Outline

1) Hypothesis testing
2) Neyman-Pearson Lemma
Hypothesis Testing

In hypothesis testing, we use data \( X \) to infer which of two submodels generated \( X \)

\[
\text{Model } \mathcal{M} = \{ \mathcal{P}_0 : \Theta \in \Theta_0 \}
\]

Null hypothesis \( H_0 : \Theta \in \Theta_0 \)

Alternative hyp. \( H_1 : \Theta \in \Theta_1 \)

(Whenever \( H_1 \) unspecified, assume \( \Theta_1 = \Theta \setminus \Theta_0 \))

\( H_0 \) 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 \( (\phi(x) = \delta_0, 13) \)
(In theory, simplifies discussions.)

For non-randomized \( \phi \), the rejection region is
\[ R = \{ x : \phi(x) = 1 \} \]
\[ A = X \setminus R \text{ called acceptance region} \]

Power function: \[ \beta(\Theta) = \text{IE}_{\Theta} [\phi(x)] \text{ rejection prob.} \]
\[ = \text{P}_{\Theta} [\text{Reject } H_0] \]

fully summarizes behavior of test.

The significance level of \( \phi \) is \( \sup_{\Theta \in \Theta_0} \beta(\Theta) \)

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 \( \beta_{\Theta}(x) \) on \( \Theta \), subject to level-\( \alpha \) constraint
Examples

\[ X \sim N(\theta, 1) \quad H_0: \theta = 0 \quad H_1: \theta \neq 0 \]

Let \( z_\alpha = \Phi^{-1}(1-\alpha) \), \( \Phi \) = normal cdf.

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

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:
$$\Theta_0 = \{\theta_0\} \text{ or } \Theta_1 = \{\theta_1\}$$

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 $$H_1$$: $$\sum_X p_1(x) dm(x)$$ **Bang**
- Sig. level: $$\sum_X p_0(x) dm(x)$$ **Buck**

[Analogy: $100 to buy as much flour as possible]
Proposition (Keener 12.1)
Suppose $c > 0$, $\phi^*$ maximizes Lagrangian:
\[ E_1[\phi(x)] - c E_0[\phi(x)] \]
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)] - c E_1[\phi^*(x)] + c \gamma
= E_1[\phi^*(x)] \qed
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:

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

$$= \int (\rho_1(x) - c \rho_0(x)) \phi(x) \, dm(x)$$

$$= \int l \rho_1 - c \rho_0 \phi \, dm_1 - \int l \rho_1 - c \rho_0 \phi \, dm_2$$

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

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

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:

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

Keener gives converse up to wiggle room if

$$\frac{\rho_1(x)}{\rho_0(x)} = c$$ for more than 1 value of $x$.  


Choosing \( c_\alpha, \sigma_\alpha \) for \( \phi^* \)

\[
\prod_0 (\frac{\rho}{\rho_0} \geq c) \uparrow
\]

\[
\begin{align*}
\alpha_1, \\
\alpha_2, \\
\alpha_3, \\
\alpha_4 (0, 1)
\end{align*}
\]

\[
\begin{align*}
c_1, \\
c_2, \\
c_3 (c_3 \text{ or too}) \quad \gamma_2, \\
\gamma_3 = 0
\end{align*}
\]

Keener Cor 12.4: If \( \rho_0 \neq \rho_1 \) and \( \phi \) is LRT with level \( \alpha \in (0, 1) \) then \( \mathbb{E}_1 \phi(x) > \alpha \)

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

\[
C \geq 1:
\frac{\mathbb{E}_1 \phi^* - \mathbb{E}_0 \phi^* = \int |\rho_1 - \rho_0| \phi^* d\mu - \int |\rho_1 - \rho_0| \phi^* d\mu}{\text{Power} - \alpha} > 0
\]

\[
C \leq 1:
\frac{\mathbb{E}_1 (1 - \phi^*) - \mathbb{E}_0 (1 - \phi^*) = -\int |\rho_1 - \rho_0| (1 - \phi^*) d\mu}{1 - \text{Power} - (1 - \alpha)} < 0
\]
Ex. 1-param. exp. fn. 

\[ X \sim \rho_\gamma(x) = e^{x T(x) - A(\gamma)} h(x) \]

\[ H_0 : \gamma = \gamma_0 \quad \text{vs.} \quad H_1 : \gamma = \gamma_1, > \gamma_0 \]

\[ \frac{\rho_1(x)}{\rho_0(x)} = \frac{e^{\gamma_1 x T(x) - A(\gamma_1)}}{e^{\gamma_0 x T(x) - A(\gamma_0)}} \]

\[ = e^{(\gamma_1 - \gamma_0) x T(x) - (A(\gamma_1) - A(\gamma_0))} \]

\[ \phi^*(x) \] rejects for large \( T(x) \):

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

choose \( c, \gamma \) to make

\[ \Pr_{\gamma_0}(T(x) > c) + \gamma \Pr_{\gamma_0}(T(x) = c) = \alpha \]

Ex. \( X_1, \ldots, X_n \ \text{\iid } \rho_\gamma(x) \), same \( H_0, H_1 \):

Reject for large \( \frac{1}{n} \sum T(X_i) \)

**Surprise:** \( \phi^*(x) \) depends only on \( \gamma_0 \) and \( \text{sign} (\gamma_1 - \gamma_0) \), not on \( \gamma_1 \).

Next topic: Uniformly most powerful (UMP) tests