Chapter 4: Some Elementary Statistical Inferences

4.1 Sampling and Statistics

- (Random Sample) The random variables $X_1, \cdots, X_n$ constitute a random sample on a random variable X if $X_1, \cdots, X_n$ are iid with the same distribution as that of X. Because their distributions are identical, they have the same expected values (means) and variances, i.e., $\mathrm{E}(X_1) = \cdots = \mathrm{E}(X_n) = \mu$ and $\mathrm{V}(X_1) = \cdots = \mathrm{V}(X_n) = \sigma^2$.
  - In theoretical statistics, we use random variables to represent observations (i.e., data). Then, we can use probability to study their properties.
  - In applied statistics, we use values. We look at their numerical results.
- (Statistic) A statistic is a function of data. It becomes a real number after you have data.
  - Before collecting the data, a statistic is a random variable. In theoretical statistics, we treat it as a random variable.
  - After collecting the data, a statistic is a real number. In applied statistics, we treat it as a number.

4.1.1. Point Estimators. Three main problems in statistics.

- Point estimation. The answer is a real number. There are three terms
  - Estimation. The entire method for the formula. It is the most important step in the derivation of the three main problems.
  - Estimator. The formula (must be a statistic).
  - Estimate. A value. After you have data, an estimator becomes an estimate.
- Confidence interval. The answer is an interval, such as $a \pm b$ or $[L, U]$.
- Hypothesis testing. The answer is *True* or *False*.

Definition 1 *Let $T = T(X_1, \cdots, X_n)$ be a statistic and we use it to estimate θ. If $\mathrm{E}(T) = \theta$, then we call it is unbiased; otherwise, we called $\mathrm{E}(T) - \theta$ as the bias of T.*

Criticism: T^2 is not an unbiased estimator of θ^2 even if T is an unbiased estimator of θ.

If $X_1, \cdots, X_n$ are random sample with common PDF (or PMF) $f(x)$ and CDF $F(x)$, the the joint PDF (or PMF) is

$$f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$$

and the joint CDF is

$$F_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = \prod_{i=1}^{n} F(x_i).$$

In addition, if a parameter is contained in $f(x)$ so that we can write $f(x) = f_\theta(x)$, then the likelihood function is defined by their joint PDF (or PMF) as

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i).$$

- The likelihood function is identical to the joint PDF or PMF.

- The focus of the likelihood function is the parameter but not the distribution.
- The most important method in statistics *maximum likelihood.* It provides point estimator of θ by maximizing the likelihood function.
- A main step in the maximum likelihood approach is the derivation of the maximizer. One method is the usage of derivatives.
- Maximum likelihood approach has also been extended to a case with more than one parameter. Then, we need to use partial derivatives (or gradient vector).
- If $\hat{\theta}$ is the MLE of θ, then for any continuous function $g(\cdot)$, $g(\hat{\theta})$ is also the MLE of $g(\theta)$.

Example 4.1.1 Suppose $X_1, \cdots, X_n$ are identically and independently collected from $Exp(\theta)$. The PDF of X_i is $f(x) = \theta e^{-\theta x}$. The likelihood function of θ is

$$L(\theta) = \prod_{i=1}^{n} f(X_i) = \prod_{i=1}^{n} (\theta e^{-\theta X_i}) = \theta^n e^{-\theta \sum_{i=1}^{n} X_i} = \theta^n e^{-n\theta \bar{X}},$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$ is called the sample mean. The loglikelihood function of θ is

$$\ell(\theta) = \log L(\theta) = n \log(\theta) - n\theta \bar{X}.$$

Taking derivative with respect to θ, we obtain the estimating equation (EE) as

$$\dot{\ell}(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \frac{n}{\theta} - n\bar{X}.$$

Solve it for θ, we obtain the maximum likelihood estimator (MLE) of θ as

$$\hat{\theta} = \frac{1}{\bar{X}}.$$

Note that the right side only depends on data. It will be a real value if data are provided. This is an important property to check whether the solution makes sense.

Based on the data given by: 359, 413, 25, 130, 90, 50, 50, 487, 102, 194, 55, 74, 97, we obtain $\bar{x} = 163.54$. Then, the maximum likelihood estimate (MLE) of θ is $\hat{\theta} = 1/163.54 = 0.006115$.

Since $\mathrm{E}(\bar{X}^{-1}) \neq \theta$, $\hat{\theta}$ is a biased estimator of θ.

Note: If I ask you maximum likelihood estimation, you need all of those. If I ask you maximum likelihood estimator, you need to provide $\hat{\theta} = 1/\bar{X}$. If I ask you maximum likelihood estimate, you need to provide 0.006115.

Example 4.1.2. Let X be $Bernoulli(\theta)$. Then, X can only be 0 or 1. Let $\theta = P(X = 1)$. Then, the PMF can be expressed as $f(x) = \theta^x (1-\theta)^{1-x}$. We write $X \sim Bernoulli(\theta)$. Suppose that $X_1, \cdots, X_n \sim^{iid} Bernoulli(\theta)$. Then, the likelihood function of θ is

$$L(\theta) = \prod_{i=1}^{n} \theta^{X_i} (1-\theta)^{1-X_i} = \theta^{n\bar{X}} (1-\theta)^{n(1-\bar{X})}.$$

The loglikelihood function of θ is

$$\ell(\theta) = \log L(\theta) = n\bar{X} \log(\theta) + n(1-\bar{X}) \log(1-\theta).$$

The estimating equation is

$$\dot{\ell}(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \frac{n\bar{X}}{\theta} - \frac{n(1-\bar{X})}{1-\theta} = 0 \Rightarrow \hat{\theta} = \bar{X}.$$

Since $\mathrm{E}(\bar{X}) = \theta$, $\hat{\theta}$ is an unbiased estimator of θ.

