Distributions of Functions of Normal Random Variables

The Unit (or Standard) Normal

The **unit** or **standard normal** random variable U is a normally distributed variable with mean zero and variance one, i.e. $U \sim \mathrm{N}(0,1)$. Note that if $x \sim \mathrm{N}(\mu,\sigma^2)$ that

$$\frac{x-\mu}{\sigma} \sim U \sim \mathrm{N}(0,1)$$

Thus, to simulate a normal random variable with mean μ and variance σ^2, we can simply transform unit normals, as

$$x \sim \mu + \sigma\, U$$

If $x \sim \mathrm{N}(\mu,\sigma^2)$, then $\overline{x} \sim \mathrm{N}(\,\mu, \sigma^2/n\,)$. Thus

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim U \sim \mathrm{N}(0,1) \tag{A5.1}$$

Central and Noncentral χ^2 Distributions

The χ^2 distribution arises from sums of squared, normally distributed, random variables — if $x_i \sim \mathrm{N}(0,1)$, then $u = \sum_{i=1}^{n} x_i^2 \sim \chi^2_n$, a **central** χ^2 distribution with n degrees of freedom. It follows that the sum of two χ^2 random variables is also χ^2 distributed, so that if $u \sim \chi^2_n$ and $v \sim \chi^2_m$, then

$$u + v \sim \chi^2_{(n+m)} \tag{A5.14a}$$

Two other useful results are that if $x_i \sim \mathrm{N}(0,\sigma^2)$, then

$$\sum_{i=1}^{n} x_i^2 \sim \sigma^2 \cdot \chi^2_n \tag{A5.14b}$$

and for $\overline{x} = n^{-1}\sum_{i=1}^{n} x_i$,

$$\sum_{i=1}^{n} (x_i - \overline{x}\,)^2 \sim \sigma^2 \cdot \chi^2_{(n-1)} \tag{A5.14c}$$

In this last case, subtraction of the mean causes the loss of one degree of freedom. The χ^2 distribution is thus vital for both hypothesis testing and construction of confidence intervals of unknown variances, as Equation A5.12c implies that if $x \sim \mathrm{N}(0, \sigma_o^2)$ then for the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\, x_i - \overline{x}\,)^2, \qquad \text{we have} \qquad \frac{(n-1)\, S^2}{\sigma_o^2} \sim \chi^2_{n-1}$$

A **noncentral** χ^2 arises when the random variables being considered have nonzero means. In particular, if $x_i \sim \mathrm{N}(\mu_i, 1)$, then $u = \sum_{i=1}^{n} x_i^2$ follows a noncentral χ^2 distribution with n degrees of freedom and **noncentrality parameter**

$$\lambda = \sum_{i=1}^{n} \mu_i^2 \tag{A5.15a}$$

and we write $u \sim \chi^2_{n,\lambda}$. As shown in Figure A5.3, increasing the noncentrality parameter λ shifts the distribution to the right. This is also seen by considering the mean and variance of u,

$$E(u) = n + \lambda \qquad \text{and} \qquad \sigma^2(u) = 2(n + 2\lambda) \tag{A5.15b}$$

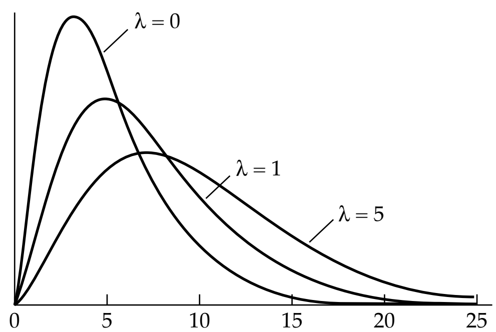

Figure A5.3 The probability distribution function for a noncentral χ^2. As the noncentrality parameter λ increases, the distribution is pulled to the right. We plot here a χ^2 random variable with n = 5 degrees of freedom and noncentrality parameters $\lambda = 0$ (a central χ^2), 1, and 5.

It follows directly from the definition that sums of noncentral χ^2 variables also follows a noncentral χ^2 distribution, so that if $u \sim \chi^2_{n,\,\lambda_1}$ and $v \sim \chi^2_{m,\,\lambda_2}$, then

$$(u + v) \sim \chi^2_{(n+m),(\lambda_1+\lambda_2)} \tag{A5.15c}$$

Finally, Equations A5.14b,c can be generalized to noncentral χ^2 random variables as follows. Suppose $x_i \sim \mathrm{N}(\mu_i, \sigma^2)$, then

$$\sum_{i=1}^{n} x_i^2 \sim \sigma^2 \cdot \chi^2_{n,\lambda} \qquad \text{where} \qquad \lambda = \sum_{i=1}^{n} \mu_i^2 \tag{A5.15d}$$

and

$$\sum_{i=1}^{n} (\, x_i - \overline{x} \,)^2 \sim \sigma^2 \cdot \chi^2_{(n-1),\lambda} \qquad \text{where} \qquad \lambda = \sum_{i=1}^{n} \frac{\mu_i^2}{\sigma^2} \tag{A5.15e}$$

