MATH1312: Lecture Note on Probability Theory and Mathematical Statistics

Shixiao W. Jiang

Institute of Mathematical Sciences, ShanghaiTech University, Shanghai 201210, China

jiangshx@shanghaitech.edu.cn

2024 年 11 月 20 日

Contents

Chapter 1

Statistics and Their Distributions

1.1 The notion of a sample

（总体、母体）

Definition 1.1.1 *The collection of all elements under investigation is called the population.*

Definition 1.1.2 *The variables representing the population is called the random variables. When we say that the population has the distribution $F(x)$, we mean that we are investigating a character X of elements of this population and this character X is a random variable with the distribution function $F(x)$.*

随机样本（简称样本），抽样　样本值、观察值，样本空间

Definition 1.1.3 *Sampling is the selection of a subset (a data sample) of individuals from a statistical population to estimate characteristics of the whole population. A random sample of size n from a population X (or a distribution, or a CDF $F(x)$) consists of i.i.d. random variables $X_1, X_2, \ldots, X_n$, each has the same distribution as the population X. The values $x_1, x_2, \ldots, x_n$ of $X_1, X_2, \ldots, X_n$ (observed and recorded from the experiments) are called sample values. The set of all possible random samples of n elements is called the sample space.*

Definition 1.1.4 *(empirical distribution function) Let $x_1, x_2, \ldots, x_n$ be observations from $F(x)$. Let $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ be the order statistics.*

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_{(k)} \leq x \leq x_{(k+1)} \quad (k = 1, \ldots, n-1), \\ 1, & x \geq x_{(n)}. \end{cases}$$

Obviously, $F_n(-\infty) = 0, F_n(+\infty) = 1$. $F_n(x)$ is called a empirical distribution function or a sample distribution function.

Proposition 1.1.5 *By law of large number, for any given x, one has $F_n(x) \xrightarrow{p} F(x)$ as $n \to \infty$.*

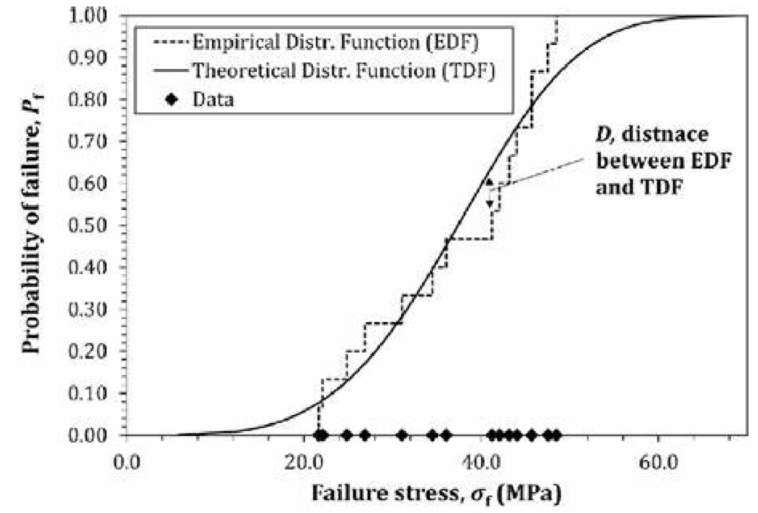

Figure 1.1: Diamonds are the observation data points, solid line is the true CDF, and dashed line is the empirical CDF.

1.2 The notion of a statistics

Definition 1.2.1 *A random variable which is a function of the observed random vector $(X_1, \ldots, X_n)$ and known parameters is called a statistics.* ***Note that a statistics is not allowed to involve unknown parameters.*** *Specifically, let $X_1, X_2, \ldots, X_n$ be samples, and then $g(X_1, \ldots, X_n)$ is a statistics. When observations are $x_1, x_2, \ldots, x_n$, then $g(x_1, \ldots, x_n)$ is an observation value of the statistics.*

Example 1.2.2 *Let $X \sim N(\mu, \sigma^2)$ with given μ but unknown σ^2. Let $(X_1, \ldots, X_n)$ be samples. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \quad \textit{and} \quad \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

are statistics, whereas $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2$ not since it involves with the unknown σ^2.

We have already encountered some of the most important statistics.We now list some below. Their denitions involve the size n of the sample, which are usually omitted when there is no confusion. Let $X_1, \ldots, X_n$ be i.i.d. samples. Then

样本均值 • The sample mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ .$$

样本方差 • The sample variance

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

无偏样本方差 • The unbiased sample variance

$$S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

样本k阶矩 • The sample moment of kth order

$$a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

样本k阶中心矩 • The sample central moment of kth order

$$b_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k.$$

Example 1.2.3 *Let $\xi_1, \ldots, \xi_n$ and $\eta_1, \ldots, \eta_m$ be independent samples from $N(0,1)$ of size n and m. Then*

$$\begin{aligned}
\chi^2 &= \sum_{i=1}^{n} \xi_i^2, \\
t &= \frac{\xi_i}{\sqrt{\chi^2/n}}, \\
F &= \frac{\sum_{j=1}^{m} \eta_j^2/m}{\sum_{i=1}^{n} \xi_i^2/n},
\end{aligned}$$

are all statistics.

1.3 Review of χ^2, t, F distributions

1.3.1 χ^2 distribution

Lemma 1.3.1 *Let X be a random variable of continuous type with density function $p(x)$. Let $y = f(x)$ be a strictly monotonic function and let $x = h(y)$ be the inverse function of $y = f(x)$ with continuous derivative. Then $Y = f(X)$ is also a continuous random variable with density function*

$$\psi(y) = \begin{cases} p(h(y)) \cdot |h'(y)|, & \alpha < y < \beta, \\ 0 & \text{else}, \end{cases}$$

where $\alpha = \min\{f(-\infty), f(+\infty)\}$ and $\beta = \max\{f(-\infty), f(+\infty)\}$.

Proof. Assume that $f(x)$ a strictly monotonically increasing function. Then $h(y)$ is also strictly monotonically increasing.

$$\begin{aligned}
F_Y(y) &= P(Y \le y) = P(f(X) \le y) \\
&= P(X \le h(y)) = \int_{-\infty}^{h(y)} p(x)dx, \quad \text{for } f(-\infty) < y < f(+\infty).
\end{aligned}$$

Thus, the pdf is

$$\psi(y) = F_Y'(y) = \begin{cases} p(h(y)) \cdot h'(y), & f(-\infty) < y < f(+\infty), \\ 0, & \text{else}. \end{cases}$$

One can follow the similar procedure here to prove for the case when $f(x)$ is strictly monotonically decreasing.
■

Remark 1.3.2 *The condition can be relaxed to that the function $f(x)$ is piecewise strictly monotonic and its inverse function is continuous and differentiable. For example, $y = x^2$.*

Example 1.3.3 *Let X be a standard normal distribution $N(0,1)$. Find the pdf of $Y = X^2$.*
Sol. For $y < 0$, $F_Y(y) = P(Y \le y) = 0$. For $y \ge 0$,

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{+\sqrt{y}} \varphi(x)dx, \end{aligned}$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Then, the pdf of Y is

$$\psi(y) = F_Y'(y) = \begin{cases} \varphi(\sqrt{y})\frac{1}{2}y^{-\frac{1}{2}} + \varphi(-\sqrt{y})\frac{1}{2}y^{-\frac{1}{2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}}\frac{1}{\sqrt{y}}, & y \ge 0, \\ 0, & y < 0. \end{cases}$$

This is the density function of $\chi^2(1)$.

