1.3　抽样分布

一、χ^2分布

二、t 分布

三、F 分布

四、概率分布的分位数

五、正态总体样本均值和方差的分布

六、一些非正态总体样本均值的分布

抽样分布的定义

统计量既然是依赖于样本的, 而后者又是随机变量, 故统计量也是随机变量, 因而就有一定的分布. 称这个分布为“抽样分布”. 也即抽样分布就是统计量的分布.

抽样分布
- 精确抽样分布 (小样本问题中使用)
- 渐近分布 (大样本问题中使用)

一、χ^2 分布

概率中学过的5类分布族：

- 二项分布族 $\{B(n,p): 0<p<1\}$,
- 泊松分布族 $\{P(\lambda): \lambda>0\}$,
- 均匀分布族 $\{U(a,b): -\infty<a<b<\infty\}$,
- 正态分布族 $\{N(\mu,\sigma^2): -\infty<\mu<\infty, \sigma^2>0\}$,
- 指数分布族 $\{Exp(\lambda): \lambda>0\}$,

本节将再介绍三类分布族，它们将在数理统计中起着重要的作用.

1. χ^2 分布

定义1.8 设随机变量$X_1, X_2, \cdots, X_n$ *i.i.d*$\sim N(0,1)$,
则称随机变量

$$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

服从自由度为n的χ^2分布，记作$\chi_n^2 \sim \chi^2(n)$.

这里的自由度是指和式中独立变量的个数.

卡尔·皮尔逊（**Karl Pearson**，**1857**年生，英国数学家，生物统计学家，数理统计学的创立者，自由思想者，对生物统计学、气象学、社会达尔文主义理论和优生学做出了重大贡献。被公认是旧派理学派和描述统计学派的代表人物，现代统计科学的创立者。

卡方分布及卡方检验，它是由被称为数理统计学之父的卡尔·皮尔逊(**Karl Pearson**)于**1900**年提出的。

Γ函数

以下积分称为Γ函数$(\alpha>0)$

$$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx\overset{令x=y^2}{=}2\int_0^{+\infty}y^{2\alpha-1}e^{-y^2}dy$$

Γ函数有下列公式

$$\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1),$$

$$\Gamma(n+1)=n!$$

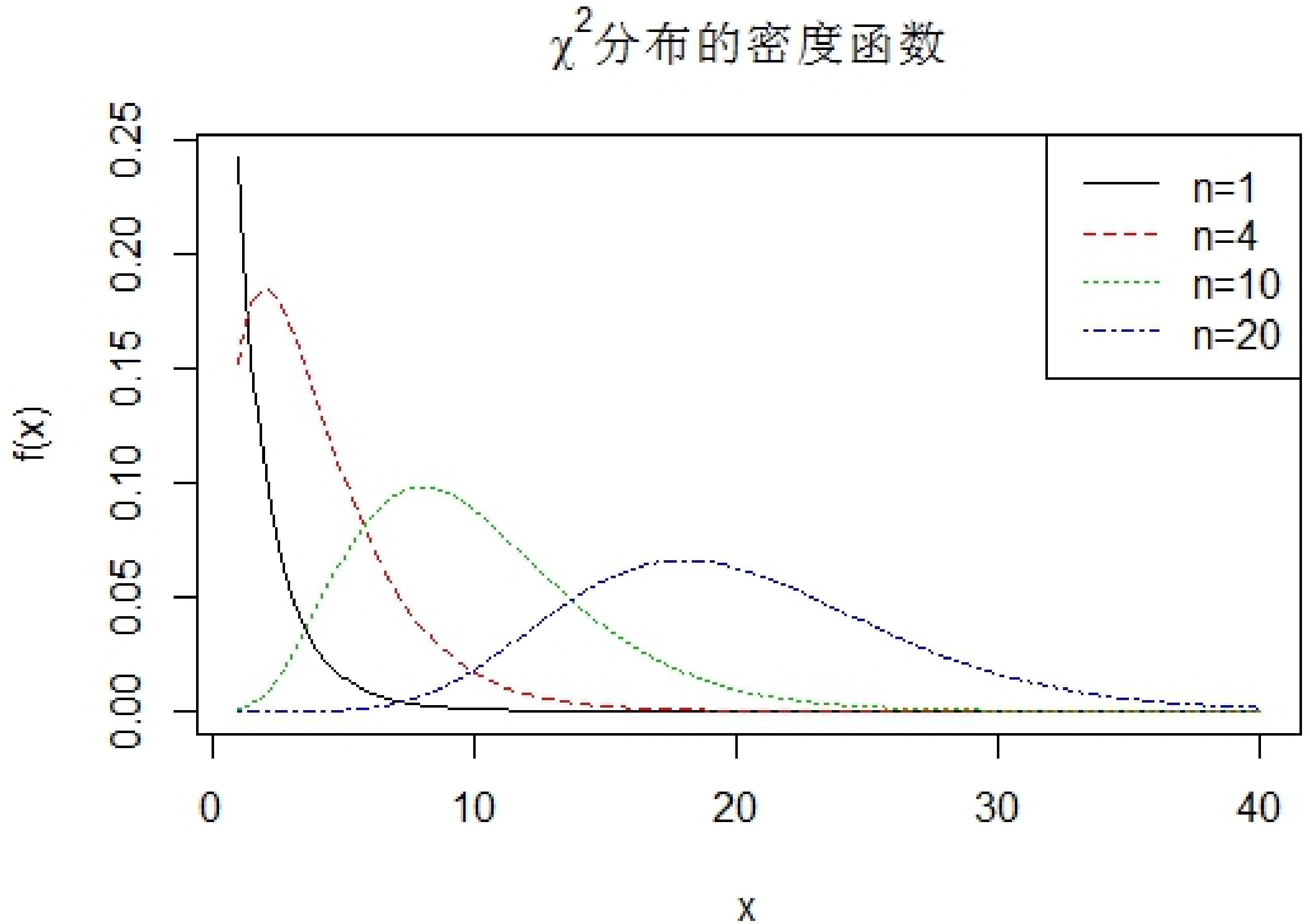
$$\Gamma(1)=\Gamma(2)=1,\ \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}.$$

