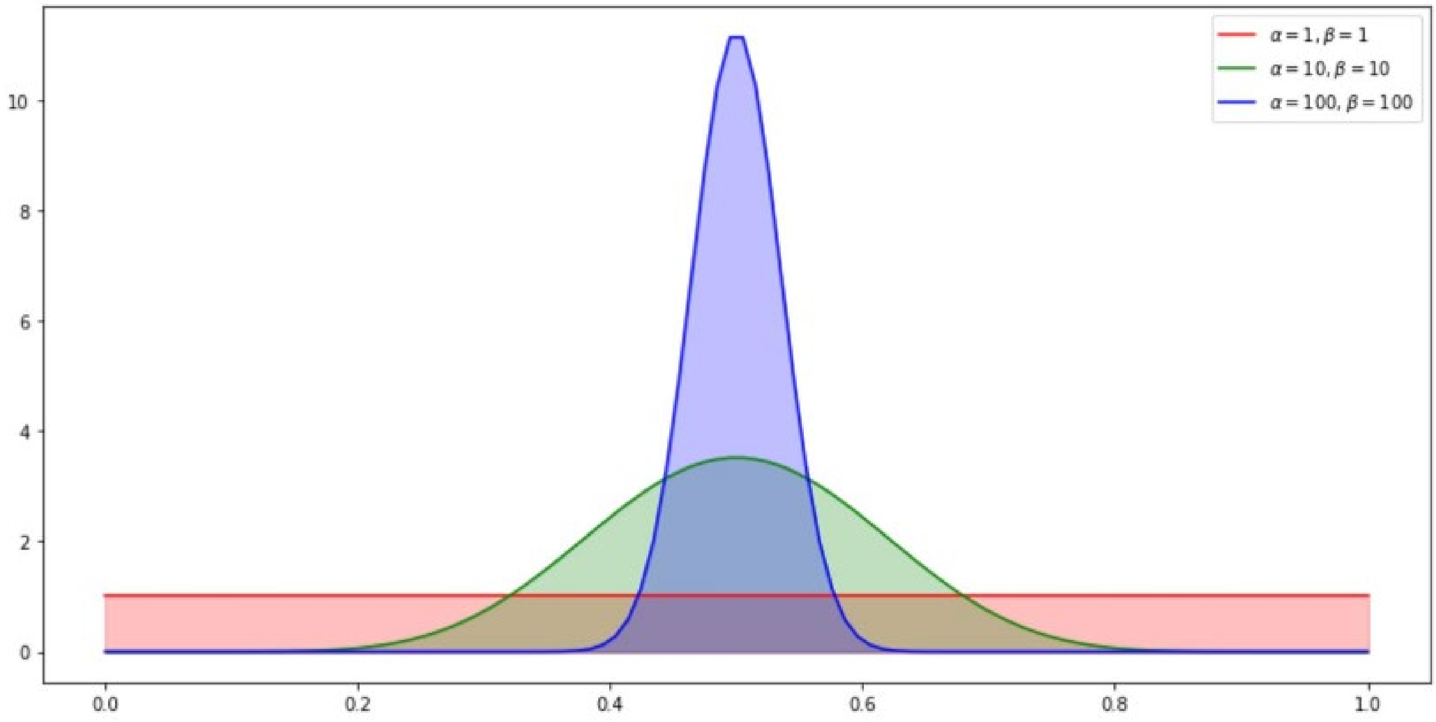
Ch 8 统计的基本概念

回顾前一次课

- 统计量: $g(X_1, X_2, \cdots, X_n)$ 关于$X_1, X_2, \cdots, X_n$连续、且不含任意参数
- **样本均值、样本方差、样本标准差、修正后的样本方差、** k阶原点矩/中心矩、次序统计量

- Beta函数：$\text{Beta}\,(\alpha_1, \alpha_2) = \int_0^1 x^{\alpha_1 - 1}\,(1 - x)^{\alpha_2 - 1} dx$
- Γ-函数为$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha - 1} e^{-x} dx$
- 两类函数的关系
- Beta分布

Beta分布图像



$\alpha = 1, \beta = 1$
$\alpha = 10, \beta = 10$
$\alpha = 100, \beta = 100$
10
8
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0