We now show the distribution for $\chi^2(n)$ and then show the detailed computation for $\chi^2(n)$. The distribution of χ^2 was obtained by Helmert (book Fiesz). The parameter n is called the number of degrees of freedom.

自由度为n的χ^2分布

Definition 1.3.4 *Let $X_1, \ldots, X_n$ be i.i.d. with normal distributions $N(0,1)$. Then $\chi^2 = X_1^2 + \cdots + X_n^2$ is said to be a χ^2 distribution with n degrees of freedom, denoted by $\chi^2(n)$. Its pdf is*

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \ge 0, \\ 0, & x < 0. \end{cases} \tag{1.1}$$

Here, Gamma function

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx, \quad t > 0,$$

with $\Gamma(1) = 1, \Gamma(\frac{1}{2}) = \sqrt{\pi}$, and $\Gamma(t+1) = t\Gamma(t)$.

Now let's see how to derive the pdf of $\chi^2(n)$. We first compute the distribution of sum. Let (X, Y) be a random variable of continuous type with joint pdf $p(x, y)$. Then the distribution function of $Z = X + Y$ is

$$F_Z(z) = \iint_{x+y \le z} p(x,y)dxdy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} p(x,y)dy \right) dx.$$

In addition, if X and Y are independent, then

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{\infty} dx \left(\int_{-\infty}^{z-x} p_X(x)p_Y(y)dy \right) \\ &= \int_{-\infty}^{\infty} dx \left(\int_{-\infty}^{z} p_X(x)p_Y(\hat{y}-x)d\hat{y} \right) \quad (y = \hat{y} - x) \\ &= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} p_X(x)p_Y(\hat{y}-x)dx \right) d\hat{y}. \end{aligned}$$

The pdf of Z is

$$p_Z(z) = F_Z'(z) = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x)dx.$$

By symmetry, one can also show that

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(z-x)p_Y(x)dx,$$

where both above formula are the convolution between $p_X(x)$ and $p_Y(y)$.

Example 1.3.5 *Let X, Y be two independent normally distributed random variables. Then their sum $Z = X + Y$ has the pdf of*

$$\begin{aligned} p_Z(z) &= \int_{-\infty}^{\infty} p_X(x) p_Y(z-x) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{-\frac{(z-x)^2}{2}} dx \\ &= \frac{1}{2\pi} e^{-\frac{z^2}{4}} \int_{-\infty}^{\infty} e^{-(x-\frac{z}{2})^2} dx = \frac{1}{2\sqrt{\pi}} e^{-\frac{z^2}{4}}. \end{aligned}$$

Thus, $Z \sim N(0, 2)$.

Remark 1.3.6 *Let X_i $(i = 1, \ldots, n)$ be independent random variable with normal distributions $N(\mu_i, \sigma_i^2)$.*
(1) Then, $\sum_{i=1}^n X_i$ is normally distributed with $N(\mu, \sigma^2)$, where $\mu = \sum_{i=1}^n \mu_i$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.
(2) For any real l_i, $\sum_{i=1}^n l_i X_i$ is also normally distributed with $N(\mu, \sigma^2)$, where $\mu = \sum_{i=1}^n l_i \mu_i$ and $\sigma^2 = \sum_{i=1}^n l_i^2 \sigma_i^2$.
(3) The above can be proved using convolustion formula by induction or using characteristic function method.

Definition 1.3.7 *If X has pdf*

$$p(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Then we say that X is a random variable with a gamma distribution with parameters (α, λ), denoted by $\Gamma(\alpha, \lambda)$.

Example 1.3.8 *Let X, Y be independent random variables with distributions $\Gamma(\alpha_1, \lambda)$ and $\Gamma(\alpha_2, \lambda)$. (Note that the second parameter λ is the same). Find the pdf of $Z = X + Y$.*
Sol. Using the convolution formula,

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(z-x) p_Y(x) dx = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} \int_0^z (z-x)^{\alpha_1-1} x^{\alpha_2-1} dx.$$

Let $\frac{x}{z} = t$. Then

$$\int_0^z (z-x)^{\alpha_1-1} x^{\alpha_2-1} dx = z^{\alpha_1+\alpha_2-1} \int_0^1 (1-t)^{\alpha_1-1} t^{\alpha_2-1} dt$$

where

$$\int_0^1 (1-t)^{\alpha_1-1} t^{\alpha_2-1} dt = B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}.$$

Then

$$\begin{aligned} p_Z(z) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} z^{\alpha_1+\alpha_2-1} \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} z^{\alpha_1+\alpha_2-1} e^{-\lambda z}. \end{aligned}$$

Thus $Z \sim \Gamma(\alpha_1+\alpha_2, \lambda)$, which means that gamma distribution has the additive property for the first parameter.

Remark 1.3.9 *One can also compute the above result using Ch.f. method.*

Remark 1.3.10 *In general, if $\{X_i\}_{i=1}^n$ are independent random variables with distributions $\{\Gamma(\alpha_i, \lambda)\}_{i=1}^n$, then $Z = X_1 + \cdots + X_n \sim \Gamma(\alpha_1 + \cdots + \alpha_n, \lambda)$.*

Remark 1.3.11 *If* $\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$ *for the gamma distribution* $\Gamma(\frac{n}{2}, \frac{1}{2})$*, then one can observe that it is the same as the pdf of* χ^2 *from (1.1), that is,* $\Gamma(\frac{n}{2}, \frac{1}{2}) = \chi^2(n)$*. Moreover,* χ^2 *distribution has the additive property as inherited from the gamma distribution. That is, if* $X \sim \chi^2(n)$ *and* $Y \sim \chi^2(m)$*, and* X *and* Y *are independent, then* $X + Y \sim \chi^2(n+m)$*.*

Remark 1.3.12 *In general, if* $X_1, \ldots, X_k$ *are independent random variables and* $X_i \sim \chi^2(m_i)$ $(i = 1, \ldots, k)$*, then* $X_1 + \cdots + X_k \sim \chi^2(m_1 + \cdots + m_k)$*.*

Remark 1.3.13 *Based on the Definition 1.3.4, if* $X_1, \ldots, X_n$ *are* n *i.i.d. random variables with normal distributions* $N(0,1)$*. Then* $X_k^2 \sim \chi^2(1)$ *for each* $k = 1, \ldots, n$ *and* $Y = X_1^2 + \cdots + X_n^2 \sim \chi^2(n)$ *is said to be a* χ^2 *distribution with* n *degrees of freedom.*

Remark 1.3.14 *If* $X \sim \chi^2(n)$*, then* $EX = n, Var(X) = 2n$*.*

Proof. We can compute the expectation

$$\begin{aligned} EX &= \int_0^\infty x \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx \\ &= \frac{2^{\frac{n+2}{2}}\Gamma\left(\frac{n+2}{2}\right)}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^\infty \frac{1}{2^{\frac{n+2}{2}}\Gamma(\frac{n+2}{2})} x^{\frac{n+2}{2}-1} e^{-\frac{x}{2}} dx \\ &= 2 \cdot \frac{n}{2} = n, \end{aligned}$$

where the integral in the second line is for the pdf of $\chi^2(n+2)$. Similarly, $EX^2 = n^2 + 2n$ and $Var(X) = 2n$. ∎

Last, we consider the case for the independent random variables X_k $(k = 1, \ldots, n)$ where the normal distribution has variance σ^2, that is, the normal density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

The expression

$$\chi^2 = \sum_{k=1}^n X_k^2,$$

is also called the statistic χ^2, whose distribution was obtained by Helmert. We can obtain the expectation and variance as

$$E\chi^2 = n\sigma^2, \quad Var(\chi^2) = 2n\sigma^4.$$

The parameter n is also called the number of degrees of freedom, which corresponds to the fact that χ^2 is the sum of n independent random variables. The tables of the χ^2 distribution usually give the values of the distribution function value with $\sigma = 1$ for different values of x and n.

