HW4

Halid Kopanski

2/10/2022

Problem 1

1.a)

$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ $\qquad Y_1, \dots, Y_n \overset{iid}{\sim} Poisson(\lambda)$

$E(Y_i) = \lambda$

$Var(Y_i) = \lambda$

$$\bar{Y}_n = E(Y_n) = \frac{1}{n} E\left(\sum_{i=1}^{n} Y_i\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(Y_i)$$

$$= \frac{1}{n}\, n \quad \lambda$$

$$= \lambda$$

$$Var(Y_n) = Var\left(\frac{1}{n} \sum_{i=1}^{n} Y_i\right)$$

$$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} Y_i\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var(Y_i)$$

$$= \cancel{n} \frac{1}{n\cancel{2}} \sigma^2$$

$$= \frac{\lambda}{n}$$

λ is the mean of the population while $\bar{Y}_n$ is the sample mean. λ is fixed and unkown while the sample is mean is a random variable. In this case $\bar{Y}_n - \lambda = 0$ Therefore this is an unbiased statistic.

b) $\bar{Y}_n \xrightarrow{P} \lambda$ as $n \to \infty$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_i$$

$$\lim_{n \to \infty} P(|X_n - x| \geq \varepsilon) = 0 \text{ for all } \varepsilon > 0$$

$$P(|\bar{Y}_n - \lambda| \geq \varepsilon) \leq \frac{Var(\bar{Y}_n)}{\varepsilon^2} = \frac{\lambda}{n\varepsilon^2}$$

$$\lim_{n \to \infty} \frac{\lambda}{n\varepsilon^2} = 0$$

Problem 2

#2

$$F_{Z^2}(z) = P(Z^2 \le z) = P(-\sqrt{z} \le Z \le \sqrt{z})$$

$$= P(-\sqrt{z} \le \min(x,y) \le \sqrt{z})$$

$$= \tfrac{1}{2}\left(P(-\sqrt{z} \le X \le \sqrt{z}\right) + \tfrac{1}{2}\left(P(-\sqrt{z} \le Y \le \sqrt{z}\right) \quad ①$$

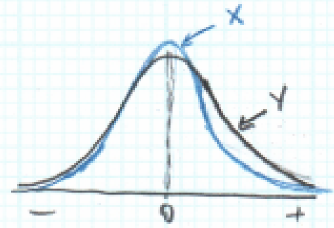
$$= \frac{1}{2} F_{X^2}(z) + \frac{1}{2} F_{Y^2}(z)$$

$$= F_{X^2}(z) = F_{Y^2}(z) \Rightarrow$$ square of standard random normal variable is a χ^2_1 when $m = 1$

① Law of total probability

$$\begin{cases} P(-\sqrt{z} \le X \le \sqrt{z} \mid X < Y) \\ P(-\sqrt{z} \le Y \le \sqrt{z} \mid Y < X) \end{cases}$$

