How to Construct a Level-α Test?

- Approximate the data by a statistical model with an unknown parameter θ, and describe the corresponding H_0 and H_a as statements on θ.

- Construct a test statistic $T(x_1, \ldots, x_n)$. Know the range of T that is against H_0 (e.g., right-tail, left-tail, or both tails), which will be the Rejection Region. For example, the Rejection Region is

$$T(x_1, \ldots, x_n) \geq c.$$

- Find the distribution or asymptotic distribution of T under H_0, then choose the cutoff value c such that

$$\mathbb{P}(T(X_1, \ldots, X_n) \geq c \mid H_0) = \alpha.$$

Remaining issues we will address in the last two lectures:

- How to find the test statistic $T(x_1, \dots, x_n)$?
  - Likelihood ratio test (LRT)
- How to construct the best test?
  - The Uniformly Most Powerful (UMP) test
- What if the distribution of $T(x_1, \dots, x_n)$ under H_0 is difficult to evaluate?
  - Wald test, score test, asymptotic LRT (not required)

UMP Test

Consider a collection of tests for testing

$$H_0 : \theta \in \Theta_0, \quad H_a : \theta \in \Theta_1.$$

- There is a trade-off between the type I and type II errors, and it's impossible to reduce both of them.
- So we focus on all level-α tests, i.e., the type I error of all tests in this collection $\mathcal{C}$ is controlled by α.
- A test in $\mathcal{C}$ is called the uniformly most powerful (UMP) test if its power $\beta(\theta)$ is bigger than the power of any other test in $\mathcal{C}$ for any $\theta \in \Theta_1$.

  That is, a UMP test has the smallest type II error among all tests in $\mathcal{C}$.

The Likelihood Ratio Test

We discuss the LRT test in the following three cases.

- (Simple vs Simple) $H_0 : \theta = \theta_0, \quad H_a : \theta = \theta_1$

  The LRT test is the UMP by the Neyman-Pearson Lemma.

- (One-sided test)

$$H_0 : \theta = \theta_0, H_a : \theta > \theta_0 \text{ or } \theta < \theta_0.$$

  Construct the UMP test based on the Simple-vs-Simple case.

- (Two-sided test)

$$H_0 : \theta = \theta_0, \quad H_a : \theta \neq \theta_0.$$

  UMP tests usually do not exist; the LRT test could be the UMP test among all unbiased level α tests.

Simple vs Simple

$$H_0 : \theta = \theta_0 \quad H_a : \theta = \theta_1$$

Likelihood Ratio:

$$\lambda(x_1, \ldots, x_n) = \frac{f(x_1; \theta_0) f(x_2; \theta_0) \cdots f(x_n; \theta_0)}{f(x_1; \theta_1) f(x_2; \theta_1) \cdots f(x_n; \theta_1)}.$$

Neyman-Pearson Lemma : Construct a test with rejection region

$$S = \{(x_1, \ldots, x_n) : \lambda(x_1, \ldots, x_n) \leq k\},$$

where $k \geq 0$ is a constant. Choose k to control the Type I error

$$\mathbb{P}_{\theta_0}((X_1, \ldots, X_n) \in S) = \alpha.$$

Then this is the UMP level-α test.

Proof. Consider another level-α test with rejection region $\tilde{S}$.

Since both tests have the same type I error,

$$\int_S f(\mathbf{x};\theta_0)d\mathbf{x} = \int_{\tilde{S}} f(\mathbf{x};\theta_0)d\mathbf{x}, \tag{1}$$

where $\mathbf{x}$ is a shorthand notation for $(x_1,\ldots,x_n)$.

To show that the LRT is the UMP test, we need to show that

$$\int_S f(\mathbf{x};\theta_1)d\mathbf{x} \geq \int_S f(\mathbf{x};\theta_1)d\mathbf{x}. \quad (?)$$

To simplify our discussion, let's first remove the integral over the shared region of S and $\tilde{S}$ from both sides.

Let $A = S \cap \tilde{S}$, and

$$S = S_1 \cup A, \quad \tilde{S} = \tilde{S}_1 \cup A.$$

- S_1: part of the region region of the LRT, but not in the rejection region of the other test.
- $\tilde{S}_1$: part of the region region of the other test, but not in the rejection region of the LRT.

