Handout 02: Sample Variance

As on the previous handout, consider a random sample of size n $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{G}$ from a distribution $\mathcal{G}$ with a population mean of μ_X and population variance of σ_X^2.[1] Last time we looked at the sample mean. Today, we define the **sample variance**, a sample statistic defined and denoted by:

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

The sample mean $\bar{X}$ and sample variane S_X^2 are independent random variables.[2] As suggested by the name, the sample variance can be used as a good point estimator for the population variance. On the worksheet, we will show that it is an unbiased and consistent estimator of the population variance.

If $\mathcal{G}$ is a normal distribution, then we can show that the sample variance has distribution equal to a scaled chi-squared distribution.[3] Specifically, we have:

$$\frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi^2(n-1).$$

This will also hold in the limit of large n for other distributions $\mathcal{G}$ due to the central limit theorem.

Assume that we have two independent random variables: $Z \sim N(0,1)$ and $C \sim \chi^2(k)$. Then, define the following ratio between the two random variables:

$$T = \frac{Z}{\sqrt{C/k}}.$$

It should be clear that this random variable has a well defined distribution that depends only on the degrees of freedom k of the chi-squared distribution. The distribution is called **Student's t-distribution** with k degrees of freedom, which is denoted by $t(k)$. As we will motivate on today's worksheet, the t-distribution has a mean of zero, is symmetric around the origin, and will be limit to a standard normal as the degrees of freedom limits to infinity. We will see next time how the t-distribution allows us to combine information about the sample mean (which has a scaled standard normal distribution) and the sample variance (which has a scaled chi-squared distribution).

[1] I am intentionally repeating much of the terminology from the first handout. If any of the statements here are unclear, make sure to go back and lookup what each term means.

[2] We will not prove this as the proof is a bit complex (it requires a n-dimenisional change of variabels formula) and not particularly insightful.

[3] Recall that a chi-squared with k degrees of freedom results from taking k independent random variables with standard normal distributions, squaring them, and taking their sum.