Example 4.1.3. Let $X_1, \cdots, X_n$ be iid from $N(\mu, \sigma^2)$. Then, the PDF is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Let $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\mu, \sigma^2)$. The likelihood function of $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} = (\frac{1}{\sqrt{2\pi}})^n (\frac{1}{\sigma^2})^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(\bar{X}-\mu)^2 + (X_i - \bar{X})^2]}.$$

The loglikelihood function of $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}[n(\bar{X}-\mu)^2 + \sum_{i=1}^n (X_i - \bar{X})^2].$$

Taking derivatives, we have

$$\dot{\ell}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2} \end{pmatrix} = \begin{pmatrix} \frac{n(\bar{X}-\mu)}{\sigma^2} \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}[n(\bar{X}-\mu)^2 + \sum_{i=1}^n (X_i - \bar{X})^2] \end{pmatrix}.$$

Solving $\dot{\ell}(\boldsymbol{\theta}) = 0$, we obtain the MLE of μ as

$$\hat{\mu} = \bar{X}$$

and the MLE of σ^2 as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2.$$

Based on the data given by the textbook (Page 229), we have $n = 24$, $\bar{X} = 53.92$ and $n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2 = 97.25$. We obtain the maximum likelihood estimate of μ as $\hat{\mu} = 53.92$ and $\hat{\sigma}^2 = 97.25$.

Note: There is another estimator of σ^2. It is given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

We call S^2 the sample variance and S the standard error (or sample standard deviation). We can show that $\mathrm{E}(S^2) = \sigma^2$. Then, $\hat{\sigma}^2$ is a biased estimator of σ^2.

Example 4.1.4. Let $X_1, \cdots, X_n$ be iid from uniform $[0, \theta]$. The PDF is

$$f(x) = \frac{1}{\theta} I(0 \le x \le \theta) = \begin{cases} 1/\theta, & 0 \le x \le \theta, \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood function is

$$\begin{aligned} L(\theta) = & \prod_{i=1}^n f(X_i) = \prod_{i=1}^n \frac{1}{\theta} I(0 \le X_i \le \theta) \\ = & \frac{1}{\theta^n} I(0 \le \min(X_i) \le \max(X_i) \le \theta) \\ = & \frac{1}{\theta^n} I(0 \le X_{(1)} \le X_{(n)} \le \theta) \\ = & \frac{1}{\theta^n} I(0 \le X_{(1)}) I(X_{(n)} \le \theta). \end{aligned}$$

where $X_{(1)} = \min(X_i)$ and $X_{(n)} = \max(X_i)$.

Now, we look at the MLE. To make $L(\theta)$ large, we need to make θ small, but θ cannot be lower than $X_{(n)}$. Therefore,

$$\hat{\theta} = X_{(n)} = \max(X_i).$$

Note: We cannot use derivative to find the maximum of the likelihood function. This example introduces an important method to find the MLE.

We next compute the CDF and PDF of $X_{(n)}$. We have a trick. Let $F(x)$ be the CDF of X. Then, $F(x) = x/\theta$ if $0 \le x \le \theta$. The CDF of $X_{(n)}$ is

$$\begin{aligned} F_n(x) =& P(X_{(n)} \le x) \\ =& P(X_1, X_2, \cdots, X_n \le x) \\ =& \prod_{i=1}^{n} P(X_i \le x) \\ =& F^n(x) \\ =& \frac{x^n}{\theta^n}. \end{aligned}$$

The PDF is

$$f_n(x) = \frac{dF_n(x)}{dx} = \frac{nx^{n-1}}{\theta^n}.$$

Thus,

$$\mathrm{E}(X_{(n)}) = \int_0^\theta x f_n(x) dx = \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n\theta}{n+1}$$

and

$$\mathrm{E}(X_{(n)}^2) = \int_0^\theta x^2 f_n(x) dx = \frac{n}{\theta^n} \int_0^\theta x^{n+1} dx = \frac{n\theta^2}{n+2}.$$

We have

$$\mathrm{V}(X_{(n)}) = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

Note: The distribution of the MLE is not normal. This is a nice example to be evaluated in the future.

Example: Let $X_1 \cdots, X_n \sim^{iid} Poisson(\theta)$. The PMF of the Poisson distribution if

$$f(x) = \frac{\theta^x}{x!} e^{-\theta}.$$

The likelihood function is the joint PMF, which is

$$\begin{aligned} L(\theta) =& \prod_{i=1}^{n} \frac{\theta^{X_i}}{X_i!} e^{-\theta} \\ =& (\prod_{i=1}^{n} \frac{1}{X_i!})(\theta^{\sum_{i=1}^n X_i})(e^{-n\theta}) \\ =& (\prod_{i=1}^{n} \frac{1}{X_i!})(\theta^{n\bar{X}})(e^{-n\theta}). \end{aligned}$$

We still study the log-likelihood function (i.e., the logarithm of the likelihood function), which is

$$\ell(\theta) = \log L(\theta) = -\log(\prod_{i=1}^{n} \frac{1}{X_i!}) + n\bar{X} \log\theta - n\theta.$$

By

$$\dot{\ell}(\theta) = \frac{n\bar{X}}{\theta} - n = 0$$

we obtain the MLE of θ as

$$\hat{\theta} = \bar{X}.$$

4.2 Confidence Intervals

I am going to focus on the first two examples and quickly go over other examples.

