Lecture 13: Noncentral χ^2-, t-, and F-distributions

The results on transformation lead to many useful results based on transformations of normal random variables.

Ratio of two normal random variables

If X_1 and X_2 are independent and both have the normal distribution $N(0,1)$, then, the pdf of $V_1 = X_1/X_2$ is

$$\begin{aligned} f_{V_1}(v_1) &= \int_{-\infty}^{\infty} |v_2| \frac{e^{-(v_1 v_2)^2/2}}{\sqrt{2\pi}} \frac{e^{-v_2^2/2}}{\sqrt{2\pi}} dv_2 \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |v_2| e^{-(1+v_1^2)v_2^2/2} dv_2 \\ &= \frac{1}{\pi} \int_0^{\infty} v_2 e^{-(1+v_1^2)v_2^2/2} dv_2 \\ &= \frac{1}{\pi} \int_0^{\infty} e^{-(1+v_1^2)s} ds \\ &= \frac{1}{\pi(1+v_1^2)} \end{aligned}$$

which is the pdf of $Cauchy(0,1)$.

The next result concerns a ratio of independent chi-squares random variables, or sums of squared independent normal random variables.

Ratio of chi-square random variables and F-distribution

Let X_1 and X_2 be independent random variables having the chi-square distributions with degrees of freedom n_1 and n_2, respectively.
By the transformation theorem, the p.d.f. of $Z = X_1/X_2$ is

$$
\begin{aligned}
f_Z(z) &= \int_0^\infty x_2 \frac{(zx_2)^{n_1/2-1} e^{-zx_2/2}}{2^{n_1/2}\Gamma(n_1/2)} \frac{x_2^{n_2/2-1} e^{-x_2/2}}{2^{n_2/2}\Gamma(n_2/2)} dx_2 \\
&= \frac{z^{n_1/2-1}}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} \int_0^\infty x_2^{(n_1+n_2)/2-1} e^{-(1+z)x_2/2} dx_2 \\
&= \frac{\Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} \frac{z^{n_1/2-1}}{(1+z)^{(n_1+n_2)/2}}, \qquad z > 0
\end{aligned}
$$

where the last equality follows from the fact that

$$
\frac{1}{2^{(n_1+n_2)/2}\Gamma[(n_1+n_2)/2]} x_2^{(n_1+n_2)/2-1} e^{-x_2/2} \qquad x_2 > 0
$$

is the pdf of the chi-square distribution with degrees of freedom $n_1 + n_2$.

Making another transformation $Y = (n_2/n_1)Z = (X_1/n_1)/(X_2/n_2)$ and applying the (univariate) transformation theorem, we obtain that the pdf of Y is

$$\frac{n_1^{n_1/2} n_2^{n_2/2} \Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} \frac{y^{n_1/2-1}}{(n_1 y + n_2)^{(n_1+n_2)/2}}, \qquad y > 0$$

which is the pdf of the well-known F-distribution with degrees of freedom n_1 and n_2.

t-distribution

Let U_1 and U_2 be independent random variables, $U_1 \sim N(0,1)$ and U_2 has the chi-square distribution with degrees of freedom n.

What is the distribution of $T = U_1/\sqrt{U_2/n}$?

Let $X_1 = U_1^2$ and $X_2 = U_2$.
Then X_1 and X_2 are independent, because U_1 and U_2 are independent.
By a result obtained previously, X_1 has the chi-square distribution with degree of freedom 1.
From the previous proof, $Y = X_1/(X_2/n)$ has the F-distribution with degrees of freedom 1 and n.

