第7章　数理统计的基本概念

第7章　数理统计的基本概念

在概率论中，我们知道，随机变量的概率分布（分布函数、分布律、密度函数等）完整地描述了随机变量的统计规律性．在概率论的许多问题中，常常假定概率分布是已知的，而一切有关的计算与推理均基于这个已知的概率分布．但在实际问题中，情况并非如此．

引例 若从一批合格率为 p 的产品中随机重复抽取 10 件来检查，用 X 表示所取 10 件产品中的合格品数，则 X 服从二项分布 $B(10, p)$．显然，p 的大小决定了该批产品的质量，因此，人们会对未知的参数 p 提出一些问题，比如：

（1）p 的大小如何．

（2）p 大概在什么范围内．

（3）能否认为 p 满足规定要求（如 $p \geq 0.90$）．

数理统计主要研究两类问题

1. 试验的设计和研究
   研究如何合理有效地获得数据资料的方法，并对这些方法进行分析

2. 统计推断
   研究如何利用获得的数据资料对所关心的问题作出尽可能精确、可靠的判断

总体与个体

定义　在数理统计中,研究对象的全体称为**总体**(Collectivity),把组成总体的每个基本单元称为**个体**. 若总体中包含有限个个体，称为**有限总体**；若总体包含无限个个体，称为**无限总体**.

样本

定义　从总体中独立地随机地抽样称为简单随机抽样；这样抽得的 n 个个体称为一个**简单随机样本**(Simple random sample)，记为 $(X_1, X_2, \cdots, X_n)$ 或 $X_1, X_2, \cdots, X_n$．其观测值记为 $(x_1, x_2, \cdots, x_n)$ 或 $x_1, x_2, \cdots, x_n$． n 称为**样本容量**．一个简单随机样本与其观测值，常统一简称为一个**样本**．样本中的个体称为**样品**.

例 **测量物体长度X, 如果在相同条件下, 独立测量n次, 测量结果为(样本)$X_1,X_2,\ldots,X_n$. 则$X_1, X_2, \ldots, X_n$相互独立且与总体X有相同分布;**

一般地, 在相同条件下对总体X进行n次重复的独立观测, 获得的样本$X_1,X_2,\ldots,X_n$是n个相互独立的且与总体X同分布的随机变量.

$X_1, X_2,\ldots, X_n$——一组样本或一个样本.

n——样本大小或样本容量.

$x_1, x_2,\ldots, x_n$——样本$X_1,X_2,\ldots,X_n$的观察值。

注：样本指某次抽样的观测值$x_1, x_2,\ldots, x_n$，也指泛指一次抽出的可能结果(随机变量) $X_1,X_2,\ldots,X_n$.

样本具有二重性：

一方面，由于样本是从总体 X 中随机抽取的，抽取前无法预知它们的数值，因此，样本

$(X_1, X_2, \cdots, X_n)$ 是随机变量；

另一方面，样本在抽取以后经观测就有确定的观测值，因此，样本 $(x_1, x_2, \cdots, x_n)$ 又是一组数值.

简单随机样本具有两条重要性质：

性质 1　样本$(X_1, X_2, \cdots, X_n)$中每个X_i与总体X具有相同的分布.

性质 2　样本$(X_1, X_2, \cdots, X_n)$中的$X_1, X_2, \cdots, X_n$相互独立.

独立同分布（大数定律、中心极限定理）

定理 如果总体 X 的分布函数为 $F(x)$，概率函数为 $p(x)$. 而 $(X_1,X_2,\cdots,X_n)$ 为来自总体 X 的样本，则

（1）样本 $(X_1,X_2,\cdots,X_n)$ 的联合分布函数为：

$$F(x_1,x_2,\cdots,x_n)=F(x_1)F(x_2)\cdots F(x_n)=\prod_{i=1}^{n}F(x_i)$$

（2）样本 $(X_1,X_2,\cdots,X_n)$ 的联合概率函数为：

$$p(x_1,x_2,\cdots,x_n)=p(x_1)p(x_2)\cdots p(x_n)=\prod_{i=1}^{n}p(x_i)$$

例 设总体 X 服从两点分布 $B(1, p)$，其中 $0 < p < 1$．$(X_1, X_2, \cdots, X_n)$ 是来自总体的样本，求 $(X_1, X_2, \cdots, X_n)$ 的联合分布律.

例 设总体 X 服从两点分布 $B(1, p)$，其中 $0 < p < 1$．$(X_1, X_2, \cdots, X_n)$ 是来自总体的样本，求 $(X_1, X_2, \cdots, X_n)$ 的联合分布律.

