Methods of Evaluating Tests (section 8.3)

How good is a testing procedure?

How to specify a rejection region?

Error Probabilities and the Power Function (Section 8.3.1)

Decision		Underlying Truth H_0	H_1
	Reject H_0	Type I error	—
	~~Accept H_0~~ (Fail to reject H_0)	—	Type II error

Def: Let R = rejection region of a testing procedure. Then

$$P(\text{Type I error}) = P_\theta(\underset{\sim}{X} \in R), \quad \text{if } \theta \in \Theta_0$$

$$P(\text{Type II error}) = P_\theta(\underset{\sim}{X} \in R^c) = 1 - P_\theta(\underset{\sim}{X} \in R), \quad \text{if } \theta \in \Theta_1$$

Def (8.3.1) The power function of a hypothesis test with rejection region R is

$$\beta(\theta) = P_\theta(\underset{\sim}{X} \in R), \quad \theta \in \Theta.$$

Remarks

1. The power function will play the same role in hypothesis testing that mean-squared error played in point estimation. ($\beta(\theta)$ will usually be our standard in assessing the goodness of a test)

2. The ideal power function is $\beta(\theta) = \begin{cases} 0 & \text{for } \theta \in \Theta_0 \\ 1 & \text{for } \theta \in \Theta_1. \end{cases}$

3. A good realistic test has "small" $\beta(\theta)$-values on Θ_0 and "large" $\beta(\theta)$-values on Θ_1.

Def (8.3.5 and 8.3.6) For $0 \le \alpha \le 1$, a test with power function $\beta(\theta)$ is a size α test if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$ and a level α test if $\sup_{\theta \in \Theta_0} \beta(\theta) \le \alpha$.

Remark: Typically, $\alpha = 0.01, 0.05$, or 0.10 and we would like a testing procedure to have a small Type I error probability with large power on Θ_1.

Ex: $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(θ), $\theta \in \Theta = \{\frac{1}{4}, \frac{3}{4}\}$.

Ho: $\theta = \frac{1}{4}$

Ha: $\theta = \frac{3}{4}$.

The LRT rejects Ho for $\bar{x} \geq k$, $k \in [\frac{1}{2}, 1]$

$\Leftrightarrow \quad s = \sum_{i=1}^{n} x_i \geq nk = d \in [\frac{n}{2}, n]$.

$$\beta(\theta) = P_\theta(\underset{\sim}{X} \in \underbrace{R}_{\text{rejection region}}) = P_\theta\left(\sum_{i=1}^{n} X_i \geq d\right) = \sum_{s=d}^{n} \binom{n}{s} \theta^s (1-\theta)^{n-s} \quad \text{since} \quad \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, \theta)$$

For $n = 10$,

I would select one of these two (d = 5, 6)

d	P(Type I error) $\beta(\frac{1}{4})$	Power on Θ_1 $\beta(\frac{3}{4})$	P(Type II error) $1-\beta(\frac{3}{4})$
5	0.078	0.98	0.020
6	0.020	0.922	0.078
7	0.004	0.776	0.224
8	0.0004	0.526	0.474
9	0.00003	0.244	0.756
10	0.00001	0.056	0.944

Ex: $X_1, \ldots, X_n \overset{iid}{\sim}$ Exponential(θ), $\theta > 0$

Ho: $\theta \leq \theta_0$

Ha: $\theta > \theta_0$

The LRT rejects Ho for $\bar{x} > k\theta_0$ when $1 < k < \infty$

$\Leftrightarrow \quad s = \sum_{i=1}^{n} x_i > kn\theta_0$...

$$\beta(\theta) = \int_{kn\theta_0}^{\infty} f(s|\theta)\,ds \quad \text{when} \quad \sum_{i=1}^{n} X_i \sim \text{Gamma}(\underset{\text{"shape"}}{n}, \underset{\text{"scale"}}{\theta}).$$

(Reminder: Want $\beta(\theta)$ "small" on Θ_0 and $\beta(\theta)$ "large" on Θ_1),

See Handout for $\theta_0 = 1$, $n = 10$.

