# STA260: Probability and Statistics II

Kris Aujla - kris.aujla@mail.utoronto.ca

Summer 2025

This is the sequel to the course notes of STA256. Extensive course notes covering all material. Includes chapters 1-6, practice problem solutions from each chapter at the end of each section (from the list), slides organized by chapter. Note that tutorial problem questions have answers keys but I will still be doing them (chapter 7). This is constantly being updated and tweaked, for any mistakes contact me.

Entirety of chapters 1-6 completed in second-sub session

Textbook: Mathematical Statistics with Applications by Wackerly, Mendenhall and Scheaffer, 7th edition. ISBN-13: 978-0-495-11081-1

*Note: This course uses a different textbook from STA256. This course covers content from chapters 6-11 from the textbook but for organization reasons I will order them from 2-5. We briefly covered the Central Limit Theorem and Normal Distribution approximations so chapter 7 (chapter 2 here) is mostly review but goes more in depth. Chapter 1 also goes over a single topic that was skipped in STA256. Thus there might be repeat definitions and theorems in these course notes. Also note that all material from the courses notes of STA256 will be treated as a mandatory prerequisite as I sometimes refer to Theorems from the STA256 notes. Also note that all mentions of "Probability Density Function" refers to the CDF. Note that all required Tables are found in the appendix of the textbook; I do not go over how to use these tables and you are expected to figure it out yourself. Note that this course and these notes put an emphasis on practice questions and examples which are highly important to perform well. As you will see there is not much content and so I encourage you to attempt all practice problems. However I did not complete all problems (since Masoud assigned hundreds) but I did a lot of them. Note that a lot of test questions were from the textbook and tutorials.*

# Contents

# 1 Order Statistics

## 1.1 Distribution of Minimum and Maximum

We begin by going over a topic that was skipped in STA256: Order Statistics. I highly recommend going over these notes for this topic for the heuristic explanations.

Given random variables, it is often useful to compare the magnitudes of these observed random variables. For example, imagine you are designing a backup power supply system for a data center. The system uses $n = 5$ independent backup batteries. Each battery has a random lifetime $X_i$ (say, in $i$ years), which follows some continuous distribution (e.g., exponential, normal). If the whole system fails when the first battery fails, then the lifetime of the system is the minimum of all batteries:

$$X_{(1)} = \min(X_1, X_2, X_3, X_4, X_5).$$

Here we say the system lifetime depends directly on the first order statistic. On the other hand, if you are interested in replacing all batteries at the same time (to avoid different ages), you might care about:

$$X_{(n)} = \max(X_1, X_2, X_3, X_4, X_5).$$

This ensures you don't have a very old battery left in use when the others have already been replaced. If you want to analyze, for example, the typical behavior, you might look at the median lifetime, which is approximately $X_{(3)}$ when $n = 5$. This helps in maintenance planning: you can say "Half of the batteries will last longer than this time." Thus we can see that order statistics arise commonly in statistical inferences. We formalize this in the following definition:

**Definition 1.1.1**

Let $X_1, X_2, \ldots, X_n$ denote independent continuous random variables with CDF $F(x)$ and pdf $f(x)$ (that is they form a random sample). We denote the ordered random variables $X_i$ by $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ where $X_{(1)} \leq X_{(2)}, \leq \ldots \leq X_{(n)}$. From this we get that

$$X_{(1)} = \min(X_1, X_2, \ldots, X_n) \quad \text{(First Order Statistic)}$$

and

$$X_{(n)} = \max(X_1, X_2, \ldots, X_n). \quad \text{(Last Order Statistic)}$$

We can find the pdf for $X_{(1)}$ and $X_{(n)}$ by using the CDF. We begin with the minimum. We see that $F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x)$. However if the minimum $X_{(1)} > x$ then it must be true that $X_i > x$ for all $i$. That is we get that

$$F_{X_{(1)}}(x) = 1 - P(X_1 > x, X_2 > x, \ldots, X_n > x).$$

Since these random variables are identically distributed and independent (iid),

$$P(X_1 > x, X_2 > x, \ldots, X_n > x) = [1 - F(x)]^n.$$

Thus

$$F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n.$$

We then differentiate to get the pdf

$$f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x)$$

Next for the maximum we see that $F_{X_{(n)}}(x) = P(X_{(n)} \leq x)$. However if the maximum random variable is less than $n$ then it must be true that $X_i \leq x$ for all $i$. Thus we get that

$$F_{X_{(n)}}(x) = P(X_i < x \text{ for all } i) = [F(x)]^n.$$

Thus differentiating this we get that

$$f_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x)$$

We now consider finding the pdf for the $k$-th order statistic. We know that $F_{X_{(k)}}(x) = P(X_{(k)} \leq x) = P(\text{atleast } k \text{ of } X_i \text{ are } \leq x)$. But

$$P(\text{exactly } k \text{ of } X_i \leq x) = \binom{n}{k}[F(x)]^k[1 - F(x)]^{n-k}.$$

since there are $\binom{n}{k}$ ways we can observe the magnitudes of $X_i \leq x$. Notice that this is binomial probability. That is $Y \sim Bin(n, p = F(x))$. We then see that

$$F_{X_{(k)}}(x) = \sum_{j=k}^{n} P(\text{exactly } k \text{ of } X_i \leq x) = \sum_{j=k}^{n} \binom{n}{k}[F(x)]^k[1 - F(x)]^{n-k}.$$

Differentiating this we get a complex expression. However in the next section we can show in a "heuristic approach" that the pdf simplifies to

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1 - F(x)]^{n-k} f(x).$$

**Example 1.1.2**

Electronic components of a certain type have a length of life $Y$, with probability density given by

$$f(y) = \begin{cases} \frac{1}{100} e^{-y/100}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

(Length of life is measured in hours.) Suppose that two such components operate independently and in series in a certain system (hence, the system fails when either component fails). Find the density function for $X$, the length of life of the system.

*Solution* Since the system fails when either component fails we see that $X = \min(Y_1, Y_2)$ where $Y_1, Y_2$ are independent random variables with common distribution $Y$. Thus we see that the distribution for $X$ is

$$f_X(y) = 2[1 - F_Y(y)] f_Y(y).$$

We find the CDF of $Y$ by integrating the pdf:

$$F_Y(y) = \frac{1}{100} \int_0^y e^{-t/100} dt = 1 - e^{-y/100}.$$

Thus we see that

$$f_X(y) = \frac{2e^{-y/100}}{100}[1 - (1 - e^{-y/100})] = \frac{e^{-y/50}}{50}$$

for $y > 0$ and zero elsewhere. Thus, the minimum of two exponentially distributed random variables has an exponential distribution. Notice that the mean length of life for each component is 100 hours, whereas the mean length of life for the system is $E(X) = E(Y_{(1)}) = 50$. □

**Example 1.1.3**

Suppose that the components in Example 1.1.2 operate in parallel (hence, the system does not fail until both components fail). Find the density function for $X$, the length of life of the system.

*Solution* In this example $X = \max(Y_1, Y_2)$ where $Y_1, Y_2$ are independent random variables with common distribution $Y$. Thus we see that the distribution for $X$ is

$$f_X(y) = 2[F_Y(y)]f_Y(y).$$

Using the CDF we found from Example 1.1.2 we see that

$$f_X(y) = \frac{2e^{-y/100}}{100}(1 - e^{-y/100})$$

for $y > 0$ and zero elsewhere. Here we see that We see here that the maximum of two exponential random variables is not an exponential random variable. □

## 1.2 Distribution of the $k$-th Order Statistic

As we left off in the previous section, the rigorous derivation of the density function of the $k$-th order statistic is advanced. However the expressions structure for the pdf is quite intuitive so once we understand the structure the expression follows naturally. Given a continuous random variable, we can say that the probability that the random variable is at a particular point is proportional to the probability that the variable is "close" to that point. That is

$$P(x \leq X \leq x + dx) \approx f(x)dx.$$

Now consider the $k$-th order statistic. For $X_{(k)}$ to be near $x$. That is $X_{(k)} \in [x, x + dx]$ then it must be true that exactly $k - 1$ of the observations must be $\leq x$, and $n - k$ observations must be $> x$. The respective probabilities for each one is $p_1 = f(x)dx, p_2 = [F(X)]^{k-1}, p_3 = (1 - F(x))^{n-k}$. So we get that

$$P(x \leq X \leq x + dx) \approx P[(k - 1) \text{ in group 1, 1 in group 2 and } n - k \text{ in group 3}]$$

where group 1,2,3 refer to the respective cases we mentioned above. Using multinational distribution/coefficients we say in STA256 (Definition 1.3.17 in STA256 Notes) we see that the number of ways we can have this are

$$\begin{pmatrix} & n & \\ k-1 & 1 & n-k \end{pmatrix} = \frac{n!}{(k-1)!1!(n-k)!}.$$

Thus we see that

$$P(x \leq X \leq x+dx) \approx \begin{pmatrix} & n & \\ k-1 & 1 & n-k \end{pmatrix} p_1^{k-1}p_2^1p_3^{n-k} = \frac{n!}{(k-1)!(n-k)!}[F(X)]^{k-1}f(x)dx[1-F(x)]^{n-k}.$$

So we get

$$f_{X_{(k)}}(x)dx \approx \frac{n!}{(k-1)!(n-k)!}[F(X)]^{k-1}f(x)dx[1-F(x)]^{n-k}.$$

We summarize this in the following theorem:

**Theorem 1.2.1**

Let $Y_1, \ldots, Y_n$ be independent identically distributed continuous random variables with common distribution function $F(y)$ and common density function $f(y)$. If $Y_{(k)}$ denotes the $k$th-order statistic, then the density function of $Y_{(k)}$ is given by

$$g_{(k)}(y_k) = \frac{n!}{(k-1)!\,(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}f(y_k), \quad -\infty < y_k < \infty.$$

If $j$ and $k$ are two integers such that $1 \leq j < k \leq n$, the joint density of $Y_{(j)}$ and $Y_{(k)}$ is given by

$$\begin{aligned} g_{(j)(k)}(y_j, y_k) = &\frac{n!}{(j-1)!\,(k-1-j)!\,(n-k)!} \cdot \left[F(y_k) - F(y_j)\right]^{j-1} \\ &\cdot \left[1 - F(y_k)\right]^{n-k} \cdot f(y_j)\, f(y_k), \end{aligned}$$

for $-\infty < y_j < y_k < \infty$.

*Proof.* We have proven the first single order $k$-th order statistic derivation and we will prove the join density $k$-th order statistic with the same heuristic approach. We look at $P(X_{(j)} \in [y_j, y_j + dy_j), X_{(k)} \in [y_k, y_k + dy_k))$. Notice that

- Exactly $j-1$ observations $\leq y_j$.
- Exactly $k-j-1$ observations in $(y_j, y_k)$.
- Exactly $n-k$ observations $> y_k$.
- Exactly one observation in $[y_j, y_j + dy_j)$ (this becomes $X_{(j)}$).
- Exactly one observation in $[y_k, y_k + dy_k)$ (this becomes $X_{(k)}$).

We then see that the probabilities for each part is

- Probability an observation $\leq y_j$: $F(y_j)$.
- Probability in $(y_j, y_k)$: $F(y_k) - F(y_j)$.
- Probability $> y_k$: $1 - F(y_k)$.
- Probability in $[y_j, y_j + dy_j)$: $f(y_j)\, dy_j$.
- Probability in $[y_k, y_k + dy_k)$: $f(y_k)\, dy_k$.

Using the multinomial distribution we used earlier we see that there are

$$\frac{n!}{(j-1)!(k-1-j)!(n-k)! \times 1! \times 1!}$$

of ways for this to happen. Putting it all together we get that

$$g_{(j)(k)}(y_j, y_k) = \frac{n!}{(j-1)!(k-1-j)!(n-k)!}\big[F(y_j)\big]^{j-1}\big[F(y_k)-F(y_j)\big]^{k-1-j}$$
$$\times\big[1-F(y_k)\big]^{n-k} f(y_j)f(y_k), \quad -\infty < y_j < y_k < \infty.$$

□

**Example 1.2.2**

Suppose that $Y_1, Y_2, \ldots, Y_5$ denotes a random sample from a uniform distribution defined on the interval $(0,1)$. That is,

$$f(y) = \begin{cases} 1, & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the density function for the second-order statistic. Also, give the joint density function for the second- and fourth-order statistics.

*Solution* We begin by finding the CDF of the random variables:

$$F(y) = \begin{cases} 0 & y < 0 \\ y & 0 \le y \le 1 \\ 1 & y > 0 \end{cases}$$

The density function of the second-order statistic can be found by using the formula from Theorem 1.2.1:

$$f_{X_{(2)}}(y) = \frac{5!}{(2-1)!(5-2)!}[F(y)]^{2-1}(1-F(y))^{5-2}f(y) = 20y(1-y)^3$$

for $0 \le y \le 1$ and zero elsewhere. Notice that this is the Beta distribution with parameters $Beta(\alpha = 2, \beta = 4)$. The joint density function for the second-order and fourth-order statistics can be obainted by the second result from Theorem 1.2.1. We see that

$$f_{X_{(2)},X_{(4)}}(y_2, y_4) = \frac{5!}{(2-1)!(4-1-2)!(5-4)!}[F(y_2)]^{2-1}[F(y_4)-F(y_2)]^{4-1-2}[1-F(y_4)]^{5-4}f(y_2)f(y_4)$$

Simplifiying this we get that

$$f_{X_{(2)},X_{(4)}}(y_2, y_4) = 5!y_2(y_4 - y_2)(1-y_4)$$

for $0 \le y_2 \le y_4 \le 1$ and zero elsewhere. □

## 1.3 Joint Distribution of All Order Statistics

We now seek to find the pdf of

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n).$$

We will begin by considering the $n = 3$ case, for which then the general formula follows intuitively. That is we seek to find the pdf of $f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3)$. We begin by fixing the values $x_1 < x_2 < x_3$ as otherwise the probability will be zero since for example the smallest value and the second smallest value can be the same value if the distributions are continuous. We will heuristically assume that

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3) \approx P(X_{(1)} = x_1, X_{(2)} = x_2, X_{(3)} = x_3).$$

This means that we are interpreting $f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3)$ as the probability that the smallest value among $X_1, X_2, X_3$ is $x_1$, the second smallest is $x_2$ and the largest is $x_3$. This can happen by

$$\begin{aligned}
&X_1 = x_1, X_2 = x_2, X_3 = x_3,\\
&\text{or}\\
&X_1 = x_2, X_2 = x_1, X_3 = x_3,\\
&\text{or}\\
&X_2 = x_2, X_2 = x_3, X_3 = x_1,\\
&\text{or} \ldots
\end{aligned}$$

By considering all all permutations of which original $X_i$ corresponds to each order statistic, for $n = 3$ we get that $P_3^3 = 3!$ total ways. Thus we multiply that by the probability for each is $f(x_1)f(x_2)f(x_3)$. Together we get that

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3) = 3!f(x_1)f(x_2)f(x_3)$$

for $0 < x_1 < x_2 < x_3$. For general $n$ we see that

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1,x_2,\ldots,x_n) = n!f(x_1)f(x_2)\cdots f(x_n.)$$

## 1.4 Practice Problems

**6.72**

Let $Y_1$ and $Y_2$ be independent and uniformly distributed over the interval $(0, 1)$. Find

(a) the probability density function of $U_1 = \min(Y_1, Y_2)$.

(b) $E(U_1)$ and $V(U_1)$.

*Solution* For (a): using Theorem 1.2.1 we see that

$$f_{X_{(1)}}(x_1) = n[1 - F(x_1)]^{n-1} f(x_1).$$

Since the distribution is uniform over (0,1) we see that

$$f(x) = \frac{1}{1-0} = 1.$$

And so it follows that $F(x_1) = x_1$. Thus we see that when $n = 2$ we have that

$$f_{X_{(1)}}(x_1) = 2(1 - x_1)$$

for $x_1 \in (0, 1)$ and zero elsewhere. For(B): We first find the mean

$$E(U_1) = \int_0^1 x_1 f_{X_{(1)}}(x_1)dx_1 = \int_0^1 2x_1(1 - x_1)dx_1 = \frac{1}{3}.$$

Next to find the variance we find $E(U_1^2)$.

$$E(U_1^2) = \int_0^1 2x_1^2(1-x_1)dx_1 = \frac{1}{6}.$$

Putting it together we get that

$$\text{Var}(U_1) = E(U_1^2) - \mu^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}.$$

□

**6.73**

As in Exercise 6.72, let $Y_1$ and $Y_2$ be independent and uniformly distributed over the interval $(0,1)$. Find

(a) the probability density function of $U_2 = \max(Y_1, Y_2)$.

(b) $E(U_2)$ and $V(U_2)$.

*Solution* For (a): Using Theorem 1.2.1 we see that the pdf of $U_2$ is

$$f_{X_{(n)}}(x_n) = n[F(x_n)]^{n-1}f(x_n).$$

It then follows from Excercise 6.72 that

$$f_{X_{(n)}}(x_n) = 2x_1$$

for $x_n \in (0,1)$ and zero elsewhere. For (B): We again first find the mean.

$$E(U_2) = \int_0^1 2x_n^2 dx_n = \frac{2}{3}.$$

We next see that

$$E(U_2^2) = \int_0^1 2x_n^3 dx_n = \frac{1}{2}.$$

Putting it together we see that

$$\text{Var}(U_2) = E(U_2^2) - \mu^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}.$$

□

**6.74**

Let $Y_1, Y_2, \ldots, Y_n$ be independent, uniformly distributed random variables on the interval $[0,\theta]$. Find the

(a) probability distribution function of $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$.

(b) density function of $Y_{(n)}$.

(c) mean and variance of $Y_{(n)}$.

*Solution* For (a/b): We see that from section Theorem 1.2.1 that

$$f_{Y_{(n)}}(y) = n[F(y)]^{n-1} f(y).$$

We see that $f(y) = 1/\theta$ and $F(y) = y/\theta$ for $0 \le y \le \theta$. Thus

$$f_{Y_{(n)}}(y) = n \left(\frac{y}{\theta}\right)^{n-1} \frac{1}{\theta} = \frac{ny^{n-1}}{\theta^n}$$

for $0 \le y \le \theta$ and zero elsewhere. Next we find the mean

$$E(Y_{(n)}) = \int_0^\theta y \cdot \frac{ny^{n-1}}{\theta^n} dy = \frac{n\theta}{n+1}.$$

We next find

$$E(Y_{(n)}^2) = \int_0^\theta y^2 \cdot \frac{ny^{n-1}}{\theta^n} dy = \frac{n\theta^2}{n+2}.$$

Together we get that

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

□

**6.75**

Refer to Exercise 6.74. Suppose that the number of minutes that you need to wait for a bus is uniformly distributed on the interval $[0, 15]$. If you take the bus five times, what is the probability that your longest wait is less than 10 minutes?

*Solution* Let $Y_1, Y_2, Y_3, Y_4, Y_5$ be the wait times for each bus. We are told that $Y_i \sim U(0, 15)$ for all $i = 0, 1, 2, 3, 4, 5$. We are asked what is the probability that $P(Y_{(n)} < 10) =$ where $Y_{(n)} = \max(Y_1, Y_2, Y_3, Y_4, Y_5)$. Using Theorem 1.2.1 we see that

$$P(Y_{(n)} < 10) = [F(10)]^5 = \left(\frac{10}{15}\right)^5 \approx 0.132.$$

□

**6.76**

Let $Y_1, Y_2, \ldots, Y_n$ be independent, uniformly distributed random variables on the interval $[0, \theta]$.

(a) Find the density function of $Y_{(k)}$, the $k$th-order statistic, where $k$ is an integer between 1 and $n$.

(b) Use the result from part (a) to find $E(Y_{(k)})$.

(c) Find $V(Y_{(k)})$.

(d) Use the result from part (c) to find $E\big(Y_{(k)} - Y_{(k-1)}\big)$, the mean difference between two successive order statistics. Interpret this result.

*Solution* For (a): Using Theorem 1.2.1 we see that the density function of the $k$-th order statistic is going to be of the form

$$f_{Y_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!}[F(y)]^{k-1} f(y)[1-F(x)]^{n-k}.$$

We see that the common pdf is $f(y) = 1/\theta$ and the cdf is $F(y) = y/\theta$. Together we get that

$$f_{Y_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!}\left(\frac{y}{\theta}\right)^{k-1}\frac{1}{\theta}\left(1-\frac{y}{\theta}\right)^{n-k}.$$

Using this we now find the mean.

$$\begin{aligned} E(Y_{(k)}) &= \int_0^\theta y \cdot \frac{n!}{(k-1)!(n-k)!}\left(\frac{y}{\theta}\right)^{k-1}\frac{1}{\theta}\left(1-\frac{y}{\theta}\right)^{n-k} dy \\ &= \frac{n!}{(k-1)!(n-k)!\theta^k}\int_0^\theta y^k\left(1-\frac{y}{\theta}\right)^{n-k} dy \end{aligned}$$

From here we do a change of variables with $y = y/\theta$ and see that this becomes the Beta Function with parameters $Beta(k+1, n-k+1)$. We then see that

$$E(Y_{(k)}) = \frac{n!}{(k-1)!(n-k)!\theta^k}\cdot\theta^{k+1}\cdot\frac{k!(n-k)!}{(n+1)!} = \frac{\theta k}{n+1}.$$

Next for the variance we find $E(Y_{(k)}^2)$. We follow a similar process and see that the integral is a Beta Integral with parameters $Beta(k+2, n-k-1)$. We then get that

$$E(Y_{(k)}^2) = \frac{n!}{(k-1)!(n-k)!}\theta^2\cdot\frac{(k+1)!(n-k)!}{(n+2)!}.$$

Simplifying

$$\frac{n!(k+1)!}{(k-1)!(n+2)!}\theta^2 = \frac{k(k+1)}{(n+1)(n+2)}\theta^2.$$

Then,

$$V(Y_{(k)}) = E(Y_{(k)}^2) - (E(Y_{(k)}))^2 = \frac{k(k+1)}{(n+1)(n+2)}\theta^2 - \left(\frac{k}{n+1}\theta\right)^2.$$

Simplified:

$$V(Y_{(k)}) = \frac{k(n+1-k)\theta^2}{(n+1)^2(n+2)}.$$

For part (d) we see that

$$E(Y_{(k)} - Y_{(k-1)}) = E(Y_{(k)}) - E(Y_{(k-1)}) = \frac{k}{n+1}\theta - \frac{k-1}{n+1}\theta = \frac{1}{n+1}\theta.$$

The expected gap between successive order statistics is uniform and equal to $\theta/(n+1)$. □

**6.78**

Refer to Exercise 6.76. If $Y_1, Y_2, \ldots, Y_n$ are independent, uniformly distributed random variables on the interval $[0, 1]$, show that $Y_{(k)}$, the $k$th-order statistic, has a beta density function with $\alpha = k$ and $\beta = n-k+1$.

*Solution* We know that

$$f_{Y_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!}[F(y)]^{k-1} f(y)[1-F(x)]^{n-k}.$$

Since they are uniformly distributed then $f(y) = 1$ and $F(y) = y$. Thus

$$f_{Y_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!} y^{k-1}(1-y)^{n-k}.$$

for $0 \leq y \leq 1$ and zero elsewhere. However notice that this is equal to

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}$$

where $X$ is a continuous random variable and $X \sim Beta(\alpha, \beta)$ where $\alpha = k$ and $\beta = n-k-1$. Thus we can conclude that

$$Y_{(k)} \sim Beta(k-1, n-k-1).$$

□

**6.80**

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables, each with a beta distribution, with $\alpha = \beta = 2$. Find

(a) the probability distribution function of $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$.

(b) the density function of $Y_{(n)}$.

(c) $E(Y_{(n)})$ when $n = 2$.

*Solution* For (a): By now we know that

$$F_{Y(n)]}(x) = [F(x)]^n$$

where $f$ has a beta distribution with parameters $Beta(2, 2)$. We first find the cdf:

$$F(x) = \int_0^y \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} t^{2-1}(1-t)^{2-1} dt = \int_0^y 6t(1-t)dt = 3y^2 - 2y^3.$$

Thus we see that then

$$F_{Y(n)]}(x) = (3y^2 - 2y^3)^n.$$

For (b): Next to find the pdf we simplify differentiate the cdf:

$$f_{Y(n)]}(x) = 6n(3y^2 - 2y^3)^{n-1}(y - y^2).$$

We next find the mean when $n = 2$. We see that

$$E(Y_{(n)}) = \int_0^\infty y \cdot 12(3y^2 - 2y^3)^{2-1}(y - y^2)dy = \frac{22}{35}.$$

□

**6.81**

Let $Y_1, Y_2, \ldots, Y_n$ be independent, exponentially distributed random variables with mean $\beta$.

(a) Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ has an exponential distribution, with mean $\beta/n$.

(b) If $n = 5$ and $\beta = 2$, find $P(Y_{(1)} \leq 3.6)$.

*Solution* For (a): We need to show that $Y_{(1)}$ has an exponential distribution with mean $\beta/n$. We see that

$$f_{Y_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x).$$

Since we are working with exponential distribution we have that then $f(x) = e^{-y/\beta}/\beta$. We then see that $F(x) = 1 - e^{-\beta y}$. Thus we see that

$$f_{Y_{(1)}}(x) = n[1 - (1 - e^{-\beta y})]^{n-1} \cdot \frac{e^{-y/\beta}}{\beta} = \frac{n}{\beta} \cdot e^{-ny/\beta}.$$

Clearly we see that $Y_{(1)} \sim Exp(\beta/n)$. For (b): We now need to compute $P(Y_{(1)} \leq 3.6)$ when $n = 5$ and $\beta = 2$. We see that

$$P(Y_{(1)} \leq 3.6) = 1 - e^{-2.5 \cdot 3.6} \approx 0.9998.$$

□

**6.85**

Let $Y_1$ and $Y_2$ be independent and uniformly distributed over the interval $(0, 1)$. Find $P(2Y_{(1)} < Y_{(2)})$.

*Solution* We begin by finding the joint pdf of All order statistics. That is we find

$$f_{Y_{(1)}, Y_{(n)}}(u, v) = n! f(x) = 2 \cdot 1 = 2$$

where $0 < u < v < 1$. Thus we simply find

$$P(2Y_{(1)} < Y_{(2)}) = \iint_{2u<v} f_{Y_{(1)}, Y_{(n)}}(u, v) du dv.$$

The bounds then imply for a fixed $v$ that $u < v/2$ and $0 < u < v$. So we integrate from $0 < u < v/2$ where $0 < v < 1$. Thus

$$P(2Y_{(1)} < Y_{(2)}) = \int_0^1 \int_0^{v/2} 2 du dv = \frac{1}{2}.$$

□

**6.86**

Let $Y_1, Y_2, \ldots, Y_n$ be independent, exponentially distributed random variables with mean $\beta$. Give the

(a) density function for $Y_{(k)}$, the $k$th-order statistic, where $k$ is an integer between 1 and $n$.

(b) joint density function for $Y_{(j)}$ and $Y_{(k)}$ where $j$ and $k$ are integers $1 \leq j < k \leq n$.

*Solution* For (a): Using Theorem 1.2.1 we know that

$$f_{Y_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!}[F(y)]^{k-1}f(y)[1-F(x)]^{n-k}.$$

Since $Y_i$ are exponentially distributed with mean $\beta$ then we know that $f(y) = \frac{1}{\beta}e^{-y/\beta}$ and $F(y) = 1 - e^{-y/\beta}$. Thus we see that

$$f_{Y_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!}(1-e^{-y/\beta})^{k-1}\left(\frac{1}{\beta}e^{-y/\beta}\right)(e^{-y/\beta})^{n-k}.$$

For (b): Again using theorem 1.2.1 we see that the pdf is of the form

$$f_{Y_{(j)},Y_{(k)}}(y_1,y_2) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!}\left[F(y_1)\right]^{j-1}\left[F(y_2)-F(y_1)\right]^{k-j-1}\left[1-F(y_2)\right]^{n-k}f(y_1)f(y_2).$$

Substituting everything in we have that

$$f_{Y_{(j)},Y_{(k)}}(y_1,y_2) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!}\left(1-e^{-y_1/\beta}\right)^{j-1}\left(e^{-y_1/\beta}-e^{-y_2/\beta}\right)^{k-j-1}\left(e^{-y_2/\beta}\right)^{n-k}\frac{1}{\beta}e^{-y_1/\beta}\frac{1}{\beta}e^{-y_2/\beta}.$$

Simplifying:

$$f_{Y_{(j)},Y_{(k)}}(y_1,y_2) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!}\left(1-e^{-y_1/\beta}\right)^{j-1}\left(e^{-y_1/\beta}-e^{-y_2/\beta}\right)^{k-j-1}\left(e^{-y_2/\beta}\right)^{n-k}\frac{e^{-(y_1+y_2)/\beta}}{\beta^2}.$$

□

**6.88**

Suppose that the length of time $Y$ it takes a worker to complete a certain task has the probability density function given by

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

where $\theta$ is a positive constant that represents the minimum time until task completion. Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of completion times from this distribution. Find

(a) the density function for $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$.

(b) $E(Y_{(1)})$.

*Solution* For (a): We know that

$$f_{Y_{(1)}}(y) = n[1-F(y)]^{n-1}f(y).$$

We find the CDF:

$$F(y) = \int_0^y e^{-(t-\theta)}dt = 1 - e^{-(y-\theta)}.$$

Thus

$$f_{Y_{(1)}}(y) = n(e^{-(y-\theta)})^{n-1}(e^{-(y-\theta)}) = ne^{-n(y-\theta)}$$

for $y > 0$. Next we find the mean:

$$E(Y_{(1)}) = \int_0^\infty y \cdot ne^{-n(y-\theta)}dy = \frac{1}{n} + \theta.$$

□

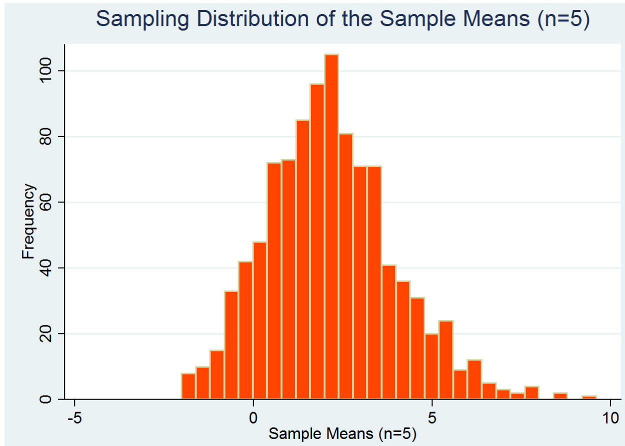


Figure 1: A histogram of the relative frequency of the possible values of a statistic (sample mean in this case) with repeated sampling (5 times in this case)

# 2 Sampling Distributions and the Central Limit Theorem

## 2.1 Introduction

Throughout this chapter we will be looking at a random samples $Y_1, Y_2, \ldots, Y_n$ from a population of interest. Recall that a random sample are random variables that are independent and identically distributed (iid). We learned in STA256 that certain random samples can be used to make decisons or estimate certain population parameters. For example lets say we want to estimate the population mean $\mu$. If we obtain a random sample with $n$ observations $Y_1, Y_2, \ldots, Y_n$ we found that we could estimate $\mu$ with the sample mean:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Moreover, in a typical statistical problem we know the underlying probability distribution but are missing the parameters. For example $X \sim Exp(\mu)$ or $X \sim N(\mu, \sigma^2)$ or $X \sim Possion(\lambda)$. We typically represent these parameters as $\theta$. For example $X \sim Exp(\theta)$ or $X \sim N(\theta_1, \theta_2)$ or $X \sim Possion(\theta)$. So if we obtain data or observations we can estimate these parameters. These estimations such as $\overline{Y}$ are what we call statistics since they are a function of observations and constants.

**Definition 2.1.1**

A statistic is a function of the observable random variables in a sample and known constants.

We have already encountered many different statistics such as sample mean $\overline{Y}$, sample variance $S^2$, $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$, $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$, the sample median and so on. Notice that since all statistics are based on random samples, which are random variables, then a statistic is also a random variable. So then a statistic also has a probability distribution. We call these sampling distributions. When you repeatedly take samples from a population and compute a statistic each time (for example, the sample mean), the distribution of those values you get for the statistic is the sampling distribution. This tells us how that statistic would vary from sample to sample, which is crucial for making inferences about the population. See Figure 1 for example. We also illustrate this in the following example.

**Example 2.1.2**

A balanced die is tossed three times. Let $Y_1, Y_2$, and $Y_3$ denote the number of spots observed on the upper face for tosses 1, 2, and 3, respectively. Suppose we are interested in

$$\overline{Y} = \frac{Y_1 + Y_2 + Y_3}{3},$$

the average number of spots observed in a sample of size 3. What are the mean, $\mu_{\overline{Y}}$, and standard deviation, $\sigma_{\overline{Y}}$, of $\overline{Y}$? How can we find the sampling distribution of $\overline{Y}$?

*Solution* We see that

$$\mu = E(\overline{Y}) = E(\frac{1}{3} \cdot (Y_1 + Y_2 + Y_3)) = \frac{E(Y_1) + E(Y_2) + E(Y_3)}{3}.$$

We know that the expected value for the number of spots observed on the upper face for tosses 1,2, and 3 are

$$E(Y_i) = \sum_{y=1}^{6} \frac{y}{6} = 3.5.$$

Thus we see that

$$\mu = E(\overline{Y}) = \frac{3 \cdot 3.5}{3} = 3.5.$$

Moreover we recall that the variance of the sample mean was $\text{Var}(\overline{Y}) = \sigma^2/n$ where $\sigma^2$ was the common variance. We see that in this case $\sigma^2 = 2.9167$ for all $Y_i$. Thus we see that $\text{Var}(\overline{Y}) = 2.9167/3 = 0.9722$ and so $\sigma_{\overline{Y}} = 0.9860$. To find the probability distribution of $\overline{Y}$ we see that first it would be a discrete distribution of the form $\overline{Y} = W/3$ where $W = Y_1 + Y_2 + Y_3$. Moreover the possible values for $W$ are going to be $3, 4, 5, \ldots, 18$. Also the probability that the balanced die takes on the values $Y_1, Y_2, Y_3$ for the three rolls will be $(1/6)^3 = 1/216$. Thus we see that

$$p_W(w) = 1/216$$

for all $w = 3, 4, 5, \ldots, 18$. Then the distribtion of $\overline{Y}$ becomes

$$P(\overline{Y} = x) = P\left(\frac{W}{3} = x\right) = P(W = 3x)$$

for all $x = 3, 4, 5, \ldots, 18$. So

$$P(\overline{Y} = 1) = P(W = 3) = p_W(1, 1, 1) = \frac{1}{216}$$

$$P(\overline{Y} = 4/3) = P(W = 4) = p_W(1, 1, 2) + p_W(1, 2, 1) + p_W(2, 1, 1) = \frac{3}{216}$$

$$\begin{aligned} P(\overline{Y} = 5/3) = P(W = 5) &= p_W(1, 1, 3) + p_W(1, 3, 1) + p_W(3, 1, 1) \\ &+ p_W(1, 2, 2) + p_W(2, 1, 2) + p_W(2, 2, 1) = \frac{6}{216} \end{aligned}$$

$$\vdots$$

The probabilities $P(\overline{Y} = i/3)$, $i = 7, 8, \ldots, 18$ are obtained similarly. $\square$

Notice that in the last example obtaining the exact distribution for the sample mean was quite tedious. However we could have instead of simulated the sampling distribution by taking independent samples of size 3 and compute $\overline{Y}$ for each sample and construct an histogram with these values. Doing so would result in a histogram of similar shape as Figure 1. We would see that most of the values accumulate in a mound shape manner around the theoretical mean we found which was 3.5 (which if you recall follows approximately the normal distribution).

## 2.2 Sampling Distributions Related to the Normal Distribution

Recall from that from STA256 we learned that many real world phenomenon can be modeled by the normal distribution. Moreover we also learned that if we have a random sample from a normal distribution then the sampling distribution of any linear combination of these random variables also has a normal distribution. Thus the sample mean is also normal.

**Theorem 2.2.1**

Let $Y_1$, $Y_2$, $\ldots$, $Y_n$ be a random sample of size $n$ from a normal distribution with mean $\mu$ and variance $\sigma^2$. Then

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

is normally distributed with mean $\mu_{\overline{Y}} = \mu$ and variance $\sigma^2_{\overline{Y}} = \sigma^2/n$.

*Proof.* Since $Y_1, Y_2, \ldots, Y_n$ are a random sample then they are iid. Also notice that

$$\begin{aligned}\overline{Y} &= \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}Y_1 + \frac{1}{n}Y_2 + \cdots + \frac{1}{n}Y_n \\ &= a_1Y_1 + a_2Y_2 + \cdots + a_nY_n\end{aligned}$$

where $a_i = 1/n$ for $i = 1, 2, \ldots, n$. Thus this a linear combination of independent normal random variables and so by Theorem 3.4.4 (from STA256 Notes) $\overline{Y}$ has a normal distribution. We find the parameters $\mu$ and $\text{Var}(\overline{Y})$.

$$E(\overline{Y}) = E(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\cdot n\mu = \mu.$$

$$\text{Var}(\overline{Y}) = \text{Var}(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(Y_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Thus we see that

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

$\square$

Since $\overline{Y}$ is normally distributed we then can standardize it by

$$Z = \frac{\overline{Y} - \mu}{\sigma_{\overline{Y}}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right).$$

**Example 2.2.2**

A bottling machine can be regulated so that it discharges an average of $\mu$ ounces per bottle. It has been observed that the amount of fill dispensed by the machine is normally distributed with $\sigma = 1.0$ ounce. A sample of $n = 9$ filled bottles is randomly selected from the output of the machine on a given day (all bottled with the same machine setting), and the ounces of fill are measured for each. Find the probability that the sample mean will be within .3 ounce of the true mean $\mu$ for the chosen machine setting.

*Solution* Let $Y_1, Y_2, \ldots, Y_9$ be the observed amount of fill for each bottle from our random sample. We are told that the fill dispensed by the machine follows a normal distribution with parameters $N(\mu, \sigma = 1)$ for all $Y_i$ where $i = 1, 2, \ldots, 9$. Therefore by Theorem 2.2.1 we know that $\overline{Y} \sim N(\mu, 1/9)$. We are asked to find $P(\mu - 0.3 \leq \overline{Y} \leq u + 0.3) = P(|\overline{Y} - \mu| \leq 0.3) = P(-0.3 \leq \overline{Y} - \mu \leq 0.3)$. We standardize this

$$P(-0.3 \leq \overline{Y} - \mu \leq 0.3) = P\left(\frac{-0.3}{\sigma/\sqrt{n}} \leq \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right) = P\left(\frac{-0.3}{1/\sqrt{9}} \leq Z \leq \frac{0.3}{1/\sqrt{9}}\right).$$

Since the normal distribution is symmetric we have that

$$P(|\overline{Y} - \mu| \leq 0.3) = 2P(Z \leq \frac{0.3}{1/\sqrt{9}} = 0.9) \approx 0.6138.$$

Thus, the probability is only 0.6318 that the sample mean will be within 0.3 ounce of the true population mean. □

**Example 2.2.3**

Refer to Example 2.2.2. How many observations should be included in the sample if we wish $\overline{Y}$ to be within 0.3 ounce of $\mu$ with probability 0.95?

*Solution* We need to find the value of $n$ such that

$$P(|\overline{Y} - \mu| \leq 0.3) = 0.95.$$

We found that

$$P(|\overline{Y} - \mu| \leq 0.3) = 2P(Z \leq \frac{0.3}{1/\sqrt{n}} = 0.3\sqrt{n}) = 0.95.$$

We see that $P(Z \leq 1.65) = 0.95$ and so we get that

$$0.3\sqrt{n} = 1.65 \quad \text{or equivalently} \quad n = \left(\frac{1.65}{0.3}\right)^2 \approx 42.68.$$

Since we cannot take a sample of 42.68 we see that with $n = 42$ the probability is slightly less than 0.95 and with $n = 43$ slightly exceeds 0.95 which is what we will go with. □

**Example 2.2.4**

A forester studying the effects of fertilization on certain pine forests in the Southeast is interested in estimating the average basal area of pine trees. In studying basal areas of similar trees for many years, he has discovered that these measurements (in square inches) are approximately normally distributed with standard deviation approximately 4 square inches. If the forester samples n = 9 trees, find the probability that the sample mean will be within 2 square inches of the population mean.

*Solution* Let $Y_!, Y_2 \ldots, Y_n$ be the observed amount of basal area of the pine trees. We are told that this random sample follows a normal distribution with parameters $N(\mu, \sigma = 4)$. We are asked to find the the probability that $P(|\overline{Y} - \mu| \leq 2)$. Like in Example 2.2.2 we see that this results to finding

$$P(-2 \leq \overline{Y} - \mu \leq 2) = P(\frac{-2}{\sigma/\sqrt{n}} \leq \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \leq \frac{2}{\sigma\sqrt{n}}).$$

Notice we standardized the expression. Thus we see that

$$P(|\overline{Y} - \mu| \leq 2) = 2P(Z \leq \frac{2}{\sigma\sqrt{n}} = \frac{2\sqrt{9}}{4}).$$

Calculating this we see that

$$P(|\overline{Y} - \mu| \leq 2) = 0.8664$$

□

**Example 2.2.5**

Suppose that $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ are independent random samples, with the variables $X_i$ normally distributed with mean $\mu_1$ and variance $\sigma_1^2$ and the variables $Y_i$ normally distributed with mean $\mu_2$ and variance $\sigma_2^2$. The difference between the sample means, $\overline{X} - \overline{Y}$, is then a linear combination of $m + n$ normally distributed random variables and hence, is also normally distributed. This additivity property of the family of normal random variables was established in STA256.

(a) Find $E(\overline{X} - \overline{Y})$.

(b) Find $V(\overline{X} - \overline{Y})$.

(c) Suppose that $\sigma_1^2 = 2$, $\sigma_2^2 = 2.5$, and $m = n$. Find the sample sizes so that $\overline{X} - \overline{Y}$ will be within 1 unit of $(\mu_1 - \mu_2)$ with probability 0.95.

*Solution* For (a): Using the linearity of Expectation we see that

$$E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = \mu_1 - \mu_2.$$

For (b): Using Theorem 2.6.1 (from STA256 Notes) we see that

$$\text{Var}(\overline{X} - \overline{Y}) = \text{Var}(\overline{X}) + \text{Var}(\overline{Y}) - 2Cov(\overline{X}, \overline{Y}).$$

However since these are random samples they are independent and so the covariance is zero. Thus we see that

$$\text{Var}(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{m^2} + \frac{\sigma_2^2}{n^2}.$$

For (c): Let $\overline{Z} = \overline{X} - \overline{Y}$. We are asked to find $P(|\overline{Z} - (\mu_1 - \mu_2)| \leq 1) = 0.95$. Standardizing this we get that

$$P\left(\frac{-1}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{n}}} \leq \frac{\overline{Z} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{n}}} \leq \frac{1}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{n}}}\right).$$

After simplifying we see that

$$P\left(Z \leq \frac{1}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{n}}}\right) = 0.95.$$

Using the Normal distribution table we find that $P(Z \leq 1.96) = 0.95$. And so we get that

$$1.96\sqrt{\frac{\sigma_1^2+\sigma_2^2}{n}} = 1 \quad \text{or equivalently} \quad n = (1.96)^2(\sigma_1^2+\sigma_2^2).$$

Substituting the given values we find that

$$n \approx 17.29.$$

Thus we round up to exceed the probability of 0.95 and find that the answer is $n = 18$. □

We now go back to a distribution we introduced in STA256 and recall a familiar theorem. If you sum the squares of $n$ independent standard normal variables, the resulting random variable follows a chi-squared distribution with $n$ degrees of freedom. In statistics, this forms the basis of tests and estimators involving variances (e.g., estimating population variance, constructing confidence intervals for variance). The sum of squares is used in constructing sample variances and leads to $t$ and $F$ distributions. As for a visual analogy: think of each $Z_i$ as a "standardized deviation" from the mean. Squaring them measures "how far" each point is from the mean, ignoring direction. Adding them up gives a total "spread" measure, and this total spread is known to follow a chi-squared distribution.

**Theorem 2.2.6**

Let $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. If $Z_i = \frac{Y_i-\mu}{\sigma}$, then

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n}\left(\frac{Y_i-\mu}{\sigma}\right)^2$$

has a $\chi^2$ distribution with $n$ degrees of freedom (df).

*Proof.* Because $Y_i \sim N(\mu, \sigma^2)$, the linear transformation $Z_i = (Y_i-\mu)/\sigma$ yields $Z_i \sim N(0,1)$. Independence of the $Y_i$ is preserved, so the $Z_i$ are independent standard normal random variables.

Let $W_i = Z_i^2$. By the Theorem 3.4.3 (from STA256 Notes), each $W_i$ possesses a Gamma density with shape parameter $\frac{1}{2}$ and common scale parameter 2; symbolically

$$W_i \sim \Gamma\left(\tfrac{1}{2}, 2\right).$$

Since the scale parameter is the same for every $W_i$, we use the theorem on the additivity of independent Gamma variables with a common scale: the sum of $n$ such variables is itself Gamma with shape equal to the sum of the individual shapes and the same scale. Consequently,

$$S_n = \sum_{i=1}^{n} W_i \sim \Gamma\left(\tfrac{1}{2}+\cdots+\tfrac{1}{2},\ 2\right) = \Gamma\left(\tfrac{n}{2},\ 2\right).$$

Finally, by definition a chi–squared distribution with $k$ degrees of freedom is precisely a Gamma distribution with shape $k/2$ and scale 2. Taking $k = n$ identifies

$$S_n \;\sim\; \Gamma\big(\tfrac{n}{2}, 2\big) = \chi^2(n),$$

□

We now prove the additivity of the Chi-squared distributions.

**Theorem 2.2.7**

If $Y_1, \ldots, Y_n$ are independent random variables that have, respectively, chi-squared distributions with $v_1, \ldots, v_n$ degrees of freedom, then the random variable

$$U = Y_1 + \cdots + Y_n$$

has a chi-squared distribution with $v = v_1 + \cdots + v_n$ degrees of freedom.

*Proof.* Recall a chi–squared distribution with $k$ degrees of freedom is a special case of the Gamma family:

$$\chi^2(k) \;=\; \Gamma\big(\tfrac{k}{2},\, 2\big),$$

where the first argument is the shape parameter and the second is the *scale* parameter. Equivalently, its moment generating function (mgf) is

$$M_{\chi^2(k)}(t) \;=\; (1-2t)^{-k/2}, \qquad t < \tfrac{1}{2}.$$

Because $Y_1, \ldots, Y_n$ are independent, the mgf of their sum is the product of their individual mgfs:

$$M_U(t) \;=\; \prod_{i=1}^{n} M_{Y_i}(t) \;=\; \prod_{i=1}^{n} (1-2t)^{-\,v_i/2} \;=\; (1-2t)^{-\,\frac{1}{2}\sum_{i=1}^{n} v_i} \;=\; (1-2t)^{-\,v/2}, \qquad t < \tfrac{1}{2}.$$

The mgf obtained is exactly the mgf of a chi–squared random variable with $v$ degrees of freedom. Since the mgf uniquely determines the distribution (for variables whose mgf exists in an open interval around the origin), we conclude

$$U \sim \chi^2(v).$$

□

**Example 2.2.8**

If $Z_1, Z_2, \ldots, Z_6$ denotes a random sample from the standard normal distribution, find a number $b$ such that

$$P\left(\sum_{i=1}^{6} Z_i^2 \le b\right) = 0.95.$$

*Solution* Using Theorem 2.2.6 we know that the $U = \sum_{i=1}^{6} Z_i^2$ has the distribution $\chi^2_{(6)} = \Gamma(6/2 = 3, 2)$. We are asked to find the 95th percentile of $\chi^2_{(6)}$. Using a table provided in the formula sheet or other software we find that $b \approx 12.592$. □

We started off this chapter by talking about estimating population parameters through random samples. We learned that the sample mean is a good estimator for the population mean $\mu$. However if we want to estimate the population variance $\sigma^2$ then a natural good estimator would be the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

The following theorem tells us that the sampling distribution of this statistic follows a Chi-squared distribution.

**Theorem 2.2.9**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$. Then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

has a $\chi^2$ distribution with $(n-1)$ degrees of freedom (df). Also, $\overline{Y}$ and $S^2$ are independent random variables.

*Proof.* The complete proof of this theorem requires some topics that are out of scope of this course but I will show the case for $n = 2$ to make this theorem more plausible just like in the textbook. We will show that $(n-1)S^2/\sigma^2$ has a $\chi^2$ distribution with 1 df and omit the independence of sample mean and sample variance. First notice that

$$\overline{Y} = \frac{Y_1 + Y_2}{2}.$$

Then we see that

$$\begin{aligned} S^2 &= \frac{1}{2-1}\sum_{i=1}^{2}(Y_i - \overline{Y})^2 \\ &= \left[Y_1 - \frac{1}{2}(Y_1 + Y_2)\right]^2 + \left[Y_2 - \frac{1}{2}(Y_1 + Y_2)\right]^2 \\ &= \left[\frac{1}{2}(Y_1 - Y_2)\right]^2 + \left[\frac{1}{2}(Y_2 - Y_1)\right]^2 \\ &= 2\left[\frac{1}{2}(Y_1 - Y_2)\right]^2 = \frac{(Y_1 - Y_2)^2}{2}. \end{aligned}$$

It follows that, when $n = 2$,

$$\frac{(n-1)S^2}{\sigma^2} = \frac{(Y_1 - Y_2)^2}{2\sigma^2} = \left(\frac{Y_1 - Y_2}{\sqrt{2\sigma^2}}\right)^2.$$

We will show that this is the square of a standard normal variable which then by Theorem 2.2.6 implies that it follows a $\chi^2$ distribution with 1 df. We begin my rewriting the expression

$$Y_1 - Y_2 = a_1 Y_1 + a_2 Y_2$$

where $a_1 = 1$ and $a_2 = 2$. Then since this a linear combination of normal random variables it then is also normal with mean $1\mu - 1\mu = 0$ and variance $(1)^2\sigma^2 + (-1)^2\sigma^2 = 2\sigma^2$. Thus we see that

$$Z = \frac{Y_1 - Y_2}{\sqrt{2\sigma^2}}$$

is a standard normal variable and so then

$$\frac{(n-1)S^2}{\sigma^2} = Z^2$$

which by Theorem 2.2.6 implies $(n-1)S^2/\sigma^2$ has a $\chi^2$ distribution with 1 df. □

**Example 2.2.10**

Let $Y_1, Y_2$ be a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$. Based on the previous theorem, can you conclude that $Y_1 + Y_2$ and $Y_1 - Y_2$ are independent?

*Solution* We are assuming that from Theorem 2.2.9 that $\overline{Y} = \frac{Y_1+Y_2}{2}$ and $S^2 = \frac{(Y_1-Y_2)^2}{2}$ are independent. Let $U_1 = Y_1 + Y_2$ and $U_2 = Y_1 - Y_2$. Then we see that

$$U_1 = 2\overline{Y} \quad \text{and} \quad U_2 = \sqrt{2S^2}.$$

Since $U_1$ is only a function of the sample mean and $U_2$ is only a function of the sample variance we can conclude that $U_1$ and $U_2$ are independent. Of course we could have also shown that $Cov(U_1, U_2) = 0$ which is also valid. □

**Exampel 2.2.11**

In Example 2.2.2, the ounces of fill from the bottling machine are assumed to have a normal distribution with $\sigma^2 = 1$. Suppose that we plan to select a random sample of ten bottles and measure the amount of fill in each bottle. If these ten observations are used to calculate $S^2$, it might be useful to specify an interval of values that will include $S^2$ with a high probability. Find numbers $b_1$ and $b_2$ such that

$$P(b_1 \leq S^2 \leq b_2) = 0.90.$$

*Solution* First we use the fact that

$$P(b_1 \leq S^2 \leq b_2) = P\left[\frac{(n-1)b_1}{\sigma^2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \frac{(n-1)b^2}{\sigma^2}\right].$$

Since $\sigma^2 = 1$, follows a $\chi^2$ distribution with $n - 1 = 9$ df. We then use the tables provided in the formula sheet to find the values

$$P(a_1 \leq (n-1)S^2 \leq a_2) = 0.9.$$

To find these values what we can to is using the table for percentage points of Chi-squared distribution (Table 6 in Appendix of Textbook), we look for the value of $a_1$ that cuts off 0.05 in the lower tail and the value of $a_2$ that cuts off 0.95 in the upper tail. This gives us $a_1 = 3.325$ and $a_2 = 16.919$. Solving for the values of $b_1$ and $b_2$ gives us

$$3.325 = a_1 = \frac{(n-1)b_1}{\sigma^2} = 9b_1 \quad \text{or} \quad b_1 = \frac{3.325}{9} = 0.369 \quad \text{and}$$

$$16.919 = a_2 = \frac{(n-1)b_2}{\sigma^2} = 9b_2 \quad \text{or} \quad b_2 = \frac{16.919}{9} = 1.880.$$

Thus, if we wish to have an interval that will include $S^2$ with probability 0.90, one such interval is $(0.369, 1.880)$. Notice that this interval is fairly wide. □

**Example 2.2.12**

Let $Y_1, Y_2, \ldots, Y_5$ be a random sample of size 5 from a normal population with mean 0 and variance 1, and let

$$\overline{Y} = \frac{1}{5}\sum_{i=1}^{5} Y_i.$$

Let $Y_6$ be another independent observation from the same population. What is the distribution of

(a) $W = \sum_{i=1}^{5} Y_i^2$? Why?

(b) $U = \sum_{i=1}^{5}(Y_i - \overline{Y})^2$? Why?

(c) $\sum_{i=1}^{5}(Y_i - \overline{Y})^2 + Y_6^2$? Why?

*Solution* For (a): Notice that $Y_i \sim N(0,1) \sim Z$ for all $i = 1, 2, \ldots, 5$. Thus using Theorem 2.2.6 we see that $W \sim \chi^2_{(5)}$. For (b): Using Theorem 2.2.9 we see that

$$U = \frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)\sum_{i=1}^{5}(Y_i - \overline{Y})^2}{(n-1)\sigma^2} = \sum_{i=1}^{5}(Y_i - \overline{Y})^2.$$

Thus we clearly see that then $U \sim \chi^2_{(n-1)}$. For (c): We know from part b that $\sum_{i=1}^{5}(Y_i - \overline{Y})^2 \sim \chi^2_{(4)}$ and using Theorem 2.2.6 we know that $Y_6^2 \sim \chi^2_{(1)}$. Thus we have a sum of chi-squared random variables which then by Theorem 2.2.7 implies that that $\chi^2_{(4)} + \chi^2_{(1)} \sim \chi^2_{(5)}$. □

### 2.2.1 $t$ and $F$ Distributions

Theorem 2.2.1 is an important theorem that provides the basis for development of inference making procedures about the mean $\mu$ of a normal population with known variance $\sigma^2$. It it tells us that $\sqrt{n}(\overline{Y} - \mu)/\sigma$ has a standard normal distribution. We also learned that when the population standard deviation $\sigma$ is unknown we can estimate it with $S = \sqrt{S^2}$ and so

$$\sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right)$$

gives us a the basis for developing methods for inferences about $\mu$. We will show that $\sqrt{n}(\overline{Y} - \mu)/\sigma$ has a distribution known as Student's $t$ distribution with $n-1$ df. The general definition is as follows

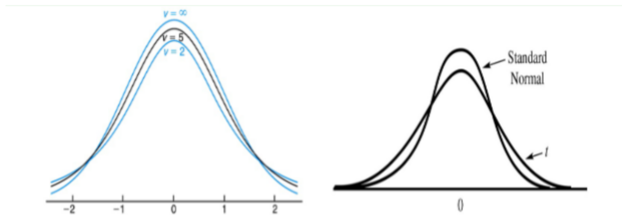


Figure 2: Graphs depicting the $t$-distribution curves with varying df and comparison to the Standard Normal Distribution.