解 $P\{X = i\} = p^i (1-p)^{1-i}$， $i = 0,1$

$$P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\}$$

$$= P\{X_1 = x_1\} P\{X_2 = x_2\} \cdots P\{X_n = x_n\}$$

$$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}$$

样本分布函数

将n个样本值按大小排成$x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$,对任意实数$x$，称$F_n(x)$为总体$X$的样本分布函数。

$$F_n(x)=\begin{cases}0, & x<x_{(1)},\\ \dfrac{k}{n} & x_{(k)}\le x<x_{(k+1)},\ k=1,2\cdots n-1,\\ 1, & x_{(n)}\le x.\end{cases}$$

定理（Glivenko,1933）：设总体分布函数为$F(x)$，样本分布函数为$F_n(x)$，则有

$$P(\lim_{n\to\infty}\sup_{-\infty<x<+\infty}|F_n(x)-F(x)|=0)=1$$

Valery Ivanovich Glivenko (Ukrainian: Валерій Іванович Гливенко, Russian: Валерий Иванович Гливенко; 2 January 1897 (Gregorian calendar) / 21 December 1896 (Julian calendar) in Kiev – 15 February 1940 in Moscow) was a Ukrainian Soviet mathematician. He worked in foundations of mathematics, real analysis, probability theory, and mathematical statistics. He taught at Moscow Industrial Pedagogical Institute[1] until his death at age 43.[2][3] Most of Glivenko's work was published in French.

样本的数字特征和样本矩

样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ $\qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

样本标准差 $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$

样本**k**阶原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, k = 1, 2, \cdots$

样本**k**阶中心矩 $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k, k = 1, 2\cdots$

定理 $S^2=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\bar{X}^2)$

$$\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}+\bar{X}^2)$$

$$=\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\bar{X}^2$$

$$=\sum_{i=1}^{n}X_i^2-n\bar{X}^2$$

$$S^2=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\bar{X}^2)$$

统计量

定义　设$(X_1, X_2, \cdots, X_n)$为来自总体 X 的一个样本，$(x_1, x_2, \cdots, x_n)$是该样本的观测值．若样本函数 $g(X_1, X_2, \cdots, X_n)$不包含任何未知参数，则称它为一个**统计量**(Statistic)．而 $g(x_1, x_2, \cdots, x_n)$称为**统计量的观测值**.

定理 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，且总体均值 $E(X)=\mu$，总体方差 $D(X)=\sigma^2$，则

$$E(\bar{X}) = \underline{\qquad}$$

$$D(\bar{X}) =$$

$$\underline{\qquad}$$

$$E(S^2) = \underline{\qquad}$$

由样本的独立性、同分布性及数学期望和方差的性质，可得

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

$$D(\bar{X}) = D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i)$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$E(S^2)=E\left\{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right\}=E\left\{\frac{1}{n-1}\sum_{i=1}^{n}[(X_i-\mu)-(\bar{X}-\mu)]^2\right\}$$

$$=E\left\{\frac{1}{n-1}[\sum_{i=1}^{n}(X_i-\mu)^2-n(\bar{X}-\mu)^2]\right\}$$

$$=\frac{1}{n-1}[\sum_{i=1}^{n}E(X_i-\mu)^2-nE(\bar{X}-\mu)^2]$$

$$=\frac{1}{n-1}[\sum_{i=1}^{n}D(X_i)-nD(\bar{X})]=\frac{1}{n-1}(n\sigma^2-n\cdot\frac{\sigma^2}{n})=\sigma^2$$