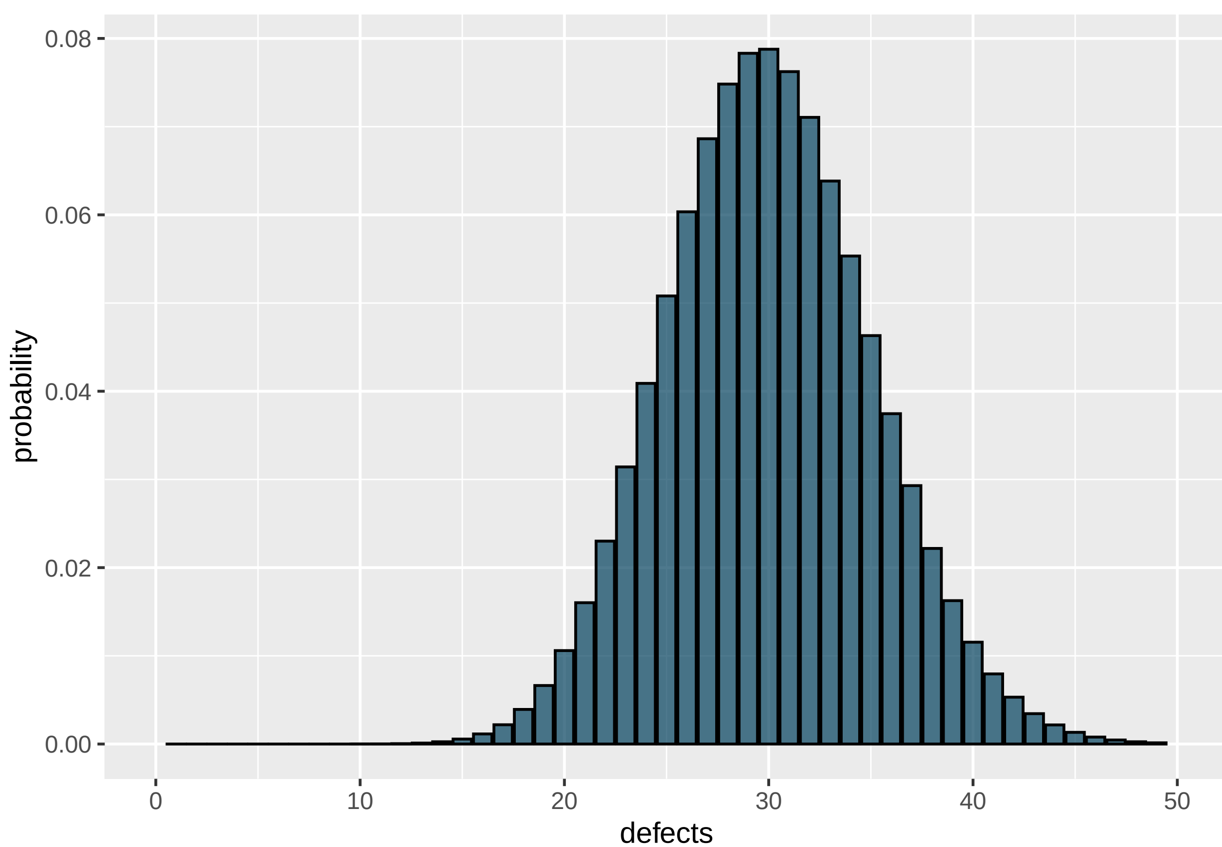
$P(X<Y) = P(Y<X)$
$= 1/2$?



Problem 3

Since p = 0.15, we can approximate the distribution to be normal with a mean of n*p and a variance of n*p*(1 - p). The results are comparable.

```
library(tidyverse)

n <- 200
p <- 0.15
probs <- c()

for (i in 0:n){
  probs <- append(probs, choose(n, i) * p^i * (1 - p)^(n - i))
}

df_factory <- data.frame("defects" = 0:n, "probability" = probs)

ggplot(data = df_factory, aes(x = defects, y = probability)) +
    geom_bar(stat = "identity", color = "black", fill = "#003f5c", alpha = 0.7) +
    xlim(0, 50)
```




```
sprintf("Binomial calculation: %0.5f", 1 - sum(df_factory[0:41, "probability"]))
```

```
## [1] "Binomial calculation: 0.02200"
```

```r
sprintf("Normal Approximation: %0.5f", 1 - pnorm(40, mean = n*p, sd = (n*p*(1-p))^0.5))
```

```
## [1] "Normal Approximation: 0.02384"
```

```r
sprintf("Binomial calculation (built in): %0.5f", 1 - pbinom(40, n, p))
```

```
## [1] "Binomial calculation (built in): 0.02200"
```

Problem 4

Problem 4

$X_1 \ldots\ldots X_n \sim \Gamma(n, \lambda)$

Since multiple random samples from an exponential distribution have a joint distribution of Gamma where $\alpha = n$ and $\lambda = \lambda$.

