MATH 503: Mathematical Statistics

Lecture 6: Intro. to Hypothesis Testing

Reading: Chapter 8

Kimberly F. Sellers

Department of Mathematics & Statistics

Today's Topics

- General framework and set-up
- Critical/rejection region
- Type I and II errors
- Power and power functions
- Steps for solving hypothesis tests
  - Critical region method
  - P-value method
- Most powerful (i.e. best) tests

Intro. to Hypothesis Testing

- ***Setup:*** rv X with pdf/pmf $f(x;\theta), \theta \in \Omega$ unknown.
- ***Goal:*** to resolve a test of the form

$$H_0: \theta \in \omega_0 \text{ vs. } H_1: \theta \in \omega_1$$

where ω_0, ω_1 subsets of Ω; $\omega_0 \cup \omega_1 = \Omega$

- The <u>null hypothesis</u> (denoted H_0) is a claim about one or more populations that is initially assumed true, i.e. “the status quo”.
- The <u>alternative hypothesis</u> (denoted either H_1 or H_a) is the assertion that is contradictory to H_0, i.e. what is to be proven. (often referred to as “researcher’s hypothesis.”)

Note

- Decision rule to take H_0 or H_1 based on sample $X_1, \ldots, X_n$.

- When performing a test, the null hypothesis is rejected in favor of the alternative ONLY if the sample evidence suggests that H_0 is false. Otherwise, we cannot draw that conclusion and therefore continue to believe that H_0 is true.

$\rightarrow$ we either "reject H_0" or "fail to reject H_0"

Critical Region

- A test of H_0 vs. H_1 is based on subset C called the <u>critical region</u>.
- The critical region and its corresponding decision rule is

Reject H_0	if $(X_1, \ldots, X_n) \in C$
Fail to reject H_0	if $(X_1, \ldots, X_n) \notin C$

- We say a critical region C is of size α if
$$\alpha = \max_{\theta \in \omega_0} P_\theta[(X_1, \ldots, X_n) \in C]$$

Type I and Type II error

- $\alpha =$ Type I error $= P$(reject H_0 | H_0 is true)
$$= P_\theta[(X_1, \ldots, X_n) \in C], \text{ for } \theta \in \omega_0;$$
this is also called the significance level of the test.

- $\beta =$ Type II error $= P$(fail to reject H_0 | H_0 false)

Power

- Goal: Type II error to be as small (i.e. power = 1-Type II error to be as big) as possible

- Power $= 1 - \beta = P$(reject H_0 | H_0 false)
  $= P_\theta[(X_1, \ldots, X_n) \in C]$, for $\theta \in \omega_1$