$$\int_0^{+\infty}xe^{-x^2}dx$$

$=?$

定理1.6 设随机变量$X_1, X_2, \cdots, X_n$独立，同服从$N(0,1)$分布，则随机变量$Y_n = \sum_{i=1}^{n} X_i^2$的概率分布密度为

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}}, & x > 0, \\ 0, & x \le 0. \end{cases}$$



χ^2分布的密度函数
0.00
0.05
0.10
0.15
0.20
0.25
f(x)
0
10
20
30
40
x
n=1
n=4
n=10
n=20

2. χ^2 分布的性质

性质1 (χ^2分布的数学期望和方差)

若 $\chi_n^2 \sim \chi^2(n)$, 则 $E(\chi_n^2) = n$, $D(\chi_n^2) = 2n$.

证明 因为 $X_i \sim N(0,1)$, 所以

$$E(X_i^2) = D(X_i) + [E(X_i)]^2 = 1,$$

故 $E(\chi_n^2) = E\left(\sum_{i=1}^{n} X_i^2\right) = \sum_{i=1}^{n} E(X_i^2) = n,$

$$E(X_i{}^4)=\int_{-\infty}^{+\infty}x^4\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}\,x=\frac{-1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^3\cdot\mathrm{d}e^{-\frac{x^2}{2}}$$

$$=\frac{-1}{\sqrt{2\pi}}[x^3e^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty}-\int_{-\infty}^{+\infty}3x^2\cdot e^{-\frac{x^2}{2}}\,\mathrm{d}\,x]\ =3$$

可利用Γ函数积分!

$$\therefore\quad D(X_i{}^2)=E(X_i{}^4)-[E(X_i{}^2)]^2$$

$$=3-1=2,\quad (i=1,2,\cdots,n)$$

$$D(\chi_n^2)=D\left(\sum_{i=1}^{n}X_i{}^2\right)=\sum_{i=1}^{n}D(X_i{}^2)=2n.$$

性质2 (χ^2 分布的可加性)

若$X\sim\chi^2(m)$,$Y\sim\chi^2(n)$,并且X与Y相互独立，则

$X+Y\sim\chi^2(m+n)$.

(此性质可以推广到多个随机变量的情形)

性质3 设$X\sim\chi^2(n)$,则对任意实数x,有

$$\lim_{n\to\infty}P\{\frac{X-n}{\sqrt{2n}}\le x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$$

即当n充分大时，$X\sim AN(n,2n)$.

定理1.7 (柯赫伦定理)设$X_1, X_2, \cdots, X_n$是n个独立、同服从标准正态分布$N(0, 1)$的随机变量，记$Q = \sum_{i=1}^{n} X_i^2$.