**Definition 2.2.13**

Let $Z$ be a standard normal random variable and let $W$ be a $\chi^2-$distributed random variable with $v$ df. Then if $Z$ and $W$ are independent then

$$T = \frac{Z}{\sqrt{W/v}}$$

is said to have a $t$ distribution with $v$ df.

We will show that $\sqrt{n}(\overline{Y} - \mu)/\sigma$ has a $t$ distribution with $n-1$ df. Let $Y_1, Y_2, \dots, Y_n$ be a random sample from a normally distributed population with mean $\mu$. If the population variance $\sigma^2$ is known then using Theorem 2.2.1 we can standardize $\overline{Y}$ to get that $\sqrt{n}(\overline{Y} - \mu)/\sigma \sim N(0, 1)$. When $\sigma$ is unknown (which is almost always true in practice), you cannot use the standard normal distribution for inference about $\mu$. Instead, you use the $t-$distribution. Since we do not know $\sigma$ we instead estimate it using the sample: $S = \sqrt{S^2}$. We then replace $\sigma$ with $S$ to get

$$\sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right).$$

Here we know that the numerator $\overline{Y} - \mu \sim N(0, \sigma^2/n)$ and the denominator is random. However Theorem 2.2.9 tells us that $\overline{Y}$ and $S$ are independent. So we rewrite it as

$$\frac{\sqrt{n}\,(\overline{Y} - \mu)}{S} = \frac{\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}} = \frac{Z}{\sqrt{W/(n-1)}},$$

where $Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ and $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. Thus we see that $\sqrt{n}(\overline{Y} - \mu)/\sigma$ has $t$ distribution with $n-1$ df. The pdf formula will not be given and calculations will have to be done using the tables provided in the formula sheets. Also note that for $v > 1$ we have that $E(T) = 0$ and that for $v > 2$ we have that $\mathrm{Var}(T) = v/(v-2)$. Here are also some more properties of the $t$-distribution:

- Each $t_\nu$ curve is bell-shaped and centered at 0.
- Each $t_\nu$ curve is spread out more than the standard normal ($Z$) curve.
- As $\nu$ increases, the spread of the corresponding $t_\nu$ curve decreases.
- As $\nu \to \infty$, the sequence of $t_\nu$ curves approaches the standard normal curve (the $Z$ curve is called a $t$ curve with df $= \infty$).

**Example 2.2.14**

The tensile strength for a type of wire is normally distributed with unknown mean $\mu$ and unknown variance $\sigma^2$. Six pieces of wire were randomly selected from a large roll; $Y_i$, the tensile strength for portion $i$, is measured for $i = 1, 2, \ldots, 6$. The population mean $\mu$ and variance $\sigma^2$ can be estimated by $\overline{Y}$ and $S^2$, respectively. Because $\sigma^2_{\overline{Y}} = \sigma^2/n$, it follows that $\sigma^2_{\overline{Y}}$ can be estimated by $S^2/n$. Find the approximate probability that $\overline{Y}$ will be within $2S/\sqrt{n}$ of the true population mean $\mu$.

*Solution* We are given a random sample $Y_1, \ldots, Y_6$ from a normally distributed population with unknown mean $\mu$ and unknown variance $\sigma^2$. We are asked to find the probability that $\overline{Y}$ will be within 2 standard deviations from the true population mean $\mu$. That is we need to find

$$P(|\overline{Y} - \mu| \leq 2S/\sqrt{n}) = P\left(\frac{-2S}{\sqrt{n}} \leq \overline{Y} - \mu \leq \frac{2S}{\sqrt{n}}\right) = P\left(-2 \leq \frac{\overline{Y} - \mu}{S/\sqrt{n}} \leq 2\right).$$

However notice that this is $t-$distribution with $n - 1 = 5$ df. We see that

$$P(-2 \leq T \leq 2) = 1 - P(T < -2) - P(T > 2).$$

Since the $t-$distribution is symmetric we have that $P(T < -2) = P(T > 2)$ so

$$P(-2 \leq T \leq 2) = 1 - 2P(T > 2).$$

Using the table we see that $P(-2.015 \leq T \leq 2) = 1 - 2P(T > 2.015) = 1 - 2(0.050) \approx 0.90$. The probability that $\overline{Y}$ will be within 2 standard deviations from the mean is slightly less than 0.90. Also note that if the variance was given we could have instead found this by using the standardized sample mean instead. □

**Example 2.2.15**

Suppose that $Z$ has a standard normal distribution and that $W$ is an independent $\chi^2$-distributed random variable with $\nu$ df. Then, according to Definition 2.2.13,

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t_\nu.$$

1. If $Z$ has a standard normal distribution, find $E(Z)$ and $E(Z^2)$.
2. Recall that $E(W^\alpha) = \frac{\Gamma(\frac{\nu}{2}+\alpha)}{\Gamma(\frac{\nu}{2})}2^\alpha$ if $\nu > -2\alpha$. Show that $E(T) = 0$, $\nu > 1$ and $V(T) = \frac{\nu}{\nu-2}$, $\nu > 2$.

*Solution* For the first part: We know that if $Z$ has a standard normal distribution then the expected value of $Z$ is zero. That is $E(Z) = 0$. Moreover since we know that the variance of any standard normal distribution is 1 we can see that $1 = E(Z^2) - \mu^2 = E(Z^2) - 0 = E(Z^2)$. For the second part: We are using the identity for the $k-$th moment of a $\chi^2$ distribution for $v$ df and we are asked to find the mean and variance. First we see that the first moment will be

$$E(T) = E\left[\frac{Z}{\sqrt{W/v}}\right] = E(Z)E\left(\frac{1}{\sqrt{W/v}}\right) = 0.$$

We next see that

$$E(T^2) = E\left[\frac{Z^2}{W/v}\right] = vE(Z^2)E\left(\frac{1}{W}\right) = 1 \cdot vE(W^{-1}).$$

Using the identity for $k-$th moment of a $\chi^2$ distribution for $v$ df we see that

$$E(W^{-1}) = \frac{\Gamma(\frac{v}{2}-1)}{2\Gamma(\frac{v}{2})} = \frac{1}{2} \cdot \frac{1}{\frac{v}{2}-1} = \frac{1}{v-2}.$$

where in the last equality we used the fact that $\Gamma(a) = a\Gamma(a-1)$. Thus we have that

$$\text{Var}(T) = \frac{v}{v-2} - 0^2 = \frac{v}{v-2}$$

for $v > 2$. □

Next suppose that we want to compare the variances of two normal populations based on information contained in independent random samples from the two populations. Lets say that sample size of $n_1$ and $n_2$ are taken from the populations with variance $\sigma_1^2$ and $\sigma_2^2$ respectively. For each sample we can estimate the variance $\sigma_i^2$ by $S_i^2$ for all $i = 1, 2$. Thus it makes sense that the ratio $S_1^2/S_2^2$ can make inferences about the relative magnitudes of $\sigma_1^2$ and $\sigma_2^2$. However this ratio is random since the sample variances are random so we consider scaled version of the sample variances

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\ \sigma_2^2} = \left(\frac{\sigma_2^2}{\sigma_1^2}\right)\left(\frac{S_1^2}{S_2^2}\right).$$

Each term in numerator and denominator is a chi-square divided by its degrees of freedom (after adjusting). Explicitly,

$$\frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)} = \frac{\chi^2_{n_1-1}/(n_1-1)}{\chi^2_{n_2-1}/(n_2-1)} \sim F(n_1-1, n_2-1).$$

Thus we see that

$$\frac{S_1^2}{S_2^2} \sim \frac{\sigma_2^2}{\sigma_1^2} F(n_1-1, n_2-1).$$

The general definition is as follows:

**Definition 2.2.16**

Let $W_1$ and $W_2$ be independent $\chi^2$-distributed random variables with $\nu_1$ and $\nu_2$ df, respectively. Then

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

is said to have an $F$ distribution with $\nu_1$ numerator degrees of freedom and $\nu_2$ denominator degrees of freedom.

The density function is not given here as we will be using the tables from the formula sheet.

**Example 2.2.17**

If we take independent samples of size $n_1 = 6$ and $n_2 = 10$ from two normal populations with equal population variances, find the number $b$ such that

$$P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 0.95.$$

*Solution* Since the population variances are equal we have that

$$\frac{S_1^2}{S_2^2} \sim F(n_1 - 1 = 5, n_2 - 1 = 9).$$

Also notice that $P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 1 - P\left(\frac{S_1^2}{S_2^2} > b\right) = 0.95$. Thus we find $P\left(\frac{S_1^2}{S_2^2} > b\right) = 0.05$. Using the table we see that $b = 3.48$. ☐

**Example 2.2.18**

If $Y$ is a random variable that has an $F$ distribution with $\nu_1$ numerator and $\nu_2$ denominator degrees of freedom, show that $U = \frac{1}{Y}$ has an $F$ distribution with $\nu_2$ numerator and $\nu_1$ denominator degrees of freedom.

*Solution* We are given that $Y$ has an F distribution with parameters $F(v_1, v_2)$. Thus there exist independent random variables $W_1, W_2$ with $\chi^2$ distributions with $v_1$ and $v_2$ df respectively. So we see that

$$U = \frac{1}{Y} = \frac{W_2/v_2}{W_1/v_1} \sim F(v_2, v_1).$$

☐

**Example 2.2.19**

Show that if $T$ has a $t$ distribution with $v$ df, then $U = T^2$ has an $F$ distribution with 1 numerator degree of freedom and $v$ denominator degrees of freedom.

*Solution* We are given that $T$ has a $t-$distribution with v df. Then that means there exist a standard normal random variable $Z$ and chi-squared random variable with $v$ df $W$ such that

$$T = \frac{Z}{\sqrt{W/v}}.$$

However notice that

$$U = T^2 = \frac{Z^2}{W/v}$$

where $Z^2 \sim \chi^2_{(1)}$ by Theorem 2.2.6. Since $W$ and $Z^2$ are independent we then have that $U \sim F(1, v)$ as required. ☐

We now show that $E(F) = v_2/(v_2 - 2)$ for $v_2 > 2$. Notice that

$$E(F) = E\left[\frac{W_1/v_1}{W_2/v_2}\right] = E\left[\frac{W_1}{v_1}\right] E\left[\frac{v_2}{W_2}\right] = v_2 E(W_2^{-1}) = v_2 \cdot \frac{1}{v_2 - 2}$$

as required. Note that in the last equality we used the identity for the $k$-th moment of a $\chi^2$ distribution as done in Example 2.2.15. We now find $\mathrm{Var}(F)$. We have to find $E(F^2)$. That is

$$E(F^2) = E\left[\frac{W_1^2/v_1^2}{W_2^2/v_2^2}\right] = \frac{E(W_1^2)}{v_1^2} \cdot v_2 E(W^{-2}).$$

We now use the identity for the $k-$th moment of a chi-squared distribution.

$$E(W_1^2) = \frac{\Gamma\left(\frac{v_1}{2} + 2\right)}{\Gamma\left(\frac{v_1}{2}\right)} \cdot 2^2 = 4\left(\frac{v_1}{2}\right)\left(\frac{v_1}{2} + 1\right)$$

where in the last equality I used the fact that $\Gamma(a+2) = \Gamma(a)\cdot a\cdot(a+1)$. Next we find $E(W_2^{-2})$.

$$E(W_2^{-2}) = \frac{\Gamma\left(\frac{v_2}{2}-2\right)}{\Gamma\left(\frac{v_2}{2}\right)}\cdot 2^{-2} = \frac{1}{4\left(\frac{v_2}{2}-2\right)\left(\frac{v_2}{2}-1\right)}$$

where in the last equality I used the fact that $\Gamma(a-2) = \frac{\Gamma(a)}{(a-2)(a-1)}$.

$$E(F^2) = \frac{4\left(\frac{v_1}{2}\right)\left(\frac{v_1}{2}+1\right)}{v_1^2}\cdot v_2\cdot\frac{1}{4\left(\frac{v_2}{2}-2\right)\left(\frac{v_2}{2}-1\right)} = \frac{v_1(v_1+2)v_2}{v_1^2(v_2-2)(v_2-4)}$$

Therefore,

$$\mathrm{Var}(F) = E(F^2) - [E(F)]^2 = \frac{v_1(v_1+2)v_2}{v_1^2(v_2-2)(v_2-4)} - \left(\frac{v_2}{v_2-2}\right)^2$$

for $v_2 > 4$. Simplifying this we get that

$$\mathrm{Var}(F) = \frac{2v_2^2(v_1+v_2-2)}{v_1(v_2-2)^2(v_2-4)}$$

## 2.3 The Central Limit Theorem

Recall from STA256 notes section 4.3 that even if we are given some random sample $Y_1, Y_2, \ldots, Y_n$ from a population of interest that doesn't have to be normally distributed, the Central Limit Theorem says we can still say that for a large enough $n$, the sample mean's sampling distribution is approximately normal. That is $\overline{Y} \sim N(\mu, \sigma^2/n)$. I formalize this and prove the theorem in STA256 Notes so I will give a less formal reiteration of the theorem, however I strongly suggest you reread section 4.3.

**Theorem 2.3.1 : Central Limit Theorem**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with mean $\mu$ and variance $\sigma^2$, but unknown (or non-normal) distribution. Then if $n$ is sufficiently large, $\overline{Y}$ is approximately normally distributed with mean $\mu$ and variance $\sigma^2/n$, i.e.,

$$\overline{Y} \approx N\big(\mu, \sigma^2/n\big).$$

Here "large $n$" usually refers to $n \geq 30$ but in some cases it might be much less. Once again we formalize what it means to be approximately normal in the STA256 but really this means that if

$$U_n = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$$

then

$$\lim_{n\to\infty} P(U_n \leq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

**Example 2.3.2**

The fracture strength of tempered glass averages 14 (measured in thousands of pounds per square inch) and has standard deviation 2.

1. What is the probability that the average fracture strength of 100 randomly selected pieces of this glass exceeds 14.5?
2. Find an interval that includes, with probability 0.95, the average fracture strength of 100 randomly selected pieces of this glass.

*Solution* For the first part: we are given that the population mean $\mu = 14$ and the population variance is $\sigma^2 = 4$. We are asked to find $P(\overline{Y} \geq 14.5)$ where $\overline{Y} = 1/n \sum_{i=1}^{100} Y_i$ where the distribution for $Y_i$ are unknown. Using the Central Limit Theorem we can approximate this with the normal distribution. We standardize this and get that

$$P(\overline{Y} \geq 14.5) = P\left[\frac{\overline{Y} - \mu}{\sigma/\sqrt{100}} \geq \frac{14.5 - 14}{2/\sqrt{100}}\right] = P(Z \geq 2.5).$$

Using the standard normal table we see that

$$P(\overline{Y} \geq 14.5) = 0.0064.$$

For the second part: We are asked to find an interval such that $P(a \leq \overline{Y} \leq b) = 0.95$. Using the Central Limit Theorem we can instead find

$$P(a \leq \overline{Y} \leq b) = P\left[\frac{a - 14}{2/\sqrt{100}} \leq Z \leq \frac{b - 14}{2/\sqrt{100}}\right] = 0.95.$$

We find $P(Z \leq p) = 0.9750$ which gives us $p = 1.96$. Thus we have that $-1.96 = \frac{a-14}{0.2}$ or $a = 13.608$ and similarly for $b = 14.392$. We see that the interval $[13.608, 14.392]$ includes the average fracture strenght of 100 randomly selected pieces of this glass with 0.95 probability. $\square$

**Example 2.3.3**

Achievement test scores of all high school seniors in a state have mean 60 and variance 64. A random sample of $n = 100$ students from one large high school had a mean score of 58. Is there evidence to suggest that this high school is inferior? (Calculate the probability that the sample mean is at most 58 when $n = 100$.)

*Solution* We are given that the population mean is $\mu = 60$ and the population variance is $\sigma^2 = 64$. We are asked to find the probability that the sample mean is at most 58. Using the Central Limit Theorem we can approximate this with the normal distribution. We standardize this and get that

$$P(\overline{Y} \leq 58) = P\left[\frac{\overline{Y} - \mu}{\sigma/\sqrt{100}} \leq \frac{58 - 60}{8/\sqrt{100}}\right] = P(Z \leq -2.5).$$

Using the standard normal table we see that

$$P(Z \leq -2.5) = 0.0062.$$

$\square$

**Example 2.3.4**

The service times for customers coming through a checkout counter in a retail store are independent random variables with mean 1.5 minutes and variance 1.0. Approximate the probability that 100 customers can be served in less than 2 hours of total service time.

*Solution* Let $T = \sum_{i=1}^{100} Y_i$ be the total service time. Since each $Y_i$ has mean 1.5 and variance 1.0, we have

$$E(T) = 100 \times 1.5 = 150$$

and

$$\text{Var}(T) = 100 \times 1.0 = 100.$$

We are asked to find $P(T < 120)$, since 2 hours = 120 minutes. Notice that

$$P(T < 120) = P\left(\overline{Y} < \frac{120}{100} = 1.2\right).$$

Since the sample size is large the Central Limit Theorem says that we can approximate this with the Normal Distirbution. Standardizing this we get that

$$P\left(Z < \frac{1.2 - 1.5}{1.0/\sqrt{100}}\right) = P(Z < -3).$$

Using the standard normal table we see that

$$P(Z < -3) = 0.0013.$$

□

## 2.4 The Normal Approximation to the Binomial Distribution

We can use the Central Limit Theorem to use the Normal Distribution to approximate discerete distributions. For example as we seen in STA256 Notes, we can approximate the Binomial distribution when the probabilities calculationsare tedious to do. Let $Y$ be a binomial distribution with probability of success $p$. If we wanted to find $P(Y \leq b)$ we can use the pdf for the binomial distirbution and find $P(Y = y)$ for each non-negative $y$ less than or equal to $b$. However we instead can let

$$Y = \sum_{i=1}^{n} X_i$$

where $X_i = 1$ if the $i$-th trial is a success and 0 otherwise. We then can easily verify that $E(X_i) = p$ and that $\text{Var}(X_i) = p(1-p)$ for all $i = 1, 2, \dots, n$. It is also true that each $X_i$ are independent since each trial is independent. Then we have that when $n$ is large,

$$\frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

has an approximately normal distirbution with mean $p$ and variance $p(1-p)/n$. We can conclude this by saying that if $Y$ has a binomial distribution $Y \sim Bin(n, p)$ then $Y \approx N(np, np(1-p))$.

Moreover when we do approximate the binomial distribution with the Normal distribution we have to use Continuity Correction which is detailed in STA256 Notes. The basic idea is to treat the discrete value $b$ as the continuous interval from $b - 0.5$ to $b + 0.5$ giving the following adjustments:

- $P(Y = b) = P(b - 0.5 \leq Y \leq b + 0.5)$
- $P(Y \leq b) = P(Y \leq b + 0.5)$
- $P(b \leq Y) = P(b - 0.5 \leq Y)$

**Example 2.4.1**

A new computer virus attacks a folder consisting of 200 files. Each file gets damaged with probability 0.2 independently of other files. What is the probability that fewer than 50 files get damaged?

*Solution* We are given that the probability that a file is damaged by the virus is $p = 0.2$. Since this follows a binomial distribution $Y \sim Bin(100, 0.2)$ we can approximate it with the Normal Distribution. We are asked to find $P(Y \leq 49)$. we can approximate it by $N(np = 40, np(1-p) = 32)$. Thus standarizing this we get that

$$P(Y \leq 49) \approx P\left(Z \leq \frac{49 + 0.5 - 40}{\sqrt{32}}\right) = P(Z \leq 1.679) = 0.9525.$$

□

**Example 2.4.2**

A pair of dice is rolled 180 times. What is the probability that a total of 7 occurs at least 25 times?

*Solution* Here we are given a balanced die with $n = 180$ rolls. Let $Y_i = 1$ if the $i-$th roll is a 7 and zero otherwise. We know that $P(Y_i = y) = p = 1/6$. We know that $Y = \sum_{i=1}^{180} Y_i \sim Bin(180, 1/6)$ We are asked to find

$$P\left(Y = \sum_{i=1}^{180} Y_i \geq 25\right) = 1 - P(Y \leq 24).$$

Thus we can approximate this with $N(np = 30, 5)$. We standardize and Continuity correct this and get that

$$P(Y \geq 25) = P\left(Z \geq \frac{24.5 - 30}{5}\right) = P(Z \geq -1.1) = 1 - P(Z \leq -1.1) = 1 - P(Z > 1.1) = 1 - 0.93570.8643.$$

□

**Example 2.4.3**

Candidate A believes that she can win a city election if she can earn at least 55% of the votes in precinct 1. She also believes that about 50% of the city's voters favor her. If $n = 100$ voters show up to vote at precinct 1, what is the probability that candidate A will receive at least 55% of their votes?

*Solution* We are given that the probability that a voter form precint 1 will vote for Candidate A is $p = 0.5$. Let $Y$ be the number of votes that Candidate A recieved. Then $Y \sim Bin(100, 0.5)$. We are asked to find $P(Y \geq 55)$. Again since the sample size is large we can approximate this with the Normal Distribution $N(np = 50, np(1-p) = 25)$. Thus using continuity correction and standarizing we get that

$$P(Y \geq 55) = P\left(Z \geq \frac{54.5 - 50}{5}\right) = P(Z \geq 0.9) = 0.1841.$$

Note that the textbook finds $P(Y \geq 0.55)$ which doesn't make sense in this question. □

## 2.5 Practice Problems

**7.9**

Refer to Example 2.2.2. The amount of fill dispensed by a bottling machine is normally distributed with $\sigma = 1$ ounce. If $n = 9$ bottles are randomly selected from the output of the machine, we found that the probability that the sample mean will be within .3 ounce of the true mean is .6318. Suppose that $\overline{Y}$ is to be computed using a sample of size $n$.

(a) If $n = 16$, what is $P(|\overline{Y} - \mu| \leq .3)$?

(b) Find $P(|\overline{Y} - \mu| \leq .3)$ when $\overline{Y}$ is to be computed using samples of sizes $n = 25$, $n = 36$, $n = 49$, and $n = 64$.

(c) What pattern do you observe among the values for $P(|\overline{Y} - \mu| \leq .3)$ that you observed for the various values of $n$?

(d) Do the results that you obtained in part (b) seem to be consistent with the result obtained in Example 2.2.3"?

*Solution* For (a): We need to find $P(|\overline{Y} - \mu| \leq 0.3)$ when $n = 16$. We know that from Theorem 2.2.6 that $\overline{Y} \sim N(\mu, 1/16)$. Then we can see that

$$P\left(\frac{-0.3}{\sqrt{1/16}} \leq \frac{\overline{Y} - \mu}{\sqrt{1/16}} \leq \frac{0.3}{\sqrt{1/16}}\right) = P(-1.2 \leq Z \leq 1.2).$$

Using the standard normal table we see that

$$P(|\overline{Y} - \mu| \leq .3) = 0.7698.$$

For (b): We repeat what we did in part (a) for each sample size.

For $n = 25$:

$$P\left(\frac{-0.3}{1/5} \leq Z \leq \frac{0.3}{1/5}\right) = P(-1.5 \leq Z \leq 1.5)$$

From the standard normal table: $P(-1.5 \leq Z \leq 1.5) = 0.8664$.

For $n = 36$:

$$P\left(\frac{-0.3}{1/6} \leq Z \leq \frac{0.3}{1/6}\right) = P(-1.8 \leq Z \leq 1.8)$$

From the standard normal table: $P(-1.8 \leq Z \leq 1.8) = 0.9281$.

For $n = 49$:

$$P\left(\frac{-0.3}{1/7} \leq Z \leq \frac{0.3}{1/7}\right) = P(-2.1 \leq Z \leq 2.1)$$

From the standard normal table: $P(-2.1 \leq Z \leq 2.1) = 0.9642$.

For $n = 64$:

$$P\left(\frac{-0.3}{1/8} \le Z \le \frac{0.3}{1/8}\right) = P(-2.4 \le Z \le 2.4)$$

From the standard normal table: $P(-2.4 \le Z \le 2.4) = 0.9836$. For (c): As $n$ increases the probability $P(|\overline{Y} - \mu| \le 0.3)$ also increases. This makes sense because increasing the sample size decreases the standard deviation of the sampling distribution of the sample mean making $\overline{Y}$ more concertrated around the mean $\mu$. For (d): the result is consistent with Example 2.2 as in Example 2.2 it was said that $n = 46$ sample size was needed for the probability of $P(|\overline{Y} - \mu| \le 0.3) = 0.95$ which matches our result. So the probabilities are: - $n = 25$: 0.8664 - $n = 36$: 0.9281 - $n = 49$: 0.9642 - $n = 64$: 0.9836 □

**7.10**

Refer to Problem 7.9. Assume now that the amount of fill dispensed by the bottling machine is normally distributed with $\sigma = 2$ ounces.

(a) If $n = 9$ bottles are randomly selected from the output of the machine, what is $P(|\overline{Y} - \mu| \le .3)$? Compare this with the answer obtained in Example 7.2.

(b) Find $P(|\overline{Y} - \mu| \le .3)$ when $\overline{Y}$ is to be computed using samples of sizes $n = 25$, $n = 36$, $n = 49$, and $n = 64$.

(c) What pattern do you observe among the values for $P(|\overline{Y} - \mu| \le .3)$ that you observed for the various values of $n$?

(d) How do the respective probabilities obtained in this problem (where $\sigma = 2$) compare to those obtained in Exercise 7.9 (where $\sigma = 1$)?

*Solution* For (a): We now have that $\overline{Y} \sim N(\mu, 2/9)$. We then see that

$$P\left(\frac{-0.3}{2/\sqrt{9}} \le Z \le \frac{0.3}{2/\sqrt{9}}\right) = 0.3472.$$

For $n = 25$:

$$P\left(\frac{-0.3}{2/5} \le Z \le \frac{0.3}{2/5}\right) = P(-0.75 \le Z \le 0.75)$$

From the standard normal table: $P(-0.75 \le Z \le 0.75) = 0.5468$.

For $n = 36$:

$$P\left(\frac{-0.3}{2/6} \le Z \le \frac{0.3}{2/6}\right) = P(-0.9 \le Z \le 0.9)$$

From the standard normal table: $P(-0.9 \le Z \le 0.9) = 0.8159 - 0.1841 = 0.6318$.

For $n = 49$:

$$P\left(\frac{-0.3}{2/7} \le Z \le \frac{0.3}{2/7}\right) = P(-1.05 \le Z \le 1.05)$$

From the standard normal table: $P(-1.05 \le Z \le 1.05) = 0.8531 - 0.1469 = 0.7062$.

For $n = 64$:

$$P\left(\frac{-0.3}{2/8} \le Z \le \frac{0.3}{2/8}\right) = P(-1.2 \le Z \le 1.2)$$

From the standard normal table: $P(-1.2 \leq Z \leq 1.2) = 0.8849 - 0.1151 = 0.7698$. For (c): Same answer for previous problem. For (d): With $\sigma = 1$ the sample sizes yielded higher probabilties so this means higher variability leads to a lower accuracy of the sample mean. $\square$

**7.12**

Suppose the forester in Example 2.2.4 would like the sample mean to be within 1 square inch of the population mean, with probability .90. How many trees must he measure in order to ensure this degree of accuracy?

*Solution* We are asked to find the minimum sample size $n$ such that

$$P(|\overline{Y} - \mu| \leq 1) = P(-1 \leq \overline{Y} - \mu \leq 1) = 0.90.$$

Since the sample mean is normally distributed we can standardize this and get that

$$P(-1 \leq \overline{Y} - \mu \leq 1) = P\left(\frac{-1}{4/\sqrt{n}} \leq Z \leq \frac{1}{4/\sqrt{n}}\right) = 1 - 2P\left(Z \leq \frac{\sqrt{n}}{4}\right) = 0.90.$$

Where in the last equality we used the fact that the Normal distirbution is symmetric. We now find the value that statisfies $P(Z > b) = 0.05$ from the standard normal table. We see that $b \approx 1.654$ satisfies this and thus we get that

$$n = (4 \cdot 1.654)^2 \approx 43.3$$

rounding up we get that $n = 44$. $\square$

**7.13**

The Environmental Protection Agency is concerned with the problem of setting criteria for the amounts of certain toxic chemicals to be allowed in freshwater lakes and rivers. A common measure of toxicity for any pollutant is the concentration of the pollutant that will kill half of the test species in a given amount of time (usually 96 hours for fish species). This measure is called LC50 (lethal concentration killing 50% of the test species). In many studies, the values contained in the natural logarithm of LC50 measurements are normally distributed, and, hence, the analysis is based on ln(LC50) data. Studies of the effects of copper on a certain species of fish (say, species A) show the variance of ln(LC50) measurements to be around .4 with concentration measurements in milligrams per liter. If $n = 10$ studies on LC50 for copper are to be completed, find the probability that the- sample mean of ln(LC50) will differ from the true population mean by no more than .5.

*Solution* We are given that the population variance is $\sigma^2 = 0.4$. Let $\overline{Y}$ denote the sample mean of $\ln(LC50)$. We are asked to find $P(|\overline{Y} - \mu| \leq 0.5) = P(-0.5 \leq \overline{Y} - \mu \leq 0.5)$. Since this data is normally distributed we can standardize this to find that

$$P(-0.5 \leq \overline{Y} - \mu \leq 0.5) = P\left(\frac{-0.5}{\sqrt{0.4}/\sqrt{10}} \leq Z \leq \frac{0.5}{\sqrt{0.4}/\sqrt{10}}\right) = P(-2.5 \leq Z \leq 2.5) = 0.9876.$$

$\square$

**7.14**

If in Problem 7.13 we want the sample mean to differ from the population mean by no more than .5 with probability .95, how many tests should be run?

*Solution* We are asked to find the required sample size $n$ such that $P(|\overline{Y}-\mu| \leq 0.5) = 0.95$. We see that

$$P(|\overline{Y}-\mu| \leq 0.5) = P\left(\frac{-\sqrt{n}0.5}{\sqrt{0.4}} \leq Z \leq \frac{\sqrt{n}0.5}{\sqrt{0.4}}\right) = 1 - 2P\left(Z > \frac{\sqrt{n}0.5}{\sqrt{0.4}}\right) = 0.95.$$

We see that $P(Z > 1.96) = 0.025$ and so

$$n = \left(\frac{\sqrt{0.4}\cdot 1.96}{0.5}\right)^2 = 6.14656$$

rounding up we find that $n = 7$ is the required sample size. □

**7.20**

Ammeters produced by a manufacturer are marketed under the specification that the standard deviation of gauge readings is no larger than .2 amp. One of these ammeters was used to make ten independent readings on a test circuit with constant current. If the sample variance of these ten measurements is .065 and it is reasonable to assume that the readings are normally distributed, do the results suggest that the ammeter used does not meet the marketing specifications? [Hint: Find the approximate probability that the sample variance will exceed .065 if the true population variance is .04.]

*Solution* The manufacturer claims that the gauge readings standard deviation $\sigma = 0.2$. We are given that the sample variance $S^2 = 0.065$ when $n = 10$. We are also given that $S^2 = \frac{1}{n-1}\sum_{i=1}^{10} Y_i$ where $Y_i$ are normally distributed for $i = 1, 2, \ldots, 10$. We are asked whether the observed sample variance of $S^2 = 0.065$ is consistent with the manufacturer's claim that the population variance is $\sigma^2 = 0.04$, based on a sample of size $n = 10$. We are asked to find $P(S^2 > 0.065)$. Using Theorem 2.2.9 since the data is normally distributed we know that $(n-1)S^2/\sigma^2 \sim \chi^2_{(n-1)}$. So we have that

$$P(S^2 > 0.065) = P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{9\cdot 0.065}{0.04}\right) = P(\chi^2_{(9)} > 14.625) \approx 0.10.$$

□

**7.21**

Refer to Exercise 7.13. Suppose that $n = 20$ observations are to be taken on ln(LC50) measurements and that $\sigma^2 = 1.4$. Let $S^2$ denote the sample variance of the 20 measurements.

(a) Find a number $b$ such that $P(S^2 \leq b) = 0.975$.

(b) Find a number $a$ such that $P(a \leq S^2) = 0.975$.

(c) If $a$ and $b$ are as in parts (a) and (b), what is $P(a \leq S^2 \leq b)$?

*Solution* For (a): This is a classic sample variance question. We will turn this into a chi-squared distribution since the data lnLC50 is normally distributed. That is

$$P(S^2 \leq b) = P\left(\frac{(n-1)S^2}{\sigma^2} \leq \frac{(n-1)b}{\sigma^2}\right) = P\left(\chi^2_{(19)} \leq \frac{(n-1)b}{\sigma^2}\right) = 0.975.$$

Using the table we get that $P(\chi^2_{(19)} \leq 32.8523) = 0.975$. Solving for $b$ we get that

$$b = \frac{(32.8523)(1.4)}{19} \approx 2.418.$$

For (b): Similarly we can instead find

$$P(S^2 \geq a) = P\left(\chi^2_{(19)} \geq \frac{19a}{1.4}\right) = 0.975.$$

We get that $P(\chi^2_{(19)} \geq 8.90655) = 0.975$. Solving for $a$ we find that

$$a = \frac{(8.90655)(1.4)}{19} \approx 0.656.$$

For (c): We are asked to find $P(0.656 \leq S^2 \leq 2.418)$. Notice that

$$P(a \leq S^2 \leq b) = P(S^2 \leq b) - P(S^2 \leq a) = 0.975 - 0.025 = 0.95.$$

$\square$

Note that questions 7.29,7.30,7.33, 7.34, 7.37 are done as examples.

**7.36**

Let $S_1^2$ denote the sample variance for a random sample1 of ten ln(LC50) values for copper and let $S_2^2$ denote the sample variance for a random sample of eight ln(LC50) values for lead, both samples using the same species of fish. The population variance for measurements on copper is assumed to be twice the corresponding population variance for measurements on lead. Assume $S_1^2$ to be independent of $S_2^2$.

(a) Find a number $b$ such that

$$P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 0.95.$$

(b) Find a number $a$ such that

$$P\left(a \leq \frac{S_1^2}{S_2^2}\right) = 0.95.$$

*[Hint: Use the result of Exercise 7.29 and notice that $P(U_1/U_2 \leq k) = P(U_2/U_1 \geq 1/k)$].*

(c) If $a$ and $b$ are as in parts (a) and (b), find

$$P\left(a \leq \frac{S_1^2}{S_2^2} \leq b\right).$$

*Solution* For (a): In order to calculate this probabiltity we can rewrite this as a $F$-distribution.

$$P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq \left(\frac{\sigma_2^2}{\sigma_1^2}\right) b\right).$$

Using the fact that $\sigma_1^2 = 2\sigma_2^2$ we get that

$$P(F \leq 0.5b) = 0.95$$

where $F \sim F(9,7)$. Using the table for the $F$-distribution we get that $0.5b = 3.68$ or $b = 7.36$.

For (b): Since our table does not give us the tail percentage point of 0.95 we instead use the fact that

$$P\left(\frac{S_1^2}{S_2^2} \geq a\right) = P\left(\frac{S_2^2}{S_1^2} \leq \frac{1}{a}\right).$$

So we instead find $P(F \leq 1/(2a)) = 0.95$ where $F \sim F(7,9)$. We get that

$$\frac{1}{2a} = 3.29 \quad \text{or} \quad a = \frac{1}{2 \times 3.29} = 0.151975683891.$$

For (c): We are asked to find $P\left(a \leq \frac{S_1^2}{S_2^2} \leq b\right)$. This is simply the probability that $\frac{S_1^2}{S_2^2}$ falls between $a$ and $b$. From parts (a) and (b), we have:

$$P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 0.95$$

and

$$P\left(a \leq \frac{S_1^2}{S_2^2}\right) = 0.95$$

Since the lower and upper tails each cut off 0.05 probability, the probability that $\frac{S_1^2}{S_2^2}$ falls between $a$ and $b$ is:

$$P\left(a \leq \frac{S_1^2}{S_2^2} \leq b\right) = 0.95 - 0.05 = 0.90$$

So, the answer is 0.90. □

**7.38**

Suppose that $Y_1, Y_2, \ldots, Y_5, Y_6, \overline{Y}, W,$ and $U$ are as defined in Example 2.2.12. What is the distribution of "

(a) $\frac{\sqrt{5}Y_6}{\sqrt{W}}$? Why?

(b) $\frac{2Y_6}{\sqrt{U}}$? Why?

(c) $2\left(\frac{5\overline{Y}^2 + Y_6^2}{U}\right)$? Why?

*Solution* For (a): We know that $W \sim \chi^2_{(5)}$ and $U \sim \chi^2_{(4)}$. Notice that

$$\frac{\sqrt{5}Y_6}{\sqrt{W}} = \frac{T}{\sqrt{W/5}}$$

where $T$ is a standard normal distribution. Thus it is clear that this is a $t$-distribution with 5 df. For (b): Similarly we have that

$$\frac{2Y_6}{\sqrt{W}} = \frac{T}{\sqrt{W/4}}$$

thus this is a $t$-distribution with 4 df. For (c): Notice that $\overline{Y} \sim N(0, 1/5)$. We can transform this so we get $Z = \sqrt{5} \cdot \overline{Y} \sim N(0,1)$ so then $Z^2 \sim \chi^2_{(1)}$. So we get that $\overline{Y}^2 \sim \chi^2_{(1)}/5$. Thus we have that

$$5\overline{Y}^2 + Y_6^2 = \chi^2_{(1)} + \chi^2_{(1)} = \chi^2_{(2)}.$$

Next we see that
$$\frac{\chi^2_{(2)}/2}{\chi^2_{(4)}/4} = 2 \cdot \frac{\chi^2_{(2)}}{\chi^2_{(4)}} = 2\left(\frac{5\overline{Y}^2 + Y_6^2}{U}\right)$$
thus we have that it is a $F$-distribution with parameters $F(1, 3)$. $\square$

**7.39**

Suppose that independent samples (of sizes $n_i$) are taken from each of $k$ populations and that population $i$ is normally distributed with mean $\mu_i$ and variance $\sigma^2$, $i = 1, 2, \ldots, k$. That is, all populations are normally distributed with the *same* variance but with (possibly) different means. Let $\overline{X}_i$ and $S_i^2$, $i = 1, 2, \ldots, k$ be the respective sample means and variances. Let

$$\theta = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k,$$

where $c_1, c_2, \ldots, c_k$ are given constants.

(a) Give the distribution of

$$\hat{\theta} = c_1\overline{X}_1 + c_2\overline{X}_2 + \cdots + c_k\overline{X}_k.$$

Provide reasons for any claims that you make.

(b) Give the distribution of

$$\frac{\text{SSE}}{\sigma^2}, \quad \text{where SSE} = \sum_{i=1}^{k}(n_i - 1)S_i^2.$$

Provide reasons for any claims that you make.

Give the distribution of

$$\frac{\hat{\theta} - \theta}{\sqrt{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \cdots + \frac{c_k^2}{n_k}\right)\text{MSE}}}, \quad \text{where MSE} = \frac{\text{SSE}}{n_1 + n_2 + \cdots + n_k - k}.$$

Provide reasons for any claims that you make.

*Solution* For (a): We begin by noticing that $\overline{X}_i \sim N(\mu_i, \sigma^2/n_i)$. Then $\hat{\theta}$ is a linear combonation of normal random variables which we know that

$$\hat{\theta} \sim N\left(\sum_{i=1}^{k} c_i\mu_i, \sum_{i=1}^{k} c_i^2 \cdot \frac{\sigma^2}{n_i}\right) = N\left(\theta, \sigma^2 \sum_{i=1}^{k} c_i^2 \frac{c_i^2}{n_i}\right).$$

For (b): The first thing to notice is that

$$\frac{\text{SSE}}{\sigma^2} = \sum_{i=1}^{k} \frac{(n_i - 1)S_i^2}{\sigma^2}.$$

Notice that $\frac{(n_i-1)S_i^2}{\sigma^2} \sim \chi^2_{(n_i-1)}$. Thus we have

$$\frac{\text{SSE}}{\sigma^2} = \sum_{i=1}^{k} \chi^2_{(n_i-1)} \sim \chi^2_{(\sum_{i=1}^{k} n_i - k)}$$

For (c): We are asked for the distribution of

$$T = \frac{\hat{\theta} - \theta}{\sqrt{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \cdots + \frac{c_k^2}{n_k}\right)\text{MSE}}}, \quad \text{where MSE} = \frac{\text{SSE}}{n_1 + n_2 + \cdots + n_k - k}.$$

From part (a),

$$\hat{\theta} \sim N\left(\theta, \sigma^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}\right)$$

So,

$$\hat{\theta} - \theta \sim N\left(0, \sigma^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}\right)$$

From part (b),

$$\text{SSE}/\sigma^2 \sim \chi^2_{(\sum_{i=1}^{k} n_i - k)}$$

So,

$$\text{MSE} = \frac{\text{SSE}}{\sum_{i=1}^{k} n_i - k}$$

is independent of $\hat{\theta}$ (since sample means and sample variances are independent for normal samples). Let $V = \sum_{i=1}^{k} \frac{c_i^2}{n_i}$, then

$$T = \frac{\hat{\theta} - \theta}{\sqrt{V \cdot \text{MSE}}}$$

Write numerator as $Z\sqrt{\sigma^2 V}$, denominator as $\sqrt{V \cdot \text{MSE}} = \sqrt{V}\sqrt{\text{MSE}}$.So,

$$T = \frac{Z\sqrt{\sigma^2 V}}{\sqrt{V}\sqrt{\text{MSE}}} = \frac{Z\sigma}{\sqrt{\text{MSE}}}$$

where $Z \sim N(0,1)$. But $\text{MSE}/\sigma^2 = \frac{\text{SSE}/\sigma^2}{\sum_{i=1}^{k} n_i - k} \sim \frac{\chi^2_{(\sum_{i=1}^{k} n_i - k)}}{\sum_{i=1}^{k} n_i - k}$. Thus,

$$T = \frac{Z}{\sqrt{W/(\sum_{i=1}^{k} n_i - k)}}$$

where $Z \sim N(0,1)$, $W \sim \chi^2_{(\sum_{i=1}^{k} n_i - k)}$, and $Z$ and $W$ are independent.By the definition of the $t$-distribution,

$$T \sim t_{(\sum_{i=1}^{k} n_i - k)}$$

That is, $T$ has a $t$-distribution with $\sum_{i=1}^{k} n_i - k$ degrees of freedom. The numerator is standard normal, the denominator is the square root of an independent chi-squared variable divided by its degrees of freedom, matching the definition of the $t$-distribution.

$$\frac{\hat{\theta} - \theta}{\sqrt{\left(\frac{c_1^2}{n_1} + \cdots + \frac{c_k^2}{n_k}\right)\text{MSE}}} \sim t_{(\sum_{i=1}^{k} n_i - k)}$$

□

### 7.43

An anthropologist wishes to estimate the average height of men for a certain race of people. If the population standard deviation is assumed to be 2.5 inches and if she randomly samples 100 men, find the probability that the difference between the sample mean and the true population mean will not exceed .5 inch.

*Solution* We are asked to find $P(|\overline{Y} - \mu| \leq 0.5) = P(-0.5 \leq \overline{Y} - \mu \leq 0.5)$. We see that we can approximate this with the Normal Distribution $N(\mu, \sigma^2/100)$ since

$$P\left(\frac{-0.5}{2.5/10} \leq Z \leq \frac{0.5}{2.5/10}\right) = 0.9544.$$

□

### 7.44

Suppose that the anthropologist of Problem 7.43 wants the difference between the sample mean and the population mean to be less than .4 inch, with probability .95. How many men should she sample to achieve this objective?

*Solution* We are asked to find the reuqired sample size $n$ such that $P(|\overline{Y} - \mu| \leq 0.4) = 0.95$. That is

$$P\left(\frac{-0.4\sqrt{n}}{2.5} \leq Zleq\frac{0.4\sqrt{n}}{2.5}\right) = 1 - 2P\left(Z > \frac{0.4\sqrt{n}}{2.5}\right) = 0.95.$$

We see that $P(Z > 1.96) = 0.025$. Solving for $n$ we have that

$$n = \left(\frac{(2.5)(1.96)}{0.4}\right)^2 = 150.06$$

round up we have that She must sample at least 151 men to be 95% confident that the sample mean is within 0.4 inches of the population mean. □

### 7.45

Workers employed in a large service industry have an average wage of \$7.00 per hour with a standard deviation of \$.50. The industry has 64 workers of a certain ethnic group. These workers have an average wage of \$6.90 per hour. Is it reasonable to assume that the wage rate of the ethnic group is equivalent to that of a random sample of workers from those employed in the service industry? [Hint: Calculate the probability of obtaining a sample mean less than or equal to \$6.90 per hour.]

*Solution* We are asked to find the probability that a random sample of worker wages from the service industry is less than or equal to \$6.90 per hour. If this probability is small then it is not reasonable to assume that the wages are equivalent. We need to find $P(\overline{Y} \leq 6.90)$. Using the Central Limit theorem we can approximate it with the Normal distribution $N(7, 0.50/64)$. We see that

$$P(\overline{Y} \leq 6.90) \approx P\left(Z \leq \frac{6.90 - 7.00}{0.50/\sqrt{64}} = P(Z \leq -1.6)\right) = P(Z \geq 1.6) = 0.0548.$$

Since the probability is small (%5.48) it is not reasonable to assume that the wages are equivalent. □

**7.46**

The acidity of soils is measured by a quantity called the pH, which may range from 0 (high acidity) to 14 (high alkalinity). A soil scientist wants to estimate the average pH for a large field by randomly selecting $n$ core samples and measuring the pH in each sample. the population standard deviation of pH measurements is not known, past experience indicates that most soils have a pH value of between 5 and 8. If the scientist selects $n = 40$ samples, find the approximate probability that the sample mean of the 40 pH measurements will be within .2 unit of the true average pH for the field. [Hint: See Exercise 1.17.]

*Solution* Let $\overline{Y}$ denote the sample mean of the pH measurements of the field. We are asked to find $P(|\overline{Y} - \mu| \leq 0.2) = P(-0.2 \leq \overline{Y} - \mu \leq 0.2)$. We do not know the sandard deviation however we can approximate it since the Range $\approx 4\sigma$. This is because around %95 of the data lies within $\pm 2$ standard deviation from the mean. Thus we can say that $\sigma \approx 0.75$. Thus we have that

$$P\left(\frac{-0.2}{0.75/\sqrt{40}} \leq Z \leq \frac{0.2}{0.75/\sqrt{40}}\right) = 1 - 2P(Z \geq 1.6865) = 1 - 2 \cdot (0.0465) = 0.9078.$$

□

**7.47**

Suppose that the scientist of Exercise 7.46 would like the sample mean to be within .1 of the true mean with probability .90. How many core samples should the scientist take?

*Solution* We are now asked to find the required sample size so that $P(-0.1 \leq \overline{Y} - \mu \leq 0.1) = 0.90$. We find that

$$P(-0.1 \leq \overline{Y} - \mu \leq 0.1) = 1 - 2P\left(Z \geq \frac{0.1\sqrt{n}}{0.75}\right) = 0.90.$$

We then find $P(Z \geq b) = 0.05$. We find that $b = 1.648$. Solving for $n$ we have that

$$n = \left(\frac{0.75 \cdot 1.648}{0.1}\right)^2 = 152.29.$$

Rounding up we get that $n = 151$ is the required sample size. □

**7.48**

An important aspect of a federal economic plan was that consumers would save a substantial portion of the money that they received from an income tax reduction. Suppose that early estimates of the portion of total tax saved, based on a random sampling of 35 economists, had mean 26% and standard deviation 12%.

(a) What is the approximate probability that a sample mean estimate, based on a random sample of $n = 35$ economists, will lie within 1% of the mean of the population of the estimates of all economists?

(b) Is it necessarily true that the mean of the population of estimates of all economists is equal to the percent tax saving that will actually be achieved?

*Solution* For (a): We are asked to find the probability that the sample mean is withint %1 percent of the true mean. That is we are asked find $P(-0.01 \leq \overline{Y} - \mu \leq 0.01)$. We can use the Central Limit Theorem here to approximate this probability with $N(\mu, \sigma/n = 0.12/35)$. we find that

$$P(-0.01 \leq \overline{Y} - \mu \leq 0.01) = P\left(\frac{-0.01}{0.12\sqrt{35}} \leq Z \leq \frac{-0.01}{0.12\sqrt{35}}\right) = 1 - 2P(Z \geq 0.493) = 1 - 2\cdot(0.3121) = 0.3758.$$

For (b): No. The mean of the population of estimates reflects what economists believe will happen not what actually happens. It is an estimate based on expert opinion, not a measurement of the true savings. So it can be biased or inaccurate. □

**7.49**

The length of time required for the periodic maintenance of an automobile or another machine usually has a mound-shaped probability distribution. Because some occasional long service times will occur, the distribution tends to be skewed to the right. Suppose that the length of time required to run a 5000-mile check and to service an automobile has mean 1.4 hours and standard deviation .7 hour. Suppose also that the service department plans to service 50 automobiles per 8-hour day and that, in order to do so, it can spend a maximum average service time of only 1.6 hours per automobile. On what proportion of all workdays will the service department have to work overtime?

*Solution* We are asked to find the proportion on all workdays where the service department will have to work overtime. The threshold for the service department to start working overtime is 1.6 hours for each automobile. Thus we find the probability $P(\overline{Y} > 1.6)$ where $\overline{Y} \sim N(1/4, \sigma/n = 0.7/50)$. Thus we have that

$$P(\overline{Y} > 1.6) = P\left(Z > \frac{1.6 - 1.4}{0.7/\sqrt{50}}\right) = P(Z > 2.02) \approx 0.0217.$$

Thus they will have to work overtime on %2.16 of workdays. □

**7.50**

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). If 100 test welds are to be measured, what is the approximate probability that the sample mean will be within 1 psi of the true population mean?

*Solution* We are asked to find $P(-1 \leq \overline{Y} - \mu \leq 1)$. Using the central limit theorem we have that

$$P(-1 \leq \overline{Y} - \mu \leq 1) = P\left(\frac{-1}{10/\sqrt{100}} \leq Z \leq \frac{-1}{10/\sqrt{100}}\right) = 1 - 2P(Z \geq 1) = 1 - 2(0.1587) = 0.6826.$$

□

**7.51**

Refer to Exercise 7.50. If the standard deviation of shear strength measurements for spot welds is 10 psi, how many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

*Solution* We now need to find the required sample size $n$ such that $P(-1 \le \overline{Y} - \mu \le 1) = 0.99$. Again we see that

$$P(-1 \le \overline{Y} - \mu \le 1) = 1 - 2P\left(Z \ge \frac{1\sqrt{n}}{10}\right) = 0.99.$$

We find $P(Z \ge a) = 0.005$ from the table. We see that $a = 2.75$. Solving for $n$ we get that

$$n = \left(\frac{10 \times 2.75}{1}\right)^2 = (27.5)^2 = 756.25$$

Rounding up, we find that at least $n = 757$ test welds should be sampled to ensure the sample mean is within 1 psi of the true mean with probability approximately 0.99. □

**7.52**

Resistors to be used in a circuit have average resistance 200 ohms and standard deviation 10 ohms. Suppose 25 of these resistors are randomly selected to be used in a circuit.

(a) What is the probability that the average resistance for the 25 resistors is between 199 and 202 ohms?

(b) Find the probability that the *total* resistance does not exceed 5100 ohms. *[Hint: see Example 7.9.]*

*Solution* For (a): We are asked to find $P(25 \le \overline{Y} \le 199)$. We can approximate this with the Normal Distritbuion by using the Central Limit Theorem. That is

$$P(25 \le \overline{Y} \le 199) = P\left(\frac{199 - 200}{10\sqrt{25}} \le Z \le \frac{202 - 200}{10\sqrt{25}}\right) = P(-0.5 \le Z \le 1) = 0.5328.$$

For (b): We need to find $P(T = \sum_{i=1}^{25} X_i \le 5100)$ where each $X_i$ are the individual resistence. We then see that $T \sim N(n\mu, \sigma^2)$. Thus we see that

$$P(T \le 5100) = P\left(Z \le \frac{5100 - 25 \cdot 200}{10^2}\right) = P(Z \le 2) = 0.9772.$$

□

**7.53**

One-hour carbon monoxide concentrations in air samples from a large city average 12 ppm (parts per million) with standard deviation 9 ppm.

(a) Do you think that carbon monoxide concentrations in air samples from this city are normally distributed? Why or why not?

(b) Find the probability that the average concentration in 100 randomly selected samples will exceed 14 ppm.

*Solution* For (a): No, it is unlikely that carbon monoxide concentrations in air samples from a large city are exactly normally distributed. Environmental measurements such as pollutant concentrations often have skewed distributions due to occasional high readings (outliers) and natural variability. However, for large sample sizes, the Central Limit Theorem allows us to approximate the sampling distribution of the sample mean by a normal distribution, even if the underlying data are not normal. For (b): We need to find $P(\overline{Y} > 14)$. Using the Central Limit Theorm we have that

$$P(\overline{Y} > 14) = P\left(Z > \frac{14-12}{9/\sqrt{100}}\right) = P(Z > 2.22) = 0.0132.$$

□

**7.55**

The downtime per day for a computing facility has mean 4 hours and standard deviation 0.8 hour.

(a) Suppose that we want to compute probabilities about the average daily downtime for a period of 30 days.

(i) What assumptions must be true to use the result of Theorem 7.4 to obtain a valid approximation for probabilities about the average daily downtime?

(ii) Under the assumptions described in part (i), what is the approximate probability that the average daily downtime for a period of 30 days is between 1 and 5 hours?

(b) Under the assumptions described in part (a), what is the approximate probability that the *total* downtime for a period of 30 days is less than 115 hours?