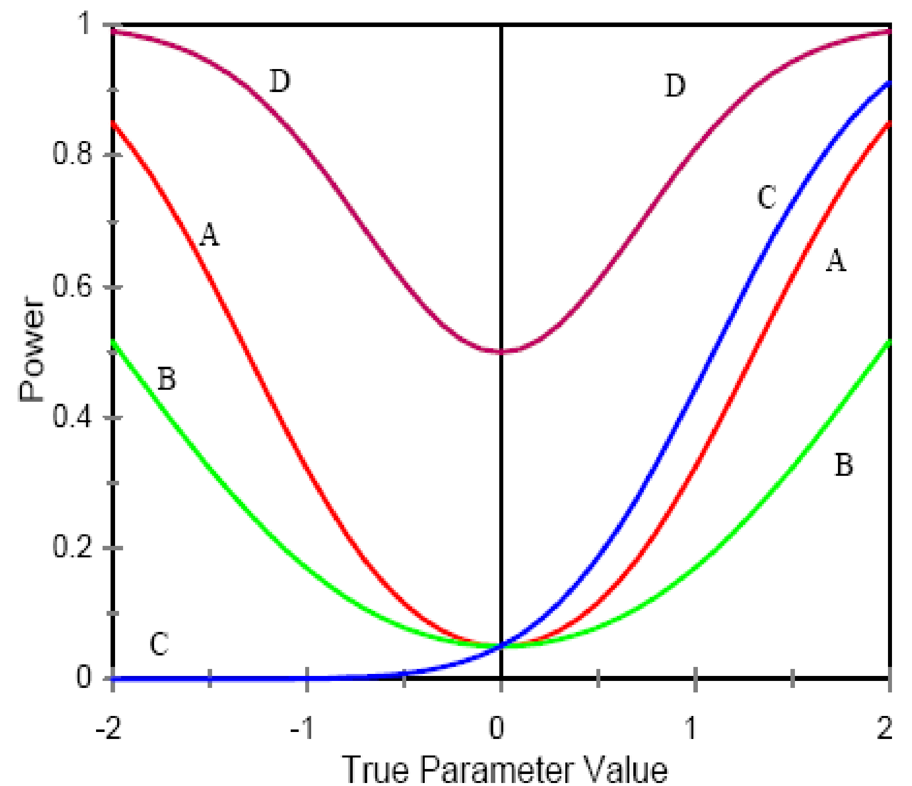
Example 1

Let X have a binomial distribution with the number of trials $n = 10$ and with p either ¼ or ½. The simple hypothesis H_0: $p = ½$ is rejected, and the alternative simple hypothesis H_1: $p = ¼$ is accepted, if the observed value of X, a random sample of size 1, is less than or equal to 3. Find the significance level and the power of the test.

Power Functions

- Power function of a critical region is
$$\gamma_C(\theta) = P_\theta[(X_1, \ldots, X_n) \in C]$$
for θ

- Given two critical regions of size α, C_1 is better than C_2 if $\gamma_{C_1}(\theta) \geq \gamma_{C_2}(\theta) \; \forall \theta$



Example 2

- Consider $X_1, \dots, X_n$ random sample with mean μ and variance $\sigma^2 < \infty$
- Test $H_0: \mu = \mu_0$ vs. $H_1: \mu > \mu_0$
- For n large (by CLT),

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \xrightarrow{d} Z \sim \mathrm{N}(0,1)$$

- Standard normal distribution table provided on Canvas

Example 2 (cont.)

- Decision rule:

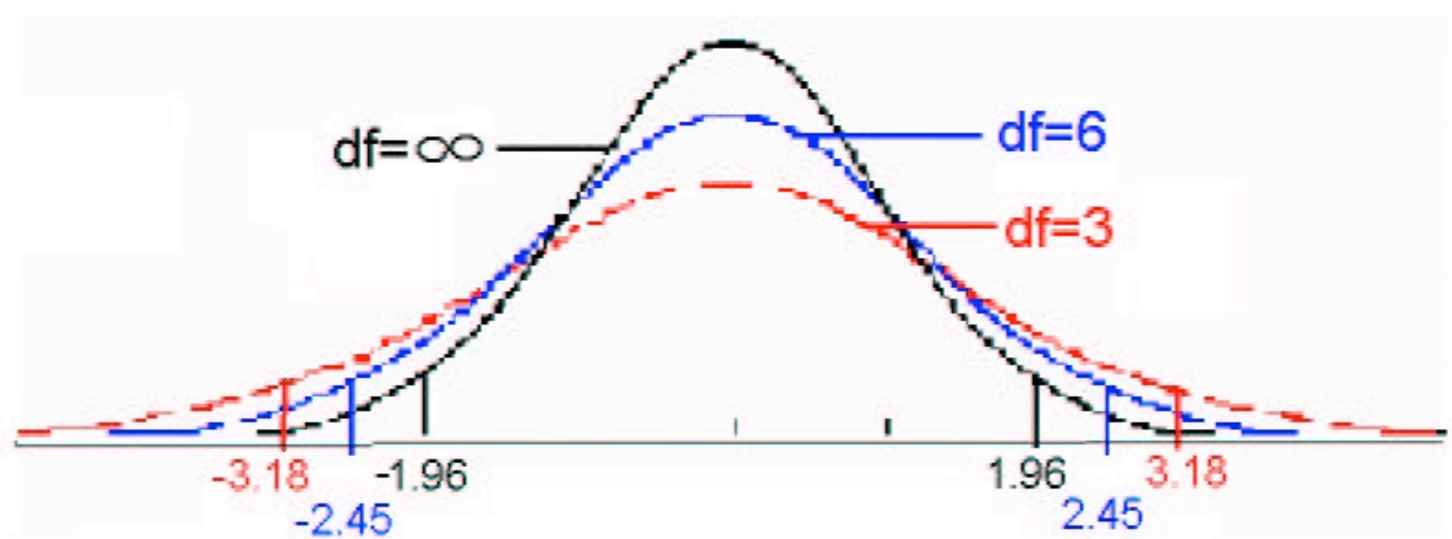
$$\text{Reject } H_0 \text{ if } \quad \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \geq z_\alpha$$

