Worksheet 02 (Solutions)

1. 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.

2. 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):

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

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, $Y \sim \chi^2(n)$.

3. 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:

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

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. This is somewhat tricky. Make sure you check the anwer before moving on.

Solution: We start by the straightforward distribution of the squared term:

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

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):

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

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.

4. 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 1.

Solution:

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

5. 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).

6. 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 that S_X^2 is an unbiased estimator of σ_X^2.

Solution: We have:

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

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.

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

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

$$\begin{aligned}
\mathrm{Var}\left[\frac{(n-1)S_X^2}{\sigma_X^2}\right] &= 2(n-1) \\
\frac{(n-1)^2}{\sigma_X^4}\cdot\mathrm{Var}\left[S_X^2\right] &= 2(n-1) \\
\mathrm{Var}\left[S_X^2\right] &= \frac{\sigma_X^4}{n-1}
\end{aligned}$$