*Solution* For (a): To use the Central Limit Theorem what must be true is that all the computers in the facility must independent and identically distributed. Under these assumptions we are asked to find $P(1 \leq \overline{Y} \leq 5)$. We can approximate this with a Normal Distribution using the Central Limit Theorem. That is

$$P(1 \leq \overline{Y} \leq 5) = p\left(\frac{1-4}{0.8/\sqrt{30}} \leq Z \leq \frac{5-4}{0.8/\sqrt{30}}\right) = P(-20.5395 \leq Z \leq 6.847) \approx 1.$$

That is it almost certain the the average daily downtime will be between 1 and 5 hours. For (b): We now find $P(T = \sum_{i=1}^{30} X_i \leq 115)$ where $T \sim N(n\mu = 120, n\sigma^2)$. We see that

$$P(T \leq 115) = P\left(Z \leq \frac{115 - 30 \cdot 4}{0.8 \cdot \sqrt{30}}\right) = P(Z \leq -1.14) \approx 0.1271.$$

□

**7.56**

Many bulk products—such as iron ore, coal, and raw sugar—are sampled for quality by a method that requires many small samples to be taken periodically as the material is moving along a conveyor belt. The small samples are then combined and mixed to form one composite sample. Let $Y_i$ denote the volume of the $i$th small sample from a particular lot and suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample, with each $Y_i$ value having mean $\mu$ (in cubic inches) and variance $\sigma^2$. The average volume $\mu$ of the samples can be set by adjusting the size of the sampling device. Suppose that the variance $\sigma^2$ of the volumes of the samples is known to be approximately 4. The total volume of the composite sample must exceed 200 cubic inches with probability approximately 0.95 when $n = 50$ small samples are selected. Determine a setting for $\mu$ that will allow the sampling requirements to be satisfied.

*Solution* We need to find a value $\mu$ such that $P(T = \sum_{i=1}^{50} Y_i \geq 200) = 0.95$. Notice that $T \sim N(n\mu = 50\mu, n\sigma^2 = 200)$. Using the Central Limit Theorem we can approximate this probability. We see that

$$P(T = \sum_{i=1}^{50} Y_i \geq 200) = P\left(Z \geq \frac{200 - 50\mu}{\sqrt{200}}\right) = 0.95.$$

Since the table only gives us the right-hand tail we can instead find $P(Z \leq a) = 0.05$. We see that $a = -1.645$. Thus we see that

$$\frac{200 - 50\mu}{\sqrt{200}} = -1.645 \quad \text{or} \quad \mu \approx 4.465.$$

□

**7.57**

Twenty-five heat lamps are connected in a greenhouse so that when one lamp fails, another takes over immediately. (Only one lamp is turned on at any time.) The lamps operate independently, and each has a mean life of 50 hours and standard deviation of 4 hours. If the greenhouse is not checked for 1300 hours after the lamp system is turned on, what is the probability that a lamp will be burning at the end of the 1300-hour period?

*Solution* Let $T = \sum_{i=1}^{25}$ be the total time the lamps can last where $X_i$ is the total time for plan $i$. Then notice that $T \sim N(n \cdot \mu = 1250, n\sigma^2 = 400)$ by the Central Limit Theorem. Thus we have to find $P(T > 1300)$. We see that

$$P(T > 1300) = P\left(Z > \frac{1300 - 1250}{\sqrt{400}}\right) = P(Z > 2.5) = 0.0062.$$

□

**7.58**

Suppose that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are independent random samples from populations with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, respectively. Show that the random variable

$$U_n = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}$$

satisfies the conditions of Theorem 7.4 and thus that the distribution function of $U_n$ converges to a standard normal distribution function as $n \to \infty$. *[Hint: Consider $W_i = X_i - Y_i$, for $i = 1, 2, \ldots, n$.]*

*Solution* We need to show that $W = \overline{W} - \overline{Y}$ forms a sample mean from a random sample with mean $\mu_W = \mu_1 - \mu_2$ and standard deviation $\sqrt{(\sigma_1^2 + \sigma_2^2)/n}$. The first thing to notice is that

$$W = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} X_i - Y_i.$$

Since $X_i$ and $Y_i$ are from different populations they are independent. Thus then this implies that $X_i - Y_i$ are independent for all $i = 1, 2, \ldots, n$. Thus this forms a random sample that is independent and identically distributed. Next we find the mean of this random variable.

$$E(W) = E(\overline{X} - \overline{Y}) = E(\overline{X}) = E(\overline{Y}) = \mu_1 - \mu_2$$

as needed. Simiariliy we see that

$$\text{Var}(W) = \text{Var}(\overline{X} - \overline{Y}) = \text{Var}(\overline{X}) + \text{Var}(\overline{Y}) - 2\text{Cov}(\overline{X}, \overline{Y}) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n} - 2 \cdot 0 = \frac{\sigma_1^2 + \sigma_2^2}{n}.$$

Thus we can clearly see that this satisfies the condtions of the Central Limit Theorem and that $U_n$ converges to the standard normal distribution. □

**7.59**

An experiment is designed to test whether operator A or operator B gets the job of operating a new machine. Each operator is timed on 50 independent trials involving the performance of a certain task using the machine. If the sample means for the 50 trials differ by more than 1 second, the operator with the smaller mean time gets the job. Otherwise, the experiment is considered to end in a tie. If the standard deviations of times for both operators are assumed to be 2 seconds, what is the probability that operator A will get the job even though both operators have equal ability?

*Solution* Let $\overline{Y}$ be the sample mean for Operator A and $\overline{X}$ be the sample mean for Operator B with both sample sizes being $n = 50$. We are assuming that $E(\overline{Y}) = E(\overline{X}) = \mu$ and $\text{Var}(\overline{Y}) = \text{Var}(\overline{X}) = \sigma$ since both are assumed to have equal ability. We are told that if $D = \overline{Y} - \overline{X} > 1$ and $\overline{Y} < \overline{X}$ then Operator A gets the job. Notice that $E(D) = 0$ and $\text{Var}(D) = \text{Var}(\overline{Y}) + \text{Var}(\overline{X}) = 4/50 + 4/50 = 0.16$. Since Operator A has to be smaller we have that

$$P(\overline{X} - \overline{Y} > 1) = P(D < -1) = P(D > 1).$$

Converting this to standard normal distribution using the Central Limit Theorem we have that

$$P\left(Z > \frac{1 - 0}{\sqrt{0.16}}\right) = P(Z > 2.5) = 0.0062.$$

Thus the probability that Operator A will get the job is %0.62. □

**7.60**

The result in Exercise 7.58 holds even if the sample sizes differ. That is, if $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ constitute independent random samples from populations with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, respectively, then $\overline{X} - \overline{Y}$ will be approximately normally distributed, for large $n_1$ and $n_2$, with mean $\mu_1 - \mu_2$ and variance $\left(\frac{\sigma_1^2}{n_1}\right) + \left(\frac{\sigma_2^2}{n_2}\right)$.
The flow of water through soil depends on, among other things, the porosity (volume proportion of voids) of the soil. To compare two types of sandy soil, $n_1 = 50$ measurements are to be taken on the porosity of soil A and $n_2 = 100$ measurements are to be taken on soil B.

Assume that $\sigma_1^2 = 0.01$ and $\sigma_2^2 = 0.02$. Find the probability that the difference between the sample means will be within 0.05 unit of the difference between the population means $\mu_1 - \mu_2$.

*Solution* We are asked to find the probability that $P(|(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)| \leq 0.05)$. As we were told, even though the sample sizes from both populations are different, since they are large, we can approximate this probability with the Normal Distribution by using the Central Limit Theorem.Thus we convert this probability into the standard normal random variable using $\mu_1 - \mu_2$ as the mean and $\left(\frac{\sigma_1^2}{n_1}\right) + \left(\frac{\sigma_2^2}{n_2}\right)$ for the variance.

$$P(-0.05 \leq (\overline{X}-\overline{Y})-(\mu_1-\mu_2) \leq 0.05) = P\left(\frac{-0.05}{\sqrt{\left(\frac{0.01}{50}\right)+\left(\frac{0.02}{100}\right)}} \leq Z \leq \frac{0.05}{\sqrt{\left(\frac{0.01}{50}\right)+\left(\frac{0.02}{100}\right)}}\right) = P(-2.5 \leq Z \leq 2.5) =$$

□

**7.61**

Refer to Exercise 7.60. Suppose that $n_1 = n_2 = n$, and find the value of $n$ that allows the difference between the sample means to be within 0.04 unit of $\mu_1 - \mu_2$ with probability 0.90.

*Solution* We are now asked to find the required sample size $n$ such that $P(-0.04 \leq (\overline{X} - \overline{Y}) - (\mu_1 - \mu_2) \leq 0.04) = 0.90$. Again using the Central Limit Theorem we have that

$$P\left(\frac{-0.04}{\sqrt{\left(\frac{0.01}{n}\right)+\left(\frac{0.02}{n}\right)}} \leq Z \leq \frac{0.04}{\sqrt{\left(\frac{0.01}{n}\right)+\left(\frac{0.02}{n}\right)}}\right) = 0.90.$$

Simplifying we have that

$$1 - 2P\left(Z > \frac{0.04}{\sqrt{\left(\frac{0.01}{n}\right)+\left(\frac{0.02}{n}\right)}}\right) = 0.90 \quad \text{or} \quad P\left(Z > \frac{0.04}{\sqrt{\left(\frac{0.01}{n}\right)+\left(\frac{0.02}{n}\right)}}\right) = 0.05.$$

Using the standard normal table we find that

$$\frac{0.04}{\sqrt{\left(\frac{0.01}{n}\right)+\left(\frac{0.02}{n}\right)}} = 1.648 \quad \text{or} \quad n = \frac{(1.648)^2}{(0.04)^2} \cdot (0.01 + 0.02) = 50.9232$$

rounding up we have that $n = 51$. □

**7.62**

The times that a cashier spends processing individual customer's order are independent random variables with mean 2.5 minutes and standard deviation 2 minutes. What is the approximate probability that it will take more than 4 hours to process the orders of 100 people?

*Solution* Let $T = \sum_{i=1}^{100} X_i$ where $X_i$ is the time it takes to process the $i$-th person. We are asked to find $P(T > 4)$. Notice that $T \sim n\overline{Y} \sim N(n\mu = 250, n\sigma^2 = 400)$. We can then convert this into a standard normal variable to get that

$$P(T > 4) = P\left(Z > \frac{4 - 250}{\sqrt{400}}\right) = P(Z > -0.5) = 1 - P(Z > 0.5) = 0.6915.$$

□

**7.63**

Refer to Exercise 7.62. Find the number of customers $n$ such that the probability that the orders of all $n$ customers can be processed in less than 2 hours is approximately .1.

*Solution* We need to find the sample size $n$ such that $P(T \leq 2) = 0.1$. We see that

$$P(T \leq 2) = P\left(Z \leq \frac{2 - 2.5n}{\sqrt{4n}}\right) = 0.1$$

Using the standard normal table we see that

$$\frac{2 - 2.5n}{2\sqrt{n}} = -1.2816 \quad \text{or} \quad 6.25n^2 - 606.57n + 14400 = 0.$$

Using the quadratic root formula we see that $n = 41$ is the required sample size. $\square$

Note that most questions 7.70-7.87 are skipped as I found most of them to be repatative. The general idea for approximating the Binomial distribution using the Normal Distribution is the same and is something that computers can do really well. Note that most of these calculations can be performed really quickly using software such as $R$, the software I use, which makes doing these questions really fast.

# 3 Estimation

## 3.1 Introduction

As we touched upon in Chapter 2, we defined statistics to be the basis to allow us to make inferences about the population parameters from which the random sample is taken from. For example we said that a typical statistical problem might be that we know the underlying probability distribution, such as $X \sim Exp(\mu)$ but we are missing the parameters. In the previous chapter we derived the sampling distribution for many statistics which will prove to be useful for us.

Estimation has many practical applications. For example, a manufacturer of washing machines might be interested in estimating the proportion $p$ of washers that can be expected to fail prior to the expiration of a 1-year guarantee time. Other important population parameters are the population mean, variance, and standard deviation. For example, we might wish to estimate the mean waiting time $\mu$ at a supermarket checkout station or the standard deviation of the error of measurement $\sigma$ of an electronic instrument. To simplify our terminology, we will call the parameter of interest in the experiment the target parameter.

There are two main types of estimates for the target parameter: point estimates and interval estimates.

A point estimate is a single value calculated from the sample data that serves as a best guess for the target parameter. For example, the sample mean $\overline{X}$ is often used as a point estimate for the population mean $\mu$. While point estimates are useful for providing a specific value, they do not give any information about the uncertainty or variability associated with the estimate.

**Definition 3.1.1**

A point estimator of the parameter $\theta$ is a function of the underlying random variables and so it is a random variable with a distribution function. Thus, a point estimate of $\theta$ is a function of the data $Y_1, Y_2, \ldots, Y_n$ only. We say $\hat{\theta}$ is a point estimator of $\theta$.

If $y_1, y_2, \ldots, y_n$ are an observed sample then $\hat{\theta} = g(y_1, y_2, \ldots, y_n)$ is a real number.

An interval estimate, on the other hand, provides a range of values within which the target parameter is likely to lie, along with a specified level of confidence. For example, a 95% confidence interval for the population mean $\mu$ might be expressed as $(\mu_1, \mu_2)$, indicating that we are 95% confident that $\mu$ falls within this interval. Interval estimates are particularly valuable because they account for the variability inherent in the sampling process and provide a measure of reliability for the estimate.

**Definition 3.1.2**

An interval estimator of the parameter $\theta$ is a pair of functions of the underlying random variables, $(L(Y_1, Y_2, \ldots, Y_n), U(Y_1, Y_2, \ldots, Y_n))$, where $L$ and $U$ represent the lower and upper bounds of the interval, respectively. Thus, an interval estimate of $\theta$ is a pair of real numbers $(l, u)$ calculated from the observed sample $y_1, y_2, \ldots, y_n$.

The interval $(l, u)$ is interpreted as the range within which the true value of $\theta$ is likely to lie, with a specified level of confidence.

In any case, the actual estimation is accomplished by using an estimator for the target parameter.

**Definition 3.1.3**

An estimator is a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.

As we already see, many different types of estimators can exist for a parameter. Thus this leads us to the fact that then there must be good estimators and bad estimators. We develop this idea in the next section.

## 3.2 The Bias and Mean Square Error of Point Estimators

To begin this section I will use an intutive example from the textbook: Point estimation can be thought of as trying to hit a target with a dart. The estimator is like the dart thrower, the estimate is the dart itself, and the parameter of interest is the bull's-eye on the target. When we draw a single sample from the population and use it to compute an estimate for the parameter, it is akin to throwing a single dart at the bull's-eye.

Imagine a person throws a dart and it lands exactly on the bull's-eye. Would we immediately conclude that they are an expert dart thrower? Likely not, as one successful throw does not provide enough evidence of skill. However, if the person consistently hits the bull's-eye with many throws, we might gain confidence in their ability. Similarly, we cannot judge the quality of a point estimation procedure based on a single estimate. Instead, we need to evaluate the procedure by observing the estimates it produces over repeated sampling.

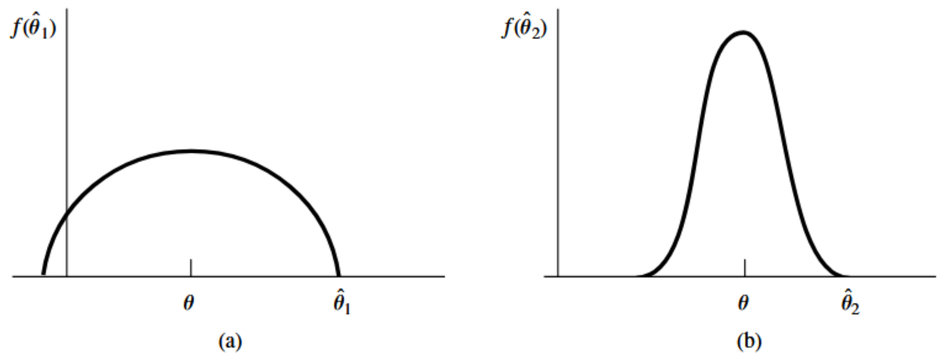


Figure 3: Sampling distributions for two unbiased estimators: (a) estimator with large variation; (b) estimator with small variation

To assess the goodness of a point estimator, we examine the distribution of estimates obtained from repeated samples. If this distribution clusters closely around the target parameter, we can consider the estimator to be reliable.

Recall that we call an denote an estiamtor of $\theta$ by $\hat{\theta}$. The parameter below the "hat" is what we are trying to estimate. Going back to the dart-throwing example, if we want the estimator's sample distribution to cluster around the target parameter, we would like the mean of our sampling distribution to be the target parameter. That is $E(\hat{\theta}) = \theta$. Point estimators that satisfy this property are said to be unbiased.

**Definition 3.2.1**

Let $\hat{\theta}$ be a point estimator for a parameter $\theta$. Then $\hat{\theta}$ is an unbiased estimator if $E(\hat{\theta}) = \theta$. If $E(\hat{\theta}) \neq \theta$, then $\hat{\theta}$ is said to be biased.

**Definition 3.2.2**

The bias of a point estimator $\hat{\theta}$ is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Figure 8.3 illustrates two potential sampling distributions for unbiased point estimators of a target parameter $\theta$. Among these, the distribution depicted in Figure 8.3(b) is preferable due to its smaller variance. A lower variance ensures that, across repeated sampling, a greater proportion of the values of $\hat{\theta}_2$ will be closer to $\theta$. Consequently, while unbiasedness is an essential property of an estimator, minimizing the variance of its sampling distribution, $V(\hat{\theta})$, is equally important. When comparing two unbiased estimators for a parameter $\theta$, assuming all other factors are equal, the estimator with the smaller variance is the better choice.

The mean squared error (MSE) of an estimator combines both bias and variance into a single measure of the estimator's quality. Intuitively, the MSE can be thought of as the average squared distance between the estimator's predicted values and the true parameter value. It is given by:

**3.2.3**

The mean squared error of a point estimator $\hat{\theta}$ is

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

Expanding this expression, we can see that the MSE is composed of two components:

$$\text{MSE}(\hat{\theta}) = \text{Bias}^2(\hat{\theta}) + V(\hat{\theta})$$

This decomposition shows that the MSE accounts for both the bias of the estimator (how far the expected value of the estimator is from the true parameter) and the variance of the estimator (how spread out the sampling distribution of the estimator is).

Returning to the dart-throwing analogy, the MSE can be thought of as a measure of how consistently the darts land close to the bull's-eye. If the dart thrower consistently hits the bull's-eye (unbiased estimator) but the darts are scattered widely (high variance), the MSE will still be large. Conversely, if the darts are tightly clustered but consistently miss the bull's-eye (biased estimator), the MSE will also be large. The ideal scenario is for the darts to be tightly clustered around the bull's-eye, minimizing both bias and variance, which results in a low MSE.

Thus, when evaluating point estimators, minimizing the MSE is a desirable goal, as it ensures both accuracy (low bias) and precision (low variance). We go over some common unbiased point estiamtors in the next section.

**Example 3.2.3**

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

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

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

*Solution* For (a): By defintion 3.2.2 we have that

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

For (b): To find a function $\hat{\theta}^*$ such that $\hat{\theta}$ is unbiased we need to construct it so that $E(\hat{\theta}) - \theta$. Obviously what we see is that $\hat{\theta}^*(\hat{\theta}) = 1/a(\hat{\theta} - b)$ satisfies these condtions. To see this we compute the expected value of $\hat{\theta}^*$.

$$\begin{aligned} E(\hat{\theta}^*) &= E(1/a(\hat{\theta} - b)) \\ &= \frac{1}{a}\left(E(\hat{\theta}) - b\right) \\ &= \frac{1}{a}(a\theta + b - b) = \theta \end{aligned}$$

as required. $\square$

**Example 3.2.4**

Suppose that $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$, $\text{Var}(\hat{\theta}_1) = \sigma_1^2$ and $\text{Var}(\hat{\theta}_2) = \sigma_2^2$. 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 $\theta$.

(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 $\hat{\theta}_3$?

*Solution* For (a): We begin by showing that $E(\hat{\theta}_3) = \theta$.

$$\begin{aligned} E(\hat{\theta}_3) &= E(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) \\ &= aE(\hat{\theta}_1) + (1-a)E(\hat{\theta}_2) \\ &= a\theta + (1-a)\theta & = \theta. \end{aligned}$$

For (b): We need to find a value for $a$ such that the variance of the estiamtor $\hat{\theta}_3$ is minimized. We begin by first finding the variance of this estiamtor.

$$\begin{aligned} \text{Var}(\hat{\theta}_3) &= \text{Var}(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) \\ &= a^2\text{Var}(\hat{\theta}_1) + (1-a)^2\text{Var}(\hat{\theta}_2) \\ &= a^2\sigma_1^2 + (1-a)^2\sigma_2^2 = V(a). \end{aligned}$$

Since this is a function of $a$ and is convex we can use calculus to find the global minimum of this function to minimize the variance. We see that

$$V'(a) = 2a\sigma_1^2 - 2\sigma_2^2(1-a).$$

Setting this equal to zero and solving for $a$ we see that the value of $a$ that minimizes the variance is

$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

□

**Example 3.2.5**

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

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

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 $\theta$ and find $\text{MSE}(\hat{\theta})$.

*Solution* We need to show that $E(\hat{\theta}) = \theta$ and then compute the MSE of $\hat{\theta}$. We see that $E(\hat{\theta}) = nE(Y_{(1)})$ so we compute the mean of the smallest-order statistic. To do this we first find the pdf. Recall that the pdf of the smallest-order statistic $Y_{(1)}$ is given by:

$$f_{Y_{(1)}}(y) = nf_Y(y)\left[1 - F_Y(y)\right]^{n-1},$$

where $f_Y(y) = \frac{1}{\theta}e^{-y/\theta}$ and $F_Y(y) = 1 - e^{-y/\theta}$ are the pdf and cdf of the exponential distribution, respectively.

Substituting these expressions, we get:

$$f_{Y_{(1)}}(y) = n\frac{1}{\theta}e^{-y/\theta}\left(e^{-y/\theta}\right)^{n-1} = n\frac{1}{\theta}e^{-ny/\theta}.$$

Notice that then $Y_{(1)} \sim Exp(\theta/n)$ and so $E(\hat{\theta}) = \theta$. This shows that $\hat{\theta} = nY_{(1)}$ is an unbiased estimator for $\theta$. Next, we compute $\text{MSE}(\hat{\theta})$. Recall:

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

Since $\hat{\theta}$ is unbiased, $\text{Bias}^2(\hat{\theta}) = 0$, so:

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}).$$

We compute $\text{Var}(\hat{\theta})$:

$$\text{Var}(\hat{\theta}) = n^2\text{Var}(Y_{(1)}).$$

For the exponential distribution, the variance of the smallest-order statistic $Y_{(1)}$ is:

$$\text{Var}(Y_{(1)}) = \frac{\theta^2}{n^2}.$$

Thus:

$$\text{Var}(\hat{\theta}) = n^2\frac{\theta^2}{n^2} = \theta^2.$$

Therefore:

$$\text{MSE}(\hat{\theta}) = \theta^2.$$

□

## 3.3 Some Common Unbiased Point Estimators

In this section, we discuss some commonly used unbiased point estimators, including the sample mean difference and the sample proportion difference.

The sample mean difference is used to estimate the difference between two population means, $\mu_1 - \mu_2$. Suppose we have two independent random samples, $Y_1, Y_2, \ldots, Y_{n_1}$ from population 1 and $X_1, X_2, \ldots, X_{n_2}$ from population 2. The sample mean difference is given by:

$$\hat{\Delta}_\mu = \overline{Y} - \overline{X},$$

where $\overline{Y}$ and $\overline{X}$ are the sample means of the two populations.

The expected value and variance of $\hat{\Delta}_\mu$ are:

$$E(\hat{\Delta}_\mu) = \mu_1 - \mu_2,$$

$$\text{Var}(\hat{\Delta}_\mu) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Thus, $\hat{\Delta}_\mu$ is an unbiased estimator for $\mu_1 - \mu_2$.

The sample proportion difference is used to estimate the difference between two binomial proportions, $p_1 - p_2$. Suppose $Y \sim \text{Bin}(n_1, p_1)$ and $X \sim \text{Bin}(n_2, p_2)$. The sample proportions are:

$$\hat{p}_1 = \frac{Y}{n_1}, \quad \hat{p}_2 = \frac{X}{n_2}.$$

To see how this is an unbiased estiamtor of the parameter $p$ notice that $E(\hat{p}) = E(Y/n) = nE(Y) = \frac{np}{n} = p$.

The sample proportion difference is given by:

$$\hat{\Delta}_p = \hat{p}_1 - \hat{p}_2.$$

The expected value and variance of $\hat{\Delta}_p$ are:

$$E(\hat{\Delta}_p) = p_1 - p_2,$$

$$\text{Var}(\hat{\Delta}_p) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$

Thus, $\hat{\Delta}_p$ is an unbiased estimator for $p_1 - p_2$.

These estimators are widely used in statistical inference and provide a foundation for comparing populations based on sample data. Note that we denote $\sigma^2_{\hat{\theta}}$ as the standard deviation for the sampling distribution of an estimator. Furthermore $\sigma_{\hat{\theta}} = \sqrt{\sigma^2_{\hat{\theta}}}$ is reffered to as the standard error. We summarize these common point estiamtors in the following table.

| **Target Parameter** $\theta$ | **Sample Size(s)** | **Point Estimator** $\hat{\theta}$ | $E(\hat{\theta})$ | **Standard Error** $\sigma_{\hat{\theta}}$ |
|---|---|---|---|---|
| $\mu$ | $n$ | $\overline{Y}$ | $\mu$ | $\frac{\sigma}{\sqrt{n}}$ |
| $p$ | $n$ | $\hat{p} = \frac{Y}{n}$ | $p$ | $\sqrt{\frac{pq}{n}}$ |
| $\mu_1 - \mu_2$ | $n_1$ and $n_2$ | $\overline{Y}_1 - \overline{Y}_2$ | $\mu_1 - \mu_2$ | $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ |
| $p_1 - p_2$ | $n_1$ and $n_2$ | $\hat{p}_1 - \hat{p}_2$ | $p_1 - p_2$ | $\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$ |

Table 1: Summary of Point Estimators and Their Properties

In conclusion, while unbiasedness and low variance are critical properties of a good point estimator, they alone may not fully capture the "goodness" of an estimation procedure. For instance, an estimator with low bias but high variance might still lead to unreliable predictions, and vice versa. This raises an intriguing question: how can we systematically evaluate the overall quality of a point estimator? In the next section, we delve deeper into this topic, exploring criteria and methods that help us assess and compare estimators in a more comprehensive manner.

## 3.4 Evaluating the Goodness of a Point Estimator

One natural way to measure the goodness of a point esitmator might be to look at the distance between the estimation producted by the esitmator and the target parameter.

**Definition 3.4.1**

The error of estimation $\epsilon$ is the distance between an estimator and its target parameter. That is, $\epsilon = |\hat{\theta} - \theta|$.

Since $\hat{\theta}$ is a random variable then so too is the error of estimation. Thus this means we cannot make any statements on how large or small this error will be for any particular estimate made by $\hat{\theta}$. However what we can do is examine the probability that this error is within some certain margin $b$. That is

$$P(|\hat{\theta} - \theta| < b) = P(-b < \hat{\theta} - \theta < b) = P(\theta - b < \hat{\theta} < b + \theta).$$

That is we are finding the probability that $\hat{\theta}$ is within $b$ units of the target parameter. Visually this is the area under the sampling distribution of $\hat{\theta}$ over the invercal $(\theta - b, \theta + b)$. And if the estimator is unbiased then the distribution will be cenetered around $\theta$. If from a practical point of view $b$ can be viewed as small then $P(\epsilon < b)$ gives us a measure of goodness of a single estimate. Looking at Figure 3 (a) we choose some $b$ such that $\theta - b$ and $\theta + b$ are at the tails of the pdf then we see that $P(\epsilon < b)$ will be high. If we now consider (b) we see that we can allow for the value of $b$ to be smaller and have still have a higher probability, which helps us say that (b) is a better point estiamtor than (a).

Whether we know the sampling distribution of our point estimator or not, if it is unbiased, we can use Chebyshev's inequality to find an upper bound for $b$ regardless of the distribution. Chebyshev's inequality states that for any random variable $X$ with mean $\mu$ and variance $\sigma^2$, the probability that $X$ deviates from its mean by more than $k\sigma$ is at most $\frac{1}{k^2}$. Applying this to our unbiased point estimator $\hat{\theta}$, we have:

$$P(|\hat{\theta} - \theta| \geq b) \leq \frac{\text{Var}(\hat{\theta})}{b^2}.$$

Thus, the probability that the error of estimation $\epsilon$ is within $b$ units of the target parameter is:

$$P(|\hat{\theta} - \theta| < b) \geq 1 - \frac{\text{Var}(\hat{\theta})}{b^2}.$$

This inequality provides a useful bound for evaluating the goodness of an unbiased estimator, even when the exact sampling distribution is unknown. It highlights the importance of minimizing the variance of the estimator to ensure that the probability of the error being within a small margin is high. This approach is particularly valuable when comparing estimators or selecting the most appropriate one for a given statistical problem.

If we let $b = 2\sigma_{\hat{\theta}}$, then using Chebyshev's inequality, the probability that the error of estimation $\epsilon$ is within $b$ units of the target parameter is at least:

$$P(|\hat{\theta} - \theta| < b) \geq 1 - \frac{\text{Var}(\hat{\theta})}{b^2}.$$

Substituting $b = 2\sigma_{\hat{\theta}}$ and $\text{Var}(\hat{\theta}) = \sigma_{\hat{\theta}}^2$, we get:

$$P(|\hat{\theta} - \theta| < 2\sigma_{\hat{\theta}}) \geq 1 - \frac{\sigma_{\hat{\theta}}^2}{(2\sigma_{\hat{\theta}})^2} = 1 - \frac{1}{4} = 0.75.$$

Thus, the probability that the error is within $2\sigma_{\hat{\theta}}$ is at least 0.75, providing a practical measure of the estimator's reliability. This is an often good choice since many distribtions in nature lie within 2 standard deviations from their mean with a probability of 0.95.

**Example 3.4.2**

If $Y_1, Y_2, \dots, Y_n$ denote a random sample from an exponential distribution with mean $\theta$, then $E(Y_i) = \theta$ and $\text{Var}(Y_i) = \theta^2$. Suggest an unbiased estimator for $\theta$ and provide an estimate for the standard error of your estimator.

*Solution* We can use the sample mean $\overline{Y}$ to be an unbiaseded estimator for $\theta$. Then we see that $E(\overline{Y}) = \theta$ and $\text{Var}(\overline{Y}) = \theta^2/n$. Thus we see that our standard error for this estiamtor is

$$\sigma_{\overline{Y}} = \frac{\theta}{\sqrt{n}}.$$

We can approximate this by using the sample mean and so

$$\sigma_{\overline{Y}} = \frac{\overline{Y}}{\sqrt{n}}.\text{''}$$

$\square$

**Example 3.4.3**

An engineer observes $n = 10$ independent length-of-life measurements on a type of electronic component. The average of these 10 measurements is 1020 hours. If these lengths of life come from an exponential distribution with mean $\theta$, estimate $\theta$ and place a 2-standard-error bound on the error of estimation.

*Solution* We are asked to esitmate the mean $\theta$ and then place a 2-standard-error bound on the error of estimation. We except this probability to be high. First we see that the sample mean is an ubiased estimator of the mean. That is

$$\overline{Y} = \frac{1}{10}\sum_{i=1}^{10} Y_i$$

where each $Y_i \sim Exp(\theta)$. We are given that $\hat{\theta} = \overline{Y} = 1020$. Thus then the standard error is $\sigma_{\overline{Y}} = \overline{Y}/\sqrt{10} \approx 322.61$. Thus a 2-standard-error bound on the error of estimation is $2 \cdot \sigma_{\overline{Y}} = 645.22$. Thus our estimate is 1020 and or 2-standard-error bound is $1020 \pm 645.22$. This is really high so this engineer sucks. $\square$

**Example 3.4.4**

A sample of $n = 1000$ voters, randomly selected from a city, showed $y = 560$ in favor of candidate Jones. Estimate $p$, the fraction of voters in the population favoring Jones, and place a 2-standard-error bound on the error of estimation.

*Solution* We are asked to estimate $p$ the sample proportion. We can do this with $\hat{p} = Y/n$. The estimate of $p$ we have is

$$\hat{p} = \frac{Y}{n} = \frac{560}{1000} = 0.56.$$

We know that the sampling distribution of $\hat{p}$ will be approximately normal because of the large sample size $n = 1000$. Thus we can say that the probability that $P(\epsilon < b) \approx 0.95$. To find this bound we see that

$$b = 2\sigma_{\hat{p}} = 2 \cdot \sqrt{\frac{pq}{n}} \approx 2 \cdot \sqrt{\frac{(0.56)(0.44)}{1000}} = 0.03.$$

The probability that the error of estimation is less than .03 is approximately .95. Consequently, we can be reasonably confident that our estimate, .56, is within .03 of the true value of $p$, the proportion of voters in the population who favor Jones. □

**Example 3.4.5**

A comparison of the durability of two types of automobile tires was obtained by road testing samples of $n_1 = n_2 = 100$ tires of each type. The number of miles until wear-out was recorded, where wear-out was defined as the number of miles until the amount of remaining tread reached a prespecified small value. The measurements for the two types of tires were obtained independently, and the following means and variances were computed:

$$\overline{y}_1 = 26,400 \text{ miles}, \quad \overline{y}_2 = 25,100 \text{ miles},$$
$$s_1^2 = 1,440,000, \quad s_2^2 = 1,960,000.$$

Estimate the difference in mean miles to wear-out and place a 2-standard-error bound on the error of estimation.

*Solution* We are asked to estimate $\mu_1 - mu_2$ and place a 2-standard-error bound on the error of estimation. We can use the sample mean difference $\overline{Y_1} - \overline{Y_2}$ to estimate this. We see that then the point estimate is

$$(\overline{Y_1} - \overline{Y_2}) = 26400 - 25100 = 1300.$$

Then we see that

$$b = 2\sigma_{(\overline{Y_1} - \overline{Y_2})} = 2 \cdot \sqrt{\frac{\sigma_1^2}{100} + \frac{\sigma_2^2}{100}} \approx 2 \cdot \sqrt{\frac{1440000}{100} + \frac{1960000}{100}} = 368.7817.$$

Thus we estimate the difference in mean to be 1300 and expect the error of estimation to be within 368.8 miles with a probability of 0.95. □

We now end this section by asking: Is unbiased estimator always good?

**Example 3.4.6**

If $Y \sim \text{Poisson}(\theta)$. Show that $(-1)^Y$ is an unbiased estimator of $e^{-2\theta}$. Explain why this estimator is not acceptable.

*Solution* We first show that $E[(-1)^Y] = (-1)^Y$ is an unbiased estiamtor of $e^{-2\theta}$.

$$\begin{aligned} E[(-1)^Y] &= \sum_{i=1}^{\infty} (-1)^y P(Y = y) \\ &= \sum_{i=1}^{\infty} (-1)^y \cdot \frac{\theta^y e^{-\theta}}{y!} \\ &= e^{-\theta} \sum_{i=1}^{\infty} \frac{(-\theta)^y}{y!}. \end{aligned}$$

Notice that the series is the taylor series for $e^y$ centered at $-\theta$. Thus then it converges or $e^{-\theta}$. So

$$E[(-1)^Y] = e^{-2\theta}.$$

Although $(-1)^Y$ is an unbiased estimator for $e^{-2\theta}$, it is not acceptable for several reasons. First, it is highly unstable and lacks practical utility. The values of $(-1)^Y$ oscillate between $-1$ and 1 depending on whether $Y$ is odd or even, which makes the estimator extremely variable. This variability renders it unreliable for estimating $e^{-2\theta}$ in practice, as it does not provide a meaningful or consistent measure of the target parameter.

Second, $(-1)^Y$ is not in the parameter space of $e^{-2\theta}$. The parameter $e^{-2\theta}$ is strictly positive for all $\theta > 0$, whereas $(-1)^Y$ takes values in $\{-1, 1\}$. This discrepancy means that $(-1)^Y$ cannot produce estimates that are consistent with the range of possible values for $e^{-2\theta}$. A good estimator should not only be unbiased but also have low variance and produce estimates within the parameter space, which $(-1)^Y$ fails to achieve. $\square$

## 3.5 Confidence Intervals

We now talk about interval estimators which we introduced in section 3.1 (recall Definition 3.1.3). As we said, for an interval estimate we find some rules (functions) of the underlying random observations to produce two numbers to form an interval $(l, u)$ where $L(Y_1, Y_2, \ldots, Y_n) = l$ and $U(Y_1, Y_2, \ldots, Y_n) = u$. This interval is interpreted as the range within the true value of the target parameter $\theta$ lies in with some specified level of confidence. Since the endpoints are functions of random variables then the endpoints themselves will vary from sample to sample so this means we cannot guarantee that the target parameter $\theta$ lies in the interval for any particular sample. However we can generate a probability that it does. Interval estimators are commonly called confidence intervals. The upper and lower endpoints of a confidence interval are called the upper and lower confidence limits, respectively. And the probability that the interval will contain the target parameter $\theta$ is called the confidence coefficient denoted by $1 - \alpha$.

Suppose $\hat{\theta_L}$ and $\hat{\theta_U}$ are the random lower and upper confidence limits for the target parameter $\theta$ respectively. Then if

$$P(\hat{\theta_L} \leq \theta \leq \hat{\theta_U}) = 1 - \alpha,$$

the probability $1 - \alpha$ is called the confidence coefficient and the resulting interval $[\hat{\theta_L}, \hat{\theta_U}]$ is a two-sided confidence interval. We then can form a one-sided confidence interval.

$$P(\hat{\theta_L} \leq \theta) = 1 - \alpha$$

is the confidence interval $[\hat{\theta_L}, \infty)$. Similarly we can have an upper one-sided confidence interval $P(\theta \leq \hat{\theta_U})$ which gives us the interval $(-\infty, \hat{\theta_U}]$.

### 3.5.1 Pivotal Method

Let $Y_1, Y_2, \ldots, Y_n$ be a ranom sample and we observe the data $y_1, y_2, \ldots, y_n$. We want to find a confidence interval $[(\hat{\theta_L}, (\hat{\theta_U}]$ such that $P(\hat{\theta_L} \leq \theta \leq (\hat{\theta_U}) = 1 - \alpha$. One way of getting this confidence interval is the Pivotal Method. We need to find a random variable, called a pivot, that is typically a function of the estimator of $\theta$ satisfying the following:

1. It is a function of the sample measurements and the unknown parameter $\theta$, where $\theta$ is the only unknown quantity.

2. Its probability distribution does not depend on the parameter $\theta$.

If the distribution for the Pivot is known then what we can do is find any constants $a, b$ such that $P(a \leq Y \leq b) = 1 - \alpha$ where $Y$ is the random variable for the pivot. Then using the fact that

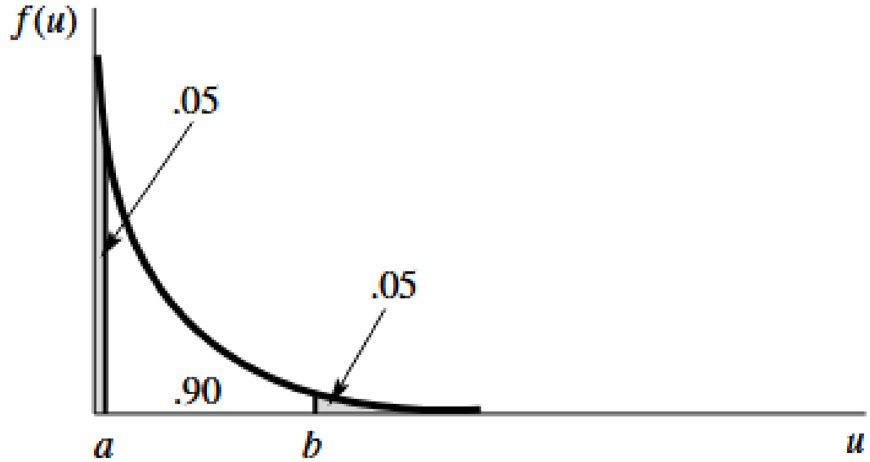


Figure 4: Illustration of the pivotal method for Example 3.5.1

$P(ca \leq cY \leq cB) = 1 - \alpha$ and $P(a + d \leq Y + d \leq b + d) = 1 - \alpha$, we try to express it so that we get $P(L(Y_1, Y_2 \ldots, Y_n) \leq \theta \leq U(Y_1, Y_2 \ldots, Y_n))$. We illustrate this in the following examples.

**Example 3.5.1**

Suppose that we are to obtain a single observation Y from an exponential distribution with mean $\theta$. Use Y to form a confidence interval for $\theta$ with confidence coefficient .90.

*Solution* The pdf of $Y$ is given by $f_Y(y) = \frac{1}{\theta}e^{-y/\theta}$ for $y \geq 0$. Let $U = Y/\theta$. Then we see that $U$ has the pdf $f_U(u) = e^{-u}$ for $u > 0$. Since $U$ is a function on th sample measurement $Y$ and the distribution does not depend on $\theta$ we can use it as the Pivot. Since we want our confidence coefficient to be 0.90 we find two constants $a, b$ such that $P(a \leq U \leq b) = 0.90$. To find this we can find $P(U < a) = \int_0^a e^{-u} du = 0.5$ and $P(U > b) = \int_b^\infty e^{-u} du = 0.5$. Solving for these values we see $a = .051$, $b = 2.996$. Thus it follows that

$$P(0.051 \leq U \leq 2.996) = 0.9 = P\left(0.051 \leq \frac{Y}{\theta} \leq 2.996\right).$$

We then manipulate the ineqaulites to get $\theta$ in the middle.

$$0.90 = P\left(0.051 \leq \frac{Y}{\theta} \leq 2.996\right) = P\left(\frac{0.051}{Y} \leq \frac{1}{\theta} \leq \frac{2.996}{Y}\right).$$

Taking the reciprocal we see that

$$P\left(\frac{Y}{2.996} \leq \theta \leq \frac{Y}{0.051}\right) = 0.90.$$

Thus, we see that $Y/2.996$ and $Y/0.051$ form the desired lower and upper confidence limits, respectively. To obtain numerical values for these limits, we must observe an actual value for Y and substitute that value into the given formulas for the confidence limits. We know that the interval $(Y/2.996, Y/0.051)$ will contain the true value of $\theta$ with probability 0.90. This method illustrates how the pivotal approach can be used to construct confidence intervals based on the known distribution of a pivot variable. $\square$

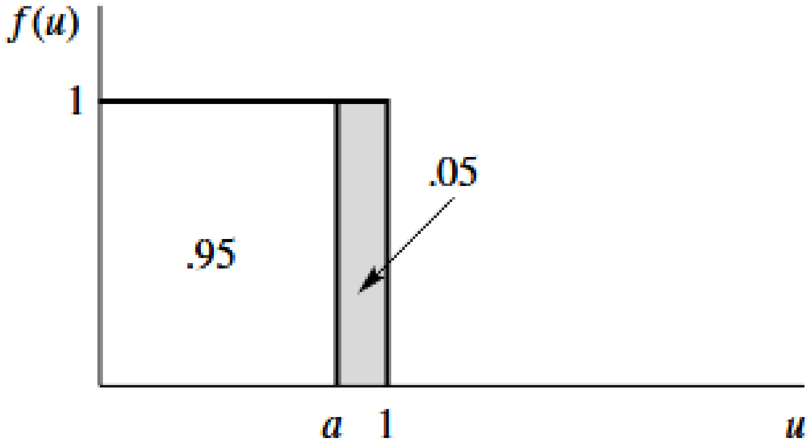


Figure 5: Illustration of the pivotal method for Example 3.5.2

**Example 3.5.2**

Suppose that we take a sample of size $n = 1$ from a uniform distribution defined on the interval $[0, \theta]$, where $\theta$ is unknown. Find a 95% lower confidence bound for $\theta$.

*Solution* We are asked to find $P(U \leq a) = 0.95$ where $U = Y/\theta$ where $Y \sim Uniform(\theta)$. Then we see that $U$ is uniform on the interval $[0, 1]$. We then see that

$$P(U \leq a) = \int_0^a 1du = a = 0.95.$$

Thus

$$P\left(\frac{Y}{\theta} \leq 0.95\right) = P\left(\frac{Y}{0.95} \leq \theta\right) = 0.95.$$

We see that $Y/0.95$ is a lower confidence limit for $\theta$, with confidence coefficient .95. $\square$

**Example 3.5.3**

Assume that $Y_1, Y_2, \ldots, Y_n$ is a sample of size $n$ from an exponential distribution with mean $\theta$.

(a) Show that

$$2\sum_{i=1}^{n} \frac{Y_i}{\theta}$$

is a pivotal quantity with a $\chi^2$ distribution with $2n$ df.

(b) Use the pivotal quantity $2\sum_{i=1}^{n} \frac{Y_i}{\theta}$ to derive a 95% confidence interval for $\theta$.

(c) If a sample of size $n = 7$ yields $\overline{y} = 4.77$, use the result from part (b) to give a 95% confidence interval for $\theta$.

*Solution* For (a): We need to show that $2\sum_{i=1}^n \frac{Y_i}{\theta}$ is a pivotal quantity. The first condtional is trivally statisfied since it is a function of the random sample and the unknown parameter $\theta$. What is next is to show that the distribution does not depend on $\theta$. The first thing to notice is that the sum of independent exponential distribtions forms a gamma distribution. That is $\sum_{i=1}^n Y_i \sim Gamma(n,\theta)$. This in turn implies that

$$\frac{2}{\theta}\sum_{i=1}^n Y_i \sim Gamma(n, \frac{2}{\theta}\cdot\theta) = Gamma(n,2) = Gamma(2n/2,2) \sim \chi^2_{(2n)}.$$

Notice that the distribution is free of $\theta$ and so we can use this as a pivotal quantity. For (b): We now use this to find a two-sided confidence interval $[\hat{\theta}_L, \hat{\theta}_U]$ such that $P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 0.95$. We begin by finding $P(a \leq \frac{2}{\theta}\sum_{i=1}^n Y_i \leq b) = 0.95$. Let $\chi^2_{1-\frac{\alpha}{2}}$ and $\chi^2_{\frac{\alpha}{2}}$ be the upper and lower critcal values (this isn't properly taught in STA256 and it is not in Chapter 8 of the textbook so you are expected to learn this yourself). Then

$$P\left(\chi^2_{\frac{\alpha}{2}} \leq \frac{2}{\theta}\sum_{i=1}^n Y_i \leq \chi^2_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Solving for $\theta$ we get that

$$P\left(\frac{2\sum_{i=1}^n Y_i}{\chi^2_{1-\frac{\alpha}{2}}} \leq \theta \leq \frac{2\sum_{i=1}^n Y_i}{\chi^2_{\frac{\alpha}{2}}}\right) = 1-\alpha.$$

Thus the confidence interval is then $\left[\frac{2\sum_{i=1}^n Y_i}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{2\sum_{i=1}^n Y_i}{\chi^2_{\frac{\alpha}{2}}}\right]$. For (c): First notice that $2\sum_{i=1}^n Y_i = 2n\overline{Y}$. Given that $n=7$ yields $\overline{Y} = 4.77$ we are asked to find the actual bounds for the confidence interval. Plugging everything in we see that

$$P\left(\frac{2\cdot 7\cdot 4.77}{\chi^2_{1-\frac{\alpha}{2}}} \leq \theta \leq \frac{2\cdot 7\cdot 4.77}{\chi^2_{\frac{\alpha}{2}}}\right) = 1-\alpha.$$

We then find $\chi^2_{1-\alpha/2} = \chi^2_{0.975}$. That is $P(\chi^2_{(14)} \leq \chi^2_{0.975}) = 0.975$. We see that $\chi^2_{0.975} \approx 5.629$. Simiarily we find that $\chi^2_{0.025} \approx 26.119$. Putting everyhing together we see that

$$\left(\frac{2\cdot 7\cdot 4.77}{26.119}, \frac{2\cdot 7\cdot 4.77}{5.629}\right) = (2.56, 11.87).$$

□

**Example 3.5.4**

Assume that $Y_1, Y_2, \ldots, Y_n$ is a sample of size $n$ from a gamma-distributed population with $\alpha = 2$ and unknown $\beta$.

(a) Show that
$$2\sum_{i=1}^{n} \frac{Y_i}{\beta}$$
is a pivotal quantity with a $\chi^2$ distribution with $4n$ df.

(b) Use the pivotal quantity $2\sum_{i=1}^{n} \frac{Y_i}{\beta}$ to derive a 95% confidence interval for $\beta$.

(c) If a sample of size $n = 5$ yields $\overline{y} = 5.39$, use the result from part (b) to give a 95% confidence interval for $\beta$.

*Solution* For (a): We need to show that $2\sum_{i=1}^{n} \frac{Y_i}{\beta}$ is pivotal quantity. The first condtional is trivally statisfied since it is a function of the random sample and the unknown parameter $\beta$. What is next is to show that the distribution does not depend on $\beta$. The first fact we need is that a sum of gamma distributions is still a gamma distribution. That is $\sum_{i=1}^{n} Y_i \sim Gamma(2n\beta)$. This in turn implies that

$$\frac{2}{\beta}\sum_{i=1}^{n} Y_i \sim Gamma(2n, \left(\frac{2}{\beta}\right)\beta) = Gamma(2n, 2) = Gamma\left(\frac{4n}{2}, 2\right) \sim \chi^2_{(4n)}.$$

Notice that the distribution is free of $\beta$ and so we can use this as a pivotal quantity. For (b): We now use this to find a two-sided confidence interval $[\hat{\beta}_L, \hat{\beta}_U]$ such that $P(\hat{\beta}_L \leq \beta \leq \hat{\beta}_U) = 0.95$. We begin by finding $P(a \leq \frac{2}{\beta}\sum_{i=1}^{n} Y_i \leq b) = 0.95$. Let $\chi^2_{1-\frac{\alpha}{2}}$ and $\chi^2_{\frac{\alpha}{2}}$ be the upper and lower crtical values. Then

$$P\left(\chi^2_{\frac{\alpha}{2}} \leq \frac{2}{\beta}\sum_{i=1}^{n} Y_i \leq \chi^2_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Solving for $\beta$ we get that

$$P\left(\frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{1-\frac{\alpha}{2}}} \leq \beta \leq \frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{\frac{\alpha}{2}}}\right) = 1 - \alpha.$$

Thus the confidence interval is then $\left[\frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{\frac{\alpha}{2}}}\right]$. For (c): First notice that $2\sum_{i=1}^{n} Y_i = 2n\overline{Y}$. Given that $n = 5$ yields $\overline{Y} = 5.39$ we are asked to find the actual bounds for the confidence interval. Plugging everything in we see that

$$P\left(\frac{2 \cdot 5 \cdot 5.39}{\chi^2_{1-\frac{\alpha}{2}}} \leq \beta \leq \frac{2 \cdot 5 \cdot 5.39}{\chi^2_{\frac{\alpha}{2}}}\right) = 1 - \alpha.$$

We then find $\chi^2_{1-\frac{\alpha}{2}]} = \chi^2_{0.975}$. That is $P(\chi^2_{(20)} \leq \chi^2_{0.975}) = 0.975$. We see that $\chi^2_{0.975} \approx 9.591$. Similarly we find that $\chi^2_{0.025} \approx 34.170$. Putting everything together we see that

$$\left[\frac{2 \cdot 5 \cdot 5.39}{34.170}, \frac{2 \cdot 5 \cdot 5.39}{9.591}\right] = [1.577, 5.621].$$

□

One thing to note is that pivotal quantities are not unique. For example, let $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$. Two pivotal quantities are:

1. $Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, free from $\mu$.

2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$, free from $\sigma^2$.

Both are valid and can be used for constructing confidence intervals.

These last few examples illustrate the pivotal method for finding confidence intervals. We develop this further in conjunction with the distributions we learned last chapter $(t, F, N)$.

## 3.6 Large-Sample Confidence Intervals

In the previous section we said that the sampling distribution for the estimators of the parameters $\mu, p, \mu_1 - \mu_2, p_1 - p_2$ are approximately normal for large sample sizes with standard errors given in Table 1. That is we can say for large enough sample sizes

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}},$$

has a standard normal distribution. Thus we can say that $Z$ forms a pivotal quantity and so the pivotal method can be used to construct confidence intervals for the target parameter $\theta$.

**Example 3.6.1**

Let $\hat{\theta}$ be a statistic that is normally distributed with mean $\theta$ and standard error $\sigma_{\hat{\theta}}$ . Find a confidence interval for $\theta$ that possesses a confidence coefficient equal to $(1-\alpha)$.

*Solution* Since $\hat{\theta}$ is normally distributed we can standardize this to get

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}.$$

Then we choose two values at the tails of this distribution $-z_{\frac{\alpha}{2}}$ and $z_{\frac{\alpha}{2}}$ such that

$$P(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}) = 1 - \alpha.$$

We then get that

$$P(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}) = P(-z_{\frac{\alpha}{2}} \le \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}]}} \le z_{\frac{\alpha}{2}}).$$

Solving for $\theta$ in the middle we find that

$$P(\hat{\theta} - z_{\frac{\alpha}{2}}\sigma_{\hat{\theta}} \le \theta \le \hat{\theta} + z_{\frac{\alpha}{2}}\sigma_{\hat{\theta}}) = 1 - \alpha.$$

Thus, the endpoints for a $100(1-\alpha)\%$ confidence interval for $\theta$ are given by:

$$\hat{\theta}_L = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \quad \text{and} \quad \hat{\theta}_U = \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}.$$

□

In general we have that the endpoints of a confidence interval are

$$\hat{\theta}_{L/U} = \text{point estimator} \pm \text{cut off (crtical value)} \cdot \text{standard error}.$$

**Example 3.6.2**

Is America's romance with movies on the wane? In a Gallup Poll 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 $p = 0.50$? Do you think that a majority of adults say that movies are getting better?

*Solution* For (a): We need to find a confidence interval for $p$ the sample proportion. We know that the estimator $\hat{p} = Y/n = Y/100$ is an unbiased estimator. We then know that the sampling distribution of this estimator is approximately normal. Thus using the general confidence interval above we have that

$$\left[\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = \left[0.45 - z_{\frac{\alpha}{2}} \cdot 0.01758, 0.45 + z_{\frac{\alpha}{2}} \cdot 0.01758\right].$$

Since we have $1 - \alpha = 0.98$ we get that $\alpha = 0.02$. Thus we find $z_{0.01}$. That is we find $P(Z \leq z_{0.01}) = 0.01$. We find that $z_{0.01]} \approx 2.33$. Thus we find that the confidence interval with confidence coefficient 0.98 is

$$[0.409, 0.491].$$

For (b): The interval does not include 0.50. So we have statistical evidence at %98 confidence that less than half of adults think movies are getting better. □

**Example 3.6.3**

The shopping times of $n = 64$ randomly selected customers at a local supermarket were recorded. The average and variance of the 64 shopping times were 33 minutes and 256 minutes$^2$, respectively. Estimate $\mu$, the true average shopping time per customer, with a confidence coefficient of $1 - \alpha = 0.90$.

