8.4 The t distribution

A typical statistical investigation will involve estimating the mean of a population using the mean of a sample. We know that this estimator is consistent, so that if we take a large sample, there is a high probability that the estimate is close to the mean. In many cases, we would like to assess the variability in the estimate (especially if we want to know **how** large a sample we need). But the variance of the sampling distribution for the mean depends on the variance of the population, which we presumably also don't know. So we also estimate this parameter using the sample variance. Miraculously, for samples from a Normal population, these two estimators are **independent**!

We can use this to fact to estimate the distribution of the sample mean.

Def: Consider two independent random variables Y and Z, where $Z \sim N(0,1)$ and $Y \sim \chi^2(m)$. Define a random variable T by

$$T = \frac{Z}{\sqrt{\frac{Y}{m}}}$$

Then the distribution of T is called the Student's t-distribution with m degrees of freedom.

A worthy aside on the t-distribution. The Student-t distribution was introduced in 1908 by William Gosset, a Master Brewer at Guinness, while working on quality control for beer. He was required by the company to publish his work under a pseudonym, and he chose the name Student.

Thm: The pdf of the t distribution with n degrees of freedom is

$$f_t(x) = \frac{\Gamma\left((m+1)/2\right)}{\sqrt{m\pi}\,\Gamma(m/2)}\left(1+\frac{x^2}{m}\right)^{-(m+1)/2}$$

Proof. Since Y, Z are independent by assumption, their joint PDF factors as the product of their marginal PDFs:

$$f_{Y,Z}(y,z) = f_Y(y) f_Z(z)$$

Let $W = Y$ and note that $Z = T\sqrt{\frac{W}{m}}$. Consider the transformation from (T, W) to (Z, Y), which has Jacobian determinant

$$\frac{\partial(z,y)}{\partial(t,w)} = \det\begin{pmatrix} \frac{\partial z}{\partial t} & \frac{\partial z}{\partial w} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial w}\end{pmatrix} = \det\begin{pmatrix} \sqrt{\frac{w}{m}} & \frac{T}{2m}\sqrt{\frac{1}{Wm}} \\ 0 & 1\end{pmatrix} = \sqrt{\frac{w}{m}}$$

By the change of variables formula,

$$\begin{aligned} f_{T,W}(t,w) =& f_{Z,Y}(z,y)\left|\frac{\partial(z,y)}{\partial(t,w)}\right| \\ =& f_Y(w) f_Z\left(t\sqrt{\frac{w}{m}}\right)\sqrt{\frac{w}{m}} \\ =& c w^{(m+1)/2-1}\exp\left\{-\frac{1}{2}\left(1+\frac{t^2}{m}\right)w\right\} \qquad \text{where } c = \left[2^{(m+1)/2}\sqrt{m\pi}\,\Gamma(m/2)\right]^{-1} \end{aligned}$$

To obtain the marginal PDF of T, we integrate out w from the joint PDF:

$$\begin{aligned} f_T(t) = \int_0^\infty f_{T,W}(t,w)\,dw = & \int_0^\infty w^{(m+1)/2-1}\exp\left\{-\frac{1}{2}\left(1+\frac{t^2}{m}\right)w\right\}dw \\ =& \int_0^\infty w^{(m+1)/2} e^{-\frac{1}{2}\beta w}\frac{1}{w}\,dw \qquad \text{where } \beta = \left(1+\frac{t^2}{m}\right) \\ =& c\frac{\Gamma((m+1)/2)}{\beta^{(m+1)/2}} \end{aligned}$$

□

Cor: The first $m-1$ moments of the t-distribution with m degrees of freedom exist. All higher moments do not exist.

While the t distribution is symmetric and bell-shaped, we say that it has heavier tails than the Normal distribution.

Thm: When $m = 1$, the t-distribution is the Cauchy distribution. As $m \to \infty$, the $t(m)$ distribution approaches the standard Normal distribution.

Proof. Recall that a variable has Cauchy distribution if it can be expressed as the ratio of two standard Normal variables. When $m = 1$, the variable $\sqrt{Y} \sim |Z_1|$, where $Z_1 \sim N(0,1)$. And so $T \sim Z/|Z_1|$. But T is symmetric, and so actually,

$$Z \sim \frac{Z}{Z_1}.$$

The second result follows from the Strong Law of Large Numbers. Consider a sequence of standard Normal random variables $Z_1, Z_2, \ldots$ and let

$$Y_m = Z_1^2 + \cdots + Z_m^2$$

By the SLLN, $\frac{Y_m}{m} \to E[Z_1^2] = 1$ with probability 1. Let $Z \sim N(0,1)$ independent of all of the Z_i, and let

$$T_m = \frac{Z}{\sqrt{\frac{Y_n}{m}}}$$

Then $T_m \sim t(m)$ by definition, and as the denominator converges to 1 with probability 1, then T_n converges to Z in distribution. □

What is the ultimate relationship between the t-distribution and samples from a Normal population?

Thm: Suppose $X_1, \ldots, X_n$ form a random sample from $N(\mu, \sigma^2)$. Let $\bar{X}$ denote the sample mean and

$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}$$

Then

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

has the t-distribution with $n-1$ degrees of freedom.

Proof. Observe

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \frac{1}{\frac{S}{\sigma}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \frac{1}{\sqrt{\frac{1}{n-1} \sum \left(\frac{X_i - \bar{X}}{\sigma}\right)^2}}$$

And

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \qquad \sum \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1).$$

□

One important feature of the preceding theorem is that neither the estimates from T nor the sampling distribution of T depend on the value of the variance σ^2.

Ex 1: A batch of stout beer is best when it has an original gravity (OG) close to 1.071. The particular OG of a batch depends on a number factors (like temperature, rest time, recipe, etc.) but is (approximately) Normally distributed. Suppose we sample 5 OG measurements from a batch of beer:

1.067 1.060 1.077 1.072 1.067 with $\bar{x} = 1.0686$ and $s = 0.0064$

What is the probability of obtaining a sample at least as extreme as this one, if the batch truly had an OG of 1.071? How would the answer differ if I **knew** $\sigma = 0.0064$?