My learning notes of Math Statistics II

Yang Ge

2021 - January - 23, 15:53

Contents

1 Concepts

1. Law of Large Numbers: $\frac{X_n}{n} = \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ converges (in a suitable sense) to $p = E[Y_i]$ as $n \to \infty$
2. Central Limit Theorem: if n is large, the distribution of X_n is approximately normal $N(\mu, \sigma^2)$
3. Parameter: a numerical characteristic of a population distribution, often unknown.
4. Statistics: a numerical summary of sample depending on sample only and does NOT involve unknown parameters. Mathematically, a quantity T is a statistics whenever it can be expressed as: $T = f(Y_1, ..., Y_n)$

2 Estimation

Point estimation: a single value estimate of a parameter θ based on the sample. The statistic $\hat{\theta}$ used to estimate θ is called an estimator of θ. Since an estimator is a statistic, it certainly does NOT contain any unknown parameter, and is solely a function of the sample. Interval estimation: construct an interval based on the sample, which hopefully contains the true parameter with certain quantified accuracy. Confidence interval (frequentist), credible interval (Bayes)

We will assume:

- θ: unknown parameter, taking value in a set Θ called parameter space (often a subset of $\mathbb{R}$).
- $\hat{\theta}$: an estimator of θ, which is a function of random sample $(Y_1, ..., Y_n)$.

Quality of the estimator:

3 Bias

Compare the center of the distribution of $\hat{\theta}$ with θ.

$$Bias(\hat{\theta};\theta) = E[\hat{\theta}] - \theta$$

Given an estimator $\hat{\theta}$, the $Bias(\hat{\theta};\theta)$ is a function of $\theta \in \Theta$. An estimator $\hat{\theta}$ of θ is said to be unbiased, if $Bias(\hat{\theta};\theta) = 0$ for all θ in the parameter space Θ of interest. Unbiasedness is a desirable property, since it says averagely speaking, the estimator captures the true parameter no matter where the latter is located. For unbiasedness, it is important to ensure that $E[\hat{\theta}] = \theta$ for every possible θ of interest, not just for a single value of θ.

Proposition: if $Bias(\hat{\theta};\theta) = 0$, then $Bias(a\hat{\theta};a\theta) = 0$, and $Bias(\hat{\theta}+a;\theta+a) = 0$, and $Bias(a\hat{\theta}+b\hat{\eta};a\theta+b\eta) = 0$

3.1 Exercise

8.3 Suppose that $\hat{\theta}$ is an estimator for a parameter θ and $E(\hat{\theta}) = a\theta + b$ for some nonzero constants a and b.

a In terms of a, b, and θ, what is $B(\hat{\theta})$?

b Find a function of $\hat{\theta}$—say, $\hat{\theta}^{\star}$—that is an unbiased estimator for θ.

a. $B(\hat{\theta};\theta) = E[\hat{\theta}] - \theta = a\theta + b - \theta = (a-1)\theta + b$

b. When $E[\hat{\theta}^{*}] = \theta$, then $B(\hat{\theta};\theta) = E[\hat{\theta}] - \theta = 0$.

$\because E[\hat{\theta}] = a\theta + b$

$\therefore E[\frac{\hat{\theta}-b}{a}] = \theta$

$\therefore E[\hat{\theta}^{*} = \frac{\hat{\theta}-b}{a}] = \theta$ is an unbiased estimator.

4 Variance

Unbiasedness is NOT the only property we seek for. The variance of the estimator $\hat{\theta}$ of θ to be:

$$Var(\hat{\theta};\theta) = Var(\hat{\theta})$$

$$S.E.(\hat{\theta}) = \sqrt{Var(\hat{\theta};\theta)}$$

4.1 Exercise

8.13 We have seen that if Y has a binomial distribution with parameters n and p, then Y/n is an unbiased estimator of p. To estimate the variance of Y, we generally use $n(Y/n)(1-Y/n)$.

a Show that the suggested estimator is a biased estimator of $V(Y)$.

b Modify $n(Y/n)(1-Y/n)$ slightly to form an unbiased estimator of $V(Y)$.

Based on binomial distribution, $Binomial(n,p)$, $Bias(p;E[\frac{\sum Y}{n}])=0$

$Bias(V(Y);E[npq])=0.$

However, $Bias(E[p^2];E[(\frac{\sum Y}{n})^2])\neq 0.$

Let $\sum Y$ denoted as Y.

a.

For $\theta = n\left(\frac{Y}{n}(1-\frac{Y}{n})\right) = Y - \frac{Y^2}{n}$

$\because E[\theta] = E[Y-\frac{Y^2}{n}] = E[Y] - \frac{1}{n}E[Y^2] = np - \frac{1}{n}E[Y^2]$

$\because E[Y^2] = V(Y) + E[Y]^2 = npq + n^2p^2$

$\therefore E[\theta] = np - pq - np^2 = np(1-p) - pq = pq(n-1) \neq pqn$

$\therefore Bias(V(Y);E[\theta]) \neq 0$

b.

$\because E[\theta] = pq(n-1)$

$\therefore E[\frac{n}{n-1}\theta] = pq(n-1)\frac{n}{n-1} = npq$

$\therefore \theta^* = \frac{n}{n-1} * n\frac{Y}{n}(1-\frac{Y}{n})$ is an unbiased estimator of V(Y)

8.6 Suppose that $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$, $V(\hat{\theta}_1) = \sigma_1^2$, and $V(\hat{\theta}_2) = \sigma_2^2$. Consider the estimator $\hat{\theta}_3 = a\hat{\theta}_1 + (1-a)\hat{\theta}_2$.

a Show that $\hat{\theta}_3$ is an unbiased estimator for θ.

b If $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, how should the constant a be chosen in order to minimize the variance of $\hat{\theta}_3$?

a.

$E[\hat{\theta}_3] = E[a\hat{\theta}_1 + (1-a)\hat{\theta}_2] = E[a\hat{\theta}_1] + E[(1-a)\hat{\theta}_2] = aE[\hat{\theta}_1] + (1-a)E[\hat{\theta}_2] = a\theta + (1-a)\theta = \theta$

b.

Because independent, $V(\hat{\theta}_3) = V(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = V(a\hat{\theta}_1) + V((1-a)\hat{\theta}_2) = a^2V(\hat{\theta}_1) + (1-a)^2V(\hat{\theta}_2) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$

$\because \frac{d}{da}V(\hat{\theta}_3) = 2(\sigma_1^2 + \sigma_2^2)a - 2\sigma_2^2$

$\frac{d^2}{da^2}V(\hat{\theta}_3) = 2(\sigma_1^2 + \sigma_2^2) > 0$

$\therefore \frac{d}{da}V(\hat{\theta}_3) = 0$ is the minimize point, and $a = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$

8.7 Consider the situation described in Exercise 8.6. How should the constant a be chosen to minimize the variance of $\hat{\theta}_3$ if $\hat{\theta}_1$ and $\hat{\theta}_2$ are not independent but are such that $\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = c \neq 0$?

$\because COV(aU + bV, cY + dZ) = acCOV(U,Y) + bcCOV(V,Y) + adCOV(U,Z) + bdCOV(V,Z)$

$\therefore COV(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = a(1-a)COV(\hat{\theta}_1, \hat{\theta}_2) = a(1-a)c$

$\therefore V(\hat{\theta}_3) = V(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = V(a\hat{\theta}_1) + V((1-a)\hat{\theta}_2) - 2COV(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = V(a\hat{\theta}_1) + V((1-a)\hat{\theta}_2) + 2a(1-a)c$

$\because \frac{d}{da}V(\hat{\theta}_3) = 2(\sigma_1^2 + \sigma_2^2 - 2c)a - 2\sigma_2^2 + 2c$, $\frac{d^2}{da^2}V(\hat{\theta}_3) = 2(\sigma_1^2 + \sigma_2^2) - 4c = 2(V(\hat{\theta}_1) + V(\hat{\theta}_2) + 2COV(\hat{\theta}_1, \hat{\theta}_2)) = 2V(\hat{\theta}_1 + \hat{\theta}_2) > 0$

$\therefore \frac{d}{da}V(\hat{\theta}_3) = 0$ is the minimize point, and $a = \frac{\sigma_2^2 - c}{\sigma_1^2+\sigma_2^2-2c}$

5 Mean Squared Error (MSE)

A measure of goodness of estimator combining both bias and variance:

$$MSE(\hat{\theta};\theta) = E[(\hat{\theta}-\theta)^2]$$

$$MSE(\hat{\theta};\theta) = Bias(\hat{\theta};\theta)^2 + Var(\hat{\theta};\theta)$$

If the estimator $\hat{\theta}$ is unbiased. then $MSE(\hat{\theta};\theta) = Var(\hat{\theta};\theta)$

5.1 Exercise

8.1 Using the identity

$$(\hat{\theta}-\theta) = [\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta] = [\hat{\theta} - E(\hat{\theta})] + B(\hat{\theta}),$$

show that

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = V(\hat{\theta}) + (B(\hat{\theta}))^2.$$

$$\begin{aligned}
MSE(\hat{\theta}) &= E[(\hat{\theta}-\theta)^2] && (1)\\
&= E\left[(\hat{\theta}-E(\hat{\theta})+B(\hat{\theta}))^2\right] && (2)\\
&= E\left[((\hat{\theta}-E(\hat{\theta}))^2 + B(\hat{\theta})^2 + 2B(\hat{\theta})(\hat{\theta}-E(\hat{\theta}))\right] && (3)\\
&= E[((\hat{\theta}-E(\hat{\theta}))^2] + E[B(\hat{\theta})^2] + E[2B(\hat{\theta})(\hat{\theta}-E(\hat{\theta}))] && (4)\\
&= V(\hat{\theta}) + B(\hat{\theta})^2 + 2B(\hat{\theta})E[(\hat{\theta}-E(\hat{\theta}))] && (5)\\
&= V(\hat{\theta}) + B(\hat{\theta})^2 && (6)
\end{aligned}$$

8.4 Refer to Exercise 8.1.

a If $\hat{\theta}$ is an unbiased estimator for θ, how does $\text{MSE}(\hat{\theta})$ compare to $V(\hat{\theta})$?

b If $\hat{\theta}$ is an biased estimator for θ, how does $\text{MSE}(\hat{\theta})$ compare to $V(\hat{\theta})$?

a.

$\because MSE(\hat{\theta}) = V(\hat{\theta}) + B(\hat{\theta})^2$

$\therefore$ if $B(\hat{\theta}) = 0$, $MSE(\hat{\theta}) = V(\hat{\theta})$

b.

if $B(\hat{\theta}) > 0$, $MSE(\hat{\theta}) = V(\hat{\theta}) + B(\hat{\theta})^2 > V(\hat{\theta})$

8.14 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a population whose density is given by

$$f(y) = \begin{cases} \alpha y^{\alpha-1}/\theta^{\alpha}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere,} \end{cases}$$

where $\alpha > 0$ is a known, fixed value, but θ is unknown. (This is the power family distribution introduced in Exercise 6.17.) Consider the estimator $\hat{\theta} = \max(Y_1, Y_2, \ldots, Y_n)$.

a Show that $\hat{\theta}$ is a biased estimator for θ.

b Find a multiple of $\hat{\theta}$ that is an unbiased estimator of θ.

c Derive MSE($\hat{\theta}$).

a.

$\because \hat{\theta} = max(Y_1, Y_2, ..., Y_n)$

$\therefore F(\hat{\theta} \le x) = P(Y_i \le x)^n = (\int_0^x \alpha \frac{y^{\alpha-1}}{\theta^\alpha}\, dy)^n = (\frac{x^\alpha}{\theta^\alpha})^n,\ f(\hat{\theta}) = F(\hat{\theta} \le x)' = \frac{n\alpha x^{n\alpha-1}}{\theta^{n\alpha}}, x \in [0, \theta]$

$\because B(\hat{\theta}) = E[\hat{\theta}] - \theta = \int_0^\theta x \frac{n\alpha x^{n\alpha-1}}{\theta^{n\alpha}}\, dx, x \in [0, \theta]$

$= n\alpha \frac{1}{\theta^{n\alpha}} \frac{1}{n\alpha+1} x^{n\alpha+1}\Big|_0^\theta - \theta = n\alpha \frac{1}{n\alpha+1}\theta - \theta \neq 0$

$\therefore$ the estimator is biased.

b.

$\because E[\hat{\theta}] = n\alpha \frac{1}{n\alpha+1}\theta$

$\therefore E[\frac{n\alpha+1}{n\alpha}\hat{\theta}] = \theta$, and it is the unbiased estimator.

c.

$$MSE(\hat{\theta}) = B(\hat{\theta})^2 + VAR(\hat{\theta}) \tag{7}$$

$$= (n\alpha\frac{1}{n\alpha+1}\theta - \theta)^2 + E[\hat{\theta}^2] - (E[\hat{\theta}])2 \tag{8}$$

$$= (n\alpha\frac{1}{n\alpha+1}\theta - \theta)^2 + \int_0^\theta x^2 \frac{n\alpha x^{n\alpha-1}}{\theta^{n\alpha}}\, dx - (n\alpha\frac{1}{n\alpha+1}\theta)^2 \tag{9}$$

$$= (\frac{-\theta}{n\alpha+1})^2 + n\alpha\frac{1}{n\alpha+2}\theta^2 - (\frac{n\alpha}{n\alpha+1}\theta)^2 \tag{10}$$

$$= \frac{2}{(n\alpha+1)(n\alpha+2)}\theta^2 \tag{11}$$

8.19 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a population with an exponential distribution whose density is given by

$$f(y) = \begin{cases} (1/\theta)e^{-y/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

If $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ denotes the smallest-order statistic, show that $\hat{\theta} = nY_{(1)}$ is an unbiased estimator for θ and find MSE($\hat{\theta}$). [*Hint:* Recall the results of Exercise 6.81.]