*Solution* We need to estimate the true average shopping time per customer $\mu$ with a confidence coefficient of 0.90. To do this we can use a common point estimator for the pivotal quantity to find the endpoints and then convert it to the standard normal disribution since the sample size is large. We will let $\hat{\mu} = \overline{Y}$. Thus we see that the interval becomes

$$\left[\hat{\mu} - z_{0.05} \cdot \sqrt{\frac{256}{64}}, \hat{\mu} + z_{0.05} \cdot \sqrt{\frac{256}{64}}\right] = [33 - 2 \cdot z_{0.05}, 33 + 2 \cdot z_{0.05}].$$

We then find the crtical values $P(Z \leq Z_{0.05}) = 0.05$ which gives us $z_{0.05} = 1.645$. Thus we see that the estimate of $\mu$ with %90 probability is the interval

$$[29.71, 36.29].$$

□

**Example 3.6.4**

Two brands of refrigerators, denoted A and B, are each guaranteed for 1 year. In a random sample of 50 refrigerators of brand A, 12 were observed to fail before the guarantee period ended. An independent random sample of 60 brand B refrigerators also revealed 12 failures during the guarantee period. Estimate the true difference $(p_1 - p_2)$ between proportions of failures during the guarantee period, with confidence coefficient approximately 0.98.

*Solution* We will use $\hat{p_1} - \hat{p_2}$ to estimate the difference between sample proportions of failtures during the gurantee period. We need a confidence interval with coefficient coefficient of $1-\alpha = 0.98$. Thus we see that the interval becomes

$$\left[(\hat{p_1} - \hat{p_2}) - z_{0.01} \cdot \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}, (\hat{p_1} - \hat{p_2}) + z_{0.01} \cdot \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}\right].$$

We have the sample sizes and the standard error all we need to find is the critcal values. That is $P(Z \leq z_{0.01}) = 0.01$. We see that $z_{0.01} = 2.33$. Substituting everything we see that the endpoints are

$$(\hat{p_1} - \hat{p_2})_L = \left(\frac{12}{50} - \frac{12}{60}\right) - 2.33 \cdot \sqrt{\frac{12/50(1-12/50)}{50} + \frac{12/60(1-12/60)}{60}} = -0.1451$$

$$(\hat{p_1} - \hat{p_2})_u = \left(\frac{12}{50} + \frac{12}{60}\right) - 2.33 \cdot \sqrt{\frac{12/50(1-12/50)}{50} + \frac{12/60(1-12/60)}{60}} = 0.2251.$$

Thus the confidence interval with confidence coefficient $1 - \alpha = 0.98$ is

$$[-0.1451, 0.2251].$$

□

In this section, we explored the pivotal method to derive large-sample confidence intervals for various parameters, including $\mu$, $p$, $\mu_1 - \mu_2$, and $p_1 - p_2$. The key formula for constructing these intervals is $\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$, where $\hat{\theta}$ is the point estimator, $\sigma_{\hat{\theta}}$ is the standard error (given in Table 1), and $z_{\alpha/2}$ is the critical value from the standard normal distribution. When the population variance $\sigma^2$ is unknown, it is replaced by the sample variance $s^2$, which introduces minimal loss of accuracy for large samples. Similarly, for proportions, the sample proportions $\hat{p}_1$ and $\hat{p}_2$ are used to estimate $p_1$ and $p_2$, respectively. When $n$ is large we can replace $p$ with $\hat{p}$ however, the resulting confidence interval will have approximately the stated confidence coefficient. These substitutions allow us to construct confidence intervals that approximate the stated confidence coefficient, ensuring reliable inference even when population parameters are unknown. The theoretical justification for these substitutions will be given in the next chapter.

## 3.7 Small-Sample Confidence Intervals for $\mu$ and $\mu_1 - \mu_2$

When the population variance $\sigma^2$ is unknown we can approximate it with the sample variance $S^2$. Thus we can use the $t$-distribution for our pivotal quantity

$$T = \frac{\overline{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Thus after isolating for $\mu$ we see that a $100(1-\alpha)\%$ is

$$\left[\overline{Y} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{Y} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right].$$

**Example 3.7.1**

A manufacturer of gunpowder has developed a new powder, which was tested in eight shells. The resulting muzzle velocities, in feet per second, were as follows:

3005 2925 2935 2965 2995 3005 2937 2905

Find a 95% confidence interval for the true average velocity $\mu$ for shells of this type. Assume that muzzle velocities are approximately normally distributed.

*Solution* Since we do not have the population variance we are going to use the sample variance instead. This will give us a $t-$distribution with $n-1=7$ df. We first find the sample mean $\overline{y}=2959$. We next find $S^2$:

$$S^2 = \frac{1}{7}\sum_{i=1}^{8}(Y_i - \overline{y})^2 = 1528.0.$$

Thus we see that the interval becomes

$$\left[2959 - t_{0.025}\sqrt{\frac{1528}{8}}, 2959 + t_{0.025}\sqrt{\frac{1528}{8}}\right].$$

To find the crtical values we find $P(T < t_{0.025}) = 0.025$ . We find that $t_{0.025]} = 2.365$. Putting everything together we get that

$$\left[2959 - 2.365\cdot\sqrt{\frac{1528}{8}}, 2959 + 2.365\cdot\sqrt{\frac{1528}{8}}\right] = [2926.32, 2991.68].$$

□

To construct a confidence interval for $\mu_1 - \mu_2$, suppose that $Y_1, Y_2, \ldots, Y_{n_1} \overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2)$ and $X_1, X_2, \ldots, X_{n_2} \overset{i.i.d.}{\sim} N(\mu_2, \sigma_2^2)$, where the two populations are independent. Assume that the population variances are equal, i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$, but unknown.

Under these conditions, the difference in sample means $\overline{Y} - \overline{X}$ is normally distributed with mean $\mu_1 - \mu_2$ and variance $\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$, i.e.,

$$\overline{Y} - \overline{X} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

Because $\sigma$ is unknown, we need to find an estimator of the common variance $\sigma^2$ so that we can construct a quantity with a $t$ distribution. The usual unbiased estimator of the common variance $\sigma^2$ is obtained by pooling the sample data to obtain the pooled estimator

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

where $S_1^2$ and $S_2^2$ are the sample variances from the first and second samples, respectively. Then, the statistic

$$T = \frac{(\overline{Y} - \overline{X}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

follows a Student's $t$-distribution with $n_1 + n_2 - 2$ degrees of freedom.

Therefore, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\overline{Y} - \overline{X}) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

**Example 3.7.2**

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:

| | Verbal | | Math | |
|---|---|---|---|---|
| | $\overline{y}$ | $s$ | $\overline{y}$ | $s$ |
| Engineering | 446 | 42 | 548 | 57 |
| Language/literature | 534 | 45 | 517 | 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) Interpret the results obtained in parts (a).

*Solution* For (a): We are asked to create a confidence interval with confidence coefficient $1 - \alpha = 0.95$ for $\overline{Y}_1 - \overline{Y}_2$ where $\overline{Y_1}$ is the average of verbal score averages of engineering and language/literature. Since the population variances for both engineering and language/literature are unknown we will have to approximate it with sample pooled variance. Thus we see that $\overline{Y}_1 - \overline{Y}_2 = 534 - 446 = 88$. Next we find $S_p$.

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} = \frac{(15-1)(42)^2 + (15-1)(45)^2}{28} = 1894.5$$

so $S_p = 43.53$. Next for the critcal values we find $P(T < t_{0.025}) = 0.025$ where $T$ has a $t$-distribution with 28 df. We see that $t_{0.025} = 2.048$. Thus the interval becomes

$$(-120.54, -55.46).$$

□

## 3.8 Confidence Intervals for $\sigma^2$

As we have said earlier, the population variance is usually unknown to an experimenter and most statistical analyses result in estimating this quantity. In previous sections, we have shown that the sample variance $S^2 = (1/n-1)\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ is an unbiased estimator for $\sigma^2$. Throughout our construction of confidence intervals for $\mu$, $S^2$ was used to estimate $\sigma^2$. In addition to using $\sigma^2$

to construct confidence intervals for the mean the mean difference, we can instead find confidence intervals for the variance as well. To do this we need a pivotal quantity. We can refer back to chapter 2 and use the fact that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Thus we then find two values $\chi^2_{1-\alpha/2}$ and $\chi^2_{\alpha/2}$ such that

$$P\left(\chi^2_{1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2}\right) = 1-\alpha.$$

Solving for $\sigma^2$ we get that

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right).$$

Thus a $100(1-\alpha)\%$ confidence interval for $\sigma^2$ is

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right]$$

**Example 3.8.1**

An experimenter wanted to check the variability of measurements obtained by using equipment designed to measure the volume of an audio source. Three independent measurements recorded by this equipment for the same sound were 4.1, 5.2, and 10.2. Estimate $\sigma^2$ with a confidence coefficient of 0.90.

*Solution* We need to estimate the variance with a confidence interval with confidence coefficient $1-\alpha = 0.90$ or $\alpha = 0.10$. To do this we can use the pivotal quantity $(n-1)S^2/\sigma^2$. Using the derviation above, we see that the interval then becomes

$$\left[\frac{2S^2}{\chi^2_{\alpha/2}}, \frac{2S^2}{\chi^2_{1-\alpha/2}}\right].$$

We now find $\chi^2_{\alpha/2} = \chi^2_{0.05}$. That is we need to find $P(\chi^2_{(2)} > \chi^2_{0.05}) = 0.05$. We see that $\chi^2_{0.05} = 0.102587$. Simiarily we see that $\chi^2_{0.950} = 5.99147$. Then we find $S^2$. To do this we first find the sample mean $\overline{Y} = 6.5$. We then see that $S^2 = 10.567$. Thus we have that

$$\left[\frac{2 \cdot 10.57}{5.991}, \frac{2 \cdot 10.57}{0.103}\right] = [3.53, 205.24].$$

□

## 3.9 Practice Problems

**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

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

*Solution* We begin by expanding the bracket within the expectation:

$$\begin{aligned}\text{MSE}(\hat{\theta}) &= E\left[(\hat{\theta}-\theta)^2\right] \\ &= E\left[\left(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta\right)^2\right] \\ &= E\left[\left(\hat{\theta}-E(\hat{\theta})\right)^2+2\left(\hat{\theta}-E(\hat{\theta})\right)\left(E(\hat{\theta})-\theta\right)+\left(E(\hat{\theta})-\theta\right)^2\right].\end{aligned}$$

Next, we use the linearity of expectation and the fact that $E[\hat{\theta}-E(\hat{\theta})]=0$ (since $\hat{\theta}-E(\hat{\theta})$ is centered around its mean). This simplifies the second term to zero:

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

The first term is the variance of $\hat{\theta}$, denoted $V(\hat{\theta})$, and the second term is the square of the bias $B(\hat{\theta})$. Thus:

$$\text{MSE}(\hat{\theta}) = V(\hat{\theta})+\left(B(\hat{\theta})\right)^2.$$

□

**8.2**

a) If $\hat{\theta}$ is an unbiased estimator for $\theta$, what is $B(\hat{\theta})$?

b) If $B(\hat{\theta})=5$, what is $E(\hat{\theta})$?

*Solution* For (a): If $\hat{\theta}$ is an unbiased estimator for $\theta$, then $E(\hat{\theta})=\theta$. Thus then

$$B(\hat{\theta})=E(\hat{\theta})-\theta=\theta-\theta=0.$$

For (b): Now if $B(\hat{\theta})=5$ then we have that $5=E(\hat{\theta})-\theta$ or $E(\hat{\theta})=5+\theta$. This means systematically $\hat{\theta}$ overestimates $\theta$ by 5 units. □

**8.4**

Refer to Exercise 8.1.

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

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

*Solution* For (a): If $\hat{\theta}$ is unbiased, then the bias is equal to zero. Thus we have that $\text{MSE}(\hat{\theta})=V(\hat{\theta})+B(\hat{\theta})=V(\hat{\theta})$. That is the MSE is just the variance. This makes sense as the MSE was thought to be the value to account for both the bias and the variance of an estimator. For (b): Now if $\hat{\theta}$ is biased then we are left with $\text{MSE}(\hat{\theta})>V(\hat{\theta})$ since MSE includes both the variance and the square of the bias, so it increases when there is bias. □

**8.5**

Refer to Exercises 8.1 and consider the unbiased estimator $\hat{\theta}^*$ that you proposed in Example 3.2.3.

a) Express $\text{MSE}(\hat{\theta}^*)$ as a function of $V(\hat{\theta})$.

b) Give an example of a value of $a$ for which $\text{MSE}(\hat{\theta}^*) < \text{MSE}(\hat{\theta})$.

c) Give an example of values for $a$ and $b$ for which $\text{MSE}(\hat{\theta}^*) > \text{MSE}(\hat{\theta})$.

*Solution* For (a): Recall that we found $\hat{\theta}^* = 1/a(\hat{\theta} - b)$. Notice that

$$\begin{aligned} V(\hat{\theta}^*) &= V(1/a(\hat{\theta} - b) \\ &= (1/a^2)V(\hat{\theta}). \end{aligned}$$

Next notice that $B(\hat{\theta}) = E(\hat{\theta}^*) - \theta = 0$. Thus we have that

$$\text{MSE}(\hat{\theta}^*) = (1/a^2)V(\hat{\theta}).$$

For (b): Recall that $\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + [(a-1)\theta + b]^2$. We need

$$\text{MSE}(\hat{\theta}^*) = (1/a^2)V(\hat{\theta}) < V(\hat{\theta}) + [(a-1)\theta + b]^2 = \text{MSE}(\hat{\theta}).$$

We see that when $a = 1$ we have that

$$\text{MSE}(\hat{\theta}^*) = V(\hat{\theta}) < V(\hat{\theta}) + b^2.$$

For (c): Letting $0 < a < 1$ be any positive constant and $b = 0$ then we have that

$$\text{MSE}(\hat{\theta}^*) = (1/a^2)V(\hat{\theta}) > V(\hat{\theta}) + [(a-1)\theta + b]^2 = \text{MSE}(\hat{\theta}$$

□

**8.7**

Consider the situation described in Example 3.2.4. 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

$$\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = c \neq 0?$$

*Solution* We now need to minimze the variance of $\hat{\theta}_3$ considering that $\hat{\theta}_1$ and $\hat{\theta}_2$ are not independent. We found that

$$\text{Var}(\hat{\theta}_3) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2 + 2a(1-a)Cov(\hat{\theta}_1, \hat{\theta}_2).$$

Since this is a function of $a$ and is convex we can use calculus to find the global minimum of this function to minimize the variance. We see that

$$V'(a) = 2\sigma_1^2 a - 2ca - 2\sigma_2^2(1-a) + 2c(1-a).$$

Setting this equal to zero and solving for $a$ we find that the value of $a$ that gives us the minimum is

$$a = \frac{\sigma_2^2 - c}{\sigma_2^2 + \sigma_1^2 - 2c}.$$

$\square$

## 8.8

Suppose that $Y_1, Y_2, Y_3$ denote a random sample from an exponential distribution with density function

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

Consider the following five estimators of $\theta$:

$$\hat{\theta}_1 = Y_1, \quad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \quad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}, \quad \hat{\theta}_4 = \min(Y_1, Y_2, Y_3), \quad \hat{\theta}_5 = \bar{Y}.$$

(a) Which of these estimators are unbiased?

(b) Among the unbiased estimators, which has the smallest variance?

*Solution* For (a): We need to see which estimator is unbiased. Recall that an estimator is unbiased if $E(\hat{\theta}) = \theta$. We begin and find the expected value for each estimator.

$$E(\hat{\theta}_1) = E(Y_1) = \theta.$$

$$E(\hat{\theta}_2) = E\left[\frac{Y_1 + Y_2}{2}\right] = \frac{E(Y_1) + E(Y_2)}{2} = \frac{2\theta}{2} = \theta$$

$$E(\hat{\theta}_3) = E\left[\frac{Y_1 + 2Y_2}{3}\right] = \frac{E(Y_1) + 2E(Y_2)}{3} = \frac{3\theta}{3} = \theta.$$

$$E(\hat{\theta}_4) = E(\min(Y_1, Y_2, Y_3)) = \frac{\theta}{3}.$$

$$E(\overline{Y}) = E\left[\frac{Y_1 + Y_2 + Y_3}{3}\right] = \frac{3\theta}{3} = \theta.$$

Thus the unbiased estimators are $\hat{\theta}_1, \hat{\theta}_2.\hat{\theta}_3.\hat{\theta}_5$. For (b): We compute the variances for each:
$\hat{\theta}_1 = Y_1$:

$$\text{Var}(\hat{\theta}_1) = \text{Var}(Y_1) = \theta^2$$

$\hat{\theta}_2 = \frac{Y_1+Y_2}{2}$:

$$\text{Var}(\hat{\theta}_2) = \frac{1}{4}(\text{Var}(Y_1) + \text{Var}(Y_2)) = \frac{1}{4}(2\theta^2) = \frac{\theta^2}{2}$$

$\hat{\theta}_3 = \frac{Y_1+2Y_2}{3}$:

$$\text{Var}(\hat{\theta}_3) = \frac{1}{9}(\theta^2 + 4\theta^2) = \frac{5\theta^2}{9}$$

$\hat{\theta}_5 = \bar{Y} = \frac{Y_1+Y_2+Y_3}{3}$:

$$\text{Var}(\bar{Y}) = \frac{\theta^2}{3}$$

Thus we can see that the one with the smallest variance is $\hat{\theta}_5$ $\square$

**8.9**

Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a population with probability density function

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

Suggest a suitable statistic to use as an unbiased estimator for $\theta$. *[Hint: Consider $\overline{Y}$ ].*

*Solution* We can use the sample mean $\overline{Y}$ to create an unbiased estimator for $\theta$. Notice that the pdf is an exponential distribution with mean $\theta + 1$. Thus then $E(\overline{Y}) = \theta + 1$. Therefore if we let $U = \overline{Y} - 1$ then we have that $E(U) = E(\overline{Y} - 1) = \theta + 1 - 1 = \theta$ as required. $\square$

**8.10**

The number of breakdowns per week for a type of minicomputer is a random variable $Y$ with a Poisson distribution and mean $\lambda$. A random sample $Y_1, Y_2, \ldots, Y_n$ of observations on the weekly number of breakdowns is available.

(a) Suggest an unbiased estimator for $\lambda$.

(b) The weekly cost of repairing these breakdowns is $C = 3Y + Y^2$. Show that $E(C) = 4\lambda + \lambda^2$.

(c) Find a function of $Y_1, Y_2, \ldots, Y_n$ that is an unbiased estimator of $E(C)$. *[Hint: Use what you know about $\bar{Y}$ and $(\bar{Y})^2$.]*

*Solution* For (a): We can use the sample mean $\overline{Y}$ to be an unbiased estimator for the mean $\lambda$. For (b): We have clearly that $E(C) = E(3Y + Y^2) = 3E(Y) + E(Y^2) = 3\lambda(\lambda + \lambda^2) = 4\lambda + \lambda^2$. Where in the last equality we used the fact that $\text{Var}(Y) = \lambda = E(Y^2) - [E(Y)]^2 = E(Y^2) - \lambda^2$. For (c): We use the fact that $E(\overline{Y}) = \lambda$ and $E(Y_i^2) = \lambda + \lambda^2$. Thus then we have that $4E(\overline{Y}) + E(Y_i^2) - \lambda = 4\lambda + \lambda^2$. However we need an unbiased estimator for $\lambda^2$. We see that $\frac{1}{n}\sum_{i=1}^{n} Y_i$ is an unbiased estimator for $\lambda^2 + \lambda$. Thus then let

$$E(\hat{C}) = 4\overline{Y} + \frac{1}{n}\sum_{i=1}^{n} Y_i - \overline{Y}.$$

Then we have that

$$E[E(\hat{C})] = 4E(\overline{Y}) + E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] - E(\overline{Y}) = 4\lambda + (\lambda^2 + \lambda) - \lambda = 4\lambda + \lambda^2.$$

$\square$

**8.11**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from a population with mean 3. Assume that $\hat{\theta}_2$ is an unbiased estimator of $E(Y^2)$ and that $\hat{\theta}_3$ is an unbiased estimator of $E(Y^3)$. Give an unbiased estimator for the third central moment of the underlying distribution.

*Solution* The third central moment is defined as $E[(Y - \mu)^3]$ or in our case $E[(Y - 3)^3]$. Expanding this we get that $E[(Y - 3)^3] = E(Y^3) - 9E(Y^2) + 27E(Y) - 27$. Thus an unbiased estimator for the third central moment is

$$\hat{\mu}_3 = \hat{\theta}_3 - 9\hat{\theta}_2 + 54.$$

□

**8.12**

The reading on a voltage meter connected to a test circuit is uniformly distributed over the interval $(\theta, \theta + 1)$, where $\theta$ is the true but unknown voltage of the circuit. Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of such readings.

(a) Show that $\overline{Y}$ is a biased estimator of $\theta$ and compute the bias.

(b) Find a function of $\overline{Y}$ that is an unbiased estimator of $\theta$.

(c) Find $\mathrm{MSE}(\overline{Y})$ when $\overline{Y}$ is used as an estimator of $\theta$.

*Solution* For (a): We show that $\overline{Y}$ is a biased estimator of $\theta$ and we will compute the bias. We see that $E(\overline{Y}) = \mu$ where $\mu$ is the mean of the uniform distribution over $(\theta, \theta+1)$. We see that then $\mu = \frac{\theta+1+\theta}{2} = \theta + 1/2$. Thus we have that then the bias is $B(\overline{Y}) = E(\overline{Y}) - \theta = \theta + 1/2 - \theta = 1/2$. For (b): To turn $\overline{Y}$ into a unbiased estimator we simply let $\hat{\theta} = \overline{Y} - B(\overline{Y}) = \overline{Y} - 1/2$. This gives us $E(\hat{\theta}) = E(\overline{Y}) - 1/2 = \theta + 1/2 - 1/2 = \theta$. For (c): We find the MSe of $\overline{Y}$. To do this we first compute the variance of $\overline{Y}$ which is

$$\sigma^2_{\overline{Y}} = \frac{\sigma^2}{n} = \frac{(\theta_2 - \theta_1)^2}{12n} = \frac{1}{12n}.$$

Putting it together we see that

$$\mathrm{MSE}(\overline{Y}) = \mathrm{Var}(\overline{Y}) - B(\overline{Y})^2 = \frac{1}{12n} + \frac{1}{4}.$$

□

**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 $V(Y)$.

*Solution* For (a): We need to show that $n(Y/n)(1 - Y/n)$ is an unbiased estimator for $V(Y) = np(1-p)$. We have that then

$$E\left[n(Y/n)(1 - Y/n)\right] = E(Y) - \frac{1}{n}E(Y^2) = np - \frac{1}{n}(np(1-p) + n^2p^2) = np(1-p)\left(1 - \frac{1}{n}\right).$$

For (b): To me this unbiased we have to divide by $1 - 1/n = (n-1)/n$. Thus the unbiased estimator for $V(Y)$ is

$$\frac{n-1}{n} \cdot \hat{p}(1 - \hat{p}).$$

□

**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 $\theta$ 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 $\theta$.

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

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

*Solution* For (a): We need to show that $\hat{\theta}$ is a biased estimator. That is $E(\hat{\theta}) \neq \theta$. To find the expected value of $\hat{\theta}$, we first need to find the pdf of $\hat{\theta} = \max(Y_1, Y_2, \ldots, Y_n)$. We know from Chapter 1 that

$$f_{\hat{\theta}}(y) = n[F(y)]^{n-1} f(y) = \frac{n\alpha[F(y)]^{n-1} y^{\alpha-1}}{\theta^\alpha}.$$

We now find $F(y)$:

$$F(y) = \frac{\alpha}{\theta^\alpha} \int_0^y t^{\alpha-1} dt = \frac{y^\alpha}{\theta^\alpha}.$$

Thus the pdf of the estimator is then

$$f_{\hat{\theta}}(y) = \frac{n(y^{\alpha n-\alpha}) y^{\alpha-1}}{\theta^{\alpha n-\alpha} \cdot \theta^\alpha} = \frac{n(y^{\alpha n-1})}{\theta^{\alpha n}}$$

for $0 \le y \le \theta$. We now find $E(\hat{\theta})$:

$$E(\hat{\theta}) = \frac{n}{\theta^{\alpha n}} \int_0^\theta y \cdot y^{\alpha n-1} dy = \frac{n}{\theta^{\alpha n}} \cdot \frac{\theta^{\alpha n+1}}{\alpha n+1} = \frac{n\theta}{\alpha n+1} \neq \theta.$$

For (b): To make $\hat{\theta}$ unbiased we simply multiply by $\alpha n + 1$. That is $\hat{\theta_2} = (\alpha n+1)\hat{\theta}$ gives us an unbiased estimator. For (c): To find the MSe for $\hat{\theta}$ we have to find the variance.

$$E(\hat{\theta}^2) = \frac{n}{\theta^{\alpha n}} \int_0^\theta y^2 \cdot y^{\alpha n-1} dy = \frac{n}{\theta^{\alpha n}} \cdot \frac{\theta^{\alpha n+2}}{\alpha n+2} = \frac{n\theta^2}{\alpha n+2}.$$

Thus we then see that

$$\text{Var}(\hat{\theta}) = E(\hat{\theta}^2) - [E(\hat{\theta})]^2 = \frac{n\theta^2}{\alpha n+2} - \left(\frac{n\theta}{\alpha n+1}\right)^2$$

Therefore the MSE of $\hat{\theta}$ is then

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + B(\hat{\theta}) = \frac{n\theta^2}{\alpha n+2} - \left(\frac{n\theta}{\alpha n+1}\right)^2 + \left[\frac{n\theta}{\alpha n+1} - \theta\right].$$

$\square$

**8.15**

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

$$f(y) = \begin{cases} 3\beta^3 y^{-4}, & \beta \le y, \\ 0, & \text{elsewhere}, \end{cases}$$

where $\beta > 0$ is unknown. (This is one of the Pareto distributions introduced in Exercise 6.18.) Consider the estimator $\hat{\beta} = \min(Y_1, Y_2, \ldots, Y_n)$.

(a) Derive the bias of the estimator $\hat{\beta}$.

(b) Derive $\text{MSE}(\hat{\beta})$.

*Solution* For (a): To find the bias of the estimator $\hat{\beta}$ we must find $E(\hat{\beta})$ first. Thus we then need to find the pdf of $\hat{\beta} = \min(Y_1, Y_2, \ldots, Y_n)$. We know from chapter 1 that

$$f_{\hat{\beta}}(y) = n[1 - F(y)]^{n-1} f(y) = 3n\beta^3 [1 - F(y)]^{n-1} y^{-4}.$$

We find the cdf $F(y)$:

$$F(y) = 3\beta^3 \int_\beta^y t^{-4} dt = 3\beta^3 \cdot \left( \frac{1}{3\beta^3} - \frac{1}{3y^3} \right) = 1 - \frac{\beta^3}{y^3}.$$

Thus the pdf for $\hat{\beta}$ is

$$f_{\hat{\beta}}(y) = 3n\beta^3 [1 - F(y)]^{n-1} y^{-4} = 3n\beta^{3n} y^{-3n-1}.$$

We now find $E(\hat{\beta})$:

$$E(\hat{\beta}) = 3n\beta^{3n} \int_\beta^\infty y \cdot y^{-3n-1} dy = 3n\beta^{3n} \cdot \left( \frac{\beta^{-3n+1}}{3n-1} \right) = \frac{3n\beta}{3n-1}.$$

Thus the bias of $\hat{\beta}$ is then

$$B(\hat{\beta}) = E(\hat{\beta}) - \beta = \frac{3n\beta}{3n-1} - \beta = \frac{\beta}{3n-1}.$$

For (b): To find the MSE of $\hat{\beta}$ we find the variance.

$$E(\hat{\beta}^2) = 3n\beta^{3n} \int_\beta^\infty y^2 \cdot y^{-3n-1} dy = 3n\beta^{3n} \int_\beta^\infty y^{-3n+1} dy = \frac{3n\beta^2}{3n-2}.$$

Then

$$\text{Var}(\hat{\beta}) = E(\hat{\beta}) - [E(\hat{\beta})]^2 = \frac{3n\beta^2}{3n-2} - \left( \frac{3n\beta}{3n-1} \right)^2$$

Thus the MSE $\hat{\beta}$ is then

$$\text{MSE}(\hat{\beta}) = \frac{3n\beta^2}{3n-2} - \left( \frac{3n\beta}{3n-1} \right)^2 + \left( \frac{\beta}{3n-1} \right)^2.$$

$\square$

**8.16**

Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with parameters $\mu$ and $\sigma^2$.

(a) Show that $S = \sqrt{S^2}$ is a biased estimator of $\sigma$. *[Hint: Recall the distribution of $(n-1)S^2/\sigma^2$ and the result given in Exercise 4.112.]*

(b) Adjust $S$ to form an unbiased estimator of $\sigma$.

(c) Find an unbiased estimator of $\mu - z_\alpha \sigma$, the point that cuts off a lower-tail area of $\alpha$ under this normal curve.

*Solution* For (a): We want to show that $E(S) \neq \sigma$. Note that

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}.$$

Since $S$ is a nonlinear function of the sample, it is not an unbiased estimator of $\sigma$. In fact, due to Jensen's inequality:

$$E(S) = E\left(\sqrt{S^2}\right) < \sqrt{E(S^2)} = \sqrt{\sigma^2} = \sigma.$$

Therefore, $S$ is a biased estimator of $\sigma$.

For (b): We know that for $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Let $c_n = E\left(\sqrt{\frac{\chi^2_{n-1}}{n-1}}\right) < 1$, which depends only on $n$. Then:

$$E(S) = \sigma \cdot c_n.$$

So, an unbiased estimator of $\sigma$ is:

$$S^* = \frac{S}{c_n}.$$

For (c): We want an unbiased estimator of $\mu - z_\alpha \sigma$. We already have $\bar{Y}$ as an unbiased estimator of $\mu$, and from part (b), $S/c_n$ is an unbiased estimator of $\sigma$ Therefore, an unbiased estimator of $\mu - z_\alpha \sigma$ is:

$$\bar{Y} - z_\alpha \cdot \frac{S}{c_n}.$$

□

**8.17**

If $Y$ has a binomial distribution with parameters $n$ and $p$, then $\hat{p}_1 = Y/n$ is an unbiased estimator of $p$. Another estimator of $p$ is $\hat{p}_2 = (Y+1)/(n+2)$.

(a) Derive the bias of $\hat{p}_2$.

(b) Derive $\text{MSE}(\hat{p}_1)$ and $\text{MSE}(\hat{p}_2)$.

(c) For what values of $p$ is $\text{MSE}(\hat{p}_1) < \text{MSE}(\hat{p}_2)$?

*Solution* For (a): To derive the bias of $\hat{p}_2$ we must find $E(\hat{p}_2)$. We see that

$$E(\hat{p}_2) = E\left(\frac{Y+1}{n+2}\right) = \frac{E(Y)+1}{n+2} = \frac{np+1}{n+2}.$$

Thus we see that then

$$B(\hat{p}_2) = E(\hat{p}_2) - p = \frac{np+1}{n+2} - p = \frac{1-2p}{n+2}.$$

For (b): We know that from Table 1 that

$$\text{MSE}(\hat{p}_1) = \text{Var}(\hat{p}_1) + B(\hat{p}_1)^2 = \frac{p(1-p)}{n} + (p-p)^2 = \frac{p(1-p)}{n}.$$

Next to find the MSE of $\hat{p}_2$ we find the variance first.

$$\text{Var}(\hat{p}_2) = \text{Var}\left(\frac{Y+1}{n+2}\right) = \frac{1}{(n+2)^2}\text{Var}(Y) = \frac{np(1-p)}{(n+2)^2}.$$

Using Part (a) we see that

$$\text{MSE}(\hat{p}_2) = \text{Var}(\hat{p}_2) + B(\hat{p}_2)^2 = \frac{np(1-p)}{(n+2)^2} + \left(\frac{1-2p}{n+2}\right)^2.$$

For (c): We need to find the values of $p$ such that

$$\frac{p(1-p)}{n} < \frac{np(1-p)}{(n+2)^2} + \left(\frac{1-2p}{n+2}\right)^2 = \frac{np(1-p)+1-2p}{(n+2)^2}.$$

Solving this inequality we see that for all values of $0 < p < 0.5$ the inequality is true. □

**8.18**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from a population with a uniform distribution on the interval $(0, \theta)$. Consider $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$, the smallest-order statistic. Use the methods of Chapter 1 to derive $E(Y_{(1)})$. Find a multiple of $Y_{(1)}$ that is an unbiased estimator for $\theta$.

*Solution* To find $E(Y_{(1)})$ we must find the pdf of $Y_{(1)}$. We know from Chapter 1 that

$$f_{Y_{(1)}}(y) = n[1-F(y)]^{n-1}f(y) = n[1-F(y)]^{n-1}\frac{1}{\theta}.$$

We then also know that since the population is uniform over $(0,\theta)$ then $F(y) = y/\theta$. Thus we have that then

$$f_{Y_{(1)}}(y) = \frac{n\left(1-\frac{y}{\theta}\right)^{n-1}}{\theta}.$$

We now solve for $E(Y_{(1)})$.

$$E(Y_{(1)}) = \int_0^\theta y \cdot \frac{n\left(1-\frac{y}{\theta}\right)^{n-1}}{\theta}dy = \frac{n}{\theta}\int_0^1 (u\theta)(1-u)^{n-1}\theta du.$$

We did the substitution $u = y/\theta$. Notice that this is the beta integral and so

$$E(Y_{(1)}) = \frac{\theta}{n+1}.$$

Thus to make this an unbiased estimator of $\theta$ we can let $\hat{\theta} = (n+1)Y_{(1)]}$. $\square$

**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

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

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 $\theta$ and find $\text{MSE}(\hat{\theta})$. *[Hint: Recall the results of Problem 6.8.1.]*

*Solution* Using the pdf we found for $Y_{(1)}$ of an exponential distribution with mean $\theta$ we see that

$$f_{Y_{(1)}}(y) = \frac{n}{\theta}e^{-ny/\theta}.$$

The notice that $E(Y_{(1)}) = \theta/n$. We then have that $E(nY_{(1)}) = n\frac{\theta}{n} = \theta$. Thus $\hat{\theta} = nY_{(1)}$ is an unbiased estimator. We next find the variance:

$$\text{Var}(\hat{\theta}) = n^2\text{Var}(Y_{(1)}) = n^2 cdot \frac{\theta^2}{n^2} = \theta^2.$$

Thus we have that

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + B(\hat{\theta})^2 = \theta^2 + 0^2 = \theta^2.$$

$\square$

**8.20**

Suppose that $Y_1,\ Y_2,\ Y_3,\ Y_4$ denote a random sample of size 4 from a population with an exponential distribution whose density is given by

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

a. Let $X = \sqrt{Y_1Y_2}$. Find a multiple of $X$ that is an unbiased estimator for $\theta$. *[Hint: Use your knowledge of the gamma distribution and the fact that $\Gamma(1/2) = \sqrt{\pi}$ to find $E(\sqrt{Y_1})$. Recall that the variables $Y_i$ are independent.]*

b. Let $W = \sqrt{Y_1Y_2Y_3Y_4}$. Find a multiple of $W$ that is an unbiased estimator for $\theta^2$. *[Recall the hint for part (a).]*

*Solution* For (a): We first find $E(X)$. Notice that $E(X) = E(\sqrt{Y_1Y_2}) = E(\sqrt{Y_1})E(\sqrt{Y_2})$. WE begin by finding $E(Y_1) = E(Y_2)$.

$$E(Y_1) = \frac{1}{\theta}\int_0^\infty \sqrt{y}e^{-y/\theta}dy$$

$$= \frac{1}{\theta}\int_0^\infty \sqrt{\theta u}e^{-u}\theta du$$

$$= \sqrt{\theta}\int_0^\infty u^{3/2-1}e^{-u}du$$

$$= \sqrt{\theta}\Gamma\left(\frac{3}{2}\right)$$

$$= \frac{\sqrt{\theta}\sqrt{\pi}}{2}.$$

So we have that then

$$E(X) = \theta \cdot \frac{\pi}{4}.$$

Thus we see that then

$$\hat{\theta} = \frac{4}{\pi}\sqrt{Y_1Y_2}$$

is an unbiased estimator of $\theta$. For (b): Using the result of $E(Y_i)$ we see that then

$$E(\sqrt{Y_1Y_2Y_3Y_4}) = \frac{\theta^2 3\pi}{16}.$$

Thus then

$$\hat{\theta}^2 = \frac{16}{3\pi}\sqrt{Y_1Y_2Y_3Y_4}$$

is an unbiased estimator of $\hat{\theta}^2$. □

**8.21**

An investigator is interested in the possibility of merging the capabilities of television and the Internet. A random sample of $n = 50$ Internet users yielded that the mean amount of time spent watching television per week was 11.5 hours and that the standard deviation was 3.5 hours. Estimate the population mean time that Internet users spend watching television and place a bound on the error of estimation.

*Solution* To estimate the population mean we can use the unbiased point estimator $\overline{Y}$. We found that the point estimate of the population was $\overline{y} = 11.5$. We are asked to place a bound on the error of estimation. We first calculate the standard error

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{50}} \approx 0.495.$$

Because the sample size is large we know the sampling distirbution for the sample mean is approximately normal. Thus we place a 2-standard-error bound on the estimation. That is

$$b = 2\sigma_{\overline{Y}} = 2 \cdot 0.495 = 0.99.$$

Thus the bound on the error of estimation is $11.5 \pm 0.99$. □

**8.22**

An increase in the rate of consumer savings frequently is tied to a lack of confidence in the economy and is said to be an indicator of a recessional tendency in the economy. A random sampling of $n = 200$ savings accounts in a local community showed the mean increase in savings account values to be 7.2% over the past 12 months, with standard deviation 5.6%. Estimate the mean percentage increase in savings account values over the past 12 months for depositors in the community. Place a bound on your error of estimation.

*Solution* To estimate the population mean we can use the unbiased point estimator $\overline{Y}$. We found that the point estimate of the population was $\overline{y} = 7.2$. We are asked to place a bound on the error of estimation. We first calculate the standard error

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{5.6}{\sqrt{200}} \approx 0.396.$$

Because the sample size is large we know the sampling distirbution for the sample mean is approximately normal. Thus we place a 2-standard-error bound on the estimation. That is

$$b = 2\sigma_{\overline{Y}} = 2 \cdot 0.396 = 0.792.$$

Thus the bound on the error of estimation is $7.2 \pm 0.792$. □

**8.23**

The Environmental Protection Agency and the University of Florida recently cooperated in a large study of the possible effects of trace elements in drinking water on kidney-stone disease. The accompanying table presents data on age, amount of calcium in home drinking water (measured in parts per million), and smoking activity. These data were obtained from individuals with recurrent kidney-stone problems, all of whom lived in the Carolinas and the Rocky Mountain states.

| | **Carolinas** | **Rockies** |
|---|---|---|
| Sample size | 467 | 191 |
| Mean age | 45.1 | 46.4 |
| Standard deviation of age | 10.2 | 9.8 |
| Mean calcium component (ppm) | 11.3 | 40.1 |
| Standard deviation of calcium | 16.6 | 28.4 |
| Proportion now smoking | 0.78 | 0.61 |

(a) Estimate the average calcium concentration in drinking water for kidney-stone patients in the Carolinas. Place a bound on the error of estimation.

(b) Estimate the difference in mean ages for kidney-stone patients in the Carolinas and in the Rockies. Place a bound on the error of estimation.

(c) Estimate and place a 2-standard-deviation bound on the difference in proportions of kidney-stone patients from the Carolinas and Rockies who were smokers at the time of the study.

*Solution* For (a): We see that the mean calcium convertration point estimation was given by $\overline{Y}_1 = 11.3$. We are asked to place a bound on the error of estimation. We first calculate the standard error

$$\sigma_{\overline{Y}_1} = \frac{16.6}{\sqrt{467}} \approx 0.768.$$

Because the sample size is large we know the sampling distirbution for the sample mean is approximately normal. Thus we place a 2-standard-error bound on the estimation. That is

$$b = 2\sigma_{\overline{Y}} = 2 \cdot 0.768 = 1.536.$$

Thus the bound on the error of estimation is $11.3 \pm 1.536$.

For (b): We see that the mean age difference point estimation was given by $\overline{Y}_2 = 46.4 - 45.1 = 1.3$. We are asked to place a bound on the error of estimation. We first calculate the standard error

$$\sigma_{\overline{Y}_2} = \sqrt{\frac{10.2^2}{467} + \frac{9.8^2}{191}} \approx 0.852.$$

Because the sample size is large we know the sampling distirbution for the sample mean is approximately normal. Thus we place a 2-standard-error bound on the estimation. That is

$$b = 2\sigma_{\overline{Y}} = 2 \cdot 0.852 = 1.704.$$

Thus the bound on the error of estimation is $1.3 \pm 1.704$. For (c): We see that the difference in proportions of kidney-stone patients from the Carolinas and Rockies point estimation was given by $\hat{p}_1 - \hat{p}_2 = 0.78 - 0.61 = 0.17$. We are asked to place a bound on the error of estimation. We first calculate the standard error

$$\sigma_{\overline{Y}_2} = \sqrt{\frac{(0.78)(1-0.78)}{467} + \frac{(0.61)(1-0.61)}{191}} \approx 0.041.$$

Because the sample size is large we know the sampling distirbution for the sample mean is approximately normal. Thus we place a 2-standard-error bound on the estimation. That is

$$b = 2\sigma_{\overline{Y}} = 2 \cdot 0.041 = 0.08.$$

Thus the bound on the error of estimation is $0.17 \pm 0.08$. □

**8.25**

A study was conducted to compare the mean number of police emergency calls per 8-hour shift in two districts of a large city. Samples of 100 8-hour shifts were randomly selected from the police records for each of the two regions, and the number of emergency calls was recorded for each shift. The sample statistics are given in the following table:

| | **Region 1** | **Region 2** |
|---|---|---|
| Sample size | 100 | 100 |
| Sample mean | 2.4 | 3.1 |
| Sample variance | 1.44 | 2.64 |

(a) Estimate the difference in the mean number of police emergency calls per 8-hour shift between the two districts in the city.

(b) Find a bound for the error of estimation.

*Solution* For (a): We see that the mean age difference point estimation was given by $\overline{Y}_1 = 2.4 - 3.1 = -0.7$". For (b): We are asked to place a bound on the error of estimation. We first calculate the standard error

$$\sigma_{\overline{Y}_2} = \sqrt{\frac{1.44}{100} + \frac{2.64}{100}} \approx 0.202.$$

Because the sample size is large we know the sampling distirbution for the sample mean is approximately normal. Thus we place a 2-standard-error bound on the estimation. That is

$$b = 2\sigma_{\overline{Y}} = 2 \cdot 0.392 = 0.404.$$

Thus the bound on the error of estimation is $-0.7 \pm 0.404$. □

**8.26**

The Mars twin rovers, *Spirit* and *Opportunity*, which roamed the surface of Mars in the winter of 2004, found evidence that there was once water on Mars, raising the possibility that there was once life on the "planet". Do you think that the United States should pursue a program to send humans to Mars? An opinion poll[3] indicated that 49% of the 1093 adults surveyed think that we should pursue such a program.

(a) Estimate the proportion of all Americans who think that the United States should pursue a program to send humans to Mars. Find a bound on the error of estimation.

(b) The poll actually asked several questions. If we wanted to report an error of estimation that would be valid for all of the questions on the poll, what value should we use? *[Hint: What is the maximum possible value for $p \times q$?]*

*Solution* For (a): We see that $\hat{p} = Y/n = 0.49$. We are asked to place a bound on the error of estimation. We first calculate the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{(0.49)(1 - 0.49)}{1090}} \approx 0.0151.$$

Then since the sample size is large it is approximately normal so we can palce a 2-standard-error bound which is approximately %95 confidence. That is

$$b = 2\sigma_{\hat{p}} = 2 \cdot 0.0151 = 0.0302.$$

Thus the bound on the error for estimation is $0.49 \pm 0.0302$. For (b): For the maximum possible error, we consider the value of $p \times q$ that maximizes the standard error. Since $q = 1 - p$, the product $p \times q$ is maximized when $p = 0.5$ and $q = 0.5$. Substituting these values into the formula for the standard error:

$$\sigma_{\hat{p}} = \sqrt{\frac{p \times q}{n}} = \sqrt{\frac{0.5 \times 0.5}{1093}} \approx 0.0151.$$

Thus, the maximum possible error of estimation, using a 2-standard-error bound, is:

$$b = 2\sigma_{\hat{p}} = 2 \times 0.0151 = 0.0302.$$

This bound would apply to all questions in the poll, regardless of the specific sample proportion.

□

**8.27**

A random sample of 985 "likely voters"—those who are judged to be likely to vote in an upcoming election—were polled during a phone-athon conducted by the Republican Party. Of those contacted, 592 indicated that they intended to vote for the Republican running in the election.

(a) According to this study, the estimate for $p$, the proportion of all "likely voters" who will vote for the Republican candidate, is $p = 0.601$. Find a bound for the error of estimation.

(b) If the "likely voters" are representative of those who will actually vote, do you think that the Republican candidate will be elected? Why? How confident are you in your decision?

(c) Can you think of reasons that those polled might not be representative of those who actually vote in the election?

*Solution* For (a): We see that $\hat{p} = Y/n = 0.601$. We are asked to place a bound on the error of estimation. We first calculate the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{(0.601)(1-0.601)}{985}} \approx 0.0156.$$

Then since the sample size is large it is approximately normal so we can palce a 2-standard-error bound which is approximately %95 confidence. That is

$$b = 2\sigma_{\hat{p}} = 2 \cdot 0.0156 = 0.0312.$$

Thus the bound on the error for estimation is $0.601 \pm 0.0312$. For (b): The estimate is 60.1%, with a margin of error $\pm 3.12\%$. This gives a confidence interval of approximately $[56.9\%, 63.2\%]$. Since the entire interval is above 50%, we can be reasonably confident (about 95%) that the Republican candidate would win if the "likely voters" truly represent actual voters. Some potential sources of this are:

- Sampling bias: The poll was conducted by the Republican Party, which may attract more Republican-leaning respondents.
- Nonresponse bias: Those who answer political polls may not reflect the general population.
- Social desirability bias: Respondents may report what they think is expected rather than their true intent.
- "Likely voter" error: People who say they will vote may not actually do so.
- Time lag: Voter intent can change before the election.
- Coverage error: People without landlines or who are less reachable may be underrepresented.

□

**8.28**

In a study of the relationship between birth order and college success, an investigator found that 126 in a sample of 180 college graduates were firstborn or only children; in a sample of 100 nongraduates of comparable age and socioeconomic background, the number of firstborn or only children was 54. Estimate the difference in the proportions of firstborn or only children for the two populations from which these samples were drawn. Give a bound for the error of estimation

*Solution* We are given that the proportion of firstborn or only chilren in the sample of college graduates were $\hat{p}_1 = 126/180 = 0.7$. We were also given that the proportion of firstborn or only chilren in the nongraduates were $\hat{p}_2 = 54/100 = 0.54$. We are asked to estimate $\hat{p_1} - \hat{p}_2 = 0.16$. We then find an bound for the error of estimation by first finding the standard error.

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{(0.7)(1-0.7)}{180} + \frac{(0.54)(1-0.54)}{100}} = 0.060.$$

Thus the error of estimation is then

$$b = 2 \cdot \sigma_{\hat{p_1}} = 2 \cdot 0.060 = 0.1208.$$

Thus the bound on the error of estimation is $0.16 \pm 0.1208$. □

Problems 8.29 - 8.33 are skipped as they are go over the same thing as these previous questions. I encourage you to take a look at Problem 8.32-8.33.

**8.34**

We can place a 2-standard-deviation bound on the error of estimation with any estimator for which we can find a reasonable estimate of the standard error. Suppose that $Y_1, Y_2, \ldots, Y_n$ represent a random sample from a Poisson distribution with mean $\lambda$. We know that $V(Y_i) = \lambda$, and hence $E(\bar{Y}) = \lambda$ and $V(\bar{Y}) = \lambda/n$. How would you employ $Y_1, Y_2, \ldots, Y_n$ to estimate $\lambda$? How would you estimate the standard error of your estimator?

*Solution* To estimate $\lambda$ we can use the sample mean $\overline{Y}$. We would estimate the standard error of our estimator by

$$\sigma_{\overline{Y}} = \sqrt{\frac{\lambda}{n}} = \sqrt{\frac{\overline{Y}}{n}}.$$

□

**8.35**

Refer to Problem 8.34. In polycrystalline aluminum, the number of grain nucleation sites per unit volume is modeled as having a Poisson distribution with mean $\lambda$. Fifty unit-volume test specimens subjected to annealing under regime A produced an average of 20 sites per unit volume. Fifty independently selected unit-volume test specimens subjected to annealing regime B produced an average of 23 sites per unit volume.

(a) Estimate the mean number $\lambda_A$ of nucleation sites for regime A and place a 2-standard-error bound on the error of estimation.

(b) Estimate the difference in the mean numbers of nucleation sites $\lambda_A - \lambda_B$ for regimes A and B. Place a 2-standard-error bound on the error of estimation. Would you say that regime B tends to produce a larger mean number of nucleation sites? Why?

*Solution* For (a): We can estimate the mean number $\lambda_A$ with the sample mean $\overline{Y}_A$. We see that $\overline{Y}_A = 20$. The standard error is

$$\sigma_{\overline{Y}_A} = \sqrt{\frac{20}{50}} = 0.6324.$$

Thus we have that

$$b = 2 \cdot \sigma_{\overline{Y}_A} = 2 \cdot 0.6324 = 1.2648.$$

Thus the bound on the error of estimation is $20 \pm 1.2648$. For (b): We can estimate the mean difference $\lambda_A - \lambda_B$ with the sample mean difference $\overline{Y}_A - \overline{Y}_B$. We see that $\overline{Y}_A - \overline{Y}_B = 20 - 23 = -3$. The standard error is

$$\sigma_{\overline{Y}_A - \overline{Y}_B} = \sqrt{\frac{\overline{Y}_A}{n_A} + \frac{\overline{Y}_B}{n_B}} = \sqrt{\frac{20}{50} + \frac{23}{50}} = 0.9274.$$

Thus we have that

$$b = 2 \cdot \sigma_{\overline{Y}_A - \overline{Y}_B} = 2 \cdot 0.9274 = 1.8548.$$

The bound on the error of estimation is $-3 \pm 1.8548$. Since the confidence interval is $[-4.8548, -1.1452]$, which does not include 0, we can conclude with approximately 95% confidence that regime B tends to produce a larger mean number of nucleation sites than regime A. $\square$

**8.39**

Suppose that the random variable $Y$ has a gamma distribution with parameters $\alpha = 2$ and an unknown $\beta$. In Exercise 6.46, you used the method of moment-generating functions to prove a general result implying that $2Y/\beta$ has a $\chi^2$ distribution with 4 degrees of freedom (df). Using $2Y/\beta$ as a pivotal quantity, derive a 90% confidence interval for $\beta$.

*Solution* Using $2Y/\beta$ as a pivotal qunatity we derive a 90% confidence interval for $\beta$. We see that

$$P\left(\chi^2_{1-\alpha/2} \leq \frac{2Y}{\beta} \leq \chi^2_{\alpha/2]}\right) = 0.90.$$

Solving for $\beta$, we get:

$$P\left(\frac{2Y}{\chi^2_{\alpha/2}} \leq \beta \leq \frac{2Y}{\chi^2_{1-\alpha/2}}\right) = 0.90.$$

Thus, the 90% confidence interval for $\beta$ is:

$$\left[\frac{2Y}{\chi^2_{\alpha/2}}, \frac{2Y}{\chi^2_{1-\alpha/2}}\right].$$

□

**8.40**

Suppose that the random variable $Y$ is an observation from a normal distribution with unknown mean $\mu$ and variance 1. Find

1. a 95% confidence interval for $\mu$.
2. a 95% upper confidence limit for $\mu$.
3. a 95% lower confidence limit for $\mu$.

*Solution* For (a): Since $Y \sim N(\mu, 1)$, we can use the pivotal quantity $Z = \frac{Y-\mu}{1} \sim N(0,1)$. To construct a 95% confidence interval, we find the critical values $-z_{0.025}$ and $z_{0.025}$ such that $P(-z_{0.025} \leq Z \leq z_{0.025}) = 0.95$. Solving for $\mu$, we get:

$$P(Y - z_{0.025} \leq \mu \leq Y + z_{0.025}) = 0.95.$$

Thus, the 95% confidence interval for $\mu$ is:

$$[Y - z_{0.025}, Y + z_{0.025}].$$

For (b): To find a 95% upper confidence limit for $\mu$, we use the pivotal quantity $Z = \frac{Y-\mu}{1} \sim N(0,1)$. We find $z_{0.05}$ such that $P(Z \leq z_{0.05}) = 0.95$. Solving for $\mu$, we get:

$$P(\mu \leq Y + z_{0.05}) = 0.95.$$

Thus, the 95% upper confidence limit for $\mu$ is:

$$Y + z_{0.05}.$$

For (c): To find a 95% lower confidence limit for $\mu$, we use the pivotal quantity $Z = \frac{Y-\mu}{1} \sim N(0,1)$. We find $-z_{0.05}$ such that $P(Z \geq -z_{0.05}) = 0.95$. Solving for $\mu$, we get:

$$P(\mu \geq Y - z_{0.05}) = 0.95.$$

Thus, the 95% lower confidence limit for $\mu$ is:

$$Y - z_{0.05}.$$

□

**8.43**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from a population with a uniform distribution on the interval $(0, \theta)$. Let $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ and $U = (1/\theta)Y_{(n)}$.

a) Show that $U$ has distribution function

$$F_U(u) = \begin{cases} 0, & u < 0, \\ u^n, & 0 \le u \le 1, \\ 1, & u > 1. \end{cases}$$

b) Because the distribution of $U$ does not depend on $\theta$, $U$ is a pivotal quantity. Find a 95% lower confidence bound for $\theta$.

*Solution* For (a): To find the distribution function $F_U(u)$, we note that $U = \frac{Y_{(n)}}{\theta}$, where $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$. The cumulative distribution function (CDF) of $Y_{(n)}$ is given by:

$$F_{Y_{(n)}}(y) = P(Y_{(n)} \le y) = P(Y_1 \le y, Y_2 \le y, \ldots, Y_n \le y) = [F_Y(y)]^n,$$

where $F_Y(y)$ is the CDF of the uniform distribution on $(0, \theta)$, given by:

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ \frac{y}{\theta}, & 0 \le y \le \theta, \\ 1, & y > \theta. \end{cases}$$

Substituting $F_Y(y) = \frac{y}{\theta}$ for $0 \le y \le \theta$, we get:

$$F_{Y_{(n)}}(y) = \left(\frac{y}{\theta}\right)^n, \quad 0 \le y \le \theta.$$

Now, since $U = \frac{Y_{(n)}}{\theta}$, we have:

$$F_U(u) = P(U \le u) = P\left(\frac{Y_{(n)}}{\theta} \le u\right) = P(Y_{(n)} \le u\theta).$$

Using the CDF of $Y_{(n)}$, we get:

$$F_U(u) = \begin{cases} 0, & u < 0, \\ u^n, & 0 \le u \le 1, \\ 1, & u > 1. \end{cases}$$

For (b): To find a 95% lower confidence bound for $\theta$, we use the pivotal quantity $U = \frac{Y_{(n)}}{\theta}$. From part (a), the CDF of $U$ is $F_U(u) = u^n$ for $0 \le u \le 1$. We need to find $u$ such that $P(U \le u) = 0.95$, which corresponds to $u = 0.95^{1/n}$.