Definition 2 *Suppose that $X_1, \cdots, X_n$ are random variables (or data). Let $L = L(X_1, \cdots, X_n)$ and $U = U(X_1, \cdots, X_n)$ be statistics. For any $\alpha \in (0,1)$. We say that the interval $[L, U]$ is $(1-\alpha)\%$ confidence interval for θ is*

$$P_\theta[\theta \in (L, U)] = 1 - \alpha,$$

where $1-\alpha$ is called the confidence level or confidence coefficient.

In confidence interval problems, we need to understand: (a) confidence level, (b) coverage probabilities, (c) length of the confidence interval. Since we need to solve both L and U based on one equation, the length of the confidence interval must be considered. The best interval should have the shortest length.

Examples 4.2.1. and 4.2.2. Suppose $X_1, \cdots, X_n$ are iid normal distributed. We use lower case to represent data after they are collected. We use upper case to represent data before they are collected. For example, $x_1, \cdots, x_n$ are observed values of $X_1, \cdots, X_n$. We write

$$X_1, \cdots, X_n \sim^{iid} N(\mu, \sigma^2).$$

We also have the observed value of the sample mean $\bar{x} = \sum_{i=1}^n x_i/n$ and the observed value of the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$. Then, s is the observed value of the sample standard deviation. We have

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$$

Thus,

$$P(-z_{\frac{\alpha}{2}} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le z_{\frac{\alpha}{2}}) = 1 - \alpha.$$

With probability $1-\alpha$, there is

$$-z_{\frac{\alpha}{2}} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le z_{\frac{\alpha}{2}}$$

which is equivalent to say that with probability $1-\alpha$ there is

$$\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}.$$

We may take the above as the confidence interval, i.e, when σ is known.

We have the following formula: suppose $x_1, \cdots, x_n$ are iid observations of a normal population and assume the standard deviation σ is known, then the $1-\alpha$ level confidence interval for μ is

$$\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = [\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}].$$

The interpretation of the confidence interval is that if we repeat the procedure many many times, with probability $1-\alpha$ the above confidence interval contains the true value of μ.

An often asked question is about the length of confidence interval. How large is the sample size n so that the $1-\alpha$ level confidence interval is less than w. Note that the length of the $1-\alpha$ level confidence interval is

$$2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}.$$

Thus, we have

$$2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le w \Rightarrow n \ge (2z_{\frac{\alpha}{2}}\frac{\sigma}{w})^2 = \frac{4z_{\frac{\alpha}{2}}^2\sigma^2}{w^2}.$$

Modification 1. Note that the previous formula requires known σ^2. If it is unknown, then we can replace σ^2 by s^2, leading the large sample confidence interval for μ as

$$\bar{X} - z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \le \mu \le \bar{X} + z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}.$$

This is recommend if n is large (e.g., $n \ge 40$).

Modification 2. If n is small, then one suggests to replace $z_{\alpha/2}$ by $t_{\alpha.2,n-1}$, leading to

$$\bar{x} \pm t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}.$$

Theoretical foundation. Suppose we observed $x_1, \cdots, x_n$ from a normal population $N(\mu, \sigma^2)$. Then

- $$\sum_{i=1}^n [(X_i - \mu)^2] \sim \sigma^2\chi_n^2$$
- $$(n-1)S^2 = \sum_{i=1}^n [(X_i - \bar{X})^2] \sim \sigma^2\chi_{n-1}^2.$$
- $$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

  and $\bar{X}$ and S^2 are independent.
- Therefore, we have
  $$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}.$$
- We denote $\chi_{\alpha,\nu}^2$ as the upper probability of χ^2 distribution with ν degrees of freedom.
- We denote $t_{\alpha,\nu}$ as the upper probability of t-distribution with ν degrees of freedom.

Coverage probability. Suppose that we use

$$\bar{X} \pm t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}$$

to compute 95% confidence interval for μ. Theoretically, we need to evaluate the formulation of the coverage probability. It is given by

$$P(\text{Coverage}) = P_{\mu,\sigma^2}(\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}).$$

This is the probability for the confidence interval to contain the true value. Generally, we say that the confidence interval is correct if it contains the true value of μ, or incorrect otherwise. Equivalently, we have

$$P(\text{Coverage}) = P(-t_{\frac{\alpha}{2},n-1} \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le -t_{\frac{\alpha}{2},n-1}) = 1 - \alpha.$$

We want to make the value identical to (or close to) $1 - \alpha$. We claim the formulation is bad if it is too high or too low. Based on the above result, we conclude that the formulation of t-confidence interval