In Fig. 1.2, the densities of χ^2 for $\sigma = 1$ and various degrees of freedom are displayed.

Remark 1.3.15 *The tables of the* $\chi^2(n)$ *distribution are usually given for no more than thirty degrees of freedom. Fisher [6] (in Fiesz book) showed that if the number of degree of freedom* n *increases to infinity,* ***the random variable*** $\sqrt{2\chi^2(n)}$ ***has the asymptotically normal distribution*** $N(\sqrt{2n-1}, 1)$*.* ***For*** $n \geq 30$*,* ***we may use the tables of the normal distribution****.*

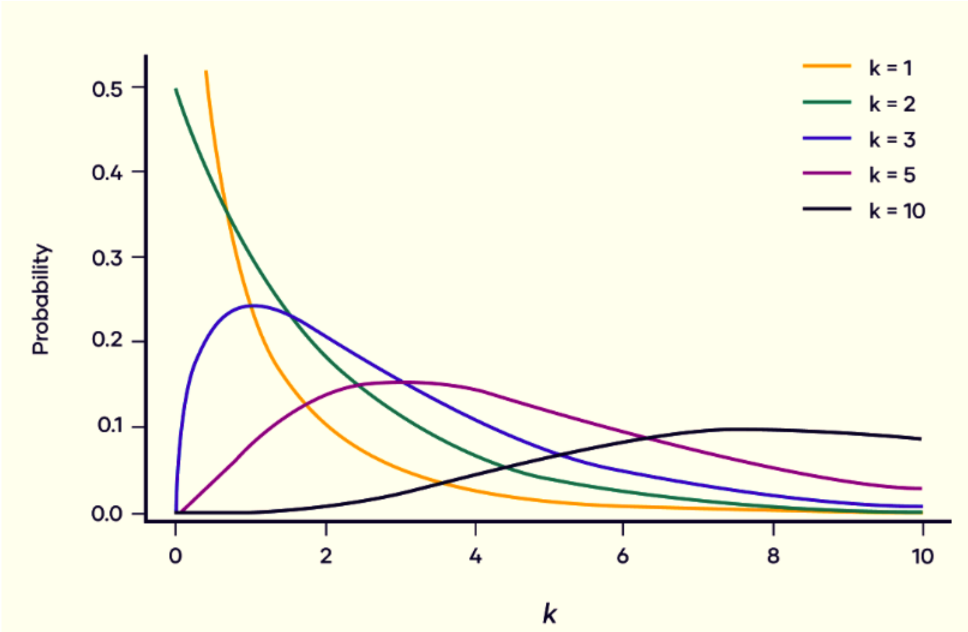

Figure 1.2: Comparison of χ^2 distribution for different n.

1.3.2 Independence and distribution of sample mean $\overline{X}$ and sample variance S_n^2

In applications, we often deal with problems of the following type. The values of sample mean $\overline{X}$ and sample variance S_n^2 are observed and found to be such that $a \leq \overline{x} < b$ and $s \geq c$. We would like to find the probability of these inequalities, or in other words, we would like to find out how often the values $\overline{x}$ of the statistic $\overline{X}$ and the values s of the statistic S_n satisfy these inequalities if we take a large series of observations. First, let us right now prove that $\overline{X}$ and S_n^2 are independent and also find their distributions as well, which will be used for the definition of the following Student's t distribution.

Proposition 1.3.16 *Let $\boldsymbol{\xi}^\top = (\xi_1, \ldots, \xi_n), \boldsymbol{\eta}^\top = (\eta_1, \ldots, \eta_n)$ be two random vectors. Let $\boldsymbol{\eta} = A\boldsymbol{\xi}$ for $A \in \mathbb{R}^{n\times n}$. Then*

$$E(\boldsymbol{\eta}) = AE(\boldsymbol{\xi}), \quad Cov(\boldsymbol{\eta}, \boldsymbol{\eta}) = ACov(\boldsymbol{\xi}, \boldsymbol{\xi})A^\top.$$

Lemma 1.3.17 *Let ξ and η be two independent random variables. Let $f(x)$ and $g(x)$ be two continuous or piecewise continuous functions. Then $f(\xi)$ and $g(\eta)$ are independent with each other. (The conclusion is intuitively correct, however, the rigorours proof is out of the scope of the course. One can refer to probability theory by fudan university 1979 for reference.)*

Theorem 1.3.18 *Let $\xi_1, \ldots, \xi_n$ $(n > 1)$ be random samples from a normal distribution $N(\mu, \sigma^2)$. The sample mean and sample variance are*

$$\overline{\xi} = \frac{1}{n}\sum_{k=1}^{n} \xi_k, \quad S_n^2 = \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \overline{\xi})^2.$$

Then $\overline{\xi}$ and S_n^2 are independent. Moreover,

$$\frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1).$$

Proof. Let

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix},$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are different by a orthogonal transformation A,

$$\boldsymbol{\eta} = A\boldsymbol{\xi},$$

where A is an orthogonal matrix,

$$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2\cdot 1}} & \frac{-1}{\sqrt{2\cdot 1}} & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{3\cdot 2}} & \frac{1}{\sqrt{3\cdot 2}} & \frac{-2}{\sqrt{3\cdot 2}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{bmatrix}.$$

Notice that A is a specific orthogonal matrix whose all row sums are zeros except for the first row. Based on this transformation,

$$\begin{aligned} \eta_1 &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n} \xi_k = \sqrt{n}\bar{\xi}, \\ \eta_1^2 &= n\bar{\xi}^2. \end{aligned} \tag{1.2}$$

Due to the orthogonality of A, one has

$$\sum_{k=1}^{n} \eta_k^2 = \boldsymbol{\eta}^\top \boldsymbol{\eta} = \boldsymbol{\xi}^\top A^\top A\boldsymbol{\xi} = \boldsymbol{\xi}^\top \boldsymbol{\xi} = \sum_{k=1}^{n} \xi_k^2 = \sum_{k=1}^{n} \left(\xi_k - \bar{\xi}\right)^2 + n\bar{\xi}^2.$$

Substituting (1.2) into above,

$$nS_n^2 = \sum_{k=1}^{n} \left(\xi_k - \bar{\xi}\right)^2 = \sum_{k=1}^{n} \eta_k^2 - n\bar{\xi}^2 = \sum_{k=1}^{n} \eta_k^2 - \eta_1^2 = \sum_{k=2}^{n} \eta_k^2.$$

We now discuss the distribution and independence of random variables $\eta_1, \ldots, \eta_n$. Since $\boldsymbol{\eta} = A\boldsymbol{\xi}$, we know that each η_k is a linear combination of normal $\xi_1, \ldots, \xi_n$. Thus $\eta_1, \ldots, \eta_n$ are all random variables with normal distributions. Moreover, one has

$$Cov(\boldsymbol{\eta}, \boldsymbol{\eta}) = ACov(\boldsymbol{\xi}, \boldsymbol{\xi})A^\top = A\sigma^2 I A^\top = \sigma^2 I,$$

where I is an identity matrix. Thus $\eta_1, \ldots, \eta_n$ are pairwise uncorrelated and thereafter independent due to they are all normally distributed. (The reason is that the joint pdf of $\boldsymbol{\eta}$ can be written in terms of the multiplication of the pdf of each η_i since the covariance of $\boldsymbol{\eta}$ is diagonal). Therefore, $\eta_1^2 = n\bar{\xi}^2$ is independent of $\sum_{k=2}^{n} \eta_k^2 = nS_n^2$, which is the main result for the first part, **$\bar{\xi}$ and S_n^2 are independent with each other**. Here we have used the conclusion from Lemma 1.3.17.

