Student's t distribution - supplement to chapter 3

For large samples,

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \tag{1}$$

has approximately a standard normal distribution. The parameter σ is often unknown and so we must replace σ by s, where s is the square root of the sample variance. This new statistic is usually denoted with a t.

$$t = \frac{\bar{X}_n - \mu}{s/\sqrt{n}} \tag{2}$$

If the sample is large, the distribution of t is also approximately the standard normal. What if the sample is not large? In the absence of any further information there is not much to be said other than to remark that the variance of t will be somewhat larger than 1. If we assume that the random variable X has a normal distribution, then we can say much more. (Often people say the population is normal, meaning the distribution of the rv X is normal.)

So now suppose that X is a **normal** random variable with mean μ and variance σ^2. The two parameters μ and σ are unknown. Since a sum of independent normal random variables is normal, the distribution of Z_n is exactly normal, no matter what the sample size is. Furthermore, Z_n has mean zero and variance 1, so the distribution of Z_n is exactly that of the standard normal. The distribution of t is not normal, but it is completely determined. It depends on the sample size and a priori it looks like it will depend on the parameters μ and σ. With a little thought you should be able to convince yourself that it does not depend on μ. We will show that it does not depend on σ either. So it only depend on n. It is called "Student's t distribution with $n-1$ degrees of freedom."

To study the distribution of t we first introduce another distribution, the chi-squared or χ^2 distribution. Let $Z_1, Z_2, \cdots, Z_n$ be independent standard normal RV's. Let

$$X = \sum_{i=1}^{n} Z_i^2 \tag{3}$$

Then X has the chi-squared distribution with n degrees of freedom. It can be shown that this is the gamma distribution with $\alpha = n/2$ and $\beta = 1/2$. (This will be a problem in homework 3.)

Theorem 1. *Let $X_1, X_2, \cdots, X_n$ be a random sample (i.i.d.) from a normal population with mean μ and variance σ^2. As before, define*

$$\begin{aligned} \bar{X}_n &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \end{aligned} \tag{4}$$

Then
(a) $\bar{X}_n$ and s^2 are independent random variables.
(b) $\bar{X}_n$ is normal with mean μ and variance σ^2.
(c) $(n-1)s^2/\sigma^2$ has a chi-squared distribution with $n-1$ degrees of freedom.

We do not give the proof. It is an interesting probability exercise.

Now consider the random variable t. We define

$$U = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \tag{5}$$

and

$$V = \frac{(n-1)s^2}{\sigma^2} \tag{6}$$

Then we can write t as

$$t = \frac{U}{\sqrt{V/(n-1)}} \tag{7}$$

The distribution of this random variable is called Student's t distribution with $n-1$ degrees of freedom. Like the normal distribution there are tables of it, and it is part of any statistics package. As $n \to \infty$ it converges to the standard normal.

Now suppose we assume that our population is normal and our sample is not necessarily lare. Then we can estimate its mean using $\bar{X}_n$ and use the t statistic to see how good this estimate is. In particular, we can find confidence intervals. For example, from the tables one finds that if $n = 10$

(so the number of degrees of freedom is 9), then $P(t \leq 1.833) = 0.95$ and $P(t \leq 2.262) = 0.975$. Thus the 90% confidence interval would be

$$[\bar{X}_n - 1.833s/\sqrt{n}, \bar{X}_n + 1.833s/\sqrt{n}] \tag{8}$$

and the 95% confidence interval would be

$$[\bar{X}_n - 2.262s/\sqrt{n}, \bar{X}_n + 2.262s/\sqrt{n}] \tag{9}$$