§2 抽样分布

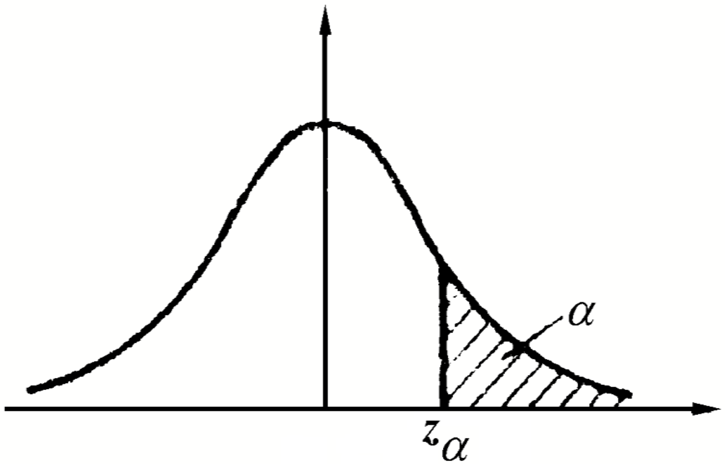
分位点

设统计量U服从某分布，如果对于α $(0<\alpha<1)$有$P(U>u_\alpha)=\alpha$，则称u_α为该分布的上α分位点。

$X \sim N(0,1)$,

要使$P\left(X > Z_\alpha\right) = \alpha$

即$\Phi(Z_\alpha) = 1-\alpha$



$N(0,1)$分布的上侧α分位点图形

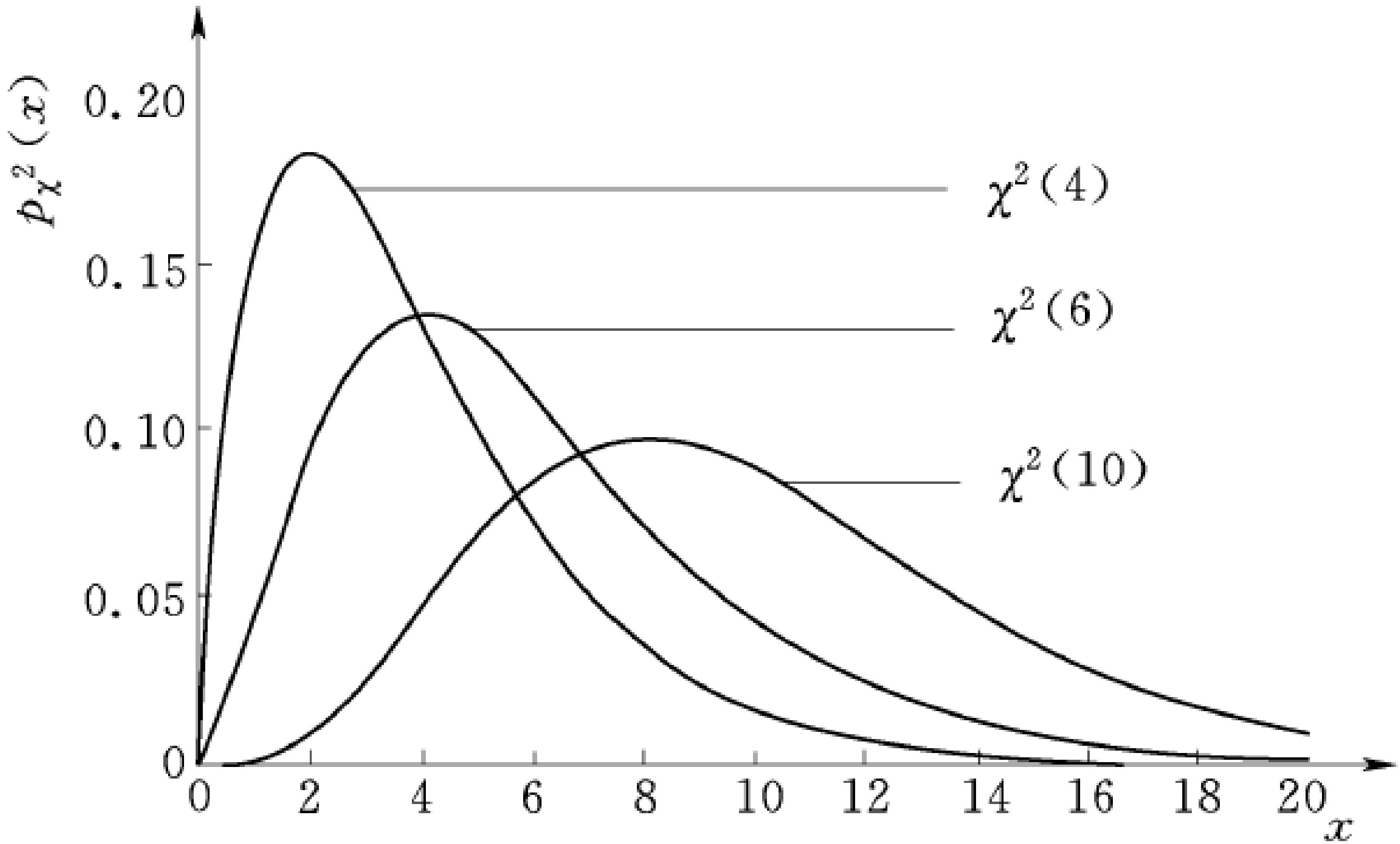
卡方分布

定义　若①每个 $Z_i \sim N(0,1)$；② $Z_1, Z_2, \cdots, Z_n$ 相互独立，则称 $\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ 服从自由度为 n 的 χ^2 分布（χ^2 distribution），记为 $\chi^2 \sim \chi^2(n)$.

密度函数

$$f(x;n) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

特例：n=1　Γ分布；　n=2 指数分布



χ^2 分布密度函数曲线

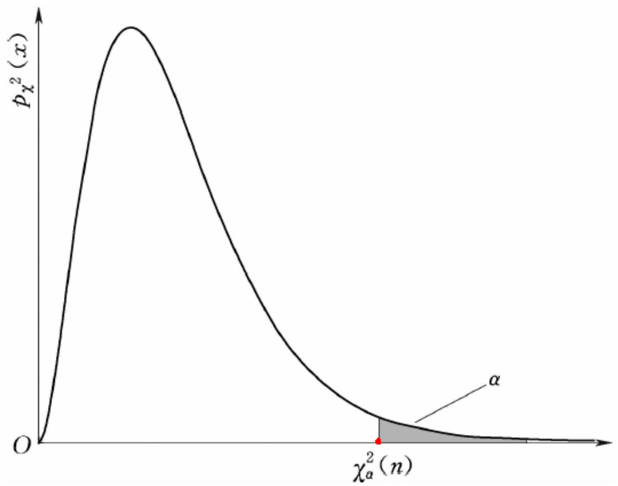
χ^2分布有如下性质：

性质 1 若 $\chi^2 \sim \chi^2(n)$，则有 n 个相互独立的 $Z_i \sim N(0,1), i=1,2,\cdots,n$，使得 $\chi^2 = \underline{Z_1^2 + Z_2^2 + \cdots + Z_n^2}$.

