

## Worksheet 02

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 observed sample variance  $s^2$ ?<sup>1</sup>

<sup>1</sup> I tried to pick numbers to make the arithmetic easy to do.

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.**

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

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  $\sigma_X^2$ ?

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.

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  $Y$ ?

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:

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

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.**

8. Divide both sides of your previous answer by  $\sigma_X^2$ . 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 1.

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

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

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  $\sigma_X^2$ , which we already proved in the general case above.

11. Take the variance of the quantity you started with in the previous question and simplify to get the variance of  $S_X^2$ .<sup>2</sup>

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