

Distributions of Functions of

Normal Random Variables

Version 25 Jan 2006

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. $U \sim N(0, 1)$. Note that if $x \sim N(\mu, \sigma^2)$ that

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

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

$$x \sim \mu + \sigma U \sim N(\mu, \sigma^2) \quad (2)$$

Consider n independent random variables $x_i \sim N(\mu, \sigma^2)$, then $\bar{x} \sim N(\mu, \sigma^2/n)$, and thus

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim U \sim N(0, 1) \quad (3)$$



Example 1. Let's construct a 95% confidence interval for the mean μ for Equation (3). First, let's use **R** to compute a value $U_{0.975}$ such that $\Pr(U \leq U_{0.975}) = 0.975$. In **R**, typing the command `qnorm(0.975)` returns **1.96**. Likewise, `qnorm(0.025)` returns **-1.96** and hence $\Pr(U \leq -1.96) = 0.025$. Hence,

$$\Pr(-1.96 \leq U \leq 1.96) = 0.95$$

Recalling Equation (3),

$$\Pr(-1.96 \leq U \leq 1.96) = \Pr\left(-1.96 \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq 1.96\right)$$

Rearranging gives

$$\Pr(-1.96\sigma/\sqrt{n} \leq \bar{x} - \mu \leq 1.96\sigma/\sqrt{n})$$

or

$$\Pr(-\bar{x} - 1.96\sigma/\sqrt{n} \leq -\mu \leq -\bar{x} + 1.96\sigma/\sqrt{n})$$

which can also be written as

$$\Pr(\bar{x} - 1.96\sigma/\sqrt{n} \leq \mu \leq \bar{x} + 1.96\sigma/\sqrt{n}) = 0.95$$

giving a 95% confidence interval for the mean μ .

Central and Noncentral χ^2 Distributions

The χ^2 distribution arises from sums of squared, normally distributed, random variables — if $x_i \sim N(0, 1)$, then $u = \sum_{i=1}^n x_i^2 \sim \chi_n^2$, 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 $u \sim \chi_n^2$ and $v \sim \chi_m^2$, then

$$u + v \sim \chi_{(n+m)}^2 \quad (4a)$$

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

$$\sum_{i=1}^n x_i^2 \sim \sigma^2 \cdot \chi_n^2 \quad (4b)$$

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

$$\sum_{i=1}^n (x_i - \bar{x})^2 \sim \sigma^2 \cdot \chi_{(n-1)}^2 \quad (4c)$$

In this last case, subtraction of the mean causes the loss of one degree of freedom. Note that a special case of Equation (4c) is the sample estimate of the variance,

$$\text{Var}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

so that

$$(n-1)\text{Var}(x) \sim \sigma^2 \cdot \chi_{(n-1)}^2, \quad \text{implying} \quad \frac{(n-1)\text{Var}(x)}{\sigma^2} \sim \chi_{(n-1)}^2 \quad (4d)$$

Example 2. We can use Equation (4d) to construct a confidence interval on the true variance σ^2 given the sample variance $\text{Var}(x)$, provided the x_i are drawn from independent normals with the same mean and variance σ^2 .

First, recall that the **R** command `qchisq(p, df)` returns a value X such that $\Pr(\chi_{df}^2 \leq X) = p$. Suppose sample size is $n = 20$. Since `qchisq(0.975, 19)` returns a value of **32.85** and `qchisq(0.025, 19)` returns **8.91**, we have

$$\Pr(8.91 \leq \chi_{19}^2 \leq 32.85) = 0.95$$

From Equation 4d,

$$\Pr(8.91 \leq \chi_{19}^2 \leq 32.85) = \Pr\left(8.91 \leq \frac{(n-1)\text{Var}(x)}{\sigma^2} \leq 32.85\right)$$

Noting that for

$$\Pr(a \leq x \leq b) = \Pr\left(\frac{1}{a} \geq \frac{1}{x} \geq \frac{1}{b}\right)$$

we have

$$\Pr\left(8.91 \leq \frac{19\text{Var}(x)}{\sigma^2} \leq 32.85\right) = \Pr\left(\frac{1}{8.91} \geq \frac{\sigma^2}{19\text{Var}(x)} \geq \frac{1}{32.85}\right)$$

or

$$\Pr\left(\frac{19\text{Var}}{8.91} \geq \sigma^2 \geq \frac{19\text{Var}}{32.85}\right) = 0.95$$

or

$$\Pr(2.13\text{Var} \geq \sigma^2 \geq 0.58\text{Var}) = 0.95$$

giving the 95% confidence interval on the variance as 0.58Var to 2.13Var.