Due the construction of the LR test, we know that there exists a constant $k > 0$ such that

$$\begin{aligned} \mathbf{x} \in S_1 &\quad : \quad f(\mathbf{x};\theta_0)/f(\mathbf{x};\theta_1) \leq k \\ \mathbf{x} \in \tilde{S}_1 &\quad : \quad f(\mathbf{x};\theta_0)/f(\mathbf{x};\theta_1) > k \end{aligned}$$

Since $\int_S \cdots = \int_{S_1} \cdots + \int_A \cdots$ and $\int_{\tilde{S}} \cdots = \int_{\tilde{S}_1} \cdots + \int_A \cdots$, (1) becomes

$$\int_{S_1} f(\mathbf{x};\theta_0)d\mathbf{x} = \int_{\tilde{S}_1} f(\mathbf{x};\theta_0)d\mathbf{x}.$$

To show that the LRT is the UMP, we need to show that

$$\int_S f(\mathbf{x};\theta_1)d\mathbf{x} \geq \int_S f(\mathbf{x};\theta_1)d\mathbf{x}. \quad (?)$$

Re-express the integrand:

$$\text{LHS} \quad : \quad \int_{S_1} f(\mathbf{x};\theta_0)\frac{f(\mathbf{x};\theta_1)}{f(\mathbf{x};\theta_0)}d\mathbf{x} \geq \frac{1}{k}\int_{S_1} f(\mathbf{x};\theta_0)d\mathbf{x}$$

$$\text{RHS} \quad : \quad \int_{\tilde{S}_1} f(\mathbf{x};\theta_0)\frac{f(\mathbf{x};\theta_1)}{f(\mathbf{x};\theta_0)}d\mathbf{x} < \frac{1}{k}\int_{\tilde{S}_1} f(\mathbf{x};\theta_0)d\mathbf{x}$$

which completes the proof of the **NP Lemma**.

Go through examples 1-5 in `Week14_LRT1.pdf`

- Simplify the likelihood ratio $\lambda(x_1, \ldots, x_n)$: cancel common terms in numerator and denominator.

- Simply the inequality $\lambda(x_1, \ldots, x_n) \leq k$ to be

$$u(x_1, \ldots, x_n) \leq c,$$

  where $u(x_1, \ldots, x_n)$ is a summary statistic (e.g., $\sum_i x_i$, $\sum_i x_i^2$, etc) and k_1 is a constant that may depend on θ_1 and θ_0.

- Find the distribution of $u(X_1, \ldots, X_n)$ under H_0, and then choose k_1 to control the type I error:

$$\mathbb{P}_{\theta_0}(u(X_1, \ldots, X_n) \leq c) = \alpha.$$

One-sided Test

$$H_0 : \theta = \theta_0, \quad H_a : \theta > \theta_0 (\text{ or } \theta < \theta_0). \tag{2}$$

Follow the **NP Lemma** to construct the UMP test for a simple-vs-simple test

$$H_0 : \theta = \theta_0, \quad H_a : \theta = \theta_1, \tag{3}$$

where θ_1 is any fixed value bigger than θ_0. In some cases, you'll find that the rejection region does not depend on θ_1, i.e., the UMP test for (3) is the same for all θ_1's as long as $\theta_1 > \theta_0$. Therefore, it's also the UMP test for the original one-sided test (2).

Example 4 in `Week14_LRT1.pdf`. Consider the following simple-vs-simple test

$$H_0 : \lambda = 5, \quad H_a : \lambda = \lambda_1,$$

where $\lambda_1 < 5$.

$$\lambda(x) = \frac{f(x;5)}{f(x;\lambda_1)} = \frac{5e^{-5x}}{\lambda_1 e^{-\lambda_1 x}} = \frac{5}{\lambda_1} e^{-(5-\lambda_1)x}.$$

$$\begin{aligned}\lambda(x) = \frac{5}{\lambda_1} e^{-(5-\lambda_1)x} \le k \quad &\Leftrightarrow \quad e^{-(5-\lambda_1)x} \le k\frac{\lambda_1}{5} \\ &\Leftrightarrow \quad -(5-\lambda_1)x \le \log\frac{k\lambda_1}{5} \\ &\Leftrightarrow \quad x \ge -\frac{1}{5-\lambda_1}\log\frac{k\lambda_1}{5} = c.\end{aligned}$$

So the rejection region is $S = \{x : x \ge c\}$.

