Problem Set 5

Troy Jennings

Introduction

As an application of the some of the properties of expected values, problems 1-7 step through a proof that the expected value of the random variable that defines sample variance is the population variance, given that the population variance is defined.

For each of these questions, let $X_1, X_2, ...X_n$ be independent, identically distributed random variables with defined mean μ and variance σ^2.

Question 8 gives examples of jointly distributed random variables that are independent and jointly distributed random variables that aren't independent.

Please complete the following tasks regarding the data in R. Please generate a solution document in R markdown and upload the .Rmd document and a rendered .doc, .docx, or .pdf document. Please turn in your work on Canvas.

These questions were rendered in R markdown through RStudio (https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf, http://rmarkdown.rstudio.com).

Question 1 (5 points)

Let $X_1, X_2, ...X_n$ be independent, identically distributed random variables with mean μ and variance σ^2, and define the random variable $\bar{X}$ by $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Justify the equality

$$E\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = E\left[\sum_{i=1}^{n} X_i^2\right] - 2E\left[\sum_{i=1}^{n} \bar{X}X_i\right] + E\left[\sum_{i=1}^{n} \bar{X}^2\right]$$

Question 1 Solution

We can use the the fact, $(a-b)^2 = a^2 - 2ab + b^2$, to show that $(X_i - \bar{X})^2 - X_i^2 - 2X_i\bar{(}X) + \bar{X}^2$.

Question 2 (5 points)

Let $X_1, X_2, ...X_n$ be independent, identically distributed random variables with mean μ and variance σ^2.

In terms of μ and σ^2, what is the value of $E[X_i^2]$? Note that $Var[X_i] = E[X_i^2] - E[X_i]^2$, while $Var[X_i] = \sigma^2$ and $E[X_i] = \mu$. Please justify your answer.

Question 2 Part A Solution

If $Var[X_i] = \sigma^2$ and $E[X_i]^2 = \mu^2$, then using basic algebra, $E[X_i^2] = \sigma^2 + \mu^2$

Confirm numerically that your answer is correct for $X_i \sim gamma(shape = 3, scale = 2)$ which has mean equal to 6 and variance equal to 12.

Question 2 Part B Solution

Since we know that $E[X_i^2]$ must be $\sigma^2 + \mu^2$, then we can plug in to solve, given the values for the mean and variance:

$$
\begin{aligned}
E[X_i^2] &= \sigma^2 + \mu^2 \\
&= 6^2 + \sqrt{12^2} \\
&= 6^2 + 12 = 48
\end{aligned}
$$

```
f2 <- function(x){ x^2*dgamma(x,shape= 3,scale= 2) }
integrate(f2, 0, Inf)$value
```

```
## [1] 48
```

Question 3 (5 points)

Assuming that $E[X_i^2] = \sigma^2 + \mu^2$ for all i, what is $E\left[\sum_{i=1}^n X_i^2\right]$. Recall that $E\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n E[Y_i]$ for any random variables $Y_i, Y_2 ... Y_n$ with defined means.

Question 3 Solution

We have $E\left[\sum_{i=1}^n X_i^2\right] = \sum_{i=1}^n E[\bar{X}^2] = \sum_{i=1}^n (\sigma^2 + \mu^2)$. Since $sigma^2 + \mu^2$ is a constant, we can rewrite as $E\left[\sum_{i=1}^n X_i^2\right] = n(\sigma^2 + \mu^2)$ using the properties of linearity.

Question 4 (5 points)

Define the random variable $\bar{X}$ by $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. What is the value of $E\left[\sum_{i=1}^n \bar{X}^2\right]$? Please justify your answer.

Recall that the mean of $\bar{X}$ equals μ and the variance equals $\frac{\sigma^2}{n}$. The fact that $E\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n E[Y_i]$ mentioned above may also be useful. Further, $\bar{X}$ is constant with respect to the index i in the sum.

Question 4 Solution

We know that $E\left[\sum_{i=1}^n \bar{X}^2\right] \equiv Var\left(E\left[\sum_{i=1}^n \bar{X}^2\right]\right) + \mu^2$. Substituting the value we are given for the variance, we are left with

$$
\begin{aligned}
E\left[\sum_{i=1}^n \bar{X}^2\right] &= \sum_{i=1}^n E[\bar{X}^2] \\
&= \sum_{i=1}^n \left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= n\left(\frac{\sigma^2}{n} + \mu^2\right)
\end{aligned}
$$

Question 5 (10 points)

Why is

$$E\left[\sum_{i=1}^{n} \bar{X} X_i\right] = E\left[\bar{X} \sum_{i=1}^{n} X_i\right] = E\left[n\bar{X}^2\right]$$

Question 5 Solution

Since our variable of indexing is i, we can use properties of linearity to "pull" $\bar{X}$ out of the summation, leaving $E[\bar{X}\sum_{i=1}^{n} \bar{X}_i]$. We know that $\sum_{i=1}^{n} \bar{X} = n\bar{X}$, which leaves $E[\bar{X}n\bar{X}] = E[n\bar{X}^2]$.