Student's *t* Distribution

If $x \sim \mathrm{N}(\mu, \sigma^2)$, then (Equation A5.1) $(\, \overline{x} - \mu \,)/(\, \sigma/\sqrt{n} \,) \sim \mathrm{N}(0, 1)$, which allows for both hypothesis testing and construction of confidence intervals when σ^2 is know. When the variance is unknown, the above test statistic replaces the true (but unknown) variance σ^2 with the sample variance S^2,

$$t = \frac{\overline{z} - \mu}{S/\sqrt{n}} \qquad \text{where} \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\, z_i - \overline{z} \,)^2$$

Notice that

$$t = \left(\frac{\overline{z} - \mu}{\sigma/\sqrt{n}} \right) \left(\frac{1}{\sqrt{S^2/\sigma^2}} \right) = \frac{U}{\sqrt{\chi^2_{n-1}/(n-1)}}$$

Thus we define a t **random variable with** ν **degrees of freedom** by

$$t_\nu \sim \frac{U}{\sqrt{\chi^2_\nu/\nu}}$$

This distribution has mean zero and variance $\sigma^2(t_\nu) = 1 + 2/(\nu - 2)$ (for $\nu > 2$). The coefficient of kurtosis is $k_4 = 6/(\nu - 4)$, implying that the t distribution has heavier tails than a normal. The **noncentral Student's t- distribution** is defined as follows. If $x \sim \mathrm{N}(\mu_o, \sigma^2)$, but we assume the correct mean in μ, then

$$t_{\nu,\lambda} = \frac{\overline{z} - \mu}{S/\sqrt{n}}$$

is distirbution an a noncentral t with μ degrees of freedom and noncentrality parameter $\lambda = (\mu - \mu_o)/\sigma$.

Central and Noncentral *F* Distributions

The ratio of two χ^2-distributed variables leads to the F distribution. In particular, if $u \sim \chi^2_n$ and $v \sim \chi^2_m$, then the ratio of these two χ^2 variables divided by their respective degrees of freedom follows a **central *F* distribution** with **numerator** and **denominator degrees of freedom** n and m (respectively), i.e., $(u/n)/(v/m) \sim F_{n,m}$. Since

$$\lim_{m \to \infty} F_{n,m} \to \frac{\chi^2_n}{n}$$

the F distribution can be approximated by a χ^2_n when the denominator degrees of freedom is large. The F distribution was introduced by Snedecor. Notice that the square of the t distribution is

$$t^2_\nu \sim \frac{U^2}{\chi^2_\nu/\nu} = \frac{\chi^2_1/1}{\chi^2_\nu/\nu} \sim F_{1,\nu}$$

Thus the square of a t distribution with n degrees of freedom is distributed as an F with 1 and n degrees of freedom.

The **noncentral *F* distribution** results when the numerator χ^2 variable is noncentral. If $u \sim \chi^2_{n,\lambda}$ and $v \sim \chi^2_m$, then $F = (u/n)/(v/m)$ follows a noncentral F distributed with noncentrality parameter λ, and we write $F \sim F_{n,m,\lambda}$. As with the noncentral χ^2, increasing λ shifts the distribution further to the right. Again, this is seen in the mean and variance, with

$$E(F) = \frac{m}{m-2}\left(1 + \frac{2\lambda}{n}\right) \tag{A5.16a}$$

$$\sigma^2(F) = 2\left(\frac{m}{n}\right)^2 \left[\frac{(n+m)^2 + (n+2\lambda)(m-2)}{(m-2)^2(m-4)}\right] \tag{A5.16b}$$

Various mathematical and statistical packages provide routines for computing cumulative probabilities of noncentral χ^2 and F random variables, and a number of approximations have been suggested (e.g., Patnaik 1949, Severo and Zelen 1960, Tiku 1965). Winer et al. (1991) offer one such approximation based on the unit normal U, with the probability that a noncentral F-distributed random variable exceeds a value F_o being approximately

$$\Pr(\, F_{n,m,\lambda} > F_o\,) \simeq \Pr(\, U > z_o\,) \tag{A5.17a}$$

where

$$z_o = \frac{\sqrt{(2m-1)\,B} - \sqrt{2(n+\lambda) - A}}{\sqrt{A+B}}, \qquad A = \frac{n+2\lambda}{n+\lambda}, \quad B = \frac{n}{m}\,F_o \tag{A5.17b}$$

From this general expression follow simplified approximations for the special cases of central F and noncentral χ^2 variables. Setting $\lambda = 0$ gives an approximation for the central F distribution as

$$\Pr(F_{n,m} > F_o) \simeq \Pr(\, U > \widetilde{z}_o\,), \quad \text{with} \quad \widetilde{z}_o = \frac{\sqrt{(2m-1)\,B} - \sqrt{2n-1}}{\sqrt{1+B}} \tag{A5.17c}$$

Likewise, taking the limit as $m \to \infty$ offers an approximation for the noncentral χ^2, since $\chi^2_{n,\lambda} \sim n \cdot F_{n,\infty,\lambda}$. Taking the limit of Equation A5.17c as $m \to \infty$ gives the probability that a noncentral χ^2 exceeds a value C_o as approximately

$$\Pr(\, \chi^2_{n,\lambda} > C_o\,) \simeq \Pr(\, U > \widetilde{z}_o\,), \quad \text{with} \quad \widetilde{z}_o = \frac{\sqrt{2C_o} - \sqrt{2(n+\lambda) - A}}{\sqrt{A}} \tag{A5.17d}$$