g. \( p_0 + p_1 \))

and \( \gamma, c \) chosen to make \( \mathbb{E}_\theta \Phi^*(x) = \alpha \)

\( \Phi^* \) is called the likelihood ratio test (LRT)

Intuition: (discrete case)

Power under \( H_1 \): \( \int_R p_1(x) dm(x) \)

Sig. level: \( \int_R p_0(x) dm(x) \)

[Analogy: $100 to buy as much flour as possible]
**Proposition (Keener 12.1)**

Suppose $c > 0$, $\phi^*$ maximizes Lagrangian:

$$E_1[\phi(X)] - cE_0[\phi(X)]$$

among all critical functions.

If $E_0[\phi^*(X)] = \alpha$ then $\phi^*$ maximizes

$$E_1[\phi(X)]$$

among all level-\( \alpha \) tests $\phi$

(Essentially, Lagrange multipliers work even though $X$ infinite/cts./etc.)

**Proof**

Suppose $E_0[\phi(X)] \leq \alpha$. Then

$$E_1[\phi(X)] \leq E_1[\phi(X)] + c(\alpha - E_0[\phi(X)])$$

$$\leq E_1[\phi^*(X)] - cE_1[\phi^*(X)] + c\gamma$$

$$= E_1[\phi^*(X)]$$

\( \Box \)
**Theorem (Neyman-Pearson Lemma)**

LRT with significance level $\alpha$ is optimal for testing $H_0: X \sim \rho_0$ vs. $H_1: X \sim \rho_1$.

**Proof**

Maximize the Lagrangian:

$$E[x \phi(x)] - c E_0[\phi(x)]$$

$$= \int_{x} (\rho_1(x) - \rho_0(x)) \phi(x) \, dm(x)$$

$$= \int_{\rho_1 > \rho_0} 1 |\rho_1 - \rho_0| \phi \, dm - \int_{\rho_1 < \rho_0} 1 |\rho_1 - \rho_0| \phi \, dm$$

Max. first term $\Rightarrow \phi(x) = 1$ when $\rho_1(x) > \rho_0(x)$

Min. second term $\Rightarrow \phi(x) = 0$ when $\rho_1(x) < \rho_0(x)$

Choose minimum $c$ s.t.

$$P_0 \left[ \frac{\rho_1(x)}{\rho_0(x)} > c \right] \leq \alpha \leq P_0 \left[ \frac{\rho_1(x)}{\rho_0(x)} \geq c \right]$$

And choose $\gamma$ to "top up" sig. level:

$$P_0 \left[ \frac{\rho_1}{\rho_0} > c \right] + \gamma P_0 \left[ \frac{\rho_1}{\rho_0} = c \right] = \alpha \quad \text{ (Apply Prop.)}$$

Keener gives converse up to wiggle room if

$$\frac{\rho_1(x)}{\rho_0(x)} = c \text{ for more than 1 value of } x.$$
Choosing $c_1, c_2$ for $\phi^*$

$$P_0(\frac{\rho_i}{\rho_0} > c)$$

Keener Cor 12.4: If $\rho_0 \neq \rho_1$ and $\phi$ is LRT with level $\alpha \in (0,1)$ then $E_1 \phi(x) > \alpha$

Proof: $m(\{\rho_i > \rho_0\})$, $m(\{\rho_i < \rho_0\})$ both $> 0$

\[ C \geq 1: \]
\[ E_1 \phi^* - E_0 \phi^* = \int |\rho_i - \rho_0| \phi^* \, dm - \int |\rho_i - \rho_0| \phi^* \, dm \]

Power - $\alpha$

\[ \frac{\rho_i}{\rho_0} > 1 \]

\[ > 0 \]

\[ C \leq 1: \]
\[ \int \rho_i - \rho_0 (1 - \phi^*) \, dm \]

1 - Power - (1-$\alpha$)

\[ \frac{\rho_i}{\rho_0} < 1 \]

\[ < 0 \]
Ex 1-param. exp. fm.

\[ X_1, \ldots, X_n \overset{iid}{\sim} \rho_\eta(x) \]

\[ = e^{\eta T(x) - A(\eta)} h(x) \]

\[ H_0: \eta = \eta_0 \quad \text{vs.} \quad H_1: \eta = \eta_1, \eta > \eta_0 \]

\[ \frac{\rho_1(x)}{\rho_0(x)} = \frac{\prod_{i=1}^{n} \rho_\eta(x_i)}{\prod_{i=0}^{n} \rho_{\eta_0}(x_i)} \]

\[ = e^{(\eta_1 - \eta_0) \sum T(x_i) - n A(\eta_1)} \]

\[ = e^{(\eta_1 - \eta_0) \sum T(x_i) - n (A(\eta_1) - A(\eta_0))} \]

\( \phi^*(X) \) rejects for large \( \sum T(x_i) \):

\[ \phi^*(X) = \begin{cases} 
0 & \sum T(x_i) < c \\
\gamma & \sum T(x_i) = c \\
1 & \sum T(x_i) > c 
\end{cases} \]

choose \( c, \gamma \) to make

\[ P_{\eta_0} [\sum T(x_i) > c] \left[ \gamma P_{\eta_0} \sum T(x_i) = c \right] = \alpha \]

**Surprise:** \( \phi^*(X) \) depends only on \( \eta_0 \) and \( \text{sgn}(\eta_1 - \eta_0) \), not on \( \eta_1 \). Next topic.