For $B(nY_{(1)})$:

$\because P(Y_{(1)} \geq x) = P(min(Y_1, Y_2, Y_3...Y_n) \geq x) = P(Y_i \geq x)^n$

$\therefore F(Y_{(1)}) = 1 - P(Y_{(1)} \geq x) = 1 - P(Y_i \geq x)^n = 1 - (1 - P(Y_i \leq x))^n$

$\because Y_i \sim exp(\theta)$, $F(Y_i) = 1 - e^{-\frac{x}{\theta}}$

$\therefore F(Y_{(1)}) = 1 - (1 - 1 + e^{-\frac{x}{\theta}})^n = 1 - e^{-\frac{nx}{\theta}}$, and $f(y_{(1)}) = -n(e^{-\frac{x}{\theta}})^{n-1}e^{-\frac{x}{\theta}}(-\frac{1}{\theta})$

$\therefore E[nY_{(1)}] = n \int\limits_0^{+\infty} x f_x(Y_{(1)}) = n\frac{\theta}{n} = \theta$

For $MSE(nY_{(1)})$:

$$\begin{aligned}
MSE(nY_{(1)}) &= B(nY_{(1)}) + V(nY_{(1)}) && (12)\\
&= n^2 V(Y_{(1)}) && (13)\\
&= n^2(E[Y_{(1)}^2] - E[Y_{(1)}]^2) && (14)\\
&= n^2(\int_0^{+\infty} x^2 f_x(Y_{(1)}) - (\frac{\theta}{n})^2) && (15)\\
&= n^2(2(\frac{\theta}{n})^2 - (\frac{\theta}{n})^2) && (16)\\
&= \theta^2 && (17)
\end{aligned}$$

6 Efficiency

Give two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of the same parameter θ, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is defined as (note the reversed order in the ratio):

$$eff(\theta_1, \theta_2) = \frac{MSE(\theta_2)}{MSE(\theta_1)}$$

6.1 Exercise

9.7 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from an exponential distribution with density function given by

$$f(y) = \begin{cases} (1/\theta)e^{-y/\theta}, & 0 < y, \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 8.19, we determined that $\hat{\theta}_1 = nY_{(1)}$ is an unbiased estimator of θ with $\text{MSE}(\hat{\theta}_1) = \theta^2$. Consider the estimator $\hat{\theta}_2 = \overline{Y}$ and find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

$\because eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}$

$\because MSE(\hat{\theta}_1) = \theta^2$

$\because MSE(\hat{\theta}_2) = B(\hat{\theta}_2)^2 + V(\hat{\theta}_2)$, and $E[\hat{\theta}_2] = E[\bar{Y}] = \theta$, $V(\hat{\theta}_2) = V(\bar{Y}) = \frac{\sigma^2}{n} = \frac{\theta^2}{n}$

$\therefore eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)} = \frac{\frac{\theta^2}{n}}{\theta^2} = \frac{1}{n}$

7 Consistency

Convergence: For non-random sequence, A sequence of non-random numbers $x_n \in \mathbb{R}$ is said to converge to $x \in \mathbb{R}$, if for any fixed $\epsilon > 0$ for all sufficiently large n, we have $|x_n - x| \leq \epsilon$. For a sequence of random variables X_n is said to converge in probability to a constant x if for any fixed $\epsilon > 0$ as $n \to \infty$, we have $P(|X_n - x| \leq \epsilon) = P(x - \epsilon \leq X_n \leq x + \epsilon) \to 1$, denoted as $X_n \xrightarrow{P} x$.

Consistency: An estimator $\hat{\theta}_n$ (strictly, a sequence of estimators) of θ based on a sample of size n is said to be consistent.

Theorem 8.3.7: if $MSE(\hat{\theta}_n; \theta) \xrightarrow{P} 0$ for all $\theta \in \Theta$, then $\hat{\theta}_n$ is consistent for θ

Proposition: 1. $X_n + Y_n \xrightarrow{P} x + y$; 2. $X_n \times Y_n \xrightarrow{P} x \times y$; 3. $X_n / Y_n \xrightarrow{P} x/y$ if $Y_n, y \neq 0$); 4. If g is a continuous function, then $g(X_n) \xrightarrow{P} g(x)$, and could be generalized to multivariate continuous function.

7.1 Exercise

Exercise 8. Show that the concept $\xrightarrow{P}$ coincides with the usual convergence of non-random numbers if X_n are not random. Hint: if X_n are not random, the event $|X_n - x| \leq \epsilon$ has probability either 0 or 1.

$\because \lim_{n\to\infty} f(n) = X_n = x$

$$\therefore P(|f(n) - x| < \epsilon) = \begin{cases} P(1 - \frac{x}{f(n)} < \frac{\epsilon}{f(n)}) \overset{n\to+\infty}{\longrightarrow} P(1 - 1 < \frac{\epsilon}{x}) \xrightarrow{P} 1 & \text{, if } \lim_{n\to\infty} f(n) > 0 \,\&\, \epsilon > 0 \\ P(1 - \frac{x}{f(n)} > \frac{\epsilon}{f(n)}) \overset{n\to+\infty}{\longrightarrow} P(1 + 1 > \frac{\epsilon}{x}) \xrightarrow{P} 1 & \text{, if } \lim_{n\to\infty} f(n) < 0 \,\&\, \epsilon > 0 \\ P(f(n) - x < \epsilon) \overset{n\to+\infty}{\longrightarrow} P(0 - 0 < \epsilon) \xrightarrow{P} 1 & \text{, if } \lim_{n\to\infty} f(n) = 0 \,\&\, \epsilon > 0 \end{cases}$$

***9.27** Use the method described in Exercise 9.26 to show that, if $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ when $Y_1, Y_2, \ldots, Y_n$ are independent uniform random variables on the interval $(0, \theta)$, then $Y_{(1)}$ is *not* a consistent estimator for θ. [*Hint:* Based on the methods of Section 6.7, $Y_{(1)}$ has the distribution function

$$F_{(1)}(y) = \begin{cases} 0, & y < 0, \\ 1 - (1 - y/\theta)^n, & 0 \leq y \leq \theta, \\ 1, & y > \theta.] \end{cases}$$

$\because P(|Y_{(1)} - \theta| \leq \epsilon) = P(\theta - \epsilon \leq Y_{(1)} \leq \theta + \epsilon) = F(\theta + \epsilon) - F(\theta - \epsilon) = (1 - \frac{\theta - \epsilon}{\theta})^n - (1 - \frac{\theta + \epsilon}{\theta})^n = (\frac{\epsilon}{\theta})^n - (\frac{-\epsilon}{\theta})^n$

$\therefore P(|Y_{(1)} - \theta| \leq \epsilon) = 0$ when n is an even number.

$\therefore P(|Y_{(1)} - \theta| \leq \epsilon) = 2(\frac{\epsilon}{\theta})^n \overset{n\to+\infty}{\longrightarrow} 0$ when n is an odd number and $0 \leq \epsilon < \theta$.

9.19 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y) = \begin{cases} \theta y^{\theta - 1}, & 0 < y < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

where $\theta > 0$. Show that $\overline{Y}$ is a consistent estimator of $\theta/(\theta + 1)$.

$\because E[y] = \int_0^1 y\theta y^{\theta-1}\, dy = \frac{\theta}{\theta+1}$

$\therefore B(\bar{Y}, \frac{\theta}{\theta+1}) = 0$

$\because V(y) = \frac{\theta}{(\theta+2)(\theta+1)^2}$

$\therefore V(\bar{Y}) = \frac{V(y)}{n}$

$\therefore MSE(\bar{Y}, \frac{\theta}{\theta+1}) \stackrel{n\to+\infty}{\longrightarrow} 0$

$\therefore \bar{Y}$ is a consistent estimator of $\frac{\theta}{\theta+1}$

8 Constructing estimators

By Method of Moments

Sample's kth moment is defined as $m_k = m_{k.n} = \frac{1}{n}\sum_{i=1}^{n} Y_i^k$. It is statistics. $m_k \xrightarrow{P} E[Y_i^k] = \mu_k, k = 1, 2, ...$

Generally:

1. Compute the population moments: $\mu_k = \mu_k(\theta_1, ..., \theta_r) = E[Y_1^k], k = 1, ..., r$ which depend on $\theta_1, ..., \theta_r$.
2. Establish a system of equations $\mu_k(\theta_1, ..., \theta_r) = m_k = \frac{1}{n}\sum_{i=1}^{n} Y_i^k, k = 1, ..., r.$
3. Solve the system of equations, namely, express $\theta_1, ..., \theta_r$ in terms of $m_1, ..., m_r$, to obtain the estimators.

8.1 Exercise

9.78 Let $Y_1, Y_2, \ldots, Y_n$ denote independent and identically distributed random variables from a power family distribution with parameters α and $\theta = 3$. Then, as in Exercise 9.43, if $\alpha > 0$,

$$f(y|\alpha) = \begin{cases} \alpha y^{\alpha-1}/3^\alpha, & 0 \le y \le 3, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $E(Y_1) = 3\alpha/(\alpha + 1)$ and derive the method-of-moments estimator for α.

$\because E[y] = \int_0^3 y \frac{\alpha y^{\alpha-1}}{3^\alpha}\, dy = \frac{\alpha}{3^\alpha}(\frac{1}{\alpha+1} y^{\alpha+1})\Big|_0^3 = \frac{3\alpha}{\alpha+1}$

$\therefore \hat{\theta} = \frac{3\alpha}{\alpha+1}$

$\therefore \alpha = \frac{\hat{\theta}}{3-\hat{\theta}}$

9.72 If $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the normal distribution with mean μ and variance σ^2, find the method-of-moments estimators of μ and σ^2.

$\because m_t = \exp(\mu t + \frac{t^2\sigma^2}{2})$

$\therefore m_1 = \exp(\mu + \frac{\sigma^2}{2})$ and $m_2 = \exp(2\mu + 2\sigma^2)$

$\because \bar{Y} \overset{L.L.N}{\longrightarrow} \mu$, and $\hat{\sigma}^2 \overset{L.L.N}{\longrightarrow} \sigma^2$

$\therefore \mu = \exp(\bar{Y} + \frac{\hat{\sigma}^2}{2})$

$\therefore \sigma^2 = m_2 - m_1^2 = \exp(2\bar{Y} + 2\hat{\sigma}^2) - 2\exp(\bar{Y} + \frac{\hat{\sigma}^2}{2})$

9 Estimation Under Parametric Models

Let Θ be a subset of $\mathbb{R}$. Let F_θ be a probability distribution which is uniquely determined by the value $\theta \in \Theta$. Then the collection $(F_\theta : \theta \in \Theta)$ is called a parametric family of distributions parameterized by θ in parameter space Θ.

Issues: If no other information of a parametric family of distributions about Y is known, then one value of θ does NOT determine one unique distribution. In a parametric family, while a single parameter CANNOT correspond to multiple distributions, it can happen that multiple parameters correspond to the same distribution (in this case we say the parameter is not identifiable).

How is a parametric family used for statistical inference?

Idea: Suppose we have IID sample $Y_1, ..., Y_n$ from one distribution F_θ of the family, but we do not know which θ. We need to construct an estimator $\hat{\theta}$ based on the sample, so that the particular distribution $F_{\hat{\theta}}$ fits the sample well.

10 MLE: Maximum likelihood estimator

Likelihood function $L(\theta) = L(\theta; Y_1, ..., Y_n) = p(Y_1;\theta) \times ... \times p(Y_n;\theta)$ or $L(\theta) = f(Y_1;\theta) \times ... \times f(Y_n;\theta)$. A likelihood function is the joint PMF/PDF. The maximum likelihood estimator (MLE) defined as $\hat{\theta} = \arg\max_{\theta\in\Theta} L(\theta)$, that is the θ maximizing the likelihood function. If multiple maximizers exist. choose one of them. The solution $\hat{\theta}$ is a statistics since it only depends on $(Y_1, ...Y_n)$. Choose the parameter which makes the sample most probable.

Since the logarithm function ln is strictly increasing, MLE can be replaced by $\hat{\theta} = \arg\max_{\theta\in\Theta} \ln L(\theta)$. Maximizing the log likelihood $\ell(\theta) = \ln L(\theta)$ is often easier.

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(Y_i;\theta)$$

10.1 Exercise

9.103 A random sample of size n is taken from a population with a Rayleigh distribution. As in Exercise 9.34, the Rayleigh density function is

$$f(y) = \begin{cases} \left(\frac{2y}{\theta}\right) e^{-y^2/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the MLE of θ.

***b** Find the approximate variance of the MLE obtained in part (a).

$$\because L(\theta) = (\frac{2y_1}{\theta} e^{-\frac{y_1^2}{\theta}})...(\frac{2y_n}{\theta} e^{-\frac{y_n^2}{\theta}})$$
$$= (\theta)^{-n} 2^n \prod y_i e^{-\frac{\sum y_i^2}{\theta}}$$

$\therefore \ell(\theta) = ln(L(\theta)) = -nln(\theta) + nln(2) + ln(\prod y_i) + (-\frac{1}{\theta}) \sum y_i^2$

$\therefore \ell'(\theta) = -\frac{n}{\theta} + 0 + 0 + \frac{1}{\theta^2} \sum y_i^2 = 0$

$= \frac{1}{\theta}(\frac{1}{\theta} \sum y_i^2 - n) = 0$

$\therefore \hat{\theta} = \frac{\sum y_i^2}{n}$

9.97 The geometric probability mass function is given by

$$p(y \mid p) = p(1-p)^{y-1}, \qquad y = 1, 2, 3, \ldots.$$

A random sample of size n is taken from a population with a geometric distribution.

a Find the method-of-moments estimator for p.

b Find the MLE for p.

a.

