数理统计

冯伟

数学科学学院

wfeng_323@buaa.edu.cn

上页 下页 返回

Γ分布族

若连续型 **r.v** ***X***具有概率密度

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

则称**X**服从参数为α,β的**Γ**分布,记作$X\sim\Gamma(\alpha,\beta)$

其中α,β均为正常数,分别称为形状参数和尺度参数。$\Gamma(\alpha)=\int_0^{+\infty}t^{\alpha-1}e^{-t}dt$是含参变量的广义积分。

上页 下页 返回

当$\alpha = n/2, \beta = 1/2$时, Γ分布则是统计学中十分重要的$\chi^2(n)$分布,其概率密度为

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

性质1 设$X \sim \Gamma(\alpha, \beta)$，则$E(X)= \alpha/\beta$, $D(X)= \alpha/\beta^2$.

性质2 设$X_i \sim \Gamma(\alpha_i, \beta)$， $i=1,2,\ldots,n$，且X_i相互独立，则$X_1+X_2+\ldots+X_n \sim \Gamma(\alpha_1+\alpha_2+\ldots+\alpha_n, \beta)$

多元正态分布族

定义　如果p维随机向量 (随机变量)

$$X=(X_1,X_2,\cdots,X_p)^T$$

(联合)概率密度函数为

$$f(x_1,x_2,\cdots,x_p)=\frac{1}{(2\pi)^{\frac{p}{2}}|V|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(X-\mu)^T V^{-1}(X-\mu)\right\}$$

则称随机向量X为p维正态随机向量，其中μ称为均值向量，V为协方差矩阵(协差阵)，且$V>0$.

上页　下页　返回

多元正态分布的性质：

（1） p维正态分布由其均值向量和协方差阵唯一确定。

（2） 对于任一 p维向量μ及p阶非负定矩阵V，必存在p维正态随机向量$X \sim N_p(\mu, V)$。

（3） 设$X \sim N_p(\mu, V)$，A是$m \times p$常数矩阵，b是m维向量，若令$Y = AX + b$，则

$$Y \sim N_m(A\mu + b, AVA^T).$$

几种常用的统计量的分布

- 正态总体样本的线性函数的分布
- χ^2—分布
- t—分布
- F—分布

正态总体样本线性函数的分布

定理1 设总体$X \sim N(\mu,\sigma^2)$,（$X_1,X_2,\ldots,X_n$)是总体的容量为n的样本,令

$$U=a_1x_1+a_2x_2+\ldots+a_nx_n,$$

其中$a_1,a_2,\ldots,a_n$是已知常数，则U也是正态随机变量，其均值、方差分别为

$$E(U)=\mu\sum_{i=1}^{n}a_i$$

$$D(U)=\sigma^2\sum_{i=1}^{n}a_i^2$$

定理2 设总体$X\sim N(\mu,\sigma^2)$，$(X_1,X_2,\ldots,X_n)$是总体的容量为n的样本,$A=(a_{ij})$是$p\times n$阶矩阵。记$Y=(Y_1,Y_2,\ldots,Y_p)'=A(X_1,X_2,\ldots,X_n)'$，则$Y_1,Y_2,\ldots,Y_p$也是正态随机变量，其均值、方差、协方差分别为

$$E(Y_i)=\mu\sum_{i=1}^{n}a_{ij}\qquad D(Y_i)=\sigma^2\sum_{i=1}^{n}a_{ij}^{2}$$

$$Cov(Y_i,Y_j)=\sigma^2\sum_{k=1}^{n}a_{ik}a_{jk}$$

- 正交变换下的不变性

当$\mu=0$,且A是一$n\times n$阶正交矩阵时，$Y_1,Y_2,\ldots,Y_p$也相互独立，且服从$N(0,\sigma^2)$

上页 下页 返回

χ^2分布

定义　设随机变量$X_1, X_2, \ldots, X_n$相互独立且服从$N(0,1)$分布，则称随机变量

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

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

$$\chi^2 \sim \chi^2(n)$$

定理1 设随机变量$\chi^2 \sim \chi^2(n)$，则χ^2的密度函数为

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

上页 下页 返回

定理1的证明梗概：

- 首先写出随机向量（$X_1,X_2,\dots,X_n$)的概率密度函数；
- 其次利用随机向量的函数的分布的知识来直接求χ^2的分布函数（此时可利用积分变换显式求到分布函数，再求导得到密度函数，或直接用夹逼定理求到密度函数）；
- 最后再利用Γ 函数的性质和密度函数的性质确定积分常数。

