第一章 注意事项(丢分或掉坑的地方)

1、互斥与独立事件的判断

1) 称 A、B 为互不相容, 若 $AB=\varnothing$. 即 A、B 不可能同时发生.

2) 称 A、B 相互独立,若 $P(AB) = P(A)P(B)$

或 $P(A|B) = P(A)$

注意: 随机变量的独立的判断　　设 (X, Y) 是二维随机变量

(都要满足 2^n-n-1 等式) $P\{X\le x, Y\le y\} = P\{X\le x\}P\{Y\le y\}$ $\forall (x, y)\in R^2$

3、全概率及贝叶斯公式的应用：教材例1.3.10

步骤：　(1)所求的事件(结果)记为 A(文字描述)

(2)样本空间有限划分(或完备事件组,导致的原因)记为 B_i(文字描述)

(3)由全概率公式：$P(A)=\sum_{i=1}^{n} P(B_i)P(A|B_i)$

或由贝叶斯公式：$P(B_i|A)=\dfrac{P(B_i)P(A|B_i)}{P(A)}$

第 一 章的核心(结合随机变量表达事件来复习)

独立事件与互斥(或互不相容)事件

1. 称A、B为互不相容, 若$AB=\varnothing$,即A、B不可能同时发生.

2. 称A、B 相互独立， 若$P(AB)=P(A)P(B)$

或 $P(A|B)=P(A)$

即事件A发生的可能性大小不受事件B的影响.

注1: 事件的独立性是否存在传递性?即事件A与事件B相互独立, 事件B与事件C相互独立, 不能推知事件A与事件C相互独立.

注2: 事件A与事件B相互独立(或互斥), 事件C是事件B的子集, 能否推知事件A与事件C相互独立(或互斥)？(独立对子集不成立封闭,互斥对子集成立)

注3: 独立与互斥的关系?任意两个事件$P(A)>0$、$P(B)>0$, 它们相互独立和互不相容不能同时成立。

注4:第三章通过随机变量的分布函数, 分布律, 密度函数判断随机变量独立的形式.

1. 事件的独立性是否存在传递性?即事件A与事件B相互独立, 事件B是事件C相互独立, 能否推知事件A与事件C相互独立？试举例说明.

解答　事件的独立性不存在传递性.　　　　(3 分)

反例　独立地抛掷出一枚硬币和一个骰子，令三个事件如下

$A=\{出现正面\}$，$B=\{掷出第6点\}$，$C=\{出现反面\}$　　　　(6 分)

则事件 A 与事件 B 相互独立，事件 B 与事件 C 相互独立，但事件 A 与事件 C 不相互独立.　　　　(8 分)

2.事件A与事件B相互独立, 事件C是事件B的子集, 能否推知事件A与事件C相互独立?

不能。设随机实验：抛一枚均匀硬币两次。A={第一次正面向上}，B={第二次正面向上}，C={第一次第二次均正面向上}。则事件A与事件B相互独立，事件C是事件B的子集。A与C不相互独立。

3. 事件A与事件B互不相容, 事件C是事件B的子集, 能否推知事件A与事件C互不相容?

能。作韦氏图。

第二和三章 注意事项(丢分或掉坑的地方)

1、(一维和二维)分布函数的性质及分段或分片表达

（1） $0\le F(x)\le 1$, 且 $\lim\limits_{x\to-\infty}F(x)=0$, $\lim\limits_{x\to+\infty}F(x)=1$

$$\lim_{x\to-\infty}F(x,y)=0 \quad \lim_{y\to-\infty}F(x,y)=0 \quad \lim_{\substack{x\to+\infty\\ y\to+\infty}}F(x,y)=1$$

$$P\{x_1<X\le x_2, y_1<Y\le y_2\}$$
$$=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)$$

例2

$$\cancel{F_Y(y)=\begin{cases}-, & y_1\le y\le y_2\\ 0, & else\end{cases}}$$

$$F_Y(y)=\begin{cases}0, & y<y_1\\ - & y_1\le y\le y_2\\ 1 & y>y_2\end{cases}$$

2、随机变量函数的密度函数：教材例3.4.4和例3.4.5

基本的分布函数法:

$$F_Y(y)=P\{Y\le y\}=P\{|X|\le y\}=\begin{cases}P\{-y\le X\le y\}, & y\ge 0\\ 0, & else\end{cases}$$

套用（必写） 大写 小写

分段

$X \sim U(0, y), i.e.,$

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{y}, & y > x > 0 \\ 0, & ellse \end{cases}$$

