Ch 9 统计的基本概念

回归前一次课

依分布收敛$Y_n \xrightarrow{d} a$: $\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$

- 林德贝格-勒维中心极限定理：独立同分布随机变量，若 $E[X_k] = \mu$和$\mathrm{Var}(X_k) = \sigma^2$, 则$\sum_{k=1}^{n} X_k \xrightarrow{d} N(n\mu, n\sigma^2)$
- 棣莫弗-拉普拉斯中心极限定理：若 $X_n \sim B(n,p)$，则 $X_n \xrightarrow{d} N\big(np, np(1-p)\big)$
- 李雅普诺夫定理: 独立不同分布中心极限定理

什么是统计学、研究的内容

基本概念：总体、个体、抽样、样本二重性、简单样本

统计量：样本均值、样本方差、修正后的样本方差、原点/中心矩

次序统计量

假设$X_1,X_2,\cdots,X_n$是来自总体X的一个样本，定义最小次序统计量和最大次序统计量分别为:

$$X_{(1)} = \min\{X_1,X_2,\cdots,X_n\}$$

$$X_{(n)} = \max\{X_1,X_2,\cdots,X_n\}$$

以及定义样本极差: $R_n = X_{(n)} - X_{(1)}$.

设总体X的分布函数为$F(x)$, 则有

$$F_{X_{(1)}}(x) = P\left(X_{(1)} \le x\right) = 1 - \left(1 - F(x)\right)^n$$

$$F_{X_{(n)}}(x) = F^n(x)$$

k次序统计量

设 $X_1, X_2, \cdots, X_n$ 是总体 X 的一个样本，总体的密度函数为$f(x)$，分布函数为$F(x)$，则第k次序统计量$X_{(k)}$的分布函数和密度函数分别为

$$F_k(x) = \sum_{r=k}^{n} \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r}$$

$$f_k(x) = \frac{n!}{(k-1)!\,(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x)$$

Γ-函数

对任意给定$\alpha > 0$, 定义Γ-函数为

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

又被称为第二类欧拉积分函数.

