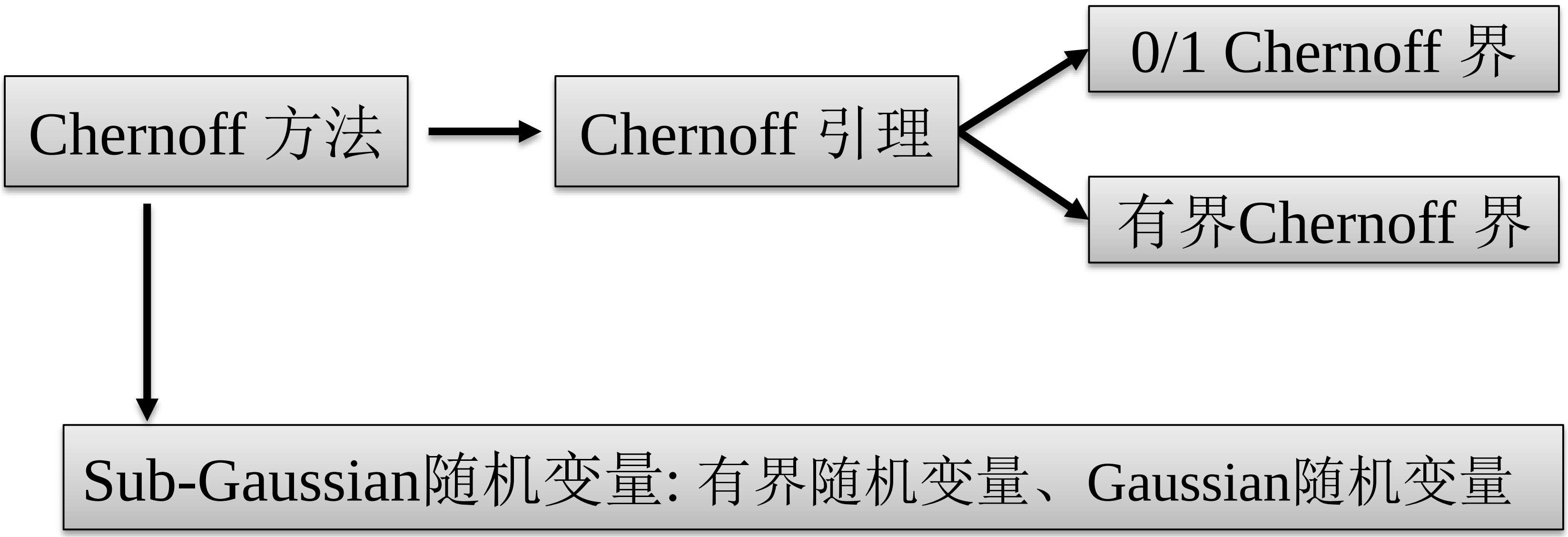
第六章 集中不等式

知识结构

基础不等式： Markov不等式、Chebyshev不等式、Hölder不等式、
Cauchy-Schwartz不等式、单边Chebyshev不等式



基础不等式

Markov不等式: $\boldsymbol{P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}}$

Chebyshev不等式: $\boldsymbol{P(|X-\mu| > \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}}$

单边Chebyshev不等式: $P(X-\mu \geq \epsilon) \leq \frac{\sigma^2}{\sigma^2+\epsilon^2}$

Hölder不等式: $E(|XY|) \leq \left(E(|X|^p)\right)^{\frac{1}{p}}\left(E(|Y|^q)\right)^{\frac{1}{q}}$

例题

随机变量X和Y满足$E(X) = 2$，$E(Y) = 2$，$\mathrm{Var}(X) = 1, \mathrm{Var}(Y) = 4, \rho_{XY} = -1/2$. 利用Chebyshev不等式估计$\mathrm{P}(|X - Y| \geq 6) \leq ???$