One can compute the expectation of $\boldsymbol{\eta}$,

$$E(\boldsymbol{\eta}) = \begin{pmatrix} E\eta_1 \\ E\eta_2 \\ \vdots \\ E\eta_n \end{pmatrix} = AE(\boldsymbol{\xi}) = \begin{pmatrix} \sqrt{n}\mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Thus each η_i $(i = 2, 3, \dots, n)$ has the normal distribution $N(0, \sigma^2)$. Then based on the definition of χ^2 distribution, one has

$$\frac{nS_n^2}{\sigma^2} = \frac{\sum_{k=2}^n \eta_k^2}{\sigma^2} = \sum_{k=2}^n \left(\frac{\eta_k}{\sigma}\right)^2 \sim \chi^2(n-1),$$

satisfying the χ^2 distribution with $n-1$ degrees of freedom. ■

We realize that the above proof is very tricky for the construction of orthogonal matrix A. Next, we provide another way to prove the above independence result and find the distribution of $\left(\overline{X}, S_n^2\right)$. We hope to have an intuitive understanding. Let $X_1, \dots, X_n$ $(n > 1)$ be independent random samples from a normal distribution $N(0, \sigma^2)$,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Here *all variables are shifted to the zero mean.* The sample mean and sample variance are

$$\overline{X} = \frac{1}{n}\sum_{k=1}^n X_k, \quad S_n^2 = \frac{1}{n}\sum_{k=1}^n (X_k - \overline{X})^2,$$

and their observation values are $\overline{x}$ and s^2, respectively.

Proof. Let $f(x_1, \dots, x_n)$ be the density of the n-dimensional random variable $(X_1, \dots, X_n)$. We call the expression

$$dP = f(x_1, \dots, x_n) dx_1 \dots dx_n,$$

the probability element of this random variable. Since $X_1, \dots, X_n$ are independent, we obtain

$$\begin{aligned} dP &= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^n x_k^2\right) dx_1 \dots dx_n \\ &= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{n\overline{x}^2 + ns^2}{2\sigma^2}\right) dx_1 \dots dx_n. \end{aligned} \tag{1.3}$$

Let us make the following transformation:

$$x_k = \overline{x} + sz_k, \quad (k = 1, \dots, n). \tag{1.4}$$

From the constraint relations

$$\sum_{k=1}^n x_k = n\overline{x}, \quad \sum_{k=1}^n x_k^2 = n\overline{x}^2 + ns^2,$$

we obtain two constraint relations for the variable z_k,

$$\sum_{k=1}^n z_k = 0, \quad \sum_{k=1}^n z_k^2 = n. \tag{1.5}$$

Thus, all z_k's are NOT independent and two variables among all the z_k, say z_n and z_{n-1}, are functions of the remaining z_k,

$$z_{n-1}, z_n(z_1, \ldots, z_{n-2}).$$

By solving (1.5) for z_n and z_{n-1}, either

$$z_{n-1} = \frac{A-B}{2}, \quad z_n = \frac{A+B}{2},$$

or $z_{n-1} = \frac{A+B}{2}, \quad z_n = \frac{A-B}{2}$, where

$$A = -\sum_{k=1}^{n-2} z_k, \quad B = \sqrt{2n - 3\sum_{k=1}^{n-2} z_k^2 - \sum_{\substack{k,j=1 \\ k\neq j}}^{n-2} z_k z_j}.$$

Hence transformation in (1.4) is not one-to-one for all z_k, but to

$$(x_1, \ldots, x_n) \leftrightarrow (\overline{x}, s, z_1, \ldots, z_{n-2}),$$

where $s > 0, \sum_{k=1}^{n-2} z_k \neq 0$, and $\sum_{k=1}^{n-2} z_k^2 < n$, there correspond two systems $(x_1, \ldots, x_n)$, namely, the system

$$\begin{aligned} x_k &= \overline{x} + s z_k, \quad (k = 1, 2, \ldots, n-2), \\ x_{n-1} &= \overline{x} + s\frac{A-B}{2}, \quad x_n = \overline{x} + s\frac{A+B}{2}, \end{aligned} \tag{1.6}$$

and the system

$$\begin{aligned} x_k &= \overline{x} + s z_k, \quad (k = 1, 2, \ldots, n-2), \\ x_{n-1} &= \overline{x} + s\frac{A+B}{2}, \quad x_n = \overline{x} + s\frac{A-B}{2}. \end{aligned} \tag{1.7}$$

One can see that thw absolute values of the Jacobians of the two transformations are equal. For the transformation (1.6), we have

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial \overline{x}} & \cdots & \frac{\partial x_n}{\partial \overline{x}} \\ \frac{\partial x_1}{\partial s} & \cdots & \frac{\partial x_n}{\partial s} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial z_{n-2}} & \cdots & \frac{\partial x_n}{\partial z_{n-2}} \end{vmatrix} = \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 & 1 \\ z_1 & z_2 & \cdots & z_{n-2} & \frac{A-B}{2} & \frac{A+B}{2} \\ s & 0 & \cdots & 0 & \frac{s}{2}\left(\frac{\partial A}{\partial z_1} - \frac{\partial B}{\partial z_1}\right) & \frac{s}{2}\left(\frac{\partial A}{\partial z_1} + \frac{\partial B}{\partial z_1}\right) \\ 0 & \ddots & \ddots & 0 & \vdots & \vdots \\ 0 & \cdots & 0 & s & \frac{s}{2}\left(\frac{\partial A}{\partial z_{n-2}} - \frac{\partial B}{\partial z_{n-2}}\right) & \frac{s}{2}\left(\frac{\partial A}{\partial z_{n-2}} + \frac{\partial B}{\partial z_{n-2}}\right) \end{vmatrix}.$$

After a few computations we obtain the absolute value of the Jacobian,

$$|J| = k s^{n-2},$$

where $k = k(z_1, \ldots, z_{n-2})$ is some complicated function independent of $\overline{x}$ and s. Notice that the density has the same value for both transformations (1.6) and (1.7), and we obtain from (1.3) that

$$dP = 2\frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{n\overline{x}^2 + ns^2}{2\sigma^2}\right) k s^{n-2} d\overline{x} ds dz_1 \ldots dz_{n-2}.$$

Let us now represent the above formula in the following form:

$$\begin{aligned} dP &= \frac{\sqrt{n}}{\sigma (2\pi)^{1/2}} \exp\left(-\frac{n\overline{x}^2}{2\sigma^2}\right) d\overline{x} \times \frac{n^{(n-1)/2} s^{n-2} \exp\left(-\frac{ns^2}{2\sigma^2}\right)}{2^{(n-3)/2}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}} ds \\ &\quad \times \frac{\Gamma\left(\frac{n-1}{2}\right)}{n^{n/2}\pi^{(n-1)/2}} k(z_1, \ldots, z_{n-2}) dz_1 \ldots dz_{n-2}. \end{aligned}$$