$\because E[p(y|p)] = \frac{1}{p}$ for geometric probability

$\therefore \mu_1 = \frac{1}{p}$

$\therefore p = \frac{1}{\mu_1}$

b.

$L(p) = p(1-p)^{y_1-1}p(1-p)^{y_2-1}...p(1-p)^{y_n-1} = p^n(1-p)^{\sum y_i - n}$

$ln(L(p)) = nln(p) + (\sum y_i - n)ln(1-p)$

$\therefore \ell(p) = \frac{1}{p}ln(p) + \frac{1}{1-p}(\sum y_i - n)(-1)$

let $\ell(p) = 0$

$\therefore \frac{1}{p}ln(p) + \frac{1}{1-p}(\sum y_i - n)(-1) = 0$

$\therefore p = \frac{n}{\sum y_i} = \frac{1}{\mu_1}$

9.104 Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from the density function

$$f(y \mid \theta) = \begin{cases} e^{-(y-\theta)}, & y > \theta, \\ 0, & \text{elsewhere} \end{cases}$$

where θ is an unknown, positive constant.

a Find an estimator $\hat{\theta}_1$ for θ by the method of moments.

b Find an estimator $\hat{\theta}_2$ for θ by the method of maximum likelihood.

c Adjust $\hat{\theta}_1$ and $\hat{\theta}_2$ so that they are unbiased. Find the efficiency of the adjusted $\hat{\theta}_1$ relative to the adjusted $\hat{\theta}_2$.

a.

$\because E[f(y|\theta)] = \int_\theta^\infty ye^{\theta-y}dy = (-ye^{\theta-y} - e^{\theta-y})\Big|_\theta^\infty = 1 + \theta$

$\therefore \mu_1 = 1 + \theta$, $\hat{\theta} = \mu_1 - 1$

b.

$L(\theta) = e^{n\theta - \sum y_i}$

$\because y > \theta$

$\therefore \max(L(\theta)) = \max(n\theta - \sum y_i)$

$\therefore \theta = min(Y_i)$

c.

$E[\hat{\theta}_1] = \mu_1 - 1 = 1 + \theta - 1 = \theta$

$E[\hat{\theta}_2] = E[\min(Y_i)] = \int_\theta^\infty xF'(\hat{\theta}_2)\,dx$

$= \int_\theta^\infty -nx(e^{\theta-x})^n dx$

$= -n(-\frac{x}{n}(e^{\theta-x})^n - \frac{1}{n^2}(e^{\theta-x})^n)\Big|_\theta^\infty$

$= \theta + \frac{1}{n} \xrightarrow{P} \theta$

$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)} = \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)}$

$V(\hat{\theta}_1) = \frac{1}{n}$

$V(\hat{\theta}_2) = \frac{1}{n^2}$

$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{1}{n}$

9.80 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Poisson distribution with mean λ.

a Find the MLE $\hat{\lambda}$ for λ.

b Find the expected value and variance of $\hat{\lambda}$.

c Show that the estimator of part (a) is consistent for λ.

d What is the MLE for $P(Y = 0) = e^{-\lambda}$?

a.

$L(\lambda) = \prod \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$

$Ln(L(\lambda)) = \sum ln(\frac{\lambda^{y_i} e^{-\lambda}}{y_i!}) = \sum(-\lambda + y_i ln(\lambda) - ln(y_i!))$

$\ell(L(\hat{\lambda})) = \sum(-1 + \frac{1}{\hat{\lambda}} y_i) = 0$

$\hat{\lambda} = \frac{\sum y_i}{n} = \bar{Y}$

b.

$V(\hat{\lambda}) = V(\bar{Y}) = \frac{\lambda}{n}$

$E(\hat{\lambda}) = E(\bar{Y}) = \lambda$

c.

$MSE(\hat{\lambda}) = 0 + v(\hat{\lambda}) \xrightarrow{P} 0$

d.

$\because \hat{\lambda}$ is MLE of λ

$\therefore \hat{\lambda}$ also is MEL of λ when $Y = 0$

$\therefore P(Y = 0) = e^{-\hat{\lambda}}$

11 Fisher information

$$I(\theta) = E\Big[-\frac{d^2}{d\theta^2}\ln p(Y_i;\theta)\Big] = Var\Big[\frac{d}{d\theta}\ln p(Y_i;\theta)\Big]$$

where p is a marginal PMF/PDF, the second equality holds under some condition. Suppose θ is the true parameter. Under some conditions, the MLE $\hat{\theta}$ satisfies, as the sample size $n \to \infty$:

1. $\hat{\theta}_n$ is consistent ($\hat{\theta} \overset{P}{\to} \theta_0$).
2. $\hat{\theta}$ is approximately distributed as $N(\theta, \frac{1}{nI(\theta)})$.
3. $\hat{\theta}$ is asymptotically unbiased, with asymptotic variance $\frac{1}{nI(\theta)}$.

In fact, no other consistent estimator can beat this asymptotic variance (Cramer-Rao lower bound).

12 Sufficiency

Suppose $(Y_1, ..., Y_n)$ is a sample from a population distribution with unknown parameter θ. A statistic T is said to be `sufficient` for θ, if the conditional distribution of the sample $(Y_1, ..., Y_n)$ given T dose Not depends on θ

`Factorization Theorem`: The statistic $T = f(Y_1, ..., Y_n)$, f is a function, is sufficient for parameter θ, if and only if the likelihood function can be factorized as $L(\theta; Y_1, ..., Y_n) = g(T, \theta) \times h(Y_1, ...Y_n)$. Namely, it can be factorized into two parts such that,

1. one part involves θ and T;
2. the other part dose not involve θ.

The MLE $\hat{\theta}$ of θ is a function of a sufficient statistics $T = f(Y_1, ..., Y_n)$ of θ. By factorization theorem $L(\theta; Y_1, ..., Y_n) = g(T, \theta) \times h(Y_1, ...Y_n)$. The h factor dose NOT depend on θ. So maximizing L with respect to θ is equivalent to maximizing $g(T, \theta)$. Hence the maximizer depends only on T. It implies that MLE automatically explores the full information (sufficient statistic) about θ.

12.1 Exercise

9.38 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a normal distribution with mean μ and variance σ^2.

a If μ is unknown and σ^2 is known, show that $\overline{Y}$ is sufficient for μ.

b If μ is known and σ^2 is unknown, show that $\sum_{i=1}^n (Y_i - \mu)^2$ is sufficient for σ^2.

c If μ and σ^2 are both unknown, show that $\sum_{i=1}^n Y_i$ and $\sum_{i=1}^n Y_i^2$ are jointly sufficient for μ and σ^2. [Thus, it follows that $\overline{Y}$ and $\sum_{i=1}^n (Y_i - \overline{Y})^2$ or $\overline{Y}$ and S^2 are also jointly sufficient for μ and σ^2.]

For a normal distribution, the PDF is:

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

a.

if only μ is unknown:

$L(\mu) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\sum \frac{(y_i-\mu)^2}{2\sigma^2}}$

$= (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2}$

$\because \sum(y_i-\mu)^2 = \sum y_i^2 + 2\sum y_i\mu + \mu^2 = \sum y_i^2 + 2n\bar{Y} + \mu^2$

$\therefore L(\mu) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2}$

$= (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}(\sum y_i^2+2n\bar{Y}+\mu^2)}$

$= (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}(2n\bar{Y}+\mu^2)} e^{-\frac{1}{2\sigma^2}(\sum y_i^2)}$

$= g(\bar{Y},\mu) \times h(y_i^2)$

$\therefore \bar{Y}$ is sufficient for μ

b.

if only σ^2 is unknown:

similar to a,

$L(\mu) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\sum\frac{(y_i-\mu)^2}{2\sigma^2}}$

$= g(\sigma^2, \sum(y_i-\mu)^2) \times 1$

$\therefore \sum(y_i-\mu)^2$ is sufficient for μ

c.

if both μ, σ^2 are unknown:

similar to a,

$L(\mu) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}(\sum y_i^2+2\sum y_i\mu+\mu^2)}$

$= g(\sum y_i^2, \sum y_i, \mu, \sigma) \times 1$

$\therefore \sum y_i^2, \sum y_i$ are sufficient for μ, σ^2

9.43 Let $Y_1, Y_2, \ldots, Y_n$ denote independent and identically distributed random variables from a power family distribution with parameters α and θ. Then, by the result in Exercise 6.17, if $\alpha, \theta > 0$,

$$f(y \mid \alpha, \theta) = \begin{cases} \alpha y^{\alpha-1}/\theta^\alpha, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

If θ is known, show that $\prod_{i=1}^n Y_i$ is sufficient for α.

$\because L(\alpha) = (\frac{\alpha}{\theta^\alpha})^n \prod y_i^{\alpha-1} I\{0 \le y_i \le \theta\}$

$= g(\prod y_i, \alpha) \times 1$

$\therefore \prod y_i$ is sufficient for α

***9.51** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y \mid \theta) = \begin{cases} e^{-(y-\theta)}, & y \ge \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

$L(\theta) = e^{\theta-\sum y_i} I\{y_i \ge \theta\}$

$\because all\, y_i \ge \theta$

$\therefore \min(y_i) \ge \theta$

$L(\theta) = e^{\theta - \sum y_i} I\{y_i \geq \theta\}$

$= e^{\theta - \sum y_i} I\{\min(y_i) \geq \theta\}$

$= g(\min(y_i), \theta) \times 1$

$\therefore \min(y_i)$ is sufficient for θ

13 Optimal: MVUE

MVUE: An unbiased estimator $\hat{\theta}$ of θ is said to be MVUE (minimum-variance unbiased estimator) or optimal unbiased estimator, if for any other unbiased estimator $\hat{\theta}'$, $Var(\hat{\theta};\theta) \leq Var(\hat{\theta}';\theta)$ for all $\theta \in \Theta$.

Procedure:

1. Find a sufficient statistic T for θ using the Factorization Theorem.
2. Find a transform f so that if $\hat{\theta} = f(T)$, then $\hat{\theta}$ is unbiased.

A sufficient statistic T for θ is said to be complete, if the transform f that can be found in the procedure of is unique.

13.1 Exercise:

9.56 Refer to Exercise 9.38(b). Find an MVUE of σ^2.

9.38 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a normal distribution with mean μ and variance σ^2.

a If μ is unknown and σ^2 is known, show that $\overline{Y}$ is sufficient for μ.

b If μ is known and σ^2 is unknown, show that $\sum_{i=1}^{n}(Y_i - \mu)^2$ is sufficient for σ^2.

c If μ and σ^2 are both unknown, show that $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n} Y_i^2$ are jointly sufficient for μ and σ^2. [Thus, it follows that $\overline{Y}$ and $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ or $\overline{Y}$ and S^2 are also jointly sufficient for μ and σ^2.]

For a normal distribution, the PDF is:

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

For MVUE: 1. sufficient, 2. Unbiased

1.

Sufficient

if only σ^2 is unknown:

$L(\mu) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\sum \frac{(y_i-\mu)^2}{2\sigma^2}} = (\frac{1}{\sigma^2 2\pi})^{\frac{n}{2}} e^{-\sum \frac{(y_i-\mu)^2}{2\sigma^2}}$

$= g(\sigma^2, \sum(y_i - \mu)^2) \times 1$

$\therefore \sum(y_i - \mu)^2$ is sufficient for σ^2

2.

Unbiased

$\because \sigma^2 = \sum(Y_i - \mu)^2$

$\therefore \hat{\theta} = \sum(y_i - \mu)^2$ is unbiased for $\sum(Y_i - \mu)^2$

$\therefore \sum(y_i - \mu)^2$ is MVUE of σ^2

9.64 Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance 1.

a Show that the MVUE of μ^2 is $\widehat{\mu^2} = \overline{Y}^2 - 1/n$.

b Derive the variance of $\widehat{\mu^2}$.

For a normal distribution, the PDF is:

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

For MVUE: 1. sufficient, 2. Unbiased

1.

Sufficient

$L(\mu) = (\frac{1}{\sqrt{2\pi}})^n e^{-\sum \frac{(y_i-\mu)^2}{2}}$

$\because \sum(y_i-\mu)^2 = \sum y_i^2 + 2\sum y_i\mu + \mu^2 = \sum y_i^2 + 2n\bar{Y} + \mu^2$

$\therefore L(\mu) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum(y_i-\mu)^2}$

$= (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}(\sum y_i^2 + 2n\bar{Y} + \mu^2)}$

$= (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}(2n\bar{Y}+\mu^2)} e^{-\frac{1}{2}(\sum y_i^2)}$

$= (\frac{1}{\sqrt{2\pi}})^n e^{-n\bar{Y}} (e^{-\mu})^2 e^{-\frac{1}{2}(\sum y_i^2)}$

$= g(\bar{Y}, \mu) \times h(y_i^2)$

$\therefore \bar{Y}$ is sufficient for μ

$\therefore \bar{Y}$ is also sufficient for μ^2

2.

Unbiased

$\because \bar{Y} \sim N(\mu, \frac{\sigma^2}{n} = \frac{1}{n})$

$\therefore E[\bar{Y}^2] - E[\bar{Y}]^2 = V(\bar{Y}) = \frac{1}{n}$

$\therefore E[\bar{Y}^2] - \frac{1}{n} = \mu^2$

$\therefore \bar{Y}^2 - \frac{1}{n}$ is unbiased estimator of μ^2

$\therefore \bar{Y}^2 - \frac{1}{n}$ is MVUE of μ^2

9.62 Refer to Exercise 9.51. Find a function of $Y_{(1)}$ that is an MVUE for θ.

***9.51** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y \mid \theta) = \begin{cases} e^{-(y-\theta)}, & y \geq \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

For MVUE: 1. sufficient, 2. Unbiased

1.