若Q分解为$Q = \sum_{i=1}^{k} Q_i$,其中$Q_i (i = 1, 2, \cdots, k)$是秩为$n_i$的关于$(X_1, X_2, \cdots, X_n)^T$的非负二次型，则$Q_i$相互独立，且$Q_i \sim \chi^2(n_i)(i = 1, 2, \cdots, k)$的充要条件为$\sum_{i=1}^{k} n_i = n$.

注：和性质2作比较。

二、t分布族

1. t分布

定义1.9 设$X \sim N(0,1), Y \sim \chi^2(n)$,且$X$与$Y$相互独立,

则称随机变量

$$T = \frac{X}{\sqrt{Y/n}}$$

服从自由度为n的t分布，记为$T \sim t(n)$，随机变量T亦称为T变量

t 分布又称**学生氏(Student)**分布.

2. t分布的密度函数

定理1.8 T变量的分布密度函数为

$$f_T(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\quad -\infty<x<+\infty$$

证 此密度函数可分两步来推导出来。

首先利用一维随机变量的函数的密度函数求解公式计算Z=$\sqrt{Y/n}$的密度函数；

再利用商的概率密度计算公式计算

$$T=\frac{X}{Z}=\frac{X}{\sqrt{Y/n}}$$的密度函数。

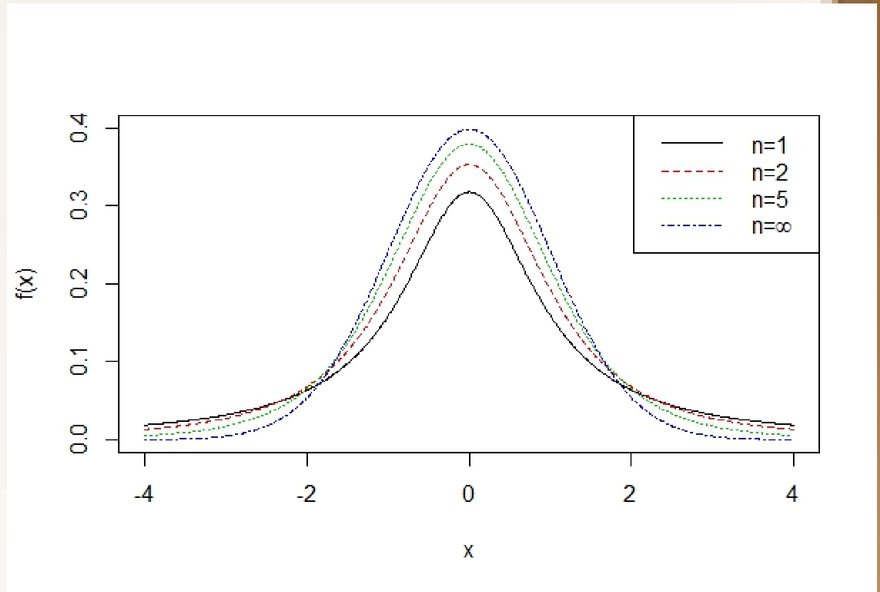
t 分布是由英国统计学家威廉·希利·戈塞特发现的。戈塞特获化学、数学双学位，依靠自己的化学知识进酿酒厂，工作期间考虑酿酒配方实验中的统计学问题，追随卡尔皮尔逊学习了一年的统计学，后来设计了一种方法来评价酒的质量。
因为行业机密，酒厂不允许他的工作内容外泄，**1908** 年哥塞特在 **Biometrics** 杂志以笔名“学生”发表了使他名垂统计史册的论文，提出**t** 分布，打破了正态分布一统天下的局面，开创了小样本统计推断的新纪元。

“t”，是伟大的Fisher为之取的名字。Fisher最早将这一分布命名为“Student's distribution”，并以“t”为之标记。

W. S. Gosset

3. t分布的图象特征

t分布的概率密度曲线如图



当n充分大时, 其图形类似于标准正态变量概率密度的图形.

因为$\lim\limits_{n\to\infty} f(x,n)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$,

所以当n足够大时t分布近似于$N(0,1)$分布,

但对于较小的n, t分布与$N(0,1)$分布相差很大.

证明 $\lim\limits_{n\to\infty} f(x,n)=\lim\limits_{n\to\infty}\dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

（1）利用Stirling公式可以证明 $\lim\limits_{n\to\infty}\dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n}\,\Gamma\left(\dfrac{n}{2}\right)}=\dfrac{1}{\sqrt{2}}$

（2）利用重要极限可以证明

$$\lim_{n\to\infty}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}=\mathrm{e}^{-\frac{x^2}{2}}$$

4. t分布的性质

性质1 $EX=0,\qquad DX=\dfrac{n}{n-2},n>2.$

性质2 自由度为1的t分布称为柯西分布，其密度函数为

$$f(x)=\frac{1}{\pi(1+x^2)},\qquad -\infty<x<\infty$$

此分布的数学期望不存在.

