第1.3节　　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 $\begin{cases} \text{精确抽样分布} & \text{(小样本问题中使用)} \\ \text{渐近分布} & \text{(大样本问题中使用)} \end{cases}$

一、χ^2 分布

首先回顾以前学过的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 $\{E(\lambda):\ \lambda>0\}$,

本节将介绍其他几类分布族，它们将在数理统计中起着重要的作用.

1. Γ函数

$$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x$$

Γ函数的性质:

$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$, （利用分部积分可以证明）

$\Gamma(1)=\Gamma(0)=1$,

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

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

2. Γ分布（补充内容）

定义设随机变量X的分布密度函数为

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\beta x}, & x>0,\\ 0, & x\le 0,\end{cases}$$

则称X服从Γ分布，记为$X \sim \Gamma(\alpha,\beta)$,其中$\alpha>0,\beta>0$为参数，Γ分布族常记为$\{\Gamma(\alpha,\beta):\alpha>0,\beta>0\}$.

注：指数分布为特殊的Γ分布，即$E(\lambda)=\Gamma(1,\lambda)$

3. Γ分布的性质

性质1 $E(X^{k})=\dfrac{\Gamma(\alpha+k)}{\Gamma(\alpha)\beta^{k}}=\dfrac{(\alpha+k-1)(\alpha+k-2)\cdots\alpha}{\beta^{k}}$

其中 $E(X)=\dfrac{\alpha}{\beta},\quad D(X)=E(X^2)-E^2(X)=\dfrac{\alpha}{\beta^2}$

证 $E(X^k)=\int_0^{\infty}x^k\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\beta x}\mathrm{d}x=\int_0^{\infty}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{k+\alpha-1}\mathrm{e}^{-\beta x}\mathrm{d}x$

$$\xlongequal{\text{换元}\beta x=t}\frac{1}{\beta^k\Gamma(\alpha)}\int_0^{\infty}t^{k+\alpha-1}\mathrm{e}^{-t}\mathrm{d}t=\frac{\Gamma(\alpha+k)}{\beta^k\Gamma(\alpha)}$$