性质 2 若 $\chi^2 \sim \chi^2(n)$，则 $E(\chi^2) = \underline{n}$，$D(\chi^2) = \underline{2n}$.

性质 3 若 $X \sim \chi^2(n)$，$Y \sim \chi^2(m)$，且相互独立，则 $X + Y \sim \underline{\chi^2(n+m)}$.

定义 设$\chi^2 \sim \chi^2(n)$，对给定的α（$0<\alpha<1$），满足$P\{\chi^2 > \chi^2_\alpha(n)\} = \alpha$的点$\chi^2_\alpha(n)$称为$\chi^2$分布的上侧分位点.



χ^2 分布的上侧分位点

$$\chi^2_{0.99}(10) = 2.558$$

$$\chi^2_{0.01}(10) = 23.209$$

$$n > 45\text{时}, \chi^2_\alpha(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$$

Sir Ronald Fisher

Born	17 February 1890 East Finchley, London, England, United Kingdom
Died	29 July 1962 (aged 72) Adelaide, South Australia, Australia
Residence	United Kingdom and Australia
Nationality	British
Alma mater	Gonville and Caius College, Cambridge
Known for	Fisher's principle, Fisher information
Awards	Weldon Memorial Prize (1930)
Scientific career	
Fields	Statistics, Genetics, and Evolutionary biology
Institutions	Rothamsted Experimental Station, University College London, Cambridge University, University of Adelaide, Commonwealth Scientific and Industrial Research Organisation
Academic advisors	James Hopwood Jeans and F. J. M. Stratton
Doctoral students	C. R. Rao, D. J. Finney, and Walter Bodmer [1]

t 分布(Student's t-distribution)

定义　若① $Z \sim N(0,1)$；② $\chi^2 \sim \chi^2(n)$；③ Z 与 χ^2 相互独立，则称 $T = \dfrac{Z}{\sqrt{\chi^2/n}}$ 服从自由度为 n 的 t **分布**(t distribution)，记为 $T \sim t(n)$.

密度函数

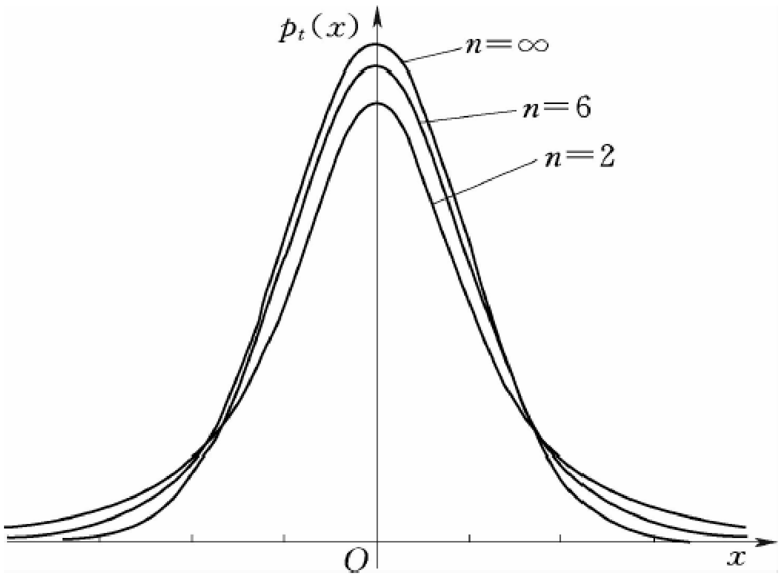
$$f_n(x) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$$

t分布是由统计学家哥威廉·戈塞在都柏林的A.吉尼斯父子酿酒厂对小样本中平均数比例对其标准误差的分布所做的研究，由于吉尼斯酿酒厂的规定禁止戈塞发表关于酿酒过程变化性的研究成果，因此戈塞不得不于1908年，首次以“学生”(Student)为笔名，发表自己的研究成果。

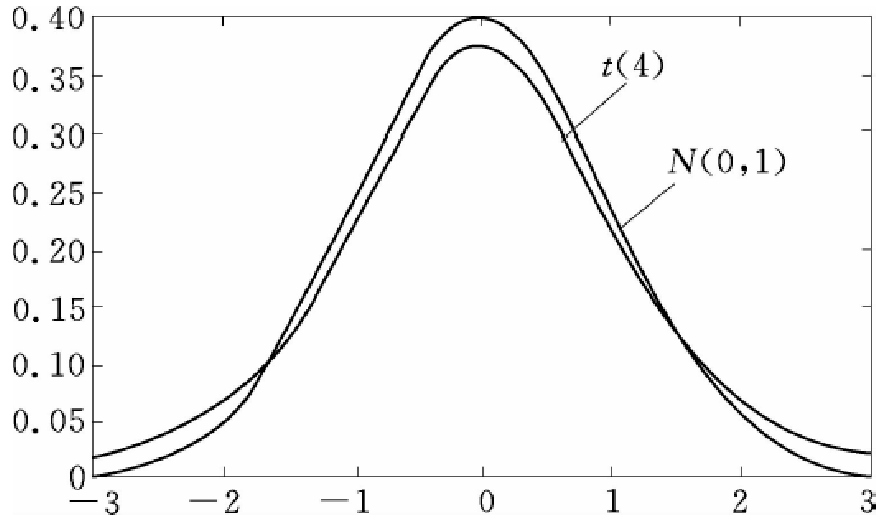
William Sealy Gosset

William Sealy Gosset (aka *Student*) in 1908 (age 32).