设随机变量$X \sim N(-1,2)$和$Y \sim E(1)$，且X和Y的相关系数为-1/2，利用Chebyshev不等式估计$P(|X + Y| \geq 6) \leq ???$

Chernoff方法

对任意$\epsilon > 0$和$t < 0$有

$$P[X \le -\epsilon] = P[tX \ge -t\epsilon] \le e^{t\epsilon} E[e^{tX}]$$

同理有

$$P[X \le -\epsilon] \le \min_{t<0} \{e^{t\epsilon} E[e^{tX}]\}.$$

称为Chernoff方法, 是证明集中不等式最重要的方法之一

有界的Chernoff不等式

Chernoff 引理: 随机变量$X \in [0,1]$的期望$\mu = E[X]$. 对$\forall t > 0$有

$$E[e^{tX}] \leq \mathrm{e}^{t\mu + \frac{t^2}{8}}$$

推论: $X \in [a, b]$期望$\mu = E[X]$, 对$\forall t > 0$有$E[e^{tX}] \leq \mathrm{e}^{t\mu + \frac{t^2(b-a)^2}{8}}$

Chernoff不等式：假设$X_1, \ldots, X_n$是n独立的随机变量、且满足$X_i \in [a, b]$. 对任意$\epsilon > 0$有

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i] \geq \epsilon\right] \leq e^{-2n\epsilon^2/(b-a)^2}$$

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i] \leq -\epsilon\right] \leq e^{-2n\epsilon^2/(b-a)^2}$$

例题

分类器f在集合$S_n = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)\}$ 的错误率为$\hat{p} > 0$，请估计分类器f在分布$\mathcal{D}$上的错误率，求f在分布$\mathcal{D}$上的错误率为$(9\hat{p}/10, 11\hat{p}/10)$ 之间的概率.

第七章 统计的基本概念

知识结构

统计学
基本概念

三大统计分布

五大抽样定理

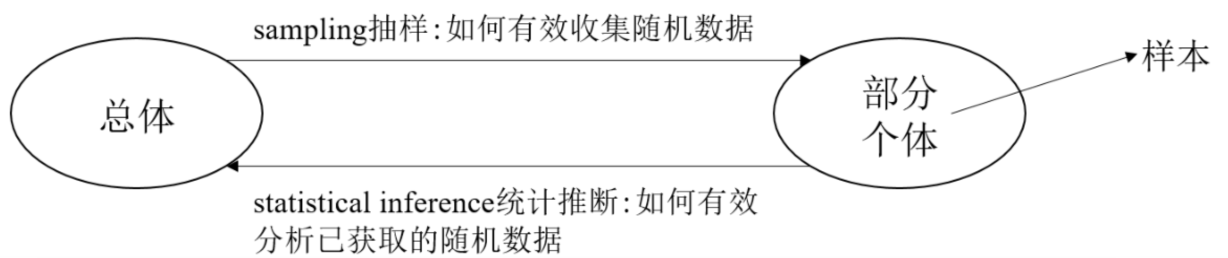
概念：分位点

统计学

统计学：以概率论为基础，研究如何有效收集研究对象的随机数据, 以及如何运用所获得的数据揭示统计规律的一门学科.

统计学的研究内容包括:

- **抽样**
- **参数估计**
- **假设检验**



统计学基本概念

总体与样本

统计量： $g(X_1,\cdots,X_n)$ 关于$X_1,\cdots,X_n$连续、且不含任意参数

常见统计量：

- **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

欧拉积分函数及其分布(**了解**)