3.概率密度函数与条件概率密度函数

$$f(x,y) = \begin{cases} e^{-y}, & x > 0, y > x \\ 0, & else \end{cases} \quad f_Y(y) = \begin{cases} ye^{-y}, & y > 0 \\ 0, & else \end{cases}$$

当$y_1 \le y \le y_2$时，$\boldsymbol{f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{y}, & y > x > 0 \\ 0, & \text{其他} \end{cases}}$ 满足非负、归一性

例4 例5 例6

4.分布函数与条件分布函数：教材例3.3.3

满足三条性质

变量连续：$F_X(x) = P\{X \le x\} = \int_{-\infty}^{x} f_X(x)dx$

$$F_{X|Y}(x|y) = P\{X \le x | Y = y\} \ne \frac{P\{X \le x, Y = y\}}{P\{Y = y\}}$$

$$= \int_{-\infty}^{x} f_{X|Y}(u|y)du$$

例3书32题

为0

$$P\{X \le x | Y > y\} = \frac{P\{X \le x, Y > y\}}{P\{Y > y\}}$$

不为0

第二和三章 注意事项(丢分或掉坑的地方)

5. $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$, $\int_{-\infty}^{+\infty} e^{-y^2} dy = \sqrt{\pi}$, $\int_{-\infty}^{y} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}\Phi(y)$

6. 对于伯努里试验，判断试验独立和重复,考察如下问题:

(1) n次试验中事件A 首次发生时的试验次数Z;

$$\{Z = k\} = \{\overline{A_1}\,\overline{A_2}\cdots\overline{A_{k-1}}A_k\}, k = 1,\cdots,n-1$$

$$\{Z = n\} = \{\overline{A_1}\,\overline{A_2}\cdots\overline{A_{n-1}}A_n\} \cup \{\overline{A_1}\,\overline{A_2}\cdots\overline{A_{n-1}}\,\overline{A_n}\}$$

注：推广$n \to +\infty$，$P\{Z = k\} = p(1-p)^{k-1}$为几何分布.

(2) 事件A 发生k 次时的试验次数Y; 负二项分布(或

$P\{Y = t\} = C_{t-1}^{k-1} p^k (1-p)^{t-k}, t = k, k+1, \cdots,$ 帕斯卡(Pascal)分布)

(3) n 次试验中事件A 发生的总次数X.

($B(n,p)$再生性和可加性)

第二和三章 注意事项(丢分或掉坑的地方)

7. 和的概率密度

(类似边缘密度的思想：均是连续型随机变量的和)

$$f_Z(z)=\int_{-\infty}^{+\infty} f(z-y,y)\,dy \qquad f_Z(z)=\int_{-\infty}^{+\infty} f(x,z-x)\,dx$$

$Z = X/Y$ 即商的分布(类似边缘密度的思想)

$$f_Z(z)=\int_{-\infty}^{+\infty}|y|\,f(zy,y)\,dy$$

8. 离散与连续随机变量混合求和的分布(不能直接套公式)

23. 设 P{X=0}=P{X=1}=1/2, Y~U(0,1)且 X,Y 相互独立，求 X+Y 的概率分布.

注意：随机变量函数的分布与第四章结合起来

二维正态分布

正态分布的重要性质

$$(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$$

1.边缘分布一定是正态分布：$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$

任意线性组合$aX+bY$仍是正态分布：

1) X与Y独立：$aX+bY\sim N(a\mu_1+b\mu_2,a^2\sigma_1^2+b^2\sigma_2^2)$

2) X与Y不独立：

$$aX+bY\sim N(a\mu_1+b\mu_2,a^2\sigma_1^2+b^2\sigma_2^2+2ab\sigma_1\sigma_2\rho)$$

2. X与Y独立$\Leftrightarrow\rho=0$

$(X,Y)\sim N(0,1;0,1;0)$　　$f(x,y)=\dfrac{1}{2\pi}e^{-\frac{x^2+y^2}{2}},x,y\in R$

设随机变量 X 和 Y 均服从标准正态分布，二者的相关系数为 0，则下列说法中正确的个数有

1）X 与 Y 一定独立；（2）$D(X+3Y)=10$；（3）$3X\sim N(0,3)$；（4）$\operatorname{cov}(2X,3Y)=0$.