Born	13 June 1876 Canterbury, Kent, England
Died	16 October 1937 (aged 61) Beaconsfield, Buckinghamshire, England
Other names	Student
Alma mater	New College, Oxford
Known for	Student's t-distribution



t 分布的密度曲线



t 分布与 $N(0,1)$ 密度曲线比较

t分布具有以下性质：

性质 1 若 $T \sim t(n)$，则有相互独立的 $Z \sim N(0,1)$，

$\chi^2 \sim \chi^2(n)$使$T = \underline{\dfrac{Z}{\sqrt{\chi^2/n}}}$.

性质 2 $\lim\limits_{n\to\infty} p(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(x)$，即$t$分布的极限分布是

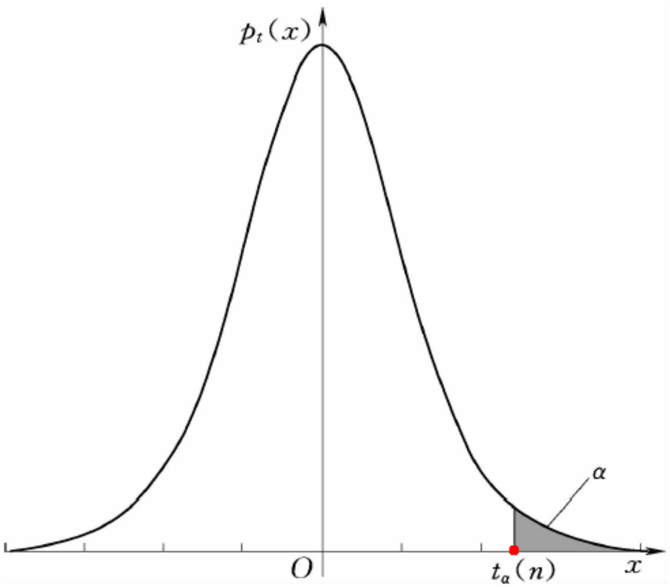
标准正态分布.

性质 3 若$T \sim t(n)$，则$n>1$时， $E(T) = \underline{\ 0\ }$ （因为$p(x)$

关于y轴对称）； $n>2$时， $D(T) > \underline{\ 1\ }$.

定义 设 $T \sim t(n)$，对给定的 α（$0<\alpha<1$），满足

$P\{T > t_\alpha(n)\} = \alpha$ 的点 $t_\alpha(n)$ 称为 t 分布的上侧分位点.



t 分布的上侧分位点

$$t_{0.05}(10) = 1.8125$$

$$z_\alpha = t_\alpha(\infty)$$

$$n > 45$$

$$t_\alpha(n) \approx z_\alpha$$

设 X_1, X_2, X_3, X_4 为来自总体 $N(1,\sigma^2)$ 的简单随机样本，则统计量 $\dfrac{X_1 - X_2}{|X_3 + X_4 - 2|}$ 的分布为（　　）.

（A）$N(0,1)$；（B）$\chi^2(1)$；（C）$t(1)$；（D）$F(1,1)$.

设 X_1, X_2, X_3, X_4 为来自总体 $N(1,\sigma^2)$ 的简单随机样本，则统计量 $\dfrac{X_1 - X_2}{|X_3 + X_4 - 2|}$ 的分布为（ C ）.

（A） $N(0,1)$；（B） $\chi^2(1)$；（C） $t(1)$；（D） $F(1,1)$.

$$X_1 - X_2 \sim N(0, 2\sigma^2) \Rightarrow \frac{X_1 - X_2}{\sqrt{2}\sigma} \sim N(0,1)$$