- Power function:

- This is an approximate or asymptotic test

How is $\bar{X}$ distributed?

- For σ known, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$
- For σ unknown, $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$ where $n-1$ is the degrees of freedom (df)



Properties of T_k Distribution

- bell curve with heavier tails than a normal distribution
- k = degrees of freedom (n minus number of estimated parameters), which determine the spread
- for k large enough, T and Z are nearly equivalent
- as $n \to \infty$, $s \to \sigma$ and $T \to \mathrm{N}(0,1)$
- Distribution table provided on Canvas

Example 3

- Now, consider $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ iid
- Test $H_0: \mu = \mu_0$ vs. $H_1: \mu > \mu_0$
- By definition,

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

- Decision rule:

$$\text{Reject } H_0 \text{ if } \quad \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \geq t_{\alpha,n-1}$$

- t critical values generally larger than z critical values $\Rightarrow$ t test conservative relative to large sample (z) test

Example 4

- Consider $X_1, \dots, X_n$ random sample with mean μ and variance $\sigma^2 < \infty$
- Test $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$
- For n large (by CLT),

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \overset{d}{\rightarrow} Z \sim \mathrm{N}(0,1)$$

- Decision rule:

$$\text{Reject } H_0 \text{ if } \left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| \geq z_{\alpha/2}$$

Example 4 (cont.)

- Power function:

General steps to solving hypothesis tests

1. Determine H_0 and H_1. Is it one-tailed or two-tailed?
2. Determine the significance level, α.
3. Compute the test statistic.
   - A test statistic is a function of the sample data used to decide whether or not we reject H_0.
4. Either determine the rejection/critical region associated with α or compute the p-value associated with the test statistic.
   - A p-value tells the probability of getting a test-statistic more extreme than the one computed in this test.

General steps to solving hypothesis tests (cont.)

5. Draw conclusions.
   - For the rejection region method:
     - If the test statistic falls in the rejection region, then we reject H_0.
     - If the test statistic does not fall in the rejection region, then we fail to reject H_0.
   - For the p-value method:
     - If p-value $< \alpha$, then we reject H_0.
     - If p-value $> \alpha$, then we fail to reject H_0.

Normal population with σ known or unknown

- The test is one of the following:

$$H_0: \mu = \mu_0 \text{ vs. } H_1: \mu > \mu_0 \text{ (one-tailed)}$$
$$H_0: \mu = \mu_0 \text{ vs. } H_1: \mu < \mu_0 \text{ (one-tailed)}$$
$$H_0: \mu = \mu_0 \text{ vs. } H_1: \mu \neq \mu_0 \text{ (two-tailed)}$$

- Test statistic:
  - For σ known, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$
  - For σ unknown, $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

Normal population with σ known or unknown (cont.)

- Rejection region:
  - For the test, $H_0: \mu = \mu_0$ vs. $H_1: \mu > \mu_0$, reject if $z \geq z_\alpha$ or $t \geq t_{\alpha;n-1}$ (upper-tailed test).
  - For the test, $H_0: \mu = \mu_0$ vs. $H_1: \mu < \mu_0$, reject if $z \leq z_\alpha$ or $t \leq t_{\alpha;n-1}$ (lower-tailed test).
  - For the test, $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$, reject if $z \leq -z_{\alpha/2}$ or $z \geq z_{\alpha/2}$, or $t \leq -t_{\alpha/2;n-1}$ or $t \geq t_{\alpha/2;n-1}$ (two-tailed test).

