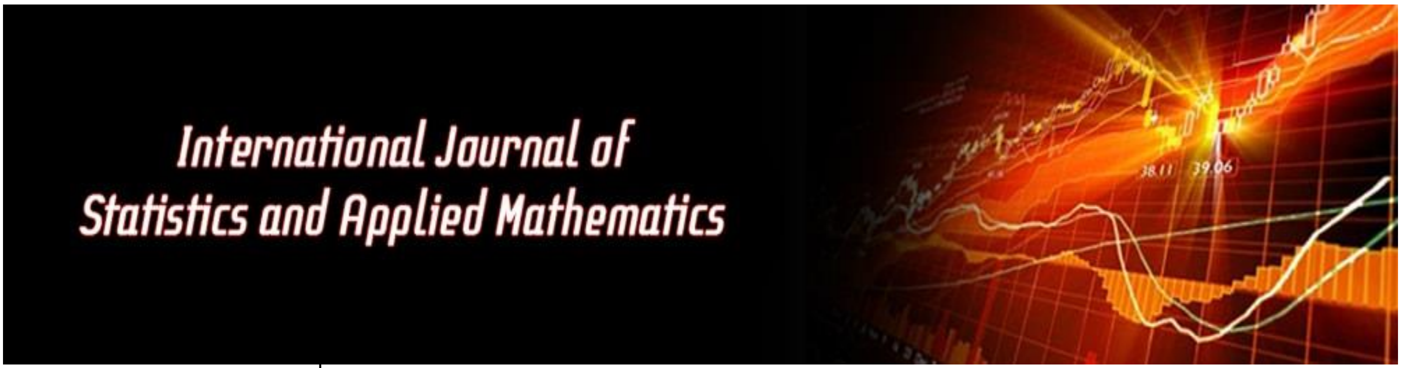
International Journal of Statistics and Applied Mathematics 2020; 5(5): 07-11



ISSN: 2456-1452
Maths 2020; 5(5): 07-11

www.mathsjournal.com
Received: 05-07-2020
Accepted: 09-08-2020

Ruijia Shi
School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao, Hebei, China

Corresponding Author:
Ruijia Shi
School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao, Hebei, China

The relationship between the three major sampling distributions and the normal distribution

Ruijia Shi

Abstract
The chi-square distribution, *t* distribution and *F* distribution are the three most important sampling distributions in probability theory and mathematical statistics. They all have a certain relationship with the normal distribution, and they are asymptotically normal under certain conditions. This article first combines the gamma distribution and the moment generating function indicate the correlation properties of the three major sampling distributions and the relationship between the three. Using the relevant functions such as the characteristic function and the central limit theorem to explain that they all converge to the normal distribution under certain conditions. At the same time, they use mathematical software MATLAB to verify.

Keywords: Three major sampling distributions Normal distribution Matlab characteristic function central limit theorem

1. Introduction
The three major sampling distributions occupy an important position in probability theory and mathematical statistics, and are often used in estimation and inference problems such as hypothesis testing, interval estimation, and analysis of variance. They are all evolved from the normal distribution, and finally belong to the normal. However, most of the textbooks do not explain the relationship between the three distributions in detail. This article summarizes and analyzes the three distributions based on the knowledge learned, highlighting their connection with the normal distribution and proving that they are all under certain conditions. Converge to the normal distribution, and perform MATLAB verification at the same time. With such a relationship, some more complex distributions can be transformed into normal distributions that are easy to accept and understand, and it is more convenient to operate.

2. Preliminary knowledge
If a random variable X follows a gamma distribution with a shape parameter of α and a scale parameter of β, we record it as $X(\alpha,\beta)$ and its probability density function is $X\sim\Gamma(\alpha,\ \beta)$ then its probability density function is

$$f(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\ \alpha,\beta>0,x\geq 0\ E(X)=\frac{\alpha}{\beta},Var(X)=\frac{\alpha}{\beta^{2}}$$

Where $\Gamma(\alpha)$ is the gamma function $\Gamma(\alpha)=\int_0^{\infty}x^{\alpha-1}e^{-x}\,dx$ and the moment generating function of the gamma distribution is $M(t)=\left(\frac{\beta}{\beta-t}\right)^{\alpha}$

3. Definition and nature of the three major sampling distributions
3.1 chi-square distribution
Definition If $Z\sim N(0,1)$, $X=Z^2$ then Xobey the chi-square distribution with 1 degree of freedom, denoted as $X\sim\chi^2_{[1]}$.The following uses the moment generating function to analyze the relationship between chi-square distribution and Γ distribution:

$$F_X(x) = P(X \le x) = P(Z^2 \le x) = P(-\sqrt{x} \le Z \le \sqrt{x})$$

$$F_X(x) = F_Z(\sqrt{x}) - F_Z(-\sqrt{x})$$

The derivative of the distribution function can be obtained:

$$f_X(x) = \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$$

It is not difficult to find that the above formula happens to be the probability density function of $\Gamma(\frac{1}{2},\frac{1}{2})$, so $M_X(t) = (1-2t)^{-\frac{1}{2}}$, which is written as $X \sim \chi^2_{[1]}$.

