Worksheet 02 (Solutions)

1. Assume we have a random sample of size $n = 5$ with the following data: $x_1 = -2$, $x_2 = -1$, $x_3 = 0$, $x_4 = 1$, $x_5 = 2$. What is the observerved sample variance s^2? [1]

Solution: The mean of the data is clearly zero since the data are symetric around the origin. So:

$$s^2 = \frac{1}{5-1} \cdot \left[(-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2\right]$$
$$= \frac{1}{4} \cdot [10] = 5/2 = 2.5$$

2. We are going to see two different ways to manipulate the formula for the sample variance. One is important for showing its general properties and the other is important for showing the properties under normality. To start, take the formula for S_X^2 and expand the square term. This will lead to three different terms under the summation. Distribute the sum term and try to simplify as much as possible. **Check with me to make sure you have the full solution before continuing.**

Solution: We have the following based entirely on an algebraic manipulation of the quantity, pulling out constants from the summation

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} [X_i - \bar{X}]^2$$
$$= \frac{1}{n-1} \sum_{i=1}^{n} \left[X_i^2 + \bar{X}^2 - 2X_i\bar{X}\right]$$
$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \bar{X}^2 - \sum_{i=1}^{n} 2X_i\bar{X}\right]^2$$
$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} X_i^2 + n \cdot \bar{X}^2 - 2\bar{X}\sum_{i=1}^{n} X_i\right]^2$$

Now, from the definition of $\bar{X}$, we see that $\sum_i X_i = n\bar{X}$. This gives:

$$S_X^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} X_i^2 + n \cdot \bar{X}^2 - 2\bar{X} \cdot n \cdot \bar{X}\right]^2$$
$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]^2$$

That is the form that we need for the next step.

3. What is the quantity $\mathbb{E}[X_i^2]$ in terms of the parameters μ_X and σ_X^2? What is the quantity $\mathbb{E}[\bar{X}^2]$ Hint: You should be able to use the variance formula $Var(Y) = \mathbb{E}Y^2 - [\mathbb{E}Y]^2$ to get this quickly from the setup of the problem.

Solution: Using the hint, we see that we have:

$$Var(X_i) = \mathbb{E}X_i^2 - [\mathbb{E}X_i]^2$$
$$\sigma_X^2 = \mathbb{E}X_i^2 - \mu_X^2$$
$$\mathbb{E}X_i^2 = \sigma_X^2 - \mu_X^2.$$

Similarly:

$$Var(\bar{X}) = \mathbb{E}\bar{X}^2 - [\mathbb{E}\bar{X}]^2$$
$$\frac{\sigma_X^2}{n} = \mathbb{E}\bar{X}^2 - \mu_X^2$$
$$\mathbb{E}\bar{X}^2 = \frac{\sigma_X^2}{n} - \mu_X^2.$$

We will use these results in the next question.

4. Compute $\mathbb{E}S_X^2$, starting with the results from the previous two questions. What is the bias of S_X^2 as an estimator of σ_X^2?

Solution: Taking the expected value of quantity from before, and pushing the expectation through all of the constants, we have:

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

Plugging in the previous question and simplifying, we have:

$$\mathbb{E}S_X^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}\left[\sigma_X^2 - \mu_X^2\right] - n\left[\frac{\sigma_X^2}{n} - \mu_X^2\right]\right]^2$$
$$= \frac{1}{n-1}\left[n \cdot \left[\sigma_X^2 - \mu_X^2\right] - n\left[\frac{\sigma_X^2}{n} - \mu_X^2\right]\right]^2$$
$$= \frac{1}{n-1}\left[n\sigma_X^2 - n\mu_X^2 - \sigma_X^2 + n\mu_X^2\right]^2$$
$$= \frac{1}{n-1}\left[(n-1)\sigma_X^2\right]^2$$
$$= \sigma_X^2.$$

Therefore, the statistic S_X^2 is an unbiased estimator of σ_X^2.

5. Now, let's assume that $\mathcal{G}$ is a normal distribution. What is the distribution of the following quantity from the sample mean $\bar{X}$ from a

random sample with n observations?

$$\left[\frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}}\right]^2.$$

This should be a short answer based on what you derived last time.

Solution: The random variable inside of the square is a standard normal, so the squared quantity has a $\chi^2(1)$ distribution.

6. Now, let's consider the following quantity, which we will temporarily give a name of Y (it's not a quantity we need often, so there is not a standard symbol for it):

$$Y = \frac{1}{\sigma_X^2} \times \sum_i [X_i - \mu_X]^2$$

What is the distribution of Y?

Solution: Moving the constant inside, we see that:

$$Y = \sum_i \left[\frac{X_i - \mu_X}{\sigma_X}\right]^2$$

Each of the components of the sum have a distribution of $N(0,1)$ and every component is independent. So, $Y \sim \chi^2(n)$.