Thus:

$$P\left(U \ge 0.95^{1/n}\right) = 0.05.$$

Substituting $U = \frac{Y_{(n)}}{\theta}$, we get:

$$P\left(\frac{Y_{(n)}}{\theta} \ge 0.95^{1/n}\right) = 0.05.$$

Rearranging for $\theta$, we find:

$$P\left(\theta \leq \frac{Y_{(n)}}{0.95^{1/n}}\right) = 0.05.$$

Therefore, the 95% lower confidence bound for $\theta$ is:

$$\frac{Y_{(n)}}{0.95^{1/n}}.$$

□

# 4 Properties of Point Estimators and Methods of Estimation

## 4.1 Introduction

In this chapter, we delve into the mathematical properties of point estimators, focusing on their efficiency and consistency. We discuss techniques for finding minimum-variance unbiased estimators and introduce two widely used approaches: the method of moments and the method of maximum likelihood. These methods, along with their properties, are examined in detail to provide a comprehensive understanding of estimation in statistical analysis. Note that Chapter 4 covers sections 9.1, 9.2, 9.3, 9.6, 9.7 from the textbook and sections 9.4 (sufficiency), 9.5 (Rao-Blackwell theorem) are done in chapter 10.

## 4.2 Relative Efficiency

As we seen in the previous chapter, there can be more than one unbiased estimator for a target parameter. Thus we mentioned (Section 3.2 Figure 3) that given two unbiasded estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ for the same target parameter $\theta$, we prefer to use the estimator with a smaller variance. We see that $\hat{\theta}_1$ is Relatively more efficient than $\hat{\theta}_2$ if $\text{Var}(\hat{\theta}_2) > \text{Var}(\hat{\theta}_1)$. We then can use the ratio $\text{Var}(\hat{\theta}_2)/\text{Var}(\hat{\theta}_1)$ to define the Relative efficiency of two unbiased estimators.

**Definition 4.2.1**

The relative efficiency of two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of a parameter $\theta$, denoted as $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$, is defined as:

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}.$$

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators for $\theta$, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$, $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$, is greater than 1 only if $\text{Var}(\hat{\theta}_2) > \text{Var}(\hat{\theta}_1)$. In this case, $\hat{\theta}_1$ is a better unbiased estimator than $\hat{\theta}_2$. For example, if $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = 1.8$, then $\text{Var}(\hat{\theta}_2) = 1.8\,\text{Var}(\hat{\theta}_1)$, and $\hat{\theta}_1$ is preferred to $\hat{\theta}_2$. Similarly, if $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$ is less than 1—say, 0.73—then $\text{Var}(\hat{\theta}_2) = 0.73\,\text{Var}(\hat{\theta}_1)$, and $\hat{\theta}_2$ is preferred to $\hat{\theta}_1$. We illustrate this in the following examples.

**Example 4.2.2**

Let $Y_1, Y_2, \dots, Y_n$ denote a random sample from the uniform distribution on the interval $(0, \theta)$. Two unbiased estimators for $\theta$ are

$$\hat{\theta}_1 = 2\overline{Y} \quad \text{and} \quad \hat{\theta}_2 = \left(\frac{n+1}{n}\right) Y_{(n)},$$

where $Y_{(n)} = \max(Y_1, Y_2, \dots, Y_n)$. Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

*Solution* From Definition 4.2.1 we see that the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is given by

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_1)}{\text{Var}(\hat{\theta}_1)}.$$

We see that $\text{Var}(\hat{\theta}_1) = 4 \cdot \frac{\sigma^2}{n}$.

$$\text{Var}(\hat{\theta}_1) = 4 \cdot \frac{\sigma^2}{n} = 4 \cdot \frac{\theta^2/12}{n} = \frac{\theta^2}{3n}.$$

Next we see that

$$\text{Var}(\hat{\theta}_2) = \left(\frac{n+1}{n}\right)^2 \text{Var}(Y_{(n)}).$$

We find $\text{Var}(Y_{(n)}) = E(Y_{(n)}^2) - [E(Y_{(n)})]^2 = E(Y_{(n)}^2) - \theta^2$. Notice that

$$E(Y_{(n)}^2) = \frac{n}{\theta^n} \int_0^\theta y^{n+1} dy = \frac{n}{n+2} \cdot \theta^2$$

Thus we find that

$$\text{Var}(Y_{(n)}) = \frac{n\theta^{n+1}}{n+2} - \theta^2 = \left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right] \theta^2$$

Thus then

$$\begin{aligned} \text{Var}(\hat{\theta}_2) = \text{Var}\left[\left(\frac{n+1}{n}\right) Y_{(n)}\right] &= \left(\frac{n+1}{n}\right)^2 \text{Var}(Y_{(n)}) \\ &= \left[\frac{(n+1)^2}{n(n+2)} - 1\right] \theta^2 = \frac{\theta^2}{n(n+2)}. \end{aligned}$$

Therefore, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is given by

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} = \frac{\theta^2/[n(n+2)]}{\theta^2/3n} = \frac{3}{n+2}.$$

We see that the efficiency is less than 1 if $n > 1$. That means that the efficiency for $\hat{\theta}_2$ generally has a smaller variance than $\hat{\theta}_1$ and so is a preferred estimator to $\hat{\theta}_1$. □

### 4.3 Consistency

Before we talk about consistency recall from STA256 the topics of Convergence in Probability and Distribution. We said that given some estimator for some target parameter, as your sample size $n$ grows large, the distribution of your estimator (like a sample mean, sample proportion, etc.) often approaches a known distribution (like the normal distribution). This is what allows use tools like Z-scores for for distributions such as the sample mean even when the underlying distribution isn't normal (Central Limit Theorem). For example lets say we wanted to estimate the parameter $p$ for a binomial distribution $Y$. We learned that $\hat{p} = Y/n$ forms an unbiased estimator for $p$. Intuitvely as $n$ gets larger then $\hat{p}$ should get closer and closer to the true value of $p$. If this happens we then can say that $\hat{p}$ is a consistent estimator of $p$.

**Definition 4.3.1**

The estimator $\hat{\theta}_n$ is said to be a consistent estimator of $\theta$ if, for all $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| < \epsilon) = 1.$$

or equivently

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| > \epsilon) = 0.$$

That is if $\hat{\theta}_n$ converges in probability to $\theta$ then it is a consistent estimator.

We then can use the following theorem to show that an unbiased estimator is consistent.

**Theorem 4.3.2**

An unbiased estimator $\hat{\theta}_n$ for $\theta$ is consistent estimator of $\theta$ if

$$\lim_{n\to\infty} \text{Var}(\hat{\theta}_n) = 0.$$

*Proof.* Since $\hat{\theta}_n$ is unbiased then $E(\hat{\theta}_n) = \theta$. Let $\text{Var}(\hat{\theta}_n) = \sigma^2$. Using Chebyshev's inequality we have that

$$P(|\hat{\theta}_n - \theta| > k\sigma) \leq \frac{1}{k^2}.$$

Let $\epsilon > 0$ be given. Let $k = \epsilon/\sigma > 0$ we have that then

$$P(|\hat{\theta}_n - \theta| > \epsilon) \leq \frac{\sigma^2}{\epsilon} = \frac{\text{Var}(\hat{\theta}_n)}{\epsilon}.$$

Thus for any fixed $n$ we have that

$$0 \leq P(|\hat{\theta}_n - \theta| > \epsilon) \leq \frac{\text{Var}(\hat{\theta}_n)}{\epsilon}.$$

Taking the limit as $n \to \infty$ we have that

$$0 \leq \lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) \leq \lim_{n\to\infty} \frac{\text{Var}(\hat{\theta}_n)}{\epsilon} = \frac{0}{\epsilon} = 0.$$

Thus $\hat{\theta}_n$ is a consistent estimator of $\theta$. □

**Example 4.3.3**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a distribution with mean $\mu$ and variance $\sigma^2 < \infty$. Show that $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$ is a consistent estimator of $\mu$.

*Solution* We first notice that $\overline{Y}_n$ is an unbiased estimator of $\mu$. Moreover we also know that $\text{Var}(\overline{Y}_n) = \sigma^2/n$. Thus using Theorem 4.3.2 we have that

$$\lim_{n\to\infty} \text{Var}(\overline{Y}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

Thus $\overline{Y}$ is a consistent estimator of $\mu$. Recall from STA256 Notes (Theorem 4.1.2) that this was also called the Week Law of Large Numbers. This theorem/result provides experimenters the theoretical justification for the averaging proccess employed by many experimenters. That is the average of many independent observations should be close to the true population mean with high probability. For example if you want to estimate the average weight of a certain animal, you might instead measure take the average of many animals to get a higher precision. □

Recall from the previous chapter that we introduced some common unbiased point estimators such as the sample mean difference $\overline{Y}_1 - \overline{Y}_2$. The following theorem will be important to help us confirm the consistency of those estimators.

**Theorem 4.3.4**

Suppose that $\hat{\theta}_n$ converges in probability to $\theta$ and that $\hat{\theta}'_n$ converges in probability to $\theta'$.

a. $\hat{\theta}_n + \hat{\theta}'_n$ converges in probability to $\theta + \theta'$.

b. $\hat{\theta}_n \times \hat{\theta}'_n$ converges in probability to $\theta \times \theta'$.

c. If $\theta' \neq 0$, $\hat{\theta}_n / \hat{\theta}'_n$ converges in probability to $\theta/\theta'$.

d. If $g(\cdot)$ is a real-valued function that is continuous at $\theta$, then $g(\hat{\theta}_n)$ converges in probability to $g(\theta)$.

*Proof.* Let $\epsilon > 0$ be given. Since $\hat{\theta}_n \xrightarrow{P} \theta$ and $\hat{\theta}'_n \xrightarrow{P} \theta'$, for any $\epsilon > 0$, there exist $N_1$ and $N_2$ such that for all $n \geq N_1$ and $n \geq N_2$, we have

$$P(|\hat{\theta}_n - \theta| < \epsilon/2) > 1 - \epsilon/2 \quad \text{and} \quad P(|\hat{\theta}'_n - \theta'| < \epsilon/2) > 1 - \epsilon/2.$$

Now, consider the sum $\hat{\theta}_n + \hat{\theta}'_n$. Using the triangle inequality, we have

$$|\hat{\theta}_n + \hat{\theta}'_n - (\theta + \theta')| \leq |\hat{\theta}_n - \theta| + |\hat{\theta}'_n - \theta'|.$$

Thus,

$$P(|\hat{\theta}_n + \hat{\theta}'_n - (\theta + \theta')| < \epsilon) \geq P(|\hat{\theta}_n - \theta| < \epsilon/2 \text{ and } |\hat{\theta}'_n - \theta'| < \epsilon/2).$$

By the union bound, we have

$$P(|\hat{\theta}_n - \theta| < \epsilon/2 \text{ and } |\hat{\theta}'_n - \theta'| < \epsilon/2) \geq 1 - P(|\hat{\theta}_n - \theta| \geq \epsilon/2) - P(|\hat{\theta}'_n - \theta'| \geq \epsilon/2).$$

Since $P(|\hat{\theta}_n - \theta| \geq \epsilon/2) < \epsilon/2$ and $P(|\hat{\theta}'_n - \theta'| \geq \epsilon/2) < \epsilon/2$ for $n \geq \max(N_1, N_2)$, it follows that

$$P(|\hat{\theta}_n + \hat{\theta}'_n - (\theta + \theta')| < \epsilon) > 1 - \epsilon.$$

Hence, $\hat{\theta}_n + \hat{\theta}'_n \xrightarrow{P} \theta + \theta'$.

For part (b), consider the product $\hat{\theta}_n \times \hat{\theta}'_n$. Using the definition of convergence in probability, we have

$$|\hat{\theta}_n\hat{\theta}'_n - \theta\theta'| = |\hat{\theta}_n\hat{\theta}'_n - \hat{\theta}_n\theta' + \hat{\theta}_n\theta' - \theta\theta'| \leq |\hat{\theta}_n||\hat{\theta}'_n - \theta'| + |\theta'||\hat{\theta}_n - \theta|.$$

Let $\epsilon > 0$. Since $\hat{\theta}_n \xrightarrow{P} \theta$ and $\hat{\theta}'_n \xrightarrow{P} \theta'$, for any $\epsilon > 0$, there exist $N_1$ and $N_2$ such that for all $n \geq N_1$ and $n \geq N_2$, we have

$$P(|\hat{\theta}_n - \theta| < \epsilon/(2|\theta'|)) > 1 - \epsilon/2 \quad \text{and} \quad P(|\hat{\theta}'_n - \theta'| < \epsilon/(2|\theta|)) > 1 - \epsilon/2.$$

Thus, for $n \geq \max(N_1, N_2)$, we have

$$P(|\hat{\theta}_n\hat{\theta}'_n - \theta\theta'| < \epsilon) > 1 - \epsilon.$$

Hence, $\hat{\theta}_n\hat{\theta}'_n \xrightarrow{P} \theta\theta'$.

For part (c), consider the ratio $\hat{\theta}_n/\hat{\theta}'_n$. Since $\hat{\theta}'_n \xrightarrow{P} \theta'$ and $\theta' \neq 0$, there exists $N_3$ such that for all $n \geq N_3$, $|\hat{\theta}'_n| > |\theta'|/2$ with probability greater than $1 - \epsilon/2$. For $n \geq \max(N_1, N_3)$, we have

$$\left|\frac{\hat{\theta}_n}{\hat{\theta}'_n} - \frac{\theta}{\theta'}\right| = \left|\frac{\hat{\theta}_n\theta' - \theta\hat{\theta}'_n}{\hat{\theta}'_n\theta'}\right| \leq \frac{|\hat{\theta}_n - \theta|}{|\hat{\theta}'_n|} + \frac{|\theta||\hat{\theta}'_n - \theta'|}{|\hat{\theta}'_n\theta'|}.$$

Since $|\hat{\theta}'_n| > |\theta'|/2$ and $\hat{\theta}_n, \hat{\theta}'_n \xrightarrow{P} \theta, \theta'$, it follows that

$$P\left(\left|\frac{\hat{\theta}_n}{\hat{\theta}'_n} - \frac{\theta}{\theta'}\right| < \epsilon\right) > 1 - \epsilon.$$

Hence, $\hat{\theta}_n/\hat{\theta}'_n \xrightarrow{P} \theta/\theta'$.

For part (d), let $g(\cdot)$ be a real-valued function that is continuous at $\theta$. Since $\hat{\theta}_n \xrightarrow{P} \theta$, for any $\epsilon > 0$, there exists $N_4$ such that for all $n \geq N_4$, $P(|\hat{\theta}_n - \theta| < \delta) > 1 - \epsilon$, where $\delta > 0$ is chosen such that $|g(\hat{\theta}_n) - g(\theta)| < \epsilon$ whenever $|\hat{\theta}_n - \theta| < \delta$. Thus,

$$P(|g(\hat{\theta}_n) - g(\theta)| < \epsilon) > 1 - \epsilon.$$

Hence, $g(\hat{\theta}_n) \xrightarrow{P} g(\theta)$. □

**Example 4.3.5**

Suppose that $Y_1, Y_2, \ldots, Y_n$ represent a random sample such that $E(Y_i) = \mu$, $E(Y_i^2) = \mu'_2$ and $E(Y_i^4) = \mu'_4$ are all finite. Show that

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$$

is a consistent estimator of $\sigma^2 = \mathrm{V}(Y_i)$. *(Note: We use subscript $n$ on both $S^2$ and $\overline{Y}$ to explicitly convey their dependence on the value of the sample size $n$.)*

*Solution* To show that $S_n^2$ is a consistent estimator of $\sigma^2$ we use the following identity:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i^2 - \overline{Y}_n^2\right).$$

The statistic $(1/n)\sum_{i=1}^{n} Y_i^2$ is the average of $n$ independent and identically distributed random variables, with $E(Y_i^2) = \mu_2'$ and $V(Y_i^2) = \mu_4' - (\mu_2')^2 < \infty$. By the law of large numbers (Example 4.3.3), we know that $(1/n)\sum_{i=1}^{n} Y_i^2$ converges in probability to $\mu_2'$.

Example 9.2 also implies that $\overline{Y}_n$ converges in probability to $\mu$. Because the function $g(x) = x^2$ is continuous for all finite values of $x$, Theorem 4.3.4(d) implies that $\overline{Y}_n^2$ converges in probability to $\mu^2$. It then follows from Theorem 4.3.4(a) that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2$$

converges in probability to $\mu_2' - \mu^2 = \sigma^2$. Because $n/(n-1)$ is a sequence of constants converging to 1 as $n \to \infty$, we can conclude that $S_n^2$ converges in probability to $\sigma^2$. Equivalently, $S_n^2$, the sample variance, is a consistent estimator for $\sigma^2$, the population variance. □

In the previous chapter, we explored large-sample confidence intervals for estimating parameters of practical interest. Specifically, for a random sample $Y_1, Y_2, \ldots, Y_n$ from a distribution with mean $\mu$ and variance $\sigma^2$, we established that the interval $\overline{Y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ provides an approximate $(1-\alpha)$ confidence level. If $\sigma^2$ is unknown but the sample size is large, substituting the sample standard deviation $S$ for $\sigma$ introduces negligible loss of accuracy. This substitution is justified theoretically, as shown in the subsequent theorem.

**Theorem 4.3.6**

Suppose that $U_n$ has a distribution function that converges to a standard normal distribution function as $n \to \infty$. If $W_n$ converges in probability to 1, then the distribution function of $U_n/W_n$ converges to a standard normal distribution function.

*Proof.* This result follows from the general result known as Slutsky's theorem which was introduced in STA256 Notes. The proof for this theorem was omitted because I didn't feel like proving it. □

**Example 4.3.7**

Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size $n$ from a distribution with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$. Define $S_n^2$ as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2.$$

Show that the distribution function of

$$\sqrt{n}\left(\frac{\overline{Y}_n - \mu}{S_n}\right)$$

converges to a standard normal distribution function.

*Solution* From the Central Limit Theorem we know that the distribution

$$U_n = \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right)$$

converges in distribution to the standard normal distribution. Moreover sice $S_n^2$ converges in probability to $\sigma^2$ and $g(x) = \pm\sqrt{x/c}$ is a continuous function for postive values of $x$ and $c$ then $S_n/\sigma = \pm\sqrt{S_n^2/\sigma^2}$ converges in probability to 1. Thus Theorem 4.3.6 implies that

$$\sqrt{n}\left(\frac{\overline{Y}-\mu}{\sigma}\right)/(S_n/\sigma) = \sqrt{n}\left(\frac{\overline{Y}_n-\mu}{S_n}\right)$$

converges in distribution to the standard normal distribution. □

We learned that the $\sqrt{n}\left(\frac{\overline{Y}_n-\mu}{S_n}\right)$ has a $t-$distribution with $n-1$ df. We also just learned that when the sample size is large then the distribution is approximately normal. That is That is, as $n$ gets large and hence as the number of degrees of freedom gets large, the $t$-distribution function converges to the standard normal distribution function.

## 4.4 The Method of Moments

We now discusss a method for finding point estimators: method of moments. This method follows a simple procedure.

Recall that the $k$-th moment of a random variable is given by $E(Y^k) = \mu_k$. We also learned that the sample mean is an unbiased estimator of the first moment. That is $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ is the first sample moment that gives us a good theoretical approximation of the true first moment $E(Y) = \mu_1$. More generally

$$m_k = \frac{1}{n}\sum_{i=1}^n Y_i^k$$

is the $k$-th sample moment. The intuitve idea here is that the sample moments provide good estimates of the corresponding population moments. That is $m_k$ is a good estimator for $\mu_k$ for $k = 1, 2, \ldots$. Then because the population moments $\mu_1, \mu_2, \ldots, \mu_k$ are functions of the population parameters, we can equate corresponding population and sample moments and solve for the desired estimators. We illustrate this in the following examples.

**Example 4.4.1**

A random sample of $n$ observations, $Y_1, Y_2, \ldots, Y_n$, is selected from a population in which $Y_i$, for $i = 1, 2, \ldots, n$, possesses a uniform probability density function over the interval $(0, \theta)$ where $\theta$ is unknown. Use the method of moments to estimate the parameter $\theta$.

*Solution* Using the method of moments what we do is we find $m_1 = \frac{1}{n}\sum_{i=1}^n Y_i$ and equate it to $\mu_1$. We see that $\mu_1 = \mu = \theta/2$. Thus we have that corresponding first sample moment is

$$m_1 = \frac{1}{n}\sum_{i=1}^n Y_i = \overline{Y}.$$

Equating the population and sample moment we have that

$$\mu_1 = \frac{\theta}{2} = \overline{Y}.$$

Solving for $\theta$ we have that the method of moments estimator is $\hat{\theta} = 2\overline{Y}$. □

For the distributions that we consider in this text, the methods of Section 4.3 can be used to show that sample moments are consistent estimators of the corresponding population moments.

Because the estimators obtained from the method of moments obviously are functions of the sample moments, estimators obtained using the method of moments are usually consistent estimators of their respective parameters.

**Example 4.4.2**

Show that the estimator $\hat{\theta} = 2\overline{Y}$ is a consistent estimator of $\theta$.

*Solution* Clearly this estimator is unbiased. Next notice that $\text{Var}(\hat{\theta}) = \theta^2/3n$. Then since $\lim_{n\to\infty} \text{Var}(\hat{\theta}) = 0$ we have that it is a consistent estimator of $\theta$. □

**Example 4.4.3**

A random sample of $n$ observations, $Y_1, Y_2, \ldots, Y_n$, is selected from a population where $Y_i$, for $i = 1, 2, \ldots, n$, possesses a gamma probability density function with parameters $\alpha$ and $\beta$ (see Section 4.6 for the gamma probability density function). Find method-of-moments estimators for the unknown parameters $\alpha$ and $\beta$.

*Solution* Since we seek to find two parameters $\alpha$ and $\beta$ we must equate two population and sample moments. That is $\mu_1 = m_1$ and $\mu_2 = m_2$. We first see that $\mu_1 = \mu = \alpha\beta$. And then $\mu_2 = E(Y^2) = \text{Var}(Y) + [E(Y)]^2 = \alpha\beta^2 + \alpha^2\beta^2$. We then equate these quantites to their corresponding sample moments and solve for $\hat{\alpha}$ and $\hat{\beta}$. We see that

$$\mu_1 = \alpha\beta = m_1 = \overline{Y}$$

$$\mu_2 = \alpha\beta^2 + \alpha^2\beta^2 = m_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2.$$

From the first equation we have $\hat{\beta} = \overline{Y}/\alpha$. substituting this into the second equation and solving for $\alpha$ we find that

$$\hat{\alpha} = \frac{\overline{Y}^2}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2} = \frac{n\overline{Y}^2}{\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2}.$$

Plugging this back into $\hat{\beta}$ we see that

$$\hat{\beta} = \frac{\overline{Y}}{\alpha} = \frac{\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2}{n\overline{Y}}.$$

□

It is easy to show that the estimators of $\alpha$ and $\beta$ from Example 4.4.3 are consistent. Using the fact that $\overline{Y}$ converges in probability to $\alpha\beta$ and $\frac{1}{n}\sum_{i=1}^{n} Y_i^2$ converges in probability to $\alpha\beta^2 + \alpha^2\beta^2$ we have that

$$\hat{\alpha} = \frac{\overline{Y}^2}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2} \quad \text{is a consistent estimator of} \quad \frac{(\alpha\beta)^2}{\alpha\beta^2 + \alpha^2\beta^2 - (\alpha\beta)^2} = \alpha,$$

and

$$\hat{\beta} = \frac{\overline{Y}}{\hat{\alpha}} \quad \text{is a consistent estimator of} \quad \frac{\alpha\beta}{\alpha} = \beta.$$

### 4.5 The Method of Maximum Likelihood

We now go over a method of maximum likelihood (MLE) which is one of the most widely used techniques for estimating the parameters of a statistical model (see section 4.8.1). The core idea is that you pick the parameter(s) that maximize the probability or likelihood (the joint probability function or joint density function) of observing the data you actually observed. I can show you this with an example.

Suppose you have a box with three balls, each either being red or white but you do not know which color they are and how many of that color are in the box. Let's say that we randomly sample two balls without replacement from the box and observe two red balls. We ask ourselves what would be a good estimator of the total number of red balls in the box? Clearly the total number of red balls in the box must either be two or three since we observed two red balls. If there are two red balls and one white ball then the probability of picking two red balls is

$$\frac{\binom{2}{2}\binom{1}{0}}{\binom{3}{2}} = \frac{1}{3}.$$

Similarly, if there are three red balls in the box, the probability of randomly selecting two red balls is

$$\frac{\binom{3}{2}}{\binom{3}{2}} = 1.$$

From this it seems reasonable to choose three as the estimate of the number of red balls in the box even though there might be two red balls in the box since the observed sample gives more confidence in three balls. More formally, the method of maximum likelihood can be stated as follows:

Suppose that the likelihood function depends on $k$ parameters $\theta_1, \theta_2, \ldots, \theta_k$. Choose as estimates those values of the parameters that maximize the likelihood

$$L(y_1, y_2, \ldots, y_n \mid \theta_1, \theta_2, \ldots, \theta_k).$$

We illustrate this method in the following examples.

**Example 4.5.1**

A binomial experiment consisting of $n$ trials resulted in observations $y_1, y_2, \ldots, y_n$, where $y_i = 1$ if the $i$th trial was a success and $y_i = 0$ otherwise. Find the MLE of $p$, the probability of a success.

*Solution* Since each trial is a Bernoulli trial then the pmf is

$$P(Y_i = y_i) = p^{y_i}(1-p)^{1-y_i}.$$

Since the observations are independent then the likelihood of the observed samples gives us

$$L(y_1, , y_2, \ldots, y_n \mid p) = \prod_{i=1}^{n} P(Y_i = y_i) = p^y (1-p)^{n-y}$$

where $y = \sum_{i=1}^{n} y_i$. We now seek to find the value of $p$ that maximizes the likelihood $L(p)$ by going over cases. When $y = 0$ we have $L(p) = (1-p)^n$ which is maximized when $p = 0$. If $y = n$ then we have $L(p) = p^n$ which is maximized when $p = 1$. We now consider the case when $y = 1, 2, \ldots, n-1$. When $p = 0$ or $p = 1$ then $L(p)$ is zero and is otherwise continuous for values of $p$ between 0 and 1. To find the maximized value we can use calculus. Now since our likelihood function is a product it will be messy to differentiate and work with numerically so instead we take the logarithm of the likelihood function instead and find the value of $p$ that maximizes the function. That is $L(p) = \ln[L(p)]$. The reason why the minimum and maximum stays the same is because log is a strictly increasing function so maximizing $L(p)$ gives the same $p$ as maximizing $\ln[L(p)]$. We see that then

$$\begin{aligned}\ln[L(p)] &= \ln[p^y(1-p)^{n-y}] \\ &= \ln(p^y) + \ln[(1-p)^{n-y}] \\ &= y\ln(p) + (n-y)\ln(1-p).\end{aligned}$$

Then differentiataing with respect to $p$ we have that

$$\frac{d\ln[L(p)]}{dp} = \frac{y}{p} - \frac{n-y}{1-p}.$$

Setting this equal to zero and solving for $p$ we get that

$$\hat{p} = \frac{y}{n}$$

is the MLE. That is the MLE is the intuitve unbiased estimator sample fraction which we used in the previous chapter. □

**Example 4.5.2**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$. Find the MLEs of $\mu$ and $\sigma^2$.

*Solution* We begin by finding the MLE of $\mu$. We first find the likelihood function $L(y_1,, y_2, \ldots, y_n \mid \mu, \sigma^2)$. Notice that each observation is normally distributed and independent thus we have that

$$\begin{aligned}L(y_1,, y_2, \ldots, y_n \mid \mu, \sigma^2) &= \prod_{i=1}^{n} P(Y_i = y_i) \\ &= \left\{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y_1-\mu)^2}{2\sigma^2}\right]\right\} \times \cdots \times \left\{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y_n-\mu)^2}{2\sigma^2}\right]\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right].\end{aligned}$$

Like in Exampl 4.5.1 we take the logarithm of $L(\mu, \sigma^2)$ to simplify it. After simplifying we see that

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

We now take the derivative with respect to $\mu$ then $\sigma^2$ and set it equal to zero to get the maximuzed values for $\mu$ and $\sigma^2$ respectively. We get

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

and

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

Setting these derivatives equal to zero and solving simultaneously, we obtain from the first equation

$$\frac{1}{\hat{\sigma}^2} \sum_{i=1}^{n} (y_i - \hat{\mu}) = 0, \quad \text{or} \quad \sum_{i=1}^{n} y_i - n\hat{\mu} = 0, \quad \text{and} \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i = \overline{Y}.$$

substituting $\overline{Y}$ for $\mu$ in the second equation and solving for $\hat{\sigma}^2$ we get that

$$-\left(\frac{n}{\hat{\sigma}^2}\right) + \frac{1}{\hat{\sigma}^4} \sum_{i=1}^{n} (y_i - \overline{y})^2 = 0, \quad \text{or} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2.$$

Thus we have that $\overline{Y}$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2$ are the MLE for $\mu$ and $\sigma^2$ respectively. Also notice that $\overline{Y}$ is unbiased but $\hat{\sigma}^2$ is not however it can easily be turned into unbiased estimator $S^2$. $\square$

**Example 4.5.3**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of observations from a uniform distribution with probability density function

$$f(y_i \mid \theta) = \frac{1}{\theta}, \quad \text{for } 0 \leq y_i \leq \theta \text{ and } i = 1, 2, \ldots, n.$$

Find the MLE of $\theta$.

*Solution* We begin by finding the likelihood function $L(y_1, , y_2, \ldots, y_n \mid \theta)$. We easily see that

$$L(y_1, , y_2, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} P(Y_i = y_i) = \frac{1}{\theta^n}$$

if $0 \leq y_i \leq \theta$ for $i = 1, 2, \ldots, n$. However notice that $L(\theta)$ is a monotonically decreasing function (graph this function to get a better visual understanding) so to make $L(\theta)$ as large as possible we want to make $\theta$ as small as possible. Thus considering the constraint we have with $0 \leq y_i \leq \theta$ and the fact that $L(\theta)$ decreases as $\theta$ increases we can instead find the smallest value of $\theta$ such that $0 \leq y_i \leq \theta$. This is of course the maximum of all observations. That is $\hat{\theta} = \max(Y_1, Y_2, \ldots, Y_n) = Y_{(n)}$ is the MLE for $\theta$. This is not an unbiased estimator but can be made to be unbiased. $\square$

## 4.6 Cramér–Rao Lower Bound and Efficiency

Note that this section is not included in this course textbook but is in the STA256 Notes textbook under section 6.2. In this section we establish an inequality that will tell us the lowest variance that an unbiased estimator can have for a parameter. Before this though we introduce the topic of Fisher Information which is vital in the derivation of this inequality. We then show that the variances of the MLE estimators acheieve this lower bound asymptotically.

Suppose $Y_1, Y_2 \ldots, Y_n$ is a random sample from a population that is distributed with density function $f(y; \theta)$ where $\theta$ is an unknown parameter and that you want to estimate this unknown parameter with an estimator $\hat{\theta}$. We want to know: what is the best possible accuracy (lowest variance) we can get when estimating $\theta$, using an unbiased estimator? This is what the Cramér–Rao Lower Bound (CRLB) tells us. To ensure the validity of the derivation of this inequality, the following regularity conditions are assumed:

1. The support of $f$ does not depend on $\theta$.
2. The pdf $f(y;\theta)$ IS differentiable with respec to $\theta$ and twice differentiable.
3. The integral $\int f(y;\theta)dy$ can be differentiated twice under the integral sign as a function of $\theta$.

We begin with the fact that
$$\int_{-\infty}^{\infty} f(y;\theta)dy = 1.$$
We then differentiate both sides with respect to $\theta$ to get
$$\frac{\partial}{\partial\theta}\left(\int_{-\infty}^{\infty} f(y;\theta)dy\right) = 0.$$
Assuming the regularity condtions hold then we can instead do
$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}[f(y;\theta)]dy = 0$$
Dividing and multiplying by $f(y;\theta)$ on the inside we get
$$\int_{-\infty}^{\infty} \frac{\partial f(y;\theta)/\partial\theta}{f(y;\theta)} f(y;\theta)dy = 0.$$
Then using the fact that
$$\frac{\partial f(y;\theta)}{\partial\theta} = f(y;\theta)\frac{\partial\log[f(y;\theta)]}{\partial\theta}$$
so we have that
$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}[f(y;\theta)]dy = \int_{-\infty}^{\infty} \frac{\partial\log[f(y;\theta)]}{\partial\theta} f(y;\theta)dy = 0.$$
However this is really just the expectation of $\frac{\partial\log[f(y;\theta)]}{\partial\theta}$. That is we can express it as
$$E\left[\frac{\partial\log[f(y;\theta)]}{\partial\theta}\right] = 0.$$

We call $\frac{\partial\log[f(y;\theta)]}{\partial\theta}$ the score function. If we differentiate the integral above again we get that
$$0 = \int_{-\infty}^{\infty} \frac{\partial^2\log f(y;\theta)}{\partial\theta^2} f(y;\theta)\,dy + \int_{-\infty}^{\infty} \frac{\partial\log f(y;\theta)}{\partial\theta}\frac{\partial\log f(y;\theta)}{\partial\theta} f(y;\theta)\,dy.$$

The right hand term is called the Fisher Information and we denote it by $\mathcal{I}(\theta)$. That is
$$\mathcal{I}(\theta) = \int_{-\infty}^{\infty} \frac{\partial\log f(y;\theta)}{\partial\theta}\frac{\partial\log f(y;\theta)}{\partial\theta} f(y;\theta)\,dy = \mathbb{E}\left[\left(\frac{\partial\log f(X;\theta)}{\partial\theta}\right)^2\right].$$
However it is usually easier to compute the left hand side so we can also write
$$\mathcal{I}(\theta) = -E\left[\frac{\partial^2\log f(y;\theta)}{\partial^2\theta}\right].$$
Moreover we can also see that
$$\mathcal{I}(\theta) = \mathrm{Var}\left(\frac{\partial\log f(y;\theta)}{\partial\theta}\right)$$

The Fisher Information measures how much information a random variable (or a sample) carries about an unknown parameter $\theta$. If we look at the curve of $\log[f(y;\theta)]$ and this curve is very peaked, then small changes in $\theta$ lead to big changes in likelihood ; you can estimate $\theta$ very precisely. If it's flat, then even large changes in $\theta$ don't affect the likelihood ; you have little information about $\theta$. So, Fisher Information quantifies the sharpness of the likelihood. A more visual intution is that imagine you're trying to find the peak of a mountain (the MLE). The sharper the peak, the easier it is to know where the true top is. A flat hill makes it hard to pinpoint the top. Fisher Information is like a measure of how sharp the mountain is.

**Example 4.6.1**

Consider a single observation from Bernoulli distribution as follows:

$$Y \sim \text{Bernoulli}(\theta), \quad 0 < \theta < 1$$

Find the Fisher Information $I(\theta)$.

*Solution* The probability mass function of a Bernoulli random variable is

$$f(y;\theta) = \theta^y(1-\theta)^{1-y}$$

where $y = 0, 1$. Then the log-likelihood is

$$\log[f(y;\theta)] = y\log(\theta) + (1-y)\log(1-\theta).$$

Then taking the der with respect to $\theta$ we find that

$$\frac{\partial \log[f(y;\theta)]}{\partial\theta} = \frac{y}{\theta} - \frac{1-y}{1-\theta}.$$

Then we know that the Fisher Information is the expected value of the score function squared:

$$\mathcal{I}(\theta) = E\left[\left(\frac{y}{\theta} - \frac{1-y}{1-\theta}\right)^2\right].$$

When $y = 1$ we get that

$$E\left[\left(\frac{y}{\theta} - \frac{1-y}{1-\theta}\right)^2\right] = \frac{1}{\theta}.$$

When $y = 0$ we get that

$$E\left[\left(\frac{y}{\theta} - \frac{1-y}{1-\theta}\right)^2\right] = \frac{1}{1-\theta}.$$

Adding both contributions we get that

$$\mathcal{I}(\theta) = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}.$$

Fisher Information increases as $\theta$ approaches 0 or 1, meaning the distribution becomes more informative at the extremes. This makes sense because when the probability $\theta$ is near 0.5 then its a unbiased random "flip of a coin" meaning the Fisher Information is minimized at that point. □

We talked about the case where it was a single observation, however what happens to the Fisher Information once we have a sample size of $n$? Suppose we have a sample $Y_1, Y_2, \ldots, Y_n$ that is iid with distribution $f(y;\theta)$. Then it makes sense that each $Y_i$ contributes its own Fisher Information for the common parameter $\theta$. Since they are iid then it makes sense that the total information will just be the sum of all the Fisher Information for each independent observations. To see this we notice that if the pdf for this random sample is $L(\theta)$ then we have that

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

Or instead

$$\log[L(\theta)] = \sum_{i=1}^{n} \log[f(y;\theta)].$$

Then we get that

$$\frac{\partial \log[L(\theta)]}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial \log[f(y;\theta)]}{\partial\theta}.$$

Using independence we get that

$$\text{Var}\left(\frac{\partial \log[L(\theta)]}{\partial\theta}\right) = \sum_{i=1}^{n} \text{Var}\left(\frac{\partial \log[f(y;\theta)]}{\partial\theta}\right) = \sum_{i=1}^{n} \mathcal{I}(\theta) = n\mathcal{I}(\theta).$$

With this in mind we now can establish the CRLB.

**Theorem 4.6.2 : Cramér–Rao Lower Bound**

Let $\hat{\theta}$ be an unbiased estiamtor $\theta$ based on a random sample of size $n$ from a distribution $f(y;\theta)$. Then

$$\text{Var}(\hat{\theta}) \geq \frac{1}{n\mathcal{I}(\theta)}$$

*Proof.* The proof of this theorem is ommited for now since it is quite verbose and my proof requires using external results. □

This theorems tells us that for any unbiased estimator, there is a fundamental limit for how precise the estimator can be. It the recriprocal of the Fisher Information. This in turn means that the more information that a sample carries, the more precise the estimator will become. Thus this implies that as $n \to \infty$ then $1/n\mathcal{I}(\theta) \to 0$ which means $\text{Var}(\hat{\theta}) \geq 0$. That is the estimator can become arbitrarily precise. With this in mind, we then can call estimators that achieve this lower bound efficient.

**Definition 4.6.3**

Let $Y$ be an unbiased estimator of a pa- rameter $\theta$ in the case of point estimation. The statistic $Y$ is called an efficient estimator of $\theta$ if and only if the variance of $Y$ attains the Cramér–Rao lower bound.

If an estimator is efficient then it is the minimum variance unbiased estimator (MVUE) of a parameter $\theta$ which has the smallest possible variance amoung all unbiased estimators of $\theta$. You might be wondering why not just denote any estimator that acheives this lower bound a MVUE and the reason why is because if an estimator is efficient then it is an MVUE, however if an estimator is MVUE it is not nescessarily efficient. That is an estimator can have the lowest variance from all unbiased estimators but it might still not reach the theoretical lower bound set by the Fisher information.

The CRLB can actually be generalized to biased estimators as well

$$\text{Var}(\hat{\theta}) \geq \frac{(1 + B'(\theta))^2}{n\mathcal{I}(\theta)}.$$

Which in the case when the estimator is unbiased we have $B(\theta) = 0$ and so the result of Theorem 4.6.2 follows. Also note that the biased estimators can have lower variances than the CRLB for the unibased estimators but the only cost is the added bias.

**Example 4.6.4**

Let $Y_1, Y_2, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$.
Let $\hat{\theta} = \bar{Y}$.

(a) Find $\text{Var}(\hat{\theta})$.

(b) Find the CR-LB for $\hat{\theta}$.

(c) Is $\hat{\theta}$ an MVUE for $\theta$?

*Solution* For (a): We know that $\text{Var}(\overline{Y}) = \sigma^2/n$. Since the distribution is Possion with parameter $\theta$ then we have that $\text{Var}(\overline{Y}) = \theta/n$. For (b): We now need to find the CRLB for $\theta$. We know that

$$\text{Var}(\overline{Y}) \geq \frac{1}{n\mathcal{I}(\theta)}.$$

So we have to find the fisher information which is given by

$$\mathcal{I}(\theta) = E\left[\left(\frac{\partial \log[f(y;\theta)]}{\partial\theta}\right)^2\right].$$

We find the score function

$$\log[f(y;\theta)] = y\log(\theta) - \log(y!) - \theta.$$

Then we have

$$\frac{\partial \log[f(y;\theta)]}{\partial\theta} = \frac{y}{\theta} - 1 = \frac{y-\theta}{\theta}.$$

We now find the expected value of the score function squared

$$\mathcal{I}(\theta)E\left[\frac{(y-\theta)^2}{\theta^2}\right] = \frac{E[(y-\theta)^2]}{\theta^2} = \frac{\text{Var}(Y)}{\theta^2} = \frac{1}{\theta}.$$

So the CRLB is

$$\frac{\theta}{n}.$$

For (c): Since $\text{Var}(\overline{Y}) = \text{CRLB}$ then it is an efficient estimator and so it is an MVUE for $\theta$. $\square$

**Example 4.6.7**

Let $Y_1, Y_2, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Beta}(\theta, 1), \quad 0 < Y < 1, \quad 0 < \theta < \infty.$

(a) Find the MLE of $\theta$, denoted by $\hat{\theta}_{\text{MLE}}$.

(b) Compute the CR-LB for $\hat{\theta}_{\text{MLE}}$.

(c) Construct an unbiased estimator of $\hat{\theta}_{\text{MLE}}$ and investigate its efficiency.

*Solution* For (a): To fine the MLE we first need to find the likelihood function. Since the sample is distributed $\text{Beta}(\theta, 1)$ then we have that $f(y_i; \theta) = \theta y_i^{\theta-1}$. From this we get that the likelihood function is

$$L(y_1, y_2, \ldots,_y n \mid \theta) = \prod_{i=1}^{n} f(y; \theta) = \theta^n \prod_{i=1}^{n} y_i^{\theta-1}.$$

Then to maximize this function with respect to $\theta$ we use calculus. We find the first derivative of the log-likelihood which is

$$\log[f(y; \theta)] = n \log(\theta) + (\theta - 1) \sum_{i=1}^{n} \log(y_i)$$

We then get

$$\frac{d \log[f(y; \theta)]}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log(y_i).$$

Setting this equal to zero and solving for $\theta$ we get that

$$\frac{n}{\theta} + \sum_{i=1}^{n} \log(y_i) = 0 \quad \text{or} \quad \hat{\theta}_{\text{MLE}} = -\frac{n}{\sum_{i=1}^{n} \log(y_i)}.$$

For (b): To find the CRLB for this estimator we must find the Fisher Information. To do so we find the score function first.

$$\frac{\partial \log[f(y_i; \theta)]}{\partial \theta} = \frac{1}{\theta} + \log(y_i).$$

From here we can find the Fisher information however finding the expected value of the score funtion squared is quite messy so in this case we instead find the second partial derivative with respect to $\theta$ of the score function instead.

$$\frac{\partial^2 \log[f(y_i; \theta)]}{\partial^2 \theta} = -\frac{1}{\theta^2}.$$

Thus we have that

$$\mathcal{I}(\theta) = -E\left(\frac{\partial^2 \log[f(y_i; \theta)]}{\partial^2 \theta}\right) = \frac{1}{\theta^2}.$$

Thus the CRLB is then

$$\text{Var}(\hat{\theta}_{\text{MLE}}) \geq \frac{\theta^2}{n}.$$

For (c): To construct an unbiased estimator of $\hat{\theta}_{\text{MLE}}$ we must find the mean of this estimator which means we have to find the distribution. Notice that since each $y_i \sim \text{Beta}(\theta, 1)$ then $-\log(y_i) \sim \text{Gamma}(1, 1/\theta)$. Thus then $W = -\sum_{i=1}^{n} \log(y_i) \sim \text{Gamma}(n.1/\theta)$. We then see that

$$\hat{\theta}_{\text{MLE}} \sim n[\text{Gamma}(n, 1/\theta)]^{-1}.$$

Thus we have that

$$E(\hat{\theta}_{\text{MLE}}) = E\left[\frac{n}{W}\right] = nE(W^{-1}) = n\theta \cdot \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{n\theta}{n-1}.$$

Where we used the identity for the $k$-th moment of a gamma distribution. Note that $\hat{\theta}_{\text{MLE}}$ is biased so to construct an unbised estimator we let $\hat{\theta} = (n-1)\hat{\theta}_{\text{MLE}}/n$ to get that

$$E(\hat{\theta}) = \frac{n-1}{n}E(\hat{\theta}_{\text{MLE}}) = \theta.$$

Now that we have an unbiased estimator we can investigate weather this estimator is efficient. To do this we find the variance.

$$\text{Var}(\hat{\theta}) = \left(\frac{n-1}{n}\right)^2 \text{Var}(\hat{\theta}_{\text{MLE}})$$

We now just have to find the variance of the MLE estimator. We see that

$$E(\hat{\theta}^2_{\text{MLE}}) = E\left[\frac{n^2}{W^2}\right] = n^2 E[(W)^{-2}] = n^2\theta^2 \cdot \frac{\Gamma(n-2)}{\Gamma(n)} = \frac{n^2}{(n-1)(n-2)} \cdot \theta^2.$$

So then we have that

$$\begin{aligned}
\text{Var}(\hat{\theta}_{\text{MLE}}) &= E(\hat{\theta}^2_{\text{MLE}}) - [E(\hat{\theta}_{\text{MLE}})]^2 \\
&= \frac{n^2}{(n-1)(n-2)} \cdot \theta^2 - \frac{n^2\theta^2}{(n-1)^2} \\
&= \frac{n^2}{(n-1)^2(n-2)}\theta^2.
\end{aligned}$$

Thus then

$$\text{Var}(\hat{\theta}) = \left(\frac{n-1}{n}\right)^2 \text{Var}(\hat{\theta}_{\text{MLE}}) = \left(\frac{n-1}{n}\right)^2 \cdot \frac{n^2}{(n-1)^2(n-2)}\theta^2 = \frac{\theta^2}{n-2} > \frac{\theta^2}{n}.$$

Thus thie unbiased estimator is not an efficient estimator. □

## 4.7 Sufficiency

In this section we talk about another property of statistics (for this section recall Section 2.4 from STA256 Notes). Suppose we collect some sample $Y = (Y_1, Y_2, \ldots, Y_n)$. What we often do is summarize this data with a statistic $G(Y)$ such as the sample mean $\overline{Y}$ or sample variance $S^2$ or the sum $T$. The question we have is after reducing the data into certain statistics like ones above, do we lose any information about the parameter in this process? That is given some statistic $G(Y)$, does this statistic contain all the information in $Y$ about the unknown parameter $\theta$? If the answer is yes then $G(Y)$ is a sufficient statistic for $\theta$. In other words Once you know $G(Y)$, knowing the rest of the data doesn't tell you anything more about $\theta$ and the "leftover randomness" in the data is just noise unrelated to $\theta$. A sufficient statistic can be thought of as a lossless compression of the data. To illustrate this I will use an example. Suppose you have a sample $Y_1, Y_2, \ldots, Y_n \sim \text{Bernoulli}(\theta)$ and you are trying to estimate $\theta$. If I tell you the sum $T = \sum_{i=1}^n Y_i$ and then give you the entire sample $Y_1, Y_2, \ldots, Y_n$, do you learn anything more about $\theta$? No since the statistic tells you how many 1's there are and thats all the data says about $\theta$. Thus we can say that $T$ is a sufficient statistic. How do we determine weather a statistic is sufficient? Well since we want to confirm that after given some statistic $G(Y)$ the rest of the data does not give any information on $\theta$ (that is the distribution for the data just not depend on $\theta$), we can just find weather the condtional distribution of the data given $G(Y)$ is independent from $\theta$.

**Definition 4.7.1**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a probability distribution with unknown parameter $\theta$. Then a statistic $U = g(Y_1, Y_2, \ldots, Y_n)$ is said to be sufficient for $\theta$ if the condtional distribution of $Y_1, Y_2, \ldots, Y_n$ given $U$ does not depend on $\theta$.

For notation we will use $L(y_1, y_2 \ldots, y_n \mid \theta)$ to denote the likelihood function of the sample with respect to the unknown parameter $\theta$. The definition above tells us how to see weather a statistic is sufficient or not, however doing it brute force seems really annoying and for some common distributions might be straight up impossible. Thus we have the following theorem that will help us determine weather a statistic is sufficient. Recall Theorem 2.44 from STA256 notes as this following theorem might be look famlier.

**Theorem 4.7.2 : Factorization Theorem**

Let $U$ be a statistic based on the random sample $Y_1, Y_2, \ldots, Y_n$. Then $U$ is a *sufficient statistic* for the estimation of a parameter $\theta$ if and only if the likelihood $L(\theta) = L(y_1, y_2, \ldots, y_n \mid \theta)$ can be factored into two nonnegative functions,

$$L(y_1, y_2, \ldots, y_n \mid \theta) = g(u, \theta) \times h(y_1, y_2, \ldots, y_n)$$

where $g(u, \theta)$ is a function only of $u$ and $\theta$ and $h(y_1, y_2, \ldots, y_n)$ is not a function of $\theta$.

*Proof.* The proof for this theorem is out of scope. □

**Example 4.7.3**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample in which $Y_i$ possesses the probability density function

$$f(y_i \mid \theta) = \begin{cases} \left(\frac{1}{\theta}\right) e^{-y_i/\theta}, & 0 \leq y_i < \infty, \\ 0, & \text{elsewhere}, \end{cases}$$

where $\theta > 0$, $i = 1, 2, \ldots, n$. Show that $\overline{Y}$ is a sufficient statistic for the parameter $\theta$.

*Solution* We first find the likelihood function of the entire sample which is just the product of the pdf since it is a random sample.

$$L(y_1, y_2 \ldots, y_n \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta}\right) e^{-y_i/\theta} = \frac{1}{\theta^n} e^{-\sum_{i=1}^{n} y_i/\theta} = \frac{1}{\theta^n} e^{-n\overline{Y}/\theta}.$$

Notice that our likelihood function is only a function of $\theta$ and $\overline{Y}$. Thus if we let $g(\theta, \overline{Y}) = \frac{1}{\theta^n} e^{-n\overline{Y}/\theta}$ and $h(y_1, y_2 \ldots, y_n) = 1$ then using Theorem 4.7.2 we have that $\overline{Y}$ is a sufficient statistic for $\theta$. □

**Example 4.7.4**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution over the interval $(0, \theta)$. Show that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for $\theta$.

*Solution* We begin by finding the likelihood function. Since $f(y_i \mid \theta) = \frac{1}{\theta}$ for $0 < y_i < \theta$, we have

$$L(y_1, y_2, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n}$$

if $0 < y_i < \theta$ for all $i$. Otherwise, the likelihood is zero. To ensure it is nonnegative holds we need $\max(y_1, y_2, \ldots y_n) < \theta$. To do this we can use the indicator function which is 1 when this is true and zero otherwise. Thus this can be written as

$$L(y_1, \ldots, y_n \mid \theta) = \frac{1}{\theta^n} \cdot I(Y_{(n)} < \theta),$$

where $I(\cdot)$ is the indicator function. This factors as $g(Y_{(n)}, \theta) = \frac{1}{\theta^n} I(Y_{(n)} < \theta)$ and $h(y_1, \ldots, y_n) = 1$, so by the factorization theorem, $Y_{(n)}$ is a sufficient statistic for $\theta$. $\square$

**Example 4.7.5**

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

$$f(y) = e^{-(y-\theta)}, \quad y \geq \theta \text{ and } 0 \text{ elsewhere.}$$

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

*Solution* We begin by finding the likelihood function. Since $f(y_i \mid \theta) = e^{-(y-\theta)}$ for $y > \theta$ and zero elsewhere we have

$$L(y_1, y_2, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} e^{-(y_i-\theta)} = e^{-(\sum_{i=1}^{n} y_i - n\theta)}.$$

where $y_i > \theta$ for all $i = 1, 2, \ldots, n$ and zero elsewhere. For this to be valid we need the $\min(y_1, y_2 \ldots, y_n) > \theta$. Thus we can use the indicator function $I(Y_{(1)} > \theta)$ which is 1 when $Y_{(1)} > \theta$ and zero otherwise. Thus we can let $g(\theta, Y_{(1)}) = e^{n\theta} I(Y_{(1)} > \theta)$ and $h(y_1, y_2 \ldots, y_n) = e^{-\sum_{i=1}^{n} y_i}$. Thus using Theorem 4.7.2 we can conclude that $Y_{(1)}$ is a sufficient statistic for $\theta$. $\square$

One thing I want to note from the above two examples is that when we factored the likelihood function into two nonnegative functions $g(\theta, Y_{(1)})/Y_{(n)})$ and $h(y_1, y_2 \ldots, y_n)$, even though the function $g(\theta, Y_{(1)})/Y_{(n)})$ does not explicitly contain the either of the pdf's for the order statistics, they are stil there implicitly through the indicator function. To clarify using Example 4.7.5 as the example,

$$I(Y_{(1)} > \theta) = \begin{cases} 1 & \text{if } Y_{(1)} > \theta \\ 0 & \text{otherwise} \end{cases}$$

This is where $Y_{(1)}$ appears in the support of the joint distribution. It's the only way $\theta$ interacts with the sample, so the only dependence of the joint PDF on $\theta$ is through this indicator, which is a function of the minimum. Even though $Y_{(1)}$ does not appear algebraically in $e^{n\theta}$, it appears in the domain restriction, which is part of the function $g$. This method allows us to sneakily use this theorem for situations like these.

**Example 4.7.6**

Let $Y_1, Y_2, \dots, Y_n$ be a random sample from $N(0, \theta)$. Here $\theta$ represents $\sigma^2$. Find a sufficient statistic for $\sigma^2$.

