Chapter 6:

18. The plastic tubes produced by company X for the irrigation system used in golf courses have a mean length of 1.5 meters and a standard deviation of 0.1 meter. The plastic tubes produced by company Y have a mean length of 1 meter and a standard deviation of 0.09 meter. Suppose that both tube lengths follow normal distributions.

(a) Calculate the probability that a random sample of 15 tubes from company X has a mean length at least 0.45 meter greater than the mean length of a random sample of size 20 from company Y.

(b) Suppose that the population variances are unknown but equal, $S_x = 0.1$, and $S_y = 0.09$. Calculate the probability that a random sample of 15 plastic tubes from company X has a mean length at least 0.45 meter greater than the mean length of a random sample of 20 plastic tubes from company Y.

Solution:

(a) Note: $\overline{X} - \overline{Y} \sim N\left((1.5-1), \sqrt{\frac{0.1^2}{15} + \frac{0.09^2}{20}}\right)$

$\mathbb{P}(\overline{X} - \overline{Y} > 0.45) = 0.9367$

```
> 1 - pnorm(0.45, 1.5 - 1, sqrt(0.1^2/15 + 0.09^2/20))

[1] 0.9366637
```

(b) When $\sigma_X^2 = \sigma_y^2$, a pooled standard deviation is appropriate.

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x+n_y-2}$$

$\mathbb{P}(\overline{X} - \overline{Y} > 0.45) = \mathbb{P}(t_{33} > -1.55115) = 0.9348$

```
> tobs <- (0.45 - 0.5)/(sqrt((14*.1^2 +19*.09^2)/33)*sqrt(1/15 + 1/20))
> tobs

[1] -1.55115

> 1 - pt(tobs, 33)

[1] 0.9347978
```

Chapter 7

4. Let X be a $Bin(n, \pi)$ random variable.

(a) Find the mean squared error of the π parameter estimators $T_1 = X/n$ and $T_2 = (X+1)/(n+2)$.

(b) When $n = 100$ and $\pi = 0.4$, which estimator, T_1 or T_2, has the smaller MSE?

(c) Plot the efficiency of T_2 relative to T_1 versus π values in $(0, 1)$ for n values from 1 to 10.

Solution:

Since $X \sim Bin(n, \pi)$, it follows that $E[X] = n\pi$ and $Var[X] = n\pi(1-\pi)$.

(a) Given $T_1 = \frac{X}{n}$,

$$E[T_1] = E\left[\frac{X}{n}\right] = \frac{E[X]}{n} = \frac{n\pi}{n} = \pi$$

and

$$Var[T_1] = Var\left[\frac{X}{n}\right] = \frac{Var[X]}{n^2} = \frac{n\pi(1-\pi)}{n^2} = \frac{\pi(1-\pi)}{n}.$$

The $MSE[T_1]$ is $Var[T_1] + (Bias[T_1])^2$.

$$\begin{aligned} MSE[T_1] &= Var[T_1] + (Bias[T_1])^2 \\ &= \frac{\pi(1-\pi)}{n} + (E[T_1] - \pi)^2 \\ &= \frac{\pi(1-\pi)}{n} + (\pi - \pi)^2 \\ &= \frac{\pi(1-\pi)}{n} \end{aligned}$$

For $T_2 = \frac{X+1}{n+2}$,

$$E[T_2] = E\left[\frac{X+1}{n+2}\right] = \frac{n\pi + 1}{n+2},$$

$$Var[T_2] = Var\left[\frac{X+1}{n+2}\right] = \frac{n\pi(1-\pi)}{(n+2)^2}, \text{ and}$$

$$Bias[T_2] = (E[T_2] - \pi) = \frac{n\pi+1}{n+2} - \pi = \frac{(1-2\pi)}{n+2}.$$

It follows that

$$\begin{aligned} MSE[T_2] &= Var[T_2] + (Bias[T_2])^2 \\ &= \frac{n\pi(1-\pi)}{(n+2)^2} + \frac{(1-2\pi)^2}{(n+2)^2} \\ &= \frac{n\pi(1-\pi) + (1-2\pi)^2}{(n+2)^2}. \end{aligned}$$

Thus, $MSE[T_1] = \frac{\pi(1-\pi)}{n}$ and $MSE[T_2] = \frac{n\pi(1-\pi)+(1-2\pi)^2}{(n+2)^2}$.