Sufficient

$L(\theta) = e^{n\theta - \sum y_i} I(\{y_i \geq \theta\})$

$= e^{n\theta - \sum y_i} I(\{\min(Y_i) \geq \theta\})$

$= e^{n\theta - \sum y_i} I(\{Y_{(1)} \geq \theta\})$

$= g(\theta; Y_{(1)}) \times h(\sum y_i)$

$\therefore Y_{(1)}$ is sufficient for θ

2.

Unbiased

$E[Y_{(1)}] = \int_{\theta}^{\infty} xF'(Y_{(1)})\, dx$

$= \int_{\theta}^{\infty} -nx(e^{\theta-x})^n dx$

$= -n(-\frac{x}{n}(e^{\theta-x})^n - \frac{1}{n^2}(e^{\theta-x})^n)\Big|_{\theta}^{\infty}$

$= \theta + \frac{1}{n} \xrightarrow{P} \theta$

$\therefore Y_{(1)}$ is unbiased estimator of θ

$\therefore Y_{(1)}$ is MVUE of θ

14 Confidence Intervals

Basic concepts: A confidence interval (CI) is given by $I = [L, U]$ (two-sided), or $I = (-\infty, U]$ or $[L, \infty)$ (one-sided), where L (lower bound) and U (upper bound) are statistics (functions of sample only).

Most often L and U have continuous distributions. In this case taking closed or open interval does NOT matter. Since the sample is random, a CI is often random. On the other hand, when a CI has been reported from the observed data (realized sample), it is NOT random any more. The notion CI can be extended to a higher dimension, in which case it is called a confidence region.

Assume θ is the true unknown parameter. Let I be a CI, then $\alpha = 1 - P(\theta \in I) = P(\theta \notin I)$ is called the confidence level. The number $P(\theta \in I) = 1 - \alpha$ is called the coverage probability or confidence coefficient. We say I is a $(1-\alpha)$-CI for θ.

$(-\infty, U]$ is a $(1-\alpha_1)$-CI for θ; $[L, \infty)$ is a $(1-\alpha_2)$-CI for θ; $L < U$. Then $[L, U]$ is a $(1-\alpha_1-\alpha_2)$-CI for θ.

15 Pivotal

A quantity $T = f(\theta; Y_1, ..., Y_n)$ which is a function of sample and unknown parameter is said to be pivotal, if assuming θ is the true parameter, the distribution of T does NOT depend on θ. A pivotal quantity is NOT a statistic in general. Although its functional depends on θ, its distribution doesn't.

Pivotal methods:

1. Construct a pivotal quantity T.
2. Set up an equation of the form $P(T \leq y) = 1 - \alpha$ or $P(T \geq y) = 1 - \alpha$. Solve to determine y.
3. Solve the inequality $T \leq y$ or $T \geq y$ for θ to get CI.

15.1 Exercise

Provide a short explanation of the concept $(1 - \alpha)$-CI to someone with little statistical training

From a random sample which generated from a distribution with a parameter of θ, the $(1-\alpha)$-CI has $(1-\alpha)$ chance to capture the true θ. E.g. if we repeated the process of using $(1-\alpha)$-CI to capture the true θ by N times with random samples each time, we would have $(1-\alpha) \times N$ times success.

8.44 Let Y have probability density function

$$f_Y(y) = \begin{cases} \frac{2(\theta - y)}{\theta^2}, & 0 < y < \theta, \\ 0, & \text{elsewhere}. \end{cases}$$

a Show that Y has distribution function

$$F_Y(y) = \begin{cases} 0, & y \leq 0, \\ \frac{2y}{\theta} - \frac{y^2}{\theta^2}, & 0 < y < \theta, \\ 1, & y \geq \theta. \end{cases}$$

b Show that Y/θ is a pivotal quantity.

c Use the pivotal quantity from part (b) to find a 90% lower confidence limit for θ.

a.

$\because \int_0^x \frac{2(\theta-y)}{\theta^2} dx = \frac{2x}{\theta} - \frac{x^2}{\theta^2}$, for $0 < x < \theta$

$$\therefore F_Y(y) = \begin{cases} 0 & , y \leq 0 \\ \frac{2x}{\theta} - \frac{x^2}{\theta^2} & , 0 < y < \theta \\ 1 & , y \geq \theta \end{cases}$$

b.

$\because U = \frac{Y}{\theta}$

$\therefore f_U(u) = f_Y[h^{-1}(u)] \frac{dh^{-1}}{du} = \frac{2(\theta - \theta u)}{\theta^2} \times \theta = 2(1-u)$, $\mu \in (0,1)$, not depend on θ

$\therefore \frac{Y}{\theta}$ is pivotal.

c.

$\because F_U(x) = P(u \leq x) = 2x - x^2$ $(u \in [0,1])$, let a and b (b>a) are two 90% cut-off

$\because F_U(a) = 2a - a^2 = 0.1$ and $F_U(b) = 2b - b^2 = 0.9$

$\therefore (a-1)^2 = 0.9$ and $(b-1)^2 = 0.1$

$\because a, b \in [0, 1]$

$\therefore a = 1 - \sqrt{0.9}$ and $b = 1 - \sqrt{0.1}$

$\therefore \frac{Y}{\theta} \geq 1 - \sqrt{0.9}$ or $\frac{Y}{\theta} \leq 1 - \sqrt{0.1}$

$\therefore \theta \leq \frac{Y}{1-\sqrt{0.9}}$ or $\theta \geq \frac{Y}{1-\sqrt{0.1}}$ are two different 90%CI

$\therefore \theta \in [\frac{Y}{1-\sqrt{0.1}}, \infty)$

8.45 Refer to Exercise 8.44.

a Use the pivotal quantity from Exercise 8.44(b) to find a 90% upper confidence limit for θ.

b If $\hat{\theta}_L$ is the lower confidence bound for θ obtained in Exercise 8.44(c) and $\hat{\theta}_U$ is the upper bound found in part (a), what is the confidence coefficient of the interval $(\hat{\theta}_L,\ \hat{\theta}_U)$?

a.

From 8.44:

$\because \theta \leq \frac{Y}{1-\sqrt{0.9}}$ or $\theta \geq \frac{Y}{1-\sqrt{0.1}}$ are two different 90%CI

$\therefore \theta \in (-\infty, \frac{Y}{1-\sqrt{0.9}}]$

b.

$$\begin{aligned}
\because P(\theta \in (\hat{\theta}_L, \hat{\theta}_U)) &= P(\theta \in (\hat{\theta}_L, \infty) \bigcap \theta \in (-\infty, \hat{\theta}_U)) \\
&= P(\theta \in (\hat{\theta}_L, \infty)) + P(\theta \in (-\infty, \hat{\theta}_U)) - P(\theta \in (\hat{\theta}_L, \infty) \bigcup \theta \in (-\infty, \hat{\theta}_U)) \\
&= 0.9 + 0.9 - 1 \\
&= 0.8
\end{aligned}$$

4. Let $Y_1, \ldots, Y_n$ be an IID sample from Uniform$(0, \theta)$. Construct a two-sided $(1-\alpha)$-confidence interval for θ.

Let $\hat{\theta} = \max(Y_1, ...Y_n)$

$$\begin{aligned}
\therefore P(\frac{\hat{\theta}}{\theta} \leq u) &= P(\frac{Y_1}{\theta} \leq u) \times P(\frac{Y_2}{\theta} \leq u) \times ... \times P(\frac{Y_n}{\theta} \leq u) \\
&= u^n, (Y_i \leq \theta, u = \frac{Y_{(n)}}{\theta} \in [0, 1])
\end{aligned}$$

$\therefore T = \frac{\hat{\theta}}{\theta} = \frac{Y_{(n)}}{\theta}$ is pivotal

$\because F(T) = (\frac{Y_{(n)}}{\theta})^n \in [\frac{\alpha}{2}, 1 - \frac{\alpha}{2}]$

$\therefore \theta \in [\frac{Y_{(n)}}{(1-\frac{\alpha}{2})^{\frac{1}{n}}}, \frac{Y_{(n)}}{(\frac{\alpha}{2})^{\frac{1}{n}}}]$

16 Z-Score

The most challenging part of the pivotal method is to come up with a good pivotal quantity. Fortunately for many common situations, when the sample size is large, one type of pivotal quantity called Z-score can be always used.

By Central Limit Theorem, we know (approximate equality in distribution):

$$\frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \overset{d}{\approx} N(0,1)$$

If σ^2 is unknown, use S^2 sample variance replace it $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2$.

Definition: suppose $\hat{\theta}$ is an estimator of θ and $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2 = Var[\hat{\theta}]$. Then the quantity $Z = \frac{\hat{\theta}-\theta}{\hat{\sigma}} \overset{d}{\approx} N(0,1)$ is called the Z-score. Then $\hat{\theta} \pm \hat{\sigma} z_{\alpha/2}$

16.1 Exercise

Construct a two-sided $(1-\alpha)$**-CI for** θ **in Uniform**$[0,\theta]$ **based on IID** $Y_1,...,Y_n$

Let $y_1 \le y_2$, and $\frac{Y}{\theta}$ is pivotal

$\because P(T \ge y_1) = 1-\frac{\alpha}{2}$ and $P(T \le y_2) = 1-\frac{\alpha}{2}$

$\therefore y_1 = z_{\frac{\alpha}{2}}$ and $y_2 = z_{1-\frac{\alpha}{2}}$

$\therefore T = \frac{Y}{\theta} \ge y_1 = z_{\frac{\alpha}{2}}$ and $T = \frac{Y}{\theta} \le y_2 = z_{1-\frac{\alpha}{2}} = -z_{\frac{\alpha}{2}}$

$\therefore \theta \le \frac{Y}{z_{\frac{\alpha}{2}}}$ and $\theta \ge \frac{Y}{-z_{\frac{\alpha}{2}}}$

$\therefore (1-\alpha)$-CI is $[\frac{Y}{-z_{\frac{\alpha}{2}}}, \frac{Y}{z_{\frac{\alpha}{2}}}]$

8.56 Is America's romance with movies on the wane? In a Gallup Poll[5] of $n = 800$ randomly chosen adults, 45% indicated that movies were getting better whereas 43% indicated that movies were getting worse.

a Find a 98% confidence interval for p, the overall proportion of adults who say that movies are getting better.

b Does the interval include the value $p = .50$? Do you think that a majority of adults say that movies are getting better?

a.

$\because$ n=800, p=0.45

$\therefore \hat{\sigma} = \sqrt{\frac{p(1-p)}{n}} = 0.0176$

$\because Z = \frac{\hat{p}-p}{\hat{\sigma}} \overset{d}{\approx} N(0,1)$

$\therefore P(Z \le z_{\alpha/2}) = 0.01$ and $P(Z \ge z_{1-\alpha/2}) = 0.01$, $z_{\alpha/2} = -z_{1-\alpha/2} = -2.32$

$\therefore z_{\alpha/2} \le Z \le z_{1-\alpha/2}$

$\therefore 0.409 = \hat{p} - \hat{\sigma} z_{\alpha/2} \le p \le p + \hat{\sigma} z_{\alpha/2} = 0.491$

b.

$\because p \leq 0.5$, less than half think it better.

$\therefore$ a majority of adults say that movies are not getting better.

17 Small-sample CI

When the sample size is small, to compensate for the lack of information due to small sample size, we shall make a strong assumption: a sample follows a normal distribution (central limit theorem).

Let $Z(Z_1, Z_2, ..., Z_n)$ be I.I.D. N(0,1):

1. $\chi^2(\nu)$, distribution of $W = Z_1^2/\sigma + Z_2^2/\sigma + ... + Z_n^2/\sigma$;
2. $t(\nu)$, distribution of $T = \frac{Z}{\sqrt{W/\nu}}$. $t(\nu)$ has a similar shape to normal distribution (e.g. $\nu > 30$).

For normal distribution $N \sim (\mu, \sigma^2)$,

THEOREM 7.1: $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim N(\mu, \sigma^2/n)$.

THEOREM 7.2: for $Z_i = (Y_i - \mu)/\sigma$, $\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n}(\frac{Y_i-\mu}{\sigma})^2$ has a χ^2 distribution with degrees of freedom (df) = n.

THEOREM 7.3: $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ has a χ^2 distribution with (n-1) d.f., S^2 is a random sample variance.

Definition 7.3: Let W_1 and W_2 be independent χ^2 distributed random variables with ν_1 and ν_2 d.f.. Then $F\frac{W_1/\nu_1}{W_2/\nu_2}$ is said to have an F distribution with ν_1 numerator d.f. and ν_2 denominator d.f..

17.1 Exercise

Derive a one-sided $(1-\alpha)$-CI $(-\infty, U]$ for μ

$\because P(T_n = \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \geq t_{\alpha,(n-1)}) = 1-\alpha$

$\therefore \mu \leq \bar{Y} - t_{\alpha,(n-1)}\frac{\sigma}{\sqrt{n}}$

$\because t_{\alpha,(n-1)} \leq 0$

$\therefore \mu \in (-\infty, \bar{Y} + |t_{\alpha,(n-1)}|\frac{\sigma}{\sqrt{n}}]$

8.90 Do SAT scores for high school students differ depending on the students' intended field of study? Fifteen students who intended to major in engineering were compared with 15 students who intended to major in language and literature. Given in the accompanying table are the means and standard deviations of the scores on the verbal and mathematics portion of the SAT for the two groups of students:[16]

	Verbal		Math	
Engineering	$\bar{y} = 446$	$s = 42$	$\bar{y} = 548$	$s = 57$
Language/literature	$\bar{y} = 534$	$s = 45$	$\bar{y} = 517$	$s = 52$

a Construct a 95% confidence interval for the difference in average verbal scores of students majoring in engineering and of those majoring in language/literature.

b Construct a 95% confidence interval for the difference in average math scores of students majoring in engineering and of those majoring in language/literature.

c Interpret the results obtained in parts (a) and (b).

d What assumptions are necessary for the methods used previously to be valid?

a.

