Homework 7

Darnell Chen

2024-11-04

Problem 7.2.6

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{7}$

a) $z = \frac{X-\mu}{\sigma_{\bar{X}}} = \frac{10-8.2}{\frac{1.5}{7}} = 8.4$

$P(Z < 8.4) \approx 1$ since the z-score is so high.

Thus, the probability that the average time waiting in line is less than 10 minutes is approximately 1.

b) $z = \frac{X-\mu}{\sigma_{\bar{X}}} = \frac{5-8.2}{\frac{1.5}{7}} \approx -14.9333$

$P(-14.933 < X < 8.4) = P(X < 8.4) - P(X > -14.933) \approx 1 - 0 = 1$

Thus, the probability that the average time waiting in line between 5 and 10 minutes is approximately 1.

c) $z = \frac{X-\mu}{\sigma_{\bar{X}}} = \frac{6-8.2}{\frac{1.5}{7}} \approx -10.267$

$P(Z < -10.267) \approx 4.979314 \times 10^{-25} \approx 0$

Thus, the probability that the average waiting time in line is less than 6 minutes is approximately 0.

```r
pnorm(10, mean=8.2, sd=(1.5/7))
```

```
## [1] 1
```

```r
pnorm(10, mean=8.2, sd=(1.5/7)) - pnorm(5, mean=8.2, sd=(1.5/7))
```

```
## [1] 1
```

```r
pnorm(6, mean=8.2, sd=(1.5/7))
```

```
## [1] 4.979314e-25
```

Problem 7.3.7

a) $\hat{\theta} = \frac{425+431+416+419+421+436+418+410+431+433+423+426+410+435+436+428+411+426+409+437+422+428+413+416}{24} \approx$ 423.33

b) $S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=0}^{24}(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{(425-423.33)^2+(431-423.33)^2+\ldots+(413-423.33)^2+(416-423.33)^2}{23}} \approx \sqrt{82.49275} \approx$ 9.082552

c) $\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n}} = \sqrt{\frac{S^2}{n}} \approx \sqrt{\frac{82.49275}{24}} \approx 1.853968$

d) $\frac{423+425}{2} = 424$

e) $\frac{7}{24}$ Notably, the following thicknesses are larger than 430: 431 431 433 435 436 436 437

```
data <- c(425, 431, 416, 419, 421, 436, 418, 410, 431, 433, 423, 426, 410, 435, 436, 428, 411, 426, 409
mean(data)
```

```
## [1] 423.3333
```

```
sqrt(var(data))
```

```
## [1] 9.082552
```

```
sqrt(var(data)/length(data))
```

```
## [1] 1.853968
```

```
sort(data)
```

```
##  [1] 409 410 410 411 413 416 416 418 419 421 422 423 425 426 426 428 428 431 431
## [20] 433 435 436 436 437
```

```
median(data)
```

```
## [1] 424
```

Problem 7.3.9

a)

$$
\begin{aligned}
E[\bar{X}_1 - \bar{X}_2] &= E[\bar{X}_1] - E[\bar{X}_2] \\
&= E[\frac{X_{1,1} + X_{1,2} + ... + X_{1,n_1}}{n_1}] - E[\frac{X_{2,1} + X_{2,2} + ... + X_{2,n_2}}{n_2}] \\
&= \frac{1}{n_1} E[X_{1,1} + X_{1,2} + ... + X_{1,n_1}] - \frac{1}{n_2} E[X_{2,1} + X_{2,2} + ... + X_{2,n_1}] \\
&= \frac{1}{n_1}(\mu_1 + \mu_1 + ... + \mu_1) - \frac{1}{n_2}(\mu_2 + \mu_2 + ... + \mu_2) \\
&= \frac{1}{n_1} n_1 \mu_1 - \frac{1}{n_2} n_2 \mu_2 \\
&= \mu_1 - \mu_2
\end{aligned}
$$

b)