$P(X_n \geq 2)$

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \qquad x \geq 0$$

$\alpha = n$

$$f(x) = \frac{\lambda^{n}}{\Gamma(n)} x^{n-1} e^{-\lambda x}$$

$$P(x \geq 2) = \int_0^2 \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}\, dx$$

$$= \frac{\lambda^n}{\Gamma(n)} \int_0^2 x^{n-1} e^{-\lambda x}\, dx$$

$$= \frac{\lambda^n}{\Gamma(n)} \left[\frac{(n-1)!}{\lambda^n} \left(1 - e^{-2\lambda} \sum_{i=0}^{n} \frac{(2\lambda)^i}{i!} \right) \right]$$

$$= \frac{(n-1)!}{\Gamma(n)} \left[1 - e^{-2\lambda} \sum_{i=0}^{n} \frac{(2\lambda)^i}{i!} \right]$$

Problem 5

5a) $\bar{X}_a = \sum_{i=1}^{n} a_i x_i$

$$\sum_{i=1}^{n} E(a_i x_i) = \sum_{i=1}^{n} a_i \, E(x_i)$$

$$= n(a\,E(x)) = \mu$$

$$\Rightarrow n a \mu = \mu$$

$$\Rightarrow a = \frac{1}{n}$$

b) $\mathrm{Var}(\bar{X}_a) = \mathrm{Var}\left(\sum_{i=1}^{n} a_i x_i\right)$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(x_i)$$

$$= \mathrm{Var}(naX)$$

$$= n^2 \, \mathrm{Var}(aX)$$

$$= n^2 a^2 \, \mathrm{Var}(X)$$

$$= n^2 a^2 = 1$$

$$a^2 = \frac{1}{n^2}$$

$$a = \frac{1}{n}$$

Problem 6

6) a) $y_i = X_i + e_i$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}(X_i + e_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} e_i$$

$$= E(X_i) + E(e_i)$$

$$= \mu + \mu_e$$

b) $E\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i\right\}$

$$= \frac{1}{n} E\left(\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(y_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i + e_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) + \frac{1}{n}\sum_{i=1}^{n} E(e_i)$$