$\bar{Y}_{e,v} = 446$, $S_{e,v} = 42$, $n_1 = 15$, and $\bar{Y}_{l,v} = 534$, $S_{l,v} = 45$, $n_2 = 15$.

$\because S^2_{e,l,verbal} = \frac{(n_1-1)S^2_{e,v}+(n_2-1)S^2_{l,v}}{30-2} = 1894.5$

$\therefore (\bar{Y}_{e,v} - \bar{Y}_{l,v}) \pm t_{\alpha/2,28}\sqrt{1894.5(1/15 + 1/15)}$

$\therefore (\mu_{e,v} - \mu_{l,v}) \in [-88 - 2.048407 * 15.89339 = -120, -88 + 2.048407 * 15.89339 = -55]$

b.

$\bar{Y}_{e,m} = 548$, $S_{e,m} = 57$, $n_1 = 15$, and $\bar{Y}_{l,m} = 517$, $S_{l,m} = 52$, $n_2 = 15$.

$\because S^2_{e,l,math} = \frac{(n_1-1)S^2_{e,m}+(n_2-1)S^2_{l,m}}{30-2} = 2976.5$

$\therefore (\bar{Y}_{e,m} - \bar{Y}_{l,m}) \pm t_{\alpha/2,28}\sqrt{2976.5(1/15 + 1/15)}$

$\therefore (\mu_{e,m} - \mu_{l,m}) \in [31 - 2.048407 * 19.92151 = -10, 31 + 2.048407 * 19.92151 = 72]$

c.

$(\mu_{e,v} - \mu_{l,v}) \in [-120, -55]$ and $(\mu_{e,m} - \mu_{l,m}) \in [-10, 72]$

$\therefore$ Engineering has lower average scores of Verbal than Language/literature, and Engineering has similar average scores of Math to Language/literature

d. These methods need the SAT scores of Engineering and Language/literature students follow Normal distribution.

18 CI for σ^2

$T = \frac{(n-1)\hat{\sigma}_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sim \chi^2(n-1)$

18.1 Exercise

8.102 The ages of a random sample of five university professors are 39, 54, 61, 72, and 59. Using this information, find a 99% confidence interval for the population standard deviation of the ages of all professors at the university, assuming that the ages of university professors are normally distributed.

$\because T = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2}{\sigma^2} = \frac{(n-1)\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-1=4)$

$\because \sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2 = 48$

$\therefore T = 48/\sigma^2$

$\chi^2_{0.005,4} = 0.207$ and $\chi^2_{0.005,4} = 14.860$

$\therefore T \in [0.207, 14.860]$

$\therefore \sigma^2 \in [3.23, 231.88]$

Provide other real-life scenarios where such imbalance (a) exists and (b) does not exist.

a. Imbalance severity of false decision. For cancer diagnosis, H_1: Having cancer, H_2: Not having cancer. Mistake on $H2$ is much severity.

b. Balance severity of false decision. For a patient has a higher Human C-Reactive Protein (CRP) Protein, the doctor given antibiotics to treat inflammation. H_1: Upper respiratory infection, H_2: Lower respiratory infection. Mistake on $H2$ is not much severity.

19 Statistical Decision

Definition 11.1.2:

1. $\{F_\theta : \theta \in \Theta\}$: a parametric family of distributions (e.g., $F_\theta = \text{Bernoulli}(\theta)$);
2. $\{Y_1, ..., Y_n\}$: samples from one (unknown) F_θ;
3. $\Theta = \Theta_1 \cup \Theta_2$ and $\Theta_1 \cap \Theta_2 = \emptyset$
4. Hypothesis $H_1 : \theta \in \Theta_1$, $H_2 : \theta \in \Theta_2$: (e.g., $\Theta = [0,1] = \theta_1 \cup \theta_2 = [0.9, 1] \cup [0, 0.9)$)
5. T: a statistic used to make decision H_1 or H_2;
6. R_1, R_2: two disjoint regions in $\mathbb{R}$, so that $R_1 \cup R_2$ cover all possible values of T. (e.g., $T = \bar{Y}_n$, $R_1 = [\theta_t, 1]$ and $R_2 = [0, \theta_t)$)

Decision rule

$$T \in R_1 \Rightarrow H_1, T \in R_2 \Rightarrow H_2$$

$$P_\theta(A) : \text{probability of event A when} \theta \text{is the true parameter}$$

Mistake is inevitable when θ is near the border between Θ_1 and Θ_2 (candidate distributions are indistinguishable).

`Mistake A`: H_1 holds but we choose H_2 (falsely reject H_1).

`Mistake B`: H_2 holds but we choose H_1 (falsely reject H_2).

Severity of false decision. If mistake on H_1 or H_2, the severity similar then it is balance, otherwise is imbalance. When the imbalance exists, `New idea`: In a statistical decision, we first guard against the mistake of more severity (e.g., ensure the probability of this mistake is ≤ 0.05), and then do our best to minimize the probability of other mistake.

20 Elements of statistical test

Definition 11.2.1.

Statistical hypothesis test (HT) is a statistical decision process which has the following elements:

1. Null hypothesis $H_0 : \theta \in \Theta_0$;
2. Alternative hypothesis $H_a : \theta \in \Theta_a$; Where disjoint $\Theta_0 \cup \Theta_a = \Theta$, and Severity of falsely accepting H_0 < Severity of falsely accepting H_a.
3. Test statistic T;
4. Rejection region RR;
5. Decision rule: $T \in RR \Rightarrow \text{reject} H_0 (\text{or accept} H_a)$; otherwise, $T \notin RR \Rightarrow \text{fail to reject} H_0$.

Definition 11.2.3.

In a HT:

1. `Type I Error`: H_0 is true but we choose H_a (falsely reject H_0), whose probability is denoted as $\alpha(\theta) = P_\theta(T \in RR), \theta \in \Theta_0$
2. `Type II Error`: H_a is true but we choose H_0 (falsely reject H_a), whose probability is denoted as $\beta(\theta) = P_\theta(T \notin RR), \theta \in \Theta_a$

3. The probability of accepting H_a when H_a is true is called `power of the test`, $\text{pw}(\theta) = P_\theta(T \in RR) = 1 - \beta(\theta), \theta \in \Theta_a$. namely,

Important: 1. In general, both $\alpha(\theta)$ and $\beta(\theta)$ are functions, but defined on different domains Θ_0 and Θ_a respectively. 2. simple null hypothesis or composite null hypothesis. 3. $\alpha = \max(\alpha(\theta) : \theta \in \Theta_0)$

20.1 Exercise

Provide an elementary explanation of the concepts in a HT: null/alternative hypotheses, Type I/II errors and power to someone with little statistical training.

Answer:

Let a range Θ can divided into two separate sub-area Θ_0, Θ_a, and we have a θ needed to be tested

1. Null hypothesis: the hypothesis to be tested, e.g θ belong to Θ_0, $\theta \in \Theta_0$.
2. Alternative hypotheses: the hypothesis we could accept when we reject the null hypothesis.
3. Type I errors: if the null hypothesis is true, the probability of falsely rejecting the null hypothesis.
4. Type II errors: if the alternative hypothesis is true, the probability of falsely rejecting the alternative hypothesis.
5. Power: if the alternative hypothesis is true, the probability of correctly accepting alternative hypothesis.

Suppose Y_i are IID N(θ, 1) as above. Suppose we test $H_0 : \theta = 0$ vs $H_a : \theta \neq 0$. Still $T = \bar{Y}_n$, but RR now takes the form $(-\infty, -c] \cup [+c, \infty)$, where c is a threshold chosen to make the test at level α. Calculate the type I, II error probabilities and the power.

Answer:

$\because Y_i \sim N(\theta, 1)$, $H_0 : \theta = 0$, $H_a : \theta \neq 0$, $T = \bar{Y} \sim N(\theta, \frac{1}{n})$, Reject region (RR) $= (-\infty, -c] \cup [+c, \infty)$

$\therefore \frac{\bar{Y}-\theta}{\frac{1}{\sqrt{n}}} \sim N(0, 1)$

Type I error: when $\theta = 0$

$$
\begin{aligned}
\therefore \alpha(\theta = 0) &= P_\theta(T \in RR) \\
&= P_\theta(T \leq -c \cup T \geq c) \\
&= 1 - P_\theta(-c \leq T \leq c) \\
&= 1 - P_\theta\Big(\sqrt{n}(-c-\theta) \leq \sqrt{n}(T-\theta) \leq \sqrt{n}(c-\theta)\Big) \\
&= 1 - \Big(\Phi(\sqrt{n}(c-\theta)) - \Phi(\sqrt{n}(-c-\theta))\Big) \\
&= 1 - \Big(\Phi(c\sqrt{n}) - \Phi(-c\sqrt{n})\Big) \\
&= \alpha
\end{aligned}
$$

$\therefore \Phi(c\sqrt{n}) - \Phi(-c\sqrt{n}) = 1 - \alpha$

$\therefore \Phi(c\sqrt{n}) = 1 - \frac{\alpha}{2}$

$\therefore c\sqrt{n} = z_{1-\frac{\alpha}{2}}$, $c = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$

Type II error: when $\theta \neq 0$

$$\begin{aligned}
\therefore \beta(\theta \neq 0) &= P_\theta(T \notin RR) \\
&= P_\theta(-c \leq T \leq c) \\
&= P_\theta\Big(\sqrt{n}(-c-\theta) \leq \sqrt{n}(T-\theta) \leq \sqrt{n}(c-\theta)\Big) \\
&= \Phi(\sqrt{n}(c-\theta)) - \Phi(\sqrt{n}(-c-\theta)), \text{if } c = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \\
&= \Phi(z_{1-\frac{\alpha}{2}} - \theta\sqrt{n}) - \Phi(-z_{1-\frac{\alpha}{2}} - \theta\sqrt{n})
\end{aligned}$$

Power: when $\theta \neq 0$, $P_\theta(T \in RR) = 1 - \beta(\theta \neq 0)$

$\therefore \text{pw}(\theta) = 1 - \Big(\Phi(z_{1-\frac{\alpha}{2}} - \theta\sqrt{n}) - \Phi(-z_{1-\frac{\alpha}{2}} - \theta\sqrt{n})\Big)$

21 Large-sample Z-tests

Idea: 1. Estimate θ with an estimator $\hat{\theta}$; 2. $\hat{\theta}$ far from $\theta_0 \Rightarrow$ reject H_0

$$Z = \frac{\hat{\theta} - \theta_0}{\hat{\alpha}}$$

H_0	H_0	Rejection Rule
$\theta = \theta_0$	$\theta \neq \theta_0$	$\lVert Z \rVert > Z_{\alpha/2}$
$\theta = \theta_0$ or $\theta \leq \theta_0$	$\theta > \theta_0$	$\lVert Z \rVert > Z_{\alpha/2}$
$\theta = \theta_0$ or $\theta \geq \theta_0$	$\theta < \theta_0$	$\lVert Z \rVert < -Z_{\alpha/2}$

If $H_0 : \theta = \theta_0$ is true, then when sample size is large, $Z \sim N(0,1)$

Type I error: $P_{\theta_0}(|Z| > z_{\alpha/2}) = \alpha$

Hypothesis test and confidence interval: if a CI fails to cover θ_0, then $\hat{\theta} = \theta_0$ is unlikely.

21.1 Exercise

8.56 Is America's romance with movies on the wane? In a Gallup Poll[5] of $n = 800$ randomly chosen adults, 45% indicated that movies were getting better whereas 43% indicated that movies were getting worse.

a Find a 98% confidence interval for p, the overall proportion of adults who say that movies are getting better.

b Does the interval include the value $p = .50$? Do you think that a majority of adults say that movies are getting better?

a.

$\because$ n=800, p=0.45

$\therefore \hat{\sigma} = \sqrt{\frac{p(1-p)}{n}} = 0.0176$

$\because Z = \frac{\hat{p}-p}{\hat{\sigma}} \overset{d}{\approx} N(0,1)$

$\therefore P(Z \leq z_{\alpha/2}) = 0.01$ and $P(Z \geq z_{1-\alpha/2}) = 0.01$, $z_{\alpha/2} = -z_{1-\alpha/2} = -2.32$

$\therefore z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}$

$\therefore 0.409 = \hat{p} - \hat{\sigma} z_{\alpha/2} \leq p \leq p + \hat{\sigma} z_{\alpha/2} = 0.491$

b.

$\because p \leq 0.5$, less than half think it better.

$\therefore$ a majority of adults say that movies are not getting better.

Derive a one-sided $(1-\alpha)$**-CI** $(-\infty, U]$ **for** μ

$\because P(T_n = \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \geq t_{\alpha,(n-1)}) = 1 - \alpha$

$\therefore \mu \leq \bar{Y} - t_{\alpha,(n-1)} \frac{\sigma}{\sqrt{n}}$

$\because t_{\alpha,(n-1)} \leq 0$

$\therefore \mu \in (-\infty, \bar{Y} + |t_{\alpha,(n-1)}|\frac{\sigma}{\sqrt{n}}]$

8.90 Do SAT scores for high school students differ depending on the students' intended field of study? Fifteen students who intended to major in engineering were compared with 15 students who intended to major in language and literature. Given in the accompanying table are the means and standard deviations of the scores on the verbal and mathematics portion of the SAT for the two groups of students:[16]

	Verbal		Math	
Engineering	$\overline{y} = 446$	$s = 42$	$\overline{y} = 548$	$s = 57$
Language/literature	$\overline{y} = 534$	$s = 45$	$\overline{y} = 517$	$s = 52$

a Construct a 95% confidence interval for the difference in average verbal scores of students majoring in engineering and of those majoring in language/literature.

b Construct a 95% confidence interval for the difference in average math scores of students majoring in engineering and of those majoring in language/literature.

c Interpret the results obtained in parts (a) and (b).

d What assumptions are necessary for the methods used previously to be valid?

a.

