Statistics 351 (Fall 2009)
The t-Test for Independent Normal Random Variables

Our goal is to explain the t-test from first-year statistics.

Theorem. *Let $Y_1, Y_2, \dots, Y_n$ be independent and identically distributed $\mathcal{N}(\mu, \sigma^2)$ random variables, and suppose that*

$$\overline{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2$$

denote the sample mean and sample variance, respectively. If we define the random variable

$$T = \frac{\overline{Y} - \mu}{S/\sqrt{n}},$$

then $T \in t(n-1)$; that is, T has a t-distribution with $n-1$ degrees of freedom.

The main step in the proof of this theorem is the independence of $\overline{Y}$ and S^2 established last lecture. However, there are a number of other preliminary results that will also be needed.

Definition. For $m = 1, 2, 3, \dots$, we say that a random variable X has a *t-distribution with m degrees of freedom* if the density function of X is

$$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi m}\,\Gamma\left(\frac{m}{2}\right)}\left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \quad -\infty < x < \infty.$$

Definition. For $m = 1, 2, 3, \dots$, we say that a random variable X has a *chi-squared distribution with m degrees of freedom* if the density function of X is

$$f_X(x) = \frac{2^{-m/2}}{\Gamma(m/2)} x^{\frac{m}{2}-1} e^{-x/2}, \quad x > 0.$$

In other words, $X \in \chi^2(m)$ if and only if $X \in \Gamma(m/2, 2)$.

Remark. Observe that $\chi^2(2) = \Gamma(1, 2) = \text{Exp}(2)$.

Example. Show that if $Z \in \mathcal{N}(0, 1)$, then $Z^2 \in \chi^2(1)$.

Solution. Suppose that $X = Z^2$. For $x > 0$, the distribution function of X is

$$\begin{aligned} F_X(x) = P\{X \le x\} &= P\{Z^2 \le x\} \\ &= P\{-\sqrt{x} \le Z \le \sqrt{x}\} \\ &= P\{Z \le \sqrt{x}\} - P\{Z \le -\sqrt{x}\} \\ &= \frac{1}{\sqrt{2\pi}}\int_0^{\sqrt{x}} \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z - \frac{1}{\sqrt{2\pi}}\int_0^{-\sqrt{x}} \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z \end{aligned}$$

so that the density of X is

$$f_X(x) = F_X'(x) = \frac{1}{\sqrt{2\pi}}e^{-x/2} \cdot \frac{1}{2\sqrt{x}} - \frac{1}{\sqrt{2\pi}}e^{-x/2} \cdot \left(-\frac{1}{2\sqrt{x}}\right) = \frac{1}{\sqrt{2\pi}}x^{-1/2}e^{-x/2}, \quad x > 0.$$

Since $\Gamma(1/2) = \sqrt{\pi}$, we recognize the density of X as the density of a $\chi^2(1)$ random variable. That is, $Z^2 \in \chi^2(1)$ as required.

Example. If $X_1 \in \Gamma(p_1, a)$ and $X_2 \in \Gamma(p_2, a)$ are independent, show $X_1 + X_2 \in \Gamma(p_1 + p_2, a)$.

Solution. We actually did this in Lecture #6 using the transformation theorem (Theorem 1.2.1); see Problem #39 of Chapter 1 as well. The easier way to verify this is to use moment generating functions. Recall that the moment generating function of a sum of independent random variables is the product of the individual moment generating functions. That is,

$$m_{X_1+X_2}(t) = \mathbb{E}(e^{t(X_1+X_2)}) = \mathbb{E}(e^{tX_1}) \cdot \mathbb{E}(e^{tX_2}) = m_{X_1}(t) \cdot m_{X_2}(t).$$

As shown on page 67, the moment generating function of $X \in \Gamma(p, a)$ is

$$m_X(t) = \frac{1}{(1-at)^p} \quad \text{for } t < \frac{1}{a}.$$

Hence,

$$m_{X_1+X_2}(t) = m_{X_1}(t) \cdot m_{X_2}(t) = \frac{1}{(1-at)^{p_1}} \cdot \frac{1}{(1-at)^{p_2}} = \frac{1}{(1-at)^{p_1+p_2}}$$

for $t < 1/a$ so that $X_1 + X_2 \in \Gamma(p_1 + p_2, a)$ as required.