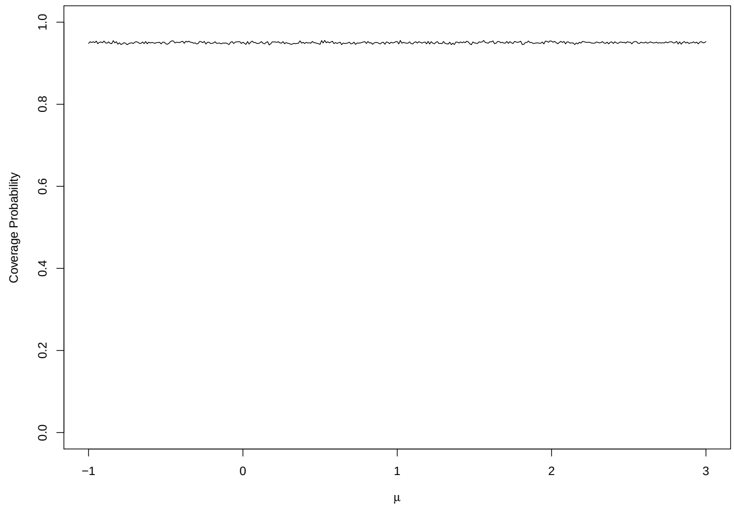

Figure 1: Coverage probability of the t-confidence intervals as functions of μ when $n = 10$ and $\sigma^2 = 1$.

is good. In this problem, I evaluate the properties of coverage probabilities and the result is display in Figure 1.

Example 4.2.3 (Confidence interval for binomial proportion). It is a large sample confidence interval (e.g., $np > 10$ and $n(1-p) > 10$). Suppose $X \sim Bin(n,p)$ and X is observed. The estimate of p is $\hat{p} = X/n$ with

$$\hat{p} \sim^{approx} N(p, \frac{p(1-p)}{n}).$$

Approximately, we have

$$P(-z_{\frac{\alpha}{2}} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq z_{\frac{\alpha}{2}}) \approx 1-\alpha.$$

Solve the inequality

$$-z_{\frac{\alpha}{2}} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq z_{\frac{\alpha}{2}}.$$

We have

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}.$$

Note that the left and the right are not statistics. We use the $1-\alpha$ level confidence interval for p as

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

This is called the Wald confidence interval. I also calculate the covarage probability of the Wald confidence interval by simulations. The result is displayed in Figure 2. Since the curve is not alway close to 0.95. The formulation may not be correct.

4.2.1. Confidence intervals for difference in means.

Assume we observed

$$X_1, X_2, \cdots, X_{n_1} \sim^{iid} N(\mu_1, \sigma_1^2)$$

and

$$Y_1, Y_2, \cdots, Y_{n_2} \sim N(\mu_2, \sigma_2^2),$$

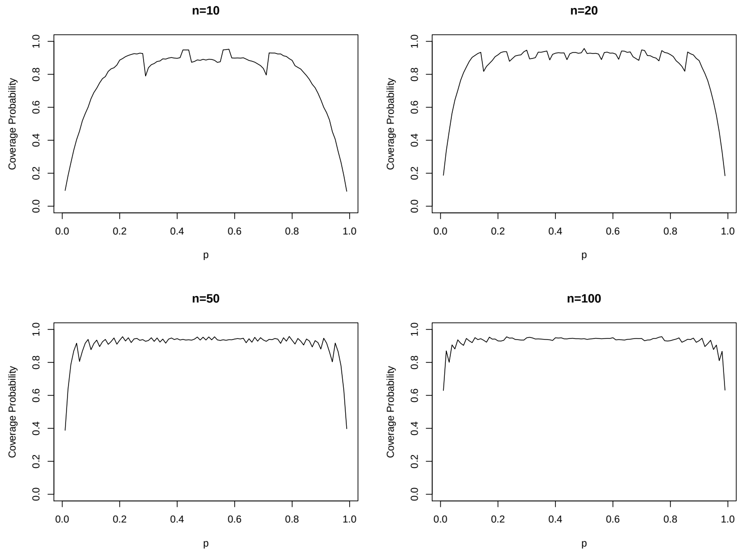

Figure 2: Coverage probability of the t-confidence intervals as functions of μ when $n = 10$ and $\sigma^2 = 1$.

where σ_1^2 and σ_2^2 are known. Then,

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$$

and

$$\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \sim N(\mu_2, \frac{\sigma_2^2}{n_2}).$$

Then,

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}).$$

Case 1: Suppose that σ_1^2 and σ_2^2 are known.

Write $\bar{x}$ and $\bar{y}$ are observed values of $\bar{X}$ and $\bar{Y}$ respectively. Then, the $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{X} - \bar{y}) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Case 2: Large Sample Case.

When σ_1^2 and σ_2^2 are unknown, but both n_1 and n_2 are large (e.g. $m, n > 40$), then we approximately have

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim^{approx} N(0, 1).$$

Then, the $(1 - \alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{x} - \bar{y}) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Case 3: Pooled t-confidence interval.

Assume $\sigma_1^2 = \sigma_2^2$. Let

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

and write s_p^2 as the observed value of S_p^2. Then,

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2(1/n_1 + 1/n_2)}} \sim t_{n_1+n_2-2}.$$

Thus, the $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is

$$\bar{x} - \bar{y} \pm t_{\frac{\alpha}{2}, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Additional. Confidence interval and test for variance ratio. In addition, we have

$$F^* = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{m-1,n-1}.$$

Thus, the $(1-\alpha)100\%$ confidence interval for σ_1^2/σ_2^2 is

$$[\frac{s_1^2/s_2^2}{F_{\alpha/2,m-1,n-1}}, \frac{s_1^2/s_2^2}{F_{1-\alpha/2,m-1,n-1}}].$$

To test

$$H_0 : \sigma_1^2 = \sigma_2^2 \leftrightarrow H_a : \sigma_1^2 \neq \sigma_2^2,$$