$\bar{Y}_{e,v} = 446$, $S_{e,v} = 42$, $n_1 = 15$, and $\bar{Y}_{l,v} = 534$, $S_{l,v} = 45$, $n_2 = 15$.

$\because S^2_{e,l,verbal} = \frac{(n_1-1)S^2_{e,v}+(n_2-1)S^2_{l,v}}{30-2} = 1894.5$

$\therefore (\bar{Y}_{e,v} - \bar{Y}_{l,v}) \pm t_{\alpha/2,28}\sqrt{1894.5(1/15+1/15)}$

$\therefore (\mu_{e,v} - \mu_{l,v}) \in [-88 - 2.048407 * 15.89339 = -120, -88 + 2.048407 * 15.89339 = -55]$

b.

$\bar{Y}_{e,m} = 548$, $S_{e,m} = 57$, $n_1 = 15$, and $\bar{Y}_{l,m} = 517$, $S_{l,m} = 52$, $n_2 = 15$.

$\because S^2_{e,l,math} = \frac{(n_1-1)S^2_{e,m}+(n_2-1)S^2_{l,m}}{30-2} = 2976.5$

$\therefore (\bar{Y}_{e,m} - \bar{Y}_{l,m}) \pm t_{\alpha/2,28}\sqrt{2976.5(1/15+1/15)}$

$\therefore (\mu_{e,m} - \mu_{l,m}) \in [31 - 2.048407 * 19.92151 = -10, 31 + 2.048407 * 19.92151 = 72]$

c.

$(\mu_{e,v} - \mu_{l,v}) \in [-120, -55]$ and $(\mu_{e,m} - \mu_{l,m}) \in [-10, 72]$

$\therefore$ Engineering has lower average scores of Verbal than Language/literature, and Engineering has similar average scores of Math to Language/literature

d.

These methods need the SAT scores of Engineering and Language/literature students follow Normal distribution.

22 Small-sample tests

Under H_0,

$$T = \frac{\bar{Y} - \mu_0}{\hat{\sigma}/\sqrt{n}} \sim t(n-1)$$

, where $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2$ is the unbiased estimate of σ^2

Two-sample meas Assumption: $\{X_1, ..., X_{n1}\}$ IID $N(\mu_1, \sigma^2)$, $\{Y_1, ..., Y_{n2}\}$ IID $N(\mu_2, \sigma^2)$, X_i's are independent of Y_i's.

$$H_0 : \mu_1 - \mu_2 = \delta_0 \text{ vs } H_a : \begin{cases} \mu_1 - \mu_2 > \delta_0; \\ \\ \mu_1 - \mu_2 \neq \delta_0; \\ \\ \mu_1 - \mu_2 < \delta_0; \end{cases} \quad \text{vs } RR = \begin{cases} t > t_{\alpha(\nu)}; \\ \\ t < -t_{\alpha/2(\nu)} \text{ or } t > t_{\alpha/2(\nu)}; \\ \\ t < -t_{\alpha(\nu)}; \end{cases}$$

$$\nu = n_1 + n_2 - 2$$

$$T = \frac{\bar{X} - \bar{Y} - \delta_0}{\hat{\sigma}/\sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2), \text{ under } H_0$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{n_1 + n_2 - 2}$$

22.1 Exercise

10.17 A survey published in the *American Journal of Sports Medicine*[2] reported the number of meters (m) per week swum by two groups of swimmers—those who competed exclusively in breaststroke and those who competed in the individual medley (which includes breaststroke). The number of meters per week practicing the breaststroke was recorded for each swimmer, and the summary statistics are given below. Is there sufficient evidence to indicate that the average number of meters per week spent practicing breaststroke is greater for exclusive breaststrokers than it is for those swimming individual medley?

	Specialty	
	Exclusively Breaststroke	Individual Medley
Sample size	130	80
Sample mean (m)	9017	5853
Sample standard deviation (m)	7162	1961
Population mean	μ_1	μ_2

a State the null and alternative hypotheses.

b What is the appropriate rejection region for an $\alpha = .01$ level test?

c Calculate the observed value of the appropriate test statistic.

d What is your conclusion?

a.

Null hypotheses: $\mu_1 \leq \mu_2 \Rightarrow \theta_0 = \mu_1 - \mu_2 \leq 0$, alternative hypotheses: $\mu_1 > \mu_2 \Rightarrow \theta_a = \mu_1 - \mu_2 > 0$

b.

$\because \hat{\sigma}^2 = \frac{\sum(X_i-\bar{X})^2+\sum(Y_i-\bar{Y})^2)}{n_1+n_2-2} = 33272854$

$\therefore \hat{\sigma} = 5768.263$, and $\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 0.1420996$

$\therefore T = \frac{\bar{X}-\bar{Y}-\delta}{5768.263/0.1420996} \sim t(130+80-2)$

$\because |t_{208,0.01}| = 2.3444$

$\therefore$ the rejection region for T, is $RR \in [2.3444, +\infty)$

c.

$\because T = \frac{\bar{X}-\bar{Y}-\delta}{5768.263/0.1420996} = \frac{9017-5853}{40593.1} = 0.07794428$

d.

Fail to reject H_0 – not enough evidence to conclude the mean distance for breaststroke is larger than individual medley.

10.30 A manufacturer claimed that at least 20% of the public preferred her product. A sample of 100 persons is taken to check her claim. With $\alpha = .05$, how small would the sample percentage need to be before the claim could legitimately be refuted? (Notice that this would involve a one-tailed test of the hypothesis.)

When $\theta_0 = 0.2$, sample size = 100, then $\sigma^2 = \frac{p(1-p)}{n}$, $\sigma = 0.04$, One-tail reject region = $[0, c]$

$$\begin{aligned}
\therefore \alpha(\theta = 0.2) &= P_\theta(\hat{\theta} \in RR) \\
&= P_\theta(\hat{\theta} \leq c) \\
&= P_\theta(\frac{\hat{\theta}-0.2}{0.04} \leq \frac{c-0.2}{0.04}) \\
&= \alpha \\
&= 0.05
\end{aligned}$$

$\because z_{0.05} = -1.645$

$\therefore \frac{c-0.2}{0.04} \leq -1.645$

$\therefore c \leq 0.1342$

$\therefore RR \in [0, 0.1342]$

10.46 A large-sample α-level test of hypothesis for $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ rejects the null hypothesis if

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_\alpha.$$

Show that this is equivalent to rejecting H_0 if θ_0 is less than the large-sample $100(1-\alpha)\%$ lower confidence bound for θ.

$\because \frac{\hat{\theta}-\theta_0}{\alpha_{\hat{\theta}}} > z_\alpha$

$\therefore \hat{\theta} - \theta_0 > \alpha_{\hat{\theta}} z_\alpha$

$\therefore \hat{\theta} - \alpha_{\hat{\theta}} z_\alpha > \theta_0$

10.48 A large-sample α-level test of hypothesis for $H_0 : \theta = \theta_0$ versus $H_a : \theta < \theta_0$ rejects the null hypothesis if

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} < -z_\alpha.$$

Show that this is equivalent to rejecting H_0 if θ_0 is greater than the large-sample $100(1-\alpha)\%$ upper confidence bound for θ.

$\because \frac{\hat{\theta}-\theta_0}{\alpha_{\hat{\theta}}} < -z_\alpha$

$\therefore \hat{\theta} - \theta_0 < -\alpha_{\hat{\theta}} z_\alpha$

$\therefore \hat{\theta} + \alpha_{\hat{\theta}} z_\alpha < \theta_0$

$$RR = \begin{cases} t > t_{\alpha(n-1)}; \\ \\ |t| > t_{\alpha/2(n-1)}; \\ \\ t < -t_{\alpha(n-1)}; \end{cases}$$

Find the CIs of μ corresponding to the three rejection regions above respectively.

Answer:

$$(1-\alpha)\%\text{CI} = \begin{cases} (-\infty, \mu + \frac{\hat{\sigma}}{\sqrt{n}} t_{\alpha(n-1)}); \\ (\mu - \frac{\hat{\sigma}}{\sqrt{n}} t_{\alpha/2(n-1)}, \mu + \frac{\hat{\sigma}}{\sqrt{n}} t_{\alpha/2(n-1)}); \\ (\mu - \frac{\hat{\sigma}}{\sqrt{n}} t_{\alpha(n-1)}, \infty,); \end{cases}$$

10.72 An Article in *American Demographics* investigated consumer habits at the mall. We tend to spend the most money when shopping on weekends, particularly on Sundays between 4:00 and 6:00 P.M. Wednesday-morning shoppers spend the least.[15] Independent random samples of weekend and weekday shoppers were selected and the amount spent per trip to the mall was recorded as shown in the following table:

Weekends	Weekdays
$n_1 = 20$	$n_2 = 20$
$\overline{y}_1 = \$78$	$\overline{y}_2 = \$67$
$s_1 = \$22$	$s_2 = \$20$

a Is there sufficient evidence to claim that there is a difference in the average amount spent per trip on weekends and weekdays? Use $\alpha = .05$.

b What is the attained significance level?

a:

$H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 \neq 0$

$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = 442$

$$T = \frac{78 - 67 - 0}{\hat{\sigma}\sqrt{1/20 + 1/20}} = \frac{11}{21 * 0.316} = 1.658$$

$\because t_{0.025,(38)} = -2.02$

$\therefore$ no sufficient evidence to claim that there is a difference in the average amount spent on weekends and weekdays.

10.51 Two sets of elementary schoolchildren were taught to read by using different methods, 50 by each method. At the conclusion of the instructional period, a reading test yielded the results $\overline{y}_1 = 74$, $\overline{y}_2 = 71$, $s_1 = 9$, and $s_2 = 10$.

a What is the attained significance level if you wish to see whether evidence indicates a difference between the two population means?

b What would you conclude if you desired an α-value of .05?

a. Let the attained significant level be $\alpha = 0.05$

b.

$H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 \neq 0$

$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = 90.5$

$T = \frac{74-71-0}{\hat{\sigma}\sqrt{1/50+1/50}} = \frac{3}{9.51*0.2} = 1.579$

$\because t_{0.025,(98)} = -1.98$

$\therefore$ there no sufficient evidence to claim that there is a difference in the average scores of two methods.

23 Variance

Assumption: $\{Y_1, ..., Y_n\}$ IID $N(\mu_2, \sigma^2)$

$$H_0 : \sigma^2 = \sigma_0^2 \text{ vs } H_a : \begin{cases} \sigma^2 > \sigma_0^2; \\ \sigma^2 \neq \sigma_0^2; \\ \sigma^2 < \sigma_0^2; \end{cases} \quad \text{vs } RR = \begin{cases} t > \chi^2_\alpha(n-1); \\ t > \chi^2_{\alpha/2}(n-1) \text{ or } t < \chi^2_{1-\alpha/2}(n-1); \\ t < \chi^2_{1-\alpha}(n-1); \end{cases}$$

$$T = \frac{\sum_{i=1}^{n_1}(Y_i - \bar{Y})^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1), \text{ under } H_0$$

where $\chi^2_\alpha(\nu)$ denotes the (1-α)-quantile.

23.1 Exercise

8.102 The ages of a random sample of five university professors are 39, 54, 61, 72, and 59. Using this information, find a 99% confidence interval for the population standard deviation of the ages of all professors at the university, assuming that the ages of university professors are normally distributed.

$\because T = \frac{\sum\limits_{i=1}^{n}(Y_i - \bar{Y}_n)^2}{\sigma^2} = \frac{(n-1)\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-1=4)$

$\because \sum\limits_{i=1}^{n}(Y_i - \bar{Y}_n)^2 = \hat{\sigma}^2(5-1)$

$\therefore T = 144.5001 * 4/\sigma^2$

$\chi^2_{0.005,4} = 0.20699$ and $\chi^2_{0.005,4} = 14.8602$

$\because 144.5001 \times 4/14.8602 = 38.89587, 144.5001 \times 4/0.20699 = 2792.407$

$\therefore \sigma^2 \in [38.89587, 2792.407]$

24 The p-values

Let two-sided α level Z-test of $H_0 : \theta = \theta_0$ vs $H_a : \theta \neq \theta_0$, reject H_0 if

$$|Z = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}}| > z_{\alpha/2}$$

1. As $\alpha \uparrow$, $z_{\alpha/2} \downarrow$, so RR expands.
2. Expands or shrink RR_α, such that RR_α just exclude Z. Denote this α as `p`.
3. This p is the smallest α could be if we want still rejecting H_0

Definition 11.6.2: In a HT, the α-level rejection regions RR_α are said to be nested, if $RR_\alpha \subset RR_{\alpha'}$ when $\alpha \leq \alpha'$, namely, the rejection region expands as $\alpha \uparrow$.

Definition 11.6.3: In a HT with nested rejection regions RR_α, the p-value is defined as the random variable $\hat{p} = \min\{\alpha : T \in RR_\alpha\}$. Because T is statistic, so $\hat{p}$ is also a statistic.

Let $u \in [0, 1]$, $\hat{p} = \min\{\alpha : T \in RR_\alpha\} > u \Longleftrightarrow T \notin RR_u$. Then $P_{\theta_0}(\hat{p} > u) = P_{\theta_0}(T \notin RR_u) = 1 - u$.

Hence $P_{\theta_0}(\hat{p} \leq u) = u$, it is a Uniform(0,1) distribution.

Hence an observed p-value can be interpreted as: the probability of the test statistic being more extreme than the observed value under H0.

24.1 Exercise

Provide an elementary explanation of p-value to someone with little statistical training.