性质2（可加性） 若$X_j\sim\Gamma(\alpha_j,\beta),\ j=1,2,\cdots,n,$

而且X_j间相互独立，则

$$\sum_{j=1}^{n}X_j\sim\Gamma\Big(\sum_{j=1}^{n}\alpha_j,\beta\Big),$$

4. χ^2分布

定义1.8 设随机变量$X_1,X_2,\cdots,X_n$相互独立且同服从标准正态分布$N(0,1)$,则称随机变量

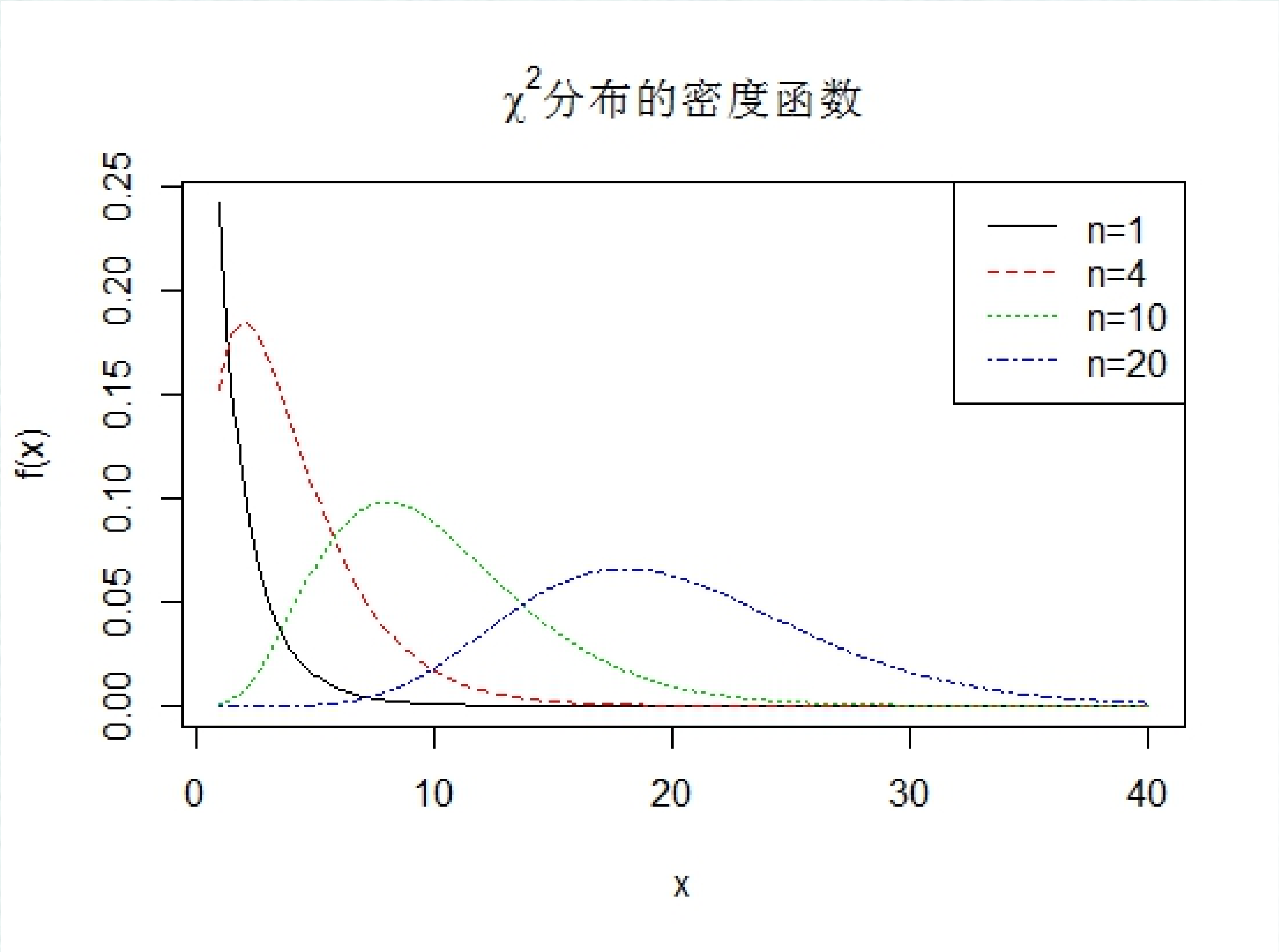
$$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

自由度为n 的χ^2分布，记为$\chi_n^2 \sim \chi^2(n)$.这里的自由度是指和式中独立变量的个数.

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

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

注 当$\alpha=\dfrac{n}{2},\beta=\dfrac{1}{2}$,则$\Gamma(\dfrac{n}{2},\dfrac{1}{2})=\chi^2(n)$


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

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

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

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

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

$$= 3\int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = 3$$

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

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

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

$$= \sum_{i=1}^{n} E(X_i^{\ 2}) = n, \qquad (\ E(X_i^{\ 2}) = 1\ )$$

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

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

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

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$\sum_{i=1}^{k} n_i = n$.

二、t分布族

学生氏

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

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

证 此问题可以利用商的概率密度计算公式计算.

首先计算$Z=\sqrt{\frac{Y}{n}}$的分布函数

$$F_Z(z)=P\{\sqrt{\frac{Y}{n}}\le z\}=P\{Y\le nz^2\}=F_Y(nz^2)$$

因此

$$\varphi_Z(z) = (F_Y(nz^2))' = F_Y'(nz^2)\times 2nz = \varphi_Y(nz^2)\times 2nz$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}(nz^2)^{\frac{n}{2}-1}\mathrm{e}^{-\frac{nz^2}{2}}\times 2nz \quad = \frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})}n^{\frac{n}{2}}z^{n-1}\mathrm{e}^{-\frac{nz^2}{2}} \qquad z>0.$$