Example 5

Lightbulbs of a certain type are advertised as having an average lifetime of 750 hours. A random sample of 50 bulbs is selected to test if the true average lifetime is actually shorter than what is advertised. The average lifetime from the bulbs sampled was 738.44 with a standard deviation of 38.20. What conclusion would be appropriate for a significance level of $\alpha = 0.05$?

Example 5 (cont.)

How would your answer change if $\alpha = 0.01$?

General steps to solving hypothesis tests

1. Determine H_0 and H_1. Is it one-tailed or two-tailed?
2. Determine the significance level, α.
3. Compute the test statistic.
   – A test statistic is a function of the sample data used to decide whether or not we reject H_0.
4. Either determine the rejection/critical region associated with α or compute the p-value associated with the test statistic.
   – A p-value tells the probability of getting a test-statistic more extreme than the one computed in this test.

General steps to solving hypothesis tests (cont.)

5. Draw conclusions.
   - For the rejection region method:
     - If the test statistic falls in the rejection region, then we reject H_0.
     - If the test statistic does not fall in the rejection region, then we fail to reject H_0.
   - For the p-value method:
     - If p-value $< \alpha$, then we reject H_0.
     - If p-value $> \alpha$, then we fail to reject H_0.

Normal population with σ known or unknown

- The test is one of the following:

$$H_0: \mu = \mu_0 \text{ vs. } H_1: \mu > \mu_0 \text{ (one-tailed)}$$

$$H_0: \mu = \mu_0 \text{ vs. } H_1: \mu < \mu_0 \text{ (one-tailed)}$$

$$H_0: \mu = \mu_0 \text{ vs. } H_1: \mu \neq \mu_0 \text{ (two-tailed)}$$

- Test statistic:
  - For σ known, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$
  - For σ unknown, $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

Normal population with σ known or unknown (cont.)

- Compute p-value:
  - For the test, $H_0: \mu = \mu_0$ vs. $H_1: \mu > \mu_0$ (upper-tailed test),

$$\mathrm{p-value} = \begin{cases} P(Z > \text{test statistic}) \text{ for } \sigma \text{ known} \\ P(T > \text{test statistic}) \text{ for } \sigma \text{ unknown} \end{cases}$$

  - For the test, $H_0: \mu = \mu_0$ vs. $H_1: \mu < \mu_0$ (lower-tailed test),

$$\mathrm{p-value} = \begin{cases} P(Z < \text{test statistic}) \text{ for } \sigma \text{ known} \\ P(T < \text{test statistic}) \text{ for } \sigma \text{ unknown} \end{cases}$$

  - For the test, $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$ (two-tailed test),

$$\mathrm{p-value} = \begin{cases} 2P(Z > |\text{test statistic}|) \text{ for } \sigma \text{ known} \\ 2P(T > |\text{test statistic}|) \text{ for } \sigma \text{ unknown} \end{cases}$$

Redo-Example 5

Lightbulbs of a certain type are advertised as having an average lifetime of 750 hours. A random sample of 50 bulbs is selected to test if the true average lifetime is actually shorter than what is advertised. The average lifetime from the bulbs sampled was 738.44 with a standard deviation of 38.20. What conclusion would be appropriate for a significance level of $\alpha = 0.05$?

How would your answer change if $\alpha = 0.01$?

Binomial Approximation

- If we think of $Y = \sum_{i=1}^{n} X_i$ where

$$X_i = \begin{cases} 1, \text{with probability } p \\ 0, \text{with probability } q = 1 - p \end{cases}$$

  are Bernoulli trials, then for "n large" (i.e., $np \geq 10$ and $nq \geq 10$), this is a special case of the Central Limit Theorem, thus

$$Y = \sum_{i=1}^{n} X_i \sim \mathrm{N}(\mu = np, \sigma^2 = npq)$$

- Continuity correction is a procedure that helps to better approximate associated probabilities.

