Chi-square, t-, and F-Distributions (and Their Interrelationship)

1 Some Genesis

$Z_1, Z_2, \ldots, Z_\kappa$ iid N(0,1) $\Rightarrow X^2 \equiv Z_1^2 + Z_2^2 + \ldots + Z_\kappa^2 \sim \chi_\kappa^2$.
Specifically, if $\kappa = 1$, $Z^2 \sim \chi_1^2$. The density function of chi-square distribution will not be pursued here. We only note that: Chi-square is a **class** of distribution indexed by its *degree of freedom*, like the t-distribution. In fact, chi-square has a relation with t. We will show this later.

2 Mean and Variance

If $X^2 \sim \chi_\kappa^2$, we show that:

$$\begin{aligned} \mathcal{E}\{X^2\} &= \kappa, \\ \mathcal{VAR}\{X^2\} &= 2\kappa. \end{aligned}$$

For the above example, $Z_j^2 \sim \chi_1^2$, if $\mathcal{E}Z_j^2 = 1, \mathcal{VAR}Z_j^2 = 2, \forall j = 1, 2 \ldots, \kappa$, then

$$\begin{aligned} \mathcal{E}\{X^2\} &= \mathcal{E}\{Z_1^2 + \ldots + Z_\kappa^2\} \\ &= \mathcal{E}Z_1^2 + \ldots + \mathcal{E}Z_\kappa^2 \\ &= 1 + 1 + \ldots + 1 = \kappa. \end{aligned}$$

Similarly, $\mathcal{VAR}\{X^2\} = 2\kappa$. So it suffices to show

$$\mathcal{E}\{Z^2\} = 1, \mathcal{VAR}\{Z^2\} = 2.$$

Since Z is distributed as N(0,1), $\mathcal{E}\{Z^2\}$ is calculated as

$$\int_{-\infty}^{\infty} z^2 \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{z^2/-2} dz,$$

which can easily be shown to be 1. Furthermore, $\mathcal{VAR}\{Z^2\}$ can be calculated through the formula: $\mathcal{VAR}\{Z^2\} = \mathcal{E}\{Z^4\} - (\mathcal{E}\{Z^2\})^2$.

3 Illustration: As an example for Law of Large Number(LLN) and Central Limit Theorem(CLT)

If $X^2 \sim \chi^2_\kappa$, X^2 can be written as $X^2 = Z_1^2 + \ldots + Z_\kappa^2$, for some $Z_1, \ldots, Z_\kappa$ iid N(0,1). Then when $\kappa \to \infty$,

$$\frac{X^2 - \mathcal{E}X^2}{\sqrt{\mathcal{VAR}(X^2)}} = \frac{X^2 - \kappa}{\sqrt{2\kappa}} \sim N(0,1) \text{ (CLT)};$$

moreover,

$$\lim_{\kappa\to\infty} \frac{X^2}{\kappa} \equiv \frac{\chi^2_\kappa}{\kappa} \to 1 \text{ (LLN)}.$$

4 Application: To Make Inference on σ^2

If $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$, then $Z_j \equiv (X_j - \mu)/\sigma \sim N(0,1), j = 1, \ldots, n$. We know, from a previous context, that $\sum_1^n Z_j^2 \sim \chi^2_n$, or equivalently,

$$\sum_{j=1}^n \{\frac{X_j - \mu}{\sigma}\}^2 = \frac{\sum_1^n (X_j - \mu)^2}{\sigma^2} \sim \chi^2_n,$$

if μ is *known*, or otherwise (if μ is unknown) μ needs to be estimated (by $\overline{X}$, say,) such that (still needs to be proved):

$$\frac{\sum_1^n (X_j - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{1}$$