上页 下页 返回

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} x_i^{\,2})$$

$$= \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} \exp(-\frac{1}{2} \sum_{i=1}^{n} x_i^{\,2})$$

$$F(y) = P(\chi^2 \le y)$$

$$= \frac{1}{(2\pi)^{n/2}} \int \underset{\sum_{i=1}^{n} X_i^{\,2} \le y}{\cdots} \int \exp(-\frac{1}{2} \sum_{i=1}^{n} x_i^{\,2}) dx_1 dx_2 \cdots dx_n$$

上页 下页 返回

$$\begin{cases} x_1 = \rho\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-1} \\ x_2 = \rho\cos\theta_1\cos\theta_2\cdots\sin\theta_{n-1} \\ \vdots \\ x_n = \rho\sin\theta_1 \end{cases}$$

$$\mathrm{P}\left\{y < \mathrm{X}_1^{\ 2} + \mathrm{X}_2^{\ 2} + \cdots + \mathrm{X}_{\mathrm{n}}^{\ 2} \le y + h\right\}$$

$$= \frac{1}{(2\pi)^{n/2}} \int \underset{G}{\cdots} \int \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^{\ 2}\right) dx_1 dx_2 \cdots dx_n$$

这里G是外半径为$\sqrt{\mathrm{y+h}}$,内半径为$\sqrt{y}$的n维球壳。

设G的体积为∇V，则

$$\frac{1}{(2\pi)^{n/2}} \mathrm{e}^{-\frac{\mathrm{y+h}}{2}} \nabla \mathrm{V} \le \mathrm{P}\left\{y < \mathrm{X}_1^{\ 2} + \mathrm{X}_2^{\ 2} + \cdots + \mathrm{X}_{\mathrm{n}}^{\ 2} \le y + h\right\}$$

$$\le \frac{1}{(2\pi)^{n/2}} \mathrm{e}^{-\frac{\mathrm{y}}{2}} \nabla \mathrm{V}$$

定理2 设$X \sim \chi^2(n)$，则$E(X)=n, D(X)=2n$

定理3 设$X_1 \sim \chi^2(n_1)$， $X_2 \sim \chi^2(n_2)$，且X_1与X_2相互独立，则$X_1 + X_2 \sim \chi^2(n_1 + n_2)$

定理4 (Cochran) 设随机变量$X_1, X_2, \ldots, X_n$相互独立且服从$N(0,1)$分布，又设

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

其中Q_j是秩为n_j的（$X_1, X_2, \ldots, X_n$)的非负定二次型。则Q_j相互独立，且分别服从于自由度为n_j的χ^2分布的充要条件是

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

定理5(抽样分布基本定理) 设（$X_1, X_2, \ldots, X_n$）是来自总体$N(\mu,\sigma^2)$的一个样本，则

（1）$\overline{X}$ 与 S^2 相互独立；

（2）$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

注：1. $\overline{X}$与S^2的独立性仅当总体分布为正态时才成立。当总体分布的三阶中心矩为零时，可以推出两者是不相关的。

2. $\overline{X}$ 服从精确的正态分布也只有在总体为正态分布时才成立。

特征函数

- 定义 称随机变量e^{itX}的数学期望$\phi(t)=Ee^{itX}$为X的特征函数。
- 随机变量的特征函数$\phi(t)$是实变量t的复值函数，总是存在的,且与随机变量一一对应。
- 当X为连续型时, $\phi(t)=\int_{-\infty}^{\infty}e^{itx}f(x)dx$
- 当X为离散型时, $\phi(t)=\sum_{k=-\infty}^{\infty}e^{itx}p_k$，$p_k=P\{X=x_k\}$.