We reject H_0 and conclude H_a if

$$\frac{s_1^2}{s_2^2} > F_{\alpha/2,m-1,n-1}$$

or

$$\frac{s_1^2}{s_2^2} < F_{1-\alpha/2,m-1,n-1}.$$

To check value in the table, we need an important property. If $F \sim F_{m,n}$, then $1/F \sim F_{n,m}$. This implies that

$$P(F_{m,n} < c) = P(F_{n,m} > 1/c)$$

which gives

$$F_{\alpha,m,n} = 1/F_{1-\alpha,n,m}$$

where $F_{\alpha,m,n}$ represents the upper α quantile of the F-distribution with m and n degrees of freedom respectively. For example, if we know

$$F_{0.05,10,8} = 3.35$$

then we have

$$F_{0.95,8,10} = \frac{1}{3.35} = 0.2985.$$

4.2.2. Confidence intervals for difference in proportions.

Assume, we have data

$$X \sim Bin(n_1, p_1)$$

and

$$Y \sim Bin(n_2, p_2),$$

and X and Y are independent. Let $\hat{p}_1 = X/m$ and $\hat{p}_2 = Y/n$. Then,

$$\hat{p}_1 - \hat{p}_2 \sim^{approx} N(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}).$$

Since we can estimate the variance

$$\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}$$

by

$$\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2},$$

the large-sample $(1-\alpha)100\%$ confidence interval for p_1-p_2 is

$$\hat{p}_1-\hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

4.4 Order Statistics

Let $X_1,\cdots,X_n$ be iid continuous random variables with common PDF $f(x)$ and CDF $F(x)$. Let $X_{(1)},\cdots,X_{(n)}$ be the order statistics. Then, the joint PDF of $X_{(1)},\cdots,X_{(n)}$ is

$$g(y_1,\cdots,y_n)=n!\prod_{i=1}^{n} f(y_i)$$

for $y_1\le y_2\le\cdots\le y_n$.

The marginal PDF of $X_{(i)}$ is

$$g_i(y_i)=\frac{n!}{(k-1)!(n-k)!}[F(y_i)]^{i-1}[1-F(y_i)]^{n-i}f(y_i).$$

The marginal PDF of $X_{(i)}$ and $X_{(j)}$ with $i<j$ is

$$g_{ij}(y_i,y_j)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(y_i)]^{i-1}[F(y_j)-F(y_i)]^{j-i-1}[1-F(y_j)]^{n-j}f(y_i)f(y_j)$$

if $y_i\le y_j$.

We call $X_{([qn])}$ is q-th quantile of $X_1,\cdots,X_n$, where $[\cdot]$ is the function of the integer part. The median is $X_{([n/2])}$.

As $n\to\infty$ for $0<q_1<1$, we have

$$\sqrt{n}[X_{([qn])}-x_q]\xrightarrow{D} N(0,\frac{q(1-q)}{f^2(x_q)}),$$

where $x_q=F^{-1}(q)$.

As $n\to\infty$, for $0<q_1<q_2<1$, we have

$$\sqrt{n}\left[\begin{pmatrix} X_{([q_1n])} \\ X_{([q_2n])}\end{pmatrix}-\begin{pmatrix} x_{q_1} \\ x_{q_2}\end{pmatrix}\right]\xrightarrow{D} N\left(0,\begin{pmatrix} \frac{q_1(1-q_1)}{f^2(x_{q_1})} & \frac{q_1(1-q_2)}{f(x_{q_1})f(x_{q_2})} \\ \frac{q_1(1-q_2)}{f(x_{q_1})f(x_{q_2})} & \frac{q_2(1-q_2)}{f^2(x_{q_2})}\end{pmatrix}\right).$$

Example 1: Assume $X_1,\cdots,X_n$ are iid random variables with common PDF $f(x)$ and CDF $F(x)$. Suppose we use $X_{([0.3n])}$ to estimate $x_{0.3}=F^{-1}(0.3)$. Then, we have

$$\sqrt{n}[X_{([0.3n])}-x_{0.3}]\xrightarrow{D} N(0,\frac{0.21}{f^2(x_{0.3})}).$$

Therefore, the 95% confidence interval for $x_{0.3}$ is approximately

$$x_{([0.3n])}\pm\frac{1.96\times\sqrt{0.21}}{f(x_{0.3})\sqrt{n}}.$$

Let x_m be the true median and $X_{([0.5m])}$ be the sample median. Then,

$$\sqrt{n}[X_{([0.5n])} - x_m] \xrightarrow{D} N(0, \frac{0.25}{f^2(x_m)})$$

ad the 95% confidence interval for x_m is

$$X_{([0.5n])} \pm \frac{0.98}{f(x_m)\sqrt{n}}.$$

Example 2: In the previous example, suppose

$$f(x) = \frac{1}{\pi[1 + (x-\theta)^2]}, -\infty < x < \infty.$$

Then, θ is the median and $\tilde{\theta} = X_{([0.5n])}$ is an estimator of θ. The confidence interval for θ is

$$X_{([0.5n])} \pm \frac{0.98\pi}{\sqrt{n}}.$$

4.5 Introduction to Hypotheses Testing

Assume the PDF (or PMF) is $f(x;\theta)$, $\theta \in \Omega$. Assume $\Omega_0 \cup \Omega_1 = \Omega$ and $\Omega_0 \cap \Omega_1 = \phi$. Suppose we consider the hypotheses

$$H_0 : \theta \in \Omega_0 \text{ versus } H_1 : \theta \in \Omega_1.$$

We will draw conclusion based on observations.