Question 6 (5 points)

Assuming that $E\left[\sum_{i=1}^{n} X_i^2\right] = n\left(\sigma^2 + \mu^2\right)$, that $E\left[\sum_{i=1}^{n} \bar{X}^2\right] = n\left(\frac{\sigma^2}{n} + \mu^2\right)$, and that $E\left[\sum_{i=1}^{n} \bar{X} X_i\right] = E\left[n\bar{X}^2\right] = n\left(\frac{\sigma^2}{n} + \mu^2\right)$, please simplify $E\left[\sum_{i=1}^{n} X_i^2\right] - 2E\left[\sum_{i=1}^{n} \bar{X} X_i\right] + E\left[\sum_{i=1}^{n} \bar{X}^2\right]$.

Question 6 Solution

Simplifying, we have

$$\begin{aligned}
E[\sum_{i=1}^{n} X_i^2] - 2E[\sum_{i=1}^{n} \bar{X} X_i] + E[\sum_{i=1}^{n} \bar{X}^2] &= n(\sigma^2 + \mu^2) - 2(n(\frac{\sigma^2}{n} + \mu^2)) + n(\frac{\sigma^2}{n} + \mu^2) \\
&= n\sigma^2 + n\mu^2 - 2\sigma^2 - 2n\mu^2 + \sigma^2 + n\mu^2 \\
&= n\sigma^2 - \sigma^2 \\
&= (n-1)\sigma^2
\end{aligned}$$

Question 7 (5 points)

If $E\left[\sum_{i=1}^{n} X_i^2\right] - 2E\left[\sum_{i=1}^{n} \bar{X} X_i\right] + E\left[\sum_{i=1}^{n} \bar{X}^2\right] = (n-1)\sigma^2$, what is the value of $E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]$?

Question 7 Solution

First, we can factor $E\left[\sum_{i=1}^{n} X_i^2\right] - 2E\left[\sum_{i=1}^{n} \bar{X} X_i\right] + E\left[\sum_{i=1}^{n} \bar{X}^2\right]$ as follows:

$$E[\sum_{i=1}^{n} X_i^2] - 2E[\sum_{i=1}^{n} \bar{X} X_i] + E[\sum_{i=1}^{n} \bar{X}^2] = E[\sum_{i=1}^{n}(X_i^2 - 2\bar{X}X_i + X^2)]$$

From this, we can see that this takes the form $(X_i - \bar{X})^2$. Using this information and, we can conclude that:

$$\begin{aligned}
E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i\bar{X})^2\right] &= E\left[\frac{1}{n-1}(n-1)\sigma^2\right] \\
&= E\left[\sigma^2\right] = \sigma^2
\end{aligned}$$

Question 8

- Consider the probability space defined by (S, M, P) where $S = \{a, b, c, d, e, f\}$, the set of events M is the power set of S, and P is defined by the density $f(s) = \frac{1}{6}$ for all $s \in S$. Let X be the random variable on this probability space defined by $X(a) = X(b) = X(c) = 1$ and $X(d) = X(e) = X(f) = 0$. Define Y by $Y(a) = Y(d) = 2$, $Y(b) = Y(c) = Y(e) = Y(f) = 3$. Are these random variables independent?

Question 8 Solution

From the problem text, we have the following probabilities:

$$f_X(0) = \frac{1}{2}$$
$$f_X(1) = \frac{1}{2}$$
$$f_Y(2) = \frac{1}{3}$$
$$f_Y(3) = \frac{2}{3}$$

Then finding the probabilities for both events, f_{XY}:

$$f_{XY}(0,2) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$
$$f_{XY}(0,3) = \frac{1}{2} \cdot \frac{2}{3} = \frac{2}{6}$$
$$f_{XY}(1,2) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$
$$f_{XY}(1,3) = \frac{1}{2} \cdot \frac{2}{3} = \frac{2}{6}$$

We have shown that $P(X|Y) \neq P(X)$ for all cases, and therefore, the random variables X and Y are not independent.

Question 9

Suppose a is a sample from a random variable A and b is a sample from a random variable B with variances v and w respectively. What weighted average $xa+(1-x)b$ with $x \in [0, 1]$ minimizes the variance of $xa+(1-x)b$?

Question 8 Solution

To minimize the variance of $xa + (1-x)b$, we first need to take the simplify the expression. We can do so by using the property of linearity and the fact that $Var[xa + (1-x)b] = Var[xa] + Var[(1-x)b]$. Treating x and $(1-x)$ as constants, we can pull those terms, squared, out of the variance: $x^2 \cdot Var[a] + (1-x)^2 Var[b]$.

Next, we can then take the derivative, set it equal to 0, and solve for x.

$$
\begin{aligned}
f(x) &= x^2a + (1-x)^2b \\
\tfrac{d}{dx}[f(x)] &= \tfrac{d}{dx}[x^v] + \tfrac{d}{dx}[x^2w] - 2\tfrac{d}{dx}[wx] + \tfrac{d}{dx}[w] \\
0 &= 2xv + 2xw - 2w + 0 \\
2w &= 2x(v+w) \\
\frac{2w}{(v+w)} &= 2x \\
\frac{2w}{2(v+w)} &= x \\
\frac{w}{(v+w)} &= x
\end{aligned}
$$