$$
\begin{aligned}
se(X_1 - X_2) &= \sqrt{V(\bar{X}_1 - \bar{X}_2)} \\
&= \sqrt{V(\bar{X}_1) + V(\bar{X}_2)} \\
&= \sqrt{V(\frac{X_{1,1} + X_{1,2} + ... + X_{1,n_1}}{n_1}) + V(\frac{X_{2,1} + X_{2,2} + ... + X_{2,n_2}}{n_2})} \\
&= \sqrt{\frac{1}{n_1^2} V(X_{1,1} + X_{1,2} + ... + X_{1,n_1}) + \frac{1}{n_2^2} V(X_{2,1} + X_{2,2} + ... + X_{2,n_1})} \\
&= \sqrt{\frac{1}{n_1^2}(\sigma_1^2 + \sigma_1^2 + ... + \sigma_1^2) + \frac{1}{n_2^2}(\sigma_2^2 + \sigma_2^2 + ... + \sigma_2^2)} \\
&= \sqrt{\frac{1}{n_1^2} n_1 \sigma_1^2 + \frac{1}{n_2^2} n_2 \sigma_2^2} \\
&= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}
\end{aligned}
$$

c) For an unbiased estimator $E[s_P^2] = \sigma^2$ has to be true:

$$
\begin{aligned}
E[S_p^2] &= E[\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}] \\
&= E[\frac{(n_1 - 1)S_1^2}{n_1 + n_2 - 2}] + E[\frac{(n_2 - 1)S_2^2}{n_1 + n_2 - 2}] \\
&= \frac{n_1 - 1}{n_1 + n_2 - 2} E[S_1^2] + \frac{n_2 - 1}{n_1 + n_2 - 2} E[S_2^2] \\
&= \frac{n_1 - 1}{n_1 + n_2 - 2} \sigma^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} \sigma^2 \\
&= \frac{n_1 + n_2 - 2}{n_1 + n_2 - 2} \sigma^2 \\
&= \sigma^2
\end{aligned}
$$

Since $E[S_p^2] = \sigma^2$, it is thus an unbiased estimator of σ^2.

Problem 7.3.11

a) We know that X_1 and X_2 represent the sample of students who own an Apple Computer at ASU and VT respectively, while n_1 and n_2 are the respective sample sizes. Thus, we know that the probability of choosing a student who owns an Apple computer at the two schools are $\frac{X_1}{n_1}$ and $\frac{X_2}{n_2}$ respectively. To show this is unbiased estimator for $p_1 - p_2$ we can take its expected value:

$$\begin{aligned}
E[\frac{X_1}{n_1} - \frac{X_2}{n_2}] &= E[\frac{X_1}{n_1}] - E[\frac{X_2}{n_2}] \\
&= \frac{1}{n_1}E[X_1] - \frac{1}{n_2}E[X_2] \\
&= \frac{1}{n_1}n_1p_1 - \frac{1}{n_2}n_2p_2 \qquad \text{,This comes from mean in binomial distribution} \\
&= p_1 - p_2
\end{aligned}$$

Since $E[\frac{X_1}{n_1} - \frac{X_2}{n_2}] = p_1 - p_2$, it is therefore an unbiased estimator.

b)

$$\begin{aligned}
\sigma &= \sqrt{V(\frac{X_1}{n_1} - \frac{X_2}{n_2})} \\
&= \sqrt{V(\frac{X_1}{n_1}) + V(\frac{X_2}{n_2})} \\
&= \sqrt{\frac{1}{n_1^2}V(X_1) + \frac{1}{n_2^2}V(X_2)} \\
&= \sqrt{\frac{1}{n_1^2}n_1p_1(1-p_1) + \frac{1}{n_2^2}n_2p_2(1-p_2)} \qquad \text{, np(1-p) is variance of binom. distribution} \\
&= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}
\end{aligned}$$

c) Since we know that $p_1 = \frac{X_1}{n_1}$ and $p_2 = \frac{X_2}{n_2}$ from part (a), we can thus plug them into the equation we got from part (b):

$$\begin{aligned}
\sqrt{\frac{\frac{X_1}{n_1}(1-\frac{X_1}{n_1})}{n_1} + \frac{\frac{X_2}{n_2}(1-\frac{X_2}{n_2})}{n_2}} &= \sqrt{\frac{\frac{X_1}{n_1} - \frac{X_1^2}{n_1^2}}{n_1} + \frac{\frac{X_2}{n_2} - \frac{X_2^2}{n_2^2}}{n_2}} \\
&= \sqrt{X_1 - X_1^2 + X_2 - X_2^2}
\end{aligned}$$

Thus we get: $\sqrt{X_1 - X_1^2 + X_2 - X_2^2}$, where we can simply plug in our values X_1 and X_2.

d) $\hat{\theta} = \frac{X_1}{n_1} - \frac{X_2}{n_2} = \frac{150}{200} - \frac{185}{250} = 0.01$.