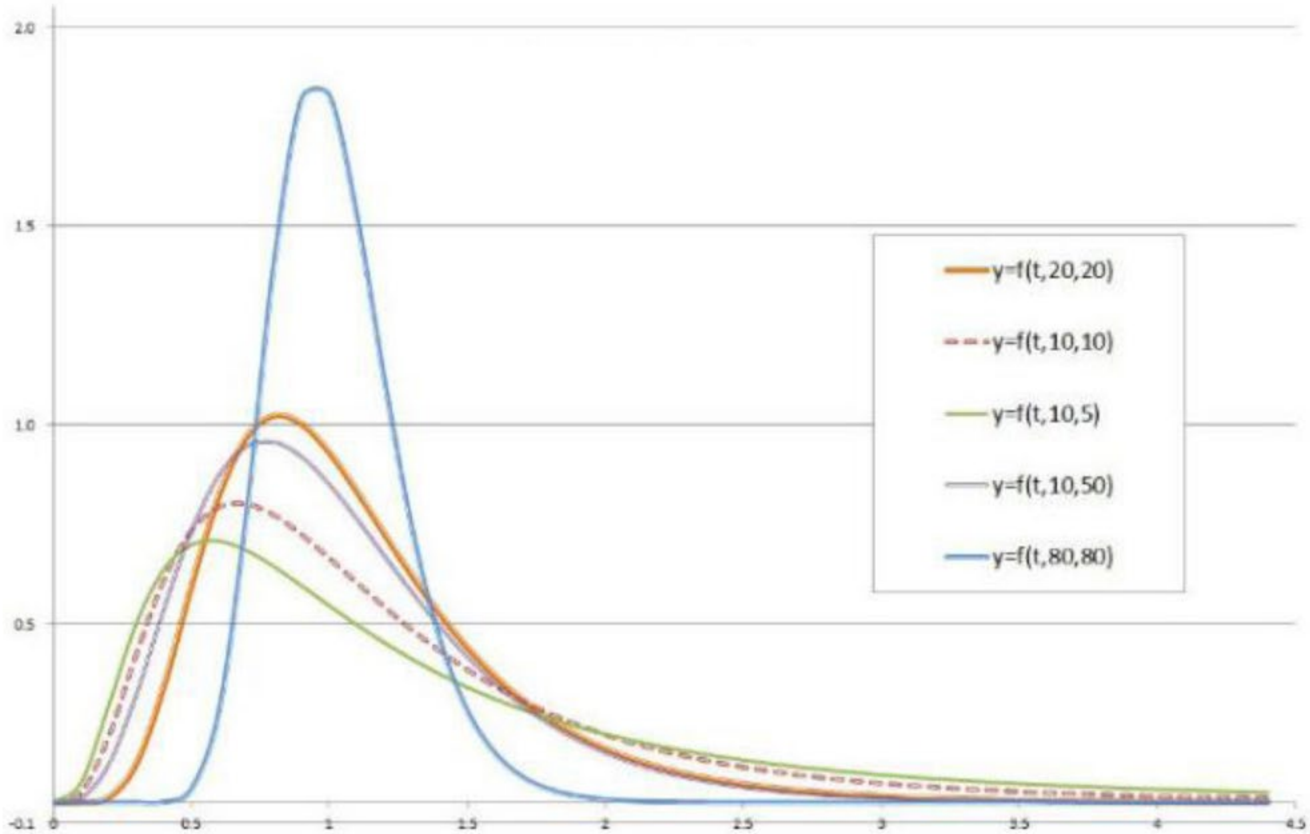
Γ分布

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$\mathrm{Var}\,(X) = \alpha/\lambda^2$

Γ分布的可加性

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

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

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

Dirichlet分布

给定$\alpha_1,\cdots,\alpha_k\in(0,+\infty)$，多元随机向量$X=(X_1,\cdots,X_k)$的密度函数为

$$f(x)=\begin{cases}\dfrac{x^{\alpha_1-1}x_2^{\alpha_2-1}\cdots x_k^{\alpha_k-1}}{\mathrm{B}(\alpha_1,\alpha_2,\ldots,\alpha_k)} & \sum_i x_i=1\quad x_i\in(0,1)\\ 0 & \text{其它}\end{cases}$$

则称X服从参数为$\alpha_1,\alpha_2,\cdots,\alpha_k$的Dirichlet分布，记$X\sim\mathrm{Dir}(\alpha_1,\alpha_2,\cdots,\alpha_k)$.

Dirichlet分布是Beta分布的一种推广，当$k=2$时Dirichlet分布退化为Beta分布.