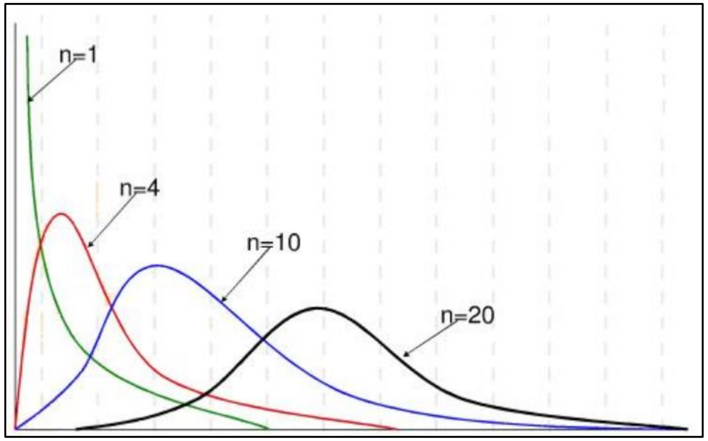
- Beta函数：Beta $(\alpha_1, \alpha_2) = \int_0^1 x^{\alpha_1 - 1}\,(1-x)^{\alpha_2 - 1} dx$
- Γ-函数为$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$
- 两类函数的关系

- Beta分布
- Γ-分布、性质、独立可加性
- 标准正态分布的平方Γ(1/2,1/2)
- Dirichlet分布、性质

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

根据$X_1^2 \sim \Gamma(1/2, 1/2)$和$\Gamma$分布的独立可加性可得$Y \sim \Gamma(n/2, 1/2)$



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

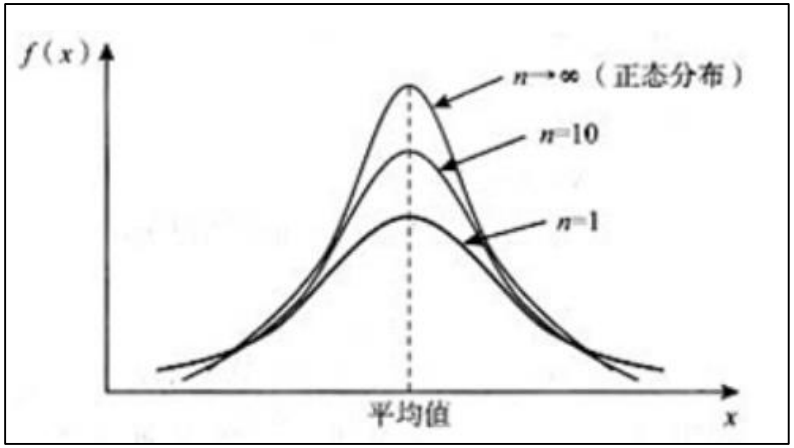
若随机变量$X \sim \chi^2(m)$和$Y \sim \chi^2(n)$相互独立，则$X + Y \sim \chi^2(m+n)$

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$T \sim t(n)$.



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

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

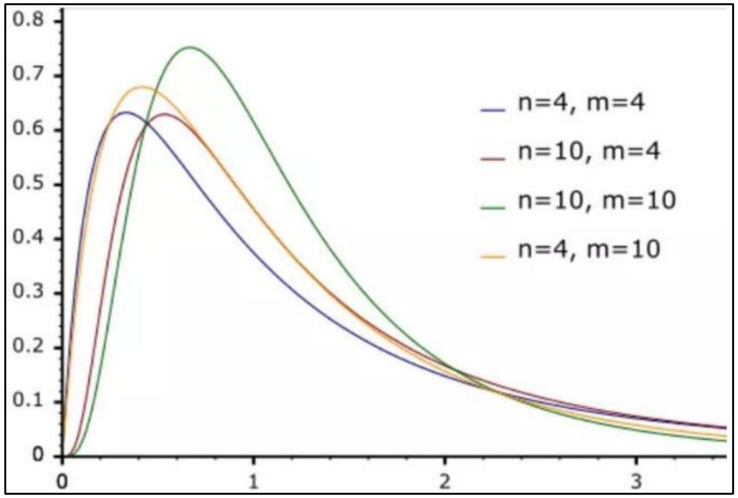
因此当n足够大时, $f(x)$ 被近似为$N(0,1)$ 的密度函数.