多维随机变量的独立性 ($X_1, X_2, \ldots X_n$ 相互独立)

定义: 设 n 维随机变量$(X_1, X_2, \ldots, X_n)$的联合分布函数

为 $F(x_1, x_2, \ldots, x_n)$, 若对任意实数$x_1, x_2, \ldots, x_n$

均有 $$F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i) \qquad (*)$$

定理: 若n 维随机变量$(X_1, X_2, \ldots, X_n)$ 相互独立, 则

1) 任意k个随机变量($2 \le k \le n$) 也相互独立.

2) 随机变量 $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$ 也 相互独立.

3) m维 随机向量$(X_1, X_2, \ldots, X_m)$ 与$n-m$维 随机向量$(X_{m+1}, X_{m+2}, \ldots, X_n)$ 也 相互独立.

4) 随机变量 $h(X_1, X_2, \ldots, X_m)$ 与 $g(X_{m+1}, X_{m+2}, \ldots, X_n)$ 也 相互独立.

随机变量X与Y相互独立的另一等价条件？

X与Y相互独立$\Leftrightarrow F(x,y)=F_X(x)F_Y(y)$

$\Leftrightarrow P\{X\le x \mid Y=y\}=F_{X|Y}(x|y)=F(x)$ 对所有$(x,y)\in R^2$成立.

两个离散型r.v.X,Y相互独立

1) $P_{ij}=P_{i.}P_{.j}$ 对所有（x_i,y_j）成立.

2) $P\{X=x_i|Y=y_j\}=P\{X=x_i\}$

3) $P\{Y=y_j\mid X=x_i\}=P\{Y=y_j\}$

两个连续型r.v.X,Y相互独立

1) $f(x,y)=f_X(x)f_Y(y)$ 2) $f_X(x)=f_{X|Y}(x|y)$

3) $f_Y(y)=f_{Y|X}(y|x)$ 在平面上除去“面积”为0 的集合外成立.

判断随机变量X与Y不相互

X与Y不相互独立 $\Leftrightarrow F(x,y)\neq F_X(x)F_Y(y)$

$\Leftrightarrow P\{X\leq x|Y=y\}\neq F_{X|Y}(x|y)=F(x)$ 对某些$(x,y)\in R^2$成立.

两个离散型r.v.X,Y 不相互独立

1)$P_{ij}\neq P_{i.}\cdot P_{.j}$ 对某些(x_i,y_j)成立.

2)$P\{X=x_i|Y=y_j\}\neq P\{X=x_i\}$

3)$P\{Y=y_j\,|\,X=x_i\}\neq P\{Y=y_j\}$

两个连续型r.v.X,Y 不相互独立

1) $f(x,y)\neq f_X(x)f_Y(y)$; 2) $f_X(x)\neq f_{X|Y}(x|y)$;

3) $f_Y(y)\neq f_{Y|X}(y|x)$ 在平面上“面积”不为0 的集合上成立(区域).

第 四章 随机变量的数字特征(期望, 方差, 协方差, 相关系数)

确定的实数

1.期望,或均值(取值的集中点或中常状态)

离散

$$E(X)=\sum_{i=1}^{+\infty} x_i P\{X=x_i\}$$

$$E[g(X,Y)]=\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} g(x_i,y_j)p_{ij}$$

$$E[g(X)]=\sum_{i=1}^{+\infty} g(x_i)p_i$$

连续

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx$$

$$E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dydx$$

$$E[g(X)]=\int_{-\infty}^{+\infty} g(x)f(x)dx$$

习题1

习题5

2. 方差:

定义: $D(X)=E\{[X-E(X)]^2\}(\geq 0)$为$X$的方差.$\sqrt{D(X)}$为标准差或均方差.

计算公式：$D(X)=E(X^2)-E(X)^2$ 重要结论：$E(X^2)=D(X)+E(X)^2$

$D(X)\leq E\left[(X-x)^2\right], x\in R$,等式在$x=E(X)$时成立.