Our point estimate for $p_1 - p_2$ is thus 0.01.

e)

$$\begin{aligned}
\hat{\sigma} &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\
&= \sqrt{\frac{0.75(1-0.75)}{200} + \frac{0.74(1-0.74)}{250}} \\
&\approx \sqrt{0.0009375 + 0.0007696} \\
&\approx 0.041317
\end{aligned}$$

Problem 7.4.1

We know that the function for geometric distribution is: $f(x) = P(X = x) = p(1-p)^{x-1}$ for $x \in \mathbb{Z}^+$. Its likelihood function is then the product of all possible x up to n:\

$L(p) = \Pi_{i=1}^{n}(1-p)^{x_i-1}p = (1-p)^{\sum_{i=1}^{n}(x_i-1)}p^{\sum_{i=1}^{n}1} = (1-p)^{\sum_{i=1}^{n}(x_i-1)}p^n$\ $l(p) = ln(L(p)) = ln((1-p)^{\sum_{i=1}^{n}(x_i-1)}p^n) = \sum_{i=1}^{n}(x_i-1)ln(1-p) + n \cdot ln(p)$

Now that we have the likelihood function $l(p)$, we want to take its derivative to find where it maximizes at:

$$\begin{aligned}
\frac{\partial l(p)}{\partial p} &= [\sum_{i=1}^{n}(x_i-1)\frac{\partial}{\partial p}] \ \cdot \ ln(1-p) + n\frac{\partial}{\partial p}ln(p) \\
&= [\sum_{i=1}^{n}(x_i-1)] \ \cdot \ (\frac{-1}{1-p}) + n\frac{1}{p} \\
&= \frac{n}{p} - \frac{\sum_{i=1}^{n}(x_i-1)}{1-p}
\end{aligned}$$

We want to set the derivative to 0 to find where the MLE function maximizes p:

$$\begin{aligned}
\frac{n}{p} - \frac{\sum_{i=1}^{n}(x_i-1)}{1-p} &= 0 \\
\frac{n}{p} &= \frac{\sum_{i=1}^{n}(x_i-1)}{1-p} \\
\frac{(1-p)(n)}{p} &= \sum_{i=1}^{n}(x_i-1) \\
(1-p)(n) &= p\sum_{i=1}^{n}(x_i-1) \\
n - np &= p\sum_{i=1}^{n}(x_i-1) \\
n &= np + p\sum_{i=1}^{n}(x_i-1) \\
n &= p[n + \sum_{i=1}^{n}(x_i-1)] \\
\frac{n}{p} &= n + \sum_{i=1}^{n}(x_i-1) \\
\frac{p}{n} &= \frac{1}{n + \sum_{i=1}^{n}(x_i-1)} \\
p &= \frac{n}{n + \sum_{i=1}^{n}(x_i-1)}
\end{aligned}$$

Thus, we get our mean likelihood estimator for p:

$\hat{p}_{MLE} = \frac{n}{n+\sum_{i=1}^{n}(X_i-1)}$.

Problem 2

A) to check that either estimator is unbiased, we want to see if $E[\hat{\mu_1}] = \mu$ or $E[\hat{\mu_2}] = \mu$. We'll first check the former:

$$\begin{aligned}
E[\hat{\mu_1}] &= E[\frac{\bar{X}_n + \bar{Y}_m}{2}] \\
&= \frac{1}{2}E[\bar{X}_n + \bar{Y}_m] \\
&= \frac{1}{2}E[\bar{X}_n] + \frac{1}{2}E[\bar{Y}_m] \\
&= \frac{1}{2}E[\frac{\sum_{i=0}^{n} X_i}{n}] + \frac{1}{2}E[\frac{\sum_{j=0}^{m} Y_j}{m}] \\
&= \frac{1}{2n}E[\sum_{i=0}^{n} X_i] + \frac{1}{2m}E[\sum_{j=0}^{m} Y_j] \\
&= \frac{1}{2n}n\mu + \frac{1}{2m}m\mu \\
&= \frac{\mu}{2} + \frac{\mu}{2} \\
&= \mu
\end{aligned}$$