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$F \sim F(m, n)$.



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

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$\chi^2$分布，并求其自由度.

- 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

- $X_1, X_2, \cdots, X_9$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

- 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

例题

- 若随机变量$X \sim t(n)$, 求$Y = X^2$的分布.

- 设$X_1, X_2, \cdots, X_5$是来自总体$N(0,1)$的样本，令$Y = c_1(X_1 +$

抽样分布定理

定理一： 设$X_1,X_2,\cdots,X_n$是来自总体$N(\mu,\sigma^2)$的样本, 则有

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

定理二： 设$X_1,X_2,\cdots,X_n$是来自总体$N(\mu,\sigma^2)$的样本,其样本均值和修正样本方差分别为

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\qquad\qquad S^2=\frac{1}{n-1}\sum_{i=1\}}^{n}(X_i-\bar{X})^2$$

则**有$\bar{X}$和S^2相互独立**, 且

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

抽样分布定理

定理三： 设$X_1, X_2, \cdots, X_m$是来自总体$N(\mu, \sigma^2)$的样本,其样本均值和修正样本方差分别为

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad\qquad S^2 = \frac{1}{n-1}\sum_{i=1\}}^{n} (X_i - \bar{X})^2$$

则

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

定理四： 设$X_1, X_2, \cdots, X_m$和$Y_1, Y_2, \cdots, Y_n$分别来自总体$N(\mu_X, \sigma_X^2)$和$N(\mu_Y, \sigma_Y^2)$的两个独立样本, 其修正样本方差分别为S_X^2和S_Y^2, 则

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

正态分布的抽样分布定理五

定理五： 设$X_1,X_2,\cdots,X_m$和$Y_1,Y_2,\cdots,Y_n$分别来自总体$N(\mu_X,\sigma^2)$和$N(\mu_Y,\sigma^2)$的两个独立样本，令其样本均值分别$\bar{X}$和$\bar{Y}$，修正样本方差分别为S_X^2和S_Y^2, 则

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2).$$

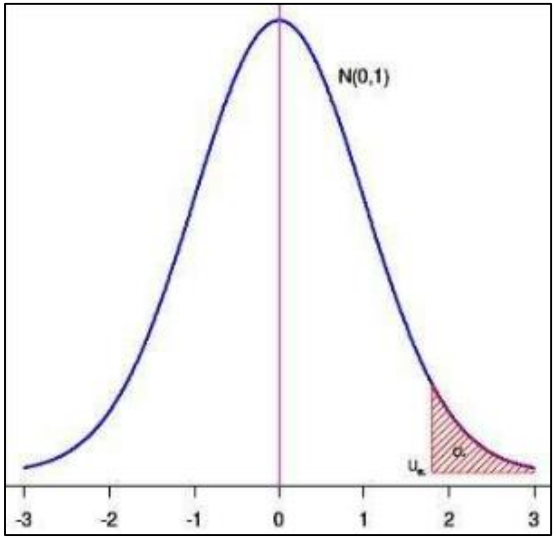
分位数

对正态分布$X \sim N(0,1)$, 给定$\alpha \in (0,1)$, 满足

$$P(X > \mu_\alpha) = \int_{\mu_\alpha}^{\infty} f(x)dx = \alpha$$

的点 $\boldsymbol{\mu_\alpha}$**称为正态分布上侧**$\boldsymbol{\alpha}$**分位点**

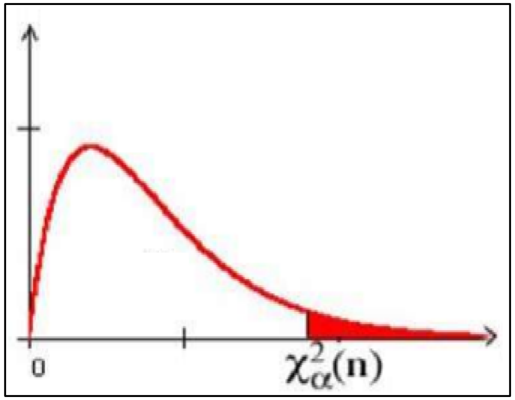
$\alpha = 1 - \Phi(\mu_\alpha)$， $\mu_{1-\alpha} = -\mu_\alpha$