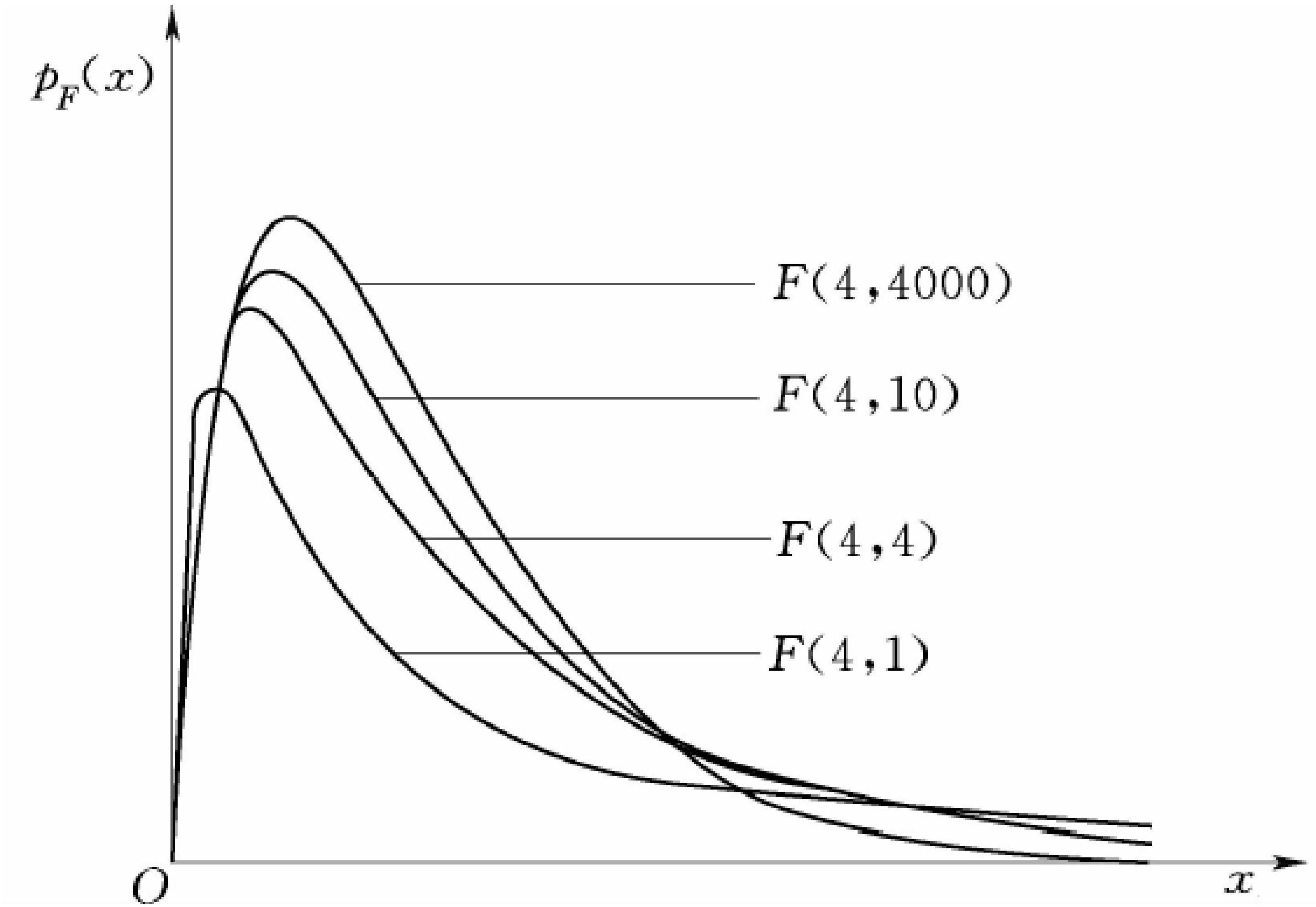
$$X_3 + X_4 - 2 \sim N(0, 2\sigma^2) \Rightarrow \left(\frac{X_3 + X_4 - 2}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)$$

F 分布

定义　若① $X \sim \chi^2(n_1)$；② $Y \sim \chi^2(n_2)$；③ X 与 Y 相互独立，则称 $F = \dfrac{X/n_1}{Y/n_2}$ 服从第一自由度为 n_1，第二自由度为 n_2 的 F 分布(F　distribution)，记为 $F \sim F(n_1, n_2)$.

$F(m,n)$的密度函数(x>0)

$$f_{m,n}(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} x^{m/2-1} \left(1 + \frac{m}{n}x\right)^{-(m+n)/2}$$



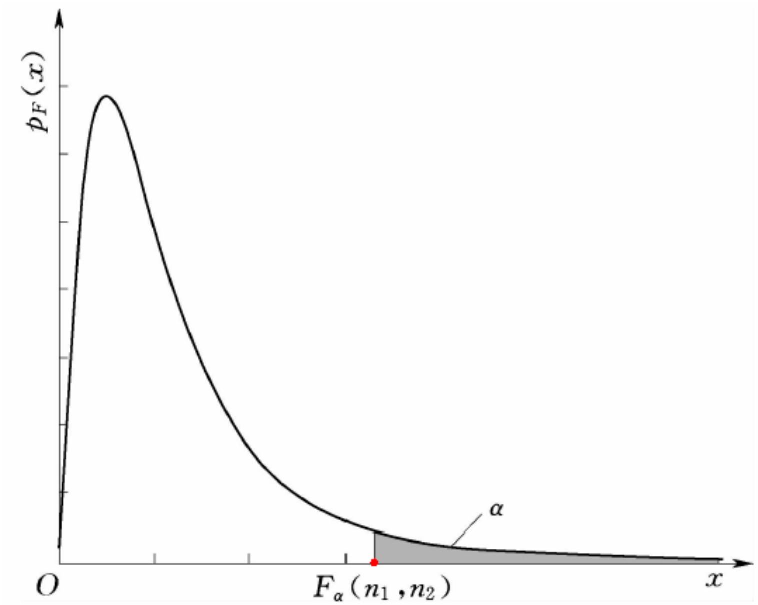
F 分布的密度曲线

定 义 设 $F \sim F(n_1, n_2)$，对给定的 α，满足

$P\{F > F_\alpha(n_1, n_2)\} = \alpha$ 的点 $F_\alpha(n_1, n_2)$ 称为 F 分布的上侧分位点.

$$F_{0.05}(3,4) = 6.59$$

$$F_{0.05}(4,3) = 9.12$$



F 分布的上侧分位点

F 分布具有以下性质：

性质 1 若 $F \sim F(n_1, n_2)$，则有相互独立的 $X \sim \chi^2(n_1)$，

$Y \sim \chi^2(n_2)$，使 $F =$ $\underline{\dfrac{X/n_1}{Y/n_2}}$.