Then the probability element of $(\overline{X}, S_n, Z_1, \ldots, Z_{n-2})$ is product of three factors, the first of which is the probability element of $\overline{X}$, the second is the probability element of S_n, and the third is the probability element of $(Z_1, \ldots, Z_{n-2})$. Hence one can see from the joint pdf that $\overline{X}, S_n$**, and** $(Z_1, \ldots, Z_{n-2})$ **are independent**. If we denote by $h(\overline{x}, s)$ the density of $\left(\overline{X}, S_n\right)$ we have

$$h(\overline{x}, s) = \begin{cases} \frac{\sqrt{n}}{\sigma(2\pi)^{1/2}} \exp\left(-\frac{n\overline{x}^2}{2\sigma^2}\right) \cdot \frac{n^{(n-1)/2} s^{n-2} \exp\left(-\frac{ns^2}{2\sigma^2}\right)}{2^{(n-3)/2}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}}, & s \geq 0, \\ 0, & s < 0. \end{cases}$$

We now find the distribution of the statistic $Z = nS_n^2/\sigma^2$, denote by the density function $f(z)$, $(2s\frac{n}{\sigma^2}ds = dz,\ z = ns^2/\sigma^2)$

$$\begin{aligned} f(z) &= \frac{n^{(n-1)/2} s^{n-2} \exp\left(-\frac{ns^2}{2\sigma^2}\right)}{2^{(n-3)/2}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}} \cdot \frac{\sigma^2}{2sn} = \frac{n^{(n-3)/2} s^{n-3} \exp\left(-\frac{ns^2}{2\sigma^2}\right)}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-3}} \\ &= \frac{z^{(n-3)/2} \exp\left(-\frac{z}{2}\right)}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)}. \end{aligned}$$

Let us compare $f(z)$ with the $\chi^2(n-1)$ distribution in Definition 1.1,

$$f(x) = \frac{1}{2^{(n-1)/2}\Gamma((n-1)/2)} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}},$$

and we find that they match each other exactly. Thus **the statistic** $Z = nS_n^2/\sigma^2$ **has the distribution as** $\chi^2(n-1)$ **with** $n-1$ **degrees of freedom**. This agrees with our intuition since

$$nS_n^2 = \sum_{k=1}^{n} (X_k - \overline{X})^2,$$

is a sum of n random variables satisfying the constraint relation

$$\sum_{k=1}^{n} X_k = n\overline{X}.$$

This result gives us a better understanding of the notion of the number of degrees of freedom. Hence

$$\begin{aligned} E(nS_n^2) &= (n-1)\,\sigma^2, \quad Var(nS_n^2) = 2\,(n-1)\,\sigma^4, \\ E(S_n^2) &= \frac{n-1}{n}\sigma^2, \quad Var(S_n^2) = \frac{2\,(n-1)}{n^2}\sigma^4. \end{aligned}$$

■

Remark 1.3.19 *If the independent random variables X_k have the same normal distribution, then the joint density of the random variables $\overline{X}$ and S_n is the product of the densities of these random variable, and hence they are independent. This extremely important and interesting result was obtained by Fisher [5] in Fiesz book.*

Remark 1.3.20 *The converse theorem is also true. If the statistics $\overline{X}$ and S_n are independent, the random variables X_k have the normal distribution. The proof of this theorem was given by Geary [1], Lukacs [1], Kawata and Sakamoto [1], and Zinger [1]. Later this theorem was generalized by Lukacs [2] and Basu and Laha [1].*

1.3.3 Student's t distribution

Definition

（商的分布） Let us first consider the distribution of the Quotient.

Let (X, Y) be a two-dimensional random variable with pdf $p(x, y)$. Find the distribution of $Z = X/Y$. The distribution function of Z is

$$\begin{aligned} F_Z(z) &= P(Z \le z) = P(\frac{X}{Y} \le z) \\ &= \iint_{\frac{x}{y} \le z} p(x, y)dxdy = \iint_{\frac{x}{y} \le z, y>0} + \iint_{\frac{x}{y} \le z, y<0} p(x, y)dxdy \\ &= \int_0^{\infty} \left(\int_{-\infty}^{yz} p(x, y)dx \right) dy + \int_{-\infty}^{0} \left(\int_{yz}^{\infty} p(x, y)dx \right) dy. \end{aligned}$$

Then the pdf of Z is

$$\begin{aligned} p_Z(z) &= F_Z'(z) = \int_0^{\infty} yp(yz, y)dy - \int_{-\infty}^{0} yp(yz, y)dy \\ &= \int_{-\infty}^{\infty} |y|\, p(yz, y)dy. \end{aligned}$$

Definition 1.3.21 *Let $\xi \sim N(0, 1)$ and $\eta \sim \chi^2(n-1)$, and ξ and η are independent. Then*

$$T = \frac{\xi}{\sqrt{\eta/(n-1)}},$$

is a student's t distribution with $(n-1)$ degrees of freedom, denoted by $T \sim t(n-1)$. Its pdf is given by

$$\begin{aligned} p_T(y) &= \frac{\Gamma(\frac{n}{2})}{\sqrt{n-1}\sqrt{\pi}\Gamma(\frac{n-1}{2})} \left(1 + \frac{y^2}{n-1}\right)^{-\frac{n}{2}}, \quad -\infty < y < \infty, \\ &= \frac{1}{\sqrt{n-1}B(\frac{1}{2}, \frac{n-1}{2})} \frac{1}{\left(1 + \frac{y^2}{n-1}\right)^{\frac{n}{2}}}. \end{aligned}$$

Proof. We now compute the pdf of student's t distribution. The pdf for $\chi^2(n-1)$ is,

$$p(x) = \begin{cases} \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

We now compute the pdf for $\sqrt{\frac{\chi^2(n-1)}{n-1}}$. Let $y = \sqrt{\frac{x}{n-1}}$ and the inverse function $h(y) = y^2(n-1)$. Using the result in Lemma 1.3.1, we obtain the pdf for $\sqrt{\frac{\chi^2(n-1)}{n-1}}$,

$$\begin{aligned} p_2(y) &= \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} \left(y^2(n-1)\right)^{\frac{n-1}{2}-1} e^{-\frac{y^2(n-1)}{2}} 2y(n-1) \\ &= \begin{cases} \frac{2(n-1)^{\frac{n-1}{2}} y^{n-2}}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} e^{-\frac{y^2(n-1)}{2}} = \frac{2(\frac{n-1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} y^{n-2} e^{-\frac{y^2(n-1)}{2}}, & y \ge 0, \\ 0, & y < 0. \end{cases} \end{aligned}$$

Using the quotient formula, we have

$$p_T(y) = \int_0^{\infty} x_2 p(x_2 y, x_2)dx_2,$$

where $p(x,y)$ is the joint pdf of a standard Gaussian and $\sqrt{\frac{\chi^2(n-1)}{n-1}}$,

$$p(x_1,x_2) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} p_2(x_2), & x_2 \geq 0, \\ 0, & x_2 < 0. \end{cases}$$

Then

$$\begin{aligned} p_T(y) &= \int_0^\infty x \frac{1}{\sqrt{2\pi}} e^{-\frac{(xy)^2}{2}} \cdot \frac{2(\frac{n-1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} x^{n-2} e^{-\frac{x^2(n-1)}{2}} dx \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{2(\frac{n-1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} x^{n-1} e^{-\frac{1}{2}x^2(y^2+n-1)} dx \\ \overset{\substack{x^2(y^2+n-1)=s \\ 2x(y^2+n-1)dx = ds}}{=} & \int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{2(\frac{n-1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \left(\frac{s}{y^2+n-1} \right)^{\frac{n-1}{2}} e^{-\frac{s}{2}} \frac{ds}{2\sqrt{\frac{s}{y^2+n-1}}(y^2+n-1)} \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{(\frac{n-1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \frac{1}{(y^2+n-1)^{\frac{n-1}{2}}} \frac{1}{\sqrt{y^2+n-1}} s^{\frac{n-1}{2}} s^{-\frac{1}{2}} e^{-\frac{s}{2}} ds \\ &= \frac{1}{\sqrt{2\pi}} \frac{(\frac{n-1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \left(y^2+n-1\right)^{-\frac{n}{2}} \int_0^\infty s^{\frac{n}{2}-1} e^{-\frac{s}{2}} ds \end{aligned}$$