定理：对Γ-函数有$\Gamma(1) = 1$和$\Gamma(1/2) = \sqrt{\pi}$, 对$\alpha > 1$有

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

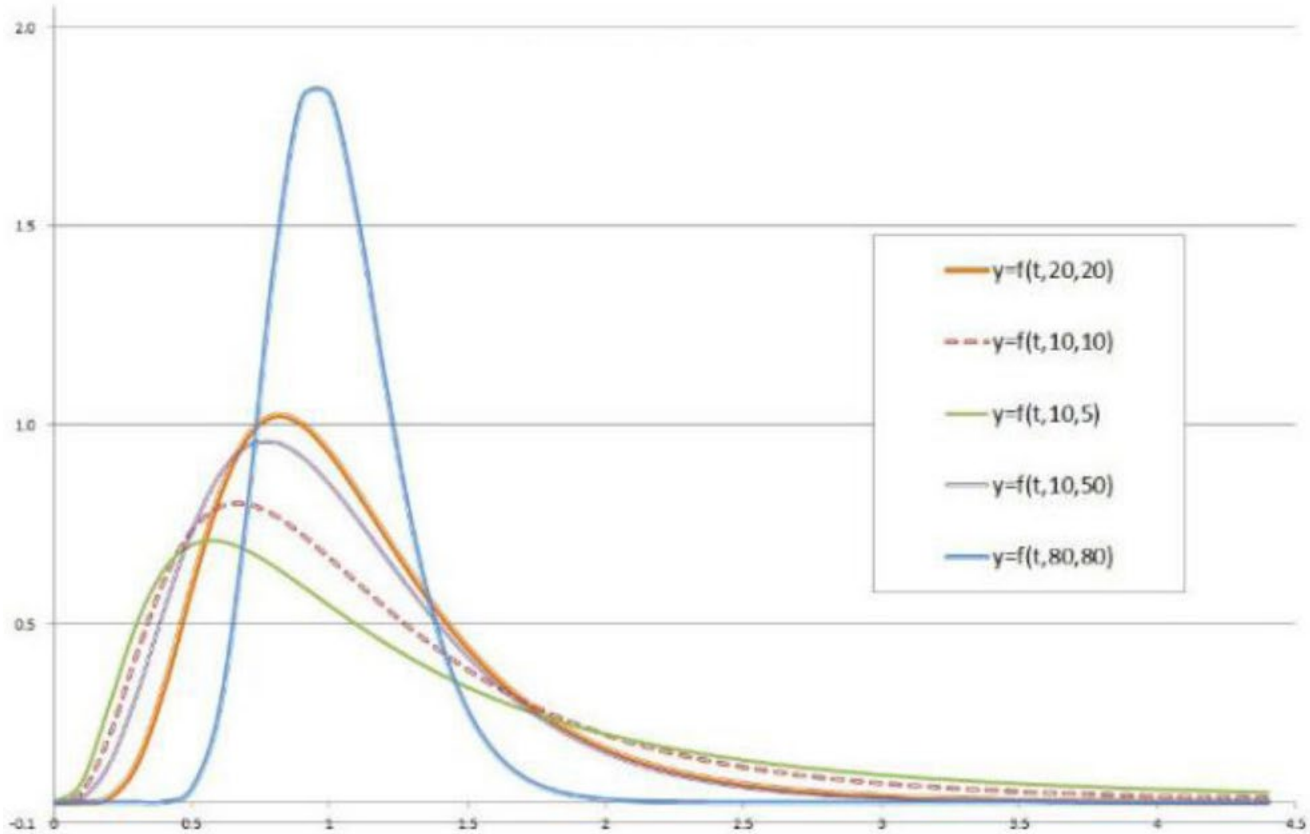
Γ分布

如果随机变量X的概率密度

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

其中$\alpha>0$和$\lambda>0$，则称随机变量X服从参数为α和λ的Γ分布，记为$X\sim\Gamma(\alpha,\lambda)$

Γ分布图像


2.0
1.5
1.0
0.5
-0.1
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
y=f(t,20,20)
y=f(t,10,10)
y=f(t,10,5)
y=f(t,10,50)
y=f(t,80,80)

Γ分布性质

定理： 若随机变量$X \sim \Gamma(\alpha, \lambda)$, 则有$E(X) = \alpha/\lambda$和$\mathrm{Var}\,(X) = \alpha/\lambda^2$

定理： 独立的随机变量$X \sim \Gamma(\alpha_1, \lambda)$和$Y \sim \Gamma(\alpha_2, \lambda)$,则