Usually μ is not known, so we use formula (1) to make inference on σ^2. Denotes $\chi^2_{\nu,p}$ as the **p-th percentile from the right** for the χ^2_ν-distribution, the two-sided confidence interval (CI) of σ^2 can be constructed as follows:

$$\begin{aligned}
& \Pr\{\chi^2_{n-1,1-\alpha/2} < \frac{\sum_1^n (X_j - \overline{X})^2}{\sigma^2} < \chi^2_{n-1,\alpha/2}\} = 100(1-\alpha)\% \\
\Longleftrightarrow \quad & \Pr\{\frac{1}{\chi^2_{n-1,1-\alpha/2}} > \frac{\sigma^2}{\sum_1^n (X_j - \overline{X})^2} > \frac{1}{\chi^2_{n-1,\alpha/2}}\} = 100(1-\alpha)\% \\
\Longleftrightarrow \quad & \Pr\{\frac{\sum_1^n (X_j - \overline{X})^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{\sum_1^n (X_j - \overline{X})^2}{\chi^2_{n-1,1-\alpha/2}}\} = 100(1-\alpha)\%. \qquad (2)
\end{aligned}$$

That is, the $100(1-\alpha)\%$ CI for σ^2 is

$$\left(\frac{\sum_1^n (X_j-\overline{X})^2}{\chi^2_{n-1,\alpha/2}}, \frac{\sum_1^n (X_j-\overline{X})^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

[**QUESTION:** How to use this formula with actual data?]

5 The Interrelationship Between t-, χ^2-, and F- Statistics

5.1 t versus χ^2

If $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$, then

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

When σ is unknown,

$$\frac{\overline{X}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}, \text{ where } \hat{\sigma} = \sqrt{\frac{\sum(X_i-\overline{X})^2}{n-1}}. \tag{3}$$

Note that

$$\begin{aligned}
\frac{\overline{X}-\mu}{\hat{\sigma}/\sqrt{n}} &= \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \cdot \frac{1}{\frac{\hat{\sigma}}{\sigma}} \\
&= Z \cdot \frac{1}{\frac{\hat{\sigma}}{\sigma}} \\
&= \frac{Z}{\sqrt{\frac{\sum(X_i-\overline{X})^2}{(n-1)\sigma^2}}} \\
&= \frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}. \qquad (4)
\end{aligned}$$

Combining (3) and (4) gives

$$t_{n-1} = \frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}},$$

or, in general,

$$t_\kappa = \frac{Z}{\sqrt{\frac{\chi^2_\kappa}{\kappa}}}.$$

5.2 F versus χ^2

From some notes about ANOVA, we have learned that

$$F_{a,b} \equiv \frac{\chi_a^2/a}{\chi_b^2/b} \text{ (Sir R. A. Fisher).} \tag{5}$$

The original concern of Fisher is to construct a **statistic** which has a sampling distribution, in some extent, free from the degrees of freedom a and b *under the null hypothesis*. With this concern, he presented his F-statistic in a way that: Since χ_a^2 has expectation a, so the numerator χ_a^2/a has expectation 1; similarly, the denominator also has expectation 1. As Fisher said, **the value of F-statistic will fluctuate near 1 under the null hypothesis** $H_0 : \mu_1 = \ldots = \mu_\kappa$ (if $\kappa = a + 1$).
From (5), we are able to express the χ^2-distribution in terms of F:

$$F_{a,\infty} = \frac{\chi_a^2/a}{\lim_{b\to\infty}(\chi_b^2/b)} = \frac{\chi_a^2/a}{1},$$

from Section 3 (LLN). So χ^2-table can be treated as a part of F-tables.

5.3 t versus F

$$\begin{aligned} t_\nu &= \frac{Z}{\sqrt{\chi_\nu^2/\nu}} \\ &= \frac{\sqrt{\chi_1^2/1}}{\sqrt{\chi_\nu^2/\nu}} \\ &= \sqrt{F_{1,\nu}}. \end{aligned}$$

Or, in other words, $t_\nu^2 = F_{1,\nu}$, a formulae useful in assessing **partial effect** in linear regression model.