Answer:

Under H0, the probability of the test statistic being more extreme than the observed value. For example, if we think x = y, then x - y should be 0. If we observed x - y = c, c > 0, then the probability of all possible results of x - y belong to [c, ∞].

25 Optimal: UMP

The goodness of a test is measured by α and β , the probabilities of type I and type II errors, respectively. Typically, the value of α is chosen in advance and determines the location of the rejection region. A related but more useful concept for evaluating the performance of a test is called the power of the test. Basically, the power of a test is the probability that the test will lead to rejection of the null hypothesis.

Suppose that W is the test statistic and RR is the rejection region for a test of a hypothesis involving the value of a parameter θ. Then the power of the test, denoted by power (θ) , is the probability that the test will lead to rejection of H_0 when the actual parameter value is θ . That is,

$$\text{power}(\theta) = P(\text{W in RR when the parameter value is } \theta)$$

Relationship Between Power and β: If θ_a is a value of θ in the alternative hypothesis H_a , then

$$\text{power}(\theta_a) = 1 - \beta(\theta_a)$$

Selecting tests with the smallest possible value of β for tests where α, the probability of a type I error, is a fixed value selected by the researcher.

THEOREM 10.1 **Neyman-Pearson Lemma**: Suppose that we wish to test the simple null hypothesis $H_0 : \theta = \theta_0$ versus the simple alternative hypothesis $H_a : \theta = \theta_a$, based on a random sample $Y_1, Y_2, ..., Y_n$ from a distribution with parameter θ. Let $L(\theta)$ denote the likelihood of the sample when the value of the parameter is θ. Then, for a given α, the test that maximizes the power at θ_a has a rejection region, **RR**, determined by

$$\frac{L(\theta_0)}{L(\theta_a)} < k$$

The value of k is chosen so that the test has the desired value for α. Such a test is a most powerful α - level test for H_0 versus H_a.

Uniformly most powerful test

EXAMPLE **10.22** Suppose that Y represents a single observation from a population with probability density function given by

$$f(y \mid \theta) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the most powerful test with significance level $\alpha = .05$ to test $H_0 : \theta = 2$ versus $H_a : \theta = 1$.

Solution Because both of the hypotheses are simple, Theorem 10.1 can be applied to derive the required test. In this case,

$$\frac{L(\theta_0)}{L(\theta_a)} = \frac{f(y|\theta_0)}{f(y|\theta_a)} = \frac{2y}{(1)y^0} = 2y, \qquad \text{for } 0 < y < 1,$$

and the form of the rejection region for the most powerful test is

$$2y < k.$$

Equivalently, the rejection region RR is $\{y < k/2\}$. Or because $k/2 = k^*$, a constant, the rejection region is RR: $\{y < k^*\}$.

Because $\alpha = .05$ is specified, the value of k^* is determined by

$$.05 = P(Y \text{ in RR when } \theta = 2) = P(Y < k^* \text{ when } \theta = 2) = \int_0^{k^*} 2y \, dy = (k^*)^2.$$

Therefore, $(k^*)^2 = .05$, and the rejection region of the most powerful test is

$$\text{RR}: \{y < \sqrt{.05} = .2236\}.$$

Among all tests for H_0 versus H_a based on a sample size of 1 and with α fixed at .05, this test has the largest possible value for power(θ_a) = power(1). Equivalently, among all tests with $\alpha = .05$ this test has the smallest type II error probability when $\beta(\theta_a)$ is evaluated at $\theta_a = 1$. What is the actual value for power(θ) when $\theta = 1$?

$$\begin{aligned} \text{power}(1) &= P(Y \text{ in RR when } \theta = 1) = P(Y < .2236 \text{ when } \theta = 1) \\ &= \int_0^{.2236} (1) \, dy = .2236. \end{aligned}$$

Even though the rejection region $\{y < .2236\}$ gives the *maximum* value for power(1) among all tests with $\alpha = .05$, we see that $\beta(1) = 1 - .2236 = .7764$ is still very large.

EXAMPLE **10.23** Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with unknown mean μ and known variance σ^2. We wish to test $H_0 : \mu = \mu_0$ against $H_a : \mu > \mu_0$ for a specified constant μ_0. Find the uniformly most powerful test with significance level α.

Solution We begin by looking for the most powerful α-level test of $H_0 : \mu = \mu_0$ versus $H_a^* : \mu = \mu_a$ for one fixed value of μ_a that is larger than μ_0. Because

$$f(y \mid \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) \exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right], \qquad -\infty < y < \infty,$$

we have

$$L(\mu) = f(y_1 \mid \mu) f(y_2 \mid \mu) \cdots f(y_n \mid \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(y_i-\mu)^2}{2\sigma^2}\right].$$

[Recall that $\exp(w)$ is simply e^w in another form.] Because both H_0 and H_a^* are *simple* hypotheses, Theorem 10.1, implies that the most powerful test of $H_0 : \mu = \mu_0$ versus $H_a^* : \mu = \mu_a$ is given by

$$\frac{L(\mu_0)}{L(\mu_a)} < k,$$

which in this case is equivalent to

$$\frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(y_i-\mu_0)^2}{2\sigma^2}\right]}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(y_i-\mu_a)^2}{2\sigma^2}\right]} < k.$$

This inequality can be rearranged as follows:

$$\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu_0)^2 - \sum_{i=1}^{n}(y_i-\mu_a)^2\right]\right\} < k.$$

Taking natural logarithms and simplifying, we have

$$-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu_0)^2-\sum_{i=1}^{n}(y_i-\mu_a)^2\right] < \ln(k)$$

$$\sum_{i=1}^{n}(y_i-\mu_0)^2-\sum_{i=1}^{n}(y_i-\mu_a)^2 > -2\sigma^2\ln(k)$$

$$\sum_{i=1}^{n}y_i^2-2n\overline{y}\mu_0+n\mu_0^2-\sum_{i=1}^{n}y_i^2+2n\overline{y}\mu_a-n\mu_a^2 > -2\sigma^2\ln(k)$$

$$\overline{y}(\mu_a-\mu_0) > \frac{-2\sigma^2\ln(k)-n\mu_0^2+n\mu_a^2}{2n}$$

or, since $\mu_a > \mu_0$,

$$\overline{y} > \frac{-2\sigma^2\ln(k)-n\mu_0^2+n\mu_a^2}{2n(\mu_a-\mu_0)}.$$

Because σ^2, n, μ_0, and μ_a are all known constants, the quantity on the right-hand side of this inequality is a constant—call it k'. Therefore, the most powerful test of $H_0: \mu = \mu_0$ versus $H_a^*: \mu = \mu_a$ has the rejection region given by

$$\text{RR} = \{\overline{y} > k'\}.$$

The precise value of k' is determined by fixing α and noting that

$$\begin{aligned}\alpha &= P(\overline{Y} \text{ in RR when } \mu = \mu_0)\\ &= P(\overline{Y} > k' \text{ when } \mu = \mu_0)\\ &= P\left(\frac{\overline{Y}-\mu_0}{\sigma/\sqrt{n}} > \frac{k'-\mu_0}{\sigma/\sqrt{n}}\right)\\ &= P\left(Z > \sqrt{n}(k'-\mu_0)/\sigma\right).\end{aligned}$$

Because, under H_0, Z has a standard normal distribution, $P(Z > z_\alpha) = \alpha$ and the required value for k' must satisfy

$$\sqrt{n}(k'-\mu_0)/\sigma = z_\alpha, \quad \text{or equivalently,} \quad k' = \mu_0 + z_\alpha\sigma/\sqrt{n}.$$

Thus, the α-level test that has the largest possible value for power(θ_a) is based on the statistic $\overline{Y}$ and has rejection region RR = $\{\overline{y} > \mu_0 + z_\alpha\sigma/\sqrt{n}\}$. We now observe that neither the test statistic nor the rejection region for this α-level test depends on the particular value assigned to μ_a. That is, for any value of μ_a greater than μ_0, we obtain exactly the same rejection region. Thus, the α-level test with the rejection region previously given has the largest possible value for power(μ_a) for *every* $\mu_a > \mu_0$. It is the *uniformly most powerful* test for $H_0: \mu = \mu_0$ versus $H_a: \mu > \mu_0$. This is exactly the test that we considered in Section 10.3.

The Neyman–Pearson lemma is useless if we wish to test a hypothesis about a single parameter θ when the sampled distribution contains other unspecified parameters.

Assumptions:

1. $H_0: \theta = \theta_0$ vs $H_a: \theta = \theta_a$
2. Level: α
3. $L(\theta)$ likelihood function, $\theta \in \{\theta_0, \theta_a\}$

25.1 Exercise

10.96 Suppose Y is a random sample of size 1 from a population with density function

$$f(y \mid \theta) = \begin{cases} \theta y^{\theta-1}, & 0 \le y \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

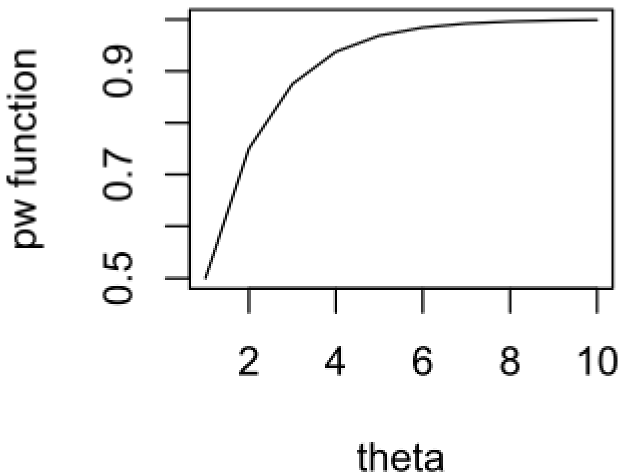
where $\theta > 0$.

a Sketch the power function of the test with rejection region: $Y > .5$.

b Based on the single observation Y, find a uniformly most powerful test of size α for testing $H_0 : \theta = 1$ versus $H_a : \theta > 1$.

a.

$P_\theta(y \in RR) = P_\theta(y > 0.5) = \int\limits_{0.5}^{1} \theta y^{\theta-1} dy = 1 - (0.5)^\theta$



b.

$\because \frac{L(\theta_0)}{L(\theta_a)} = \frac{L(1)}{\theta_a y^{\theta_a - 1}}$

Let $T = \frac{L(1)}{\theta_a y^{\theta_a - 1}} < k$, then $P_{\theta_0}(y > (\frac{1}{k\theta_a})^{1/(\theta_a - 1)} = c) = \alpha$

$\because P_{\theta_0}(y > (\frac{1}{k\theta_a})^{1/(\theta_a - 1)} = c) = \int\limits_{c}^{1} \theta y^{\theta-1} = 1 - c = \alpha$

$\therefore RR = y > 1 - \alpha$, not depend on a specific θ_a

$\therefore y > 1 - \alpha$ is a uniformly most powerful (UMP) decision rule.

26 Likelihood ratio test

The procedure works for simple or composite hypotheses and whether or not other parameters with unknown values are present.

A likelihood Ratio Test: $H_0 : \Theta \in \Omega_0$ v.s. $H_a : \Theta \in \Omega_a$ employs λ as a test statistic, and the rejection region is determined by $\lambda \leq k$.

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{\max_{\Theta \in \Omega_0} L(\Theta)}{\max_{\Theta \in \Omega} L(\Theta)}$$

Let $L(\hat{\Omega}_0)$ denote the maximum (actually the supremum) of the likelihood function for all $\Theta \in \Omega_0$. That is, $L(\hat{\Omega}_0) = \max_{\Theta \in \Omega_0} L(\Theta)$. Notice that $L(\hat{\Omega}_0)$ represents the best explanation for the observed data for all $\Theta \in \Omega_0$ and can be found by using methods similar to those used in Section 9.7. Similarly, $L(\hat{\Omega}) = \max_{\Theta \in \Omega} L(\Theta)$ represents the best explanation for the observed data for all $\Theta \in \Omega = \Omega_0 \cup \Omega_a$. If $L(\hat{\Omega}_0) = L(\hat{\Omega})$, then a best explanation for the observed data can be found inside Ω_0, and we should not reject the null hypothesis $H_0 : \Theta \in \Omega_0$. However, if $L(\hat{\Omega}_0) < L(\hat{\Omega})$, then the best explanation for the observed data can be found inside Ω_a, and we should consider rejecting H_0 in favor of H_a. A likelihood ratio test is based on the ratio $L(\hat{\Omega}_0)/L(\hat{\Omega})$.

EXAMPLE **10.24** Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. We want to test $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$. Find the appropriate likelihood ratio test.

Solution In this case, $\Theta = (\mu, \sigma^2)$. Notice that Ω_0 is the set $\{(\mu_0, \sigma^2) : \sigma^2 > 0\}$, $\Omega_a = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma^2 > 0\}$, and hence that $\Omega = \Omega_0 \cup \Omega_a = \{(\mu, \sigma^2) : \mu \geq \mu_0, \sigma^2 > 0\}$. The constant value of the variance σ^2 is completely unspecified. We must now find $L(\hat{\Omega}_0)$ and $L(\hat{\Omega})$.