Look at the following 2×2 table.

	Truth	
Conclusion	H_0	H_1
Accept H_0	Correct	Type II Error
Reject H_0	Type I Error	Correct

We call

$$P(\text{Reject } H_0|H_0)$$

is the type I error probability and

$$P(\text{Accept } H_0|H_1)$$

is the type II error probability. We call the maximum of type I error probability is the significance level, which is usually denoted by α. That is

$$\alpha = \max_{\theta\in\Omega_0} P(\text{Reject } H_0|H_0).$$

The power function of a test is defined by

$$P(\text{Reject } H_0|\theta),$$

whic is a function of θ.

For a given α, we need to find the rejection region C based on a test statistic T. We reject H_0 if $T \in C$ and we accept H_0 if $T \notin C$.

Please understand the above concepts based on the following examples:

Example: Suppose $X_1, \cdots, X_n$ are iid $N(\mu, 1)$. Let μ_0 be a given number. We can test

$$(a): \ H_0 : \mu \le \mu_0 \leftrightarrow H_1 : \mu > \mu_0$$

or

$$(b): \ H_0: \mu \geq \mu_0 \leftrightarrow H_1 < \mu_0.$$

or

$$(c): \ H_0: \mu = \mu_0 \leftrightarrow H_0 \neq \mu_0.$$

- Suppose that $n = 10$ in (a). Given the rejection region $C = \{\bar{X} > \mu_0 + 0.7\}$, compute type I error probability when $\mu = \mu_0 - 0.5$, type II error probability when $\mu = \mu_0 + 0.5$, the power function as a function of μ, and the significance level.

  Solution: Note that $\bar{X} \sim N(\mu, 1/10)$. The type I error probability when $\mu = \mu - 0.5$ is

$$\begin{aligned}
P(\text{Type I}|\mu = \mu_0 - 0.5) =& P(\text{Conclude } \mu > \mu_0|\mu = \mu_0 - 0.5) \\
=& P(\bar{X} > \mu_0 + 0.7|\mu = \mu_0 - 0.5) \\
=& 1 - \Phi(\frac{\mu_0 + 0.7 - (\mu_0 - 0.5)}{\sqrt{1/10}}) \\
=& 1 - \Phi(\frac{1.2}{\sqrt{1/10}}) \\
=& 1 - \Phi(3.79) \\
=& 7.53 \times 10^{-5}.
\end{aligned}$$

  The type II error probability when $\mu = \mu + 0.5$ is

$$\begin{aligned}
P(\text{Type II}|\mu = \mu_0 + 0.5) =& P(\text{Conclude } \mu \leq \mu_0|\mu = \mu_0 + 0.5) \\
=& P(\bar{X} \leq \mu_0 + 0.7|\mu = \mu_0 + 0.5) \\
=& \Phi(\frac{\mu_0 + 0.7 - (\mu_0 + 0.5)}{\sqrt{1/10}}) \\
=& \Phi(\frac{0.2}{\sqrt{1/10}}) \\
=& \Phi(0.63) \\
=& 0.7356.
\end{aligned}$$

  As a function of μ, the power function is

$$\begin{aligned}
P(\text{Conclude } H_1|\mu) =& P(\bar{X} > \mu_0 + 0.7|\mu) \\
=& P_\mu(\bar{X} > \mu_0 + 0.7) \\
=& 1 - \Phi(\frac{\mu_0 + 0.7 - \mu}{\sqrt{1/10}}).
\end{aligned}$$

  I display the power function in Figure 3 when $\mu_0 = 1$, where we have $C = \{\bar{X} > 1.7\}$. The siginficance level is

$$\begin{aligned}
\alpha =& \max_{H_0} P(\text{Type I}) \\
=& P(\text{Type I}|\mu = \mu_0) \\
=& 1 - \Phi(0.7/\sqrt{1/10}) \\
=& 1 - \Phi(2.21) \\
=& 0.0135.
\end{aligned}$$

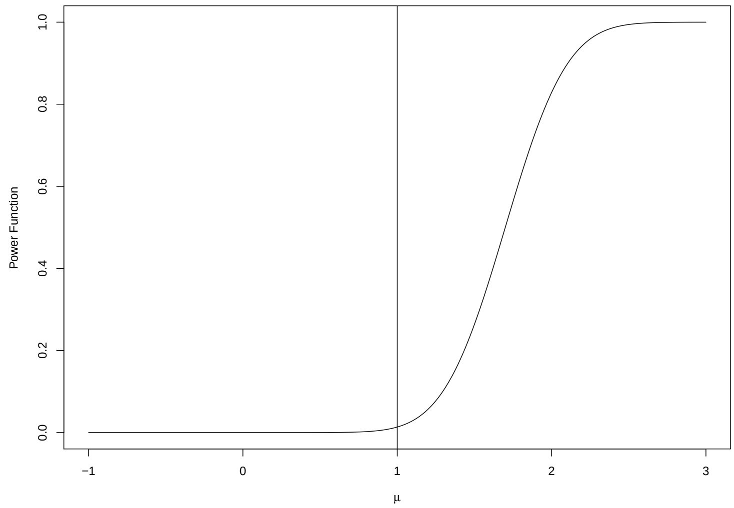

Figure 3: Power functions of the normal problem. The left is P(type I). The right is $1 - P$(type II).