For a level-α test (see 4(d) in `Week14_LRT1.pdf`), c is chosen such that

$$0.05 = \mathbb{P}(X > c \mid \lambda = 5) \Longrightarrow c = 0.599$$

which does not depend on the value of λ_1.

That is, $S = \{x : x > 0.599\}$ is the optimal rejection region for any $\lambda_1 < 5$, so it's the optimal rejection region for the following one-sided test

$$H_0 : \lambda = 5, \quad H_a : \lambda < 5.$$

Example 6 in `Week14_LRT1.pdf`. Consider the following simple-vs-simple test where $\theta_1 > 0.5$

$$H_0 : \theta = 0.5, \quad H_a : \theta = \theta_1.$$

$$\lambda(x_1, \ldots, x_n) = \frac{\prod_{i=1}^n \frac{1}{0.5} e^{-x_i/(0.5)}}{\prod_{i=1}^n \frac{1}{\theta_1} e^{-x_i/\theta_1}} = (2\theta_1)^n \exp\left\{ \left(\frac{1}{\theta_1} - 2 \right) \sum_{i=1}^n x_i \right\} \leq k$$

$$\Longleftrightarrow \sum_i x_i \geq c.$$

Due to the additivity of Gamma distribution, we have $\sum_{i=1}^n X_i \sim \mathsf{Ga}(n, \theta)$. To control the type I error, we choose c to be the 95% quantile of $\mathsf{Ga}(n, \theta)$.

In R, you can use command

```
> qgamma(0.95, 7, scale=0.5)
[1] 5.921198
```

You can use R to answer other questions

```
## 6(b)
> 1 - pgamma(5.92, 7, scale=0.75)
[1] 0.3265752
## 6(c)
> 1 - pgamma(6, 7, scale=0.5)
[1] 0.04582231
```

Go through examples in `Week14_LRT2.pdf`

Two-sided Test

$$H_0 : \theta = \theta_0, \quad H_a : \theta \neq \theta_0.$$

LRT: Reject H_0 if

$$\Lambda = \frac{L(\theta_0)}{\max_\theta L(\theta)} = \frac{L(\theta_0)}{L(\hat{\theta})} \leq k$$

where $\hat{\theta}$ is the MLE.

Go through Examples from `Week14_LRT3.pdf`

- **Example 2.** A useful equality (You can expand both sides to check that they are equal):

$$\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - a)^2.$$

In class, I'll use **Example 2** to explain why we cannot find the UMP test for two-sided test. Instead, we could find the most powerful test among unbiased level-α tests. A level-α test is unbiased if its power is bigger than or equal to α.

We have encountered a similar situation in estimation: it's impossible to find the estimator whose MSE is the smallest among all estimators (the collection of ALL estimators is just too large). Instead, we focus on all unbiased estimators, and the best one among all unbiased estimators is the one achieving the CRLB.

Suppose $n = 1$ and $\sigma^2 = 1$, i.e., $X_1 \sim \mathsf{N}(\mu, 1)$. Consider the following three rejection regions for testing

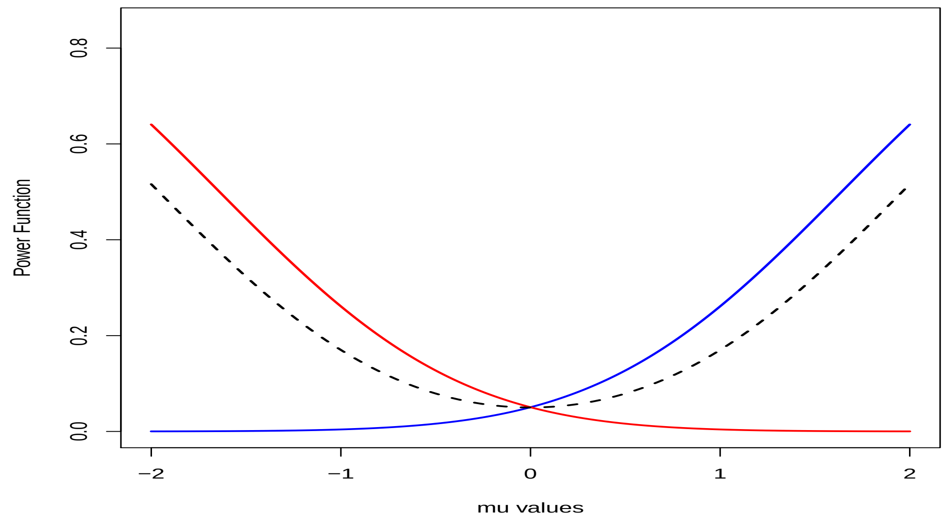
$$H_0 : \mu = 0, \quad H_a : \mu \neq 0. \tag{4}$$