Nature 1: Additivity of chi-square distribution

If $Z_1, Z_2, \ldots, Z_n$ is an independent random variable, and $Z_i \sim N(0,1)$, let $X = \sum_{i=1}^{n} {Z_i}^2$, then $X \sim \chi^2_{(n)}$

Prove: Using moment generating function and independence can be obtained:

$$\mathrm{M_X(t) = M_{Z_1^2}(t) \times M_{Z_2^2}(t) \times \ldots \times M_{Z_n^2}(t),}$$

Because of $Z_i^2 \sim \chi^2_{(1)}$, its moment generating function is $M_{Z_i^2}(t) = (1-2t)^{-\frac{1}{2}}$

Therefore, $M_Y(t) = (1-2t)^{-\frac{n}{2}}$ this is the moment generating function of $\Gamma(\frac{n}{2},\frac{1}{2})$, so$X \sim \Gamma(\frac{n}{2},\frac{1}{2})$, at this time we record as $X \sim \chi^2_{(n)}$

Nature 2: Chi-square distribution asymptotically normal distribution
When $Z_i^2 \sim \chi^2_{(1)}$, we can know $Z_i^2 \sim \Gamma(\frac{1}{2},\frac{1}{2})$, according to the expected variance calculation formula of the gamma distribution, we can get:

$$E(Z_i^2) = \frac{\alpha}{\beta} = 1, Var(Z_i^2) = \frac{\alpha}{\beta^2} = 2$$

According to the central limit theorem, for any x, the distribution function

$F_n(x) = P\left\{\frac{\sum_{i=1}^{n} Z_i^2 - nE(Z_i^2)}{\sqrt{Var(Z_i^2)n}} \le x\right\}$ the following formulas are all established

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} Z_i^2 - n\,E(Z_i^2)}{\sqrt{Var(Z_i^2)n}} \le x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = \phi(x)$$

The theorem states that when n approaches infinity, the random variable

$$Y_n = \frac{\sum_{i=1}^{n} Z_i^2 - nE(Z_i^2)}{\sqrt{Var(Z_i^2)n}} = \frac{\sum_{i=1}^{n} Z_i^2 - n}{\sqrt{2n}} \to N(0,1)$$

So $\sum_{i=1}^{n} Z_i^2 \to N(n, 2n)$, and because of $\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$, namely when n approaches infinity, $\chi^2_{(n)}$ gradually approaching the $N(n, 2n)$

Matlab Drawing Verification

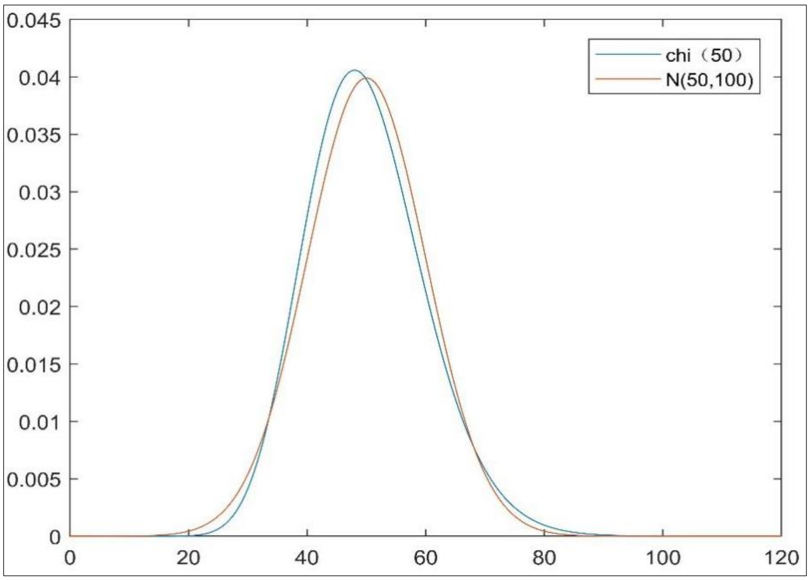


Fig 1: The chi-square distribution converges to normal distribution

From the figure 1 above, it can be seen that the larger the n, the closer the chi-square distribution density function is to the normal distribution density function, which is consistent with the proven conclusion.