- Given significance level $\alpha(1, \alpha)$, provide the rejection region for the three testing problems.

  Solution: We reject H_0 if $\bar{X} > \mu_0 + z_\alpha/\sqrt{10}$ in (a), $\bar{X} < \mu - z_\alpha/\sqrt{10}$, or $|\bar{X}| \geq z_{\alpha/2}/\sqrt{10}$ in (c). If we chooce $\alpha = 0.05$, then we have $\bar{X} > \mu_0 + 1.645/\sqrt{10}$ in (a), $\bar{X} < \mu - 1.645/\sqrt{10}$, or $|\bar{X}| \geq 1.96/\sqrt{10}$ in (c).

Example: Suppose $X \sim Bin(n, p)$. We can test

$$(a)\ H_0 : p \leq p_0 \leftrightarrow H_1 : p > p_0$$

or

$$(b)\ H_0 : p \geq p_0 \leftrightarrow H_1 : p < p_0$$

or

$$(c)\ H_0 : p = p_0 \leftrightarrow H_1 : p \neq p_0.$$

- Suppose that $n = 30$ in (a) and $p_0 = 0.5$. Given the rejection region $C = \{X \geq 19\}$, compute type I error probability when $\mu = 0.3$, type II error probability when $\mu = 0.7$, the power function as a function of μ, and the significance level.

  Solution: Note that $X \sim Bin(n, p)$. We have

$$\begin{aligned} P(\text{Type I}|p = 0.3) =& P(X \geq 19|p = 0.3) \\ =& P(Bin(30, 0.3) \geq 19) \\ =& 1.62 \times 10^{-4} \end{aligned}$$

  and

$$\begin{aligned} P(\text{Type II}|p = 0.7) =& P(X < 19|p = 0.7) \\ =& P(Bin(30, 0.7) \leq 18) \\ =& 0.1593. \end{aligned}$$

  As a function of p, the power function is

$$\begin{aligned} P(\text{Conclude } H_1|p) =& P(X \geq 19|p) \\ =& P(Bin(30, p) \geq 19). \end{aligned}$$

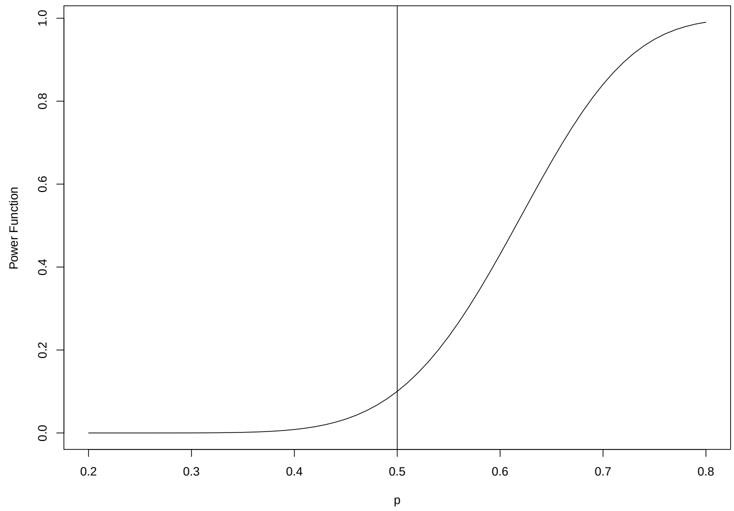

Figure 4: Power functions of the binomial problem when $p_0 = 0.5$ and $n = 30$. The left is $P(\text{type I})$. The right is $1 - P(\text{type II})$.

- Given significance level $\alpha \in (0, 1)$, provide the rejection region by the Wald method.

  Solution: Let
$$Z = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}}.$$
  We call Z the test statistic. We reject H_0 if $Z > z_\alpha$ in (a). We reject H_0 if $Z < -z_\alpha$ in (b). We reject H_0 if $|Z| > z_{\alpha/2}$ in (c).

4.6 Additional Comments About Statistical Tests

We will focus on the following examples:

Example 4.6.1: Let $X_1, \cdots, X_n$ be iid sample with mean μ and variance σ^2. Test
$$H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0.$$
Let α be the significance level. Then, we reject H_0 if
$$\left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \right| > t_{\frac{\alpha}{2}, n-1}.$$

Example 4.6.2: Assume $X_1, \cdots, X_{n_1}$ are iid $N(\mu_1, \sigma^2)$ and $Y_1, \cdots, Y_{n_2}$ are iid $N(\mu_2, \sigma^2)$. Test
$$H_0 : \mu_1 = \mu_2 \leftrightarrow H_1 : \mu_1 \neq \mu_2.$$
Suppose n is large. We reject H_0 if
$$\left| \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \right| > z_{\frac{\alpha}{2}}.$$
Suppose that n is small but we assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Let
$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}.$$

We reject H_0 is
$$\left|\frac{\bar{X}-\bar{Y}}{S_p\sqrt{1/n_1+1/n_2}}\right| > t_{\frac{\alpha}{2},n_1+n_2-2}.$$

Example 4.6.3: Suppose $X_1, \cdots, X_n$ are iid $Bernoulli(p)$. Test
$$H_0 : p = p_0 \leftrightarrow H_1 : p \neq p_0.$$
We reject H_0 if
$$\left|\frac{\bar{X}-p_0}{\sqrt{\hat{X}(1-\bar{X})/n}}\right| > z_{\frac{\alpha}{2}}.$$

Example 4.6.4: Suppose $X_1, \cdots, X_{10}$ are iid sample from $Poisson(\theta)$. Suppose we reject
$$H_0 : \theta \leq 0.1 \leftrightarrow H_1 : \theta > 0.1$$
if
$$Y = \sum_{i=1}^{10} X_i \geq 3.$$
Find the type I error probability, type II error probability and significance level.