The pdf of Y is

$$\frac{n^{n/2}\Gamma[(n+1)/2]y^{-1/2}}{\sqrt{\pi}\Gamma(n/2)(n+y)^{(n+1)/2}} \qquad y>0$$

Applying the univariate transformation theorem, we obtain that the pdf of $W=\sqrt{Y}$ is ($dy = 2wdw$ when $y = w^2$)

$$\frac{2n^{n/2}\Gamma[(n+1)/2]}{\sqrt{\pi}\Gamma(n/2)(n+w^2)^{(n+1)/2}} \qquad w>0$$

Note that

$$T = \begin{cases} W & U_1 \geq 0 \\ -W & U_1 < 0. \end{cases}$$

and

$$P(T < -t) = P(T > t). \qquad t > 0.$$

Hence the pdf of T is

$$\frac{n^{n/2}\Gamma[(n+1)/2]}{\sqrt{\pi}\Gamma(n/2)(n+t^2)^{(n+1)/2}} \qquad t \in \mathscr{R}$$

This is the pdf of the well-known t-distribution with degrees of freedom n.

When $n = 1$, the pdf of the t-distribution is

$$\frac{\Gamma(1)}{\sqrt{\pi}\Gamma(1/2)(1+t^2)} = \frac{1}{\pi(1+t^2)} \qquad t \in \mathscr{R}$$

which is the pdf of $Cauchy(0,1)$.

Hence, if $X_1 \sim N(0,1)$ and $X_2 \sim N(0,1)$ are independent, then we have just shown that $X_1/|X_2| \sim Cauchy(0,1)$.

But previously we showed that $X_1/X_2 \sim Cauchy(0,1)$.

If they are both true, then $X_1/X_2 \sim X_1/|X_2|$.

This is in fact true, because for $t \in \mathscr{R}$,

$$\begin{aligned} P(X_1/X_2 < t) &= P(X_1 < tX_2, X_2 > 0) + P(X_1 > tX_2, X_2 < 0) \\ &= P(X_1 < t|X_2|, X_2 > 0) + P(X_1 > -t|X_2|, X_2 < 0) \\ &= P(X_1 < t|X_2|, X_2 > 0) + P(-X_1 < t|X_2|, X_2 < 0) \\ &= P(X_1 < t|X_2|, X_2 > 0) + P(X_1 < t|X_2|, X_2 < 0) \\ &= P(X_1 < t|X_2|) = P(X_1/|X_2| < t) \end{aligned}$$

and the 4th equality holds since $(X_1, X_2) \sim (-X_1, X_2)$.

The previously defined chi-square (χ^2-), t-, and F-distributions previously defined are special cases of the non-central χ^2-, t-, and F-distributions, respectively, which are useful in statistics.

Definition (the noncentral chi-square distribution)

Let $X_1,...,X_n$ be independent random variables and $X_i \sim N(\mu_i, \sigma^2)$, $i = 1,...,n$. The distribution of the random variable $Y = (X_1^2 + \cdots + X_n^2)/\sigma^2$ is called the noncentral chi-square distribution with degrees of freedom n and the noncentrality parameter $\delta = (\mu_1^2 + \cdots + \mu_n^2)/\sigma^2$.

The chi-square distribution defined earlier is a special case of the noncentral chi-square distribution with $\delta = 0$ and, therefore, is sometimes called a central chi-square distribution.

It follows from the definition of noncentral chi-square distributions that if $Y_1,...,Y_k$ are independent random variables and Y_i has the noncentral chi-square distribution with degrees of freedom n_i and the noncentrality parameter δ_i, $i = 1,...,k$, then $Y = Y_1 + \cdots + Y_k$ has the noncentral chi-square distribution with degrees of freedom $n_1 + \cdots + n_k$ and the noncentrality parameter $\delta_1 + \cdots + \delta_k$.

Theorem (properties of the noncentral chi-square distribution)

Let Y be a random variable having the noncentral chi-square distribution with degrees of freedom k and noncentrality parameter δ.

(i) The pdf of Y is

$$g_{\delta,k}(x) = e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} f_{2j+k}(x),$$

where $f_v(x)$ is the pdf of the central chi-square distribution with degrees of freedom v, $v = 1,2,...$;

(ii) The mgf of Y is

$$\frac{e^{\delta t/(1-2t)}}{(1-2t)^{k/2}}, \qquad t < 1/2;$$

(iii) $E(Y) = k + \delta$;

(iv) $\mathrm{Var}(Y) = 2k + 4\delta$.