Example 6

Let p equal the proportion of drivers who use a seat belt in a state that does not have a mandatory seat belt law. It was claimed that $p = 0.14$. An advertising campaign was conducted to increase this proportion. Two months after the campaign, $y = 104$ out of a random sample of $n = 590$ drivers were wearing their seatbelts. Was the campaign successful?

a) Define the null and alternative hypotheses.
b) Define a critical region with an $\alpha = 0.01$ significance level.
c) Determine the approximate *p*-value and state your conclusion.

Best Critical Region (ie Best Test)

- Goal: to create a "best test"
- Suppose we have rv X with pdf/pmf $f(x;\theta)$, and want to test $H_0\!: \theta = \theta'$ vs. $H_1\!: \theta = \theta''$, where $\theta \in \Omega = \{\theta', \theta''\}$

- Let C denote a subset of the sample space. Then we say that C is a best critical region of size α for testing the simple hypothesis H_0 vs. H_1 if
  1. $P_{\theta'}[(X_1, \ldots, X_n) \in C] = \alpha$, and
  2. for every subset A of sample space, $P_{\theta'}[(X_1, \ldots, X_n) \in A] = \alpha$ implies
     $P_{\theta''}[(X_1, \ldots, X_n) \in C] \geq P_{\theta''}[(X_1, \ldots, X_n) \in A]$

How do we determine the best critical region?

Neymann-Pearson Thm.: Let $X_1, \dots, X_n$ (n, *a* positive fixed integer) denote a random sample from a distribution that has pdf/pmf $f(x;\theta)$. Then the likelihood of $X_1, \dots, X_n$ is

$$L(\theta;\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i;\theta).$$

Let θ' and θ'' be distinct fixed values of θ s.t. $\Omega = \{\theta\colon \theta = \theta', \theta''\}$, and let k be a positive number.

Neymann-Pearson Thm. (cont.)

Let C be a subset of the sample space s.t.

a) $\frac{L(\theta';\boldsymbol{x})}{L(\theta'';\boldsymbol{x})} \leq k$, for each point $\boldsymbol{x} \in C$

b) $\frac{L(\theta';\boldsymbol{x})}{L(\theta'';\boldsymbol{x})} \geq k$, for each point $\boldsymbol{x} \in C^c$

c) $\alpha = P_{H_0}[\boldsymbol{X} \in C]$

Then C is a best critical region of size α for testing the simple hypothesis $H_0: \theta = \theta'$ vs. $H_1: \theta = \theta''$.

Example 7

Let $X_1, \dots, X_{10}$ be a random sample of size 10 from a normal distribution $\mathrm{N}(0, \sigma^2)$. Find a best critical region of size $\alpha = 0.05$ for testing $H_0: \sigma^2 = 1$ vs. $H_1: \sigma^2 = 2$. Is this a best critical region of size $\alpha = 0.05$ for testing $H_0: \sigma^2 = 1$ vs. $H_1: \sigma^2 = 4$. Against $H_1: \sigma^2 = \sigma_1^2 > 1$?

Example 8

Let $X_1, \dots, X_n$ be iid with pmf $f(x;p) = p^x(1-p)^{1-x},\ \ x = 0,1.$ Show that $C = \{(x_1, \dots, x_n): \sum_{i=1}^n x_i \le c\}$ is a best critical region for testing $H_0: p = \frac{1}{2}$ against $H_1: p = \frac{1}{3}$. Use the CLT to find n and c so that approximately $P_{H_0}(\sum_{i=1}^n X_i \le c) = 0.10$ and $P_{H_1}(\sum_{i=1}^n X_i \le c) = 0.80$.

Neymann-Pearson Corollary

Let C be the critical region of the best test of $H_0: \theta = \theta'$ vs. $H_1: \theta = \theta''$. Suppose the significance level of the test is α. Let $\gamma_C(\theta'') = P_{\theta''}[\boldsymbol{X} \in C]$ denote the power of the test. Then $\alpha \leq \gamma_C(\theta'')$.