Example. In particular, combining the last two examples yields the following fact. If $Z_1, \ldots, Z_n$ are independent and identically distributed $\mathcal{N}(0, 1)$ random variables, then

$$Z_1^2 + \cdots + Z_n^2 \in \chi^2(n).$$

Example. Suppose that $Y_1, \ldots, Y_n$ are independent random variables with $Y_j \in \mathcal{N}(\mu_j, \sigma_j^2)$ for $j = 1, \ldots, n$. Normalizing implies

$$Z_j = \frac{Y_j - \mu_j}{\sigma_j} \in \mathcal{N}(0, 1)$$

so that we conclude

$$\sum_{j=1}^{n} \left(\frac{Y_j - \mu_j}{\sigma_j}\right)^2 \in \chi^2(n).$$

In particular, if $Y_1, \ldots, Y_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$, then

$$\frac{1}{\sigma^2} \sum_{j=1}^{n} (Y_j - \mu)^2 \in \chi^2(n). \qquad (*)$$

Example. If $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed $\mathcal{N}(\mu, \sigma^2)$, show that

$$\overline{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j \in \mathcal{N}(\mu, \sigma^2/n).$$

Solution. This can be shown using moment generating functions. That is, recall that if $Y \in \mathcal{N}(\mu, \sigma^2)$, then

$$m_Y(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}.$$

Since the moment generating function of a sum of independent random variables is the product of the individual moment generating functions, we conclude

$$m_{\overline{Y}}(t) = \prod_{j=1}^{n} m_{Y_j}(t/n) = \exp\left\{\sum_{j=1}^{n}\left(\mu\frac{t}{n} + \frac{\sigma^2 t^2}{2n^2}\right)\right\} = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2n}\right\}$$

which we recognize as the moment generating function of a $\mathcal{N}(\mu, \sigma^2/n)$ random variable.

Example. Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed $\mathcal{N}(\mu, \sigma^2)$ random variables, and let

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2$$

be the sample variance. We write

$$(Y_j - \overline{Y})^2 = (Y_j - \mu + \mu - \overline{Y})^2 = (Y_j - \mu)^2 + (\overline{Y} - \mu)^2 - 2(Y_j - \mu)(\overline{Y} - \mu)$$

and observe that

$$\sum_{j=1}^{n}(Y_j - \mu)(\overline{Y} - \mu) = (\overline{Y} - \mu)\sum_{j=1}^{n}(Y_j - \mu) = (\overline{Y} - \mu)(n\overline{Y} - n\mu) = n(\overline{Y} - \mu)^2$$

which gives

$$\sum_{j=1}^{n}(Y_j - \overline{Y})^2 = \sum_{j=1}^{n}(Y_j - \mu)^2 + \sum_{j=1}^{n}(\overline{Y} - \mu)^2 - 2n(\overline{Y} - \mu)^2 = \sum_{j=1}^{n}(Y_j - \mu)^2 - n(\overline{Y} - \mu)^2.$$

We now write

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \mu)^2 - \frac{n}{\sigma^2}(\overline{Y} - \mu)^2,$$

or equivalently,

$$\frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \mu)^2 = \frac{(n-1)S^2}{\sigma^2} + \frac{n}{\sigma^2}(\overline{Y} - \mu)^2.$$

Let

$$U = \frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \mu)^2, \quad U_1 = \frac{(n-1)S^2}{\sigma^2}, \quad U_2 = \frac{n}{\sigma^2}(\overline{Y} - \mu)^2$$

so that $U = U_1 + U_2$, and observe from $(*)$ that

$$U = \frac{1}{\sigma^2} \sum_{j=1}^{n} (Y_j - \mu)^2 \in \chi^2(n).$$

We also observe that

$$U_2 = \frac{n}{\sigma^2}(\overline{Y} - \mu)^2 = \left(\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \right)^2 \in \chi^2(1)$$

since

$$\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \in \mathcal{N}(0,1).$$

Since $\overline{Y}$ and S^2 are independent, we conclude that U_1 and U_2 are independent. Thus, using the fact that the moment generating function of a sum of independent random variables is the product of the individual moment generating functions, we see that $\psi_U(t) = \psi_{U_1}(t) \cdot \psi_{U_2}(t)$ and so using the facts that $U \in \chi^2(n) = \Gamma(n/2, 2)$ and $U_2 \in \chi^2(1) = \Gamma(1/2, 2)$, we conclude

$$\psi_{U_1}(t) = \frac{\psi_U(t)}{\psi_{U_2}(t)} = \frac{\frac{1}{(1-2t)^{n/2}}}{\frac{1}{(1-2t)^{1/2}}} = \frac{1}{(1-2t)^{(n-1)/2}} \quad \text{for} \quad t < \frac{1}{2}$$

That is, $U_1 \in \Gamma((n-1)/2, 2) = \chi^2(n-1)$ or, in other words,

$$\frac{(n-1)S^2}{\sigma^2} \in \chi^2(n-1).$$

Example. Show that if $Z \in \mathcal{N}(0,1)$ and $Y \in \chi^2(m)$ are independent random variables, then

$$\frac{Z}{\sqrt{Y/m}} \in t(m).$$

Solution. This was actually Problem #9 of Chapter 1 and given on Assignment #3.

We can finally prove our desired theorem and establish the t-test.

Proof. The fact that $\overline{Y}$ and S^2 are independent implies that

$$Z = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \in \mathcal{N}(0,1)$$

and

$$Y = \frac{(n-1)S^2}{\sigma^2} \in \chi^2(n-1)$$

are also independent. Thus, by the previous example,

$$\frac{Z}{\sqrt{Y/(n-1)}} = \frac{\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\overline{Y} - \mu}{S/\sqrt{n}} \in t(n-1)$$

and the proof is complete. □