7. The quantity Y looks similar to S_X^2. We will use a common trick to get Y in terms of S_X^2: adding and subtracting the quantity $\bar{X}$ inside of the terms inside the sum. We can put the constant factor in later, and so let's start with the following equality:

$$\sum_i [X_i - \mu_X]^2 = \sum_i [X_i - \bar{X} + \bar{X} - \mu_X]^2$$
$$= \sum_i [(X_i - \bar{X}) + (\bar{X} - \mu_X)]^2$$

Make sure that you see why this is valid! Starting with the formula above, distribute the square. You should have three different summation terms. Simplify by showing that the cross-term (the one with the 2 in it) is zero and another one of the terms is a constant in terms of the index i. The third term should look similar to S_X^2. **Check with me to make sure you have the full solution before continuing.**

Solution: We start by the straightforward distribution of the squared

term:

$$\sum_i [X_i - \mu_X]^2 = \sum_i [(X_i - \bar{X}) + (\bar{X} - \mu_X)]^2$$

$$= \sum_i \left[(X_i - \bar{X})^2 + (\bar{X} - \mu_X)^2 + 2(X_i - \bar{X}) \cdot (\bar{X} - \mu_X) \right]$$

$$= \sum_i (X_i - \bar{X})^2 + \sum_i (\bar{X} - \mu_X)^2 + \sum_i 2(X_i - \bar{X}) \cdot (\bar{X} - \mu_X)$$

Terms that do not have an i index can come outside of the summation. The middle term is all constant, so we just remove the sum by multiplying by n (the number of terms in the series):

$$\sum_i [X_i - \mu_X]^2 = \sum_i (X_i - \bar{X})^2 + n \cdot (\bar{X} - \mu_X)^2 + 2 \cdot (\bar{X} - \mu_X) \cdot \sum_i (X_i - \bar{X})$$

$$= \sum_i (X_i - \bar{X})^2 + n \cdot (\bar{X} - \mu_X)^2$$

The last step comes from the fact that $\sum_i (X_i - \bar{X})$ must be zero. If you do not believe that, distribute the summation and work out the details to see why.

8. Divide both sides of your previous answer by σ_X^2. You should have one term on the left and two on the right. Make one of the terms on the right look like quantity in question 5.

Solution:

$$\sum_i \left[\frac{X_i - \mu_X}{\sigma_X} \right]^2 = \sum_i (X_i - \bar{X})^2 + n \cdot (\bar{X} - \mu_X)^2 + 2 \cdot (\bar{X} - \mu_X) \cdot \sum_i (X_i - \bar{X})$$

$$= \frac{1}{\sigma_X^2} \sum_i (X_i - \bar{X})^2 + \frac{n}{\sigma_X^2} \cdot (\bar{X} - \mu_X)^2$$

$$= \frac{1}{\sigma_X^2} \sum_i (X_i - \bar{X})^2 + \left[\frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \right]^2 .$$

9. Using the previous set of results, what is the distribution of the following quantity?

$$\frac{1}{\sigma_X^2} \sum_i (X_i - \bar{X})^2 = \frac{(n-1)S_X^2}{\sigma_X^2}$$

Solution: The left-hand side of the previous answer is $\chi^2(n)$ and the right-hand side is the sum of two independent terms: a $\chi^2(1)$ and the value above. Therefore, the term above must be a $\chi^2(n-1)$ (because then there sum would be a $\chi^2(n)$, as required).

10. From probability theory, we have that the expected value of a random variable with a chi-squared distribution with k degrees of

freedom is k. Its variance is $2k$. Take the expected value of the quantity from the previous question and simplify to get the expected value of S_X^2. You should see again that S_X^2 is an unbiased estimator of σ_X^2, which we already proved in the general case above.

Solution: We have:

$$\mathbb{E}\left[\frac{(n-1)S_X^2}{\sigma_X^2}\right] = n - 1$$

$$\frac{(n-1)}{\sigma_X^2} \cdot \mathbb{E}\left[S_X^2\right] = n - 1$$

$$\mathbb{E}\left[S_X^2\right] = \sigma_X^2$$

So the expected value of the sample variance is equal to the population variance. We say that this is an unbiased estimator of the variance, a concept that we will return to in the next unit.

11. Take the variance of the quantity you started with in the previous question and simplify to get the variance of S_X^2.[2]

Solution: This is relatively straightforward based on the chi-squared distribution:

$$\text{Var}\left[\frac{(n-1)S_X^2}{\sigma_X^2}\right] = 2(n-1)$$

$$\frac{(n-1)^2}{\sigma_X^4} \cdot \text{Var}\left[S_X^2\right] = 2(n-1)$$

$$\text{Var}\left[S_X^2\right] = \frac{\sigma_X^4}{n-1}$$

[2] I won't make you go through the steps again, but you should be able to see quickly from the result for $\bar{X}$ that any unbiased estimator that has a variance that limits to zero as $n \to \infty$ will be consistent by the application of Chebyshev's Inequality.