Chi-Square Distribution Revisited

Shoichi Midorikawa

1 Mathematical Preliminaries

Let us first consider the integral $\int_0^\infty e^{-\alpha x} x^{s-1}\, dx$ provided that α is a real number. By setting $t = \alpha x$ we find

$$\int_0^\infty e^{-\alpha x} x^{s-1}\, dx = \frac{1}{\alpha^s} \int_0^\infty e^{-t} t^{s-1}\, dt.$$

Recalling that the gamma function $\Gamma(s)$ is defined by

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1}\, dt,$$

we rewrite this integral as

$$\int_0^\infty e^{-\alpha x} x^{s-1}\, dx = \frac{\Gamma(s)}{\alpha^s}. \tag{1}$$

We can show that the equation (1) still holds even if α is some complex number $\alpha = p + iq$ with $p > 0$. Namely,

$$\int_0^\infty e^{-(p+iq)x} x^{s-1}\, dx = \frac{\Gamma(s)}{(p+iq)^s} \tag{2}$$

We multiply both sides of (2) by e^{iqy}, and integrate over q from $-\infty$ to ∞

$$\int_0^\infty dx\, e^{-px} x^{s-1} \int_{-\infty}^\infty e^{iq(y-x)}\, dq = \Gamma(s) \int_{-\infty}^\infty \frac{e^{iqy}}{(p+iq)^s}\, dq. \tag{3}$$

Since

$$\int_{-\infty}^\infty e^{iq(y-x)}\, dq = 2\pi\delta(y-x)$$

we obtain

$$2\pi e^{-py} y^{s-1} = \Gamma(s) \int_{-\infty}^\infty \frac{e^{iqy}}{(p+iq)^s}\, dq.$$

In other words,

$$\int_{-\infty}^\infty \frac{e^{iqy}}{(p+iq)^s}\, dq = \frac{2\pi}{\Gamma(s)} e^{-py} y^{s-1}. \tag{4}$$

2 Chi-Square Distribution

The standard normal distribution is given by

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}. \tag{5}$$

We define $x = u^2$, and wish to find the distribution $T_1(x)$ of the random variable x.

$$\begin{aligned} T_1(x) &= \int_{-\infty}^{\infty} \delta(x - u^2) f(u)\, du \\ &= \int_{-\infty}^{\infty} \frac{1}{2\sqrt{x}} \left(\delta(u - \sqrt{x}) + \delta(u + \sqrt{x}\right) f(u)\, du \\ &= \frac{1}{2\sqrt{x}} \left(f(\sqrt{x}) + f(-\sqrt{x})\right) \end{aligned}$$

Using (5) we obtain

$$T_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x/2} x^{-1/2} \qquad (x \geq 0).$$

This probability density function is the so called chi-square (χ^2) distribution with 1 degree of freedom.

Let $x_1,\ x_2,\ \cdots,\ x_n$ are statistically independent random vaiables each of which has the probability distribution $T_1(x)$. Let

$$y = x_1 + x_2 + \cdots + x_n.$$

The probability distribution $T_n(y)$ of y can then be written formally as

$$T_n(y) = \int_0^{\infty} \delta(y - x_1 - x_2 - \cdots - x_n) T_1(x_1) T_1(x_2) \cdots T_1(x_n)\, dx_1 dx_2 \cdots dx_n.$$

We use the integral representation of the Dirac δ-function

$$\delta(y - x_1 - x_2 - \cdots - x_n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(y - x_1 - x_2 - \cdots - x_n)}\, dk$$

to carry out the integral.

$$\begin{aligned} T_n(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{iky} \prod_{i=1}^{n} \int_0^{\infty} e^{-ikx_i} T_1(x_i)\, dx_i \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{iky} \left[\int_0^{\infty} e^{-ikx} T_1(x)\, dx \right]^n \end{aligned} \tag{6}$$

We find by using (2) that

$$\begin{aligned}\int_0^\infty e^{-ikx} T_1(x)\,dx &= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\left(\frac{1}{2}+ik\right)x} x^{\frac{1}{2}-1}\,dx \\ &= \frac{1}{\sqrt{2\pi}} \frac{\Gamma\left(\frac{1}{2}\right)}{\left(\frac{1}{2}+ik\right)^{1/2}} \\ &= \frac{1}{\sqrt{2}\left(\frac{1}{2}+ik\right)^{1/2}}\end{aligned}$$

where we used the relation $\Gamma(1/2) = \sqrt{\pi}$. Then substituting this result to (6), we obtain

$$T_n(y) = \frac{1}{2\pi} \frac{1}{2^{n/2}} \int_{-\infty}^\infty \frac{e^{iky}}{\left(\frac{1}{2}+ik\right)^{n/2}}\,dk$$

To evaluate the integral we use (4), and finally find that

$$T_n(y) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}.$$

This function is called the chi-square distribution with n degrees of freeedom.