Proof.

We first prove result (ii).

Let X_k be a random variable having the standard normal distribution and $\mu = \sqrt{\delta}$ be a positive number.

For $t < 1/2$, the mgf of $(X_k + \mu)^2$ is

$$\begin{aligned}\psi_\delta(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} e^{t(x+\mu)^2} dx \\ &= \frac{e^{\mu^2 t/(1-2t)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1-2t)[x-2\mu t/(1-2t)]^2/2} dx \\ &= \frac{e^{\delta t/(1-2t)}}{\sqrt{1-2t}}.\end{aligned}$$

By definition, $Y \sim X_1^2 + \cdots + X_{k-1}^2 + (X_k + \sqrt{\delta})^2$, where X_i's are independent and have the standard normal distribution.

From the obtained result, the mgf of Y is

$$E\left\{e^{t[X_1^2+\cdots+X_{k-1}^2+(X_k+\sqrt{\delta})^2]}\right\} = [\psi_0(t)]^{k-1} \psi_{\sqrt{\delta}}(t) = \frac{e^{\delta t/(1-2t)}}{(1-2t)^{k/2}}.$$

We now use the proved result (ii) to show result (i).

By the uniqueness of the mfg, it suffices to show that the pdf $g_{\delta,k}(x)$ in (i) has mgf exactly the same as the one in (ii).

Note that the mgf of the central chi-square distribution with degrees of freedom v is,

$$\frac{1}{(1-2t)^{v/2}}, \qquad t < 1/2,$$

The mgf of the pdf $g_{\delta,k}(x)$ in (i) is, for $t < 1/2$,

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{tx} g_{\delta,k}(x)dx &= e^{-\delta/2} \int_{-\infty}^{\infty} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} e^{tx} f_{2j+k}(x)dx \\
&= e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \int_{-\infty}^{\infty} e^{tx} f_{2j+k}(x)dx \\
&= e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!(1-2t)^{(j+k/2)}} \\
&= \frac{e^{-\delta/2}}{(1-2t)^{k/2}} \sum_{j=0}^{\infty} \frac{\{\delta/[2(1-2t)]\}^j}{j!}
\end{aligned}$$

$$= \frac{e^{-\delta/2+\delta/[2(1-2t)]}}{(1-2t)^{k/2}} = \frac{e^{\delta t/(1-2t)}}{(1-2t)^{k/2}}$$

To establish (iii), let X_i's be as defined in the proof of (i).

Then,

$$\begin{aligned} E(Y) &= E(X_1^2)+\cdots+E(X_{k-1}^2)+E(X_k+\sqrt{\delta})^2 \\ &= k-1+E(X_k^2)+\delta+E(2\sqrt{\delta}X_k) \\ &= k+\delta \end{aligned}$$

because $E(X_i^2)=1$ and $E(X_i)=0$.

To show (iv), note that $\text{Var}(X_i^2)=2$ and $\text{Cov}(X_k^2,X_k)=E(X_k^3)=0$.

Then,

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(X_1^2)+\cdots+\text{Var}(X_{k-1}^2)+\text{Var}\big((X_k+\sqrt{\delta})^2\big) \\ &= 2(k-1)+\text{Var}(X_k^2+2\sqrt{\delta}X_k) \\ &= 2(k-1)+\text{Var}(X_k^2)+4\delta\text{Var}(X_k)+4\sqrt{\delta}\text{Cov}(X_k^2,X_k) \\ &= 2k+4\delta \end{aligned}$$

Definition (the noncentral t-distribution)

Let $X \sim N(\delta, 1)$, $\delta \in \mathscr{R}$, $U \sim$ the central chi-square with degrees of freedom n, and X and U be independent. The distribution of $T = X/\sqrt{U/n}$ is the noncentral t-distribution with degrees of freedom n and noncentrality parameter δ.