$$= \mu + \mu_e$$

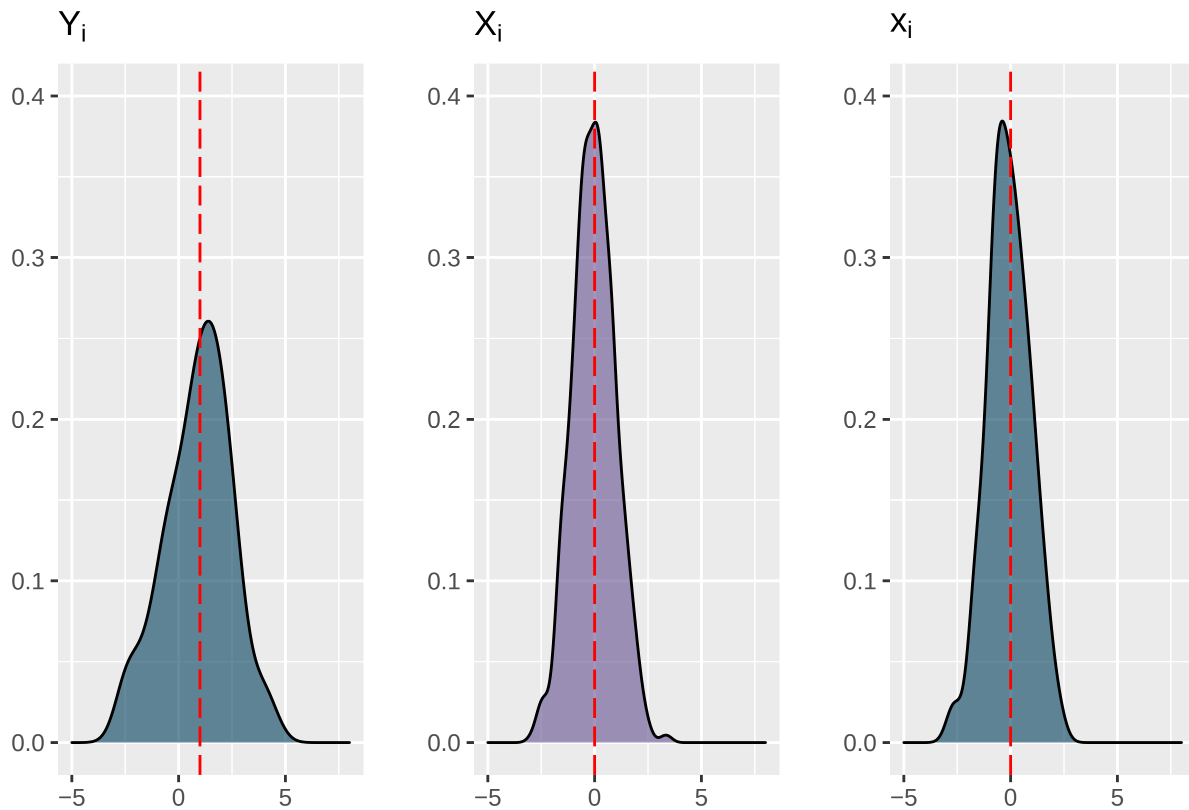
c) $\frac{1}{n}\sum_{i=1}^{n} Y_i$

$E(Y_i) = \mu + \mu_e$

only if μ_e is zero

d) See Code

```r
library(tidyverse)
library(cowplot)

options(repr.plot.width = 16, repr.plot.height = 12, repr.plot.res = 100)

set.seed(27092021)

chart_colors <- c("#003f5c", "#2f4b7c", "#665191", "#a05195",
                  "#d45087", "#f95d6a", "#ff7c43", "#ffa600")

mu <- 0
sigma <- 1
N <- 1000
n <- 100

X_i <- rnorm(N, mu, sigma)
e_i <- rnorm(n, mu + 1, sigma)
x_samples <- sample(X_i, n)
Y_i <- x_samples + e_i

df_sample <- data.frame("Samples" = x_samples, "Error" = e_i, "Y_i" = Y_i)

y_plot <-  ggplot(data = df_sample, aes(x = Y_i)) +
           geom_density(fill = chart_colors[1], alpha = 0.6) +
           ggtitle(label = expression(Y[i])) + xlab("") + ylab("") + xlim(-5, 8) + ylim(0, 0.4) +
           geom_vline(mapping = NULL, xintercept = mu + 1, linetype = "longdash", color = "red")

x_plot <-  ggplot() +
           geom_density(fill = chart_colors[3], alpha = 0.6, aes(x = as.data.frame(X_i)[,1])) +
           ggtitle(label = expression(X[i])) + xlab("") + ylab("") + xlim(-5, 8) + ylim(0, 0.4) +
           geom_vline(mapping = NULL, xintercept = mu, linetype = "longdash", color = "red")

noerror_plot <-  ggplot(data = df_sample, aes(x = x_samples)) +
           geom_density(fill = chart_colors[1], alpha = 0.6) +
           ggtitle(label = expression(x[i])) + xlab("") + ylab("") + xlim(-5, 8) + ylim(0, 0.4) +
           geom_vline(mapping = NULL, xintercept = mu, linetype = "longdash", color = "red")

plot_grid(y_plot, x_plot, noerror_plot, align = 'vh', hjust = -1, nrow = 1, ncol = 3)
```



From the plots above, it can be seen that the sample distribution without error(x_i, where $\mu_e = 0$) is more similar to the population distribution. The sample distribution with error (Y_i, $\mu_e = 1$) shows a wider distribution, indicating that the variance is much higher than the populations.

The the following sample distribution comparison, we can see that Y_i is both skewed and exhibits larger variance than the sample distribution where $\mu_e = 0$. When μ_e is zero, we can say the estimator Y_i is unbiased.

```
df_long <- as_tibble(df_sample)

df_long <- df_long %>% select(Samples, Y_i) %>% gather(source, values, Samples:Y_i)

ggplot(data = df_long) +
           geom_density(alpha = 0.6, aes(fill = source, x = values)) +
           ggtitle(label = expression(X[i])) + xlab("") + ylab("") + xlim(-5, 8) + ylim(0, 0.4) +
           scale_fill_manual( values = chart_colors[c(1,5)])
```

X_i
0.4
0.3
0.2
0.1
0.0
−5
0
5
source
Samples
Y_i