For the normal distribution, we have

$$L(\Theta) = L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2}\right].$$

Restricting μ to Ω_0 implies that $\mu = \mu_0$, and we can find $L(\hat{\Omega}_0)$ if we determine the value of σ^2 that maximizes $L(\mu, \sigma^2)$ subject to the constraint that $\mu = \mu_0$. From Example 9.15, we see that when $\mu = \mu_0$ the value of σ^2 that maximizes $L(\mu_0, \sigma^2)$ is

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_0)^2.$$

Thus, $L(\hat{\Omega}_0)$ is obtained by replacing μ with μ_0 and σ^2 with $\hat{\sigma}_0^2$ in $L(\mu, \sigma^2)$, which gives

$$L(\hat{\Omega}_0) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_0^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n}\frac{(y_i - \mu_0)^2}{2\hat{\sigma}_0^2}\right] = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_0^2}\right)^{n/2} e^{-n/2}.$$

We now turn to finding $L(\hat{\Omega})$. As in Example 9.15, it is easier to look at $\ln L(\mu,\ \sigma^2)$,

$$\ln[L(\mu,\ \sigma^2)] = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2.$$

Taking derivatives with respect to μ and σ^2, we obtain

$$\frac{\partial\{\ln[L(\mu,\ \sigma^2)]\}}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu),$$

$$\frac{\partial\{\ln[L(\mu,\ \sigma^2)]\}}{\partial\sigma^2} = -\left(\frac{n}{2\sigma^2}\right) + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2.$$

We need to find the maximum of $L(\mu,\ \sigma^2)$ over the set $\Omega = \{(\mu, \sigma^2) : \mu \geq \mu_0,\ \sigma^2 > 0\}$. Notice that

$$\begin{aligned}
\partial L(\mu,\ \sigma^2)/\partial\mu &< 0, & &\text{if } \mu > \overline{y},\\
\partial L(\mu,\ \sigma^2)/\partial\mu &= 0, & &\text{if } \mu = \overline{y},\\
\partial L(\mu,\ \sigma^2)/\partial\mu &> 0, & &\text{if } \mu < \overline{y}.
\end{aligned}$$

Thus, over the set $\Omega = \{(\mu, \sigma^2) : \mu \geq \mu_0,\ \sigma^2 > 0\}$, $\ln L(\mu,\ \sigma^2)$ [and also $L(\mu,\ \sigma^2)$] is maximized at $\hat{\mu}$ where

$$\hat{\mu} = \begin{cases} \overline{y}, & \text{if } \overline{y} > \mu_0,\\ \mu_0, & \text{if } \overline{y} \leq \mu_0. \end{cases}$$

Just as earlier, the value of σ^2 in Ω that maximizes $L(\mu,\ \sigma^2)$, is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2.$$

$L(\hat{\Omega})$ is obtained by replacing μ with $\hat{\mu}$ and σ^2 with $\hat{\sigma}^2$, which yields

$$L(\hat{\Omega}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n}\frac{(y_i - \hat{\mu})^2}{2\hat{\sigma}^2}\right] = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}^2}\right)^{n/2} e^{-n/2}.$$

Thus,

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}$$

$$= \begin{cases} \left[\dfrac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \mu_0)^2}\right]^{n/2}, & \text{if } \overline{y} > \mu_0 \\ 1, & \text{if } \overline{y} \le \mu_0. \end{cases}$$

Notice that λ is always less than or equal to 1. Thus, "small" values of λ are those less than some $k < 1$. Because

$$\sum_{i=1}^{n}(y_i - \mu_0)^2 = \sum_{i=1}^{n}[(y_i - \overline{y}) + (\overline{y} - \mu_0)]^2$$

$$= \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - \mu_0)^2$$

if $k < 1$, it follows that the rejection region, $\lambda \le k$, is equivalent to

$$\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \mu_0)^2} < k^{2/n} = k'$$

$$\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - \mu_0)^2} < k'$$

$$\frac{1}{1 + \dfrac{n(\overline{y} - \mu_0)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}} < k'.$$

This inequality in turn is equivalent to

$$\frac{n(\overline{y} - \mu_0)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} > \frac{1}{k'} - 1 = k''$$

$$\frac{n(\overline{y} - \mu_0)^2}{\dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2} > (n-1)k''$$

or, because $\overline{y} > \mu_0$ when $\lambda < k < 1$,

$$\frac{\sqrt{n}(\overline{y} - \mu_0)}{s} > \sqrt{(n-1)k''},$$

where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

Notice that $\sqrt{n}(\overline{Y} - \mu_0)/S$ is the t statistic employed in previous sections. Consequently, the likelihood ratio test is equivalent to the t test of Section 10.8.

26.1 Exercise

10.102 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Bernoulli-distributed population with parameter p. That is,

$$p(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}, \qquad y_i = 0, 1.$$

a Suppose that we are interested in testing $H_0 : p = p_0$ versus $H_a : p = p_a$, where $p_0 < p_a$.

i Show that

$$\frac{L(p_0)}{L(p_a)} = \left[\frac{p_0(1-p_a)}{(1-p_0)p_a}\right]^{\sum y_i} \left(\frac{1-p_0}{1-p_a}\right)^n.$$

ii Argue that $L(p_0)/L(p_a) < k$ if and only if $\sum_{i=1}^n y_i > k^*$ for some constant k^*.

iii Give the rejection region for the most powerful test of H_0 versus H_a.

b Recall that $\sum_{i=1}^n Y_i$ has a binomial distribution with parameters n and p. Indicate how to determine the values of any constants contained in the rejection region derived in part [a(iii)].

c Is the test derived in part (a) uniformly most powerful for testing $H_0 : p = p_0$ versus $H_a : p > p_0$? Why or why not?

a.

b.

$\because L(p) = p^{\sum y_i}(1-p)^{n-\sum y_i}$

$T = \frac{L(p_0)}{L(p_a)} = \frac{p_0^{\sum y_i}(1-p_0)^{n-\sum y_i}}{p_a^{\sum y_i}(1-p_a)^{n-\sum y_i}} = (\frac{p_0(1-p_a)}{p_a(1-p_0)})^{\sum y_i}(\frac{1-p_0}{1-p_a})^n$

ii.

$\because T = (\frac{p_0(1-p_a)}{p_a(1-p_0)})^{\sum y_i}(\frac{1-p_0}{1-p_a})^n < k$

$\therefore \ln(T) = \sum y_i \ln(\frac{p_0(1-p_a)}{p_a(1-p_0)}) + n\ln(\frac{1-p_0}{1-p_a}) < \ln(k)$

$\therefore \sum y_i > \left(\ln(k) - n\ln(\frac{1-p_0}{1-p_a})\right)\left(\ln(\frac{p_0(1-p_a)}{p_a(1-p_0)})\right)^{-1} = k^*$

iii.

$P_{\theta_0}(T < k_a) = P_{\theta_0}(\sum y_i > k^*)$

$\therefore$RR is $\sum y_i > k^*$

b.

$P_{\theta_0}(T < k_a) = P_{\theta_0}(\sum y_i > k^*) = \alpha$, $\sum y_i$ is binomial distribution with parameters n and p_0

c.

Because $P_{\theta_0}(\sum y_i > k^*) = \alpha$ can solved the value of k^*, therefore RR not depend on p_a, and it is a `uniformly most powerful (UMP)` decision rule.

10.111 Suppose that we are interested in testing the *simple* null hypothesis $H_0: \theta = \theta_0$ versus the *simple* alternative hypothesis $H_a: \theta = \theta_a$. According to the Neyman–Pearson lemma, the test that maximizes the power at θ_a has a rejection region determined by

$$\frac{L(\theta_0)}{L(\theta_a)} < k.$$

In the context of a likelihood ratio test, if we are interested in the *simple* H_0 and H_a, as stated, then $\Omega_0 = \{\theta_0\}$, $\Omega_a = \{\theta_a\}$, and $\Omega = \{\theta_0, \theta_a\}$.

a Show that the likelihood ratio λ is given by

$$\lambda = \frac{L(\theta_0)}{\max\{L(\theta_0), L(\theta_a)\}} = \frac{1}{\max\left\{1, \dfrac{L(\theta_a)}{L(\theta_0)}\right\}}.$$

b Argue that $\lambda < k$ if and only if, for some constant k',

$$\frac{L(\theta_0)}{L(\theta_a)} < k'.$$

c What do the results in parts (a) and (b) imply about likelihood ratio tests when both the null and alternative hypotheses are simple?

a.

$\because \max_{\theta \in \theta_0} L(\theta_0) \leq \max_{\theta \in \theta_a} L(\theta_a)$

$\therefore \lambda = \frac{L(\theta_0)}{\max[L(\theta_0), L(\theta_a)]} = \frac{1}{\max[1, L(\theta_a)/L(\theta_0)]}$

b.

$\because \lambda = \frac{1}{\max[1, L(\theta_a)/L(\theta_0)]} = \min[1, L(\theta_0)/L(\theta_a)]$

if $\lambda < k$, then $L(\theta_0)/L(\theta_a)$ should be also smaller than some value k'

if $\frac{L(\theta_0)}{L(\theta_a)} < k' \leq= 1$

c.

LRT coincides with the test given in the Neyman-Pearson Lemma, namely, LRT is the most powerful test in this case.

10.108 Suppose that $X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2}$, and $W_1, W_2, \ldots, W_{n_3}$ are independent random samples from normal distributions with respective unknown means μ_1, μ_2, and μ_3 and variances σ_1^2, σ_2^2, and σ_3^2.

a Find the likelihood ratio test for $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ against the alternative of at least one inequality.

b Find an approximate critical region for the test in part (a) if n_1, n_2, and n_3 are large and $\alpha = .05$.

a.

Because for normal distribution

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

$$\therefore L(\mu_1,\mu_2,\mu_3,\sigma_1^2,\sigma_2^2,\sigma_3^2) = \prod_{i=1}^{n_1} f(\mu_1,\sigma_1)\prod_{i=1}^{n_2} f(\mu_2,\sigma_2)\prod_{i=1}^{n_3} f(\mu_3,\sigma_3)$$
$$= (\frac{1}{\sqrt{2\pi}})^{(n_1+n_2+n_3)}\frac{1}{\sigma_1^{n_1}\sigma_2^{n_2}\sigma_3^{n_3}}\exp\Big(-\frac{\sum_{n_1}(x_i-\mu_1)^2}{2\sigma_1^2}-\frac{\sum_{n_2}(y_i-\mu_2)^2}{2\sigma_2^2}-\frac{\sum_{n_3}(w_i-\mu_3)^2}{2\sigma_3^2}\Big)$$
$$= (\frac{1}{\sqrt{2\pi}})^{(n_1+n_2+n_3)}\frac{1}{\sigma_1^{n_1}\sigma_2^{n_2}\sigma_3^{n_3}}\exp\Big(-\frac{\sigma_1^2 n_1}{2\sigma_1^2}-\frac{\sigma_2^2 n_2}{2\sigma_2^2}-\frac{\sigma_3^2 n_3}{2\sigma_3^2}\Big)$$

$$\ell(\mu_1,\mu_2,\mu_3,\sigma_1^2,\sigma_2^2,\sigma_3^2) = \ln\Big(L(\mu_1,\mu_2,\mu_3,\sigma_1^2,\sigma_2^2,\sigma_3^2)\Big)$$
$$= (n_1+n_2+n_3)\ln(\frac{1}{\sqrt{2\pi}})-\frac{n_1}{2}\ln(\sigma_1^2)-\frac{n_2}{2}\ln(\sigma_2^2)-\frac{n_3}{2}\ln(\sigma_3^2)-$$
$$\frac{\sum_{n_1}(x_i-\mu_1)^2}{2\sigma_1^2}-\frac{\sum_{n_2}(y_i-\mu_2)^2}{2\sigma_2^2}-\frac{\sum_{n_3}(w_i-\mu_3)^2}{2\sigma_3^2}$$

Under $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$

$$\therefore \frac{\partial\ell(\mu_1,\mu_2,\mu_3,\sigma_1^2,\sigma_2^2,\sigma_3^2)}{\partial\sigma^2} = -\frac{n_1+n_2+n_3}{2\sigma^2}+\frac{\sum_{n_1}(x_i-\mu_1)^2}{2\sigma_1^4}+\frac{\sum_{n_2}(y_i-\mu_2)^2}{2\sigma_2^4}+\frac{\sum_{n_3}(w_i-\mu_3)^2}{2\sigma_3^4}$$
$$= 0$$

$$\therefore \sigma^2 = (\sum_{n_1}(x_i-\mu_1)^2+\sum_{n_2}(y_i-\mu_2)^2+\sum_{n_3}(w_i-\mu_3)^2)/(n_1+n_2+n_3)$$

Under $H_a : \sigma_1^2, \sigma_2^2, \sigma_3^2$ at least one inequality

$\therefore \frac{\partial\ell(\mu_1,\mu_2,\mu_3,\sigma_1^2,\sigma_2^2,\sigma_3^2)}{\partial\sigma_1^2} = -\frac{n_1}{2\sigma_1^2}+\frac{\sum_{n_1}(x_i-\mu_1)^2}{2\sigma_1^4} = 0$

$\therefore \sigma_1^2 = \frac{\sum_{n_1}(x_i-\mu_1)^2}{n_1}$ and σ_2^2 and σ_3^2 are defined similar as $\sigma_2^2 = \frac{\sum_{n_2}(y_i-\mu_2)^2}{n_2}$ and $\sigma_3^2 = \frac{\sum_{n_3}(y_i-\mu_3)^2}{n_3}$

$\therefore R = \frac{\max_{H_0} L(\sigma^2)}{\max_{H_a} L(\sigma_1^2,\sigma_2^2,\sigma_3^2)} = \frac{1/(\sigma^{(n_1+n_2+n_3)})}{1/(\sigma_1^{(n_1)}\sigma_2^{(n_2)}\sigma_3^{(n_3)})} = \frac{\sigma_1^{(n_1)}\sigma_2^{(n_2)}\sigma_3^{(n_3)}}{\sigma^{(n_1+n_2+n_3)}}$

If $R < k_\alpha$ where the k_α is chosen to ensure the level of α, we would reject H_0.

b.

$\because -2\ln(R) \xrightarrow{d} \chi^2(d-d_0 = 3-1 = 2)$, $\chi^2_{0.05}(2) = 5.99$

$\therefore -2\ln(R) > 5.99$