易知$z\le 0$时，$\varphi_Z(z)=0.$

再由商的概率密度计算公式可得

$$\varphi_T(x) = \int_{-\infty}^{\infty}|z|\varphi_Z(z)\varphi_X(zx)\mathrm{d}z$$

$$= \int_0^{\infty} z\frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})}n^{\frac{n}{2}}z^{n-1}\mathrm{e}^{-\frac{nz^2}{2}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2x^2}{2}}\mathrm{d}z$$

$$=\frac{n^{\frac{n}{2}}}{\sqrt{\pi}\,2^{\frac{n-1}{2}}\Gamma(\frac{n}{2})}\int_0^{\infty} z^n e^{-\frac{z^2}{2}(n+x^2)}\mathrm{d}z\text{（令}u=\frac{z^2}{2}(n+x^2)\text{）}$$

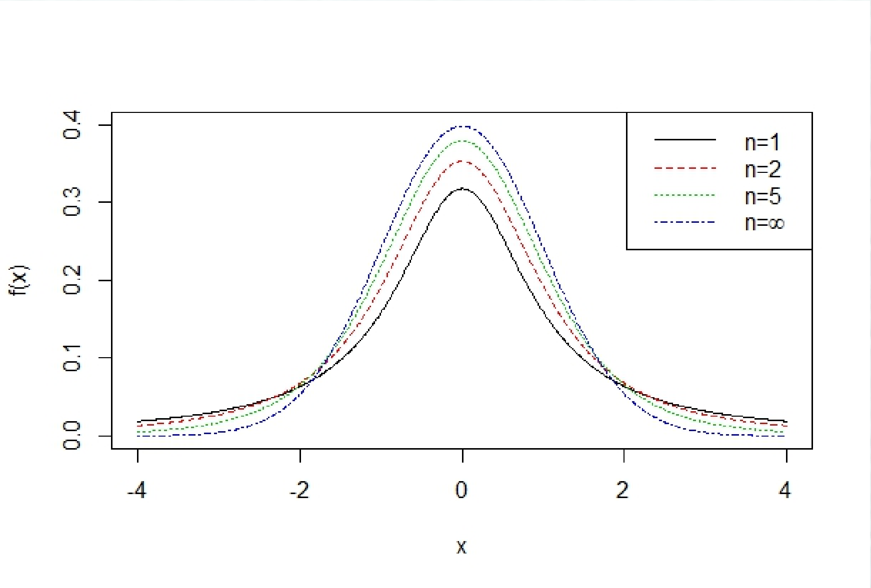
$$=\frac{1}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})(1+\frac{x^2}{n})^{\frac{n+1}{2}}}\int_0^{\infty} u^{\frac{n+1}{2}-1}e^{-u}\mathrm{d}u$$

$$=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$$

因而定理1.8成立。

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$$\lim_{n\to\infty}\varphi_T(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$$

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.

利用Stirling公式可以证明

$$\lim_{n\to\infty}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\,\Gamma\left(\frac{n}{2}\right)}=\frac{1}{\sqrt{2}}$$

利用重要极限可以证明

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

因而

$$\lim_{n\to\infty}\varphi_T(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},$$

4. t分布的性质

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

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

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

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

证明

令$U=\dfrac{X}{n_1}, V=\dfrac{Y}{n_2}$,且$U$与$V$相互独立，其分布密度函数分别为

$$f_U(u)=\begin{cases}\dfrac{n_1^{\frac{n_1}{2}}}{2^{\frac{n_1}{2}}\Gamma(\frac{n_1}{2})}u^{\frac{n_1}{2}-1}\mathrm{e}^{-\frac{n_1}{2}u} & u>0\\ 0 & u\le 0\end{cases}$$

$$f_V(v)=\begin{cases}\dfrac{n_2^{\frac{n_2}{2}}}{2^{\frac{n_2}{2}}\Gamma(\frac{n_2}{2})}v^{\frac{n_2}{2}-1}\mathrm{e}^{-\frac{n_2}{2}v} & v>0\\ 0 & v\le 0\end{cases}$$

