APMA 1655 Chapter 10 Notes: Hypothesis Testing

Basics of Hypothesis Testing

- Null hypothesis (H_0): claim to be tested
- Alternative hypothesis (H_a): the converse of the null hypothesis
- Test statistic: is a function of the sample measurements on which the statistical decision will be based, like an estimator
- Rejection region (RR): specifies the values of the test statistic for which the null hypothesis is to be rejected in favor of the alternative hypothesis (given a particular sample, if the computed value of the test statistic falls in RR, reject H_0 and accept H_a, otherwise accept H_0)
- Approach One: report testing result via p-value
  - P-value: probability of observing test statistics that are as extreme or more extreme than the present empirical data assuming H_0 is valid
- Approach Two: Report testing result via rejection/acceptance and significance level
  - Determine rejection region
  - Report Testing Result

Errors and Power

- Type 1 Error: made if H_0 is rejected when H_0 is true; probability of type 1 error is α, which is called the level of the test; $P(RR|H_0) = \alpha$
- Type 2 Error: made if H_0 is accepted when H_a is true; probability of type 2 error is denoted by β; $\beta = 1 - P(RR|H_a)$
- Power of test: defined to be $1 - \beta = P(RR|H_a)$

Testing Means of Normals

- Let $\{X_1, ..., X_n\}$ be iid samples from $N(\mu, \sigma^2)$, where σ^2 is known but μ is unknown; want to perform hypothesis testing on μ; there are three scenarios:
- One-sided test: $H_0 : \mu = \mu_0$, $H_a : \mu > \mu_0$
  - Test statistic: sample mean $\bar{X}$, ($\bar{x}$ is the observed sample mean)
  - P-Value Approach: Under H_0, $\bar{X}$ is $N(\mu_0, \sigma^2/n)$:
$$\text{P-value} = P(\bar{X} \geq \bar{x}) = \phi(-\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}})$$
  - Rejection Region: $\text{RR} = \{\bar{X} > c\}$; want $P(RR) = \alpha$ under H_0; this leads to:
$$c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$
  - Conclusion:
    - P-value $\leq \alpha$ or $\bar{x} \in$ RR: reject H_0 and accept H_a because of statistically significant evidence in data
    - P-value $> \alpha$ or $\bar{x} \notin$ RR: accept H_0 and reject H_a because evidence in data not statistically significant
  - Power: by definition is $P(RR|H_a)$ but depends on value of μ; say $\mu = \mu_a$, then:
$$\text{Power} = P(\hat{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}|\mu = \mu_a) = \phi(\frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} - z_\alpha)$$
- One-Sided Test; $H_0 : \mu = \mu_0$, $H_a : \mu < \mu_0$
  - Test statistic: sample mean $\bar{X}$, ($\bar{x}$ is the observed sample mean)
  - P-Value Approach: Under H_0, $\bar{X}$ is $N(\mu_0, \sigma^2/n)$:
$$\text{P-value} = P(\bar{X} \leq \bar{x}) = \phi(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}})$$
  - Rejection Region: $\text{RR} = \{\bar{X} > c\}$; want $P(RR) = \alpha$ under H_0; this leads to:
$$c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$$
  - Conclusion:
    - P-value $\leq \alpha$ or $\bar{x} \in$ RR: reject H_0 and accept H_a because of statistically significant evidence in data
    - P-value $> \alpha$ or $\bar{x} \notin$ RR: accept H_0 and reject H_a because evidence in data not statistically significant
  - Power: by definition is $P(RR|H_a)$ but depends on value of μ; say $\mu = \mu_a$, then:
$$\text{Power} = P(\hat{X} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}|\mu = \mu_a) = \phi(\frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}} - z_\alpha)$$
- Two-Sided Test; $H_0 : \mu = \mu_0$, $H_a : \mu \neq \mu_0$
  - Test statistic: sample mean $\bar{X}$, ($\bar{x}$ is the observed sample mean)
  - P-Value Approach: Under H_0, $\bar{X}$ is $N(\mu_0, \sigma^2/n)$:
$$\text{P-value} = P(|\bar{X} - \mu_0| \geq |\bar{x} - \mu_0|) = 2\phi(-\frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}})$$
  - Rejection Region: $\text{RR} = \{|\bar{X} - \mu_0| > c\}$; want $P(RR) = \alpha$ under H_0; this leads to:
$$c = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$
  - Conclusion:
    - P-value $\leq \alpha$ or $\bar{x} \in$ RR: reject H_0 and accept H_a because of statistically significant evidence in data
    - P-value $> \alpha$ or $\bar{x} \notin$ RR: accept H_0 and reject H_a because evidence in data not statistically significant

- Power: by definition is $P(RR|H_a)$ but depends on value of μ; say $\mu = \mu_a$, then:

$$\text{Power} = P(|\bar{X} - \mu_0| > z_{\alpha/2}\frac{\sigma}{\sqrt{n}}|\mu = \mu_a) =$$

$$\phi(\frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} - z_{\alpha/2}) + \phi(\frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}} - z_{\alpha/2})$$

Extensions to Large Samples: Testing Mean

- all discussions on testing means of normals can be extended to large sample test where the test statistic is approximately normally distributed
- For example: let $\{X_1, ..., X_n\}$ be iid samples from some population distribution with unknown mean μ
  - One sided test: $H_0 : \mu - \mu_0$, $H_a : \mu > \mu_0$
  - One sided test: $H_0 : \mu - \mu_0$, $H_a : \mu < \mu_0$
  - Two sided test: $H_0 : \mu - \mu_0$, $H_a : \mu \neq \mu_0$
- Test statistic is sample mean $\bar{X}$, by central limit theorem, $\bar{X}$ is approximately $N(\mu, \sigma^2/n)$; all previous formulae are valid; when σ is unknown, can use sample standard deviation s in place of σ

Extensions to Large Samples: Testing Proportion

- let $\{X_1, ..., X_n\}$ be iid Bernoulli samples such that $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ where p is uknown
  - One sided test: $H_0 : p = p_0$, $H_a : p > p_0$
  - One sided test: $H_0 : p = p_0$, $H_a : p < p_0$
  - Two sided test: $H_0 : p = p_0$, $H_a : p \neq p_0$
- Test statistic is sample mean $\bar{X}$, by central limit theorem, $\bar{X}$ is approximately $N(p, p(1-p)/n)$; all previous formulae are valid except for power, with p_0 in place of μ_0 and $\sqrt{p_0(1-p_0)}$ in place of σ

Neyman-Pearson Lemma

- Setup: Suppose $\{X_1, ..., X_n\}$ are iid samples with common density $f(x)$, consider the following hypotheses $H_0 : f(x) = f_0(x)$, $H_a : f(x) = f_a(x)$ where f_0 and f_a are 2 given densities
- Question: fix an arbitrary significance level α, among all possible rejection regions RR such that the type 1 error satisfies $P(RR|H_0) \leq \alpha$, which one gives the minimal type 2 error or maximal power?
- Theorem: Neyman-Pearson Lemma
  - Let $\alpha \in (0, 1)$ be any given significance level
  - For each $b > 0$ define a rejection region $RR_b \subseteq \mathbb{R}^n$ of form

    $$RR_b = (x_1, ..., x_n) : \frac{f_a(x_1)...f_a(x_n)}{f_0(x_1)...f_0(x_n)} \geq b$$

  - Let b^* be such that $P(RR_{b^*}|H_0) = \alpha$. Then RR_{b^*} yields the minimal type 2 error among all rejectio regions whose type 1 error is bounded by α