1. 数学期望的性质:

1) 线性法则: $E(cX+b)=cE(X)+b$

当b=0时, $E(cX)=cE(X)$ 当c=0时, $E(b)=b$

2) 加法法则: $E(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}E(X_i)$

3) 若$X_1, X_2, \ldots, X_n$ 相互独立, 则 $E\left(\prod_{i=1}^{n}X_i\right)=\prod_{i=1}^{n}E(X_i)$

2.方差的性质:

注: $E(E(X))=E(X)$; $D(E(X))=0.$

1) $D(cX+b)=c^2D(X)$

特别地, $D(b)=0$ $D(-X)=D(X)$

2)若$X_1, X_2, \ldots, X_n$ 相互独立 $D(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}D(X_i)$

3) $D(X)=0$ $\Longleftrightarrow$ $P\{X=E(X)\}=1$

(方差为0的随机变量必为常数)

4) $D(X\pm Y)=D(X)+D(Y)\pm 2E\{[X-E(X)][Y-E(Y)]\}$

常用分布的期望和方差表（记住）

两个随机变量间关系的数字特征: 协方差和相关系数

随机变量X,Y的协方差: $cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$

常用的计算公式：$cov(X,Y)=E(XY)-E(X)E(Y)$

注: 协方差为$[X-E(X)][Y-E(Y)]$的均值, 依赖于X,Y的度量单位, 选择适当单位使X,Y的方差为1, 则协方差就是相关系数.

相关系数:更好地反映X,Y之间的关系,不受所用单位的影响.

随机变量X与Y的相关系数: 设$D(X)>0,D(Y)>0$

$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=E\left[\frac{X-E(X)}{\sqrt{D(X)}}\cdot\frac{Y-E(Y)}{\sqrt{D(Y)}}\right]=\operatorname{cov}\left(X^*,Y^*\right)$$

相关系数又叫“标准尺度下的协方差”，是一无量纲的量.

3、随机变量之间的关系：

协方差：协方差性质：3条

$cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}=E(XY)-E(X)E(Y)$

相关系数：$\rho_{XY}=\dfrac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=cov(X^*,Y^*)$

相关系数性质: 3条(与第九章的样本相关系数对应)

4. 矩(与第六章的总体矩对应)

$\gamma_k=E(X^k)$, $k=1,2,3$…..为X的 k 阶原点矩.

$\mu_k=E\{[X-E(X)]^k\}$, $k=1,2,3$…..为X的k阶中心矩.

$\alpha_k=E(|X|^k)$, $k=1,2,3$…..为X的 k 阶绝对原点矩.

$\beta_k=E[|X-E(X)|^k]$, $k=1,2,3$…..为X 的k阶绝对中心矩.

5. 实际概率意义

数学期望—随机变量的平均值；

方差—刻划随机变量 X 围绕它的数学期望的偏离程度的数字特征.

相关系数—衡量两个随机变量之间线性相关程度的数字特征.

6. 两个随机变量的相关性概念

$\rho_{XY}=0$, 称X与Y不相关

二者无线性关系

7. 不相关与相互独立概念间的关系

1) 随机变量X 与Y 相互独立 $\Longrightarrow$ X 与Y 不相关

一般逆不真.

2）$(X,Y)\sim N(\mu_1,\sigma^2_1;\ \mu_2,\sigma^2_2;\ \rho)$ 则

习题21

X, Y相互独立 ⟺ $\rho=0$（不相关）

二、数字特征计算

1. 利用数字特征的性质；

$$X\sim N(0,1)\Rightarrow E(X^2)=D(X)+[E(X)]^2=1$$

$$E(X^4)=\int_{-\infty}^{+\infty}x^4\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=-\frac{1}{\sqrt{2\pi}}x^3e^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}3x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

$=3E(X^2)=3$(教材例4.4.7的重要结论)

注：$X\sim N(0,\sigma^2)$,则$E\left(X^n\right)=\begin{cases}0 & n=1,3,5,\cdots\\ \sigma^n(n-1)(n-3)\cdots1 & n=2,4,6,\cdots\end{cases}$

$$\Rightarrow D(X^2)=E(X^4)-[E(X^2)]^2=2\Rightarrow X^2\sim\chi^2(1)$$

2. 利用特殊分布的可加性(独立条件).

记住：正态、二项、泊松、均匀、指数的数字特征.

非退化n维正态随机变量$(X_1,X_2,\ldots,X_n)$的性质：

1) 相互独立的(一维)正态随机变量的有限线性组合仍服从一维正态分布.(正态分布具有可加性)

2) 随机变量$(X_1,X_2,\ldots,X_n)$服从n维正态分布的充分必要条件是 $X_1,X_2,\ldots,X_n$的任意非零线性组合,即

$$l_1X_1+l_2X_2+\ldots l_nX_n\ (l_1, l_2,\ldots, l_n\text{不全为}0)$$

服从一维的正态分布.

