Assignment 2

Brendan Lucas

02/22/2024

Question1: Let $Y_1, \cdots, Y_n \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d, and let $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ be the sample variance with $\bar{Y}$ representing sample mean. Show that

1. $E(S^2) = \sigma^2$.

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

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}\left\langle \sum_{i=1}^{n}(Y_i - \overline{Y})^2 \right\rangle$$

Add $\mu - \mu = 0$ to each term of the sum.

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}\left\langle \sum_{i=1}^{n}\left((Y_i - \mu) - (\overline{Y} - \mu)\right)^2 \right\rangle$$

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}\left\langle \sum_{i=1}^{n}\left((Y_i - \mu)^2 - 2(Y_i - \mu)(\overline{Y} - \mu) + (\overline{Y} - \mu)^2\right) \right\rangle$$

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}\left\langle \sum_{i=1}^{n}(Y_i - \mu)^2 - 2\sum_{i=1}^{n}(Y_i - \mu)(\overline{Y} - \mu) + \sum_{i=1}^{n}(\overline{Y} - \mu)^2 \right\rangle$$

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}\left[n\sigma^2 + \left\langle -2\sum_{i=1}^{n}(Y_i - \mu)(\overline{Y} - \mu) + \sum_{i=1}^{n}(\overline{Y} - \mu)^2 \right\rangle\right]$$

$$\sum_{i=1}^{n}\left\langle(\overline{Y} - \mu)^2\right\rangle = nV(\overline{Y}) = n\frac{\sigma^2}{n} = \sigma^2$$

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}\left[n\sigma^2 + \left\langle -2\sum_{i=1}^{n}(Y_i - \mu)(\overline{Y} - \mu) \right\rangle + \sigma^2\right]$$

$$\left\langle -2\sum_{i=1}^{n}(Y_i - \mu)(\overline{Y} - \mu) \right\rangle = -2\left\langle (\overline{Y} - \mu)\sum_{i=1}^{n}(Y_i - \mu) \right\rangle = -2n\left\langle(\overline{Y} - \mu)^2\right\rangle$$

$$= -2nV(\overline{Y}) = -2n\frac{\sigma^2}{n} = -2\sigma^2$$

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}\left[n\sigma^2 + -2\sigma^2 + \sigma^2\right]$$

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}(n-1)\sigma^2$$

$$\therefore \left\langle S^2 \right\rangle = \sigma^2$$

Q.E.D. S^2 is an unbiased estimator of σ^2.

2. (graduate students): Show that $S^2 \sim \chi^2_{n-1}$. When $Y_1, \dots, Y_n$ are i.i.d.,

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

has a χ^2 distribution with n degrees of freedom (Wackerly et al. Theorem 7.2). So, in order to show that S^2 has a χ^2 distribution with $n-1$ degrees of freedom, we have to show that it can be written in the form Z_i^2.

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

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

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left[(Y_i - \overline{Y})^2 - 2(Y_i - \overline{Y})(\overline{Y} - \mu) + (\overline{Y} - \mu)^2\right]$$

$$\sum_{i=1}^{n} \left[-2(Y_i - \overline{Y})(\overline{Y} - \mu)\right] = -2(\overline{Y} - \mu) \sum_{i=1}^{n} (Y_i - \overline{Y})$$

$$= -2(\overline{Y} - \mu)(n\overline{Y} - n\overline{Y}) = 0$$

$$\therefore \quad \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n} (Y_i - \overline{Y})^2 + \sum_{i=1}^{n} (\overline{Y} - \mu)^2\right]$$

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

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

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

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

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

The first term on the right is a χ^2 random variable with n degrees of freedom.

$$\mathcal{Q}_n \equiv \sum_{i=1}^{n} \frac{(Y_i - \mu)^2}{\sigma^2} \sim \chi^2_n$$

$$S^2 = \frac{\sigma^2}{n-1} \mathcal{Q}_n - \frac{n}{n-1}(\overline{Y} - \mu)^2$$

$$\frac{(n-1)S^2}{\sigma^2} = \mathcal{Q}_n - \frac{n(\overline{Y} - \mu)^2}{\sigma^2}$$

The second term on the right-hand side is the square of a normal random variable.

$$\sqrt{\frac{n(\overline{Y} - \mu)^2}{\sigma^2}} = \frac{(\overline{Y} - \mu)}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \quad \because \quad E(\overline{Y}) = \mu, \;\; Var(\overline{Y}) = \frac{\sigma^2}{n}$$

The square of a normal random variable is a chi-square variable with 1 degree of freedom.

$$\frac{n(\overline{Y}-\mu)^2}{\sigma^2} = \mathcal{Q}_1 \sim \chi_1^2$$
$$\frac{(n-1)S^2}{\sigma^2} = \mathcal{Q}_n - \mathcal{Q}_1 \sim \chi_{n-1}^2$$