对χ^2分布$X \sim \chi^2(n)$, 给定$\alpha \in (0,1)$, 满足

$$P\left(X \geq \chi_\alpha^2(n)\right) = \alpha$$

的点$\boldsymbol{\chi_\alpha^2(n)}$**称为**$\boldsymbol{\chi^2(n)}$**分布上侧**$\boldsymbol{\alpha}$**分位点**



t−分布和F−分布的分位数

对t-分布$X\sim t(n)$, 给定$\alpha\in(0,1)$, 满足$P\big(X>t_\alpha(n)\big)=\alpha$

的点$\boldsymbol{t_\alpha(n)}$**称为**$\boldsymbol{t(n)}$**-分布上侧**$\boldsymbol{\alpha}$**分位点**

由对称性可知$t_{1-\alpha}(n)=-t_\alpha(n)$

对F-分布$X\sim F(m,n)$, 给定$\alpha\in(0,1)$, 满足

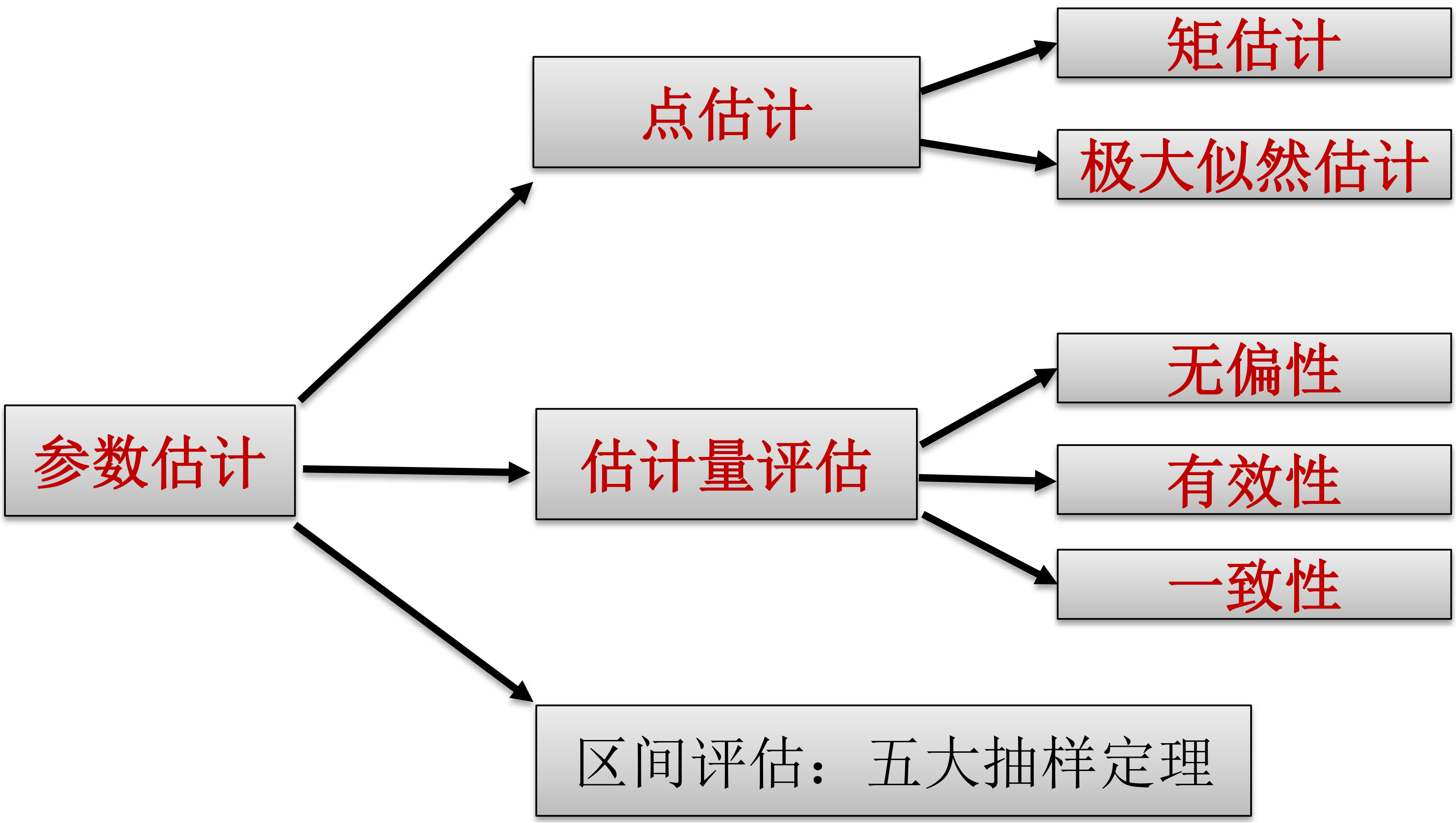
$$P[X>F_\alpha(m,n)]=\alpha$$

的点$\boldsymbol{F_\alpha(m,n)}$**称为**$\boldsymbol{F(m,n)}$ **分布上侧**$\boldsymbol{\alpha}$**分位点**

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

第八章 参数估计

知识结构



参数估计

设总体X的分布/密度函数为$F(X,\theta)$, 其中θ为未知参数

现从总体中抽取一样本 $X_1,X_2,\cdots,X_n$

问题：如何依据样本$X_1,X_2,\cdots,X_n$估计参数θ，或θ的函数$g(\theta)$,
此类问题称为 **参数估计问题**

- 研究内容包括: 点估计, 估计量标准, 区间估计
- 点估计包括：矩估计法、极大似然估计法