Using the fact that

$$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^\infty s^{\frac{n}{2}-1} e^{-\frac{s}{2}} ds = 1,$$

then we obtain that

$$\begin{aligned} p_T(y) &= \frac{1}{\sqrt{2\pi}} \frac{(\frac{n-1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \left(y^2+n-1\right)^{-\frac{n}{2}} 2^{\frac{n}{2}} \Gamma(\frac{n}{2}) \\ &= \frac{1}{\sqrt{2\pi}} \frac{2^{-\frac{n-1}{2}} 2^{\frac{n}{2}} \Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} (n-1)^{\frac{n-1}{2}} \left(y^2+n-1\right)^{-\frac{n}{2}} \\ &= \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \frac{1}{\sqrt{n-1}} \left(1 + \frac{y^2}{n-1}\right)^{-\frac{n}{2}}. \end{aligned}$$

■

Motivation and Properties

We now provide the motivation for the t distribution. We have considered the distribution of the statistic $\overline{X}$ given by

$$\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k,$$

where the random variables X_k ($k = 1, \ldots, n$) are independent and have the same normal distribution $N(\mu, \sigma^2)$. We established that $\overline{X}$ has the normal distribution $N(\mu, \sigma^2/n)$; hence for a known μ and **unknown** σ^2, the distribution of $\overline{X}$ is unknown. Of course, we cannot directly replace σ^2 by the value obtained from a sample variance since the sample variance itself is a random variable and can take on different values in different samples. **In order to deduce anything about μ without the knowledge of σ^2, we have to consider a statistic which is a function of μ and with a distribution independent of σ^2.**

This problem was solved by Gosset (pseudonym: Student [1] in Fiesz), who introduced the statistic called **Student's t-statistic**.

Let X_k $(k = 1, \ldots, n)$ be independent and have the same normal distribution $N(\mu, \sigma^2)$. Then Student's t is defined by

$$T = \frac{\overline{X} - \mu}{S_n}\sqrt{n-1} = \frac{\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{nS_n^2}{(n-1)\sigma^2}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2(n-1)\sigma^2}{(n-1)\sigma^2}}}, \tag{1.8}$$

where

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad S_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2.$$

In fact, here we need to further **verify that $\overline{X}$ and S_n^2 are independent which will be shown later**. Thus the density of Student's t is independent of σ^2. As we have already mentioned, it is this fact that makes possible many applications of the t-distribution. We say that a random variable with density of T in (1.8) has **the t-distribution with $n-1$ degrees of freedom**.

Proposition 1.3.22 *The density of the random variable t is symmetric with respect to $t = 0$.*

Proposition 1.3.23 *Student's t-distribution has only moments of order $k < n - 1$. Thus, for $n = 2$ no moments exist. The reader can verify that for $n = 2$, Student's t-distribution is a particular case of the Cauchy distribution which, as we know, has no moments.*

The graph of the density of Student's t-distribution is shown in Fig. 1.3. If we compare the Student's t-distribution for a number of degrees of freedom close to 30 with the normal distribution $N(0,1)$, we see that they are almost identical (see Fig. 1.3). This is because the t-distribution approaches the normal distribution rapidly, as the number of degrees of freedom tends to infinity. We now prove this.

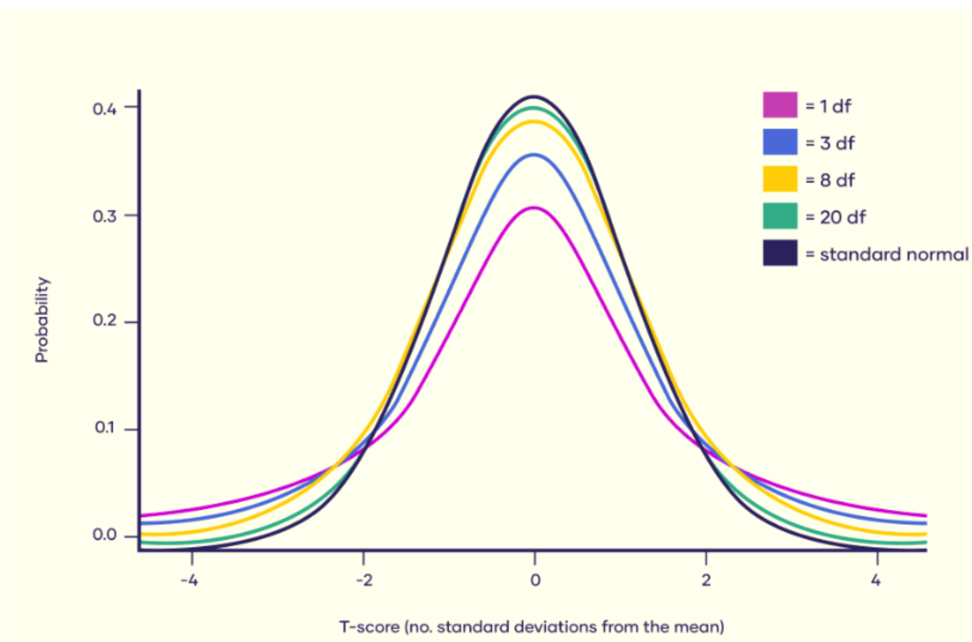

Figure 1.3: Comparison of Student's t-distribution for different n.

Lemma 1.3.24 *Let $\{X_n\}\,(n = 1, 2, \dots)$ be an arbitrary sequence of random variables (dependent or not) and let the corresponding sequence of distribution functions $\{F_n(x)\}$ converge as $n \to \infty$ to the CDF $F(x)$. Further, let $\{Y_n\}\,(n = 1, 2, \dots)$ be another sequence of random variables stochastically convergent (in probability) to a constant a. Then*
(1) the sequence of the CDFs of $X_n + Y_n$ converges to the CDF $F(x - a)$.
(2) the sequence of the CDFs of $X_n - Y_n$ converges to the CDF $F(x + a)$.
(3) the sequence of the CDFs of $X_n Y_n$ converges to the CDF $F(x/a)$ if $a > 0$ and to the CDF $1 - F(x/a)$ if $a < 0$.
(4) the sequence of the CDFs of X_n/Y_n converges to the CDF $F(ax)$ if $a > 0$ and to the CDF $1 - F(ax)$ if $a < 0$.

Theorem 1.3.25 *The sequence $\{F_n(t)\}$ of distribution functions of Student's t with n degrees of freedom satisfies for every t the relation*

$$\lim_{n\to\infty} F_n(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{s^2}{2}}\, ds.$$

Proof. Let us write the t-distribution in the form

$$T_n = \frac{Y_n}{\sqrt{\frac{X_n}{n}}} = \frac{Y_n}{\sqrt{Z_n}} = \frac{Y_n}{V_n},$$

where for every n the random variable Y_n has the normal distribution $N(0,1)$ and X_n has the $\chi^2(n)$ with n degrees of freedom. We shall prove that the sequence $\{V_n\}$ converges stochastically to the constant one.