Solution: Note that $Y \sim Poisson(10\theta)$. The type I error probability is
$$P(Y \geq 3|\theta \leq 0.1) = P(Poisson(10\theta) \geq 3|\theta \leq 0.1).$$
The type II error probability is
$$P(Y \leq 2|\theta > 0.1) = P(Poisson(10\theta) \leq 2|\theta > 0.1).$$
Significance level is
$$\max TypeI = \max P(Poisson(10\theta) \geq 3|\theta \leq 0.1) = P(Poisson(1) \geq 3) = 0.01899.$$

Example 4.6.5: Let $X_1, \cdots, X_{25}$ be iid sample from $N(\mu, 4)$. Consider the test
$$H_0 : \mu \geq 77 \leftrightarrow H_1 : \mu < 77.$$
Then, we reject H_0 is
$$\frac{\bar{X}-77}{\sqrt{4/25}} \leq -z_\alpha.$$
Suppose we observe $\bar{x} = 76.1$. The p-value is
$$P_{\mu=77}(\bar{X} \leq 76.1) = \Phi(\frac{76.1-77}{\sqrt{4/25}}) = \Phi(-2.25) = 0.012.$$

4.7 Chi-Square Tests.

Consider a test
$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1.$$
Suppose under H_0 we estimate $\mu_i = \mathrm{E}(X_i)$ by $\hat{\mu}_i$ and we estimate $\sigma_i^2 = \mathrm{V}(X_i)$ by $\hat{\sigma}_i^2$.

Pearson χ^2 statistic. The Pearson χ^2 statistic for independent random samples is
$$Y = \sum_{i=1}^{n} \frac{(X_i-\hat{\mu}_i)^2}{\sqrt{\hat{\sigma}_i^2}}.$$

The idea is motivated from independent normal distributions. Assume that $X_1, \cdots, X_n$ are independent $N(\mu_i, \sigma_i^2)$, respectively. Then,

$$X^2 = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2} \sim \chi_n^2.$$

Loglikelihood ratio statistic. Let $\ell(\theta)$ be the likelihood function. Then, the loglikelihood ratio statistic is defined by

$$\Lambda = 2 \log \frac{\sup_{\theta \in \Theta} \ell(\theta)}{\sup_{\theta \in \Theta_0} \ell(\theta)} = 2[\log \sup_{\theta \in \Theta} \ell(\theta) - \sup_{\theta \in \Theta_0} \ell(\theta)].$$

We can show both X^2 and Λ are approximately chi-square distributed. In general, we call X^2 *Pearson goodness of fit* and Λ *deviance goodness of fit* statistics. Particularly, their degrees of freedom equal to the difference of degrees of freedom between Θ and Θ_0. Let us try to understand them in the following examples for X^2. We will look at Λ in detail in Chapter 6.

Example 4.7.1 Suppose we flip a die n times. Let X_i be the number observed at the i-th time. Find Pearson χ^2 statistic X^2.

Solution: If the die is balanced, then $P(1) = P(2) = \cdots = P(6) = 1/6$. The Pearson χ^2 statistic is

$$X^2 = \sum_{i=1}^{6} \frac{(X_i - n/6)^2}{n/6}.$$

Under H_0 it approximately follows χ_5^2 distribution. In the example, we have $X_1 = 13$, $X_2 = 19$, $X_3 = 11$, $X_4 = 8$, $X_5 = 5$ and $X_6 = 4$. We have $X^2 = 15.6$. Since $15.6 > \chi^2_{0.05,5} = 11.07$, we conclude that the die is significantly unbalanced.

Example 4.7.2 Suppose we have $X_1, \cdots, X_n$ samples from a distribution taking values over $[0,1]$ with PDF $f(x) = 2x$. How to find the Pearson χ^2 statistic X^2 to test whether the distribution is uniform. Suppose we partition $[0,1]$ into four intervals $[0,1/4]$, $(1/4,1/2]$, $(1/2,3/4]$ and $(3/4,1]$.

Solution: Let p_i be the probabilities within the four intervals, respectively. Then, $p_1 = \int_0^{1/4} 2x dx = 1/16$, $p_2 = \int_{1/4}^{1/2} 2x dx = 3/16$, $p_3 = \int_{1/2}^{3/4} 2x dx = 5/16$, and $p_4 = \int_{3/4}^{1} 2x dx = 7/16$. Let n_i be the total counts in the intervals, respectively. Then,

$$X^2 = \frac{(n_1 - n/16)^2}{n/16} + \frac{(n_2 - 3n/16)^2}{3n/16} + \frac{(n_3 - 5n/16)^2}{5n/16} + \frac{(n_4 - 7n/16)^2}{7n/16}.$$

If the true distribution is the given distribution, then $X^2 \sim \chi_3^2$ approximately. Based on data $n_1 = 6$, $n_2 = 18$, $n_3 = 20$, and $n_4 = 36$. We obtain $X^2 = 1.83$. Since it is less than $\chi^2_{0.05,3} = 7.81$, we conclude that the true distribution is not significantly different from the given distribution.