$$X + Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$$

Γ分布的可加性

若随机变量$X \sim \Gamma(1/2,1/2)$, 则其密度函数为

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

性质：若随机变量$X \sim N(0,1)$, 则有$X^2 \sim \Gamma(1/2,1/2)$

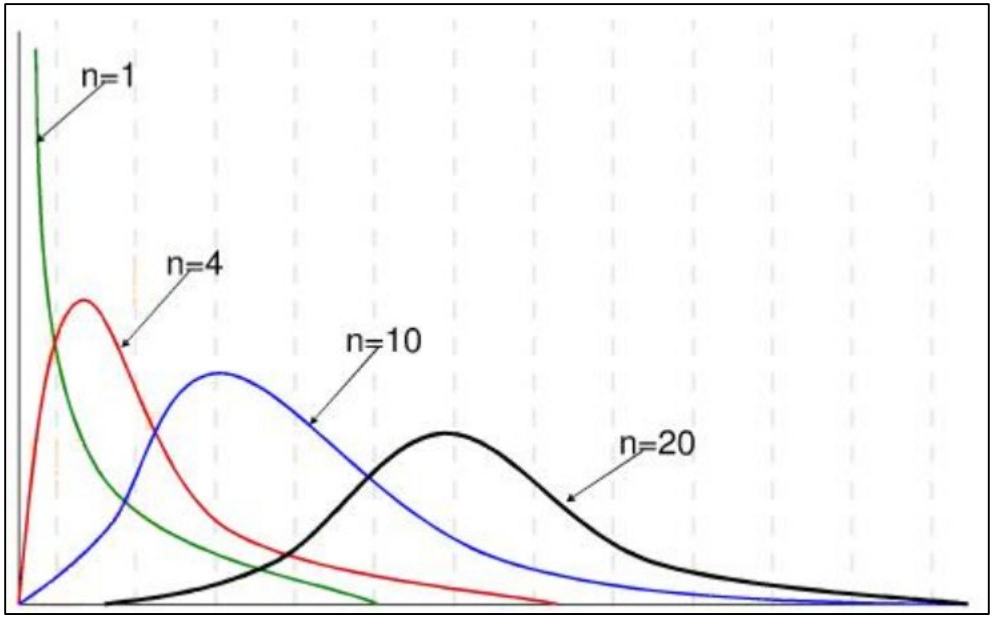
Ch 9.4 正态总体抽样分布定理

χ^2分布

若$X_1, X_2, \cdots, X_n$是来自总体$X \sim \mathrm{N}(0,1)$ 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χ^2分布, 记$Y \sim \chi^2(n)$



χ^2分布的概率密度函数

根据$X_1^2 \sim \Gamma(1/2,1/2)$和$\Gamma$分布的独立可加性可得$Y \sim \Gamma(n/2,1/2)$，于是有随机变量$Y$的概率密度为

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

χ^2分布的性质

定理： 随机变量$X\sim\chi^2(n)$, 则$E(X)=n$和$\mathrm{Var}(X)=2n$;

若随机变量$X\sim\chi^2(m)$和$Y\sim\chi^2(n)$相互独立，则$X+Y\sim\chi^2(m+n)$

若随机变量$X\sim \mathrm{N}(0,1)$, 则

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

其中 $(2k)!!=2k\cdot(2k-2)\cdots\cdot\ 2$

$(2k+1)!!=(2k+1)\cdot(2k-1)\cdots\cdot\ 1.$

例题

设X_1,X_2,X_3,X_4是来自于总体$N(0,4)$ 的样本, 以及

$$Y=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$$

求a,b取何值时, Y服从χ^2分布, 并求其自由度.

分布可加性

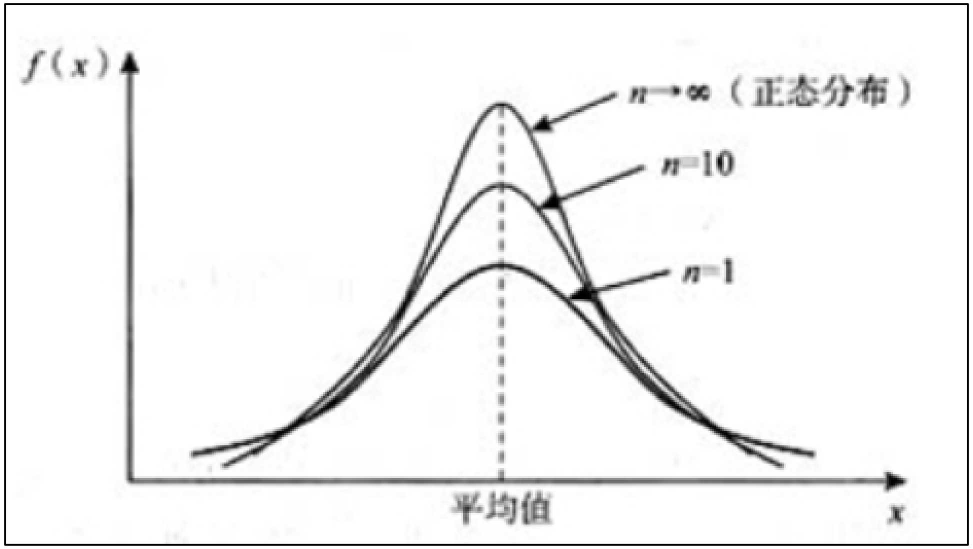
- 如果$X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ 和$Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$，且$X$与$Y$独立，那么$X \pm Y \sim \mathrm{N}(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$;
- 如果$X \sim B(n_1, p)$和$Y \sim B(n_2, p)$，且X与Y独立，那么$X + Y \sim B(n_1 + n_2, p)$;
- 如果$X \sim P(\lambda_1)$和$Y \sim P(\lambda_2)$,且X与Y独立, 那么$X + Y \sim P(\lambda_1 + \lambda_2)$;
- 如果$X \sim \Gamma(\alpha_1, \lambda)$ 和$Y \sim \Gamma(\alpha_2, \lambda)$，且$X$与$Y$独立，那么$X + Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$;
- 如果$X \sim \chi^2(m)$和$Y \sim \chi^2(n)$,且X与Y独立, 那么$X + Y \sim \chi^2(m + n)$.