矩估计法

总体X的k阶矩: $a_k = E[X^k]$　　样本k阶矩: $A_k = \sum_{i=1}^{n} X_i^k$

用**样本矩去估计总体矩**求参数θ的方法称为 **矩估计法**

总体X的k阶中心矩: $b_k = E\left[\left(X - E(X)\right)^k\right]$

样本k阶中心矩: $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$

用**样本中心矩去估计总体中心矩**求参数θ的方法亦称为 **矩估计法**

基础：独立同分布变量$X_1, \cdots, X_n$, 若$E(X) = \mu$, 则$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$

矩估计方法

总体X的分布函数F包含m个未知参数$\theta_1, \theta_2, \cdots, \theta_m$

- 计算总体X的k阶矩：$a_k = a_k(\theta_1, \theta_2, \cdots, \theta_m) = E[X^k] \;\; k \in [m]$ (a_k一般为$\theta_1, \theta_2, \cdots, \theta_m$的函数)

- 计算样本的k阶矩：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

- 令样本矩等于总体矩：

$$A_k = a_k = a_k(\theta_1, \theta_2, \cdots, \theta_m) \; k \in [m]$$

得到m个关于$\theta_1, \theta_2, \cdots, \theta_m$的方程组

- 求解方程组得到估计量$\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$

最大似然估计（离散）

设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本.