3 Distribution of the Sample Variance

Let n quantities $u_1,\ u_2,\ \cdots,\ u_n$ are independent random variables selected from the standard normal distribution $N(0,1)$, and let

$$\bar{u} = \frac{u_1 + u_2 + \cdots + u_n}{n} \tag{7}$$

is the sample mean of the n variables.

We then consider the distribution of a sum of sqaures defined by

$$x = (u_1 - \bar{u})^2 + (u_2 - \bar{u})^2 + \cdots + (u_n - \bar{u})^2. \tag{8}$$

This distribution is written formally as

$$\begin{aligned}f_X(x) &= \int \delta\left(x - \sum_{i=1}^n (u_i - \bar{u})^2\right) \\ &\quad \times \delta\left(\bar{u} - \sum_{i=1}^n u_i/n\right) \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-u_i^2/2}\,du_i\,d\bar{u}\end{aligned} \tag{9}$$

As we introduce redundant variable $\bar{u}$, it is necessary to insert delta function $\delta\left(\bar{u} - \sum_{i=1}^n u_i/n\right)$ in order to assure the constraint (7).

Note that

$$\sum_{i=1}^{n}(u_i - \bar{u})^2 = \sum_{i=1}^{n} u_i^2 - n\bar{u}^2.$$

Furthermore, we use an integral representation of the δ function

$$\begin{aligned}\delta\left(x - \sum_{i=1}^{n}(u_i - \bar{u})^2\right) &= \delta\left(x - \sum_{i=1}^{n} u_i^2 + n\bar{u}^2\right) \\ &= \frac{1}{2\pi}\int e^{ip\left(x - \sum_{i=1}^{n} u_i^2 + n\bar{u}^2\right)}\, dp. \end{aligned} \tag{10}$$

to evaluate the integral (9). We also use the following relation to the other δ function,

$$\delta\left(\bar{u} - \frac{\sum_{i=1}^{n} u_i}{n}\right) = \frac{1}{2\pi}\int e^{iq\left(\bar{u} - \sum_{i=1}^{n} u_i/n\right)}\, dq. \tag{11}$$

Substituting (10) and (11) into (9), we obtain

$$f_X(x) = \frac{1}{(2\pi)^2}\int e^{ipx}\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(1+2ip)}{2}u_i^2 - iqu_i/n}\, du_i\right) e^{i(pn\bar{u}^2 + q\bar{u})}\, d\bar{u}dpdq. \tag{12}$$

The integral in the parentheses can be carried out in the following way:

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}}\int e^{-\frac{(1+2ip)}{2}u_i^2 - iqu_i/n}\, du_i \\ &= \frac{1}{\sqrt{2\pi}}\int \exp\left\{-\frac{(1+2ip)}{2}\left(u_i + i\frac{q}{n(1+2ip)}\right)^2 - \frac{q^2}{2n^2(1+2ip)}\right\} du_i \\ &= \frac{1}{(1+2ip)^{1/2}}\exp\left\{-\frac{q^2}{2n^2(1+2ip)}\right\} \end{aligned} \tag{13}$$

Substituting (13) into (12) gives

$$f_X(x) = \frac{1}{(2\pi)^2}\int \frac{e^{ipx}}{(1+2ip)^{n/2}}\exp\left\{-\frac{q^2}{2n(1+2ip)} + ipn\bar{u}^2 + iq\bar{u}\right\} d\bar{u}dpdq. \tag{14}$$

The expression in the parentheses $\{\cdots\}$ in (14) can be written as

$$-\frac{q^2}{2n(1+2ip)} + ipn\bar{u}^2 + iq\bar{u} = -\frac{1}{2n(1+2ip)}\Big[q - in(1+2ip)\bar{u}\Big]^2 - \frac{n}{2}\bar{u}^2 \tag{15}$$

and doing the q integral gives

$$\int \exp\left\{-\frac{1}{2n(1+2ip)}\Big[q - in(1+2ip)\bar{u}\Big]^2\right\} dq = \sqrt{\frac{2\pi}{n(1+2ip)}}\ . \tag{16}$$

This leads (14) to

$$f_X(x) = \frac{1}{(2\pi)^{3/2}}\int \frac{\sqrt{n}\, e^{ipx}}{(1+2ip)^{(n-1)/2}}\, e^{-n\bar{u}^2/2}\, d\bar{u}dp \tag{17}$$

The integral over $\bar{u}$ then becomes

$$\int e^{-n\bar{u}^2/2}\,d\bar{u} = \sqrt{\frac{2\pi}{n}}$$

Thus (17) turns out to be

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi}\int \frac{e^{ipx}}{(1+2ip)^{(n-1)/2}}\,dp \\ &= \frac{1}{2\pi}\frac{1}{2^{(n-1)/2}}\int \frac{e^{ipx}}{\left(\frac{1}{2}+ip\right)^{(n-1)/2}}\,dp \; . \end{aligned} \tag{18}$$

Finally, applying (4) to (18), we obtain

$$f_X(x) = \frac{1}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{x}{2}} x^{\frac{n-1}{2}-1} \; . \tag{19}$$

This is **the probability density function for the χ^2 distribution with $n-1$ degrees of freedom.**