*Solution* We begin by findinf the likelihood function. Since

$$f(y_i; \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{y_i^2}{2\theta}\right)$$

we have that

$$L(y_1, \dots, y_n \mid \theta) = \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left(-\frac{1}{2\theta}\sum_{i=1}^n y_i^2\right).$$

Let $T = \sum_{i=1}^n y_i^2$ then we have

$$f(y_1, \dots, y_n; \theta) = \underbrace{\left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left(-\frac{T}{2\theta}\right)}_{g(T;\theta)} \cdot \underbrace{1}_{h(y_1,\dots,y_n)}$$

Thus we have that $T = \sum_{i=1}^n y_i^2$ is a sufficient statistic for $\theta = \sigma^2$. □

### 4.7.1 Completeness and Uniqueness

Before diving into the next section we talk about Completeness (not the Real Analysis kind). Let's head straight into the definition and then try to understand it.

**Definition 4.7.7**

Let $T(X)$ be a statistic. We say that $T$ is complete for the family of distributions $\{f(x; \theta)\}$ if:

For every function $g(T)$, if E[g(T)] = 0 for all $\theta$ then $g(T) = 0$ (a.s).

We've learned that a sufficient statistic $T(X)$ captures all the information about the parameter $\theta$ that's in the data but sufficiency alone doesn't guarantee uniqueness of estimators. There could be many different unbiased estimators, all functions of a sufficient statistic and all with different variances so we might have no way to tell which is best. Completeness eliminates that ambiguity.

Going back to our definition, a statistic $T(X)$ is complete if the only function $g$ of $T$ that has zero expectation for every possible $\theta$ is the trivial function:

$$E[g(T)] = 0 \forall\theta \implies g(T) = 0(a.s).$$

Let's say that $T(X)$ is sufficient for $\theta$ and suppose you are looking for an unbiased estimator of something lets say $\tau(\theta)$. We want

$$E(h(T)) = \tau(\theta).$$

But what if there are multiple such functions $h$? Which one is best? If $T$ is not complete, then multiple functions $h_1(T), h_2(T)\dots.$ might all satisfy the same expected value. But if $T$ is complete, then there is only one such function (up to probability-zero sets). Completeness ensures uniqueness of unbiased estimators within the class of functions of the sufficient statistic. The reason why this is true is because suppose :

- $T(X)$ is a complete and sufficient statistic
- $h_1(T)$ and $h_2(T)$ are both unbiased estimators of the same quantity $\tau(\theta)$, so:

$$E_\theta[h_1(T)] = E_\theta[h_2(T)] = \tau(\theta), \quad \text{for all } \theta$$

Now define a function:

$$g(T) := h_1(T) - h_2(T)$$

Then for all $\theta$:

$$E_\theta[g(T)] = E_\theta[h_1(T) - h_2(T)] = \tau(\theta) - \tau(\theta) = 0$$

But by definition of completeness $E_\theta[g(T)] = 0$ for all $\theta \Rightarrow g(T) = 0$ almost surely. Thus:

$$h_1(T) - h_2(T) = 0 \Rightarrow h_1(T) = h_2(T) \quad \text{a.s.}$$

The reason why this is important is because of the following theorem

**Theorem 4.7.8 : Lehmann–Scheffé Theorem**

Let $T(Y)$ be a complete and sufficient statistic for the parameter $\theta$. If $h(T)$ is an unbiased estimator of a function $\tau(\theta)$, then $h(T)$ is the unique minimum variance unbiased estimator (MVUE) of $\tau(\theta)$.

*Proof.* The proof is omitted for brevity. The key idea is that completeness ensures uniqueness, and sufficiency ensures that $h(T)$ uses all information about $\theta$ in the data. □

This theorem is extremely useful in practice: if you can find a statistic that is both sufficient and complete, then any unbiased estimator that is a function of that statistic is automatically the MVUE.

**Example 4.7.9**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution over the interval $(0, \theta)$. Show that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is complete.

*Solution* We are going to suppose that $E[g(Y_{(n)})] = 0$ for all $\theta$. We now need to show that $g(Y_{(n)}) = 0$. First recall that the pdf of $Y_{(n)}$ is

$$f_{Y_{(n)}}(y) = \frac{n}{\theta^n} y^{n-1}.$$

Thus since $E[g(Y_{(n)})] = 0$ we get that

$$\begin{aligned} E[g(Y_{(n)})] &= \int_0^\theta g(y) \cdot \frac{n}{\theta^n} y^{n-1} dy = 0 \\ &= \frac{n}{\theta^n} \int_0^\theta g(y) y^{n-1} dy = 0. \end{aligned}$$

We now can differentiate both sides with respect to $\theta$ using the Fundamental Theorem of calculus to get

$$\frac{n}{\theta^n} g(\theta)\theta^{n-1} - \frac{n}{\theta^n} g(0)0^{n-1} = 0.$$

simplifying we get that

$$\frac{n}{\theta} g(\theta) = 0g$$

for all $\theta > 0$. Thus we have that $g(Y_{(n)}) = 0$ a.s and so it is complete. □

## 4.8 The Rao-Blackwell Theorem and MVUE

The Rao-Blackwell theorem is a fundamental result in statistical estimation that shows how to improve any unbiased estimator by "conditioning" it on a sufficient statistic. If $\hat{\theta}$ is an unbiased estimator for $\theta$ and if U is a statistic that is sufficient for $\theta$, then there is a function of $U$ that is also an unbiased estimator for $\theta$ and has no larger variance than $\hat{\theta}$. This is the general idea of the Rao-Blackwell Theorem.

**Theorem 4.8.1 : Rao-Blackwell**

Let $\hat{\theta}$ be an unbiased estimator for $\theta$ such that $V(\hat{\theta}) < \infty$. If $U$ is a sufficient statistic for $\theta$, define $\hat{\theta}^* = E(\hat{\theta} \mid U)$. Then, for all $\theta$,

$$E(\hat{\theta}^*) = \theta \quad \text{and} \quad V(\hat{\theta}^*) \leq V(\hat{\theta}).$$

If $U$ is complete then $\hat{\theta}^*$ is the unique MVUE (Lehmann–Scheffé).

*Proof.* Because $U$ is sufficient for $\theta$, the conditional distribution of any statistic (including $\hat{\theta}$), given $U$, does not depend on $\theta$. Thus, $\hat{\theta}^* = E(\hat{\theta} \mid U)$ is not a function of $\theta$ and is therefore a statistic. Moreover we know from STA256 that

$$E(\hat{\theta}^*) = E[E(\hat{\theta} \mid U)] = E(\hat{\theta}) = \theta.$$

Thus we see that $\hat{\theta}^*$ is an unbiased estimator of $\theta$. Then recall from STA256 notes Theorem 2.3.9 we have that

$$\mathrm{Var}(\hat{\theta}^*) = \mathrm{Var}[E(\hat{\theta} \mid U)] \leq \mathrm{Var}(\hat{\theta})$$

as required. Thus using the result of Theorem 4.7.8 we have that this will also be the UMVUE if $U$ is complete. □

As we said at the start, the Rao-Blackwell theorem tells us that if we have some unbiased estimator for $\theta$ it can be made into a function of a sufficient statistic. Then it may improve by having a smaller variance. Now you might think why not keep repeating this process to get smaller and smaller variances. Lets say that we apply Rao-Blackwell theorem so that $\hat{\theta}^* = E(\hat{\theta} \mid U)$ will be a function of the statistic $U$ say $h(U)$. Then if we try apply Rao-Blackwell again with the same sufficient statistic $U$ we have that $E(h(u) \mid U) = h(U)$ (to see why this is true call Defintion 2.3.2 and 2.3.5 from STA256 notes). Thus nothing happens once we apply Rao-Blackwell again. For it to get better must use a different sufficient statistic each time.

Now if our goal is to find the unique minimum variance unbiased estimator we can go about it in two ways. The first way is to find an unbiased estimator $\hat{\theta}$ and a sufficient and complete statistic $U$, then apply Rao-Blackwell theorem along with Lehmann–Scheffé to get the UMVUE

$E(\hat{\theta} \mid U)$. The second way is to find a sufficient and complete statistic $U$ then find some function of the statistic $U$ say $\phi(U)$ such that this function is unbiased and so is the UVMUE.

**Example 4.8.2**

Let $Y_1, Y_2, \dots, Y_n$ denote a random sample from the uniform distribution over the interval $(0, \theta)$.

1. Show that $Y_{(n)} = \max(Y_1, Y_2, \dots, Y_n)$ is sufficient and complete.
2. Find the UMVUE of $\theta$.

*Solution* For (1): We have already shown that $Y_{(n)}$ in Example 4.7.4 using the Factorization Theorem and have shown that it is also complete. For (2): To find the UMVUE all we have to do is find a function $U$ that is unbiased (second way). We first find the expected value of $Y_{(n)}$. We see that

$$E(Y_{(n)}) = \int_0^\theta y \cdot \frac{1}{\theta^n} y^{n-1} dy = \frac{n}{n+1}\theta.$$

Let $\hat{\theta} = (n+1)Y_{(n)}/n$ and notice that $E(\hat{\theta}) = \theta$. By Rao-Blackwell theorem $\hat{\theta}$ is the UMVUE. □

**Example 4.8.3**

Let $Y_1, Y_2, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$.

(a) Show that $T = \sum_{i=1}^n Y_i$ is sufficient and complete.

(b) Find the UMVUE of $\lambda$.

*Solution* For (a): To show that $T$ is sufficient we begin by finding the likelihood function which is

$$L(y_1, y_2, \dots, y_n \mid \lambda) = \prod_{i=1}^n \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}}{y_1! y_2! \cdots y_n!} = \frac{\lambda^T e^{-n\lambda}}{y_1! y_2! \cdots y_n!}.$$

Then notice that $g(T, \lambda) = \lambda^T e^{-n\lambda}$ and $h(y_1, y_2, \dots, y_n) = \frac{1}{y_1! y_2! \cdots y_n!}$. By the Factorization Theorem, $T = \sum_{i=1}^n Y_i$ is sufficient for $\lambda$. To show that $T$ is complete assume that $E[g(T)] = 0$ for all $\lambda > 0$. Moreover notice that $T \sim \text{Poisson}(n\lambda)$. Then we have that

$$\begin{aligned} E[g(T)] &= \sum_{u=1}^n g(u) \frac{(n\lambda)^u e^{-n\lambda}}{u!} = 0 \\ &= e^{-n\lambda} \sum_{u=1}^n g(u) \frac{(n\lambda)^u}{u!}. \end{aligned}$$

Notice that this is a power series cenetered at $n\lambda$. Thus for this to be zero then the power series coeffiecients must be zero and thus

$$\frac{g(u)}{u!} = 0 \quad \text{or} \quad g(u) = 0$$

for $u = 1, 2, \dots$. Thus we can conclude that $T$ is complete. For (b): To find a UMVUE of $\lambda$ we can find a function of $T$ such that it is unbiased and then using Rao-Blackwell conclude that it is the UMVUE. We begin by finding that

$$E(T) = n\lambda = \overline{Y}.$$

Letting $\overline{Y} = T/n$ we get that $E(\overline{Y}) = \lambda$. Thus we have that $\overline{Y}$ This aligns with our result from Example 4.6.4. □

**Example 4.8.3**

Let $Y_1, Y_2, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} Exp(\theta)$. Find MVUE of $\theta$.

*Solution* Recall from Example 4.7.3 we found that $\overline{Y}$ was a sufficient statistic for $\theta$. Moreover we can easily show that it is also complete (although the question does not ask for UMVUE). Thus we have that $E(\overline{Y}) = \theta$ so it is unbiased. Thus using Rao-Blackwell theorem (without Lehmann–Scheffé) we have that $\overline{Y}$ is the MVUE. □

As we seen in the question above, how would we normally find the UMVUE for a statistic? We know that we can either find some estimator such that it is efficient and so is a MVUE and then check for completeness or we can instead find a sufficient statistic and then find some function of the statistic that is unbiased and then use the Rao-Blackwell theorem (see section 4.8.1). We discuss the second option. Well recall from the Factorization Theorem that we can easily identify sufficient statistics by factoring the likelihood function. Once we have a sufficient statistic, we check if it's complete (often this requires showing that the only function of the statistic with zero expectation is the zero function). If the sufficient statistic is also complete, then any unbiased function of that statistic is the UMVUE by the Lehmann-Scheffé theorem.

The general procedure for finding a UMVUE is therefore:

1. Find the likelihood function $L(y_1, y_2, \ldots, y_n \mid \theta)$
2. Use the Factorization Theorem to identify a sufficient statistic $T$
3. Verify that $T$ is complete
4. Find a function $h(T)$ such that $E[h(T)] = \theta$ (or whatever parameter you're estimating)
5. By Lehmann-Scheffé, $h(T)$ is the UMVUE

This systematic approach allows us to find the best possible unbiased estimator for a given parameter, combining the concepts of sufficiency, completeness, and the Rao-Blackwell theorem into a powerful framework for statistical estimation.

### 4.8.1 Properties of Maximum Likelihood Estimators

An important property of maximum likelihood estimators is their relationship to sufficient statistics. As we will see, MLEs are always functions of sufficient statistics, which has important implications for their efficiency properties (I included this section myself it is not in the course outline - this is because I believe it is practically important and doesn't undermine section 4.5).

**Theorem 4.8.4**

If $U$ is a sufficient statistic for $\theta$ and the MLE $\hat{\theta}_{MLE}$ exists, then $\hat{\theta}_{MLE}$ is a function of $U$.

*Proof.* By the Factorization Theorem, if $U$ is sufficient for $\theta$, then the likelihood function can be written as:

$$L(y_1, y_2, \ldots, y_n \mid \theta) = g(u, \theta) \times h(y_1, y_2, \ldots, y_n)$$

where $g(u, \theta)$ depends only on $u$ and $\theta$, and $h(y_1, y_2, \ldots, y_n)$ does not depend on $\theta$.

To find the MLE, we maximize $L(\theta)$ with respect to $\theta$. Since $h(y_1, y_2, \ldots, y_n)$ does not depend on $\theta$, maximizing $L(\theta)$ is equivalent to maximizing $g(u, \theta)$. Thus:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} g(u, \theta)$$

This shows that $\hat{\theta}_{MLE}$ depends on the data only through $u$, making it a function of the sufficient statistic $U$. □

This theorem has profound implications. Since MLEs are functions of sufficient statistics, they automatically capture all the information about $\theta$ contained in the data. This is one reason why MLEs often perform well in practice. Furthermore, when the MLE is biased but can be adjusted to be unbiased, the resulting estimator often turns out to be the MVUE. This happens because:

1. The MLE is a function of a sufficient statistic (by Theorem 4.8.4).
2. When we adjust the MLE to make it unbiased, we obtain an unbiased function of a sufficient statistic.
3. If the sufficient statistic is also complete, then by the Lehmann-Scheffé theorem, this unbiased function is the UMVUE.

For example, consider the uniform distribution on $(0, \theta)$ from Example 4.8.2. We found that:

- The MLE is $\hat{\theta}_{MLE} = Y_{(n)}$
- $Y_{(n)}$ is sufficient and complete for $\theta$
- The MLE is biased: $E(\hat{\theta}_{MLE}) = \frac{n}{n+1}\theta$
- The unbiased version is $\hat{\theta} = \frac{n+1}{n} Y_{(n)}$
- This unbiased estimator is the MVUE

Another example is the normal distribution $N(\mu, \sigma^2)$ with both parameters unknown, we found in Example 4.5.2 that:

- The MLEs are $\hat{\mu}_{MLE} = \overline{Y}$ and $\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \overline{Y})^2$
- The sufficient statistic is $(\overline{Y}, \sum_{i=1}^n (Y_i - \overline{Y})^2)$
- $\hat{\mu}_{MLE} = \overline{Y}$ is already unbiased and is the MVUE for $\mu$
- $\hat{\sigma}^2_{MLE}$ is biased, but the unbiased version $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y})^2$ is the MVUE for $\sigma^2$

This connection between MLEs and sufficient statistics explains why maximum likelihood estimation is such a powerful and widely used method. Not only do MLEs have desirable asymptotic properties (consistency, asymptotic normality, and asymptotic efficiency), but they also have the finite-sample property of being functions of sufficient statistics, ensuring they use all available information about the parameter. It is also usually the case that finding a sufficient and complete statistic first is hard which is why MLE are very important in statistical inferences.

### 4.8.2 Exponential Family Property

As we know by now, when we are searching for a UMVUE we need a sufficient and complete statistic for the target parameter. One way we can find such a statistic is using the Exponential Family Property.

**Definition 4.8.5**

Consider the pdf/pmf $f(y \mid \theta)$.

1. $f(y \mid \theta)$ is said to be a member of the **exponential class** if it can be written in the form:
$$f(y \mid \theta) = e^{p(\theta)K(y)+q(\theta)+S(y)}, \quad \text{if } a < y < b.$$
2. $f(y \mid \theta)$ is said to be a **regular** member of the exponential class (family) if in addition to (1) above:
   (a) $a$ and $b$ are free of $\theta$.
   (b) $p(\theta)$ and $K(y)$ are non-trivial $[\iff p'(\theta) \neq 0$ and $K'(y) \neq 0]$.
   (c) $S(y)$ is continuous.

In this section we concern ourselves with regular members of the exponential family distribution. The reason why is because of the following theorem that will prove useful for us.

**Theorem 4.8.6**

Consider the pdf/pmf $f(y \mid \theta)$, where $\theta \in \Omega$. If $f(y \mid \theta)$ is a regular member of the exponential class, then
$$U = \sum_{i=1}^{n} K(y_i)$$
is sufficient and complete.

*Proof.* To prove sufficiency, we use the Factorization Theorem. For a random sample $Y_1, Y_2, \ldots, Y_n$ from a regular exponential family, the likelihood function is:

$$\begin{aligned}
L(y_1, \ldots, y_n \mid \theta) &= \prod_{i=1}^{n} f(y_i \mid \theta) \\
&= \prod_{i=1}^{n} e^{p(\theta)K(y_i)+q(\theta)+S(y_i)} \\
&= e^{p(\theta)\sum_{i=1}^{n} K(y_i)+nq(\theta)+\sum_{i=1}^{n} S(y_i)} \\
&= e^{p(\theta)U+nq(\theta)} \cdot e^{\sum_{i=1}^{n} S(y_i)}
\end{aligned}$$

Setting $g(U, \theta) = e^{p(\theta)U+nq(\theta)}$ and $h(y_1, \ldots, y_n) = e^{\sum_{i=1}^{n} S(y_i)}$, we see that the likelihood factors as required by the Factorization Theorem. Therefore, $U$ is sufficient for $\theta$.

To prove completeness, suppose $E[g(U)] = 0$ for all $\theta \in \Omega$. We need to show that $g(U) = 0$ almost surely. Since $U = \sum_{i=1}^{n} K(Y_i)$ and each $K(Y_i)$ is a function of a regular exponential family member, the distribution of $U$ has a probability density/mass function of the form:

$$f_U(u \mid \theta) = e^{p(\theta)u+nq(\theta)} h_U(u)$$

for some function $h_U(u)$ that doesn't depend on $\theta$. The condition $E[g(U)] = 0$ becomes:

$$\int g(u)e^{p(\theta)u+nq(\theta)}h_U(u)du = 0$$

for all $\theta$. Since $p'(\theta) \neq 0$ (regularity condition), as $\theta$ varies over $\Omega$, $p(\theta)$ takes on an interval of values. Define $\phi = p(\theta)$ and rewrite the integral as:

$$e^{nq(\theta)} \int g(u)h_U(u)e^{\phi u}du = 0$$

Since $e^{nq(\theta)} > 0$, we have:

$$\int g(u)h_U(u)e^{\phi u}du = 0$$

for all $\phi$ in some interval. This means the moment generating function of the random variable with density proportional to $g(u)h_U(u)$ is zero on an interval. By the uniqueness of moment generating functions, this implies $g(u)h_U(u) = 0$ almost everywhere. Since $h_U(u) > 0$ wherever the distribution has support, we conclude $g(u) = 0$ almost surely. Therefore, $U$ is complete. $\square$

**Example 4.8.7**

Show the following distributions belong to the exponential class. Find a complete sufficient statistic for the parameter $\theta$.

(a) Binomial$(n, \theta)$.

(b) Poisson$(\theta)$.

(c) $N(0, \theta)$. Note: Here $\theta$ represents $\sigma^2$.

*Solution* For (a): We begin and see that

$$\begin{aligned} f(y \mid \theta) &= \binom{n}{y}\left(\frac{\theta}{1-\theta}\right)^y (1-\theta)^n \\ &= e^{\ln\left[\binom{n}{y}\left(\frac{\theta}{1-\theta}\right)^y (1-\theta)^n\right]} \\ &= e^{\ln[\binom{n}{y}]+y\ln(\frac{\theta}{1-\theta})+n\ln(1-\theta)}. \end{aligned}$$

Let $k(y) = y$ and $p(\theta) = \ln(\frac{\theta}{1-\theta})$, $q(\theta) = n\ln(1-\theta)$ and $S(y) = \ln[\binom{n}{y}]$. Then letting $U = \sum_{i=1}^n y_i$ is sufficient and complete. For (b): We again rewrite the pmf of Poisson$(\theta)$ in the form as required in Definition 4.8.5.

$$\begin{aligned} f(y \mid \theta) &= \frac{\theta^y e^{-\theta}}{y!} \\ &= e^{\ln(\theta^y)} \cdot e^{-\theta} \cdot e^{\ln[(y!)^{-1}]} \\ &= e^{y\ln(\theta)-\theta-\ln(y!)}. \end{aligned}$$

Letting $k(y) = y$ and $p(\theta) = \ln(\theta)$, $q(\theta) = -\theta$ and $S(y) = -\ln(y!)$ we can conclude that it is a member of the regular class of exponential family. Then we also see that $\sum_{i=1}^n k(y_i) = \sum_{i=1}^n y_i$ is a sufficient and complete statistic. For (c): We begin with the pdf of $N(0, \theta)$ where $\theta = \sigma^2$:

$$\begin{aligned} f(y \mid \theta) &= \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{y^2}{2\theta}} \\ &= e^{\ln\left(\frac{1}{\sqrt{2\pi\theta}}\right) - \frac{y^2}{2\theta}} \\ &= e^{-\frac{1}{2\theta} y^2 - \frac{1}{2}\ln(2\pi\theta)}. \end{aligned}$$

Let $K(y) = y^2$, $p(\theta) = -\frac{1}{2\theta}$, $q(\theta) = -\frac{1}{2}\ln(2\pi\theta)$, and $S(y) = 0$. We can verify this is a regular member of the exponential family since The support $(\infty, \infty)$ is free of $\theta$, $p'(\theta) = \frac{1}{2\theta^2} \neq 0$ and $K'(y) = 2y \neq 0$ (except at $y = 0$). $S(y) = 0$ is continuous). Therefore, $U = \sum_{i=1}^n K(Y_i) = \sum_{i=1}^n Y_i^2$ is sufficient and complete for $\theta = \sigma^2$. □

One final useful fact about regular members of the exponential class is summarized in the following theorem.

**Proposition 4.8.8**

For a regular member of the exponential class, show that

$$E(K(y)) = -\frac{q'(\theta)}{p'(\theta)}.$$

Hence,

$$E(U) = E\left(\sum_{i=1}^n K(y_i)\right) = -n\frac{q'(\theta)}{p'(\theta)}.$$

*Proof.* Recall that

$$f(y \mid \theta) = e^{p(\theta)K(y)+q(\theta)+S(y)}.$$

Since this is a probability density function, we have:

$$\int_{-\infty}^{\infty} f(y \mid \theta)dy = \int_{-\infty}^{\infty} e^{p(\theta)K(y)+q(\theta)+S(y)}dy = 1.$$

Differentiating both sides with respect to $\theta$:

$$\frac{d}{d\theta}\int_{-\infty}^{\infty} e^{p(\theta)K(y)+q(\theta)+S(y)}dy = \frac{d}{d\theta}(1) = 0.$$

Assuming we can interchange differentiation and integration:

$$\int_{-\infty}^{\infty} \frac{d}{d\theta} e^{p(\theta)K(y)+q(\theta)+S(y)}dy = 0.$$

Computing the derivative:

$$\int_{-\infty}^{\infty} [p'(\theta)K(y) + q'(\theta)]e^{p(\theta)K(y)+q(\theta)+S(y)}dy = 0.$$

This can be rewritten as:

$$\int_{-\infty}^{\infty} [p'(\theta)K(y) + q'(\theta)]f(y \mid \theta)dy = 0.$$

Taking the expectation:

$$p'(\theta)E[K(Y)] + q'(\theta) = 0.$$

Solving for $E[K(Y)]$:

$$E[K(Y)] = -\frac{q'(\theta)}{p'(\theta)}.$$

For the sum $U = \sum_{i=1}^{n} K(Y_i)$, by linearity of expectation:

$$E[U] = E\left[\sum_{i=1}^{n} K(Y_i)\right] = \sum_{i=1}^{n} E[K(Y_i)] = nE[K(Y)] = -n\frac{q'(\theta)}{p'(\theta)}.$$

□

## 4.9 Practice Problems

**9.1**

In Exercise 8.8, we considered a random sample of size 3 from an exponential distribution with density function given by

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

and determined that $\hat{\theta}_1 = Y_1$, $\hat{\theta}_2 = (Y_1 + Y_2)/2$, $\hat{\theta}_3 = (Y_1 + 2Y_2)/3$, and $\hat{\theta}_5 = \overline{Y}$ are all unbiased estimators for $\theta$. Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_5$, of $\hat{\theta}_2$ relative to $\hat{\theta}_5$, and of $\hat{\theta}_3$ relative to $\hat{\theta}_5$.

*Solution* To find the relative efficiency of all of these estimators we will find the variance for each one. First

$$\text{Var}(\hat{\theta}_1) = \text{Var}(Y_1) = \theta^2.$$

Next

$$\text{Var}(\hat{\theta}_2) = \text{Var}\left(\frac{Y_1 + Y_2}{2}\right) = \frac{1}{4}(\text{Var}(Y_1) + \text{Var}(Y_2)) = \frac{\theta^2}{2}.$$

Then we have

$$\text{Var}(\hat{\theta}_3) = \frac{1}{9}(\text{Var}(Y_1) + 4\text{Var}(Y_2)) = \frac{1}{9} \cdot 5\theta^2 = \frac{5\theta^2}{9}.$$

Finally

$$\text{Var}(\hat{\theta}_5) = \text{Var}(\overline{Y}) = \frac{\sigma^2}{n} = \frac{\theta^2}{3}.$$

Using this we find that

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_5) = \frac{\theta^2}{\theta^2/3} = \frac{1}{3},$$

$$\text{eff}(\hat{\theta}_2, \hat{\theta}_5) = \frac{\theta^2/2}{\theta^2/3} = \frac{2}{3},$$

$$\text{eff}(\hat{\theta}_3, \hat{\theta}_5) = \frac{5\theta^2/9}{\theta^2/3} = \frac{3}{5}.$$

Since all of these are less than 1 we can conclude that the sample mean $\overline{Y}$ is the most efficient (we later see that this is the UMVUE). □

**9.2**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a population with mean $\mu$ and variance $\sigma^2$. Consider the following three estimators for $\mu$:

$$\hat{\mu}_1 = \frac{1}{2}(Y_1 + Y_2), \quad \hat{\mu}_2 = \frac{1}{4}Y_1 + \frac{Y_2 + \cdots + Y_{n-1}}{2(n-2)} + \frac{1}{4}Y_n, \quad \hat{\mu}_3 = \overline{Y}.$$

a. Show that each of the three estimators is unbiased.

b. Find the efficiency of $\hat{\mu}_3$ relative to $\hat{\mu}_2$ and $\hat{\mu}_1$, respectively.

*Solution* For (a): We see that

$$E(\hat{\mu}_1) = \frac{E(Y_1) + E(Y_2)}{2} = \frac{2\mu}{2} = \mu.$$

Next we see that

$$E(\hat{\mu}_2) = \frac{E(Y_1)}{4} + \frac{E(Y_2 + \cdots + E(Y_{n-1})}{2(n-2)} + \frac{E(Y_n)}{4} = \frac{\mu}{4} + \frac{(n-2)\mu}{2(n-2)} + \frac{\mu}{4} = \mu.$$

And we know that $E(\hat{\mu}_3) = \mu$. For (b): We now find the variances for all unbiased estimators first.

$$\text{Var}(\hat{\mu}_1) = \frac{\sigma^2}{2},$$

$$\text{Var}(\hat{\mu}_2) = \frac{n}{8(n-2)}\sigma^2,$$

$$\text{Var}(\hat{\mu}_3) = \frac{\sigma^2}{n}.$$

Then we see that

$$\text{eff}(\hat{\mu}_3, \hat{\mu}_2) = \frac{n\sigma^2/8(n-2)}{\sigma^2/n} = \frac{n^2}{8(n-2)},$$

$$\text{eff}(\hat{\mu}_3, \hat{\mu}_1) = \frac{\sigma^2/2}{\sigma^2/n} = \frac{n}{2}.$$

As expected we see that $\hat{\mu}_3$ is the most efficient relative to the others. □

**9.4**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from a uniform distribution on the interval $(0, \theta)$. If $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$, the result of Exercise 8.18 is that $\hat{\theta}_1 = (n+1)Y_{(1)}$ is an unbiased estimator for $\theta$. If $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$, the results of Example 9.1 imply that $\hat{\theta}_2 = \left[\frac{(n+1)}{n}\right] Y_{(n)}$ is another unbiased estimator for $\theta$. Show that the efficiency of $\hat{\theta}_1$ to $\hat{\theta}_2$ is $1/n^2$. Notice that this implies that $\hat{\theta}_2$ is a markedly superior estimator.

*Solution* To show that $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = 1/n^2$ we first find the variances of these estimators. To find the variance we first let $U = \hat{\theta}_2/\theta$. Then we see that

$$f_{Y_{(n)}}(y) = nu^{n-1} = \frac{\Gamma(n+1)}{\Gamma(n)\Gamma(1)} u^{n-1}(1-u)^{1-1}.$$

That is we have that $U \sim \text{Beta}(n, 1)$. Thus we see that

$$\text{Var}(U) = \frac{n}{(n+1)^2(n+2)}.$$

Then we have that

$$\text{Var}(Y_{(n)}) = \theta^2 \text{Var}(U) = \theta^2 \cdot \frac{n}{(n+1)^2(n+2)}.$$

Following similar steps for the minimum by letting $U = Y_{(1)}/\theta \sim \text{Beta}(1, n)$. Thus we have

$$\text{Var}(Y_{(1)}) = \theta^2 \cdot \frac{n}{(n+1)^2(n+2)}.$$

Thus we have that then

$$\text{Var}(\hat{\theta}_1) = (n+1)^2 \cdot \text{Var}(Y_{(1)}) = \frac{n}{n+2} \cdot \theta^2,$$

$$\text{Var}(\hat{\theta}_2) = \left[\frac{(n+1)}{n}\right]^2 \cdot \text{Var}(Y_n) = \left[\frac{(n+1)}{n}\right]^2 \cdot \theta^2 \cdot \frac{n}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)}.$$

Thus putting everything together we have that

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

□

**9.5**

Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$. Two unbiased estimators of $\sigma^2$ are

$$\hat{\sigma}_1^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \quad \text{and} \quad \hat{\sigma}_2^2 = \frac{1}{2}(Y_1 - Y_2)^2.$$

Find the efficiency of $\hat{\sigma}_1^2$ relative to $\hat{\sigma}_2^2$.

*Solution* We first find the variances of the two estimators. For the first estimator, recall that $(n-1)S^2/\sigma^2$ has a $\chi^2$ distribution with $n-1$ df. Thus then

$$\text{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = \frac{(n-1)^2}{\sigma^4} \text{Var}(S^2) = 2(n-1).$$

Solving for $\text{Var}(S^2)$ we find that

$$\text{Var}(\hat{\sigma}_1^2) = \text{Var}(S^2) = \frac{2\sigma^4}{n-1}.$$

For the next variance note that $Y_1 - Y_2 \sim N(0, 2\sigma^2)$. Then we can standardize this and get that $Y_1 - Y_2 = \sqrt{2}\sigma Z$. Thus we have that

$$\mathrm{Var}(\hat{\sigma}_2^2) = \frac{\sigma^2}{2}\mathrm{Var}(Z^2) = \frac{2\sigma^4}{2}\mathrm{Var}(\chi^2_{(1)}) = 2\sigma^4.$$

Thus we have that

$$\mathrm{eff}(\hat{\sigma}_1^2, \hat{\sigma}_2^2) = \frac{2\sigma^4}{2\sigma^4/(n-1)} = n - 1.$$

So it is clear that $S^2$ is more efficient compared to the other estimator. $\square$

**9.15**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from a uniform distribution on the interval $(\theta, \theta + 1)$. Show that $\hat{\theta}_1 = \overline{Y} - 1/2$ and $\hat{\theta}_2 = Y_{(n)} - n/n + 1$ are both consistent estimators of $\theta$.

*Solution* To show these estimators are consistent, we need to show that they converge in probability to $\theta$. To do this we use the Weak Law of Large Numbers and note that $\overline{Y} \xrightarrow{P} E[Y_1] = \theta + \frac{1}{2}$. Thus we have that $\hat{\theta}_1 \xrightarrow{P} \theta + 1/2 - 1/2 = \theta$ and so it is consistent. For the next estimator we use the fact that

$$\mathrm{Var}(Y_{(n)}) = \frac{n}{(n+1)^2(n+2)}$$

which then using Theorem 4.3.2 we have that

$$\lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_2) = \lim_{n\to\infty} \frac{n}{(n+1)^2(n+2)} = 0$$

which shows that it is consistent. $\square$

**9.16**

Refer to Exercise 9.5. Is $\hat{\sigma}_2^2$ a consistent estimator of $\sigma^2$?

*Solution* Recall that $\hat{\sigma}_2^2$ is consistent estimator of $\sigma^2$ if for every $\epsilon > 0$,

$$P(|\hat{\sigma}_2^2 - \sigma^2| > \epsilon) = 0.$$

Using ChebyShev's inequality we have that

$$P(|\hat{\sigma}_2^2 - \sigma^2| > \epsilon) \leq \frac{\mathrm{Var}(\hat{\sigma}_2^2)}{\epsilon^2} = \frac{2\sigma^4}{\epsilon^2}$$

$\square$

**9.17**

Suppose that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are independent random samples from populations with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, respectively. Show that $\overline{X} - \overline{Y}$ is a consistent estimator of $\mu_1 - \mu_2$.

*Solution* By using the Weak Law of Large Numbers we know that $\overline{X} \xrightarrow{P} \mu_1$ and $\overline{Y} \xrightarrow{P} \mu_2$ thus using the continous mapping theorem we have that $\overline{X} - \overline{Y} \xrightarrow{P} \mu_1 - \mu_2$. Easy $\square$

**9.18**

In Exercise 9.17, suppose that the populations are normally distributed with $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Show that
$$\frac{\sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{i=1}^n (Y_i - \overline{Y})^2}{2n-2}$$
is a consistent estimator of $\sigma^2$.

*Solution* We use Weak Law of Large Numbers to show that $\overline{X} \xrightarrow{P} \mu_1$ and $\overline{Y} \xrightarrow{P} \mu_2$. We also use the fact that
$$(n-1)S_X^2 = \sum_{i=1}^n (X_i - \overline{X})^2.$$
Thus we have
$$\frac{\sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{i=1}^n (Y_i - \overline{Y})^2}{2n-2} = \frac{(n-1)S_X^2 + (n-1)S_Y^2}{2n-2}.$$
Moreover we use the fact that
$$\frac{(2n-2)\hat{\sigma}^2}{\sigma^2} = \frac{(n-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi^2_{2n-2}.$$
Thus we have that then
$$E\left[\frac{(2n-2)\hat{\sigma}^2}{\sigma^2}\right] = 2n-2 \quad \text{or} \quad E(\hat{\sigma^2}) = \sigma^2$$
which means it is unbiased. Next we see that
$$\text{Var}(\hat{\sigma^2}) = \frac{(2n-2)^2}{\sigma^4}\text{Var}(\hat{\sigma^2}) = 2(2n-2) \quad \text{or} \quad \text{Var}(\hat{\sigma^2}) = \frac{2\sigma^4}{2n-2}.$$
From which we can clearly see the variance limit is zero as $n \to \infty$. Thus it is consistent. □

**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 $\frac{\theta}{\theta+1}$.

*Solution* First notice that $Y_i \sim \text{Beta}(\theta, 1)$. Then we have that
$$E(Y_i) = \frac{\theta}{\theta+1}.$$
Thus using the Weak Law of Large numbers we have that $\overline{Y} \xrightarrow{P} \theta/(\theta+1)$. □

**9.20**

If $Y$ has a binomial distribution with $n$ trials and success probability $p$, show that $Y/n$ is a consistent estimator of $p$.

*Solution* We find the variance of $Y/n$:

$$\operatorname{Var}\left(\frac{Y}{n}\right) = \frac{\operatorname{Var}(Y)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

from which we can clearly see has limit zero as $n \to \infty$. Thus it is consistent. □

**9.21**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size $n$ from a normal population with mean $\mu$ and variance $\sigma^2$. Assuming that $n = 2k$ for some integer $k$, one possible estimator for $\sigma^2$ is given by

$$\hat{\sigma}^2 = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2.$$

a. Show that $\hat{\sigma}^2$ is an unbiased estimator for $\sigma^2$.

b. Show that $\hat{\sigma}^2$ is a consistent estimator for $\sigma^2$.

*Solution* For (a): We first show $\hat{\sigma}^2$ is an unbiased estimator. First notice that $D_i = Y_{2i} - Y_{2i-1} \sim N(0, 2\sigma^2)$. Thus we have that

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{2k}\sum_{i=1}^{k} E(D_i^2) \\ &= \frac{1}{2k}\sum_{i=1}^{k}[2\sigma^2 - \mu] \\ &= \frac{1}{2k}\cdot 2k\sigma^2 = \sigma^2 \end{aligned}$$

For (b): We find the variance of the estimator. We see that

$$\begin{aligned} \operatorname{Var}(\hat{\sigma}^2) &= \frac{1}{4k^2}\sum_{i=1}^{k}\operatorname{Var}\left[(\sqrt{2}\sigma Z)^2\right] \\ &= \frac{\sigma^2}{k^2}\cdot 2k \\ &= \frac{2\sigma^2}{k}. \end{aligned}$$

Thus we have that as $n = 2k \to \infty$ then the variance limit is zero and so it is consistent. □

**9.24**

Let $Y_1, Y_2, Y_3, \ldots, Y_n$ be independent standard normal random variables.

a. What is the distribution of $\sum_{i=1}^{n} Y_i^2$?

b. Let $W_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$. Does $W_n$ converge in probability to some constant? If so, what is the value of the constant?

*Solution* For (a): We know that $Y_i^2 \sim \chi^2_{(1)}$ and so $\sum_{i=1}^n Y_i^2 \sim \chi^2_{(n)}$. For (b): We can see that $W_n$ is the sample mean of $n$ independent $\chi^2_{(1)}$ random variables. And so by the Weak Law of Large Numbers we have that $W_n \xrightarrow{P} E(\chi^2_{(1)}) = 1$. □

**9.25**

Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from a normal distribution with mean $\mu$ and variance 1. Consider the first observation $Y_1$ as an estimator for $\mu$.

a. Show that $Y_1$ is an unbiased estimator for $\mu$.

b. Find $P(|Y_1 - \mu| \leq 1)$.

c. Look at the basic definition of consistency given in Definition 4.3.1. Based on the result of part (b), is $Y_1$ a consistent estimator for $\mu$?

*Solution* For (a): We can clearly see that $E(Y_1) = \mu$. For (b): We now find $P(|Y_1 - \mu| \leq 1)$.

$$\begin{aligned}
P(|Y_1 - \mu| \leq 1) &= P(-1 \leq Y_1 - \mu \leq 1) \\
&= P(-1 \leq Z \leq 1) \\
&= 2P(Z \leq 1) - 1 \\
&= 0.6827.
\end{aligned}$$

For (c): Consistency requires $P(|Y_1 - \mu| < \varepsilon) \to 1$ as $n \to \infty$. But this probability doesn't depend on $n$ (it's $2\Phi(\varepsilon) - 1 < 1$ for any fixed $\varepsilon > 0$). Therefore it does not approach 1. Hence $Y_1$ is not a consistent estimator of $\mu$ (it ignores the growing sample size). □

**9.26**

It is sometimes relatively easy to establish consistency or lack of consistency by appealing directly to Definition 4.3.1, evaluating $P(|\hat{\theta}_n - \theta| \leq \varepsilon)$ directly, and then showing that $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1$. Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size $n$ from a uniform distribution on the interval $(0, \theta)$. If $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$, we showed in Exercise 6.74 that the probability distribution function of $Y_{(n)}$ is given by

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

a. For each $n \geq 1$ and every $\varepsilon > 0$, it follows that

$$P(|Y_{(n)} - \theta| \leq \varepsilon) = P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon).$$

If $\varepsilon > \theta$, verify that $P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = 1$, and that for every positive $\varepsilon < \theta$, we obtain

$$P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = 1 - \left[\frac{\theta - \varepsilon}{\theta}\right]^n.$$

b. Using the result from part (a), show that $Y_{(n)}$ is a consistent estimator for $\theta$ by showing that, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|Y_{(n)} - \theta| \leq \varepsilon) = 1.$$

*Solution* For (a): We have to show that when $\varepsilon > \theta$ we have that $P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = 1$. Thus we get that

$$P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = F_{Y_{(n)}}(\theta + \epsilon) - F_{Y_{(n)}}(\theta - \epsilon) = 1 - 0 = 1.$$

This is because when $\varepsilon > \theta$, $\theta - \varepsilon < 0$ and $\theta + \varepsilon > \theta$. Now Similarly when $\varepsilon < \theta$ we have that

$$P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = F_{Y_{(n)}}(\theta + \epsilon) - F_{Y_{(n)}}(\theta - \epsilon) = 1 - \left[\frac{\theta - \varepsilon}{\theta}\right]^n.$$

For (b): Let $\varepsilon > 0$ be given. First assume that $\varepsilon > \theta$ We then have that

$$\lim_{n\to\infty} P(|Y_{(n)} - \theta| \leq \varepsilon) = \lim_{n\to\infty} P(|Y_{(n)} P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = 1.$$

Next assume that $\varepsilon < \theta$. We then have that

$$\begin{aligned} \lim_{n\to\infty} P(|Y_{(n)} - \theta| \leq \varepsilon) &= \lim_{n\to\infty} P(|Y_{(n)} P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) \\ &= \lim_{n\to\infty} 1 - \left[\frac{\theta - \varepsilon}{\theta}\right]^n \\ &= 1 - 0 = 1, \end{aligned}$$

as required. □

**9.30**

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables, each with probability density function

$$f(y) = \begin{cases} 3y^2, & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $\overline{Y}$ converges in probability to some constant and find the constant.

*Solution* We know that by the Weak Law of Large Numbers, $\overline{Y} \xrightarrow{P} E(Y_i)$. Thus we find the mean.

$$E(Y_i) = \int_0^1 3y^3 dy = \frac{3}{4}.$$

□

**9.36**

Suppose that $Y$ has a binomial distribution based on $n$ trials and success probability $p$. Then $\hat{p}_n = Y/n$ is an unbiased estimator of $p$. Use Theorem 4.3.6 to prove that the distribution of

$$\frac{\hat{p}_n - p}{\sqrt{\hat{p}_n \hat{q}_n / n}}$$

converges to a standard normal distribution.

*Solution* We know that $\hat{p}_n \xrightarrow{P} p$ and $\hat{q}_n \xrightarrow{P} q$. Thus using the continuous mapping theorem we have that $\sqrt{\hat{p}_n \hat{q}_n / n} \xrightarrow{P} \sqrt{pq/n}$. Moreover, by the CLT we have that

$$W_n = \frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{pq}} \xrightarrow{D} N(0,1).$$

Thus using Slutsky's theorem we have that

$$\frac{\hat{p}_n - p}{\sqrt{\hat{p}_n \hat{q}_n / n}} = \frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{pq}} \cdot \frac{\sqrt{pq}}{\sqrt{\hat{p}_n \hat{q}_n}} \xrightarrow{D} N(0,1),$$

since the first factor converges in distribution to $N(0,1)$ and the second factor converges in probability to 1. □

**9.37**

Let $X_1, X_2, \ldots, X_n$ denote $n$ independent and identically distributed *Bernoulli* random variables such that

$$P(X_i = 1) = p \quad \text{and} \quad P(X_i = 0) = 1 - p,$$

for each $i = 1, 2, \ldots, n$. Show that $\sum_{i=1}^n X_i$ is sufficient for $p$ by using the factorization criterion.

*Solution* We begin by finding the likelihood function. We see that

$$L(x_1, x_2 \ldots, x_n \mid p) = \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n p^{x_i}(1-p)^{n-x_i} = p^T(1-p)^{n-T}$$

where $T = \sum_{i=1}^n x_i$. Thus letting $g(T,p) = p^T(1-p)^{n-T}$ and $h(x_1, x_2, \ldots, x_n) = 1$ we can conclude that $T$ is sufficient for $p$. □

**9.38**

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

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

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

c. If $\mu$ and $\sigma^2$ are both unknown, show that $\sum_{i=1}^n Y_i$ and $\sum_{i=1}^n Y_i^2$ are jointly sufficient for $\mu$ and $\sigma^2$. [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 $\mu$ and $\sigma^2$.]

*Solution* For (a): We first find the likelihood function:

$$
\begin{aligned}
L(y_1, y_2, \ldots, y_n \mid \mu) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2\sigma^2}\right)(y_i - \mu)^2\right] && (1)\\
&= \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \exp\left[-\left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^n (y_i - \mu)^2\right] && (2)\\
&= \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \exp\left[-\left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^n (y_i - \overline{y})^2 + n(\overline{y} - \mu)^2\right] && (3)\\
&= \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \exp\left[-\left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^n (y_i - \overline{y})^2\right] \cdot \exp\left[-\left(\frac{n}{2\sigma^2}\right) n(\overline{y} - \mu)^2\right]. && (4)
\end{aligned}
$$

From this we can clearly see that it factors into a function of $g(\overline{Y}, \mu)$ and $h(y_1, y_2, \ldots, y_n)$. Thus $\overline{Y}$ is sufficient for $\mu$. For (b): From (2) in part a we can clearly see that using the Factorization Theorem that $\sum_{i=1}^n (Y_i - \mu)^2$ is sufficient for $\sigma^2$. For (c): Note that

$$
\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n Y_i^2 - 2\mu \sum_{i=1}^n Y_i + n\mu^2.
$$

Thus the likelihood function becomes

$$
\begin{aligned}
\left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \exp\left[-\left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^n (y_i - \mu)^2\right] &= \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \exp\left[-\left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^n Y_i^2\right] \cdot \exp\left[\left(\frac{1}{\sigma^2}\right)\sum_{i=1}^n Y_i + n\mu^2\right]\\
&= h(y_1, y_2, \ldots, y_n) \cdot g\left(\sum_{i=1}^n Y_i \sum_{i=1}^n Y_i^2; \mu, \sigma^2\right).
\end{aligned}
$$

Thus using the Factorization Theorem we can conclude they are jointly sufficient for $\mu$ and $\sigma^2$. Then it follows that $\overline{Y}$ and $S^2$ are also joinlty sufficient for $\mu$ and $\sigma^2$. □

**9.39**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Poisson distribution with parameter $\lambda$. Show by conditioning that $\sum_{i=1}^n Y_i$ is sufficient for $\lambda$.

*Solution* We first find the likelihood function and notice that

$$L(y_1, y_2, \dots, y_n \mid \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} = \frac{e^{-n\lambda}\lambda^{\sum y_i}}{\prod_{i=1}^{n} y_i!}.$$

We then show that $L(\lambda \mid S)$ does not depend on $\lambda$. Notice since $S$ is a sum of poisson random varaibles it is also poisson distributed with mean $n\lambda$. That is

$$P(S = s) = \frac{e^{-n\lambda}e^(n\lambda)^s}{s!}.$$

Thus we see that

$$L(\lambda \mid S) = \frac{\frac{e^{-n\lambda}\lambda^s}{\prod_{i=1}^{n} y_i!}}{\frac{e^{-n\lambda}(n\lambda)^s}{s!}} = \frac{s!}{n^s \prod_{i=1}^{n} y_i!}.$$

Thus it is sufficient. □

**9.43**

Let $Y_1, Y_2, \dots, Y_n$ denote independent and identically distributed random variables from a power family distribution with parameters $\alpha$ and $\theta$. 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 $\theta$ is known, show that $\prod_{i=1}^{n} Y_i$ is sufficient for $\alpha$.

*Solution* We begin by finding the likelihood function.

$$L(y_1, y_2, \dots, y_n \mid \alpha) = \prod_{i=1}^{n} \alpha y_i^{\alpha-1}/\theta^\alpha = \left(\frac{\alpha^n}{\theta^{n\alpha}}\right)\left[\prod_{i=1}^{n} y_i\right]^{\alpha-1} 1_{0 \le y_i \le \theta \forall i}.$$

Thus we can see this factors into $g(\prod_{i=1}^{n} Y_i, \alpha) = \left(\frac{\alpha^n}{\theta^{n\alpha}}\right)[\prod_{i=1}^{n} y_i]^{\alpha-1}$ and $h(y_1, y_2, \dots, y_n) = 1_{0 \le y_i \le \theta \forall i}$. Thus it is sufficient. □

**9.50**

Let $Y_1, Y_2, \dots, Y_n$ denote a random sample from the uniform distribution over the interval $(\theta_1, \theta_2)$. Show that $Y_{(1)} = \min(Y_1, Y_2, \dots, Y_n)$ and $Y_{(n)} = \max(Y_1, Y_2, \dots, Y_n)$ are jointly sufficient for $\theta_1$ and $\theta_2$.

*Solution* We begin by finding the likelihood functiion. We have that

$$L(y_1, y_2, \dots, y_n \mid \theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} = \frac{1}{(\theta_2 - \theta_1)^n} \cdot 1_{Y_{(n)} \le \theta_2, Y_{(1)} \ge \theta_1}.$$

Thus letting $g(Y_{(n)}, Y_{(1)}; \theta_1, \theta_n) = \frac{1}{(\theta_2-\theta_1)^n} \cdot 1_{Y_{(n)} \le \theta_2, Y_{(1)} \ge \theta_1}$ and $h(y_1, y_2, \dots, y_n) = 1$ we can conclude that it is sufficient. □

**9.54**

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

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

Show that $\max(Y_1, Y_2, \ldots, Y_n)$ and $\prod_{i=1}^{n} Y_i$ are jointly sufficient for $\alpha$ and $\theta$.

*Solution* We begin by finding the likelihood function.

$$L(y_1, y_2, \ldots, y_n \mid \alpha) = \prod_{i=1}^{n} \alpha y_i^{\alpha-1}/\theta^{\alpha} = \left(\frac{\alpha^n}{\theta^{n\alpha}}\right) \left[\prod_{i=1}^{n} y_i\right]^{\alpha-1} \cdot 1_{0 \le Y_{(1)} \le Y_{(n)} \le \theta \forall i}."$$

Thus letting $g\left(\prod_{i=1}^{n} Y_i, \max(Y_1, Y_2, \ldots, Y_n); \alpha, \theta)\right) = \prod_{i=1}^{n} \alpha y_i^{\alpha-1}/\theta^{\alpha} = \left(\frac{\alpha^n}{\theta^{n\alpha}}\right) \left[\prod_{i=1}^{n} y_i\right]^{\alpha-1} \cdot 1_{Y_{(n)} \le \theta \forall i}$ and $h(y_1, y_2, \ldots, y_n) = 1_{0 < Y_{(1)} < Y_{(n)}}$ we can conclude they are jointly sufficient. □

**9.56**

Refer to Problem 9.38(b). Find an MVUE of $\sigma^2$.

*Solution* We found that $U = \sum_{i=1}^{n} (Y_i - \mu)^2$ was sufficient for $\sigma^2$. Thus we can use Rao-Blackwell theorem to find a MVUE for $\sigma^2$. What we can do is find a function $h(U)$ such that $E(h(U))$ is unbiased. We first find $E(U)$.

$$E(U) = \sum_{i=1}^{n} E\left[(Y_i - \mu)^2\right] = \sum_{i=1}^{n} \mathrm{Var}(Y_i) = n\sigma^2.$$

Thus letting $h(U) = U/n$ we can conclude that it is an MVUE. □

**9.59**

The number of breakdowns $Y$ per day for a certain machine is a Poisson random variable with mean $\lambda$. The daily cost of repairing these breakdowns is given by $C = 3Y^2$. If $Y_1, Y_2, \ldots, Y_n$ denote the observed number of breakdowns for $n$ independently selected days, find an MVUE for $E(C)$.

*Solution* So we first note that $E(C) = 3E(Y^2) = 3(\lambda + \lambda^2) = 3\lambda(1 + \lambda)$. What we can do is use Theorem 4.7.8. That is we can find a sufficient statistic $U$ for $\lambda$ and then find an function $h(U)$ such that $E(h(U)) = E(C)$. We know that from Problem 9.39 and Example 4.8.7 that $U = \sum_{i=1}^{n} Y_i$ is sufficient for $\lambda$. Then we find that

$$E(U) = \sum_{i=1}^{n} E(Y_i) = n\lambda.$$

To find some function $h(U)$ such that $E[h(U)] = E(C)$, we find $\lambda$ and $\lambda^2$ in terms of $U$. That is we see that

$$E(U) = n\lambda \quad \text{or} \quad \lambda = \frac{E(U)}{n}.$$

Moreover notice that $E(U^2) = n\lambda + n^2\lambda^2$. Thus $E(U^2/n) = \lambda + n\lambda^2$. From this we see that

$$\frac{E\left[\frac{U^2}{n} - U\right]}{n} = \lambda^2.$$

Thus letting

$$h(U) = 3\left[\frac{U}{n} + \frac{\frac{U^2}{n} - U}{n}\right]$$

we see that $h(U) = E(C)$ as required and so $h(U)$ is the MVUE. $\square$

**9.60**

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

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

a. Show that this density function is in the (one-parameter) exponential family and that $\sum_{i=1}^n -\ln(Y_i)$ is sufficient for $\theta$. (See Exercise 9.45.)

b. If $W_i = -\ln(Y_i)$, show that $W_i$ has an exponential distribution with mean $1/\theta$.

c. Show that $2\theta\sum_{i=1}^n W_i$ has a $\chi^2$ distribution with $2n$ df.

d. Show that
$$E\left(\frac{1}{2\theta\sum_{i=1}^n W_i}\right) = \frac{1}{2(n-1)}.$$

e. What is the MVUE for $\theta$?

