上述定义中Γ函数为:

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

根据上式有$\Gamma(1)=1$, $\Gamma(1/2)=\sqrt{\pi}$.
τ分布的可加性:

定理8.2. *若$X\sim\Gamma(\alpha_1,\lambda)$, $Y\sim\Gamma(\alpha_2,\lambda)$, 且$X$与$Y$独立, 则$X+Y\sim\Gamma(\alpha_1+\alpha_2,\lambda)$.*

上述定理的证明留作习题. 特别地, 当$\alpha=1/2$和$\lambda=1/2$时, 有

$$f(x)=\begin{cases}\frac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}e^{-\frac{1}{2}} & x>0\\ 0 & x\le 0\end{cases}$$

例8.2. *若$X\sim\mathcal{N}(0,1)$, 则$X^2\sim\Gamma(1/2,1/2)$.*

解. 首先求解随机变量函数$Y=X^2$的分布函数: 当$y>0$时,

$$F_Y(y)=\Pr(X^2\le y)=\Pr(-\sqrt{y}\le X\le\sqrt{y})=\int_{-\sqrt{y}}^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

由此得到概率密度为$f_Y(y)=\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-\frac{y}{2}}$. 当$y\le 0$时有$f_Y(y)=0$. 从而得到$X^2\sim\Gamma(1/2,1/2)$. □

8.3 正态总体抽象分布定理

8.3.1 χ^2分布

定义8.4. *若$X_1,X_2,\cdots,X_n$是总体为$\mathcal{N}(0,1)$的一个样本, 称$Y=X_1^2+X_2^2+\cdots+X_n^2$为服从自由度为$n$的$\chi^2$分布, 记为$Y\sim\chi^2(n)$.*

根据$X_1^2\sim\Gamma(1/2,1/2)$, 以及Γ函数的可加性, 可得$Y\sim\Gamma(n/2,1/2)$. 因此Y的概率密度为

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

下面研究χ^2分布的性质:

定理8.3. *若$X\sim\chi^2(n)$, 则$E(X)=n$, $Var(X)=2n$; 若$X\sim\chi^2(m)$, $Y\sim\chi^2(n)$且独立, 则$X+Y\sim\chi^2(m+n)$;*

Proof. 若$X\sim\chi^2(n)$, 设$X=X_1+X_2+\cdots+X_n$, 其中$X_1,X_2,\cdots,X_n$独立同分布于$\mathcal{N}(0,1)$, 则有

$$\begin{aligned}E[X] &= E[X_1^2+X_2^2+\cdots+X_n^2]=nE[X_1^2]=n,\\ \mathrm{Var}(X) &= n\mathrm{Var}(X_1^2)=n[E(X_1^4)-(E(X_1^2))^2]=n(E(X_1^4)-1).\end{aligned}$$

计算

$$E(X_1^4)=\int_{-\infty}^{+\infty}\frac{x^4}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=3\int_{-\infty}^{+\infty}\frac{x^2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=3$$

可得$\mathrm{Var}(X)=2n$. □

更一般的结论: 若$X\sim\mathcal{N}(0,1)$, 则

$$E(X^k)=\begin{cases}(k-1)!! & k\text{为偶数}\\ 0 & k\text{为奇数}\end{cases}$$

其中$(2k)!!=2k\cdot(2k-2)\cdot\dots\cdot 2$, $(2k+1)!!=(2k+1)\cdot(2k-1)\cdot\dots\cdot 1$.

例8.3. 设X_1,X_2,X_3,X_4是来自于$\mathcal{N}(0,4)$的样本, $Y=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$. 求$a,b$ 取何值时, Y服从χ^2分布, 并求其自由度.

解. 根据正太分布的性质有$X_1-2X_2\sim\mathcal{N}(0,20)$和$3X_3-4X_4\sim\mathcal{N}(0,100)$, 因此

$$\frac{X_1-2X_2}{2\sqrt{5}}\sim\mathcal{N}(0,1),\qquad \frac{3X_3-4X_4}{10}\sim\mathcal{N}(0,1),$$

所以当$a=\frac{1}{2\sqrt{5}}$, $b=\frac{1}{10}$时有$Y\sim\chi^2(2)$成立. □

分布可加性:

- 如果$X\sim\mathcal{N}(\mu_1,a_1^2)$和$Y\sim\mathcal{N}(\mu_2,a_2^2)$, 且$X$与$Y$独立, 那么$X\pm Y\sim\mathcal{N}(\mu_1\pm\mu_2,a_1^2+a_2^2)$;
- 如果$X\sim B(n_1,p)$和$Y\sim B(n_2,p)$, 且X与Y独立, 那么$X+Y\sim B(n_1+n_2,p)$;
- 如果$X\sim P(\lambda_1)$和$Y\sim P(\lambda_2)$, 且X与Y独立, 那么$X+Y\sim P(\lambda_1+\lambda_1)$;
- 如果$X\sim\Gamma(\alpha_1,\lambda)$和$Y\sim\Gamma(\alpha_2,\lambda)$, 且$X$与$Y$独立, 那么$X+Y\sim\Gamma(\alpha_1+\alpha_2,\lambda)$.