性质 2 若 $F \sim F(n_1, n_2)$，则 $\dfrac{1}{F} \sim$ $\underline{F(n_2, n_1)}$.

性质 3 $F_{1-\alpha}(n_1, n_2) =$ $\underline{\dfrac{1}{F_\alpha(n_2, n_1)}}$.

$$F_{0.95}(3,4) = \frac{1}{F_{0.05}(4,3)} = \frac{1}{9.12} = 0.1097$$

提问 设 $X_1,X_2,\cdots,X_m$ 与 $Y_1,Y_2,\cdots,Y_n$ 是来自 $X\sim N(\mu,\sigma^2)$ 的两个独立样本，则

（1）$X_1+X_2+\cdots+X_m\sim$ ________（分布）.

（2）$Z=\dfrac{X_1+X_2+\cdots+X_m-m\mu}{\sqrt{m}\sigma}\sim$ ________（分布）.

（3）$T=\sum_{i=1}^{n}\left(\dfrac{Y_i-\mu}{\sigma}\right)^2\sim$ ________（分布）.

（4）Z 与 T ________（是否独立）.

（5）$\dfrac{Z}{\sqrt{T/n}}\sim$ ________（分布）.

提问 设 $X_1, X_2, \cdots, X_m$ 与 $Y_1, Y_2, \cdots, Y_n$ 是来自 $X \sim N(\mu, \sigma^2)$ 的两个独立样本，则

（1）$X_1 + X_2 + \cdots + X_m \sim$ $\underline{N(m\mu, m\sigma^2)}$（分布）.

（2）$Z = \dfrac{X_1 + X_2 + \cdots + X_m - m\mu}{\sqrt{m}\sigma} \sim$ $\underline{N(0,1)}$（分布）.

（3）$T = \sum_{i=1}^{n}\left(\dfrac{Y_i - \mu}{\sigma}\right)^2 \sim$ $\underline{\chi^2(n)}$（分布）.

（4）Z 与 T <u>独立</u>（是否独立）.

（5）$\dfrac{Z}{\sqrt{T/n}} \sim$ $\underline{t(n)}$（分布）.

设 $X_1, X_2, \cdots, X_9$ 和 $Y_1, Y_2, \cdots, Y_9$ 是来自同一总体 $N(0,9)$ 的两个独立的样本，统计量

$$Z = \left(\sum_{i=1}^{9} X_i\right) \Big/ \sqrt{\sum_{i=1}^{9} Y_i^2}$$，试确定 Z 的分布.

$$X_i \sim N(0,9) \qquad \frac{1}{9}\sum_{i=1}^{9} X_i \sim N(0,1)$$

$$Y_i \sim N(0,9)$$

$$\sum_{i=1}^{9}\left(\frac{Y_i}{3}\right)^2 = \sum_{i=1}^{9}\frac{Y_i^2}{9} \sim \chi^2(9)$$

$$\frac{\frac{1}{9}\sum_{i=1}^{9} X_i}{\sqrt{\sum_{i=1}^{9}\frac{Y_i^2}{9} \Big/ 9}} = \frac{\sum_{i=1}^{9} X_i}{\sqrt{\sum_{i=1}^{9} Y_i^2}} \sim t(9) \qquad Z \sim t(9)$$

正态总体下的抽样分布

定理 设$(X_1, X_2, \cdots, X_n)$是来自总体$X \sim N(\mu, \sigma^2)$的

样本， $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$， $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$， 则

（1） $\bar{X}$ 与 S^2 相互独立；

（2） $\bar{X} \sim \underline{N\left(\mu, \dfrac{\sigma^2}{n}\right)}$ ； $\Longleftrightarrow$ $\dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \underline{N(0,1)}$

（3） $\dfrac{(n-1)S^2}{\sigma^2} \sim \underline{\chi^2(n-1)}$.

定理 $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

定理 $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim$ $t(n-1)$

证明 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

$\bar{X}$ 与 S^2 相互独立，显然 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ 与 $\dfrac{(n-1)S^2}{\sigma^2}$ 也相互独立.

$$\frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} \sim t(n-1) \Rightarrow \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

两个正态总体下的抽样分布

设样本 $X_1, X_2, \ldots, X_{n_1}$ 来自总体 $X \sim N(\mu_1, \sigma_1^2)$

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \qquad S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(X_i - \bar{X})^2$$

设样本 $Y_1, Y_2, \cdots, Y_{n_2}$ 来自总体 $Y \sim N(\mu_2, \sigma_2^2)$

$$\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \qquad S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$$

两个样本相互独立.