The t-distribution previously defined can be called a central t-distribution, since it is a special case of the noncentral t-distribution with $\delta = 0$.

Using the formula for the ratio of two independent random variables, we can show that T has the following pdf:

$$\frac{1}{2^{(n+1)/2}\Gamma(n/2)\sqrt{\pi n}} \int_0^\infty y^{(n-1)/2} e^{-[(x\sqrt{y/n}-\delta)^2+y]/2} dy$$

Since $U \sim$ the central chi-square with degrees of freedom n, $W = \sqrt{U/n}$ has pdf

$$\frac{2n^{n/2}}{\Gamma(n/2)2^{n/2}} w^{n-1} e^{-nw^2/2}, \qquad w > 0$$

Then, T has pdf

$$\int_0^\infty w\left(\frac{n^{n/2}}{\Gamma(n/2)2^{n/2-1}}w^{n-1}e^{-nw^2/2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-(xw-\delta)^2/2}\right)dw$$

$$= \frac{n^{n/2}}{\Gamma(n/2)2^{(n-1)/2}\sqrt{\pi}}\int_0^\infty w^n e^{[(xw-\delta)^2+nw^2]/2}dw$$

Letting $y = nw^2$, we obtain the desired result.

By the independence of X and U and the fact that $X \sim N(\delta, 1)$ and $U \sim$ the central chi-square,

$$E(T) = E\left(\frac{X}{\sqrt{U/n}}\right) = \sqrt{n}E(X)E\left(\frac{1}{\sqrt{U}}\right)$$

$$= \sqrt{n}\delta\int_0^\infty \frac{1}{\sqrt{u}}\left(\frac{1}{\Gamma(n/2)2^{n/2}}u^{n/2-1}e^{-u/2}\right)du$$

$$= \frac{\sqrt{n}\delta}{\Gamma(n/2)2^{n/2}}\int_0^\infty u^{(n-1)/2-1}e^{-u/2}du = \frac{\sqrt{n}\delta\Gamma((n-1)/2)}{\sqrt{2}\Gamma(n/2)}$$

when $n > 1$ and is ∞ when $n = 1$.

$$E(T^2) = E\left(\frac{X^2}{U/n}\right) = nE(X^2)E\left(\frac{1}{U}\right)$$
$$= n(1+\delta^2)\frac{1}{\Gamma(n/2)2^{n/2}}\int_0^\infty u^{n/2-2}e^{-u/2}du$$
$$= \frac{n(1+\delta^2)}{\Gamma(n/2)2^{n/2}}\Gamma(n/2-1)2^{n/2-1} = \frac{n(1+\delta^2)}{n-2}$$

when $n > 2$ and is ∞ when $n \le 2$.

Hence, when $n > 2$,

$$\mathrm{Var}(T) = \frac{n(1+\delta^2)}{n-2} - \frac{n\delta^2}{2}\left[\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}\right]^2$$

Definition (the noncentral F-distribution)

Let $X_1 \sim$ the noncentral chi-square distribution with degrees of freedom n_1 and noncertrality parameter $\delta \ge 0$, $X_2 \sim$ the central chi-square distribution with degrees of freedom n_2, and X_1 and X_2 be independent. The distribution of $F = (X_1/n_1)/(X_2/n_2)$ is called the noncentral F-distribution with degrees of freedom n_1 and n_2 and noncentrality parameter δ.

The F-distribution introduced previously can be called a central F-distribution, since it is a special case of the noncentral F-distribution with $\delta = 0$.

Using the formula for the ratio of two independent random variables and the pdf of X_1 we derived previously, we can obtain the pdf for $F = (X_1/n_1)/(X_2/n_2)$.

Let f_v denote the pdf of the central chi-square with degrees of freedom v and $f_{k_1,k_2}(x)$ be the pdf of the central F-distribution with degrees of freedom k_1 and k_2.