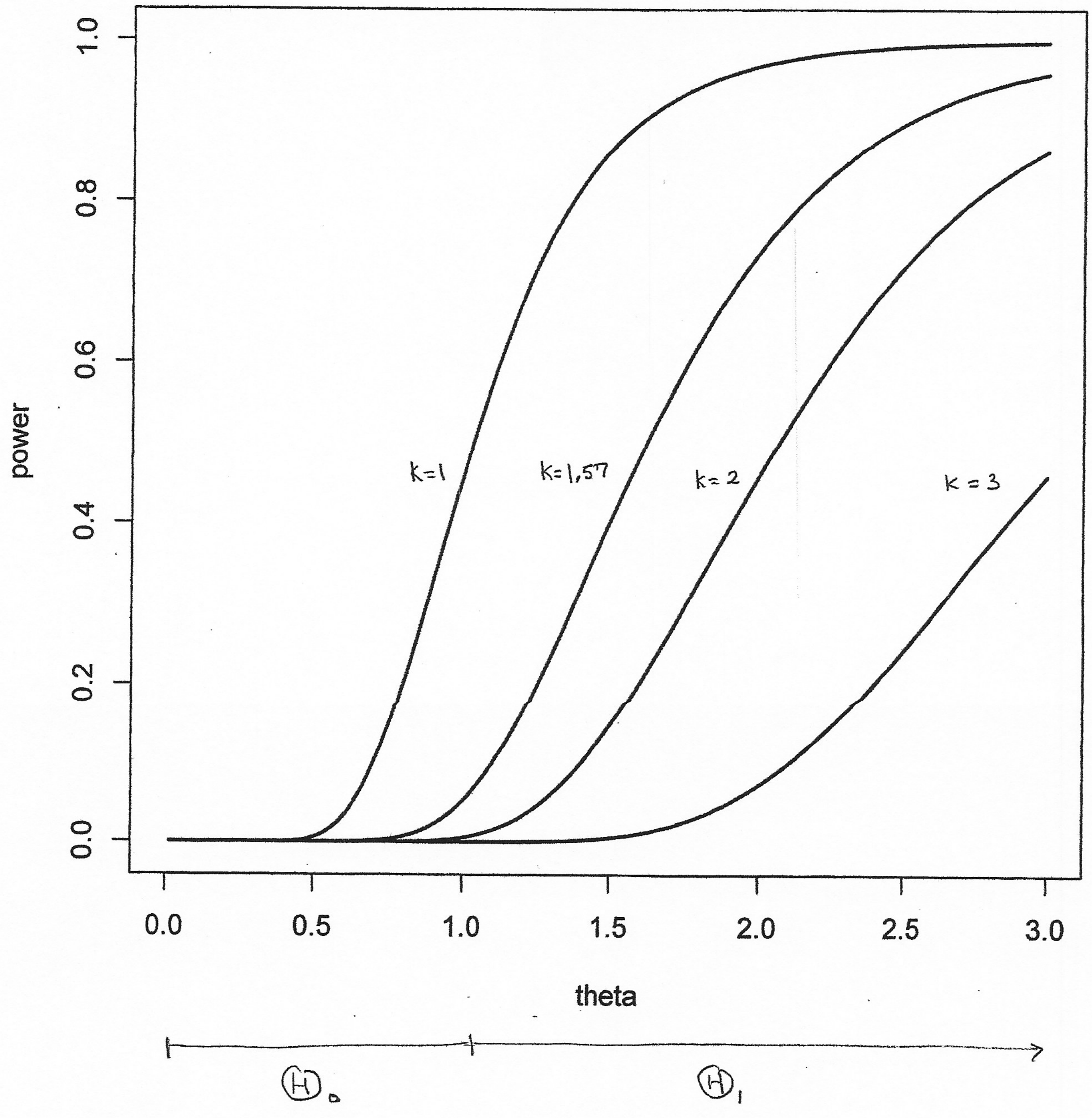
Which testing procedure would you prefer? A size $\alpha = 0.05$ test would use $k = 1.57$

$$\beta(\theta) = 1 - F_S(kn\theta_0|\theta) \quad \text{when} \quad S = \sum_{i=1}^{n} X_i$$

