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 $S = \{ P_0 : \theta \in \Theta \}$

Null hypothesis $H_0 : \theta \in \Theta_0$.
Alternative hyp. $H_1 : \theta \in \Theta_1$.

(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 \((\Phi(x) = 50, 13)\)
(In theory, simplifies discussions.)

For non-randomized \(\phi\), the rejection region is

\[
R = \{ x : \Phi(x) = 13 \}
\]

\[ A = \mathcal{X} \setminus R \text{ called acceptance region} \]

Power function:

\[
\beta_\phi(\theta) = \text{IE}_{\theta} [\Phi(x)] \text{ rejection prob.} = \Pr_{\theta} [\text{Reject } H_0]
\]

fully summarizes behavior of test.

The significance level of \(\phi\) is \(\sup_{\theta \in \Theta_0} \beta_\phi(\theta)\)

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 \(H_1\), subject to level-\(\alpha\) constraint.
**Examples**

**Ex:** $X \sim N(\theta, 1)$ \quad $H_0: \theta = 0$ \quad $H_1: \theta \neq 0$

Let $ \bar{z}_\alpha = \Phi^{-1}(1-\alpha)$, $\Phi$ = normal cdf.

- $\phi_2(x) = 1 \{ |x| > \bar{z}_{\alpha/2} \}$ \quad (2-sided test)
- $\phi_1(x) = 1 \{ x > \bar{z}_\alpha \}$ \quad (1-sided test)
- $\phi_3(x) = 1 \{ x < -\bar{z}_{\alpha/2} \text{ or } x > \bar{z}_{2\alpha/3} \}$

Sometimes a unique best test exists:

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

$\phi$ is best possible level-$\alpha$ test.
Likelihood Ratio Test

A simple hypothesis is a single distribution:

\( \Theta_0 = \{ \theta \} \) or \( \Theta_1 = \{ \theta \} \)

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 \( c, \gamma \) chosen to make \( E_0 \phi^*(x) = \alpha \)

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

Intuition: (discrete case)

Power under \( H_1 \): \( \sum \text{support of } p_1 \)

Sig. level: \( \sum \text{support of } p_0 \)

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

Suppose \( c > 0 \), \( \phi^* \) maximizes Lagrangian:
\[
\mathbb{E}_1 [\phi(X)] - c \mathbb{E}_0 [\phi(X)]
\]
among all critical functions.

If \( \mathbb{E}_0 [\phi^*(X)] = \alpha \) then \( \phi^* \) maximizes
\[
\mathbb{E}_1 [\phi(X)] \quad \text{among all level-}\alpha \text{ tests } \phi
\]
(Essentially, Lagrange multipliers work even though \( X \) infinite/cts./etc.)

Proof. Suppose \( \mathbb{E}_0 [\phi(X)] \leq \alpha \). Then
\[
\mathbb{E}_1 [\phi(X)] \leq \mathbb{E}_1 [\phi(X)] + c (\alpha - \mathbb{E}_0 [\phi(X)])
\]
\[
\leq \mathbb{E}_1 [\phi^*(X)] - c \mathbb{E}_1 [\phi^*(X)] + c \gamma
\]
\[
= \mathbb{E}_1 [\phi^*(X)] \quad \blacksquare
\]
**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:

\[
E_\alpha [\phi(X)] - c E_\rho_0 [\phi(X)]
\]

\[
= \int_X (\rho_1(x) - \rho_0(x)) \phi(x) \, dm(x)
\]

\[
= \int_{\rho_1 > \rho_0} 1 \rho_1 - \rho_0 \phi \, dm - \int_{\rho_1 < \rho_0} 1 \rho_1 - \rho_0 \phi \, dm
\]

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 \( \text{minimum} \ c \ s.t. \)

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

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

\[
P_0 \left[ \frac{\rho_1}{\rho_0} > c \right] + \gamma P_0 \left[ \frac{\rho_1}{\rho_0} = c \right] = \alpha
\]

Keener gives converse up to wiggle room if

\[\frac{\rho_1(x)}{\rho_0(x)} = c \quad \text{for more than 1 value of } X\]
Choosing \( c_\alpha, \gamma_d \) for \( \phi^* \)

\[ P_0 \left( \frac{\rho_1}{\rho_0} > c \right) \]

\[ 0 \leq \gamma \leq 1 \]

\[ \alpha_1 \]

\[ \alpha_2 \]

\[ \alpha_3 \]

\[ \gamma_1 \]

\[ \gamma_2 \]

\[ \gamma_3 \]

\[ \gamma_4 \]

\[ \gamma_5 \]

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: \( m(\{\rho_1 > \rho_0\}), m(\{\rho_1 < \rho_0\}) \) both \( \geq 0 \)

\[ C \geq 1: \]

\[ E_1 \phi^* - E_0 \phi^* = \int_1^{\frac{\rho_1}{\rho_0}} |\rho_1 - \rho_0| \phi^* dm - \int^{\frac{\rho_1}{\rho_0}} |\rho_1 - \rho_0| \phi^* dm \]

\[ \text{Power} - \alpha \]

\[ \frac{\rho_1}{\rho_0} > 1 \]

\[ \frac{\rho_1}{\rho_0} \leq 1 \]

\[ > 0 \]

\[ C \leq 1: \]

\[ E_1 (1 - \phi^*) - E_0 (1 - \phi^*) = -\int_1^{\frac{\rho_1}{\rho_0}} |\rho_1 - \rho_0| (1 - \phi^*) dm \]

\[ 1 - \text{Power} - (1 - \alpha) \]

\[ \frac{\rho_1}{\rho_0} < 1 \]

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

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

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

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

\[ \frac{\rho_1(x)}{\rho_0(x)} = \frac{\prod_{i=1}^{n} \rho_{\eta_1}(x_i)}{\prod_{i=0}^{n} \rho_{\eta_0}(x_i)} \]

\[ = e^{(\eta_1 - \eta_0) \Xi T(x_i) - nA(\eta_1)} \]

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

\[ \phi^*(x) \text{ rejects for large } \Xi T(x_0): \]

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

choose \( c, \gamma \) to make

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

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