8.3.2 t分布

定义8.5. 随机变量$X\sim\mathcal{N}(0,1)$, $Y\sim\chi^2(n)$, X与Y独立, 则$T=X/\sqrt{Y/n}$服从自由度为n的t-分布, 记$T\sim t(n)$.

随机变量$T\sim t(n)$, 其概率密度为

$$f(x)=\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}\qquad(x\in\mathbb{R}).$$

由此可知t分布的密度函数$f(x)$是偶函数. 当$n\to\infty$时,

$$f(x)\to\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

因此当n足够大时, $f(x)$可被近似为$\mathcal{N}(0,1)$的密度函数.

8.3.3 F分布

定义8.6. 设随机变量$X \sim \chi^2(m)$, $Y \sim \chi^2(n)$, 且X与Y独立, 称$F = \frac{X/m}{Y/n}$服从自由度为(m,n)的F-分布, 记$F \sim F(m,n)$,

随机变量$F \sim F(m,n)$的概率密度为

$$f(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})(\frac{m}{n})^{\frac{m}{2}} x^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(1+\frac{mx}{n})^{\frac{m+n}{2}}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

注: 若$F \sim F(m,n)$, 则$\frac{1}{F} = F(n,m)$.

课题练习:

1) 随机变量$X_1, X_2, \cdots, X_n$相互独立, 且$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, 求$\sum_{i=1}^n (X_i - \mu_i)^2/\sigma_i^2$的分布.

2) 总体$X \sim \mathcal{N}(0,9)$与总体$Y \sim \mathcal{N}(0,9)$独立, $X_1, X_2, \cdots, X_9$和$Y_1, Y_2, \cdots, Y_9$分别为来自总体X和Y的两样本, 求$(X_1 + X_2 + \cdots + X_9)/\sqrt{Y_1^2 + Y_1^2 + \cdots + Y_n^2}$的分布.

3) 设X_1, X_2, X_{2n}来自总体$\mathcal{N}(0, \sigma_2)$的样本, 求$(X_1^2 + X_3^2 + \cdots + X_{2n-1}^2)/(X_2^2 + X_4^2 + \cdots + X_{2n}^2)$的分布.

8.3.4 正态分布的抽样分布定理

定理8.4. 设$X_1, X_2, \cdots, X_n$是取自总体$\mathcal{N}(\mu, \sigma^2)$的样本, 则有

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}), \qquad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1).$$

定理8.5. 随机变量$X_1, X_2, \cdots, X_n$是取自总体$\mathcal{N}(\mu, \sigma^2)$的样本, 设$\bar{X} = \sum_{i=1}^n X_i/n$和$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, 则有

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1), \qquad \text{且 } \bar{X} \text{ 和 } S^2 \text{ 相互独立.}$$

此定理证明参考书的附件.

定理8.6. 设$X_1, X_2, \cdots, X_n$是取自总体$\mathcal{N}(\mu, \sigma^2)$的样本, 则$\bar{X} = \sum_{i=1}^n X_i/n$与$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$相互独立, 且

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

Proof. 有前面两个定理可知$(\bar{X} - \mu)/\sigma\sqrt{n} \sim \mathcal{N}(0,1)$和$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, 于是有

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

□

定理8.7. 设$X \sim \mathcal{N}(\mu_1, \sigma^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, 且$X$与$Y$独立. 设$X_1, X_2, \cdots, X_m$, $Y_1, Y_2, \cdots, Y_n$分别来自总体X和Y的两个样本, $\bar{X}, \bar{Y}$为样本均值, S_X^2和S_Y^2为修正样本方差, 则

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(n_1 + n_2 - 2).$$

Proof. 对正太分布有$\bar{X} \sim \mathcal{N}(\mu_1, \sigma^2/m)$和$\bar{Y} \sim \mathcal{N}(\mu_2, \sigma^2/n)$, 并$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2\right)$, 进一步有

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma^2\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim \mathcal{N}(0, 1).$$

根据定理 8.5 有$\frac{(m-1)S_X^2}{\sigma^2} \sim \chi^2(m-1)$和$\frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2(n-1)$, 由此得到

$$\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi^2(m + n - 2).$$

从而完成证明. □

定理8.8. 设$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, 且$X$与$Y$独立. 设$X_1, X_2, \cdots, X_m$, $Y_1, Y_2, \cdots, Y_n$分别来自总体X和Y 的两个样本, S_X^2和S_Y^2为样本的修正样本方差, 则有

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1, n-1).$$

Proof. 根据定理 8.5有$\frac{(m-1)S_X^2}{\sigma_X^2} \sim \chi^2(m-1)$和$\frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2(n-1)$, 所以

$$\frac{\frac{(m-1)S_X^2}{\sigma_X^2}/(m-1)}{\frac{(n-1)S_Y^2}{\sigma_Y^2}/(n-1)} \sim F(m-1, n-1).$$

□

课题练习题:

1) 若$X \sim t(n)$, 求$Y = X^2$的分布.