$H_0: \theta \leq 1$

$H_1: \theta > 1$

$n = 10$



Def (8.3.9) A test is unbiased if its power function $\beta(\theta)$ has the property

$\beta(\theta') \geq \beta(\theta'')$ for every $\theta' \in \Theta_1$ and $\theta'' \in \Theta_0$. i.e.,

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \inf_{\theta \in \Theta_1} \beta(\theta).$$

(least upper bound) (greatest lower bound)

(Remark: Testing procedures in last two examples are unbiased.)

Remark: One approach to hypothesis test selection is to focus on tests within a specific class of tests. e.g., consider tests where the size (or level) of the test $\alpha = \sup_{\theta \in \Theta_0} \beta(\theta) =$ max probability of a Type I error is fixed a priori.

(Refer to last two examples.)

Most Powerful Tests (Section 8.3.2)

Def (8.3.11) Let $\mathcal{C}$ be a class of tests for testing $H_0: \theta \in \Theta_0$ versus $H_1: \theta \in \Theta_0^c$.

A test in $\mathcal{C}$, with power function $\beta(\theta)$, is a uniformly most powerful (UMP) class $\mathcal{C}$ test if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $\mathcal{C}$.

(Remark: Refer back to the last two examples.)

Remark: UMP tests do not exist for all applications.

Theorem 8.3.12 (Neyman-Pearson Lemma)

Consider testing $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$. In other words, $H_0: \underline{X} \sim f(\underline{x}|\theta_0)$ vs $H_1: \underline{X} \sim f(\underline{x}|\theta_1)$. The test which rejects H_0 for $\underline{x} \in R = \{\underline{x} : L(\theta_1|\underline{x}) > k\, L(\theta_0|\underline{x})\}$ for some $k \geq 0$ where $\alpha = P_{\theta_0}(\underline{X} \in R)$ is a UMP level α test.

proof: see textbook. (see Ray's notes p. 233-234 of MGB notes)

Def: For a given test with rejection region R, define the test function $\phi(\underline{x})$ by

(aka critical function)

$$\phi(\underline{x}) = I_R(\underline{x}) = \begin{cases} 1, & \text{for } \underline{x} \in R \\ 0, & \text{for } \underline{x} \notin R. \end{cases}$$

Remark: $\beta(\theta) = P_\theta(X \in R) = E_\theta[\phi(\underline{X})]$ $\left[= \int_R \phi(x) f(\underline{x}|\theta)\,dx\right.$ for continuous case.

Corollary: In testing simple vs simple hypotheses, the LRT of level α is UMP among tests of level α.

(See problem 8.24)

Corollary 8.3.13: If T is a sufficient statistic for the family $\{f(x|\theta) : \theta \in \Theta\}$, $\Theta = \{\theta_0, \theta_1\}$, then the Neyman-Pearson Lemma UMP level α test is a function of T.

(Do UMP tests exist for simple vs composite or composite vs composite tests?)

Def (~~similar to~~ see 8.3.16): The family $\{f(\underline{x}|\theta) : \theta \in \Theta\}$, Θ an interval, is said to have a monotone likelihood ratio (MLR) if $\exists$ a unidimensional statistic T such that $\dfrac{L(\theta_2|\underline{x})}{L(\theta_1|\underline{x})}$ is monotone in T $\forall$ $\theta_2 > \theta_1$.

Remark: Ratio above can be either ~~non-decreasing~~ (monotone increasing) or ~~non-increasing~~ (monotone decreasing).