The difference between these χ_n^2 and χ_1^2 random variables can be demonstrated to be χ_{n-1}^2-distributed with the help of moment-generating functions.

$$\mathcal{Q}_n = \frac{(n-1)S^2}{\sigma^2} + \mathcal{Q}_1$$
$$M_{\mathcal{Q}_n}(t) = E(e^{t\mathcal{Q}_n}) = E\left[e^{t\frac{(n-1)S^2}{\sigma^2}+\mathcal{Q}_1}\right] = E\left[e^{\left(t\frac{(n-1)S^2}{\sigma^2}\right)}\right] E\left[e^{t\mathcal{Q}_1}\right]$$
$$M_{\mathcal{Q}_n}(t) = (1-2t)^{-n/2}, \quad M_{\mathcal{Q}_1}(t) = (1-2t)^{-1/2}$$
$$E\left[e^{\left(t(n-1)S^2/\sigma^2\right)}\right] \equiv M_{(n-1)S^2/\sigma^2}(t)$$
$$(1-2t)^{-n/2}(1-2t)^{1/2} = M_{(n-1)S^2/\sigma^2}(t)$$
$$M_{(n-1)S^2/\sigma^2}(t) = (1-2t)^{-(n-1)/2} = M_{\mathcal{Q}_{(n-1)}}(t) \iff \mathcal{Q}_{(n-1)} \sim \chi_{n-1}^2$$

That is to say, $\frac{(n-1)S^2}{\sigma^2}$ has the same moment-generating function as a χ_{n-1}^2-distributed random variable. Moment-generating functions are unique, so $\frac{(n-1)S^2}{\sigma^2}$ is a chi-squared random variable with $n-1$ degrees of freedom.

$$M_{(n-1)S^2/\sigma^2}(t) = M_{\mathcal{Q}_{(n-1)}}(t) \therefore \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Q.E.D. $S^2 \sim \chi_{n-1}^2$

3. Show that $\bar{Y}$ and S^2 are independent random variables. $\overline{Y}$ and S^2 are independent implies $\overline{Y}$ and $Y-\overline{Y}$ are independent, because S^2 is a function of $Y-\overline{Y}$.

   (a) Joint distribution $f_{\{Y_1,\ldots,Y_n\}}(y_1,\ldots,y_n)$:

$$f_{\{Y_1,\ldots,Y_n\}}(y_1,\ldots,y_n) = \prod_{i=1}^n f_{Y_i}(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{y_i-\mu}{\sqrt{2}\sigma}\right)^2$$
$$\therefore f_{\{Y_1,\ldots,Y_n\}}(y_1,\ldots,y_n) = \frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left(-\frac{1}{2}\sum_{i=1}^n\left(\frac{y_i-\mu}{\sigma}\right)^2\right)$$

   (b) Define $\Delta Y_i \equiv Y_i - \overline{Y}$ to transform the joint distribution with a Jacobian of constant n:

$$\Delta Y_1 = \overline{Y}$$
$$\Delta Y_2 = Y_2 - \overline{Y}, \qquad Y_2 = \Delta Y_2 + \Delta Y_1$$
$$\Delta Y_3 = Y_3 - \overline{Y}, \qquad Y_3 = \Delta Y_3 + \Delta Y_2$$
$$\cdots$$
$$\Delta Y_n = Y_n - \overline{Y}, \qquad Y_n = \Delta Y_n + \Delta Y_1$$
$$f_{\{Y_1,\ldots,Y_n\}}(y_1,\ldots,y_n) = \frac{1}{n} f_{\{\Delta Y_1,\ldots,\Delta Y_n\}}(\Delta y_1,\ldots,\Delta y_n)$$

   A density function on this difference ΔY is $C \times f_{\{Y_1,\ldots,Y_n\}}(y_1,\ldots,y_n)$ where C is a constant that we need not compute at present.

$$f_{\{\Delta Y_1,\ldots,\Delta Y_n\}}(\Delta y_1,\ldots,\Delta y_n) = C\exp\left(-\frac{1}{2}\sum_{i=1}^n\left(\frac{y_i-\mu}{\sigma}\right)^2\right)$$

(c) We know from the previous derivation ($Y_i \to y_i,\;\; S^2 \to s^2$ as we are looking at observations in the density function):

$$\sum_{i=1}^{n} \left(\frac{y_i - \mu}{\sigma}\right)^2 = \frac{(n-1)s^2}{\sigma^2} + \frac{n(\overline{y}-\mu)^2}{\sigma^2}$$

$$f_{\{\Delta Y_1,\ldots,\Delta Y_n\}}(\Delta y_1,\ldots,\Delta y_n) = C\exp\left(-\frac{1}{2}\left[\frac{(n-1)s^2}{\sigma^2} + \frac{n(\overline{y}-\mu)^2}{\sigma^2}\right]\right)$$