2) 若总体$X \sim \mathcal{N}(0, 1)$, $X_1, X_2, \cdots, X_5$为一样本. 设$Y = c_1(X_1 + X_3)^2 + c_2(X_2 + X_4 + X_5)^2$. 求常数$c_1, c_2$使$Y$服从$\chi^2$分布.

3) 设X_1, X_2是总体$\mathcal{N}(0, \sigma^2)$的样本, 求$\frac{(X_1+X_2)^2}{(X_1-X_2)^2}$的分布.

4) 设$X_1, X_2, \cdots, X_{10}$是总体$\mathcal{N}(\mu, \frac{1}{4})$的样本, i) 若$\mu = 0$, 求$\Pr(\sum_{i=1}^{10} X_i^2 \geq 4)$; ii) 若$\mu$未知, 求$\Pr(\sum_{i=1}^{10}(X_i - \bar{X})^2 \geq 2.85)$.

5) 设$X_1, X_2, \cdots, X_{25}$是总体$\mathcal{N}(12, \sigma^2)$的样本, i) 若$\sigma = 2$, 求$\Pr(\frac{1}{25}\sum_{i=1}^{25} X_i \geq 12.5)$; ii) 若$\sigma$未知, 求$\Pr(\frac{1}{25}\sum_{i=1}^{25} X_i \geq 12.5)$(此时样本方差$S^2 = 5.57$).

8.3.5 分位数(点)

定义8.7. 设X为一随机变量, 给定$\alpha \in (0,1)$, 称满足$\Pr(X > \lambda_\alpha) = \alpha$的实数$\lambda_\alpha$为上侧$\alpha$分位数*(点)*.

正态分布的分位点: $X \sim \mathcal{N}(0,1)$, 对给定$\alpha \in (0,1)$, 满足$\Pr(X > \mu_\alpha) = \int_{\mu_\alpha}^{\infty} f(x)dx = \alpha$的点$\mu_\alpha$称为正态分布上侧$\alpha$分位点. 由对称性可知$\mu_{1-\alpha} = -\mu_\alpha$.

$\chi^2(n)$分位点: $X \sim \chi^2(n)$, 对给定$\alpha \in (0,1)$, 满足$\Pr(X \geq \chi^2_\alpha(n)) = \alpha$的点$\chi^2_\alpha(n)$称为$\chi^2(n)$分布上侧$\alpha$分位点. 当$n \to \infty$时, 有$\chi^2_\alpha(n) \approx \frac{1}{2}(\mu_\alpha + \sqrt{2n-1})^2$, 其中$\mu_\alpha$表示正态分布上侧$\alpha$分位点.

t分布的分位点: $X \sim t(n)$, 对给定$\alpha \in (0,1)$, 满足$\Pr(X > t_\alpha(n)) = \alpha$的点$t_\alpha(n)$称为$t(n)$分布上侧$\alpha$分位点. 由对称性可知$t_{(1-\alpha)}(n) = -t_\alpha(n)$.

F分布的分位点: $X \sim F(m,n)$, 对给定$\alpha \in (0,1)$, 满足$\Pr[X > F_\alpha(m,n)] = \alpha$的点$F_\alpha(m,n)$称为$F(m,n)$分布上侧$\alpha$分位点. 对于$F$分布, 我们有如下性质:

引理8.2. 对F分布的分位点有$F_{(1-\alpha)}(m,n) = \frac{1}{F_\alpha(n,m)}$.

Proof. 设$X \sim F(m,n)$, 根据定义有

$$1 - \alpha = \Pr(X > F_{1-\alpha}(m,n)) = \Pr\left(\frac{1}{X} < \frac{1}{F_{1-\alpha}(m,n)}\right) = 1 - \Pr\left(\frac{1}{X} \geq \frac{1}{F_{1-\alpha}(m,n)}\right).$$

再根据$1/X \sim F(n,m)$, 结合上式有

$$\alpha = \Pr\left(\frac{1}{X} \geq \frac{1}{F_{1-\alpha}(m,n)}\right) = \Pr\left(\frac{1}{X} > \frac{1}{F_{1-\alpha}(m,n)}\right)$$

于是有$F_\alpha(n,m) = \frac{1}{F_{1-\alpha}(m,n)}$. □