t分布(student distribution)

随机变量$X \sim \mathrm{N}(0,1)$和$Y \sim \chi^2(n)$ 相互独立, 则随机变量

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

服从自由度为n的t-分布, 记$T \sim t(n)$.



t分布概率密度

随机变量$T \sim t(n)$的概率密度为

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

由此可知t-分布的密度函数$f(x)$ 是偶函数.

t分布的性质

定理： 当$n \to \infty$时, 随机变量$T \sim t(n)$的概率密度

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

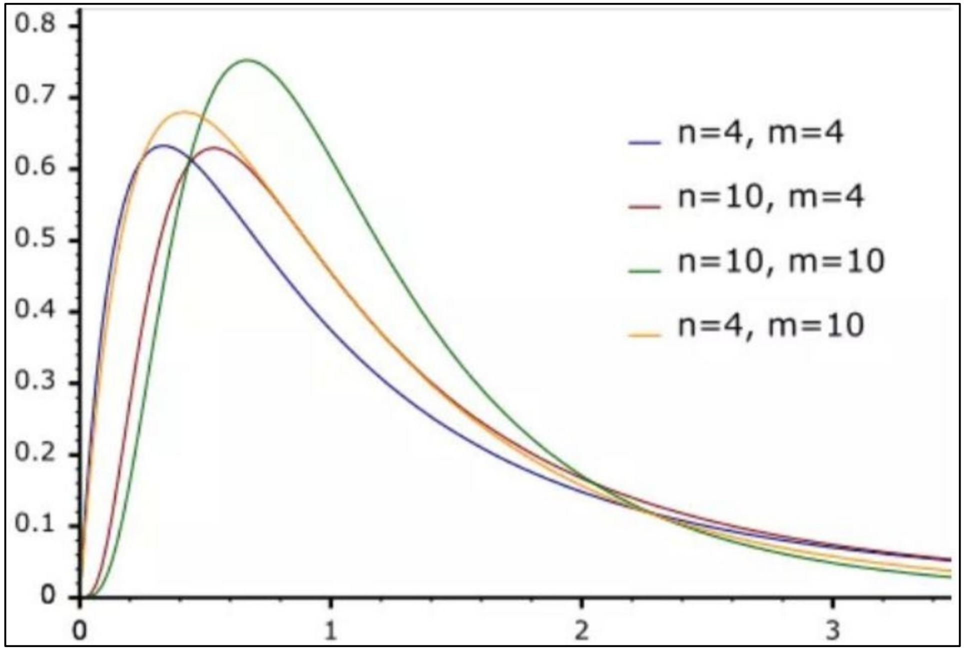
因此当n足够大时, $f(x)$ 被近似为$N(0,1)$ 的密度函数.

F分布

随机变量$X \sim \chi^2(m)$和$Y \sim \chi^2(n)$相互独立, 称随机变量

$$F = \frac{X/m}{Y/n}$$

服从自由度为(m,n) 的F-分布, 记$F \sim F(m,n)$.



F分布的性质

随机变量$F \sim F(m,n)$的概率密度为

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

定理： 若随机变量$F \sim F(m,n)$, 则$1/F = F(n,m)$.

例题

1. 独立同分布随机变量 $X_1, X_2, \cdots, X_n$ 满足 $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$，求 $\sum_{i=1}^{n}(X_i - \mu_i)^2/\sigma_i^2$的分布

2. $X_1, X_2, \cdots, X_9$和$Y_1, Y_2, \cdots, Y_9$是来自总体$N(0,9)$两样本，求$(X_1 + X_2 + \cdots + X_9)/\sqrt{Y_1^2 + Y_1^2 + \cdots + Y_9^2}$的分布.

3. 设$X_1, X_2, \ldots,\ X_{2n}$来自总体$N(0, \sigma_2^2)$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.