(d) We find that this joint density function can be written as the product of two density functions, that of the sample mean $\overline{Y}$ and that of the sample variance S^2.

$$f_{S^2}(s^2) = C_1 \exp\left(-\frac{1}{2}\left(\frac{(n-1)s^2}{\sigma^2}\right)\right)$$

$$f_{\overline{Y}}(\overline{y}) = C_2 \exp\left(-\frac{1}{2}\left(\frac{n(\overline{y}-\mu)^2}{\sigma^2}\right)\right)$$

$$C = C_1C_2$$

$$\therefore \;\; f_{\{\Delta Y_1,\ldots,\Delta Y_n\}}(\Delta y_1,\ldots,\Delta y_n) = f_{S^2}(s^2)f_{\overline{Y}}(\overline{y})$$

(e) That the joint density function can be written as the product of two distributions, one of the S^2 and one of the $\overline{Y}$, is the very definition of independent random variables.

(f) Q.E.D. S^2 and $\overline{Y}$ are independent random variables.

Question2: Use `R` to perform the following.

```r
library(ggplot2)
library(gridExtra)
```

1. Find $t_{.07}$ for a t-distributed random variable with 5 df.

```r
t_07_5df <- qt(0.07, df = 5)
print(t_07_5df)
```

```
## [1] -1.75289
```

1. Find $\chi^2_{.95}$ for a χ^2-distributed random variable with 4 df.

```r
chisq_95_4df <- qchisq(0.95, df = 4)
print(chisq_95_4df)
```

```
## [1] 9.487729
```

1. Find $Pr(\sum_{i=1}^{11} Z_i^2 \leq 11)$, where $Z_i \sim \mathcal{N}(0,1)$ are i.i.d.

$\sum_{i=1}^{11} Z_i^2$ is the definition of a χ^2 variable with 11 degrees of freedom, and we want to cumulative distribution function.

```r
prob_chisq_11leq_11df <- pchisq(11, df = 11)
print(prob_chisq_11leq_11df)
```

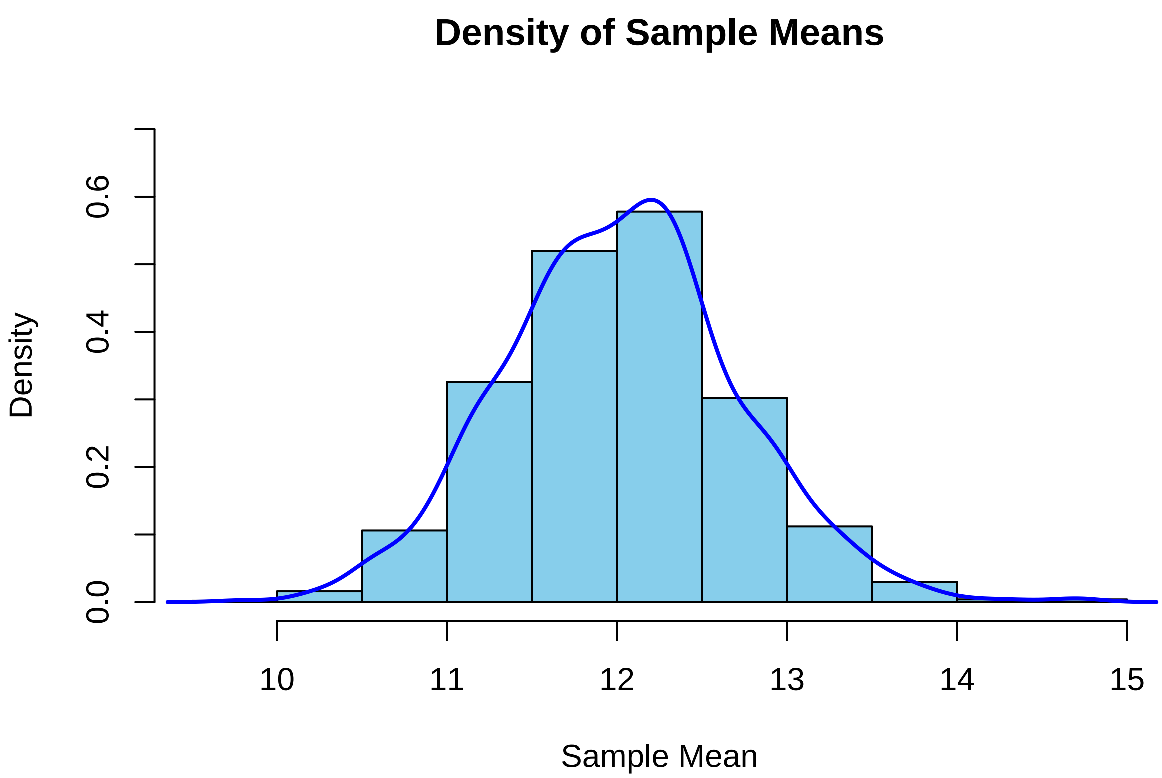
```
## [1] 0.5567367
```

1. Simulate a random draws ($n = 100$) from a Gamma distribution with parameters $\alpha = 3$, and $\lambda = 1/4$. Calculate the sample mean and sample variance. Repeat the process $k = 1000$ times and plot the density of the sample mean and sample variance. Comment on your observations.