3) n维随机变量$(X_1,X_2,\ldots,X_n)$服从n维正态分布,设$Y_1,Y_2,\ldots,Y_m$是$X_1,X_2,\ldots,X_n$的非零线性组合,则$(Y_1,Y_2,\ldots,Y_m)$是m维正态随机变量.(线性变换不变性)

例: (X_1,X_2,X_3)是三维正态随机变量, 则

$X_1+X_2-X_3$, X_1-X_2服从正态分布.

(X_1+X_2, X_1-X_2)是二维正态随机变量.

n维正态随机变量$(X_1,X_2,\dots X_n)$的性质

1) 相互独立的一维正态随机变量的有限线性函数仍服从一维正态分布. (正态分布具有可加性)

2) n维随机变量$(X_1,X_2,\dots X_n)$服从n维正态分布的充分必要条件是$X_1,X_2,\dots X_n$的任意非零线性组合

首先构成

n维随机变量

$l_1X_1+l_2X_2+\dots l_nX_n$ ($l_1, l_2,\dots, l_n$不全为0) 服从一维的正态分布.

3) n维随机变量$(X_1,X_2,\dots X_n)$服从n维正态分布,

设$Y_1,Y_2,\dots,Y_m$是$X_1,X_2,\dots X_n$的非零线性组合,

则$(Y_1,Y_2,\dots,Y_m)$是m维正态随机变量.(线性变换不变性)

4) $X_1,X_2,\dots X_n$ 相互独立 $\Longleftrightarrow$ $\rho_{ij}=0$ $(i\neq j)$

一维正态随机变量X的性质：

相互独立的正态随机变量的线性组合仍服从正态分布.(可加性)

二维正态随机变量(X_1,X_2)的性质： (首先要构成二维)

1) $\Longleftrightarrow$ $l_1X_1+l_2X_2$ (l_1, l_2不全为0)是正态随机变量.

$\Longrightarrow$ X_1，X_2 均是是一维的正态随机变量.

2)X_1,X_2相互独立 $\Longleftrightarrow \rho_{12}=0$

3)设Y_1,Y_2是X_1,X_2的非零线性组合，则(Y_1,Y_2)是二维正态随机变量.

思考：若Z_1，Z_2均是一维的正态随机变量

~~$\Longrightarrow$~~ $l_1Z_1+l_2Z_2$ (l_1, l_2不全为0) 是正态随机变量.

若X,Y都服从一维正态分布，则X与Y相互独立不等价于X与Y不相关。(X与Y均为方差非0的随机变量)

设随机变量 X 和 Y 都服从正态分布，则

(A) $X+Y$ 一定服从正态分布； (B) (X,Y) 不一定服从二维正态分布；

(C) X 与 Y 相互独立等价于不相关； (D) 若 X 和 Y 相互独立，则函数 $f(X+Y)$ 服从正态分布

第三章习题册最后一道：

24. 随机变量(X,Y)的联合概率密度函数是：

$$f(x,y)=\frac{e^{-\frac{x^2+y^2}{2}}}{2\pi}+\frac{e^{-\pi^2}}{2\pi}g(x)g(y),(x,y)\in R^2$$

其中 $g(x)=\begin{cases}\cos x,|x|\le\pi\\0,|x|>\pi\end{cases}$

1.证明X和Y都服从正态分布；2.求随机变量Y关于X的条件概率密度；3.讨论X与Y是否相互独立？4.根据本题的结果，你能总结出什么结论？

$$1.f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-\infty}^{+\infty}\frac{e^{-\frac{x^2+y^2}{2}}}{2\pi}dy+\int_{-\infty}^{+\infty}\frac{e^{-x^2}}{2\pi}g(x)g(y)dy$$

$$=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\ (x\in R)$$

$$\Rightarrow X\sim N(0,1)$$

同理 $Y\sim N(0,1)$

$$2.f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}+\frac{e^{\frac{x^2}{2}-\pi^2}}{\sqrt{2\pi}}g(x)g(y)$$

$$3.f_X(x)f_Y(y)=\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}*\frac{e^{-\frac{x^2+y^2}{2}}}{\sqrt{2\pi}}\ne f(x,y)$$

4.两个一维正态分布的联合发布不一定是正态分布。

第 五 章 大数定律和中心极限定理

一、概念

1. 依概率收敛 ⟹ 定义大数定律；

2. 依分布收敛 ⟹ 定义中心极限定理；

3. 切比雪夫不等式

期望和方差存在

- 对概率做粗略估计；

  习题1、2

- 用来验证估计量的相合性(按概率收敛).(优良估计量准则)