3.2 t distribution

Definition If $Z \sim N(0,1), U \sim \chi^2(n), Z$ and U are independent of each other, then

$$\frac{Z}{\sqrt{\frac{U}{n}}} \sim t(n)$$

Nature 1: If $X_1, X_2, \ldots, X_n$ is random variables of independent and the same distribution, and $X_i \sim N(\mu, \sigma^2)$, then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, because of $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.By using the definition of the t-distribution we have $\frac{\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$

Nature 2: If the random variable $X \sim t(n)$, then when n is greater than or equal to 2, $E(X) = 0$. when n is greater than or equal to 3, $Var(X) = \frac{n}{n-2}$

Nature 3: When n tends to infinity, $\lim_{n\to\infty} t_n(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

Prove: Setting $T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim t(n)$, in this expression $Z \sim N(0,1)$, $U \sim \chi^2(n)$ and they're independent of each other. Because the characteristic function of U is $\varphi_U(t) = (1-2it)^{-\frac{n}{2}}$, the characteristic function of $\frac{U}{n}$ is $\varphi_{\frac{U}{n}}(t) = (1-\frac{2}{n}it)^{-\frac{n}{2}}$. According to the special limit formula available: $\lim_{n\to\infty} \varphi_{\frac{U}{n}}(t) = \lim_{n\to\infty}(1-\frac{2}{n}it)^{-\frac{n}{2}} = \lim_{n\to\infty}\left[(1-\frac{2}{n}it)^{-\frac{n}{2it}}\right]^{it} = e^{it}$

It is not difficult to find that e^{it} is a characteristic function of single-point distribution $P(X_0 = 1) = 1$, so the distribution function of $\frac{U}{n}$

$$F_n(x) \xrightarrow{W} F(x) = \begin{cases} 0, x < 0 \\ 1, x \geq 0 \end{cases} (n \to \infty)$$

namely, $\frac{U}{n} \xrightarrow{P} X_0 = 1 (n \to \infty), \sqrt{\frac{U}{n}} \xrightarrow{P} \sqrt{X_0} = 1 (n \to \infty)$. So, $T = \frac{Z}{\sqrt{\frac{U}{n}}} \xrightarrow{P} Z (n \to \infty)$

According to the above deductions, we can see $\lim_{n\to\infty} t_n(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

MATLAB drawing verification

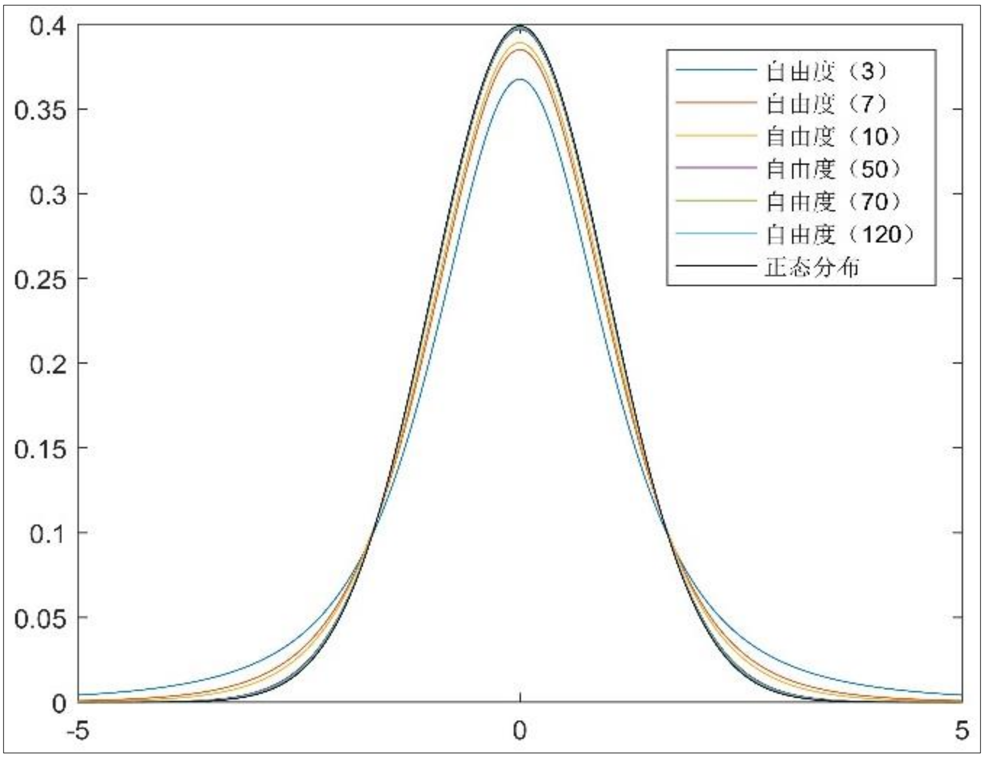


Fig 2: The t distribution converges to normal distribution

3.3 *F* distribution

Definition If $U \sim \chi^2_{(n)}, V \sim \chi^2_{(m)}$, U and V are independent of each other. Then

$$\frac{\frac{U}{n}}{\frac{V}{m}} \sim F(n,m)$$

Nature 1: If $X \sim F(n,m)$,we can deduce $\frac{1}{X} \sim F(m,n)$

Nature 2: If $X \sim t(n)$,we can deduce $X^2 \sim F(1,n)$

Nature 3: $F_\alpha(m,n) = \frac{1}{F_{1-\alpha}(n,m)}$. This formula can be used to solve the values of some of the quartile not listed on the F distribution table (α represents the upper quartile).