*Solution* For (a): We will show that $f(y \mid \theta)$ is a regular member of the exponential family. Notice that

$$f(y \mid \theta) = e^{\ln(\theta)+\theta\ln(y)-\ln(y)} = e^{p(\theta)k(y)+q(\theta)+S(y)}.$$

Since it is a regular member of the exponential family we know that then $U = \sum_{i=1}^n k(y_i) = \sum_{i=1}^n -\ln(Y_i)$ is sufficient for $\theta$. For (b): We do change of variables here since the transformation is one-to-one. We see that $Y = e^{-W}$ and so $dy = -e^{-w}dw$. Thus we find that

$$f_W(w) = f_Y(e^{-w})|-e^{-w}| = \theta e^{-w\theta-w+w} = \theta e^{-\theta w} \sim \text{Exp}(1/\theta),$$

as required. For (c): We recall that $\sum_{i=1}^n W_i \sim \text{Gamma}(n, 1/\theta)$. And so $U = 2\theta\sum_{i=1}^n W_i \sim \text{Gamma}(2n/2, 2) = \chi^2_{(2n)}$. For (d): We show that $E(U^{-1}) = 1/2(n-1)$. To do this we use the identity for the $k-$th moment of chi-squared.

$$E(U^{-1}) = 2^{-1}\frac{\Gamma\left(\frac{2n}{2}-1\right)}{\Gamma(2n/2)} = \frac{\Gamma(n-1)}{2\Gamma(n)} = \frac{1}{2(n-1)}.$$

For (e): Since we know that $S = \sum_{i=1}^n W_i \sim \text{Gamma}(n, 1/\theta)$ then using the identity for the $k$-th moment of Gamma distribution we have

$$E(S^{-1}) = \frac{(1/\theta)^{-1}\Gamma(n-1)}{\Gamma(n)} = \frac{\theta}{n-1}.$$

So then

$$\hat{\theta} = (n-1)U^{-1}$$

is a MVUE for $\theta$. $\square$

**9.64**

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

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

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

*Solution* For (a): We know that $\overline{Y}$ is sufficient and complete for $\mu^2$. Thus we find some function $h(\overline{Y})$ such that $E[h(\overline{Y})] = \mu^2$. We first note that

$$E(\overline{Y}^2) = \frac{1}{n} + \mu^2.$$

Thus letting $h(\overline{Y}) = \overline{Y}^2 - 1/n$ we see that it is the MVUE. For (b): WE now find the variance of $\hat{\mu}$. Recall that the sample mean is normally distributed with mean $\mu$ and variance $\sigma^2/n = 1/n$. Thus we know that then

$$\sqrt{n}(\overline{Y} - \mu) = Z \quad \text{or} \quad \overline{Y} = \frac{Z}{\sqrt{n}} - \mu.$$

Thus then

$$\overline{Y}^2 = \left(\frac{Z}{\sqrt{n}} - \mu\right)^2 = \frac{Z^2}{n} - \frac{2\mu Z}{\sqrt{n}} + \mu^2.$$

Computing the variance we find that

$$\text{Var}(\overline{Y}^2) = \frac{\text{Var}(Z)}{n} + \frac{4\mu^2 \text{Var}(Z)}{n} = \frac{2 + 4\mu^2}{n}$$

$\square$

**9.69**

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

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

Find an estimator for $\theta$ by the method of moments. Show that the estimator is consistent. Is the estimator a function of the sufficient statistic $T = -\sum_{i=1}^n \ln(Y_i)$ that we can obtain from the factorization criterion? What implications does this have?

*Solution* Using the method of moments, we know that $m_k = \frac{1}{n}\sum_{i=1}^n Y_i^k$ is the $k-$th sample moment. We first find $E(Y_i)$.

$$E(Y_i) = (\theta+1)\int_0^1 y^{\theta+1} dy = \frac{\theta+1}{\theta+2}.$$

Thus we equate the first sample moment to $E(Y_i)$.

$$E(Y_i) = \frac{\theta+1}{\theta+2} = \overline{Y} \quad \text{or} \quad \hat{\theta} = \frac{2\overline{Y}-1}{1-\overline{Y}}.$$

To show that this estimator is consistent we use the WLLN and note that since $\overline{Y} \xrightarrow{P} E(Y_i)$ we have

$$\hat{\theta} \xrightarrow{P} \frac{2\left(\frac{\theta+1}{\theta+2}\right)-1}{1-\frac{\theta+1}{\theta+2}} = \theta.$$

Thus it is consistent. No it depends on the sample mean and not on the sufficient statistic. In a one-parameter regular exponential family with parameter-independent support (as here), the sufficient statistic is complete. Hence any UMVUE must be a function of $T$. Since the MoM estimator is not a function of $T$, it cannot be the UMVUE. By Rao–Blackwell, conditioning any unbiased estimator on $T$ would yield a (uniquely) better unbiased estimator; e.g., the MLE $\hat{\theta}_{\text{MLE}} = -1 - \frac{n}{\sum \ln Y_i}$ is a function of $T$ and is consistent. □

**9.75**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from the probability density function given by

$$f(y \mid \theta) = \begin{cases} \frac{\Gamma(2\theta)}{[\Gamma(\theta)]^2}(y^{\theta-1})(1-y)^{\theta-1}, & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the method-of-moments estimator for $\theta$.

*Solution* First notice that the pdf is the beta distribution with parameters Beta$(\theta,\theta)$. Thus we have that

$$E[Y] = \frac{\alpha}{\alpha+\beta}, \quad E[Y^2] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.$$

With $\alpha = \beta = \theta$,

$$E[Y] = \frac{1}{2}, \quad E[Y^2] = \frac{\theta(\theta+1)}{(2\theta)(2\theta+1)} = \frac{\theta+1}{2(2\theta+1)}.$$

Let $m_2 = \frac{1}{n}\sum_{i=1}^n Y_i^2$ be the second sample moment.
Method of moments sets $m_2 = E[Y^2]$ and solves for $\theta$ :

$$m_2 = \frac{\theta+1}{2(2\theta+1)} \quad \text{or} \quad (4m_2-1)\theta = 1-2m_2 \quad \text{or} \quad \hat{\theta}_{\text{MoM}} = \frac{1-2m_2}{4m_2-1}.$$

Equivalently, using $\text{Var}(Y) = E[Y^2] - \frac{1}{4} = \frac{1}{8\theta+4}$, and the sample variance with divisor $n$, $s_n^2 = m_2 - \frac{1}{4}$, you can write:

$$\hat{\theta}_{\text{MoM}} = \frac{1}{8s_n^2} - \frac{1}{2}.$$

□

**9.79**

Let $Y_1, Y_2, \dots, Y_n$ denote independent and identically distributed random variables from a Pareto distribution with parameters $\alpha$ and $\beta$, where $\beta$ is known. Then, if $\alpha > 0$,

$$f(y \mid \alpha, \beta) = \begin{cases} \alpha\beta^{\alpha}y^{-(\alpha+1)}, & y \geq \beta, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $E(Y_i) = \alpha\beta/(\alpha - 1)$ if $\alpha > 1$ and $E(Y_i)$ is undefined if $0 < \alpha < 1$. Thus, the method-of-moments estimator for $\alpha$ is undefined.

*Solution* We begin by showing $E(Y_i) = \alpha\beta/(\alpha - 1)$ when $\alpha > 1$.

$$\begin{aligned} E(Y_i) &= \int_{\beta}^{\infty} \alpha\beta^{\alpha} y \cdot y^{-(\alpha+1)} dy \\ &= \alpha\beta \int_{\beta}^{\infty} y^{-\alpha} dy \\ &= \alpha\beta \cdot \frac{\beta^{1-\alpha}}{\alpha - 1} = \frac{\alpha\beta}{\alpha - 1}, \end{aligned}$$

which is well defined only when $\alpha > 1$ since when $\alpha = 1$ the improper integral diverges and when $0 < \alpha < 1$ it also diverges. Thus the MOM estimator for the first sample moment does not exist since it only exist for certain values of $\alpha$ and not all. $\square$

**9.80**

Suppose that $Y_1, Y_2, \dots, Y_n$ denote a random sample from the Poisson distribution with mean $\lambda$.

(a) Find the MLE $\hat{\lambda}$ for $\lambda$.

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

(c) Show that the estimator of part (a) is consistent for $\lambda$.

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

*Solution* For (a): To find the MLE we first find the likelihood function. We see that

$$L(y_1, y_2, \dots, y_n \mid \theta) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{y} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!},$$

where $y = \sum_{i=1}^{n} y_i$. We now maximize this function with respect to $\lambda$. We instead maximize the log likelihood instead.

$$\ln\left[\frac{\lambda^{y} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!}\right] = y\ln(\lambda) - n\lambda - \ln(\prod_{i=1}^{n} y_i!).$$

Then we see that

$$\frac{\partial \ln[L(\lambda)]}{\partial\lambda} = \frac{y}{\lambda} - n.$$

Setting this equal and solving for $\lambda$ we have that

$$\hat{\lambda}_{MLE} = \frac{y}{n} = \overline{Y}.$$

For (b): We then know that the sample mean has mean $\mu = \lambda$ and variance $\sigma^2/n = \lambda/n$. For (c): Using the WLLN we can conclude part c easily. Using the invariance property. That is the MLE of $g(\lambda) = e^{-n\lambda}$ is $g(\hat{\lambda}_{MLE})$. So

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

□

**9.82**

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

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

where $r$ is a known positive constant.

(a) Find a sufficient statistic for $\theta$.

(b) Find the MLE of $\theta$.

(c) Is the estimator in part (b) an MVUE for $\theta$?

*Solution* For (a): To find a sufficient statistic we will use the Factorization Theorem. we begin by finding the likelihood function.

$$L(y_1, y_2, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta}\right) r y^{r-1} e^{-y^r/\theta} = \left(\frac{1}{\theta^n}\right) r^n \left[\prod_{i=1}^{n} y_i\right]^{r-1} e^{\left(-\sum_{i=1}^{n} y_i^r\right)/\theta}.$$

Letting $g(U, \theta) = e^{\left(-\sum_{i=1}^{n} y_i^r\right)/\theta}$ and $h(y_1, y_2 \ldots, y_n) = \left(\frac{1}{\theta^n}\right) r^n \left[\prod_{i=1}^{n} y_i\right]^{r-1}$ we can conclude that $T = \sum_{i=1}^{n} y_i^r/\theta$ is a sufficient statistic. For (b): We first find the log-likelihood.

$$\ln[f(y \mid \theta)] = -n\ln(\theta) + n\ln(r) + (r-1)\sum_{i=1}^{n} \ln(y_i) - \frac{1}{\theta}\sum_{i=1}^{n} y_i^r.$$

differentiataing this with respect to $\theta$ we find that

$$\frac{\partial \ln[f(y \mid \theta)]}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} y_i^r.$$

Setting equal to zero and solving for $\theta$ we find that

$$\hat{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} y_i^r.$$

After showing that this estimator is unbiased (through finding out that $Y_i^r$ is distributed Exp($\theta$) through change-of-variables) we can conclude that since it is a function of a sufficient statistic it is an MVUE for $\theta$. □

**9.84**

A certain type of electronic component has a lifetime $Y$ (in hours) with probability density function given by

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{1}{\theta^2}\right) y e^{-y/\theta}, & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

That is, $Y$ has a gamma distribution with parameters $\alpha = 2$ and $\theta$. Let $\hat{\theta}$ denote the MLE of $\theta$. Suppose that three such components, tested independently, had lifetimes of 120, 130, and 128 hours.

(a) Find the MLE of $\theta$.

(b) Find $E(\hat{\theta})$ and $V(\hat{\theta})$.

(c) Suppose that $\theta$ actually equals 130. Give an approximate bound that you might expect for the error of estimation.

(d) What is the MLE for the variance of $Y$?

*Solution* For (a): We begin by finding the likelihood function. We see that

$$L(y_1, y_2, \dots, y_n \mid \theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta^2}\right) y_i e^{-y_i/\theta} = \frac{1}{\theta^{2n}} \prod_{i=1}^{n} y_i e^{-\sum_{i=1}^{n} y_i/\theta}.$$

Finding the log-likelihood function:

$$\ln[L(\theta)] = -2n\ln(\theta) + \sum_{i=1}^{n} \ln(y_i) - \frac{1}{\theta}\sum_{i=1}^{n} y_i.$$

differentiataing with respect to $\theta$:

$$\frac{\partial \ln[f(y \mid \theta)]}{\partial \theta} = \frac{-2n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} y_I.$$

Setting equal to zero and solving for $\theta$ we find that

$$\hat{\theta}_{MLE} = \frac{\overline{Y}}{2}.$$

For (b): We then easily find that

$$E(\hat{\theta}_{MLE}) = \frac{E(\overline{Y})}{2} = \frac{2\theta}{2} = \theta.$$

We also see that

$$\text{Var}(\hat{\theta}_{MLE}) = \frac{\text{Var}(\overline{Y})}{4} = \frac{\sigma^2/n}{4} = \frac{2\theta^2/n}{4} = \frac{\theta^2}{2n}.$$

For (c): We see that the standard error is then $SE = \sqrt{\text{Var}(\overline{Y})} \approx 53.1$. Since the sample mean is normally distirbuted we can place a 2 $SE$ bound on our estimator and see that $|\hat{\theta}_{MLE} - \theta| \leq$

$1.96 \times SE = 104$ (Example 3.6.1). For (d): Using the invariance property since $\text{Var}(Y) = 2\theta^2$, we then know that $g(\hat{\theta}_{MLE}) = 2\hat{\theta}^2_{MLE}$ is an MLe for the variance. Thus

$$\hat{\text{Var}(Y)} = 2\left(\frac{\sum_{i=1}^{n}}{2n}\right)^2.$$

$\square$

**9.92**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \frac{3y^2}{\theta^3}, & 0 \leq y \leq \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.52, you showed that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for $\theta$.

(a) Find the MLE for $\theta$.

(b) Find a function of the MLE in part (a) that is a pivotal quantity.

(c) Use the pivotal quantity from part (b) to find a $100(1-\alpha)\%$ confidence interval for $\theta$.

*Solution* For (a): We begin by finding the likelihood function

$$L(y_1, y_2, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} \frac{3y^2}{\theta^3} = \frac{3^n \prod_{i=1}^{n} y_i^2}{\theta^{3n}}.$$

Finding the log-likelihood we see that

$$\ln[L(\theta)] = n\ln(3) + 2\sum_{i=1}^{n} \ln(y_i) - 3n\ln\theta.$$

differentiataing with respect to $\theta$ we see that

$$\frac{\partial \ln[L(\theta)]}{\partial\theta} = \frac{-3n}{\theta}.$$

Note that this function is decreasing monotonically so it does not have a maximum. So to maximize it we choose the smallest possible value of $\theta$ that is consistent with our data ($0 \leq y_i \leq Y_{(n)} \leq \theta$) $\hat{\theta} = \max(y_1, y_2 \ldots, y_n)$. For (b): To find a pivotal quantity we need a function of the sample data and $\theta$ whose distribution does not depend on $\theta$. We see that

$$F_{Y_{(n)}}(y) = [F(y)]^n = \frac{y^{3n}}{\theta^{3n}}$$

which means

$$f_{Y_{(n)}}(y) = 3n\frac{y^{3n-1}}{\theta^{3n}}.$$

Letting $W = Y_{(n)}/\theta$ we see that $W \sim \text{Beta}(3n, 1)$. For (c): And so we get that

$$P\left(-q\alpha/2 \leq \frac{Y_{(n)}}{\theta} \leq q1 - \alpha\right) = 1 - \alpha.$$

We then see that the CI becomes

$$\left[\frac{Y_{(n)}}{q_{1-\alpha/2}}, \frac{Y_{(n)}}{q_{\alpha/2}}\right]$$

where $q_p = F_W(w)^{-1} = p^{1/3n}$. $\square$

# 5 Hypothesis Testing

## 5.1 Introduction

As we have been constantly repeating throughout this course, one of the main basis that we defined statistics to be was to allow us to make inferences about the population parameters from which a random sample is taken from (Recall Section 3.1). We went over making inferences about this parameters through estimation. So this means finding confidence intervals or point estimators (Section 3) as well as finding the best estimators that we can use (Section 4). However another way we can make inferences are through hypothesis test of the values.

The basic idea behind hypothesis testing is quite intuitive. Imagine you have a claim or belief about some characteristic of a population - perhaps the average height of students at a university, or the proportion of defective items produced by a factory. Rather than simply estimating these values, you want to test whether a specific claim about them is reasonable given the data you've collected. The process works by assuming a particular value or range for the parameter of interest, then examining whether your sample data is consistent with this assumption. If the data you observe would be very unlikely under this assumption, then you have evidence that the assumption is probably wrong. On the other hand, if the data is reasonably consistent with the assumption, then you don't have strong evidence against it.

This approach allows us to make decisions and draw conclusions in situations where we need to choose between competing explanations or validate claims. It provides a systematic framework for determining whether observed differences or patterns in our data are meaningful or could simply be due to random variation. This type of testing is applicable in all fields in which theory can be tested against observations. Before you move on I reccommend you to watch this video which explains hypothesis testing at a high level.

## 5.2 Elements of a Statistical Test

As we said above, we usually make a hypothesis based on the values of the population parameters. For example, a lightbulb manufacturer claims that their bulbs have a mean lifespan of at least 1200 hours. Let's say that a retailer does not believe in this claim. That is the reatiler suspects the average might actually be less than what's advertised. This is called the alternative hypothesis. To support the alternative hypothesis we try to show, using data, that the opposite statement, called the null hypothesis (the claim that the mean lifespan is 1200 hours), is unlikely. In this case we are trying to show that the probability that the mean lifespan is at least 1200 hours ($\mu = 1200$) is low. Support for the alternative hypothesis comes from showing that there is strong enough statistical evidence against the null hypothesis. In other words, we reason somewhat like a proof by contradiction: if the data are highly inconsistent with the null, we reject it and favor the alternative. How do we use that data to decide between the null hypothesis and the alternative hypothesis?

Suppose the retailer tests $n = 40$ randomly chosen bulbs and records their lifespans. The sample mean turns out to be $\bar{x} = 1150$ hours with a sample standard deviation of $s = 120$ hours. The question becomes: is this sample evidence strong enough to conclude that the true mean lifespan is less than 1200 hours?

To answer this systematically, we need to establish the key elements of any statistical test:

1. **1. Null Hypothesis ($H_0$):** This is the statement we assume to be true initially. It usually represents the status quo, no effect, or the claim being tested. In our example: $H_0 : \mu \geq 1200$.
2. **2. Alternative Hypothesis ($H_1$ or $H_a$):** This is what we suspect might be true instead. In our example: $H_1 : \mu < 1200$.
3. **3. Test Statistic:** A function of the sample data that helps us measure how consistent the data is with the null hypothesis. Common examples include $t$-statistics and $z$-statistics.
4. **4. Significance Level ($\alpha$):** The probability threshold below which we will reject the null hypothesis. Common choices are $\alpha = 0.05$ or $\alpha = 0.01$.
5. **5. P-value:** The probability of observing a test statistic as extreme or more extreme than what we actually observed, assuming the null hypothesis is true.
6. **6. Decision Rule:** We reject $H_0$ if the p-value $\leq \alpha$, otherwise we fail to reject $H_0$.

The logic is straightforward: if what we observed would be very unlikely under the null hypothesis (small p-value), then we have evidence against the null hypothesis and reject it in favor of the alternative. Going back to our example we choose a test statistic. Since population $\sigma$ is unknown we use the $t$-statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where $\mu_0$ is the hypothesized value under $H_0$. In our case, $\mu_0 = 1200$, so:

$$t = \frac{1150 - 1200}{120/\sqrt{40}} = \frac{-50}{120/6.32} = \frac{-50}{18.97} \approx -2.64$$

This test statistic follows a $t$-distribution with $n - 1 = 39$ degrees of freedom under the null hypothesis. The p-value is the probability of observing a $t$-statistic of $-2.64$ or smaller (since we have a left-tailed test with $H_1 : \mu < 1200$). Using a $t$-table or statistical software, we find that $p \approx 0.006$. If we choose $\alpha = 0.05$, then since $p = 0.006 < 0.05$, we reject the null hypothesis and conclude that there is sufficient evidence to support the retailer's suspicion that the mean lifespan is less than 1200 hours.

**Definition 5.2.1**

A statistical hypothesis test is a formal procedure for making decisions about population parameters based on sample data. The essential elements of any hypothesis test are:

1. **Null Hypothesis ($H_0$):** A statement about the population parameter that represents the status quo, no effect, or the claim being tested. It is assumed to be true until evidence suggests otherwise.

2. **Alternative Hypothesis ($H_1$ or $H_a$):** A statement that contradicts the null hypothesis and represents what the researcher suspects might be true instead.

3. **Test Statistic:** A function of the sample data that measures the discrepancy between the observed data and what would be expected under the null hypothesis.

4. **Significance Level ($\alpha$):** The maximum probability of rejecting a true null hypothesis (Type I error rate). It serves as the threshold for decision-making.

5. **P-value:** The probability of observing a test statistic as extreme or more extreme than the one calculated, assuming the null hypothesis is true.

6. **Decision Rule:** Reject $H_0$ if p-value $\leq \alpha$; otherwise, fail to reject $H_0$.

The rejection region, which will henceforth be denoted by RR, specifies the values of the test statistic for which the null hypothesis is to be rejected in favor of the alternative hypothesis. f for a particular sample, the computed value of the test statistic falls in the rejection region RR, we reject the null hypothesis $H_0$ and accept the alternative hypothesis $H_a$. If the value of the test statistic does not fall into the RR, we accept $H_0$.

Going back to our example above, the RR was that the $t$-statistic should be less than some critical value $t_{\alpha,n-1}$. Since we have a left-tailed test with $\alpha = 0.05$ and $df = 39$, the critical value is approximately $t_{0.05,39} \approx -1.685$. Therefore, the rejection region is $RR = \{t : t < -1.685\}$. Since our calculated test statistic $t = -2.64$ falls in this rejection region (because $-2.64 < -1.685$), we reject the null hypothesis, which is consistent with our p-value approach. Another exmaple: If $Y_1, Y_2, \ldots, Y_n \overset{\text{iid}}{\sim} N(\theta, 25)$, consider $H_0 : \theta = 70$ vs $H_a : \theta > 70$. Some possible **rejection regions** for this test are as follows:

1. Reject $H_0$ if $\overline{Y} > 72$.

2. Reject $H_0$ if $\max_i Y_i > 90$.

3. Reject $H_0$ if $\sum_{i=1}^{n} Y_i^2 > 390$.

Just like any sort of scentific based test, errors can be made. We define the following two types of errors. The first is rejecting the null hypothesis $H_0$ when it is actually true. The second is failing to reject the null hypothesis $H_0$ when the alternative hypothesis $H_a$ is actually true. We call thse Type I ($\alpha$) and Type II ($1 - \beta$) error rates respectively and denote them. Thus we have that

$$\alpha = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

and

$$\beta = P(\text{accept } H_0 \text{ when } H_0 \text{ is false}).$$

We then denote

$$\text{Power}(\theta) = W(\theta) = 1 - \beta = P(\text{reject } H_0 \text{ when } H_a \text{ is true}).$$

**Example 5.2.2**

Suppose you are testing $H_0 : p = \frac{1}{2}$ against $H_1 : p = \frac{2}{3}$ for a Binomial variable $X$ with $n = 3$. What values of $X$ would you assign to the rejection region (RR) if you wish to have $\alpha \leq \frac{1}{8}$ and you wish to minimize $\beta$ corresponding to the value of $\alpha$ selected?

*Solution* Recall that

$$P(X = x) = \binom{3}{x} p^x (1-p)^{3-x}.$$

Also recall that

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}) \quad \text{and} \quad \beta = P(\text{accept } H_0 \mid H_1 \text{ is true}).$$

Here are the pmf values in a table:

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P_0(X = x)$ $(p = \frac{1}{2})$ | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ |
| $P_1(X = x)$ $(p = \frac{2}{3})$ | $\frac{1}{27}$ | $\frac{6}{27}$ | $\frac{12}{27}$ | $\frac{8}{27}$ |

We see that

$$\alpha = p(\text{reject } H_0 \mid p = 0.5) = P(X \in RR \mid p = 0.5).$$

Thus we want to choose the rejection region to minimize $\beta$ while keeping $\alpha \leq \frac{1}{8}$.

Let's examine all possible rejection regions:

- RR = $\{0\}$: $\alpha = P(X = 0 \mid p = \frac{1}{2}) = \frac{1}{8}$, $\beta = P(X \neq 0 \mid p = \frac{2}{3}) = \frac{1}{27} + \frac{6}{27} + \frac{12}{27} = \frac{19}{27}$
- RR = $\{1\}$: $\alpha = P(X = 1 \mid p = \frac{1}{2}) = \frac{3}{8} > \frac{1}{8}$ (invalid)
- RR = $\{2\}$: $\alpha = P(X = 2 \mid p = \frac{1}{2}) = \frac{3}{8} > \frac{1}{8}$ (invalid)
- RR = $\{3\}$: $\alpha = P(X = 3 \mid p = \frac{1}{2}) = \frac{1}{8}$, $\beta = P(X \neq 3 \mid p = \frac{2}{3}) = \frac{1}{27} + \frac{6}{27} + \frac{12}{27} = \frac{19}{27}$
- RR = $\{0, 1\}$: $\alpha = \frac{1}{8} + \frac{3}{8} = \frac{1}{2} > \frac{1}{8}$ (invalid)
- RR = $\{0, 2\}$: $\alpha = \frac{1}{8} + \frac{3}{8} = \frac{1}{2} > \frac{1}{8}$ (invalid)
- RR = $\{0, 3\}$: $\alpha = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} > \frac{1}{8}$ (invalid)
- RR = $\{1, 2\}$: $\alpha = \frac{3}{8} + \frac{3}{8} = \frac{3}{4} > \frac{1}{8}$ (invalid)
- RR = $\{1, 3\}$: $\alpha = \frac{3}{8} + \frac{1}{8} = \frac{1}{2} > \frac{1}{8}$ (invalid)
- RR = $\{2, 3\}$: $\alpha = \frac{3}{8} + \frac{1}{8} = \frac{1}{2} > \frac{1}{8}$ (invalid)

Thus we see that the valid RR are $\{0\}$ and $\{3\}$. So we have that the RR with smallest $\beta$ is $\{3\}$. $\square$

**Example 5.2.3**

We are interested in testing whether or not a coin is balanced based on the number of heads $Y$ on 36 tosses of the coin. ($H_0 : p = 0.5$ versus $H_a : p = 0.7$). If we use the rejection region $|y - 18| \geq 4$, what is

**(a)** the value of $\alpha$?

**(b)** the value of $\beta$ if $p = 0.7$?

*Solution* For (a): We are given that the RR is $|y - 18| \geq 14$ or $y - 18 \leq -4$ or $y - 18 \geq 4$. Thus we have that

$$\begin{aligned}\alpha &= P(\text{reject } H_0 \mid H_0 \text{ is true}) \\ &= P(y - 18 \leq -4 \text{ or } y - 18 \geq 4 \mid p = 0.5) \\ &= P(Y \geq 22 \mid p = 0.5) + P(Y \leq 14 \mid p = 0.5) \\ &= 0.1214 + 0.1214 = 0.2428.\end{aligned}$$

We calculated this using the normal approximation to the Binomial distribution. For (b): We then see that

$$\beta = P(\text{accept } H_0 \mid H_1 \text{ is true}) = P(15 \leq Y \leq 21 \mid p = 0.7) = 0.0916$$

□

## 5.3 Neyman-Pearson Lemma

When you do hypothesis testing, you often have to decide between two competing simple hypotheses:

$$H_0 : \theta = \theta_0 \quad \text{or} \quad H_1 : \theta = theta_1.$$

The Neyman-Pearson lemma tells us how to choose the most powerful test for a given size $\alpha$ (probability of Type I error). What does "most powerful" mean? We first define power:

**Definition 5.3.1**

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 $\theta$. Then the *power* of the test, denoted by power($\theta$), is the probability that the test will lead to rejection of $H_0$ when the actual parameter value is $\theta$. That is,

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

A test is most powerful (MP) at level $\alpha$ if:

- It has size $\alpha$ (probability of rejecting $H_0$ when $H_0$ is true is $\alpha$)
- And among all tests of size $\alpha$, it has the largest power:

$$\text{Power}(\theta_a) = 1 - \beta(\theta_a) = 1 - P(\text{accept } H_0 \mid \theta = \theta_a).$$

In other words if you fix the allowed false positive rate, the MP test gives you the highest true positive rate. The power is defined by:

**Lemma 5.3.2: Neyman-Pearson**

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 $Y_1, Y_2, \ldots, Y_n$ from a distribution with parameter $\theta$.
Let $L(\theta)$ denote the likelihood of the sample when the value of the parameter is $\theta$. Then, for a given $\alpha$, the test that maximizes the power at $\theta_a$ 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 $\alpha$. Such a test is a most powerful $\alpha$-level test for $H_0$ versus $H_a$.

*Proof.* The proof is ommitted. □

From this we do a few examples to show the usefulness.

**Example 5.3.3**

Suppose that $Y$ represents a single observation from a population with probability density function given by

$$f(y \mid \theta) = \theta y^{\theta-1}, \quad 0 < y < 1$$

and zero otherwise.
Find the most powerful test with significance level $\alpha = 0.05$ to test

$$H_0 : \theta = 2 \quad \text{versus} \quad H_a : \theta = 1.$$

Provide the Rejection Region (RR) explicitly.

*Solution* Using the Neyman-Pearson lemma we know that most powerfuul test with significance level $\alpha = 0.05$ is

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{2y^{2-1}}{Y^{1-1}} = 2y < k \quad \text{or} \quad y < \frac{k}{2}.$$

Thus the RR is $\{y : y < k/2\}$. Then we find $k$ based on the desired significance level.

$$\begin{aligned} \alpha &= P(\text{reject } H_0 \mid H_0 \text{ is true}) \\ 0.05 &= P(y < k/2 \mid \theta = 2) \\ &= \int_0^{k/2} 2y dy \\ &= k^2. \end{aligned}$$

Solving for $k$ we find that $k = \sqrt{0.05} \approx 0.2236$. Thus the RR is $\{y : y < 0.2236\}$. □

**Example 5.3.4**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from $N(\mu, \sigma^2 = 1)$. Find the best critical region (most powerful test) of size $\alpha = 0.05$ for testing the hypothesis

$$H_0: \ \mu = 0 \quad \text{vs} \quad H_a: \ \mu = 5.$$

(In this example, $\mu_0 = 0$ and $\mu_a = 5$.)

*Solution* Using the Neyman-Pearson lemma we know that most powerfuul test with significance level $\alpha = 0.05$ is

$$\begin{aligned}
\frac{L(\theta_0)}{L(\theta_1)} &= \frac{\prod_{i=1}^n Z}{\prod_{i=1}^n N(5,1)} \\
&= \frac{\frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2}\right)\sum_{i=1}^n (y_i)^2\right]}{\frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2}\right)\sum_{i=1}^n (y_i-5)^2\right]} \\
&= \exp\left[-\left(\frac{1}{2}\right)\sum_{i=1}^n (y_i)^2 + \left(\frac{1}{2}\right)\sum_{i=1}^n (y_i-5)^2\right] \\
&= \exp\left[\frac{-1}{2}\sum_{i=1}^n (y_i)^2 + \frac{1}{2}\sum_{i=1}^n (y_i)^2 - 5\sum_{i=1}^n y_i + \frac{25n}{2}\right] \\
&= \exp\left[-5\sum_{i=1}^n y_i + \frac{25}{n}\right] < k
\end{aligned}$$

Let $k_1 = k \cdot e^{-25/n}$. Thus we have

$$\begin{aligned}
e^{-5\sum_{i=1}^n y_i} &< k_1 \\
-5\sum_{i=1}^n y_i &< \ln(k_1) \\
\sum_{i=1}^n y_i &\geq -\frac{\ln(k_1)}{5} \\
\overline{Y} &\geq -\frac{\ln(k_1)}{5n} = k_2.
\end{aligned}$$

Thus the RR is $\{y : \overline{y} \geq k_2$. Then we find $k_2$ based on the desired significance level.

$$\begin{aligned}
\alpha &= P(\text{reject } H_0 \mid H_0 \text{ is true}) \\
0.05 &= (\overline{y} \geq k_2 \mid \mu = 0)\,.
\end{aligned}$$

To ccompute this we standardize sample mean since $\overline{Y} \sim N(0, 1/n)$. Thus

$$\begin{aligned}
0.05 &= P\left(\frac{\overline{Y}-0}{1/\sqrt{n}} \geq \frac{k_2 - 0}{1\sqrt{n}}\right) \\
&= P\left(Z \geq \frac{k_2}{1\sqrt{n}}\right).
\end{aligned}$$

Notice that

$$z_\alpha = \frac{k_2}{1\sqrt{n}} \quad \text{or} \quad \frac{z_\alpha}{\sqrt{n}} = k_1.$$

Note we have that $z_\alpha = 1.645$ and so the RR is $\{y : \overline{y} \geq 1.645/\sqrt{n}\}$ □

Note we have this final definition which is sometimes used in hypothesis testing.

**Definition 5.3.5**

If a random sample is taken from a distribution with parameter $\theta$, a hypothesis is said to be a simple hypothesis if that hypothesis uniquely specifies the distribution of the population from which the sample is taken. Any hypothesis that is not a simple hypothesis is called a composite hypothesis.

### 5.3.1 Uniformly Most Powerful

The Neyman-Pearson lemma gives us the most powerful level $\alpha$ test when testing two simple hypotheses: $H_0 : \theta = \theta_0$ vs $H_a : \theta = \theta_a$. However if the resulting test is independent of $\theta_a$ (that is we get the same result for all $\theta_a \in \Omega_a$, for example all $\mu \in (-\infty, \mu)$ reffering to our first example in the section) we then say the resulting test is a "Uniformly Most Powerful (UMP)" level $\alpha$ test. If the test obtained from Neyman-Pearson lemma changes with (depends on ) with $\theta_a$ then we say there does not exist a UMP for thet testing hypothesis.

**Example 5.3.6**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from $N(\theta, \sigma^2 = 1)$. Find the uniformly most powerful test with significance level $\alpha$, if it exists, for testing

$$H_0 : \theta = 0 \quad \text{vs} \quad H_a : \theta > 0.$$

Provide Rejection Region (RR), **explicitly**.

*Solution* To find the UMP we must first use the Neyman-Pearson lemma. We first fix a $\theta_a > 0$ for the alternative hypothesis since we have a simple hypothesis vs a composite hypothesis.

$$\begin{aligned}
\frac{L(\theta_0)}{L(\theta_a)} &= \frac{\frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2}\right)\sum_{i=1}^{n} y^2\right]}{\frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2}\right)\sum_{i=1}^{n} (y-\theta_a)^2\right]} \\
&= \exp\left[\frac{-1}{2}\sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} y_i^2 - \theta_a \sum_{i=1}^{n} y_i + \frac{n\theta_a^2}{2}\right] \\
&= e^{-n\theta_a \overline{Y}} \cdot e^{n\theta_a^2/2} < k.
\end{aligned}$$

Letting $k_1 = e^{-n\theta_a^2/2} \cdot k$ we have that

$$\begin{aligned}
e^{-n\theta_a \overline{Y}} &< k_1 \\
-n\theta_a \overline{Y} &< \ln(k_1) \\
\overline{Y} &\geq k_2
\end{aligned}$$

where $k_2 = -\ln(k_1)/n\theta_a$. Thus the RR is $\{y : \overline{y} \geq k_2\}$. We now find $k_2$ by finding the $\alpha$ level.

$$\begin{aligned}
\alpha &= P(\text{reject } H_0 \mid H_0 \text{ is true}) \\
0.05 &= P\left(\overline{y} \geq k_2 \mid \mu = 0\right) \\
0.05 &= P\left(Z \geq \frac{k_2 - 0}{1/\sqrt{n}}\right).
\end{aligned}$$

Thus we have that $Z_\alpha = \sqrt{n}k_2$ or $k_2 = z_\alpha/\sqrt{n}$. Thus the RR is $\{y : \overline{y} \geq z_\alpha/\sqrt{n}\}$. □

**Example 5.3.7**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from $N(\theta, \sigma^2 = 1)$. Find the uniformly most powerful test with significance level $\alpha$, if it exists, for testing

$$H_0 : \theta = 0 \quad \text{vs} \quad H_a : \theta < 0.$$

*Solution* Again we have a simple hypothesis vs a composite hypothesis. We can use the Neyman-Pearson Lemma to find a most powerful test and see if it depends on $\theta_a$ to see if it is uniformly most powerful test. First fix $\theta_a < 0$. Then we follow the same identical steps in Example 5.3.6 to get

$$e^{n\theta_a \overline{Y}} \cdot e^{-n\theta_a^2/2} < k.$$

Letting $k_1 = e^{n\theta_a^2/2}k$ we get that

$$\begin{aligned}
e^{n\theta_a \overline{Y}} &< k_1 \\
n\theta_a \overline{Y} &< \ln(k_1) \\
\overline{Y} &< k_2
\end{aligned}$$

where $k_2 = \ln(k_1)/n\theta_a$. Thus the RR is $\{y : \overline{y} < k_2\}$. We now solve for $k_2$ by finding the $\alpha$ level.

$$\begin{aligned}
\alpha &= P(\text{reject } H_0 \mid H_0 \text{ is true}) \\
0.05 &= P\left(\overline{y} < k_2 \mid \mu = 0\right) \\
0.05 &= P\left(Z < \frac{k_2 - 0}{1/\sqrt{n}}\right).
\end{aligned}$$

Thus we have that $z_{1-\alpha} = \sqrt{n}k_2$ or $k_2 = z_{1-\alpha}/\sqrt{n}$. Thus the RR is $\{y : \overline{y} < z_{1-\alpha}/\sqrt{n}\}$. □

**Example 5.3.8**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from $N(\theta, \sigma^2 = 1)$. Find the uniformly most powerful test with significance level $\alpha$, if it exists, for testing

$$H_0 : \theta = 0 \quad \text{vs} \quad H_a : \theta \neq 0.$$

*Solution* Because the most powerful region for $\theta > 0$ is $\overline{Y} > z_{1-\alpha}/\sqrt{n}$ and for $\theta < 0$ is $\overline{Y} < -z_{1-\alpha}/\sqrt{n}$, there is no single test that is simultaneously most powerful for all $\theta \neq 0$. Thus, a nontrivial UMP test does not exist. The best we can do is the UMPU test, which rejects when $|\overline{Y}| > z_{1-\alpha/2}/\sqrt{n}$. □

## 5.4 Likelihood Ratio Tests

So far we have established the Neyman-Pearson lemma which provides us a method of constructing most powerful tests for simple hypothesis when the underlying probability distribution is known except for a single unknown parameter. We also found that in certain we can use this to find Uniformly Most Powerful tests when we have composite hypothesis with again just one parameter. However in many cases we have that there is more than one parameter. In this section we provide a general method that works for simple and composite hypothesis and one or more than one unknown parameters.

Suppose that we have a ranom sample that is selected from a distribution with a likelihood function $L(y_1, y_2, \ldots, y_n \mid \theta_1, \theta_2, \ldots, \theta_k)$ with $k$ unknown parameters. For notation I will use $\Theta$ to denote a $k-$vector containing the unknown parameters so we have $L(\Theta)$. As we have seen in examples in the previous section, we sometimes only care about one of the parameters and not the others. For example in the previous section we looked at a random sample that was distributed normally with mean $\theta$ and variance $\sigma^2$ and $\Theta = (\mu, \sigma^2)$. Our hypothesis tests were only for $\mu$ in which case $\sigma^2$ did not matter to us. We call these unknown parameters that are not of interest to use nuisance parameters. Formally, let $\Omega_0$ denote the parameter set specified by $H_0$ and let $\Omega_a$ be the (disjoint) set specified by $H_a$. Their union $\Omega = \Omega_0 \cup \Omega_a$ is the collection of all parameter values under consideration. Define

$$L(\Omega_0) = \sup_{\Theta \in \Omega_0} L(\Theta), \qquad L(\Omega) = \sup_{\Theta \in \Omega} L(\Theta).$$

(Write "sup" but in regular finite problems it is a maximum.) If the best attainable likelihood under $H_0$ equals the overall maximum ($L(\Omega_0) = L(\Omega)$), then the data are as well explained without leaving $\Omega_0$ and there is no evidence against $H_0$. If instead $L(\Omega_0)$ is noticeably smaller than $L(\Omega)$, the data are better explained by parameter values outside $\Omega_0$ (in $\Omega_a$) and we have evidence against $H_0$.

This motivates the Likelihood Ratio Test (LRT).

**Definition 5.4.1 : Likelihood Ratio Test (LRT)**

Define $\lambda$ by

$$\lambda = \frac{L(\Omega_0)}{L(\Omega)} = \frac{\max\limits_{\Theta \in \Omega_0} L(\Theta)}{\max\limits_{\Theta \in \Omega} L(\Theta)}.$$

A likelihood ratio test of $H_0 : \Theta \in \Omega_0$ versus $H_a : \Theta \in \Omega_a$ employs $\lambda$ as a test statistic, and the rejection region is determined by $\lambda \leq k$.

It can be shown that $0 \leq \lambda \leq 1$. A value of $\lambda$ close to zero indicates that the likelihood of the sample is much smaller under $H_0$ than it is under $H_a$. Therefore, the data suggest favoring $H_a$ over $H_0$. The actual value of $k$ is chosen so that $\alpha$ achieves the desired value. We illustrate the mechanics of this method with the following example.

**Example 5.4.2**

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

*Solution* First we see that $\Theta = (\mu, \sigma^2)$ and $\Omega_0 = \{(\mu_0, \sigma^2) : \sigma^2 > 0\}$ and $\Omega_a = \{(\mu, \sigma^2) : \mu > 0, \sigma^2 > 0\}$. Thus then $\Omega = \Omega_0 \cup \Omega_a = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma^2 > 0\}$. Here we can consider the constant value of the variance as a nuisance parameter. We now find $L(\hat{\Omega}_0)$ and $L(\hat{\Omega})$. We first get that

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

Looking at when $\Theta \in \Omega_0$ we have that $\mu = \mu_0$ and the variance $\sigma^2 > 0$. So then to find $L(\hat{\Omega}_0)$, we simply find the variance that maximizes $L(\mu, \sigma^2)$ with considering the constraint of $\mu = \mu_0$. Reffering to Example 4.5.2 we see that the value of $\sigma^2$ that maximizes the likelihood function $L(\mu_0, \sigma^2)$ is

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

Thus we have that $L(\hat{\Omega}_0)$ is obtained when we let $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_0)^2$ which gives us

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

We now find $L(\hat{\Omega})$. Reffering again to Example 4.5.2 we find the log-likelihood and find the partial derivatives with respect to $\mu$ and $\sigma^2$

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

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

We then just need to maximize this over the set $\Omega = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma^2 > 0\}$. We see that. For the unrestricted normal model the MLEs are $\hat{\mu} = \overline{y}$ and $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$. Under the constraint $\mu \geq \mu_0$ the constrained MLE of $\mu$ is

$$\hat{\mu}_c = \max(\overline{y}, \mu_0),$$

and the corresponding $\hat{\sigma}_c^2 = \dfrac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_c)^2$.

Thus the likelihood function is obtained by replacing $\mu$ with $\hat{\mu}_c$ and $\sigma^2$ with $\hat{\sigma^2}_c$.

$$L(\hat{\Omega}) = \begin{cases} L(\mu_0, \hat{\sigma}_0^2) & \text{if } \overline{y} < \mu_0, \\ L(\overline{y}, \hat{\sigma}^2) & \text{if } \overline{y} \geq \mu_0, \end{cases}$$

where

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

Therefore the likelihood ratio is

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \begin{cases} 1, & \overline{y} < \mu_0, \\ \left(\dfrac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}, & \overline{y} \geq \mu_0. \end{cases}$$

For $\overline{y} \geq \mu_0$ use the identity

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

so

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + (\overline{y} - \mu_0)^2} = \left(1 + \frac{(\overline{y} - \mu_0)^2}{\hat{\sigma}^2}\right)^{-1}.$$

Let $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$ (the usual unbiased sample variance), then $\hat{\sigma}^2 = \dfrac{n-1}{n}S^2$ and for $\overline{y} \geq \mu_0$,

$$\lambda = \left(1 + \frac{n(\overline{y} - \mu_0)^2}{(n-1)S^2}\right)^{-n/2} = \left(1 + \frac{T^2}{n-1}\right)^{-n/2},$$

where

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}.$$

Piecewise,

$$\lambda = \begin{cases} 1, & T < 0, \\ \left(1 + \dfrac{T^2}{n-1}\right)^{-n/2}, & T \geq 0. \end{cases}$$

Because $\lambda$ is a decreasing function of $T$ for $T \geq 0$, the rejection region $\{\lambda \leq k\}$ is equivalent to $\{T \geq c\}$ for some constant $c$. Choose $c$ so that

$$P_{H_0}(T \geq c) = \alpha,$$

i.e. $c = t_{n-1,\,1-\alpha}$ (the upper $100\alpha\%$ critical value of a $t_{n-1}$ distribution).

Hence the likelihood ratio test rejects $H_0$ when

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}} \geq t_{n-1,\,1-\alpha},$$

which is exactly the usual one–sided $t$-test. □

The likelihood ratio test for most practical problems produces the best possible test in terms of powefullness. However the distribution for the test statistic produced by the likelihood ratio test is not always known. However if the sample size is large enough we can obtain an approximation to the distribution of $\lambda$ if some regularity condtions conditions are satisfied - recall from Fisher Information. For most distributions we encounter these condtions will be satistified.

**Theorem 5.4.3**

Let $Y_1, Y_2, \ldots, Y_n$ have joint likelihood function $L(\Theta)$. Let $r_0$ denote the number of free parameters that are specified by $H_0 : \Theta \in \Omega_0$ and let $r$ denote the number of free parameters specified by the statement $\Theta \in \Omega$. Then, for large $n$, $-2\ln(\lambda)$ has approximately a $\chi^2$ distribution with $r_0 - r$ df.

*Proof.* The proof of this theorem is beyond the scope of this text. □

Since the RR is $\{\lambda < k\}$ we can write $\{-2\ln(\lambda) > -2\ln(k) = k_1\}$. If we want the $\alpha$-level test them the above theorem tells us that $k_1 \approx \chi^2_\alpha$ with $r_o - r$ df.

**Example 5.4.4**

Suppose that an engineer wishes to compare the number of complaints per week filed by union stewards for two different shifts at a manufacturing plant. One hundred independent observations on the number of complaints gave means $\bar{x} = 20$ for shift 1 and $\bar{y} = 22$ for shift 2. Assume that the number of complaints per week on the $i$th shift has a Poisson distribution with mean $\theta_i$, for $i = 1, 2$. Use the likelihood ratio method to test $H_0 : \theta_1 = \theta_2$ versus $H_a : \theta_1 \neq \theta_2$ with $\alpha \approx 0.01$.

*Solution* To find the likelihood ratio we find the likelihood function of the samples. We have that

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} \frac{\theta_1^{x_i} e^{-\theta_1}}{x_i!} \cdot \prod_{i=1}^{n} \frac{\theta_2^{y_i} e^{-\theta_2}}{y_i!} = \left(\frac{1}{k}\right) \theta_1^{\sum_{i=1}^{100} x_i} e^{-100\theta_1} \theta_1^{\sum_{i=1}^{100} y_i} e^{-100\theta_2},$$

where $k = x_1! \cdot x_2! \dots x_n! y_1! \cdot y_2! \dots y_n!$. In this example we have that $\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2 = \theta\}$ where $\theta$ is unknown. Thus considering the likelihood function under the set null hypothesis we get that

$$L(\Theta) = \left(\frac{1}{k}\right) \theta^{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i} e^{-200\theta}.$$

We now maximize this over the set $\Omega_0$. We know that this is maximized when $\theta$ is equal to its MLE. We find that the MLE is $\hat{\theta} = \frac{1}{2}(\overline{x} + \overline{y})$. Next we see that $\{(\theta_1, \theta_2) : \theta_1 > 0, \theta_2 > 0\}$. Using the general likelihood $L(\theta_1, \theta_2)$, a function of both $\theta_1$ and $\theta_2$, we see that $L(\theta_1, \theta_2)$ is maximized when $\hat{\theta}_1 = \bar{x}$ and $\hat{\theta}_2 = \bar{y}$, respectively. That is, $L(\theta_1, \theta_2)$ is maximized when both $\theta_1$ and $\theta_2$ are replaced by their maximum likelihood estimates. Thus,

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{k^{-1} (\hat{\theta})^{n\bar{x} + n\bar{y}} e^{-2n\hat{\theta}}}{k^{-1} (\hat{\theta}_1)^{n\bar{x}} (\hat{\theta}_2)^{n\bar{y}} e^{-n\hat{\theta}_1 - n\hat{\theta}_2}} = \frac{(\hat{\theta})^{n\bar{x} + n\bar{y}}}{(\bar{x})^{n\bar{x}} (\bar{y})^{n\bar{y}}}.$$

Since $\lambda$ is complicated finding its exact distribution is hard. So we instead use the approximation we established above. We find that

$$\lambda = \frac{21^{(100)(20+22)}}{20^{(100)(20)} \cdot 22^{(100)(22)}}.$$

Thus we have that

$$-2\ln(\lambda) = -2\left[4200\ln(21) - 2000\ln(20) - 2200\ln(22)\right] = 9.53.$$

The above theorem tells us that $-2\ln(\lambda)$ has a $\chi^2$ distribution with $r_0 - r = 1 - 0$ df. So the rejection region is $\{-2\ln(\lambda) > \chi^2_{0.01} = 6.635\}$. Since our observed p-value is greater than that we can reject the null hypothesis and conclude at a $\alpha = 0.01$ significance that the mean of the numbers of complaints filed by the union stewards differ. $\square$

## 5.5 Practice Problems

**10.2**

An experimenter has prepared a drug dosage level that she claims will induce sleep for 80% of people suffering from insomnia. After examining the dosage, we feel that her claims regarding the effectiveness of the dosage are inflated. In an attempt to disprove her claim, we administer her prescribed dosage to 20 insomniacs and we observe $Y$, the number for whom the drug dose induces sleep. We wish to test the hypothesis $H_0 : p = .8$ versus the alternative, $H_a : p < .8$. Assume that the rejection region $\{y \leq 12\}$ is used.

(a) In terms of this problem, what is a type I error?

(b) Find $\alpha$.

(c) In terms of this problem, what is a type II error?

(d) Find $\beta$ when $p = .6$.

(e) Find $\beta$ when $p = .4$.

*Solution* For (a): The Type I error is defined as

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}) = P(y < 12 \mid p = 0.8).$$

For (b): Assuming that $Y \sim Bin(n, p)$ we can see that

$$\alpha = \sum_{i=1}^{12} \binom{20}{y} (0.8)^y 0.2^{20-y} \approx 0.0321.$$

For (c): The Type II error is defined as

$$\beta = P(\text{accept } H_0 \mid H_a \text{ is true}).$$

That is concluding that $p = 0.8$ when the true $p < 0.8$. For (d): We see that then

$$P(Y \geq 13 \mid p = 0.6) = \sum_{i=13}^{20} \binom{20}{y} (0.6)^y 0.4^{20-y} \approx 0.4159.$$

For (d):

$$P(Y \geq 13 \mid p = 0.4) = \sum_{i=13}^{20} \binom{20}{y} (0.4)^y 0.6^{20-y} \approx 0.0210.$$

□

**10.3**

Refer to Exercise 10.2.

(a) Find the rejection region of the form $\{y \leq c\}$ so that $\alpha \approx .01$.

(b) For the rejection region in part (a), find $\beta$ when $p = .6$.

(c) For the rejection region in part (a), find $\beta$ when $p = .4$.

*Solution* We need to find the rejection region so that

$$\alpha = 0.01 = P(y \leq c \mid p = 0.8) = \sum_{i=1}^{c} \binom{20}{y} (0.8)^y 0.2^{20-y}.$$

Letting $c = 11$ we see that

$$P_{0.8}(Y \leq 11) = 0.00998 \approx 0.01.$$

For (b): We now find $\beta$ when $p = 0.6$.

$$\beta = P(Y \geq 12 \mid p = 0.6) = \sum_{i=12}^{20} \binom{20}{y} (0.6)^y 0.4^{20-y} \approx 0.5956,$$

$$\beta = P(Y \geq 12 \mid p = 0.4) = \sum_{i=12}^{20} \binom{20}{y} (0.4)^y 0.6^{20-y} \approx 0.0565.$$

$\square$

**10.91**

Let $Y_1, Y_2, \ldots, Y_{20}$ be a random sample of size $n = 20$ from a normal distribution with unknown mean $\mu$ and known variance $\sigma^2 = 5$. We wish to test $H_0\colon \mu = 7$ versus $H_a\colon \mu > 7$.

(a) Find the uniformly most powerful test with significance level .05.

(b) For the test in part (a), find the power at each of the following alternative values for $\mu : \mu_a = 7.5, 8.0, 8.5$ and $9.0$.

*Solution* To find the UMP we first find the MP using Neyman-Pearson Lemma. First fix a $\mu_a > 7$.

$$\lambda = \frac{L(\mu = 7)}{L(\mu = \mu_a)} = \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - 7)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu_a)^2\right]} = \exp\left\{-\frac{1}{2\sigma^2}\left[\sum(y_i - 7)^2 - \sum(y_i - \mu_a)^2\right]\right\}.$$

With $\sigma^2 = 5$,

$$\sum_{i=1}^{n}(y_i - 7)^2 - \sum_{i=1}^{n}(y_i - \mu_a)^2 = (2\mu_a - 14)\sum_{i=1}^{n} y_i + n(49 - \mu_a^2),$$

so

$$\lambda = \exp\left\{-\frac{1}{10}\left[(2\mu_a - 14)\sum_{i=1}^{n} y_i + n(49 - \mu_a^2)\right]\right\}.$$

For any fixed $\mu_a > 7$, $2\mu_a - 14 > 0$, hence $\lambda$ is a (strictly) decreasing function of $\sum y_i$ and the most powerful level $\alpha$ test rejects for large $\sum y_i$, i.e. large $\overline{Y}$. Under $H_0$, $\overline{Y} \sim N\big(7, \frac{5}{n}\big)$, so the size-$\alpha$ critical value is

$$c = 7 + z_{1-\alpha}\sqrt{\frac{5}{n}}.$$

Because this does not depend on the particular $\mu_a > 7$, the test is UMP for $H_a : \mu > 7$:

$$\text{Reject } H_0 \text{ if } \overline{Y} > 7 + z_{1-\alpha}\sqrt{\tfrac{5}{n}}.$$

With $n = 20$ and $\alpha = 0.05$, $z_{0.95} = 1.6449$ and $\sqrt{5/20} = 0.5$, so

$$c = 7 + 1.6449(0.5) = 7.82245.$$

Thus (a) reject $H_0$ if $\overline{Y} > 7.82245$.