特征函数基本性质

(1) $\phi(0)=1, \phi(-t)=\overline{\phi(t)}, |\phi(t)| \leq \phi(0)=1$;

(2) 特征函数在$(-\infty, \infty)$上一致连续；

(3) 若$Y=aX+b, a, b$是常数，则Y的特征函数

$\phi_Y(t) = e^{ibt}\phi_X(at)$, 其中$\phi_X(t)$是$X$的特征函数；

(4) 两个相互独立的随机变量之和的特征函数等于各个特征函数之积；

(5) 两个分布函数相等当且仅当它们所对应的特征函数相等；

（6）在$F(x)$的连续点a,b上，有

$$F(b)-F(a)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\phi(t)dt;$$

（7）函数$\phi(t)$为特征函数的充要条件是：

$\phi(t)$非负定，连续且$\phi(0)=1$;

（8）设随机变量X的n阶矩存在，则X的特征函数$\phi(t)$的k阶导数$\phi^{(k)}(t)$存在，且

$$E(X^k)=i^{(-k)}\phi^{(k)}(0),(k\le n)$$

上页 下页 返回

几个常见随机变量的特征函数

(1) 设X服从两点分布，则其特征函数为$\phi(t)=q+pe^{it}$;

(2) 设X服从二项分布$B(n,p)$，则其特征函数为

$$\phi(t)=(q+pe^{it})^n;$$

(3) 设$X\sim U[-a,a]$, 则其特征函数为$\phi(t)=\dfrac{\sin at}{at}$;

(4) 设X服从参数为λ的指数分布，则其特征函数为

$$\phi(t)=(1-\frac{it}{\lambda})^{-1};$$

(5) 设$X\sim N(\mu,\sigma^2)$, 则其特征函数为$\phi(t)=e^{i\mu t-\frac{1}{2}\sigma^2t^2}$;

上页 下页 返回

χ^2分布中几个定理的证明

- **定理2的证明：**

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

$$=\sum_{i=1}^{n}(D(X_i)+(E(X_i))^2)=n$$

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

$$=\sum_{i=1}^{n}(E(X_i^{\,4})-(E(X_i^{\,2}))^2)=2n$$

这里用了$E(X_i^{\,4})=3$这个结果，大家自己验证。

上页 下页 返回

引理 设$X \sim \chi^2(n)$，则X的特征函数为

$$\phi(t)=(1-2it)^{-n/2}.$$

- **定理3的证明**：

由引理知，X_1的特征函数为$\phi_1(t)=(1-2it)^{-\frac{n_1}{2}}$，

X_2的特征函数为$\phi_2(t)=(1-2it)^{-\frac{n_2}{2}}$.

由特征函数的性质，

X_1+X_2的特征函数为$\phi(t)=\phi_1(t)\phi_2(t)=(1-2it)^{-(n_1+n_2)/2}$，

由一一对应性，知X_1+X_2服从自由度为n_1+n_2的χ^2分布。

- **定理2的又一证明：根据引理及特征函数性质8**

得$E(X)=n, E(X^2)=n^2+2n$,所以$D(X)=2n$

上页 下页 返回

t一分布

- 定义 设$X\sim N(0,1)$, $Y\sim \chi^2(n)$,且X和Y相互独立，则称随机变量 $T=\dfrac{X}{\sqrt{Y/n}}$ 所服从的分布是自由度为n的t分布，记为$T\sim t(n)$.
- 定理1 设$T\sim t(n)$,则T的概率密度为

$$f(t)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

上页 下页 返回

定理2 设（$X_1,X_2,\ldots,X_n$)是来自总体$N(\mu,\sigma^2)$的一个样本，则有 $\frac{(\overline{X}-\mu)\sqrt{n}}{S} \sim t(n-1)$ 。

定理3 设（$X_1,X_2,\ldots,X_m$)和（$Y_1,Y_2,\ldots,Y_n$)是分别来自总体$N(\mu_1,\sigma^2)$ 和$N(\mu_2,\sigma^2)$的样本，且假定它们相互独立，则有

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{(m-1)S_1^2+(n-1)S_2^2}} \cdot \sqrt{\frac{mn(m+n-2)}{m+n}}$$

$$\sim t(m+n-2)$$

F—分布

定义　设随机变量X和Y是自由度分别为n_1和n_2的相互独立的χ^2分布随机变量，则称随机变量

$$F=\frac{X/n_1}{Y/n_2}$$

所服从的分布为自由度是(n_1,n_2)的F分布，记为$F \sim F(n_1,n_2)$. 其中n_1称为第一自由度，n_2称为第二自由度。

定理1 设$F \sim F(n_1,n_2)$,则F的概率密度为

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

定理2 若$X/\sigma^2 \sim \chi^2(n_1)$, $Y/\sigma^2 \sim \chi^2(n_2)$, 且相互独立，则$F=\dfrac{X/n_1}{Y/n_2} \sim F(n_1,n_2)$.