利用两个独立随机变量商的概率密度函数计算公式可得

$$f_F(x)=\int_{-\infty}^{+\infty}|v|\,f_U(xv)f_V(v)\mathrm{d}v$$

$$=\frac{n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\int_0^{+\infty}v(xv)^{\frac{n_1}{2}-1}\mathrm{e}^{-\frac{n_1}{2}(xv)}v^{\frac{n_2}{2}-1}\mathrm{e}^{-\frac{n_2}{2}v}\mathrm{d}v$$

$$=\frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} x^{\frac{n_1}{2}-1}\int_0^{+\infty} v^{\frac{n_1+n_2}{2}-1} \mathrm{e}^{-\frac{n_1 x+n_2}{2}v}\,\mathrm{d}v$$

（换元 $w=\frac{n_1 x+n_2}{2}v,\quad \mathrm{d}v=\frac{2}{n_1 x+n_2}\mathrm{d}w$）

$$=\frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} x^{\frac{n_1}{2}-1}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\left(\frac{2}{n_1 x+n_2}\right)^{\frac{n_1+n_2}{2}}\int_0^{+\infty} w^{\frac{n_1+n_2}{2}-1}\mathrm{e}^{-w}\,\mathrm{d}w$$

$$=\frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} x^{\frac{n_1}{2}-1}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\left(\frac{2}{n_1 x+n_2}\right)^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1+n_2}{2})$$

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

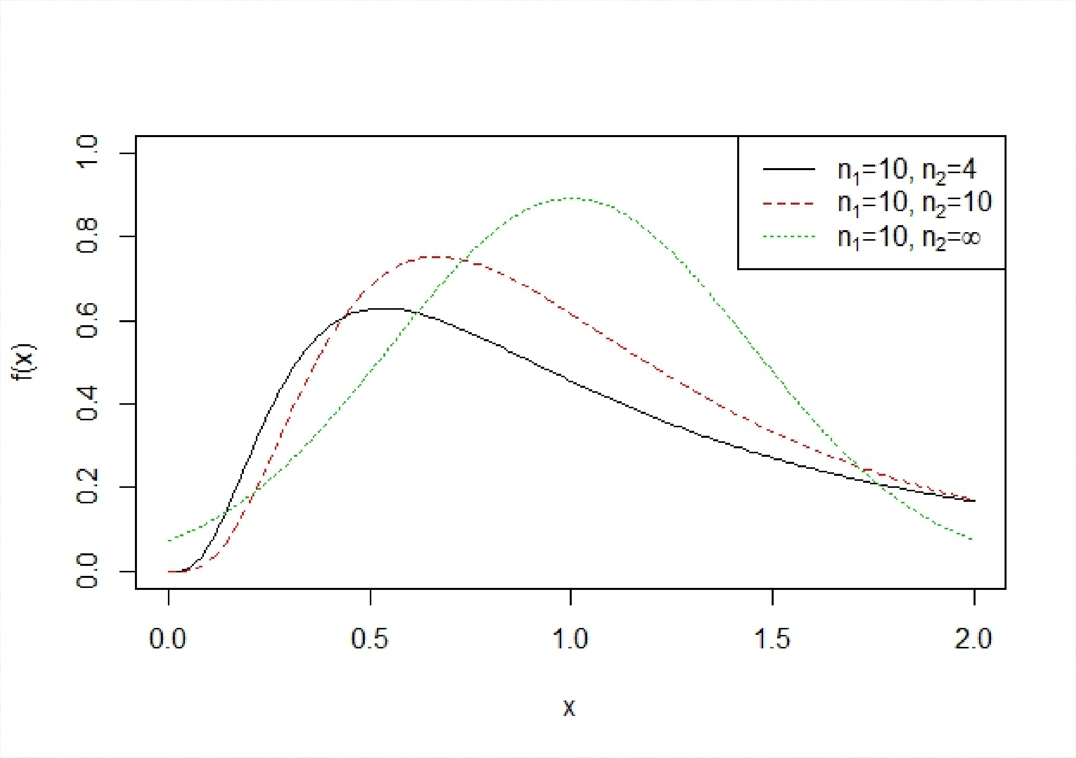
这就是$F(n_1,n_2)$的概率密度函数,由此证明定理1.9

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

Thank You!