Since $E[\hat{\mu}_1] = \mu$ it is an unbiased estimator.\ Next we will check if $E[\hat{\mu_2}]$ is unbiased:\

$$\begin{aligned}
E[\hat{\mu_2}] &= E[\frac{n \cdot \bar{X}_n + m \cdot \bar{Y}_m}{n+m}] \\
&= \frac{1}{n+m}E[n \cdot \bar{X}_n + m \cdot \bar{Y}_m] \\
&= \frac{1}{n+m}E[n \cdot \bar{X}_n] + \frac{1}{n+m}E[m \cdot \bar{Y}_m] \\
&= \frac{n}{n+m}E[\frac{\sum_{i=0}^{n} X_i}{n}] + \frac{m}{n+m}E[\frac{\sum_{j=0}^{m} Y_j}{m}] \\
&= \frac{1}{n+m}E[\sum_{i=0}^{n} X_i] + \frac{1}{n+m}E[\sum_{j=0}^{m} Y_j] \\
&= \frac{1}{n+m}n\mu + \frac{1}{n+m}m\mu \\
&= \frac{\mu(n+m)}{n+m} \\
&= \mu
\end{aligned}$$

Since $E[\hat{\mu_2}] = \mu$, our second estimator is also unbiased. Thus $\hat{\mu}_1$ and $\hat{\mu}_2$ are both unbiased.

B) Let's begin with the variance of $\hat{\mu_1}$:

$$
\begin{aligned}
V[\hat{\mu_1}] &= V[\frac{\bar{X}_n + \bar{Y}_m}{2}] \\
&= \frac{1}{4} V[\bar{X}_n + \bar{Y}_m] \\
&= \frac{1}{4} V[\bar{X}_n] + \frac{1}{4} V[\bar{Y}_m] \\
&= \frac{1}{4} V[\frac{\sum_{i=0}^{n} X_i}{n}] + \frac{1}{4} V[\frac{\sum_{j=0}^{m} Y_j}{m}] \\
&= \frac{1}{4n^2} V[\sum_{i=0}^{n} X_i] + \frac{1}{4m^2} V[\sum_{j=0}^{m} Y_j] \\
&= \frac{1}{4n^2} n\sigma^2 + \frac{1}{4m^2} m\sigma^2 \\
&= \frac{1}{4n}\sigma^2 + \frac{1}{4m}\sigma^2 \\
&= \frac{\sigma^2}{4}(\frac{1}{n} + \frac{1}{m})
\end{aligned}
$$

Then the variance of $\hat{\mu_2}$:

$$
\begin{aligned}
V[\hat{\mu_2}] &= V[\frac{n \cdot \bar{X}_n + m \cdot \bar{Y}_m}{(n+m)}] \\
&= \frac{1}{(n+m)^2} V[n \cdot \bar{X}_n + m \cdot \bar{Y}_m] \\
&= \frac{1}{(n+m)^2} V[n \cdot \bar{X}_n] + \frac{1}{(n+m)^2} V[m \cdot \bar{Y}_m] \\
&= \frac{n^2}{(n+m)^2} V[\bar{X}_n] + \frac{m^2}{(n+m)^2} V[\bar{Y}_j] \\
&= \frac{n^2}{(n+m)^2} \frac{\sigma^2}{n} + \frac{m^2}{(n+m)^2} \frac{\sigma^2}{m} \\
&= \frac{\sigma^2}{n+m}
\end{aligned}
$$

C)

$$
\begin{aligned}
\frac{MSE(\hat{\mu_1})}{MSE(\hat{\mu_2})} &= \frac{V(\hat{\mu_1})}{V(\hat{\mu_2})} \\
&= \frac{\frac{\sigma^2}{4}(\frac{1}{n} + \frac{1}{m})}{\frac{\sigma^2}{n+m}} \\
&= \frac{\sigma^2}{4}(\frac{1}{n} + \frac{1}{m})(\frac{n+m}{\sigma^2}) \\
&= \frac{1}{4}(\frac{1}{n} + \frac{1}{m})(\frac{n+m}{1}) \\
&= \frac{n+m}{4} \cdot (\frac{1}{n} + \frac{1}{m}) \\
&= \frac{n+m}{4} \cdot (\frac{n+m}{nm}) \\
&= \frac{(n+m)^2}{4nm}
\end{aligned}
$$

d) Using the formula for part 3, we know that if $\frac{(n+m)^2}{4nm} < 1$ then $\hat{\mu_1}$ is more efficient. On the other hand, if $\frac{(n+m)^2}{4nm} > 1$, then $\hat{\mu_2}$ is more efficient.