(b) Power at $\mu = \mu_a$:

$$\text{Power}(\mu_a) = P_{\mu_a}(\overline{Y} > 7.82245) = P\left(Z > \frac{7.82245 - \mu_a}{0.5}\right), \quad Z \sim N(0,1).$$

$$\begin{aligned}
\mu_a = 7.5 : z &= \frac{7.82245 - 7.5}{0.5} = 0.6449 \;\Rightarrow\; \text{Power} = 1 - \Phi(0.6449) \approx 0.260. \\
\mu_a = 8.0 : z &= \frac{7.82245 - 8.0}{0.5} = -0.3551 \;\Rightarrow\; \text{Power} = \Phi(0.3551) \approx 0.639. \\
\mu_a = 8.5 : z &= \frac{7.82245 - 8.5}{0.5} = -1.3551 \;\Rightarrow\; \text{Power} = \Phi(1.3551) \approx 0.912. \\
\mu_a = 9.0 : z &= \frac{7.82245 - 9.0}{0.5} = -2.3551 \;\Rightarrow\; \text{Power} = \Phi(2.3551) \approx 0.991.
\end{aligned}$$

□

**10.95**

Suppose that we have a random sample of four observations from the density function

$$f(y \mid \theta) = \begin{cases} \left(\frac{1}{2\theta^3}\right) y^2 e^{-y/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the rejection region for the most powerful test of $H_0\colon \theta = \theta_0$ against $H_a\colon \theta = \theta_a$, assuming that $\theta_a > \theta_0$. *[Hint: Make use of the $\chi^2$ distribution.]*

(b) Is the test given in part (a) uniformly most powerful for the alternative $\theta > \theta_0$?

*Solution* Each $Y_i \sim \text{Gamma}(3, \theta)$ (shape 3, scale $\theta$). The joint likelihood (ignoring factors not involving $\theta$):

$$L(\theta) \propto \theta^{-12} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{4} Y_i\right).$$

For simple hypotheses,

$$\frac{L(\theta_0)}{L(\theta_a)} = \left(\frac{\theta_a}{\theta_0}\right)^{12} \exp\left[-\left(\frac{1}{\theta_0} - \frac{1}{\theta_a}\right)\sum_{i=1}^{4} Y_i\right].$$

Since $\theta_a > \theta_0$ then $(1/\theta_0 - 1/\theta_a) > 0$, this ratio is a decreasing function of $T = \sum_{i=1}^4 Y_i$. Thus the NP most powerful test rejects $H_0$ for large $T$.

Distribution under $H_0$: $T \sim \text{Gamma}(12, \theta_0)$. Note that

$$\frac{2T}{\theta_0} \sim \chi^2_{24},$$

because a $\chi^2_\nu$ is Gamma$(\nu/2, 2)$. Size $\alpha$ test:

$$\text{Reject } H_0 \text{ if } T > c, \quad P_{\theta_0}(T > c) = \alpha.$$

Equivalently,

$$\frac{2T}{\theta_0} > \chi^2_{24,\,1-\alpha} \quad \text{or} \quad c = \frac{\theta_0}{2}\chi^2_{24,\,1-\alpha}.$$

So the rejection region is

$$RR = \left\{ \sum_{i=1}^{4} Y_i > \frac{\theta_0}{2}\chi^2_{24,\,1-\alpha} \right\}.$$

(b) The family $\{f(y \mid \theta)\}$ with fixed shape (known) and scale parameter $\theta$ is an exponential family with monotone likelihood ratio in $T = \sum Y_i$. Thus rejection region of the form $T > c$ is UMP for testing $H_0 : \theta = \theta_0$ vs $H_a : \theta > \theta_0$. Hence the test is uniformly most powerful for the one–sided composite alternative $\theta > \theta_0$. □

# 6 Tutorial Problems

## 6.1 Tutorial 1

**Question 1**

Let $X$ have a $\chi^2(r)$ distribution. If $k > -\frac{r}{2}$, then $E\left[X^k\right]$ exists. Prove that

$$E\left[X^k\right] = \frac{2^k \Gamma\left(\frac{r}{2} + k\right)}{\Gamma\left(\frac{r}{2}\right)}, \quad \text{if } k > -\frac{r}{2}$$

*Solution* We know that if $X \sim \chi^2(r)$, then its probability density function is

$$f_X(x) = \frac{1}{2^{r/2}\Gamma(r/2)} x^{r/2-1} e^{-x/2}, \quad x > 0.$$

The expectation is

$$E[X^k] = \int_0^\infty x^k f_X(x)\, dx = \int_0^\infty x^k \frac{1}{2^{r/2}\Gamma(r/2)} x^{r/2-1} e^{-x/2} dx.$$

Combine powers of $x$:

$$= \frac{1}{2^{r/2}\Gamma(r/2)} \int_0^\infty x^{k+r/2-1} e^{-x/2} dx.$$

Let $y = x/2$, so $x = 2y$, $dx = 2dy$:

$$= \frac{1}{2^{r/2}\Gamma(r/2)} \int_0^\infty (2y)^{k+r/2-1} e^{-y} \cdot 2dy$$

$$= \frac{1}{2^{r/2}\Gamma(r/2)} 2 \cdot 2^{k+r/2-1} \int_0^\infty y^{k+r/2-1} e^{-y} dy$$

$$= \frac{2^k}{\Gamma(r/2)} \int_0^\infty y^{k+r/2-1} e^{-y} dy$$

The integral is the definition of the Gamma function:

$$\int_0^\infty y^{a-1}e^{-y}dy = \Gamma(a)$$

with $a = k + r/2$. Therefore,

$$E[X^k] = \frac{2^k\Gamma(k + r/2)}{\Gamma(r/2)}$$

as required. □

**Question 2**

Let $X$ have the uniform distribution with pdf

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the cdf of $Y = -2\log X$. What is the pdf of $Y$?

*Solution* This is a simple transformation question from STA256. Let $Y = -2\log X$, so the inverse is $X = e^{-Y/2}$. The derivative (Jacobian) is $\frac{dX}{dY} = -\frac{1}{2}e^{-Y/2}$, so the absolute value is $\frac{1}{2}e^{-Y/2}$.

Since $X$ is uniform on $(0,1)$, $Y$ ranges from 0 (when $X = 1$) to $\infty$ (when $X \to 0^+$). The pdf is:

$$f_Y(y) = f_X(e^{-y/2})\left|\frac{d}{dy}e^{-y/2}\right| = 1 \cdot \frac{1}{2}e^{-y/2} = \frac{1}{2}e^{-y/2}$$

for $y > 0$, and zero elsewhere. □

## 6.2 Tutorial 2

**Question 1**

1. If $U$ has a $\chi^2$ distribution with $v$ degrees of freedom, find $E(U)$ and $\mathrm{Var}(U)$.
2. Let $Y_1, \dots, Y_n$ be a random sample from the normal distribution with mean $\mu$ and variance $\sigma^2$. Find $E(S^2)$ and $\mathrm{Var}(S^2)$, where $S$ is defined as:

$$S^2 = \frac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{n-1}$$

*Solution* For (1): We begin by finding $E(U)$.

If $U \sim \chi^2(v)$, then the mean and variance are well-known:

$$E(U) = v, \qquad \mathrm{Var}(U) = 2v$$

For (2): Let $Y_1, \dots, Y_n$ be a random sample from $N(\mu, \sigma^2)$. The sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$$

It is a standard result that

$$E(S^2) = \sigma^2$$

and

$$\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

$\square$

**Question 2**

Let $\overline{Y}$ and $S^2$ be the mean and the variance of a random sample of size 25 from $N(\mu = 3, \sigma^2 = 100)$. Find

$$P\left((0 < \overline{Y} < 6) \cap (55.2 < S^2 < 145.6)\right).$$

**Hint:** recall the following facts:

1. $\overline{Y}$ and $S^2$ are independent.
2. $\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.
3. $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1)$.

*Solution* Since the sample mean and sample variance are independent we see that

$$P\left((0 < \overline{Y} < 6) \cap (55.2 < S^2 < 145.6)\right) = P(0 < \overline{Y} < 6) \cdot P(55.2 < S^2 < 145.6).$$

Since the population is normally distributed we can easily find the first probability: For $\overline{Y}$, since $\overline{Y} \sim N(3, 100/25) = N(3, 4)$, we standardize:

$$P(0 < \overline{Y} < 6) = P\left(\frac{0-3}{2} < Z < \frac{6-3}{2}\right) = P(-1.5 < Z < 1.5)$$

where $Z \sim N(0,1)$. Using standard normal tables:

$$P(-1.5 < Z < 1.5) = \Phi(1.5) - \Phi(-1.5) = 2\Phi(1.5) - 1$$

where $\Phi$ is the standard normal cdf.

For $S^2$, recall that $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1)$ with $n = 25$, $\sigma^2 = 100$:

$$24 \cdot \frac{S^2}{100} \sim \chi^2(24)$$

So,

$$P(55.2 < S^2 < 145.6) = P\left(24 \cdot \frac{55.2}{100} < Q < 24 \cdot \frac{145.6}{100}\right)$$

where $Q \sim \chi^2(24)$. Compute the bounds:

$$24 \cdot \frac{55.2}{100} = 13.248, \qquad 24 \cdot \frac{145.6}{100} = 34.944$$

So,

$$P(55.2 < S^2 < 145.6) = P(13.248 < Q < 34.944)$$

where $Q \sim \chi^2(24)$. This is

$$P(13.248 < Q < 34.944) = F_Q(34.944) - F_Q(13.248)$$

where $F_Q$ is the cdf of $\chi^2(24)$.

Putting it all together:

$$P\left((0 < \overline{Y} < 6) \cap (55.2 < S^2 < 145.6)\right) = [2\Phi(1.5) - 1] \cdot [F_Q(34.944) - F_Q(13.248)] \approx 0.73644.$$

□

## 6.3 Tutorial 3

**Question 1**

If the random variable $F$ has an F-distribution with $r_1 = 5$, $r_2 = 10$, the degrees of freedom of numerator and denominator respectively. Find $a, b$ such that

$$P(F \leq a) = 0.05, \quad P(a < F < b) = 0.9.$$

*Solution* We first find the value of $a$ such that $P(F \leq a) = 0.05$. However we only have the tail probabilties for percentage points for this distribution so we instead use the identity: $P(U_1/U_2 \leq a) = P(U_2 \leq U_1 \geq 1/a)$. Thus we see that

$$P(F \leq a) = P\left(\frac{1}{F} \geq 1/a\right) = 0.05.$$

Then we see that $1/F \sim F(10, 5)$. Thus using the table we find that

$$\frac{1}{a} = 4.74 \quad \text{or} \quad a = \frac{1}{4.74} \approx 0.211.$$

□

**Question 2**

Let $Y_1, Y_2, \ldots, Y_5$ be a random sample of size 5 from a normal population with mean 0 and variance 1. Let $\overline{Y} = \frac{1}{5}\sum_{i=1}^5 Y_i$. Let $Y_6$ be another independent observation from the same population. Let $W = \sum_{i=1}^5 Y_i^2$, $U = \sum_{i=1}^5 (Y_i - \overline{Y})^2$. What is the distribution of

(a) $\frac{\sqrt{5}Y_6}{\sqrt{W}}$

(b) $\frac{2Y_6}{\sqrt{U}}$

(c) $\frac{2\left(5\overline{Y}^2 + Y_6^2\right)}{U}$

*Solution* Question done in Practice Problem 7.38 in Chapter 2. □

### 6.4 Tutorial 4

**Question 1**

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with the probability density function

$$f(x) = \begin{cases} e^{-(x-\mu)} & \text{if } x \geq \mu \\ 0 & \text{otherwise} \end{cases}$$

Let $\hat{\mu} = \min(X_1, X_2, \ldots, X_n)$. Calculate the bias of $\hat{\mu}$.

*Solution* The bias of an estimator is defined as $\text{Bias}(\hat{\mu}) = E[\hat{\mu}] - \mu$. To calculate $E[\hat{\mu}]$, note that $\hat{\mu} = \min(X_1, X_2, \ldots, X_n)$. The cumulative distribution function (CDF) of $\hat{\mu}$ is given by:

$$F_{\hat{\mu}}(x) = P(\hat{\mu} \leq x) = 1 - P(\hat{\mu} > x) = 1 - P(X_1 > x, X_2 > x, \ldots, X_n > x).$$

Since the $X_i$ are independent, we have:

$$P(X_1 > x, X_2 > x, \ldots, X_n > x) = P(X_1 > x) \cdot P(X_2 > x) \cdot \ldots \cdot P(X_n > x).$$

The survival function of $X_i$ is $P(X_i > x) = 1 - F_X(x)$, where $F_X(x)$ is the CDF of $X_i$. For the given PDF $f(x)$, the CDF is:

$$F_X(x) = \begin{cases} 1 - e^{-(x-\mu)} & \text{if } x \geq \mu \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $P(X_i > x) = e^{-(x-\mu)}$ for $x \geq \mu$. Therefore:

$$P(X_1 > x, X_2 > x, \ldots, X_n > x) = \left(e^{-(x-\mu)}\right)^n = e^{-n(x-\mu)}.$$

So the CDF of $\hat{\mu}$ is:

$$F_{\hat{\mu}}(x) = \begin{cases} 1 - e^{-n(x-\mu)} & \text{if } x \geq \mu \\ 0 & \text{otherwise.} \end{cases}$$

The PDF of $\hat{\mu}$ is the derivative of the CDF:

$$f_{\hat{\mu}}(x) = \begin{cases} ne^{-n(x-\mu)} & \text{if } x \geq \mu \\ 0 & \text{otherwise.} \end{cases}$$

To find $E[\hat{\mu}]$, compute the expectation:

$$E[\hat{\mu}] = \int_\mu^\infty x f_{\hat{\mu}}(x)\, dx = \int_\mu^\infty x n e^{-n(x-\mu)}\, dx.$$

Let $y = x - \mu$, so $x = y + \mu$ and $dx = dy$:

$$E[\hat{\mu}] = \int_0^\infty (y+\mu) n e^{-ny}\, dy = \int_0^\infty y n e^{-ny}\, dy + \int_0^\infty \mu n e^{-ny}\, dy.$$

The first term is:

$$\int_0^\infty y n e^{-ny}\, dy = \frac{1}{n}.$$

The second term is:

$$\int_0^\infty \mu n e^{-ny}\,dy = \mu \int_0^\infty n e^{-ny}\,dy = \mu.$$

Thus:

$$E[\hat{\mu}] = \frac{1}{n} + \mu.$$

The bias is:

$$\text{Bias}(\hat{\mu}) = E[\hat{\mu}] - \mu = \frac{1}{n}.$$

Therefore, the bias of $\hat{\mu}$ is $\frac{1}{n}$. □

**Question 2**

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with the probability density function

$$f(x) = \begin{cases} e^{-(x-\mu)} & \text{if } x \geq \mu \\ 0 & \text{otherwise} \end{cases}$$

Calculate the mean square error of $\hat{\mu}$.

*Solution* The mean square error (MSE) of an estimator is defined as:

$$\text{MSE}(\hat{\mu}) = \text{Var}(\hat{\mu}) + \text{Bias}^2(\hat{\mu}).$$

From the previous solution, the bias of $\hat{\mu}$ is $\frac{1}{n}$. To calculate $\text{Var}(\hat{\mu})$, we use the formula:

$$\text{Var}(\hat{\mu}) = E[\hat{\mu}^2] - (E[\hat{\mu}])^2.$$

First, compute $E[\hat{\mu}^2]$:

$$E[\hat{\mu}^2] = \int_\mu^\infty x^2 f_{\hat{\mu}}(x)\,dx = \int_\mu^\infty x^2 n e^{-n(x-\mu)}\,dx.$$

Let $y = x - \mu$, so $x = y + \mu$ and $dx = dy$:

$$E[\hat{\mu}^2] = \int_0^\infty (y+\mu)^2 n e^{-ny}\,dy = \int_0^\infty (y^2 + 2y\mu + \mu^2) n e^{-ny}\,dy.$$

Split the integral into three parts:

$$E[\hat{\mu}^2] = \int_0^\infty y^2 n e^{-ny}\,dy + 2\mu \int_0^\infty y n e^{-ny}\,dy + \mu^2 \int_0^\infty n e^{-ny}\,dy.$$

The first term is:

$$\int_0^\infty y^2 n e^{-ny}\,dy = \frac{2}{n^2}.$$

The second term is:

$$2\mu \int_0^\infty y n e^{-ny}\,dy = 2\mu \cdot \frac{1}{n} = \frac{2\mu}{n}.$$

The third term is:

$$\mu^2 \int_0^\infty n e^{-ny}\,dy = \mu^2.$$

Thus:

$$E[\hat{\mu}^2] = \frac{2}{n^2} + \frac{2\mu}{n} + \mu^2.$$

Now compute $\text{Var}(\hat{\mu})$:

$$\text{Var}(\hat{\mu}) = E[\hat{\mu}^2] - (E[\hat{\mu}])^2 = \left(\frac{2}{n^2} + \frac{2\mu}{n} + \mu^2\right) - \left(\frac{1}{n} + \mu\right)^2.$$

Expand $\left(\frac{1}{n} + \mu\right)^2$:

$$\left(\frac{1}{n} + \mu\right)^2 = \frac{1}{n^2} + \frac{2\mu}{n} + \mu^2.$$

So:

$$\text{Var}(\hat{\mu}) = \frac{2}{n^2} + \frac{2\mu}{n} + \mu^2 - \left(\frac{1}{n^2} + \frac{2\mu}{n} + \mu^2\right) = \frac{1}{n^2}.$$

Finally, compute the MSE:

$$\text{MSE}(\hat{\mu}) = \text{Var}(\hat{\mu}) + \text{Bias}^2(\hat{\mu}) = \frac{1}{n^2} + \left(\frac{1}{n}\right)^2 = \frac{2}{n^2}.$$

Thus, the mean square error of $\hat{\mu}$ is $\frac{2}{n^2}$. □

**Question 3**

Let $Y_1, Y_2, \dots, Y_n$ denote a random sample from a Uniform$(\theta, \theta + 1)$ distribution. Consider the following estimators:

$$\hat{\theta}_1 = \overline{Y} - \frac{1}{2} \qquad \hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}$$

Which estimators are unbiased?

*Solution* To determine whether the estimators are unbiased, we calculate the expected value of each estimator and compare it to $\theta$.

For $\hat{\theta}_1 = \overline{Y} - \frac{1}{2}$:

$$E[\hat{\theta}_1] = E\left[\overline{Y} - \frac{1}{2}\right] = E[\overline{Y}] - \frac{1}{2}.$$

The mean of $Y_i \sim \text{Uniform}(\theta, \theta + 1)$ is $\theta + \frac{1}{2}$. Since $\overline{Y}$ is the sample mean:

$$E[\overline{Y}] = \theta + \frac{1}{2}.$$

Thus:

$$E[\hat{\theta}_1] = \theta + \frac{1}{2} - \frac{1}{2} = \theta.$$

Therefore, $\hat{\theta}_1$ is an unbiased estimator.

For $\hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}$, where $Y_{(n)}$ is the maximum of the sample: The expected value of $Y_{(n)}$ for $Y_i \sim \text{Uniform}(\theta, \theta + 1)$ is:

$$E[Y_{(n)}] = \theta + \frac{n}{n+1}.$$

Thus:

$$E[\hat{\theta}_2] = E\left[Y_{(n)} - \frac{n}{n+1}\right] = E[Y_{(n)}] - \frac{n}{n+1} = \theta + \frac{n}{n+1} - \frac{n}{n+1} = \theta.$$

Therefore, $\hat{\theta}_2$ is also an unbiased estimator.

In conclusion, both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of $\theta$. □

## 6.5 Tutorial 5

**Question 1**

Let $\overline{X}$ be the mean of a random sample of size $n$ from a distribution that is $N(\mu, 9)$. Find $n$ such that

$$P(\overline{X} - 1 < \mu < \bar{X} + 1) = 0.90,$$

approximately.

*Solution* We are given that $\overline{X} \sim N(\mu, \frac{9}{n})$. The interval $\overline{X} - 1 < \mu < \overline{X} + 1$ corresponds to a probability of 0.90. This means the width of the interval is determined by the standard normal distribution.

Standardizing the interval:

$$P\left(-\frac{1}{\sqrt{9/n}} < Z < \frac{1}{\sqrt{9/n}}\right) = 0.90,$$

where $Z \sim N(0, 1)$. Using the symmetry of the standard normal distribution:

$$P\left(Z < \frac{1}{\sqrt{9/n}}\right) = 0.95.$$

From standard normal tables, $\Phi^{-1}(0.95) \approx 1.645$. Thus:

$$\frac{1}{\sqrt{9/n}} = 1.645.$$

Solve for $n$:

$$\sqrt{9/n} = \frac{1}{1.645}, \quad n = \frac{9 \cdot 1.645^2}{1}.$$

Compute $n$:

$$n \approx 24.4.$$

Since $n$ must be an integer, we round up to $n = 25$.

Thus, the required sample size is $n = 25$. □

## 6.6 Tutorial 6

**Question 1**

Solid copper produced by sintering a powder under specified environmental conditions is then measured for porosity in a laboratory. A sample of $n_1 = 4$ independent porosity measurements have mean $\bar{y}_1 = 0.22$ and variance $s_1^2 = 0.001$. A second laboratory repeats the same process on solid copper formed from an identical powder and gets $n_2 = 5$ independent porosity measurements with $\bar{y}_2 = 0.17$ and $s_2^2 = 0.002$. Estimate the true difference between the population mean $(\mu_1 - \mu_2)$ for these two laboratories, with confidence coefficient 0.95.

*Solution* Point estimate:

$$\hat{\mu}_1 - \hat{\mu}_2 = \bar{y}_1 - \bar{y}_2 = 0.22 - 0.17 = 0.05.$$

Use Welch two–sample $t$ interval (unequal variances):

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.001}{4} + \frac{0.002}{5}} = \sqrt{0.00025 + 0.0004} = \sqrt{0.00065} \approx 0.02550.$$

Welch d.f.:

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} = \frac{0.00065^2}{\frac{0.00025^2}{3} + \frac{0.0004^2}{4}} \approx 6.95 \approx 7.$$

Critical value $t_{0.975,7} \approx 2.365$. Margin:

$$ME = 2.365(0.02550) \approx 0.0603.$$

95% CI:

$$(0.05 - 0.0603,\ 0.05 + 0.0603) = (-0.0103,\ 0.1103).$$

Thus a 95% confidence interval for $\mu_1 - \mu_2$ is approximately $(-0.01, 0.11)$. □

**Question 2**

Suppose that $S^2$ is the sample variance based on a sample of size $n$ from a normal population with unknown mean and variance.

1. Derive a $100(1-\alpha)\%$ upper confidence limit for $\sigma^2$.
2. Derive a $100(1-\alpha)\%$ lower confidence limit for $\sigma^2$.

*Solution* For (1): We know that for a sample of size $n$ from a normal population with unknown mean and variance, the sample variance $S^2$ follows the distribution:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

For a $100(1-\alpha)\%$ upper confidence limit, we want to find a value $U$ such that:

$$P(\sigma^2 < U) = 1 - \alpha$$

Using the chi-square distribution property:

$$P\left(\frac{(n-1)S^2}{\sigma^2} > \chi^2_{\alpha,n-1}\right) = \alpha$$

Rearranging the inequality:

$$P\left(\sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha,n-1}}\right) = 1 - \alpha$$

Therefore, the $100(1-\alpha)\%$ upper confidence limit for $\sigma^2$ is:

$$U = \frac{(n-1)S^2}{\chi^2_{\alpha,n-1}}$$

For (2): For a $100(1-\alpha)\%$ lower confidence limit, we want to find a value $L$ such that:

$$P(\sigma^2 > L) = 1 - \alpha$$

Using the chi-square distribution property:

$$P\left(\frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha,n-1}\right) = 1 - \alpha$$

Rearranging the inequality:

$$P\left(\sigma^2 > \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}\right) = 1 - \alpha$$

Therefore, the $100(1-\alpha)\%$ lower confidence limit for $\sigma^2$ is:

$$L = \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}$$

$\square$

**Question 3**

Let $Y$ have probability density function

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

1. Show that $Y$ has a distribution function

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

2. Show that $\frac{Y}{\theta}$ is a pivotal quantity.

3. Use the pivotal quantity from part (b) to find a 90% lower confidence limit for $\theta$.

*Solution*

For (a): To find the cumulative distribution function (CDF), we integrate the PDF:

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} f_Y(t)\,dt$$

For $y \leq 0$: Since $f_Y(t) = 0$ for $t \leq 0$, we have $F_Y(y) = 0$. For $0 < y < \theta$:

$$F_Y(y) = \int_0^y \frac{2(\theta-t)}{\theta^2}\,dt = \frac{2}{\theta^2}\int_0^y (\theta - t)\,dt$$

$$= \frac{2}{\theta^2}\left[\theta t - \frac{t^2}{2}\right]_0^y = \frac{2}{\theta^2}\left(\theta y - \frac{y^2}{2}\right)$$

$$= \frac{2\theta y}{\theta^2} - \frac{2y^2}{2\theta^2} = \frac{2y}{\theta} - \frac{y^2}{\theta^2}$$

For $y \geq \theta$: Since the PDF is zero for $t > \theta$, we have:

$$F_Y(y) = \int_0^\theta \frac{2(\theta - t)}{\theta^2}\, dt = 1$$

Therefore, the CDF is:

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

For (b): Let $U = \frac{Y}{\theta}$. To show this is a pivotal quantity, we need to show that the distribution of $U$ does not depend on $\theta$. Using the transformation method, if $Y$ has CDF $F_Y(y)$, then $U = \frac{Y}{\theta}$ has CDF:

$$F_U(u) = P(U \leq u) = P\left(\frac{Y}{\theta} \leq u\right) = P(Y \leq u\theta) = F_Y(u\theta)$$

For $0 < u < 1$ (since $0 < y < \theta$ implies $0 < \frac{y}{\theta} < 1$):

$$F_U(u) = F_Y(u\theta) = \frac{2(u\theta)}{\theta} - \frac{(u\theta)^2}{\theta^2} = 2u - u^2$$

The PDF of $U$ is:

$$f_U(u) = \frac{d}{du}F_U(u) = \frac{d}{du}(2u - u^2) = 2 - 2u$$

Therefore:

$$f_U(u) = \begin{cases} 2(1-u) & 0 < u < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Since this distribution does not depend on $\theta$, $U = \frac{Y}{\theta}$ is indeed a pivotal quantity. For (c): For a 90% lower confidence limit, we need to find $u_{0.1}$ such that $P(U \leq u_{0.1}) = 0.1$ Using the CDF $F_U(u) = 2u - u^2$:

$$0.1 = 2u_{0.1} - u_{0.1}^2$$

$$u_{0.1}^2 - 2u_{0.1} + 0.1 = 0$$

Using the quadratic formula:

$$u_{0.1} = \frac{2 \pm \sqrt{4 - 0.4}}{2} = \frac{2 \pm \sqrt{3.6}}{2} = 1 \pm \frac{\sqrt{3.6}}{2}$$

Since $0 < u < 1$, we take the smaller root:

$$u_{0.1} = 1 - \frac{\sqrt{3.6}}{2} \approx 1 - 0.949 = 0.051$$

From the pivotal quantity relationship:

$$P\left(\frac{Y}{\theta} \geq u_{0.1}\right) = 0.9$$

$$P\left(Y \geq u_{0.1}\theta\right) = 0.9$$

$$P\left(\theta \leq \frac{Y}{u_{0.1}}\right) = 0.9$$

Therefore, the 90% lower confidence limit for $\theta$ is:

$$L = \frac{Y}{u_{0.1}} = \frac{Y}{1 - \frac{\sqrt{3.6}}{2}} \approx \frac{Y}{0.051}$$

□

## 6.7 Tutorial 7

**Question 1**

Recall that in a previous tutorial, we showed that if $Y_1, Y_2, \ldots, Y_n$ denote a random sample from Uniform$(\theta, \theta + 1)$, then

$$\hat{\theta}_1 = \bar{Y} - \frac{1}{2} \quad \text{and} \quad \hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}$$

are unbiased estimators for $\theta$. Determine which estimator is better by calculating the relative efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

*Solution* The relative efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is defined as:

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

We need to find the variances of both estimators.
**Finding Var($\hat{\theta}_1$):**
Since $\hat{\theta}_1 = \bar{Y} - \frac{1}{2}$, we have:

$$\text{Var}(\hat{\theta}_1) = \text{Var}\left(\bar{Y} - \frac{1}{2}\right) = \text{Var}(\bar{Y}) = \frac{\text{Var}(Y_i)}{n}$$

For $Y_i \sim$ Uniform$(\theta, \theta + 1)$, the variance is:

$$\text{Var}(Y_i) = \frac{(\theta + 1 - \theta)^2}{12} = \frac{1}{12}$$

Therefore:

$$\text{Var}(\hat{\theta}_1) = \frac{1}{12n}$$

**Finding Var($\hat{\theta}_2$):**
Since $\hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}$, we have:

$$\text{Var}(\hat{\theta}_2) = \text{Var}\left(Y_{(n)} - \frac{n}{n+1}\right) = \text{Var}(Y_{(n)})$$

For the maximum order statistic $Y_{(n)}$ from Uniform$(\theta, \theta + 1)$: The PDF of $Y_{(n)}$ is: $f_{Y_{(n)}}(y) = n(y - \theta)^{n-1}$ for $\theta < y < \theta + 1$. $E[Y_{(n)}] = \theta + \frac{n}{n+1}$. $E[Y_{(n)}^2] = \int_\theta^{\theta+1} y^2 \cdot n(y - \theta)^{n-1} dy$. Let $u = y - \theta$, so $y = u + \theta$ and $dy = du$:

$$E[Y_{(n)}^2] = \int_0^1 (u + \theta)^2 \cdot n u^{n-1} du$$

$$= \int_0^1 (u^2 + 2\theta u + \theta^2) \cdot nu^{n-1} du$$

$$= n \int_0^1 u^{n+1} du + 2n\theta \int_0^1 u^n du + n\theta^2 \int_0^1 u^{n-1} du$$

$$= n \cdot \frac{1}{n+2} + 2n\theta \cdot \frac{1}{n+1} + n\theta^2 \cdot \frac{1}{n}$$

$$= \frac{n}{n+2} + \frac{2n\theta}{n+1} + \theta^2$$

Therefore:

$$\text{Var}(Y_{(n)}) = E[Y_{(n)}^2] - (E[Y_{(n)}])^2$$

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

Expanding the squared term:

$$\left(\theta + \frac{n}{n+1}\right)^2 = \theta^2 + \frac{2n\theta}{n+1} + \frac{n^2}{(n+1)^2}$$

Thus:

$$\text{Var}(Y_{(n)}) = \frac{n}{n+2} - \frac{n^2}{(n+1)^2}$$

$$= \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}$$

$$= \frac{n[(n+1)^2 - n(n+2)]}{(n+2)(n+1)^2}$$

$$= \frac{n[n^2 + 2n + 1 - n^2 - 2n]}{(n+2)(n+1)^2}$$

$$= \frac{n}{(n+2)(n+1)^2}$$

Therefore:

$$\text{Var}(\hat{\theta}_2) = \frac{n}{(n+2)(n+1)^2}$$

Thus we have that

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} = \frac{\frac{n}{(n+2)(n+1)^2}}{\frac{1}{12n}}$$

$$= \frac{n}{(n+2)(n+1)^2} \cdot 12n = \frac{12n^2}{(n+2)(n+1)^2}$$

For large $n$, this approaches:

$$\lim_{n\to\infty} \frac{12n^2}{(n+2)(n+1)^2} = \lim_{n\to\infty} \frac{12n^2}{n^4} = 0$$

Since $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) < 1$ for all finite $n$, we conclude that $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$. The estimator $\hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}$ is the better estimator. $\square$

**Question 2**

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

$$f(y;\theta) = \begin{cases} \frac{2}{\theta^2}(\theta - y) & \text{if } 0 \le y \le \theta \\ 0 & \text{otherwise} \end{cases}$$

Find an estimator for $\theta$ by using the method of moments.

*Solution* We begin by equating the first sample moment to the first population moment.

Compute $E[Y]$:

$$E[Y] = \int_0^\theta y \frac{2(\theta-y)}{\theta^2}\,dy = \frac{2}{\theta^2}\int_0^\theta (\theta y - y^2)\,dy = \frac{2}{\theta^2}\left[\frac{\theta y^2}{2} - \frac{y^3}{3}\right]_0^\theta = \frac{2}{\theta^2}\left(\frac{\theta^3}{2} - \frac{\theta^3}{3}\right) = \frac{2}{\theta^2}\cdot\frac{\theta^3}{6} = \frac{\theta}{3}.$$

Set the sample mean $\overline{Y}$ equal to $E[Y]$:

$$\overline{Y} = \frac{\theta}{3} \quad \text{or} \quad \hat{\theta}_{\text{MM}} = 3\overline{Y}.$$

(Optional properties.) Since $E[\overline{Y}] = \theta/3$, we have $E[\hat{\theta}_{\text{MM}}] = 3(\theta/3) = \theta$ (unbiased). Also,

$$E[Y^2] = \int_0^\theta y^2 \frac{2(\theta-y)}{\theta^2}dy = \frac{2}{\theta^2}\int_0^\theta (\theta y^2 - y^3)dy = \frac{2}{\theta^2}\left[\frac{\theta y^3}{3} - \frac{y^4}{4}\right]_0^\theta = \frac{2}{\theta^2}\left(\frac{\theta^4}{3} - \frac{\theta^4}{4}\right) = \frac{2\theta^2}{12} = \frac{\theta^2}{6}.$$

Thus

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2 = \frac{\theta^2}{6} - \frac{\theta^2}{9} = \frac{\theta^2}{18}, \qquad \text{Var}(\hat{\theta}_{\text{MM}}) = 9\cdot\frac{\text{Var}(Y)}{n} = 9\cdot\frac{\theta^2}{18n} = \frac{\theta^2}{2n}.$$

□

**Question 3**

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

$$f(y;\theta) = \begin{cases} \dfrac{\Gamma(2\theta)}{[\Gamma(\theta)]^2}\, y^{\theta-1}(1-y)^{\theta-1} & \text{if } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find an estimator for $\theta$ by using the method of moments.

*Solution* Done in Practice Problem 9.75 in Chapter 4. □

## 6.8 Tutorial 8

Tutorial 8 Question 1 is skipped for Brevity - please take a look at posted solutions.

**Question 2**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with common probability mass function

$$p_Y(y \mid \theta) = \theta^y (1-\theta)^{1-y}, \quad y = 0 \text{ or } 1,$$

where $0 \leq \theta \leq 1$ is a parameter.

(a) Find an estimator $\hat{\theta}_{\text{MOM}}$ of $\theta$ using the method of moments.

(b) Derive the Fisher information $I_n(\theta)$ of this distribution.

(c) Prove that $\hat{\theta}_{\text{MOM}}$ is MVUE of $\theta$ using the Cramer–Rao theorem.

*Solution* For (a): Each $Y_i \sim \text{Bernoulli}(\theta)$ so $E[Y_i] = \theta$. Equate the first sample moment to the first population moment:

$$\overline{Y} = \theta \quad \text{, that is,} \quad \hat{\theta}_{\text{MOM}} = \overline{Y}.$$

For (b): The log–likelihood:

$$\ell(\theta) = \sum_{i=1}^{n} \big[Y_i \log\theta + (1 - Y_i)\log(1-\theta)\big].$$

Score:

$$\ell'(\theta) = \sum_{i=1}^{n} \left(\frac{Y_i}{\theta} - \frac{1-Y_i}{1-\theta}\right).$$

Second derivative:

$$\ell''(\theta) = -\sum_{i=1}^{n} \left(\frac{Y_i}{\theta^2} + \frac{1-Y_i}{(1-\theta)^2}\right).$$

Fisher information:

$$I_n(\theta) = -E[\ell''(\theta)] = n\left(\frac{E[Y_1]}{\theta^2} + \frac{1-E[Y_1]}{(1-\theta)^2}\right) = n\left(\frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2}\right) = n\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) = \frac{n}{\theta(1-\theta)}.$$

For (c): The estimator $\hat{\theta}_{\text{MOM}} = \overline{Y}$ is unbiased:

$$E[\overline{Y}] = \theta.$$

Its variance:

$$\text{Var}(\overline{Y}) = \frac{\theta(1-\theta)}{n}.$$

Cramér–Rao lower bound (CRLB) for any unbiased estimator of $\theta$:

$$\frac{1}{I_n(\theta)} = \frac{\theta(1-\theta)}{n}.$$

Thus $\text{Var}(\overline{Y})$ attains the CRLB, so $\overline{Y}$ is efficient. Since $T = \sum_{i=1}^n Y_i$ is (i) sufficient (factorization) and (ii) complete for the Bernoulli family, and $\overline{Y}$ is a function of $T$ and unbiased, it is the unique MVUE (Lehmann–Scheffé). Therefore

$$\hat{\theta}_{\text{MVUE}} = \overline{Y}.$$

□

### 6.9 Tutorial 9

**Question 1**

Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Poisson($\lambda$) distribution.

(a) Find the MLE $\hat{\lambda}$ for $\lambda$.

(b) Find the expected value and variance for $\hat{\lambda}$.

(c) Show that the estimator of part (a) is consistent for $\lambda$.

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

*Solution* For (a): We begin by finding the likelihood function:

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{Y_i}}{Y_i!} = e^{-n\lambda}\lambda^{\sum Y_i} \prod_{i=1}^{n} \frac{1}{Y_i!}.$$

Then the Log–likelihood:

$$\ell(\lambda) = -n\lambda + \left(\textstyle\sum_{i=1}^{n} Y_i\right) \log \lambda + C.$$

Differentiate and set to 0:

$$\ell'(\lambda) = -n + \frac{\sum Y_i}{\lambda} = 0 \implies \hat{\lambda} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

For (b): Since $Y_i \sim$ Poisson($\lambda$) and are independent,

$$E[\hat{\lambda}] = E[\bar{Y}] = \lambda, \qquad \mathrm{Var}(\hat{\lambda}) = \frac{\mathrm{Var}(Y_i)}{n} = \frac{\lambda}{n}.$$

For (c): $\hat{\lambda}$ is unbiased and $\mathrm{Var}(\hat{\lambda}) = \lambda/n \to 0$ as $n \to \infty$, hence $\hat{\lambda} \xrightarrow{P} \lambda$ (consistent). Also by the LLN, $\bar{Y} \to \lambda$ a.s. For (d): We see that $P(Y = 0) = e^{-\lambda}$. Plug in $\hat{\lambda}$ using invariance:

$$\widehat{P(Y = 0)} = e^{-\hat{\lambda}} = e^{-\bar{Y}}.$$

□

**Question 2**

Let $Y_1, \ldots, Y_n$ be a random sample from a Bernoulli($p$) distribution. Find the MVUE of $(1 - p)^2$.

*Solution* We have already shown that $T = \sum_{i=1}^{n} Y_i$ is a sufficient (and complete) statistic; $T \sim \mathrm{Bin}(n, p)$. We seek $h(T)$ with $E[h(T)] = (1 - p)^2 = 1 - 2p + p^2$.

First,

$$E\left[1 - \frac{2T}{n}\right] = 1 - 2p.$$

For $T \sim \mathrm{Bin}(n, p)$,

$$E(T^2) = \mathrm{Var}(T) + [E(T)]^2 = np(1 - p) + n^2p^2,$$

so

$$E(T^2 - T) = \big(np(1-p) + n^2p^2\big) - np = -np^2 + n^2p^2 = p^2n(n-1).$$

Hence

$$E\left[\frac{T^2 - T}{n(n-1)}\right] = p^2.$$

Define

$$h(T) = 1 - \frac{2T}{n} + \frac{T^2 - T}{n(n-1)}.$$

Then

$$E[h(T)] = 1 - 2p + p^2 = (1-p)^2.$$

Thus $h(T)$ is unbiased for $(1-p)^2$ and, being a function of the complete sufficient statistic $T$, it is the MVUE:

$$h(T) = 1 - \frac{2T}{n} + \frac{T^2 - T}{n(n-1)} = \frac{(n-T)(n-T-1)}{n(n-1)}.$$

$\square$

**Question 3**

Let $Y_1, \ldots, Y_n$ be a random sample from a Gamma$(2, \beta)$ distribution. Find the MVUE of $\beta(\beta + 2)$.

*Solution* Let $Y_i \sim \Gamma(2, \beta)$ (shape 2, scale $\beta$), so the pdf is

$$f(y; \beta) = \frac{1}{\Gamma(2)\beta^2} y^{2-1} e^{-y/\beta}, \qquad y > 0.$$

The joint pdf of the sample depends on the data only through

$$S = \sum_{i=1}^{n} Y_i \sim \Gamma(2n, \beta),$$

so $S$ is complete and sufficient for $\beta$ (one–parameter full exponential family).

Moments of $S$:

$$E[S] = 2n\beta, \qquad \mathrm{Var}(S) = 2n\beta^2, \qquad E[S^2] = \mathrm{Var}(S) + [E(S)]^2 = 2n\beta^2 + 4n^2\beta^2 = 2n(2n+1)\beta^2.$$

Hence

$$\beta = \frac{E[S]}{2n}, \qquad \beta^2 = \frac{E[S^2]}{2n(2n+1)}.$$

An unbiased estimator of $\beta^2 + 2\beta$ is therefore

$$\hat{g}(S) = \frac{S^2}{2n(2n+1)} + \frac{S}{n},$$

since

$$E[\hat{g}(S)] = \frac{E[S^2]}{2n(2n+1)} + \frac{E[S]}{n} = \beta^2 + 2\beta = \beta(\beta + 2).$$

As a function of the complete sufficient statistic $S$, $\hat{g}(S)$ is the MVUE. In terms of $\bar{Y} = S/n$:

$$\hat{g} = \bar{Y} + \frac{n\bar{Y}^2}{2(2n+1)}.$$

Thus

$$\text{MVUE of } \beta(\beta+2): \ \hat{g}(S) = \frac{S^2}{2n(2n+1)} + \frac{S}{n} = \bar{Y} + \frac{n\bar{Y}^2}{2(2n+1)}.$$

□

**Question 3**

Let $Y_1, \ldots, Y_n$ be a random sample from a $N(\mu, 1)$ distribution. Show that the MVUE of $\mu^2$ is $\hat{\mu}^2 = \bar{Y}^2 - \frac{1}{n}$.

*Solution* Done in Practice Problem 9.64 in Chapter 4. □

## 6.10 Tutorial 10

**Question 1**

Let $X_1, \ldots, X_n$ be a random sample from Poisson($\lambda$). Find the UMVUE for $\lambda$.

*Solution* Done in Example 4.8.3. □

**Question 2**

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \frac{3y^2}{\theta^3}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $Y_{(n)} = \max(Y_1, \ldots, Y_n)$ is complete.

*Solution* CDF of a single $Y_i$:

$$F(y) = P(Y_i \le y) = \int_0^y \frac{3t^2}{\theta^3} dt = \left(\frac{y}{\theta}\right)^3, \qquad 0 \le y \le \theta.$$

Hence the pdf of the maximum $Y_{(n)}$ is

$$f_{Y_{(n)}}(y) = n[F(y)]^{n-1} f(y) = n\left(\frac{y}{\theta}\right)^{3(n-1)} \frac{3y^2}{\theta^3} = \frac{3n\, y^{3n-1}}{\theta^{3n}}, \qquad 0 < y < \theta.$$

To show completeness: Suppose $g$ satisfies $E_\theta[g(Y_{(n)})] = 0$ for all $\theta > 0$. Then

$$0 = \int_0^\theta g(y) \frac{3n\, y^{3n-1}}{\theta^{3n}}\, dy = \frac{3n}{\theta^{3n}} \int_0^\theta g(y) y^{3n-1} dy, \qquad \forall \theta > 0.$$

Thus

$$\int_0^\theta g(y) y^{3n-1} dy = 0 \quad \forall \theta > 0.$$

Define $H(\theta) = \int_0^\theta g(y)y^{3n-1}dy$. Then $H(\theta) \equiv 0$, so differentiating (where differentiation is justified by standard conditions),

$$H'(\theta) = g(\theta)\theta^{3n-1} = 0 \quad \forall\theta > 0 \text{ or } g(\theta) = 0 \text{ a.e.}$$

Hence $g(Y_{(n)}) = 0$ almost surely, and $Y_{(n)}$ is complete. (Equivalently, with the change of variable $y = \theta t$, $E[g(Y_{(n)})] = 3n\int_0^1 g(\theta t)t^{3n-1}dt = 0$ for all $\theta$, leading to the same conclusion.) Therefore $Y_{(n)}$ is complete. □

**Question 3**

Let $Y$ be a single observation, which comes from a distribution with density

$$f(y;\theta) = 1 - \theta^2(y - 0.5), \quad 0 < y < 1$$

where $-1 < \theta < 1$ is the parameter.

(a) Determine a rejection region of the most powerful test of size $\alpha$ for testing $H_0 : \theta = 0$ vs $H_A : \theta = \frac{1}{2}$.

(b) Determine the power of the test in terms of some fixed $\alpha$.

*Solution* For (a) Under $H_0 : \theta = 0$, $f_0(y) = 1$ for $0 < y < 1$ (Uniform$(0,1)$). Under $H_A : \theta = 1/2$,

$$f_1(y) = 1 - \left(\tfrac{1}{2}\right)^2(y - 0.5) = 1 - \tfrac{1}{4}y + \tfrac{1}{8} = \frac{9}{8} - \frac{y}{4}, \quad 0 < y < 1,$$

a decreasing function of $y$. The likelihood ratio

$$\Lambda(y) = \frac{f_1(y)}{f_0(y)} = \frac{9}{8} - \frac{y}{4}$$

is strictly decreasing in $y$, so by the Neyman–Pearson lemma, the most powerful size-$\alpha$ test rejects for small $y$: $R = \{y : y < c_\alpha\}$. Choose $c_\alpha$ so that $P_{H_0}(Y < c_\alpha) = \alpha$. Since $Y \sim$ Uniform$(0,1)$ under $H_0$, $c_\alpha = \alpha$. Thus the rejection region is

$$\boxed{R = \{y : 0 < y < \alpha\}}.$$

For (b) The power at $\theta = 1/2$ is

$$\text{power}(\tfrac{1}{2}) = P_{1/2}(Y < \alpha) = \int_0^\alpha \left(\frac{9}{8} - \frac{y}{4}\right)dy = \frac{9}{8}\alpha - \frac{1}{8}\alpha^2 = \frac{9\alpha - \alpha^2}{8}, \qquad 0 < \alpha < 1.$$

So

$$\text{Power} = \frac{9\alpha - \alpha^2}{8}.$$

□

### 6.11 Tutorial 11

**Question 1**

Let $Y_1, \dots, Y_8$ be a random sample from the probability density function given by:

$$f(x \mid \beta) = \begin{cases} \frac{3}{\beta} e^{-y^3/\beta}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

What is the Rejection Region (RR) for the Uniformly Most Powerful (UMP) test of:

$$H_0 : \beta = 2 \quad \text{vs.} \quad H_a : \beta > 2$$

with significance level $\alpha = 0.05$ ?
**Hint:**

$$\sum_{i=1}^{8} Y_i^3 \sim \text{Gamma}(8, \beta)$$

*Solution* We test $H_0 : \beta = 2$ vs $H_a : \beta > 2$. For $W \sim \text{Gamma}(8, \beta)$ (shape 8, scale $\beta$), $E_{H_0}[W] = 16$ and larger $\beta$ shifts $W$ stochastically upward, so the UMP (Neyman–Pearson / monotone likelihood ratio) test rejects for large $W$:

$$\text{Reject } H_0 \text{ if } W \geq c_\alpha,$$

where $c_\alpha$ satisfies

$$P_{H_0}(W \geq c_\alpha) = \alpha, \qquad W \mid H_0 \sim \text{Gamma}(8, 2).$$

Noting $\text{Gamma}(8, 2) \equiv \chi^2_{16}$, choose $c_\alpha$ as the $(1 - \alpha)$ quantile of $\chi^2_{16}$:

$$c_\alpha = \chi^2_{16,\, 1-\alpha}.$$

For $\alpha = 0.05$,

$$c_{0.05} = \chi^2_{16, 0.95} \approx 26.30.$$

Thus the rejection region is

$$\sum_{i=1}^{8} Y_i^3 \geq 26.30.$$

□

**Question 2**

Suppose that $Y_1, \dots, Y_n$ denote a random sample from a population having an exponential distribution with mean $\theta$. Derive the most powerful test for $H_0 : \theta = \theta_0$ against $H_a : \theta = \theta_a$, where $\theta_a < \theta_0$.

*Solution* The likelihood for $\theta$ is $L(\theta) = \theta^{-n} \exp\big( - S/\theta \big)$, where $S = \sum_{i=1}^{n} Y_i$. The likelihood ratio is

$$\frac{L(\theta_a)}{L(\theta_0)} = \left( \frac{\theta_0}{\theta_a} \right)^n \exp\Big\{ - \Big( \frac{1}{\theta_a} - \frac{1}{\theta_0} \Big) S \Big\},$$

which is strictly decreasing in $S$ because $\theta_a < \theta_0$. By the Neyman–Pearson lemma, the MP size-$\alpha$ test rejects for small $S$:

$$\text{Reject } H_0 \ \text{ if } \ S \le c_\alpha,$$

where $c_\alpha$ satisfies $P_{\theta_0}(S \le c_\alpha) = \alpha$. Since $S \mid \theta_0 \sim \text{Gamma}(n, \theta_0)$ and $2S/\theta_0 \sim \chi^2_{2n}$,

$$c_\alpha = \frac{\theta_0}{2}\,\chi^2_{2n,\alpha}.$$

Power at a general $\theta$ is

$$\beta(\theta) = P_\theta(S \le c_\alpha) = F_{\chi^2_{2n}}\Big(\frac{2c_\alpha}{\theta}\Big) = F_{\chi^2_{2n}}\Big(\frac{\theta_0}{\theta}\,\chi^2_{2n,\alpha}\Big),$$

and at $\theta_a$: $\beta(\theta_a) = F_{\chi^2_{2n}}\big((\theta_0/\theta_a)\chi^2_{2n,\alpha}\big)$. □

**Question 3**

A company wants to know if its product is used equally or not in different cities. To see this, they conduct a random sample in each of the three cities and ask if their product is used or not. They collect the following data:

| | City 1 | City 2 | City 3 |
|---|---|---|---|
| Sample size: | 100 | 100 | 200 |
| Number who use product: | 14 | 27 | 47 |

Use the likelihood ratio at level $\alpha = 0.05$ to determine if the null hypothesis that the same percentage of the population in each city uses the product.

*Solution* Skipped for brevity - please see samples solutions posted. □

## 6.12 Tutorial 12

**Question 1**

Let $Y_1, Y_2, \dots, Y_n$ denote a random sample from a distribution with the following probability density function with parameters $\alpha > 0$ and $\beta > 0$, where $\beta$ is known.

$$f(y|\alpha,\beta) = \begin{cases} \alpha\beta^\alpha y^{-(\alpha+1)} & y \ge \beta, \\ 0 & \text{otherwise.} \end{cases}$$

Find the MLE for $\alpha$.

*Solution* This is a Pareto$(\alpha, \beta)$ model with known scale $\beta$. For data with $Y_i \ge \beta$,

$$L(\alpha) = \prod_{i=1}^n \alpha\beta^\alpha Y_i^{-(\alpha+1)} = \alpha^n \beta^{n\alpha} \prod_{i=1}^n Y_i^{-(\alpha+1)}.$$

Log–likelihood:

$$\ell(\alpha) = n\log\alpha + n\alpha\log\beta - (\alpha+1)\sum_{i=1}^n \log Y_i.$$

Differentiate and set to zero:

$$\ell'(\alpha) = \frac{n}{\alpha} + n\log\beta - \sum_{i=1}^{n}\log Y_i = 0 \text{ or } \hat{\alpha} = \frac{n}{\sum_{i=1}^{n}\log(Y_i/\beta)}.$$

Since $\ell''(\alpha) = -n/\alpha^2 < 0$, this is the unique maximizer. Thus,

$$\hat{\alpha}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n}\log\left(\frac{Y_i}{\beta}\right)}, \qquad \text{provided } Y_{(1)} \geq \beta.$$

□

## Question 2

Let $T_1, T_2, \ldots, T_n$ denote a random sample from a distribution with probability density function:

$$f(t \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log(t)-\mu)^2}{2\sigma^2}\right)\frac{1}{t}$$

where $t > 0$ and $\mu$ is known. Find the MLE for $\sigma^2$.

*Solution* Let $Z_i = \log T_i$. Then $Z_i \sim N(\mu, \sigma^2)$ with $\mu$ known. The log–likelihood (up to constants) is

$$\ell(\sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Z_i-\mu)^2 + C,$$

where $Z_i = \log t_i$. Differentiate w.r.t. $\sigma^2$ and set to zero:

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(Z_i-\mu)^2 = 0 \text{ or } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\log T_i - \mu)^2.$$

Since the second derivative at this point is negative, this is the MLE.

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}\left(\log T_i - \mu\right)^2.$$

□

## Question 3

A survey was conducted in four districts of a city to compare the proportion of households that recycle their waste regularly. Random samples of 150 households were taken in each of the four districts, and the results are shown in the accompanying table. The numbers of households recycling regularly in the four districts can be regarded as four independent binomial random variables.
Construct a likelihood ratio test for the hypothesis that the proportions of households recycling regularly are the same across all four districts. Use $\alpha = 0.05$.

| Opinion | District 1 | District 2 | District 3 | District 4 | Total |
|---|---|---|---|---|---|
| Recycle Regularly | 45 | 60 | 50 | 55 | 210 |
| Do Not Recycle | 105 | 90 | 100 | 95 | 390 |
| Total | 150 | 150 | 150 | 150 | 600 |

*Solution* We test $H_0 : p_1 = p_2 = p_3 = p_4 = p$ vs $H_a$ : not all equal. Let $x_i$ be the number recycling in district $i$ and $n_i = 150$. Under $H_a$, $\hat{p}_i = x_i/n_i$; under $H_0$, the pooled $\hat{p} = \frac{\sum x_i}{\sum n_i} = \frac{210}{600} = 0.35$.

The LR statistic (for a $2 \times 4$ table) is

$$G^2 = 2\sum_{i=1}^{4}\left[x_i \log\frac{x_i}{n_i\hat{p}} + (n_i - x_i)\log\frac{n_i - x_i}{n_i(1-\hat{p})}\right].$$

Under $H_0$, $G^2 \overset{approx}{\sim} \chi^2_{4-1} = \chi^2_3$.

Expected counts under $H_0$: successes $E_i = n_i\hat{p} = 52.5$, failures $n_i(1-\hat{p}) = 97.5$ (same for each district). Using the table:

$$G^2 \approx 3.65 \quad (\text{Pearson } X^2 = \sum (O-E)^2/E \approx 3.66).$$

With $df = 3$, the critical value at $\alpha = 0.05$ is $\chi^2_{3,0.95} = 7.815$ and

$$p\text{-value} \approx 0.30.$$

Since $G^2 < 7.815$ (and $p > 0.05$), so we fail to reject $H_0$. $\square$

If you made it at the end these notes you are completed with course - congrats. I believe that STA256 and STA260 are fundamental learning experiences if you want to explore the field of Data Analysis, Machine Learning, Quantatative finance and more. However notes alone are not enough to solidify any concept, so I encourage you to actually do something with the topics you have learnt. In fact I found myself coming back to these notes outside of school and only then do you realize the importance of statistical inference.