(b) For $n = 100$ and $\pi = 0.4$,

$$MSE(T_1) = \frac{\pi(1-\pi)}{n} = \frac{(0.4)(0.6)}{100} = 0.0024, \text{ and}$$

$$MSE[T_2] = \frac{n\pi(1-\pi)+(1-2\pi)^2}{(n+2)^2} = \frac{100(0.4)(0.6)+[1-2(0.4)]^2}{(100+2)^2} = 0.0023,$$

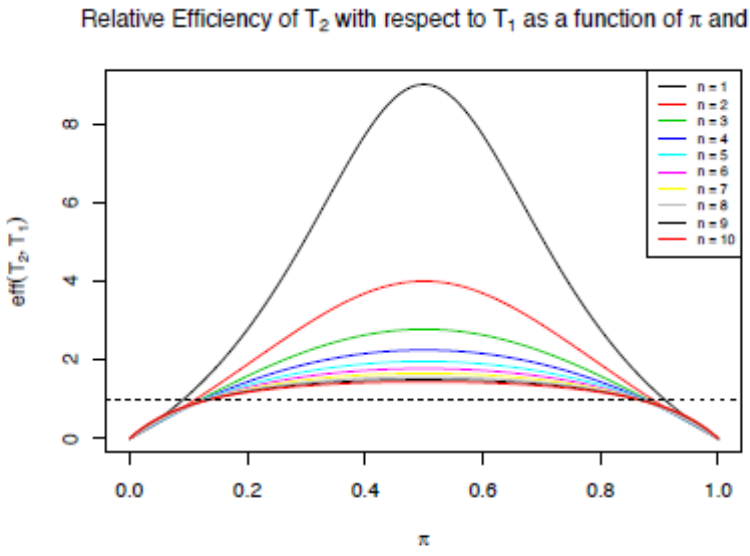
so T_2 has the smaller MSE.

(c) The efficiency of T_2 relative to T_1 is

$$eff(T_2, T_1) = \frac{MSE[T_1]}{MSE[T_2]} = \frac{\frac{\pi(1-\pi)}{n}}{\frac{n\pi(1-\pi)+(1-2\pi)^2}{(n+2)^2}}.$$

The following code is used to graph the efficiency of T_2 relative to T_1 for values of π in $(0, 1)$ and n values from 1 to 10.

```
> p <- seq(from = 0, to = 1, by = 0.01)
> rel.ef <- function(p, n){
+   (p*(1 - p)/n) / ((n*p*(1 - p)+(1 - 2*p)^2)/(n + 2)^2)
+ }
> plot(p, rel.ef(p, 1), type = "n", xlab = expression(pi),
+ ylab = expression(eff(T[2], T[1])), ylim = c(0, 9))
> for(i in 1:10){
+   lines(p, rel.ef(p, i), col = i)
+ }
> abline(h = 1, lty = 2)
> names <- numeric(10)
> for(i in 1:10){
+   names[i] <- paste("n =", i)
+ }
> legend("topright", legend = names, lty = 1, col = c(1:10), cex = 0.7)
> title(expression(paste(plain("Relative Efficiency of "),T[2],
+ plain(" with respect to "),T[1],plain(" as a function of ")*pi,
+ plain(" and " )*n)))
```



Based on the graph, the relative efficiencies of T_2 relative to T_1 are very similar for π values close to 0 and 1. For values of π between 0.2 and 0.8, the graph illustrates the superiority of the T_2 estimator, which for $n = 1$ and $\pi = 0.5$ is almost 10 times more efficient than T_1.