Dirichlet分布的期望与协方差

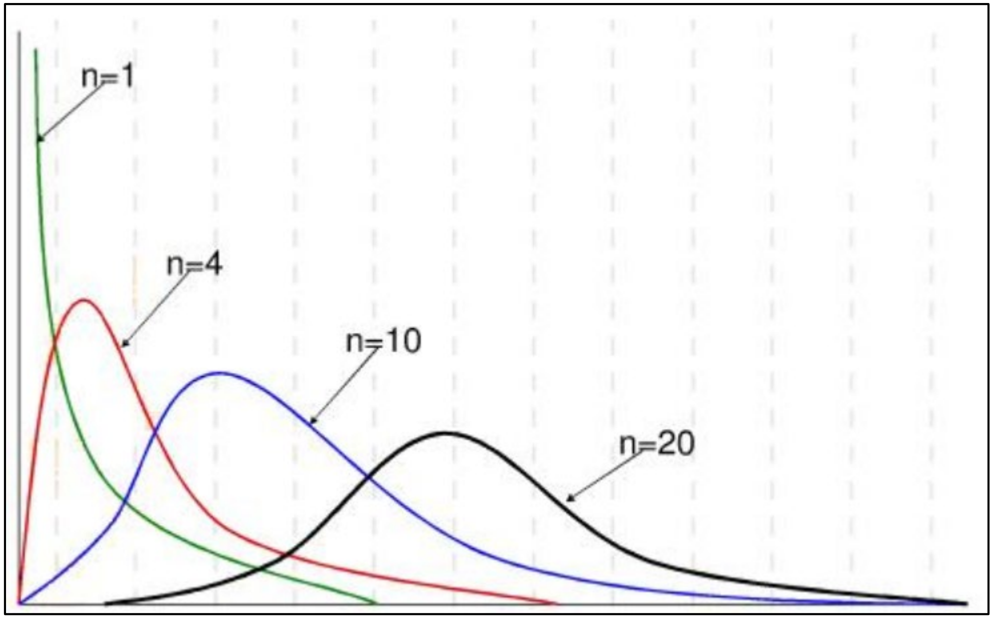
定理： 若随机向量$X=(X_1,\cdots,X_k)\sim \mathrm{Dir}(\alpha_1,\cdots,\alpha_k)$，设$\tilde{\alpha}=\alpha_1+\alpha_2+\cdots+\alpha_k$和$\tilde{\alpha}_i=\alpha_i/\tilde{\alpha}$, 则

$$E[X_i]=\tilde{\alpha}_i \qquad \mathrm{Cov}(X_i,X_j)=\begin{cases}\dfrac{\tilde{\alpha}_i(1-\tilde{\alpha}_i)}{\tilde{\alpha}+1} & i=j\\[2ex] \dfrac{\tilde{\alpha}_i\tilde{\alpha}_j}{\tilde{\alpha}+1} & i\neq j\end{cases}$$

Ch 8.4 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$\chi^2$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

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

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

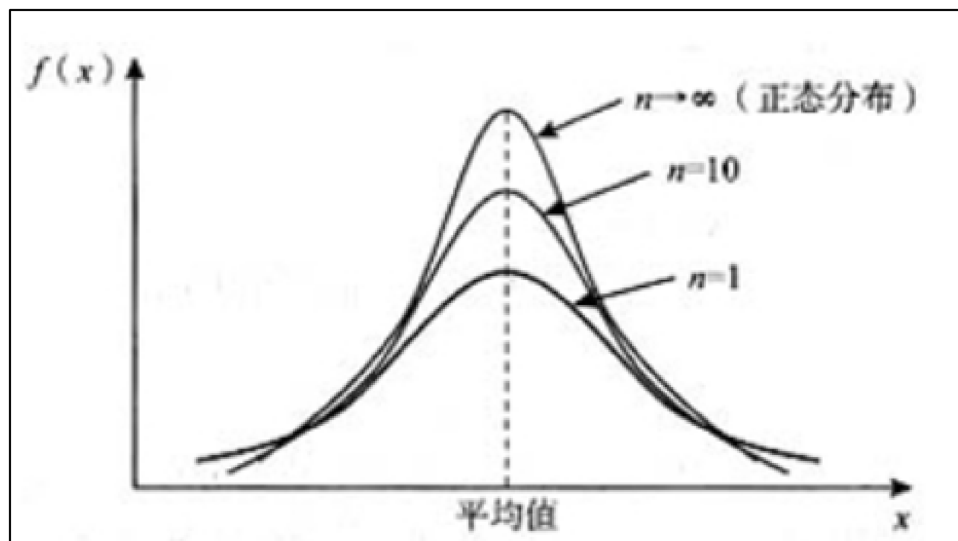
- 如果$X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ 和$Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$，且$X$与$Y$独立，那么$X \pm Y \sim \mathrm{N}(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$;
- 如果$X \sim B(n_1, p)$和$Y \sim B(n_2, p)$,且X与Y独立，那么$X + Y \sim B(n_1 + n_2, p)$;
- 如果$X \sim P(\lambda_1)$和$Y \sim P(\lambda_2)$,且X与Y独立，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