Recall that for a gamma distribution with pdf $\Gamma(\alpha, \lambda)$,

$$p(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \ge 0, \\ 0, & x < 0, \end{cases}$$

the corresponding Ch.f. is

$$\phi_\Gamma(t) = \frac{1}{(1 - it/\lambda)^\alpha}.$$

The random variable X_n has the distribution of $\chi^2(n)$, which is the gamma distribution with $\Gamma(\frac{n}{2}, \frac{1}{2})$. Hence the Ch.f. of X_n is $\frac{1}{(1-2it)^{\frac{n}{2}}}$ and then the Ch.f. of $Z_n = \frac{X_n}{n}$ is

$$\phi_n(t) = \frac{1}{(1 - \frac{2it}{n})^{\frac{n}{2}}}.$$

Hence

$$\lim_{n\to\infty} \phi_n(t) = e^{it}.$$

It follows from the last equation that the sequence $\{Z_n\}$ is stochastically convergent to the distribution of $P(X = 1) = 1$ (Two points here. First, $\phi(t) = \sum_k P(X = x_k) e^{itx_k} = e^{it} \cdot 1$. Second, the convergence in distribution to a constant is equivalent to the stochastic convergence in probability to that constant.). By continuous mapping theorem, the sequence $\{V_n\}$ is also stochastically convergent to one. From Lemma 1.3.24, we arrive at the conclusion. ■

Applications

Here we show one example for Student's t-distribution.

Example 1.3.26 *We now introduce one statistic of great importance in applications. Let* $X_1, \ldots, X_{n_1} \sim N(m_1, \sigma^2)$ *and* $Y_1, \ldots, Y_{n_2} \sim N(m_2, \sigma^2)$ *be independent random variables with the same variance but different expectation. Let*

$$\begin{aligned} \overline{X} &= \frac{1}{n_1}\sum_{k=1}^{n_1} X_k, \quad \overline{Y} = \frac{1}{n_2}\sum_{l=1}^{n_2} Y_l, \\ S_{n,1}^2 &= \frac{1}{n_1}\sum_{k=1}^{n_1}(X_k - \overline{X})^2, \quad S_{n,2}^2 = \frac{1}{n_2}\sum_{l=1}^{n_2}(Y_l - \overline{Y})^2. \end{aligned}$$

As we know, $\overline{X}$ *and* $\overline{Y}$ *have, respectively, the normal distributions*

$$N(m_1, \frac{\sigma^2}{n_1}) \text{ and } N(m_2, \frac{\sigma^2}{n_2}).$$

Hence the random variable $(\overline{X} - \overline{Y} - (m_1 - m_2))/\sigma$ *has the distribution*

$$N(0, \frac{n_1 + n_2}{n_1 n_2}).$$

It follows that

$$Z = \frac{\overline{X} - \overline{Y} - (m_1 - m_2)}{\sigma}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

has the distribution $N(0,1)$. *Moreover, the random variable*

$$W = \frac{n_1 S_{n,1}^2 + n_2 S_{n,2}^2}{\sigma^2}$$

has the χ^2 *distribution with* $n_1 + n_2 - 2$ *degrees of freedom, which follows from the addition theorem for* χ^2, *since* $S_{n,1}^2$ *and* $S_{n,2}^2$ *are independent.*

Let us consider the random variable U *defined as*

$$U = \frac{Z}{\sqrt{\frac{W}{n_1+n_2-2}}} = \frac{\frac{\overline{X}-\overline{Y}-(m_1-m_2)}{\sigma}\sqrt{\frac{n_1 n_2}{n_1+n_2}}}{\sqrt{\frac{n_1 S_{n,1}^2 + n_2 S_{n,2}^2}{\sigma^2(n_1+n_2-2)}}} = \frac{\left[\overline{X} - \overline{Y} - (m_1 - m_2)\right]}{\sqrt{n_1 S_{n,1}^2 + n_2 S_{n,2}^2}}\sqrt{\frac{n_1 n_2\,(n_1 + n_2 - 2)}{n_1 + n_2}},$$

where Z *and* W *are defined above. We see that* U *has Student's* t*-distribution with* $n_1 + n_2 - 2$ *degrees of freedom. Thus the distribution of* U *is independent of* m_1, m_2 *and* σ. *This result was obtained by Fisher [5].*

Remark 1.3.27 *The generalization of Student's* t *to multi-dimensional random variables is Hotelling's* T^2 *(Hotelling [2] in Fiesz book).*

1.3.4 Fisher's Z distribution or F distribution

Historical notes

The leading statistician R. A. Fisher (1890-1962) invented analysis of variance. According to Miller, he tabulated probabilities not for what is now called an F statistic but for $z = (\log F)/2$. Snedecor (1934) gave the name "F distribution" in honor of Fisher and tabulated F itself.

From the Wikipedia article on Fisher, he worked at the Rothamsted Experimental Station in England from 1919 to 1933, designing and performing agricultural experiments and analyzing them statistically. He published books and became a world-famous statistician during that time. In 1931 and later in 1936 he visited Iowa State College in Ames, Iowa, where there was also great interest in agricultural experiments and statistics. Fisher met Snedecor there.

According to Wikipedia on Snedecor, the book by Snedecor and Cochran (1937 and later editions) was at least for a while the most often cited book in all of science. On "F-distribution" it's said sometimes to be called "Snedecor's F distribution" or the "Fisher-Snedecor distribution," but "F distribution" is the standard terminology in textbooks.

Snedecor lived from 1881 to 1974 and Cochran from 1909 to 1980. So even the 8th edition of their book (1989) was posthumous for both of them.

Definition

Definition 1.3.28 *Let $\xi \sim \chi^2(n), \eta \sim \chi^2(m)$, and ξ, η are independent. Then*

$$F = \frac{\xi/n}{\eta/m},$$

is F distribution with degrees of freedom of (n, m), denoted by $F \sim F(n, m)$. The pdf is given by

$$p_F(y) = \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{y^{\frac{n}{2}-1}}{(ny+m)^{\frac{m+n}{2}}}.$$

Definition 1.3.29 *If $F \sim F(n, m)$, then $\frac{1}{F} \sim F(m, n)$.*

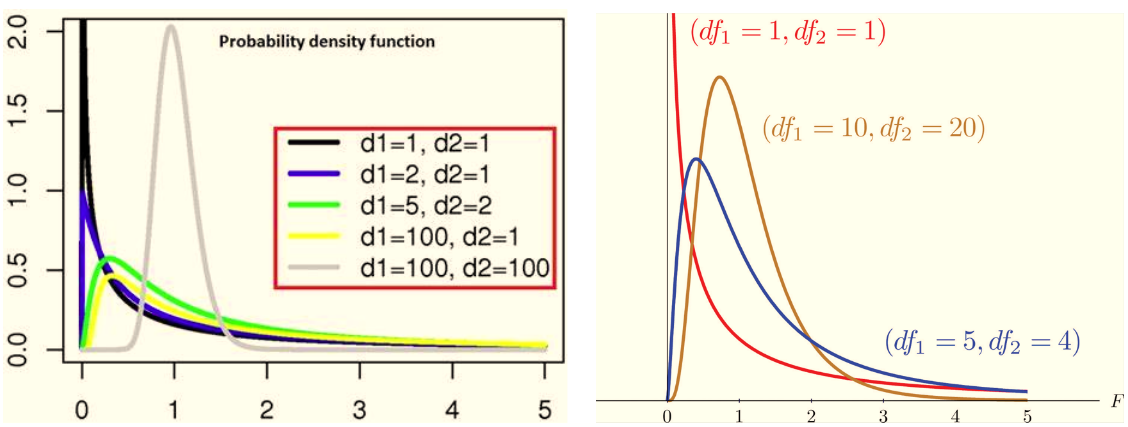

Figure 1.4: Comparison of F distribution for different (n, m).