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}$$

(括号内外不等式方向相反)

或者 $P\{|X-E(X)|<\varepsilon\}\geq 1-\dfrac{D(X)}{\varepsilon^2}$,

二、大数定律 (“频率收敛于概率”引申而来)

1. 概率意义

随机变量序列$\{X_k\}$，$k=1,2\ldots$的前n项算术平均将紧密地聚集在其数学期望的附近。

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)|<\varepsilon\}=1$$

即$\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\xrightarrow{P}0$

2. 掌握

- 切比雪夫大数定律；
- 独立同分布大数定律；
- 贝努里(Bernulli)大数大数定律；
- 辛钦大数大数定律.

3. 作用：1)矩估计；2)假设检验:小概率实际推断原理.

三、中心极限定理（重点）

(和的极限分布就是正态分布)即“和的分布收敛于正态分布”

1. 概率意义

随机变量序列的前n项和$\sum_{k=1}^{n} X_k$的标准化随机变量的极限分布为标准正态分布.

2. 掌握
- 独立同分布中心极限定理 　习题7、8
- 棣莫佛—拉普拉斯中心极限定理(正态分布逼近二项分布)
- 与泊松逼近定理区别 　习题4—6,10

3. 作用
- 概率近似计算;
- 确定大样本估计量(枢轴变量和检验统计量)的分布.(第七和八章)教材例7.3.7

第四章 注意事项(丢分或掉坑的地方)

注1： 要防止计算中常见的两个错误!

$$D(cX)=cD(X) \qquad D(X\pm Y)=D(X)\pm D(Y)$$

改为：$D(cX)=c^2D(X)$

$$D(X\pm Y)=D(X)+D(Y)\pm 2E\{[X-E(X)][Y-E(Y)]\}$$

注2： $D\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}D(X_i)$

改为：此式只有在诸随机变量相互独立的条件下才成立.

3) $\because \operatorname{cov}(X,Y)=0 \therefore \rho_{XY}=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=0$ 需要：$\sqrt{D(X)}\sqrt{D(Y)}>0$

设随机变量$X_1, X_2, \ldots, X_n$独立同分布, 且$E(X_i)=\mu$

$D(X_i)=\sigma^2$, 则$D(\bar{X})=D[\dfrac{1}{n}(X_1+X_2+\ldots+X_n)]=\dfrac{1}{n}\sigma^2$

$D(\bar{X})=\dfrac{1}{n}\sigma^2$

(第六 – 九章常用)

第五章 注意事项(丢分或掉坑的地方)

1. 切比雪夫不等式 (括号内外不等式方向相反)

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2} \text{ 或者 } P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2},$$

2.中心极限定理

注：(1)中心极限定理中 "≈" ⟶ (1~3分)

例：$P\{x_1<\sum_{i=1}^{n}x_i\leq x_2\}=P\{\frac{x_1-n\mu}{\sigma\sqrt{n}}<Z_n\leq\frac{x_2-n\mu}{\sigma\sqrt{n}}\}$

例3

$$\approx \Phi(\frac{x_2-n\mu}{\sigma\sqrt{n}})-\Phi(\frac{x_1-n\mu}{\sigma\sqrt{n}})$$

设随机变量 X 的均值 $\mu=3$，标准差 $\sigma=1/2$，用切比雪夫不等式估计概率 $P\{X\leq 1.5$或$X\geq 4.5\}$：(　　).

(A) $\geq\frac{8}{9}$;　(B) $\geq\frac{3}{4}$;　(C) $\leq\frac{1}{4}$;　(D) $\leq\frac{1}{9}$.

例2: 将一枚均匀硬币连续抛 n 次, 试用中心极定理来估计n, 使下式成立. $P\{|f_n(A)-P(A)|<0.01\}\geq 0.99$

其中 $A=\{$ 出现正面 $\}$.