若总体X为离散型随机变量，其分布列为$P(X=x)=P(X=x;\theta)$, 则样本$X_1, X_2, \cdots, X_n$的分布列为

$$L(\theta)=L(x_1, x_2, \cdots, x_n; \theta)=\prod_{i=1}^{n} P(x_i; \theta)$$

这里$L(\theta)$表示样本$X_1=x_1, \cdots, X_n=x_n$发生的概率

称**$\boldsymbol{L(\theta)}$为样本$\boldsymbol{X_1, X_2, \cdots, X_n}$的似然函数**

最大似然估计（连续）

若总体X为连续型随机变量，其概率密度为$f(x;\theta)$，则$X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$的联合概率密度为

$$L(\theta) = L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

根据概率密度定义可知$L(\theta)$ 越大，样本$X_1, X_2, \cdots, X_n$落入$x_1, x_2, \cdots, x_n$的邻域内概率越大

称$\boldsymbol{L(\theta)}$**为样本**$\boldsymbol{X_1, X_2, \cdots, X_n}$**的似然函数**

最大似然函数及不变性

综合离散和连续随机变量，统称$\boldsymbol{L(\theta)}$**为样本**$\boldsymbol{X_1, X_2, \cdots, X_n}$**的似然函数**，若

$$\hat{\theta} = \mathrm{argmax}_{\theta} L(x_1, x_2, \cdots, x_m; \theta)$$

则称 $\boldsymbol{\hat{\theta}}$**为**$\boldsymbol{\theta}$**的最大似然估计量**

最大似然估计量$\boldsymbol{\hat{\theta}}$**是使观测值**$\boldsymbol{X_1 = x_1, \cdots, X_n = x_n}$**出现概率最大**

定理： 设$\mu(\theta)$为θ的函数且存在反函数$\theta = \theta(\mu)$. 若$\hat{\theta}$是θ的最大似然估计，则$\hat{\mu} = \mu\left(\hat{\theta}\right)$是$\mu$的最大似然估计，

称为**最大似然估计的不变性**

求解与例题

求解步骤

- 计算对数似然函数$\ln\left(L(x_1,x_2,\cdots,x_m;\theta)\right)$
- 求对数似然函数中参数θ的一阶偏导, 令其等于零
- 求解方程组得到最大似然估计量$\hat{\theta}$

设总体X的概率密度函数

$$f(x)=\begin{cases}(\alpha+1)x^{\alpha} & x\in(0,1)\\ 0 & \text{其它}\end{cases}$$

设$X_1,X_2,\cdots,X_n$是来自总体X的样本, 求α的矩估计和极大似然估计

例题

设$X_1, X_2, \cdots, X_n$是来自总体X的样本, 以及总体X的密度函数为

$$f(x) = \begin{cases} \dfrac{1}{\theta} e^{-(x-\mu)/\theta} & x \geq \mu \\ 0 & \text{其它} \end{cases}$$

其中$\theta > 0$, 求μ和θ的极大似然估计

设$X_1, \cdots, X_n$是取自总体$X \sim B(1, p)$的样本, 求p的最大似然估计

估计量的评价标准

不同的估计方法可能得到不同的估计值，哪一种估计量更好，或更好的标准是什么呢?

无偏性、有效性、一致性

设$X_1, X_2, \cdots, X_n$来自总体X的样本，令$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的一个估计量, 若

$$E_{X_1,X_2,\cdots,X_n}\left[\hat{\theta}\right] = E_{X_1,X_2,\cdots,X_n}\left[\hat{\theta}(X_1, X_2, \cdots, X_n)\right] = \theta$$

则称**$\hat{\theta}$为θ的无偏估计**

无偏估计要求在期望的情形下有$E(\hat{\theta}) = \theta$成立,无系统性偏差

无偏估计不具有函数不变性: $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的无偏估计, 但并不一定有$g(\hat{\theta})$是$g(\theta)$的无偏估计

有效性

设$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和$\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$是$\theta$的两个无偏估计, 若

$$\mathrm{Var}\left(\hat{\theta}_1\right) \leq \mathrm{Var}\left(\hat{\theta}_2\right)$$