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

$$\lambda = \sum_{i=1}^n \mu_i^2 \tag{5a}$$

and we write $u \sim \chi_{n,\lambda}^2$. As shown in Figure 1, increasing the noncentrality parameter λ shifts the distribution to the right. This is also seen by considering the mean and variance of u ,

$$E(u) = n + \lambda \quad \text{and} \quad \sigma^2(u) = 2(n + 2\lambda) \tag{5b}$$

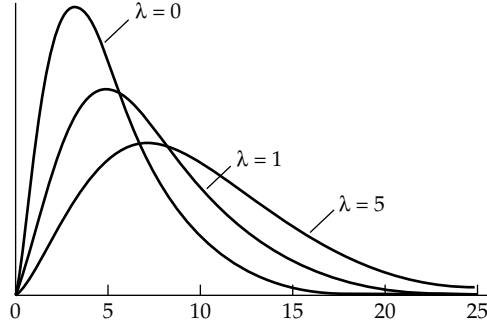


Figure 1 The probability distribution function for a noncentral χ^2 . As the noncentrality parameter λ increases, the distribution is pulled to the right. We plot here a χ^2 random variable with $n = 5$ degrees of freedom and noncentrality parameters $\lambda = 0$ (a central χ^2), 1, and 5.

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

$$(u + v) \sim \chi_{(n+m), (\lambda_1 + \lambda_2)}^2 \quad (5c)$$

Finally, Equations 4b,c can be generalized to noncentral χ^2 random variables as follows. Suppose $x_i \sim N(\mu_i, \sigma^2)$, then

$$\sum_{i=1}^n x_i^2 \sim \sigma^2 \cdot \chi_{n, \lambda}^2 \quad \text{where} \quad \lambda = \sum_{i=1}^n \frac{\mu_i^2}{\sigma^2} \quad (5d)$$

Turning the distribution of $\sum_{i=1}^n (x_i - \bar{x})^2$, defining

$$\lambda^* = \sum_{i=1}^n \frac{(\mu_i - \bar{\mu})^2}{\sigma^2}, \quad \text{where} \quad \bar{\mu} = \frac{1}{n} \sum_{i=1}^n \mu_i$$

then

$$\sum_{i=1}^n (x_i - \bar{x})^2 \sim \sigma^2 \cdot \chi_{n-1, \lambda^*}^2 \quad (5e)$$

Note that if all the x_i have the same mean ($\mu_i = \mu = \bar{\mu}$), $\lambda^* = 0$ and the χ^2 is central, while if there is some variance in the means of the x_i , then distribution is a noncentral χ^2 .

R provides commands for quantities of interest with noncentral χ^2 distributions.

- **qchisq(p, df, ncp)** returns a value X such that $\Pr(\chi_{df, ncp}^2 \leq X) = p$

- `pchisq(x,df,ncp)` returns the probability that $\Pr(\chi_{df,ncp}^2 \leq X)$
- leaving out the field for `ncp` returns these values for a central χ^2 .

Student’s t Distribution

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

$$t = \frac{\bar{x} - \mu}{\sqrt{\text{Var}/n}} \tag{6}$$

Notice that

$$t = \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right) \left(\frac{1}{\sqrt{\text{Var}/\sigma^2}} \right) = \frac{U}{\sqrt{\chi_{n-1}^2/(n-1)}}$$

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

$$t_\nu = \frac{U}{\sqrt{\chi_\nu^2/\nu}} \tag{7a}$$

A *t* random variable has mean zero and variance

$$\sigma^2(t_\nu) = 1 + \frac{2}{\nu - 2} \quad \text{for } \nu > 2 \tag{7b}$$

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 $x \sim N(\mu_o, \sigma^2)$, but we assume the correct mean is μ , then

$$t_{\nu,\lambda} = \frac{\bar{x} - \mu}{\sqrt{\text{Var}/n}}$$

is distributed as a noncentral *t* random variable with $n - 1$ degrees of freedom and noncentrality parameter $\lambda = (\mu - \mu_0)/\sigma$.

Central and Noncentral F Distributions

The ratio of two χ^2 -distributed variables leads to the *F* distribution. In particular, if $u \sim \chi_n^2$ and $v \sim \chi_m^2$, 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

$$\lim_{m \rightarrow \infty} F_{n,m} \rightarrow \frac{\chi_n^2}{n}$$

the F distribution can be approximated by a χ_n^2 when the denominator degrees of freedom is large.

R provides commands for quantities of interest for F distributions.

- **qf(p,df1,df2)** returns a value X such that $\Pr(F_{df1,df2} \leq X) = p$
- **pf(x,df1,df2)** returns the probability that $\Pr(F_{df1,df2} \leq X)$

The **noncentral F distribution** results when the numerator χ^2 variable is noncentral. If $u \sim \chi_{n,\lambda}^2$ and $v \sim \chi_m^2$, 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

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

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

R provides commands for quantities of interest for noncentral F distributions.

- **pf(x,df1,df2,ncp)** returns the probability that $\Pr(F_{df1,df2,ncp} \leq X)$
- the obvious command **qf(p,df1,df2,ncp)** does not work, as the same value is returned for all values of **ncp**.