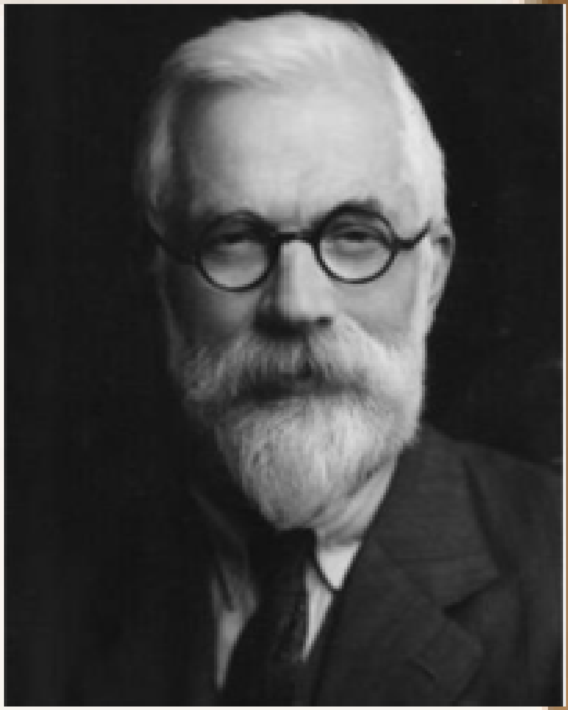
三、F 分布

1. F分布

定义1.10 设 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 且相互独立, 则随机变量 $F = \dfrac{X/n_1}{Y/n_2}$ 服从自由度为 (n_1, n_2) 的 F 分布, 记为 $F \sim F(n_1, n_2)$.

其中n_1为第一自由度，n_2为第二自由度

Fisher费歇尔(1890～1962)，英国统计与遗传学家，现代统计科学的奠基人之一，“一位几乎独自建立现代统计科学的天才”，对达尔文进化论做出了基础澄清的工作。



1922年发布的《回归公式的拟合优度即回归系数的分布》和1924年发布的《关于一个引出若干周知统计量的误差函数的分布》,正式提出方差分析方法和F-分布。

F分布是以Fisher的首字母命名。

2. F分布的密度函数

定理1.9 随机变量F的分布密度函数为

$$f_F(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)\left(\dfrac{n_1}{n_2}\right)}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}\left(\dfrac{n_1}{n_2}x\right)^{\frac{n_1}{2}-1}\left[1+\left(\dfrac{n_1x}{n_2}\right)\right]^{-\frac{n_1+n_2}{2}}, & x>0,\\ 0, & x\le 0.\end{cases}$$