则称$\boldsymbol{\theta_1}$**比**$\boldsymbol{\theta_2}$**有效**

有效统计量（了解）

随机变量X的概率密度为$f(x;\theta)$或分布函数为$F(x;\theta)$, 令

$$\mathrm{Var}_0(\theta)=\left[nE\left(\frac{\partial\ln f(X;\theta)}{\partial\theta}\right)^2\right]^{-1}$$

$$\mathrm{Var}_0(\theta)=\left[nE\left(\frac{\partial\ln F(X;\theta)}{\partial\theta}\right)^2\right]^{-1}$$

对任意的无偏估计量$\hat{\theta}$有$\mathrm{Var}\left(\hat{\theta}\right)\geq\mathrm{Var}_0(\theta)$,称$Var_0(\theta)$为估计量$\hat{\theta}$方差的下界.

当$Var\left(\hat{\theta}\right)=Var_0(\theta)$时称$\hat{\theta}$为达到方差下界的无偏估计量，此时$\hat{\theta}$为最有效估计量, 简称**有效估计量**

一致性

设$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$是$\theta$的一个估计量，若当$n \to \infty$时有$\hat{\theta}_n \xrightarrow{P} \theta$成立，即对任意$\epsilon > 0$有$\lim\limits_{n\to\infty} P[|\hat{\theta}_n - \theta| > \epsilon] = 0$，则称$\hat{\theta}_n$为$\theta$的一致估计量

定理： 设$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$是$\theta$的一个估计量, 若满足:

$$\lim_{n\to\infty} E\left[\hat{\theta}_n\right] = \theta \qquad \lim_{n\to\infty} \mathrm{Var}\left(\hat{\theta}_n\right) = 0$$

则$\hat{\theta}_n$为θ的一致估计量.

综合例题

设$X_1, \cdots, X_n$是取自总体X的一个样本，且总体X的密度函数为

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta} e^{-\frac{x}{\theta}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

1) 求θ的极大似然函数估计$\hat{\theta}$

2) 设$Z = n\min\{X_1, X_2, \dots, X_n\}$，求证$Z$和$\hat{\theta}$均为$\theta$的无偏估计量

3) 请问$\hat{\theta}$和Z哪个估计量更有效

4) $\hat{\theta}$是否为一致统计量

第九章 假设检验

知识结构（概念+原理）

假设检验基本思想: 小概率原理(小概率事件在一次实验中不应该发生)

假设检验概念

两类错误

显著性检验

显著性检验基本步骤

假设检验

根据样本信息来检验关于总体的某个假设是否正确，此类问题称为 **假设检验问题**, 可分为参数检验问题和非参数检验问题

假设检验方法(反证)： 先假设所做的假设H_0成立，然后从总体中取样, 根据样本来判断是否有 **不合理** 的现象出现, 最后做出接受或者拒绝所做假设的决定.

不合理的现象: 小概率事件在一次事件中几乎不会发生

假设检验的两类错误

- 第一类错误：拒绝实际真的假设H_0（弃真）
- 第二类错误：接收实际不真的假设H_0（存伪）

小事件

假设检验中需要对 不合理 的小事件给出一个定性描述，通常给出一上界α, 当一事件发生的概率小于α时则成为小概率事件

通常取α =0.05,0.1,0.01，其具体取值根据实际问题而定．在假定H_0成立下，根据样本提供的信息判断出不合理的现象 (概率小于α的事件发生), 则认为假设H_0不显著, α被称为显著水平.

注意：不否定假设H_0并不是肯定假设H_0一定成立，只能说差异不够显著, 没达到否定的程度, 所以假设检验被称为``**显著性检验**“

假设检验的一般步骤

- 根据实际问题提出原假设H_0和备择假设H_0
- 确定检验统计量 (分布已知)
- 确定显著性水平α, 并给出拒绝域
- 由样本计算统计量的实测值, 判断是否接受原假设H_0

问题：正态总体情形下的期望与方差检验 + 五大抽样定理