- $S_1 = \{x : x > 1.64\}$
- $S_2 = \{x : x < -1.64\}$
- $S_3 = \{x : |x| > 1.96\}$

They all are legit rejection regions for (4) with type I error 5%.

Evaluate their power.

$$\begin{aligned}
\beta_1(\mu) &= \mathbb{P}_\mu(X > 1.64) = \mathbb{P}_\mu(X - \mu > 1.64 - \mu) = 1 - \Phi(1.64 - \mu) \\
\beta_2(\mu) &= \mathbb{P}_\mu(X < -1.64) = \mathbb{P}_\mu(X - \mu < -1.64 - \mu) = \Phi(-1.64 - \mu) \\
\beta_3(\mu) &= \mathbb{P}_\mu(X < -1.96) + \mathbb{P}_\mu(X > -1.96) \\
&= \Phi(-1.96 - \mu) + 1 - \Phi(1.96 + \mu)
\end{aligned}$$



– In some regions, red test is more powerful; in some regions blue test is more powerful, but no one is dominating the others in all range of μ values.

– Neither red test nor blue test is preferred since $\beta(\mu) \leq 5\%$ for some μ values from H_a, that is, the chance of having samples from the rejection region for some alternative μ values is less than the chance under H_0, which contradicts the intuition that rejection region provides evidence supporting H_a.

– We could restrict to unbiased level-α tests, which require

$$\beta(\theta) \geq \alpha$$

for any θ from H_a.

– It's beyond the scope of Stat410, but one show that for **Example 2**, the LRT (with rejection region S_3) is indeed the UMP test among all unbiased level-α tests.

- **Example 3.**

  Typos: the estimates are for $\hat{\sigma}_0^2$ and $\hat{\sigma}^2$.

  Due to the way we estimate them, you'll find that

$$\exp\left\{-\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right\} = \exp\left\{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\} = e^{-n/2}.$$

- **Example 6.** To understand what the rejection region $\{x : \Lambda(x) \leq c\}$ looks like, rewrite the rejection region as