Proof. We now compute the pdf of $F(n, m)$ distribution. The pdf of ξ/n is

$$p_{\xi/n}(x) = \begin{cases} \frac{n}{2^{n/2}\Gamma(n/2)} (nx)^{\frac{n}{2}-1} e^{-\frac{nx}{2}}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The pdf of η/m is

$$p_{\eta/m}(x) = \begin{cases} \frac{m}{2^{m/2}\Gamma(m/2)} (mx)^{\frac{m}{2}-1} e^{-\frac{mx}{2}}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Then the pdf of F distribution becomes

$$\begin{aligned}
p_F(y) &= \int_{-\infty}^{\infty} |x|\, p(yx,x)dx = \int_0^{\infty} xp(yx,x)dx = \int_0^{\infty} xp_{\xi/n}(yx)p_{\eta/m}(x)dx \\
&= \int_0^{\infty} x \cdot \frac{n}{2^{n/2}\Gamma(n/2)}\,(nyx)^{\frac{n}{2}-1}\, e^{-\frac{nyx}{2}} \cdot \frac{m}{2^{m/2}\Gamma(m/2)}\,(mx)^{\frac{m}{2}-1}\, e^{-\frac{mx}{2}}\,dx \\
&= \frac{n}{2^{n/2}\Gamma(n/2)} \frac{m}{2^{m/2}\Gamma(m/2)}\,(ny)^{\frac{n}{2}-1}\,(m)^{\frac{m}{2}-1} \int_0^{\infty} x\,(x)^{\frac{n}{2}-1}\,(x)^{\frac{m}{2}-1}\, e^{-\frac{nyx}{2}-\frac{mx}{2}}\,dx \\
&= \frac{n^{n/2}m^{m/2}}{2^{n/2+m/2}\Gamma(n/2)\Gamma(m/2)}\,(y)^{\frac{n}{2}-1} \int_0^{\infty} x^{\frac{n}{2}+\frac{m}{2}-1} e^{-\frac{1}{2}(ny+m)x} dx.
\end{aligned}$$

Let $x(ny+m)=t$, then

$$\begin{aligned}
p_F(y) &= \frac{n^{n/2}m^{m/2}}{2^{n/2+m/2}\Gamma(n/2)\Gamma(m/2)} y^{\frac{n}{2}-1} \int_0^{\infty} \left(\frac{t}{ny+m}\right)^{\frac{n+m}{2}-1} e^{-\frac{t}{2}}\,\frac{dt}{ny+m} \\
&= \frac{n^{n/2}m^{m/2}}{2^{n/2+m/2}\Gamma(n/2)\Gamma(m/2)} \frac{y^{\frac{n}{2}-1}}{(ny+m)^{\frac{n+m}{2}}} \int_0^{\infty} t^{\frac{n+m}{2}-1} e^{-\frac{t}{2}} dt.
\end{aligned}$$

Notice that the following is pdf of $\chi^2(\frac{n+m}{2})$,

$$\frac{1}{2^{(n+m)/2}\Gamma((n+m)/2)} \int_0^{\infty} t^{\frac{n+m}{2}-1} e^{-\frac{t}{2}} dt = 1.$$

Then the $p_F(y)$ becoms

$$\begin{aligned}
p_F(y) &= \frac{n^{n/2}m^{m/2}}{2^{n/2+m/2}\Gamma(n/2)\Gamma(m/2)} \frac{y^{\frac{n}{2}-1}}{(ny+m)^{\frac{n+m}{2}}} 2^{(n+m)/2}\Gamma(\frac{n+m}{2}) \\
&= \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{y^{\frac{n}{2}-1}}{(ny+m)^{\frac{n+m}{2}}}.
\end{aligned}$$

■

Applications

Example 1.3.30 *Let $X_1, \ldots, X_{n_1} \sim N(m_1, \sigma_1^2)$ and $Y_1, \ldots, Y_{n_2} \sim N(m_2, \sigma_2^2)$ be independent random variables with normal distributions. Let*

$$\begin{aligned}
\overline{X} &= \frac{1}{n_1}\sum_{k=1}^{n_1} X_k, \quad \overline{Y} = \frac{1}{n_2}\sum_{l=1}^{n_2} Y_l, \\
S_{n,1}^2 &= \frac{1}{n_1}\sum_{k=1}^{n_1} (X_k - \overline{X})^2, \quad S_{n,2}^2 = \frac{1}{n_2}\sum_{l=1}^{n_2} (Y_l - \overline{Y})^2.
\end{aligned}$$

Then when σ_1^2 and σ_2^2 are known,

$$F = \frac{\frac{n_1 S_{n,1}^2}{(n_1-1)\sigma_1^2}}{\frac{n_2 S_{n,2}^2}{(n_2-1)\sigma_2^2}},$$

is a statistic with $F(n_1-1, n_2-1)$ distribution. In particular, when σ_1^2 and σ_2^2 are unknown but $\sigma_1^2 = \sigma_2^2$,

$$F = \frac{\frac{n_1 S_{n,1}^2}{(n_1-1)}}{\frac{n_2 S_{n,2}^2}{(n_2-1)}} = \frac{n_1(n_2-1)S_{n,1}^2}{n_2(n_1-1)S_{n,2}^2},$$

is a statistic with $F(n_1-1, n_2-1)$ distribution.

1.4 Quantile

样本极差，样本中位数，样本α-分位数

Definition 1.4.1 *Let $X_1, \ldots, X_n$ be a random sample from a population whose CDF is $F(x)$. Let $X_{(1)} \leq \ldots \leq X_{(n)}$ be the corresponding order statistics. The sample range is defined to be*

$$R_n = X_{(n)} - X_{(1)},$$

and the sample median is defined to be

$$M_n = \begin{cases} X_{(n/2)}, & \text{if } n \text{ is even,} \\ \left(X_{(\frac{n-1}{2})} + X_{(\frac{n+1}{2})} \right)/2, & \text{if } n \text{ is odd.} \end{cases}$$

For $0 < \alpha < 1$, let x_α be the α-quantile of the population of $F(x)$, if

$$P(X < x_\alpha) = \alpha.$$

The order statistic $m_\alpha = X_{(k)} + [(n+1)\alpha - k][X_{(k+1)} - X_{(k)}]$ is called sample α-quantile, $k = \lfloor (n+1)\alpha \rfloor$. Here, $\lfloor \cdot \rfloor$ means the floor integer function of the enclosed number. (see Zongshu Wei and https://mathworld.wolfram.com/Quantile.html)

In real applications, the order statistic of some population is very complicated, so that we would like to know the limit distribution of the sample α-quantile as $n \to \infty$. We show the following theorem without justification. The proof can be found in reference in Wei Zongshu.

Theorem 1.4.2 *Suppose $X_1, \ldots, X_n$ be a random sample from a population X with pdf $f(x)$. For $0 < \alpha < 1$, let x_α be the α-quantile of X. If $f(x_\alpha) > 0$ and $f(x)$ is continuous at $x = x_\alpha$, then the following statements hold for the sample α-quantile m_α as $n \to \infty$:*
(1) m_α converges to x_α in probability;
(2) m_α has an asymptotic normal distribution (converges in distribution to a normal distribution)

$$N\left(x_\alpha, \frac{1}{f^2(x_\alpha)} \frac{\alpha(1-\alpha)}{n}\right).$$