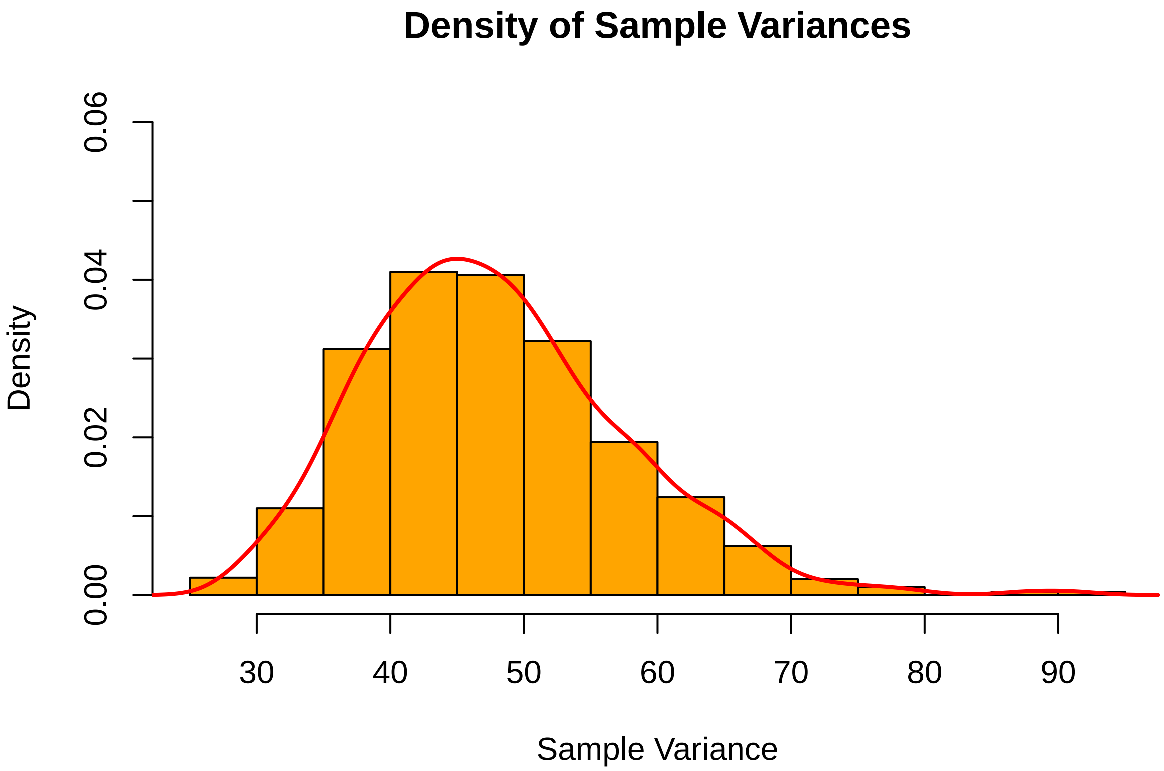
```r
#define the parameters
k <- 1000
n <- 100
alpha=3
lambda=0.25

#run the simulation
samples <- replicate(k, rgamma(n, shape=alpha, rate=lambda))
sample_means <- colMeans(samples)
sample_variances <- apply(samples, 2, var)
results <- data.frame(sample_means, sample_variances)

# Plot of Sample Mean
hist(results$sample_means, prob = TRUE, main = "Density of Sample Means",
    xlab = "Sample Mean", ylab = "Density", ylim = c(0, 0.7),
    border = "black", col = "skyblue")
lines(density(results$sample_means), col = "blue", lwd = 2)
```




```r
# Plot of Sample Variance
hist(results$sample_variances, prob = TRUE, main = "Density of Sample Variances",
    xlab = "Sample Variance", ylab = "Density", ylim = c(0, 0.06),
    border = "black", col = "orange")
lines(density(results$sample_variances), col = "red", lwd = 2)
```



My observation is that the density of the sample mean is shaped much more like a normal distribution, and the density of the sample variance is shaped much more like a χ^2 distribution, as is to be expected from the content of the derivations in the first part of the assignment.