证明 令$U=\dfrac{X}{n_1},V=\dfrac{Y}{n_2}$，且$U$与$V$相互独立。

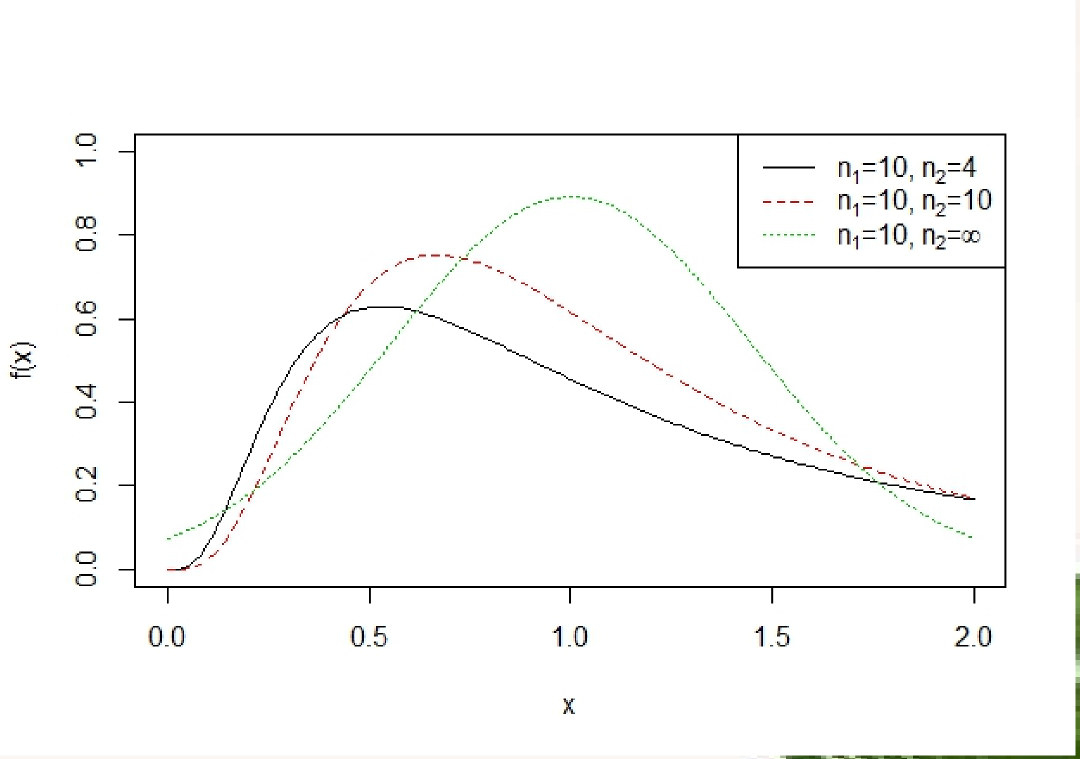
利用两个独立随机变量商的概率密度函数计算公式可得。

3. F分布的几何特征

根据定义可知,

若$F \sim F(n_1, n_2)$,

则$\dfrac{1}{F} \sim F(n_2, n_1)$.



4. F分布的性质

性质1 $E(F)=\dfrac{n_2}{n_2-2}$, $(n_2>2)$,

$$D(F)=\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)},n_2>4.$$

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

性质3 若$T\sim t(n)$,则$T^2\sim F(1,n)$.

定理1.10 设$X_1, X_2, \cdots, X_n$相互独立，且同服从$N(0,\sigma^2)$分布，$Q_i(i=1,2,\cdots,k)$是关于$(X_1, X_2, \cdots, X_n)^T$的秩(即自由度）为$n_i$的非负二次型，且

$$Q_1 + Q_2 + \cdots + Q_k = \sum_{i=1}^{n} X_i^2$$

$$n_1 + n_2 + \cdots + n_k = n$$

则

$$F_{ij} = \frac{Q_i / n_i}{Q_j / n_j} \sim F(n_i, n_j)$$

意义：在方差分析中有重要作用

例1 设X_1, X_2, X_3, X_4来自总体$N(0,\sigma^2)$,则统计量

$$T = \frac{X_1 + X_2}{\sqrt{X_3^2 + X_4^2}}$$的分布为?

解 $X_1 + X_2 \sim N(0, 2\sigma^2)$, 于是 $\dfrac{X_1 + X_2}{\sqrt{2\sigma^2}} \sim N(0,1)$

$\dfrac{X_3}{\sqrt{\sigma^2}}$与$\dfrac{X_4}{\sqrt{\sigma^2}}$独立同分布于$N(0,1)$,于是

$$\frac{X_3^2}{\sigma^2} + \frac{X_4^2}{\sigma^2} \sim \chi^2(2)$$

由定义1.9可知

$$\frac{\dfrac{X_1+X_2}{\sqrt{2\sigma^2}}}{\sqrt{\dfrac{X_3^2+X_4^2}{\sigma^2}\Big/2}} \sim t(2)$$

即 $$\frac{X_1+X_2}{\sqrt{X_3^2+X_4^2}} \sim t(2)$$

例2 设$X_1,\cdots,X_{15}$服从$N(0,2^2)$且相互独立，试求随机变量

$$Y=\frac{X_1^2+\cdots+X_{10}^2}{2(X_{11}^2+\cdots+X_{15}^2)}$$的分布。

解 $\because \dfrac{X_1^2+\cdots+X_{10}^2}{2^2}\sim\chi^2(10),\dfrac{X_{11}^2+\cdots+X_{15}^2}{2^2}\sim\chi^2(5)$

且相互独立，所以

$$Y=\frac{(X_1^2+\cdots+X_{10}^2)/10}{(X_{11}^2+\cdots+X_{15}^2)/5}=\frac{X_1^2+\cdots+X_{10}^2}{2(X_{11}^2+\cdots+X_{15}^2)}\sim F(10,5).$$

单选题 1分

设置

设X服从$N(0,1)$，则X^2服从

A $\chi^2(1)$

B $t(1)$

C $N(0,2)$

D $N(0,1)$

提交

Thank You!