Then the pdf of F is

$$\int_0^\infty y \left[e^{-\delta/2} \sum_{j=0}^\infty \frac{(\delta/2)^j}{j! n_1 n_2} f_{2j+k} \left(\frac{xy}{n_1} \right) \right] f_{n_2} \left(\frac{y}{n_2} \right) dy$$

$$= e^{-\delta/2} \sum_{j=0}^\infty \frac{(\delta/2)^j}{j!} \int_0^\infty \frac{y}{n_1 n_2} f_{2j+k} \left(\frac{xy}{n_1} \right) f_{n_2} \left(\frac{y}{n_2} \right) dy$$

$$= e^{-\delta/2} \sum_{j=0}^\infty \frac{(\delta/2)^j}{j!} \frac{n_1}{(2j+n_1)} f_{2j+n_1,n_2} \left(\frac{n_1 x}{2j+n_1} \right)$$

Here, we used the following result: if $g_j(x) \geq 0$ for all $j = 0, 1, 2, ...$ and $x \in \mathscr{R}$, then

$$\int_{-\infty}^{\infty} \sum_{j=0}^{\infty} g_j(x)dx = \sum_{j=0}^{\infty} \int_{-\infty}^{\infty} g_j(x)dx$$

which holds even when one of side is ∞.

To show this, note that $G_n(x) = \sum_{j=0}^{n} g_j(x)$ is increasing in n for each x.

By the monotone convergence theorem,

$$\begin{aligned}
\sum_{j=0}^{\infty} \int_{-\infty}^{\infty} g_j(x)dx &= \lim_{n\to\infty} \sum_{j=0}^{n} \int_{-\infty}^{\infty} g_j(x)dx = \lim_{n\to\infty} \int_{-\infty}^{\infty} \sum_{j=0}^{n} g_j(x)dx \\
&= \lim_{n\to\infty} \int_{-\infty}^{\infty} G_n(x)dx = \int_{-\infty}^{\infty} \lim_{n\to\infty} G_n(x)dx \\
&= \int_{-\infty}^{\infty} \lim_{n\to\infty} \sum_{j=0}^{n} g_j(x)dx = \int_{-\infty}^{\infty} \sum_{j=0}^{\infty} g_j(x)dx
\end{aligned}$$

Finally, let's calculate the mean and variance of F.

From the previous calculation,

$$E\left(\frac{1}{X_2}\right) = \begin{cases} \frac{1}{n_2-2} & n_2 > 2 \\ \infty & n_2 \le 2 \end{cases}$$

Then, when $n_2 > 2$,

$$E(F) = E\left(\frac{X_1/n_1}{X_2/n_2}\right) = E\left(\frac{X_1}{n_1}\right) E\left(\frac{n_2}{X_2}\right) = \frac{n_1+\delta}{n_1}\frac{n_2}{n_2-2} = \frac{n_2(n_1+\delta)}{n_1(n_2-2)}$$

Also,

$$E\left(\frac{1}{X_2^2}\right) = \frac{1}{\Gamma(n_2/2)2^{n_2/2}} \int_0^\infty x^{n_2/2-3} e^{-x/2} dx = \begin{cases} \frac{1}{(n_2-2)(n_2-4)} & n_2 > 4 \\ \infty & n_2 \le 4 \end{cases}$$

Thus, when $n_2 > 4$,

$$\begin{aligned} \text{Var}(F) &= E\left(\frac{X_1^2/n_1^2}{X_2^2/n_2^2}\right) - [E(F)]^2 = \frac{n_2^2}{n_1^2} E(X_1^2) E\left(\frac{1}{X_2}\right) - \left(\frac{n_2(n_1+\delta)}{n_1(n_2-2)}\right)^2 \\ &= \frac{n_2^2}{n_1^2}\left(\frac{2n_1+4\delta+(n_1+\delta)^2}{(n_2-2)(n_2-4)} - \frac{(n_1+\delta)^2}{(n_2-2)^2}\right) \\ &= \frac{2n_2^2[(n_1+\delta)^2+(n_2-2)(n_1+2\delta)]}{n_1^2(n_2-2)^2(n_2-4)} \end{aligned}$$