解: $P(A)=1/2$, 令 $X_i=\begin{cases}1, & 第i次出现正面; \\ 0, & 其它.\end{cases}$ $i=1,2,\cdots n$

则随机变量序列$\{X_i\}, i=1,2,\ldots$是相互独立且同分布的, 且有

$$f_n(A)=\frac{1}{n}\sum_{i=1}^{n}X_i$$

$$E(X_i)=\frac{1}{2},\quad D(X_i)=\frac{1}{4},\quad i=1,2,\cdots$$

所以随机变量序列$\{X_i\}$, 满足独立同分布中心极限定理.

$$f_n(A)=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad E(\sum_{i=1}^{n}X_i)=\frac{n}{2} \qquad D(\sum_{i=1}^{n}X_i)=\frac{1}{4}n$$

解： $P\{|f_n(A)-P(A)|<0.01\}$

$$=P\{-0.01<\frac{\sum_{i=1}^{n}X_i}{n}-\frac{1}{2}<0.01\}=P\{-0.01n+\frac{n}{2}<\sum_{i=1}^{n}X_i<\frac{n}{2}+0.01n\}$$

$$=P\{-\frac{0.01n}{1/2\sqrt{n}}<\underline{\frac{\sum_{i=1}^{n}X_i-\frac{n}{2}}{1/2\sqrt{n}}}<\frac{0.01n}{1/2\sqrt{n}}\}$$

$\dfrac{\sum_{i=1}^{n}X_i-\frac{n}{2}}{1/2\sqrt{n}}$ 可由 $N(0,1)$ 来近似.

设 X 为独立重复抛掷一枚均匀硬币 100 次的试验中正面出现的次数，用中心极限定理计算正面出现的次数 X 与平均次数 50 的误差不超过 10 次的概率（最终结果精确到小数点后两位）为（　　）.

已知：$\Phi(1)=0.8413, \Phi(2)=0.9772$.

（A）0.90;　　（B）0.85;　　（C）0.99;　　（D）0.95.

随机变量序列$\{X_i\}$, 满足独立同分布中心极限定理:

$$0.99 \le P\{-\frac{0.01n}{1/2\sqrt{n}} < \frac{\sum_{i=1}^{n} X_i - \frac{n}{2}}{1/2\sqrt{n}} < \frac{0.01n}{1/2\sqrt{n}}\} \approx 2\Phi(0.02\sqrt{n}) - 1$$

$$\Rightarrow \Phi(0.02\sqrt{n}) \ge 0.995$$

$Y \sim N(0,1)$:

$P\{-a < Y < a\} = 2\Phi(a) - 1$

因$\Phi(x)$单调不降

$$\Rightarrow 0.02\sqrt{n} \ge 2.58$$

解得 $n \ge 16,641$ (次).

设 X 为独立重复抛掷一枚均匀硬币 100 次的试验中正面出现的次数，用中心极限定理计算正面出现的次数 X

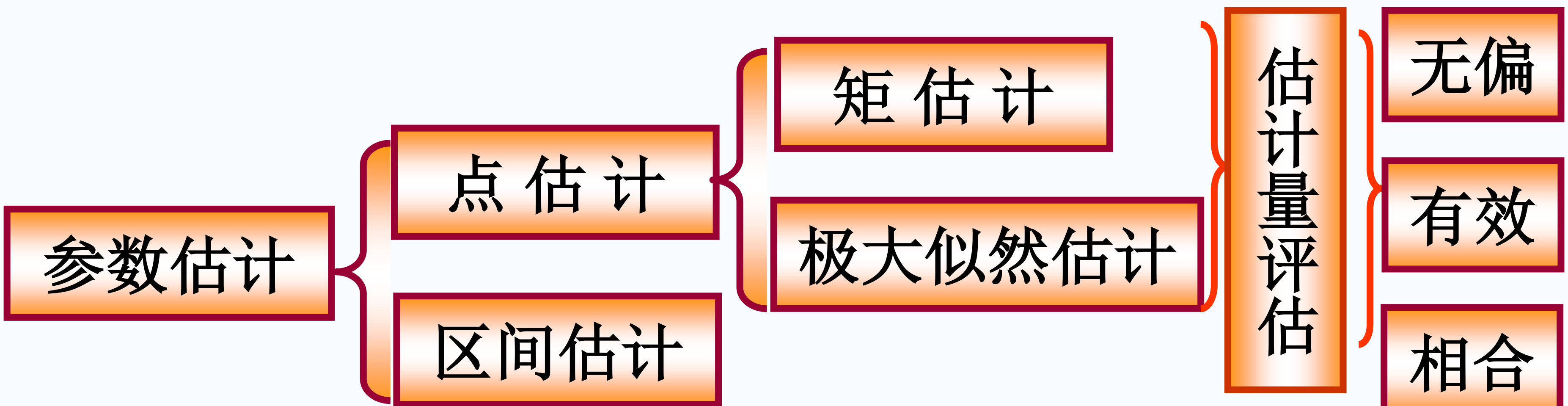
与平均次数 50 的误差不超过 10 次的概率（最终结果精确到小数点后两位）为（　　）.