9. Verify that $Var\left[\frac{\partial \ln f(X|\theta)}{\partial\theta}\right] = E\left[\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)^2\right]$. (Hint: show that $E\left[\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)\right] = 0$.)

Solution:

$$Var\left[\frac{\partial \ln f(X|\theta)}{\partial\theta}\right] = E\left[\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)^2\right] - E\left[\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)\right]^2$$

So, it is sufficient to show that $E\left[\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)\right] = 0$.

$$\begin{aligned} E\left[\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)\right] &= E\left[\left(\frac{\frac{\partial}{\partial\theta} f(X|\theta)}{f(X|\theta)}\right)\right] \\ &= \int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial\theta} f(X|\theta)}{f(X|\theta)} \cdot f(X|\theta)dx \\ &= \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(X|\theta)dx \\ &= \frac{\partial}{\partial\theta}(1) = 0 \quad \blacksquare \end{aligned}$$

So, $Var\left[\frac{\partial \ln f(X|\theta)}{\partial\theta}\right] = E\left[\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)^2\right]$.

Extra 1.If X has a binomial distribution with $n = 3$ and $\pi = 1/3$, find the probability distributions of

(a) $Y = \frac{X}{1+X}$;

(b) $Y = (X - 1)^4$.

Solution:

$$f(0) = 1 \cdot \frac{8}{27} = \frac{8}{27},\quad f(1) = 3 \cdot \frac{1}{3} \cdot \frac{4}{9} = \frac{12}{27},\quad f(2) = 3 \cdot \frac{1}{9} \cdot \frac{2}{3} = \frac{6}{27},\quad f(3) = 1 \cdot \frac{1}{27} = \frac{1}{27}$$

(a)

$x = 0$	1	2	3
$y = 0$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{4}$

y	0	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{4}$
$g(y)$	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$

(b)

$x = 0$	1	2	3
$y = 1$	0	1	16

y	0	1	16
$g(y)$	$\frac{12}{27}$	$\frac{14}{27}$	$\frac{1}{27}$

Extra 2.If $X = \ln Y$ has a normal distribution with the mean μ and the standard deviation σ, find the probability density of Y which is said to have the log-normal distribution.

Solution:

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$. Substituting $y = e^x$, with $x = \ln y$ and $\frac{dx}{dy} = y^{-1}$, we obtain

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma} \cdot y^{-1} e^{(\ln y - \mu)^2/2\sigma^2} \text{ for } y > 0, \text{ and } g(y) = 0 \text{ elsewhere.}$$

Extra 3.Let X_1 and X_2 be two continuous random variables having the joint probability density

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 & 0 < x_1 < 1, \quad 0 < x_2 < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the joint probability density of $Y_1 = {X_1}^2$ and $Y_2 = X_1X_2$.

Solution:

$f(x_1, x_2) = 4x_1x_2$

$x_1 = \sqrt{y}$

$x_2 = y_2 / \sqrt{y_1}$

$y_1 = x_1^2$ $\quad \frac{\partial x_1}{\partial y_1} = \frac{1}{2\sqrt{y_1}}$ $\quad \frac{\partial x_2}{\partial y_1} = -\frac{1}{2} y_2 y_1^{-3/2}$

$y_2 = x_1x_2$ $\quad \frac{\partial x_1}{\partial y_2} = 0$ $\quad \frac{\partial x_2}{\partial y_2} = \frac{1}{\sqrt{y_1}}$

$$g(y_1, y_2) = 4\sqrt{y_1} \frac{y_2}{\sqrt{y_1}} \cdot \frac{1}{2y_1} = \frac{2y_2}{y_1}$$

$$J = \begin{vmatrix} \frac{1}{2\sqrt{y_1}} & 0 \\ -\frac{1}{2} y_2 y_1^{-3/2} & \frac{1}{\sqrt{y_1}} \end{vmatrix} = \frac{1}{2y_1}$$

over region bounded by $y_1 = 0$, $y_1 = 1$, $y_2 = 0$, and $y_1 = {y_2}^2$.