定理

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1)$$

____

定理 $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim \underline{N(0,1)}$$

证明 $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$ $\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$ 相互独立

$$E(\bar{X}-\bar{Y}) = E(\bar{X}) - E(\bar{Y}) = \mu_1 - \mu_2$$

$$D(\bar{X}-\bar{Y}) = D(\bar{X}) + D(\bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

定理 当σ_1^2,σ_2^2未知，但两者相等时，

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim \underline{t(n_1+n_2-2)}$$

其中 $S_W^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

证明 设σ_1^2,σ_2^2都等于σ^2 $\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2/n_1+\sigma^2/n_2}} \sim N(0,1)$

$\dfrac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2(n_1-1)$ $\dfrac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_2-1)$ 相互独立

$$\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n^2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$$

$$\frac{\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2/n_1+\sigma^2/n_2}}}{\sqrt{\left[\dfrac{(n_1-1)S_1^2}{\sigma^2}+\dfrac{(n_2-1)S_2^2}{\sigma^2}\right]/(n_1+n_2-2)}} \sim t(n_1+n_2-2)$$

化简整理即得

定理 $\dfrac{S_1^2}{S_2^2} \cdot \dfrac{\sigma_2^2}{\sigma_1^2} \sim \underline{\; F(n_1 - 1, n_2 - 1) \;}$

定理 $\dfrac{S_1^2}{S_2^2}\cdot\dfrac{\sigma_2^2}{\sigma_1^2} \sim F(n_1-1, n_2-1)$

证明 $\dfrac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2(n_1-1)$ $\dfrac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_2-1)$ 相互独立

$$\frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)} \sim F(n_1-1, n_2-1)$$

$$\frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} \sim F(n_1-1, n_2-1)$$

例　设(X_1, X_2, X_3, X_4)是从总体$X \sim N(0,1)$中抽取的一个样本，令

$$Y_1 = X_1 + X_2 + X_3 + X_4，\quad Y_2 = X_1 - X_2 + X_3 - X_4$$

证明Y_1与Y_2相互独立.

$$
\begin{aligned}
Cov(Y_1,Y_2) &= Cov(X_1,X_1)-Cov(X_1,X_2) \\
&+Cov(X_1,X_3)-Cov(X_1,X_4) \\
&+Cov(X_2,X_1)-Cov(X_2,X_2) \\
&+Cov(X_2,X_3)-Cov(X_2,X_4) \\
&+Cov(X_3,X_1)-Cov(X_3,X_2) \\
&+Cov(X_3,X_3)-Cov(X_3,X_4) \\
&+Cov(X_4,X_1)-Cov(X_4,X_2) \\
&+Cov(X_4,X_3)-Cov(X_4,X_4) \\
&=Cov(X_1,X_1)-Cov(X_2,X_2) \\
&+Cov(X_3,X_3)-Cov(X_4,X_4) \quad = 0
\end{aligned}
$$

$\Rightarrow$ Y_1与Y_2不相关 $\Leftrightarrow$ Y_1与Y_2独立

例　设 $(X_1,X_2,\cdots,X_m,\cdots,X_n)$（$m<n$）是从总体 $X\sim N(0,1)$ 中抽取的一个样本，令

$$Y=a(X_1+X_2+\cdots+X_m)^2+b(X_{m+1}+\cdots+X_n)^2$$

试求 a,b 的值，使 Y 服从 χ^2 分布.

解 $X_1+X_2+\cdots+X_m$ 与 $X_{m+1}+\cdots+X_n$ 相互独立

$$X_1+X_2+\cdots+X_m \sim N(0,m) \qquad \frac{X_1+X_2+\cdots+X_m}{\sqrt{m}} \sim N(0,1)$$

$$X_{m+1}+\cdots+X_n \sim N(0,n-m) \qquad \frac{X_{m+1}+\cdots+X_n}{\sqrt{n-m}} \sim N(0,1)$$

$$\frac{(X_1+X_2+\cdots+X_m)^2}{m} + \frac{(X_{m+1}+\cdots+X_n)^2}{n-m} \sim \chi^2(2)$$

$\Rightarrow$ $a=\dfrac{1}{m}$ $\qquad b=\dfrac{1}{n-m}$

与样本函数相关的概率问题

例　设$(X_1, X_2, \cdots, X_n)$是从总体 $X \sim N(\mu, \sigma^2)$ 中抽取的一个样本，$\bar{X}$ 为样本均值，S^2 为样本方差.如果n很大，试求$P\left\{(\bar{X}-\mu)^2 \le \dfrac{2S^2}{n}\right\}$.

解

$$P\left\{(\bar{X}-\mu)^2 \le \frac{2S^2}{n}\right\} = P\left\{\frac{(\bar{X}-\mu)^2}{S^2/n} \le 2\right\}$$

$$= P\left\{\frac{|\bar{X}-\mu|}{S/\sqrt{n}} \le \sqrt{2}\right\} \approx \Phi(\sqrt{2}) - \Phi(-\sqrt{2})$$

$$\approx 2\Phi(1.414\sqrt{2}) - 1 \approx 2 \times 0.9213 - 1$$

$$= 0.8426$$