定理3 若$X \sim F(n_1,n_2)$,则$1/X \sim F(n_2,n_1)$.

定理4 若$X \sim t(n)$, 则$X^2 \sim F(1,n)$.

定理5 设（$X_1,X_2,\ldots,X_m$)和（$Y_1,Y_2,\ldots,Y_n$)是分别来自总体$N(\mu_1,\sigma_1^2)$ 和$N(\mu_2,\sigma_2^2)$的样本，且假定它们相互独立，则有

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

定理6 设$X_n \sim F(m,n)$, 则当$n \to \infty$时,

$$X_n \xrightarrow{L} \frac{1}{m}\chi_m^2.$$

定理7 设随机变量$X_1, X_2, \ldots, X_n$相互独立且服从 $N(0, \sigma^2)$，又设

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

其中Q_j是秩为n_j的（$X_1, X_2, \ldots, X_n$)的非负定二次型。若$n_1 + n_2 + \ldots + n_k = n$, 则$Q_j$相互独立, 且

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

上页 下页 返回

分位点（分位数）

定义1 设随机变量X的分布函数为$F(x)$，$0<\alpha<1$，若x_α使

$$P\{X \le x_\alpha\}=F(x_\alpha)=\alpha,$$

则称x_α为此概率分布的(下侧)α分位点（或分位数）。

分位点（分位数）

- 当$X \sim N(0,1)$,将其分位点记为u_α,或z_α.
- 当$X \sim \chi^2(n)$,将其分位点记为$\chi_\alpha^2(n)$.
- 当$X \sim t(n)$,将其分位点记为$t_\alpha(n)$.
- 当$X \sim F(m,n)$,将其分位点记为$F_\alpha(m,n)$.

上面几类分位点的性质

1. $-z_\alpha = z_{1-\alpha}$, $-t_\alpha(n) = t_{1-\alpha}(n)$
2. $F_\alpha(m,n) = 1/F_{1-\alpha}(n,m)$

非正态总体大样本的渐近分布

当样本容量n趋于无穷时，若统计量的分布趋于一定的分布，则称后者为该统计量的极限分布。它提供了统计推断的一种近似解法。所谓大样本指样本容量$n>30$,最好大于50或100.

定理1 设总体X的分布函数为$F(x)$,

$$E(X)=\mu_F,\ D(X)=\sigma_F^2,\ 0<\sigma_F^2<+\infty,$$

$X_1,X_2,\ldots,X_n$ 为来自总体X的样本，则样本的均值的渐近分布为 $N(\mu_F,\dfrac{\sigma_F^2}{n})$.

定理2 设总体X的分布函数为$F(x)$,

$$E(X)=\mu_F,\ D(X)=\sigma_F^2,\ 0<\sigma_F^2<+\infty,$$

$X_1,X_2,\ldots,X_n$ 为来自总体X的样本，则

$$\sqrt{n}\,\frac{\overline{X}_n-\mu_F}{S_n}\xrightarrow{L}N(0,1)\quad(n\to\infty)$$

上页 下页 返回

定理3 设（$X_1,X_2,\dots,X_m$) 与（$Y_1,Y_2,\dots,Y_n$) 是来自总体X与Y的两独立样本，$EX=\mu_1,DX=\sigma_1^2$，$EY=\mu_2,DY=\sigma_2^2$，则当n趋于无穷，m趋于无穷时有

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{m}+\dfrac{S_2^2}{n}}}\xrightarrow{L}N(0,1)$$

上页 下页 返回

作业**1.　设$X\sim\Gamma(\alpha,\beta)$，求 $D(X)$.**

2.　证明抽样分布基本定理