Nature 4: When x is greater than or equal to zero, F distribution will converge to $N(1,\frac{2}{n})$.

Prove When x is greater than or equal to zero, the density function of the F distribution is known as follows:

$$f(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{x^{\frac{n}{2}-1}}{(nx+m)^{\frac{m+n}{2}}}$$

When m tends to infinity, $p(x) = \lim_{m \to +\infty} f(x) = \frac{(\frac{n}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{n}{2}x}$. According to the density function expression of the gamma distribution, $p(x)$ is the density function of $\frac{\chi^2_{(n)}}{n}$.

Therefore, $E(\frac{\chi^2_{(n)}}{n}) = 1$, $Var(\frac{\chi^2_{(n)}}{n}) = \frac{2}{n}$. According to the central limit theorem, when n tends to infinity, $\frac{\frac{\chi^2_{(n)}}{n}-1}{\sqrt{\frac{2}{n}}} \to N(0,1)$.Namely, $\frac{\chi^2_{(n)}}{n} \to N(1,\frac{2}{n})$.

In summary, when n and m tends to infinity, $F(m,n)$ converges to $N(1,\frac{2}{n})$.

MATLAB drawing verification

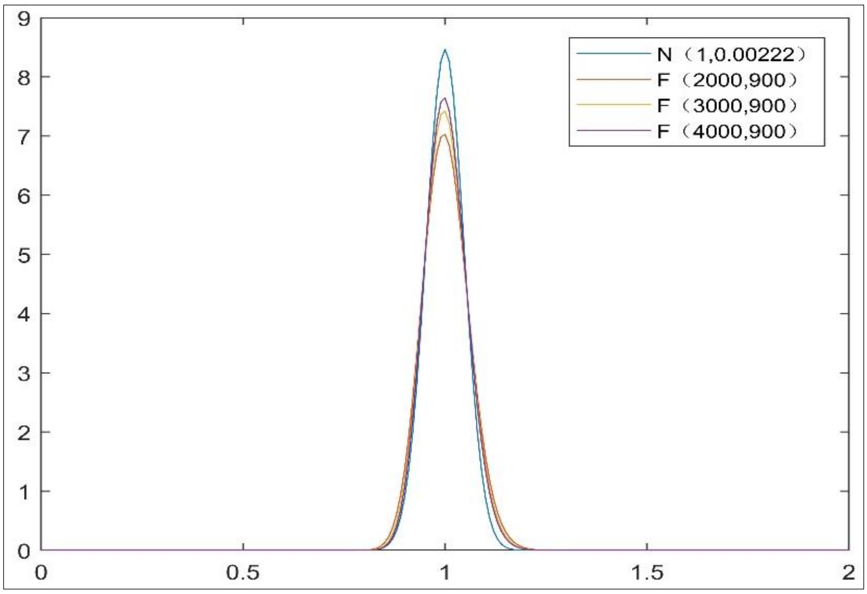


Fig 3: The F distribution converges to normal distribution

4. Conclusion

In this paper, through the learned knowledge, with the help of gamma distribution, characteristic function, and central limit theorem, it is proved that the three distributions converge to the normal distribution under certain conditions. It can be seen that the normal distribution plays an important role in mathematical statistics. When the sample size is large, We can use the normal distribution to approximate the three sampling distributions, and use mathematical software to verify. By further summarizing, studying multiple proof methods for one property is very helpful for understanding the relationship between the three distributions.

5. References

1. Sri Guleng. The role of feature function [J] Inner Mongolia Normal University News. 2000; (2):33-37.
2. Zeng Zhen, Zhang YU. The relationship between the chi-square distribution, t distribution, F distribution and the normal distribution [J] Journal of Hubei Normal University. 2015; 35(3):62-66.
3. Zhou Wei. The relationship between the chi-square distribution, t distribution and F distribution，（Engineering Mathematics [M]. 1992; 8(2):93-95.
4. Li Kaican, Meng Zhaoling. Uniform asymptotic normality of the chi-square distribution, t distribution and F distribution [J] Journal of Beijing Institute of Printing Technology. 2004; 12(3):30-33.