EX: $X_1, \dots, X_n \overset{iid}{\sim}$ Exponential(θ), $\theta > 0$. Then $T = \sum_{i=1}^{n} X_i \sim$ Gamma(n, θ).

$$\frac{L(\theta_2|\underline{x})}{L(\theta_1|\underline{x})} = \frac{\frac{1}{\theta_2^n} e^{-t/\theta_2}}{\frac{1}{\theta_1^n} e^{-t/\theta_1}} = \left(\frac{\theta_1}{\theta_2}\right)^n e^{t\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)}$$

is an increasing function of T for $\theta_2 > \theta_1$ so $\{\text{Gamma}(n, \theta) : \theta \in (0, \infty)\}$ has a MLR ~~test~~.

Theorem (~~similar to~~ see Theorem 8.3.17, Karlin-Rubin Theorem)

Suppose T is a sufficient statistic for the family $\{f(x|\theta) : \theta \in \Theta\}$. Consider testing $H_0: \theta \le \theta_0$ vs $H_1: \theta > \theta_0$.

i) If the family has a ~~non-decreasing~~ (monotone increasing) MLR, then the test with rejection region $R = \{T : T > t_0\}$ is a UMP level α test with $\alpha = P_{\theta_0}(T > t_0)$.

ii) If the family has a ~~non-increasing~~ (monotone decreasing) MLR, then the test with rejection region $R = \{T : T < t_0\}$ is a UMP level α test with $\alpha = P_{\theta_0}(T < t_0)$.

Remark: If in above Theorem $H_0: \theta \ge \theta_0$ vs $H_1: \theta < \theta_0$, theorem remains valid if inequalities in (i) and (ii) are reversed.

Ex: $X_1, \dots, X_n \overset{iid}{\sim}$ Exponential (θ), $\theta > 0$

$H_0: \theta \le \theta_0$

$H_1: \theta > \theta_0$

Sufficient statistic $T = \sum_{i=1}^{n} X_i$.

Since $\exists$ a MLR that is (monotone increasing) ~~non-decreasing~~ in T, the UMP level α test

rejects H_0 for $T > t_0$ where $\alpha = P_{\theta_0}(T > t_0) = \int_{t_0}^{\infty} f_T(t|\theta_0)\,dt$

is computed using $T \sim$ Gamma (n, θ_0).

Remark: Above Theorem holds when $f(x|\theta)$ is a member of a one-parameter exponential family, i.e., $f(x|\theta) = h(x)c(\theta)e^{w(\theta)t(x)}$ where $w(\theta)$ is monotone in θ.

Ex: $X_1, \dots, X_n \overset{iid}{\sim}$ Uniform $[0, \theta]$, $\theta > 0$.

$H_0: \theta \le \theta_0$

$H_1: \theta > \theta_0$

Sufficient statistic is $T = X_{(n)}$. Let $\theta_2 > \theta_1$.

$$\frac{L(\theta_2|\underline{x})}{L(\theta_1|\underline{x})} = \frac{\frac{1}{\theta_2^n} I_{[0,\theta_2]}(t)}{\frac{1}{\theta_1^n} I_{[0,\theta_1]}(t)} = \begin{cases} \left(\frac{\theta_1}{\theta_2}\right)^n, & 0 \le t \le \theta_1 \\ \infty, & \theta_1 < t \le \theta_2 \end{cases}$$

is a (monotone increasing) ~~non-decreasing~~ MLR.

By previous Theorem, $\exists$ a UMP level α test that rejects H_0 for $T = X_{(n)} > t_0$ where

$$\alpha = P_{\theta_0}(X_{(n)} > t_0) = \int_{t_0}^{\theta_0} f_{X_{(n)}}(t|\theta_0)\,dt = \int_{t_0}^{\theta_0} n\left(\frac{t}{\theta_0}\right)^{n-1} \frac{1}{\theta_0}\,dt$$

$$= 1 - \left(\frac{t_0}{\theta_0}\right)^n.$$

Fix α and solve for t_0 to get $t_0 = \theta_0(1-\alpha)^{1/n}$.