Chapter 6: Distributions Derived from the Normal Distribution

November 4, 2009

1 χ^2, t and F Distributions

Definition 1.1 *If Z is a standard normal random variable, the distribution of $U = Z^2$ is called the chi-square distribution with 1 degree of freedom.*

If $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim N(0, 1)$, and therefore $[(X - \mu)/\sigma]^2 \sim \chi_1^2$.

Definition 1.2 *If $U_1, U_2, \ldots, U_n$ are independent chi-square random variables with 1 degree of freedom, the distribution of $Y = U_1 + U_2 + \cdots + U_n$ is called the chi-square distribution with n degrees of freedom and is denoted by χ_n^2.*

The density of χ_n^2 is

$$f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{\frac{n}{2}-1} e^{-\frac{v}{2}},$$

which is a Gamma density with parameters $\alpha = n/2$ and $\lambda = 1/2$. Its moment-generating function is

$$\psi(t) = (1 - 2t)^{-n/2}.$$

A notable property of χ^2-distribution is that if U and V are independent and $U \sim \chi_n^2$ and $V \sim \chi_m^2$, then $U + V \sim \chi_{m+n}^2$.

Definition 1.3 *If $Z \sim N(0,1)$ and $U \sim \chi_n^2$ and Z and U are independent, then the distribution of $Z/\sqrt{U/n}$ is called the t distribution with n degrees of freedom.*

Proposition 1.1 *The density function of the t-distribution with n degrees of freedom is*

$$f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}.$$

As the number of degrees of freedom approaches infinity, the t-distribution tends to the standard normal distribution; in fact, for more than 20 or 30 degrees of freedom, the distributions are very close.

Definition 1.4 *Let U and V be independent chi-square random variables with m and n degrees of freedom, respectively. The distribution of*

$$W = \frac{U/m}{V/n}$$

is called the F-distribution with m and n degrees of freedom and is denoted by $F_{m,n}$.

Proposition 1.2 *The density function of W is given by*

$$f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \left(1 + \frac{m}{n}w\right)^{-(m+n)/2}, \quad w \geq 0.$$

From the definition of the t and F distributions, it follows that the square of a t_n random variable follows an $F_{1,n}$ distribution.

2 The sample mean and the sample variance

Theorem 2.1 *Let* $X_1, \ldots, X_n$ *be a random sample from a* $N(\mu, \sigma^2)$ *distribution, and let* $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$ *and* $S^2 = [1/(n-1)]\sum_{i=1}^{n}(X_i - \bar{X})^2$*. Then*

(a) $\bar{X}$ *and* S^2 *are independent random variables.*

(b) $\bar{X}$ *has a* $N(\mu, \sigma^2/n)$ *distribution.*

(c) $(n-1)S^2/\sigma^2$ *has a chi-squared distribution with* $n-1$ *degrees of freedom.*

PROOF: Without loss of generality, we assume that $\mu = 0$ and $\sigma = 1$. Parts (a) and (c) are proved as follows.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}[(X_1 - \bar{X})^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2)]$$

$$= \frac{1}{n-1}[\Big(\sum_{i=2}^{n}(X_i - \bar{X})\Big)^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2)]$$

The last equality follows from the fact $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$. Thus, S^2 can be written as a function only of $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$. We will now show that these random variables are independent of $\bar{X}$. The joint pdf of the sample $X_1, \dots, X_n$ is given by

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-(1/2)\sum_{i=1}^{n} x_i^2}, \quad -\infty < x_i < \infty.$$

Make the transformation

$$\begin{aligned} y_1 &= \bar{x}, \\ y_2 &= x_2 - \bar{x}, \\ &\vdots \\ y_n &= x_n - \bar{x}. \end{aligned}$$

This is a linear transformation with a Jacobian equal to $1/n$. We have

$$\begin{aligned} f(y_1, \dots, y_n) &= \frac{n}{(2\pi)^{n/2}} e^{-(1/2)(y_1 - \sum_{i=2}^{n} y_i)^2} e^{-(1/2)\sum_{i=2}^{n}(y_i + y_1)^2}, \quad -\infty < y_i < \infty \\ &= \left[\left(\frac{n}{2\pi}\right)^{1/2} e^{(-n y_1^2)/2} \right] \left[\frac{n^{1/2}}{(2\pi)^{(n-1)/2}} e^{-(1/2)[\sum_{i=2}^{n} y_i^2 + (\sum_{i=2}^{n} y_i)^2]} \right], \quad -\infty < y \end{aligned}$$

Hence, Y_1 is independent of $Y_2, \dots, Y_n$, and $\bar{X}$ is independent of S^2.

Since

$$\bar{x}_{n+1} = \frac{\sum_{i=1}^{n+1} x_i}{n+1} = \frac{x_{n+1} + n\bar{x}_n}{n+1} = \bar{x}_n + \frac{1}{n+1}(x_{n+1} - \bar{x}_n),$$

we have

$$\begin{aligned}
nS_{n+1}^2 &= \sum_{i=1}^{n+1}(x_i - \bar{x}_{n+1})^2 = \sum_{i=1}^{n+1}[(x_i - \bar{x}_n) - \frac{1}{n+1}(x_{n+1} - \bar{x}_n)]^2 \\
&= \sum_{i=1}^{n+1}[(x_i - \bar{x}_n)^2 - 2(x_i - \bar{x}_n)(\frac{x_{n+1} - \bar{x}_n}{n+1}) + \frac{1}{(n+1)^2}(x_{n+1} - \bar{x}_n)^2] \\
&= \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + (x_{n+1} - \bar{x}_n)^2 - 2\frac{(x_{n+1} - \bar{x}_n)^2}{n+1} + \frac{(n+1)}{(n+1)^2}(x_{n+1} - \bar{x}_n)^2 \\
&= (n-1)S^2 + \frac{n}{n+1}(x_{n+1} - \bar{x}_n)^2.
\end{aligned}$$

Now consider $n = 2$, $S_2^2 = \frac{1}{2}(X_2 - X_1)^2$. Since $(X_2 - X_1)/\sqrt{2} \sim N(0,1)$, part (a) of Lemma **??** shows that $S_2^2 \sim \chi_1^2$. Proceeding with the induction, we assume that for $n = k$, $(k-1)S_k^2 \sim \chi_{k-1}^2$. For $n = k+1$, we have

$$kS_{k+1}^2 = (k-1)S_k^2 + \frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2.$$

Since S_k^2 is independent of X_{k+1} and $\bar{X}_k$, and $X_{k+1}-\bar{X}_k \sim N(0, \frac{k+1}{k})$, $kS_{k+1}^2 \sim \chi_k^2$. $\square$

Corollary 2.1 *let* $\bar{X}$ *and* S^2 *be as given in Theorem 2.1. Then*

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$