5.3.2 The Derived Distributions: Student's t and Snedecor's F

Definition Let $X_1, \dots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. The quantity $(\bar{X} - \mu)/(S/\sqrt{n})$ has Student's t distribution with $n-1$ degrees of freedom. Equivalently, a random variable T has Student's t distribution with p degrees of freedom, and we write $T \sim t_p$ if it has pdf

$$f_T(t) = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1+t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty.$$

Notice that if $p = 1$, then $f_T(t)$ becomes the pdf of the Cauchy distribution, which occurs for samples of size 2.

The derivation of the t pdf is straightforward. Let $U \sim N(0,1)$, and $V \sim \chi^2_p$. If they are independent, the joint pdf is

$$f_{U,V}(u,v) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \frac{1}{\Gamma(p/2)2^{p/2}} v^{\frac{p}{2}-1} e^{-v/2}, \quad -\infty < u < \infty, \quad 0 < v < \infty.$$

Make the transformation

$$t = \frac{u}{\sqrt{v/p}}, \quad w = v,$$

and integrate out w, we can get the marginal pdf of t.

Student's t has no mgf because it does not have moments of all orders. In fact, if there are p degrees of freedom, then there are only $p-1$ moments. It is easy to check that

$$ET_p = 0, \quad \text{if } p > 1,$$

$$\mathrm{Var}T_p = \frac{p}{p-2}, \quad \text{if } p > 2.$$

Example Let $X_1, \dots, X_n$ be a random sample from $N(\mu_X, \sigma^2_X)$ population, and let $Y_1, \dots, Y_m$ be a random sample from an independent $N(\mu_Y, \sigma^2_Y)$ population. If we were interested in comparing the variability of the populations, one quantity of interest would be the ratio σ^2_X/σ^2_Y. Information about this ratio is contained in S^2_X/S^2_Y, the ratio of sample variances. The F distribution allows us to compare these quantities by giving us a distribution of

$$\frac{S^2_X/S^2_Y}{\sigma^2_X/\sigma^2_Y} = \frac{S^2_X/\sigma^2_X}{S^2_Y/\sigma^2_Y}.$$

Definition Let $X_1, \ldots, X_n$ be a random sample from $N(\mu_X, \sigma_X^2)$ population, and let $Y_1, \ldots, Y_m$ be a random sample from an independent $N(\mu_Y, \sigma_Y^2)$ population. The random variable $F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$ has Snedecor's F distribution with $n-1$ and $m-1$ degrees of freedom. Equivalently, the random variable F has the F distribution with p and q degrees of freedom if it has pdf

$$f_F(x) = \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})}\left(\frac{p}{q}\right)^{p/2}\frac{x^{(p/2)-1}}{[1+(p/q)x]^{(p+q)/2}}, \quad 0 < x < \infty.$$

A variance ratio may have an F distribution even if the parent populations are not normal. Kelker (1970) has shown that as long as the parent populations have a certain type of symmetric, then the variance ratio will have an F distribution.

Example To see how the F distribution may be used for inference about the true ratio of population variances, consider the following. The quantity $\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$ has an $F_{n-1,m-1}$ distribution. We can calculate

$$\begin{aligned} EF_{n-1,m-1} &= E\left(\frac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)}\right) \\ &= E(\chi_{n-1}^2/(n-1))E((m-1)/(\chi_{m-1}^2)) = (m-1)/(m-3). \end{aligned}$$

Note this last expression is finite and positive only if $m > 3$. Removing expectations, we have for reasonably large m,

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \approx \frac{m-1}{m-3} \approx 1,$$

as we might expect.

The F distribution has many interesting properties and is related to a number of other distributions.

Theorem 5.3.8

a. If $X \sim F_{p,q}$, then $1/X \sim F_{q,p}$; that is, the reciprocal of an F random variable is again an F random variable.

b. If $X \sim t_q$, then $X^2 \sim F_{1,q}$.

c. If $X \sim F_{p,q}$, then $(p/q)X/(1+(p/q)X) \sim beta(p/2, q/2)$.