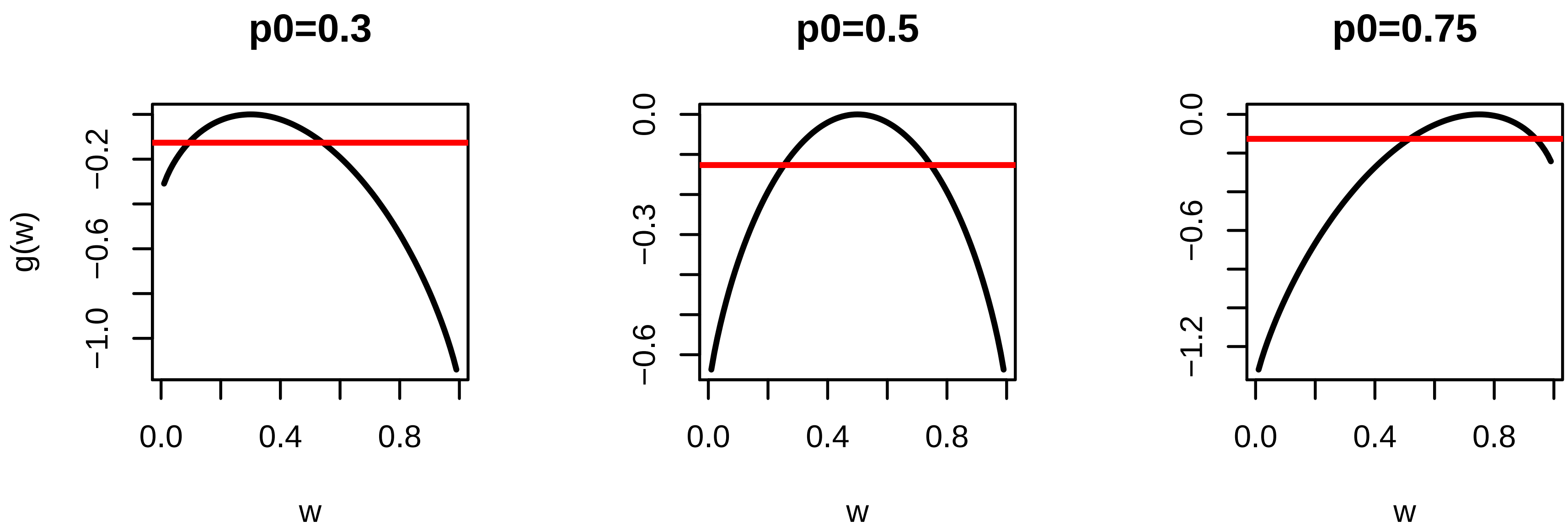
$$\frac{1}{n}\log\Lambda(x) = \frac{x}{n}\log\frac{p_0}{x/n} + (1 - n/x)\log\frac{(1-p_0)}{1-x/n} \leq \frac{\log c}{n} = c'.$$

View $\frac{1}{n}\log\Lambda(x)$ as a function of x/n, denoted by

$$g(w) = w\log\frac{p_0}{w} + (1-w)\log\frac{(1-p_0)}{1-w}, \quad 0 \leq w \leq 1.$$

Although x/n can only take (n+1) different values: $0, 1/n, \ldots, (1-1/n), 1$, it's useful to discuss function $g(w)$ with $w \in [0,1]$.

So what does $g(w)$ look like? It's concave and achieves the maximum at $w = p_0$. So the rejection region $g(x/n) \leq c'$ corresponds to the two tails: $x < c_1$ and $x > c_2$.



The red line corresponds to choosing $c = 0.15$, i.e., the rejection region will be all the x values such that $g(x/n)$ is below the red line.

- **Examples 6-9.** For discrete samples, it is possible we cannot find a test whose type I error is exactly α, due to the discreteness of the distribution.

Tests based on Asymptotic Results

- Previously, after we find the rejection region, say $T(x_1, \ldots, x_n) > c$, we need to find the distribution of the test statistic $T(X_1, \ldots, X_n)$ under H_0, and then choose the cut-off value c such that

$$\mathbb{P}(T(X_1, \ldots, X_n) > c \mid H_0) = \alpha.$$

- What if the pdf of T (under H_0) does not have a closed-form, or difficult to evaluate? Then we can work with an approximation of the distribution of T (under H_0), which should work out well when n is large.

- There are several tests that are based on the asymptotic results of the MLE $\hat{\theta}_n$ or the log-likelihood function: Wald test, LRT test, and Score test.

- **Wald test**

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathsf{N}\Big(0, \frac{1}{I(\theta_0)}\Big)$$

$$\Leftrightarrow \quad \text{Reject } H_0, \text{ if } |\hat{\theta}_n - \theta_0| \geq \frac{1}{\sqrt{nI(\theta_0)}} z_{\alpha/2}.$$

- **Likelihood ratio test**

$$\Lambda = \frac{L(\theta_0)}{\max_\theta L(\theta)}, \quad -2\ln\Lambda \xrightarrow{D} \chi^2(1)$$

$$\Leftrightarrow \quad \text{Reject } H_0, \text{ if } -2\ln\Lambda \geq \chi^2_\alpha(1)$$

- **Score test**

$$\left(\frac{\frac{d}{d\theta}\ln L(\theta_0)}{\sqrt{nI(\theta_0)}}\right)^2 \geq \chi^2_\alpha(1).$$

An advantage of the Score test is that we do not need to compute the MLE; what's matter is just the evaluation with $\theta = \theta_0$.