已知：$\Phi(1) = 0.8413, \Phi(2) = 0.9772$.

（A）0.90;　　（B）0.85;　　（C）0.99;　　（D）0.95.

统计推断三个方面：

1. 抽样分布(精确分布)；
2. 参数估计；(已知分布类型)
3. 假设检验。



1. 随机变量X的$E(X), D(X)$存在, 且$D(X)>0$,则$X^* = \dfrac{X-E(X)}{\sqrt{D(X)}}$ $E(X^*)=0,\ D(X^*)=1$

样本: $(X_1, X_2, \cdots, X_n)$

(1) X_i与总体X相同分布; (2) $X_1, X_2, \cdots, X_n$相互独立.

$E(X_i)=E(X),\ D(X_i)=D(X)$

如果矩存在的话: $E(X_i^k)=E(X^k)$,

$E[X_i-E(X_i)]^k = E[X-E(X)]^k$

$E(\overline{X}_n)=E(X)$ $D(\overline{X}_n)=\dfrac{1}{n}D(X)$ $\Leftarrow \overline{X}_n=\dfrac{1}{n}\sum_{i=1}^{n}X_i$

统计学中最常用的公式

(1) $\sum_{i=1}^{n}(X_i-\overline{X})=0$; (2) $\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$

$M_2=A_2-A_1^2=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=\dfrac{n-1}{n}S^2$

1. 总体、个体、样本、样本值、统计量(统计值);

简单样本:(1)X_i与总体同分布;(2)$X_1, X_2, \cdots, X_n$相互独立

$E(X_i)=E(X), D(X_i)=D(X)$ $\quad E(X_i^k)=E(X^k),\ E(|X_i^k|)=E(|X^k|)$

$$E\left(\overline{X}_n\right)=E\left(X\right) \qquad D\left(\overline{X}_n\right)=\frac{1}{n}D\left(X\right) \Leftarrow \overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$$

样本是一组随机变量，其具体试验（观察）数值记为：$x_1, x_2, \cdots, x_n$，称为样本观测值, 简称样本值.

$$A_1=\overline{X}$$

$$M_2=\frac{n-1}{n}S^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=A_2-A_1^2 \qquad S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \qquad M_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$$

$$(1)\sum_{i=1}^{n}(X_i-\overline{X})=0;$$

$$(2)\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2.$$

2. 三个结构定理：χ^2，T，F;

3. 两个抽样定理：单正态总体，双正态总体.

三大统计分布结构定理

定理6.2.1(英国的Pearson) 设 $X_1, X_2, \ldots, X_n$ 相互独立且都服从标准正态分布, 则

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$$

定理6.2.2(英国的Cosset) 设 随机变量X,Y 相互独立, $X \sim N(0,1), Y \sim \chi^2(n)$, 则

$$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$$

定理6.2.3(英国的Fisher) 设随机变量X, Y 相互独立, $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$

$$F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$$

$\chi^2 \sim \chi^2(n)$:性质1(数字特征) 则有 $E(\chi^2) = n$, $D(\chi^2) = 2n$

性质2(可加性)设Y_1, Y_2相互独立且$Y_1 \sim \chi^2(n_1), Y_1 \sim \chi^2(n_2)$,则 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$

设随机变量 X 和 Y 都服从标准正态分布，则必有（　　）.

（A） $X + Y$ 服从正态分布 $N(0,2)$；　（B） $\dfrac{X}{|Y|}$ 服从 $t(1)$ 分布；

（C） X^2 和 Y^2 都服从 $\chi^2(1)$ 分布；　（D） $\dfrac{X^2}{Y^2}$ 服从 $F(1,1)$ 分布.

上侧分位数 (0< α<1):

$$\boxed{t_{1-\alpha}(n) = -t_{\alpha}(n)}$$

n 较大时(n>45),

$$t_{\alpha}(n) \approx u_{\alpha}$$

$$\chi_{\alpha}^{2}(n) \approx n + u_{\alpha}\sqrt{2n},$$

$X \sim N(0,1)$ $P\{X > u_{\alpha}\} = \alpha$

$\chi^2 \sim \chi^2(n)$ $P\{\chi^2 > \chi_{\alpha}^{2}(n)\} = \alpha$

$T \sim t(n)$: $P\{T > t_{\alpha}(n)\} = \alpha$

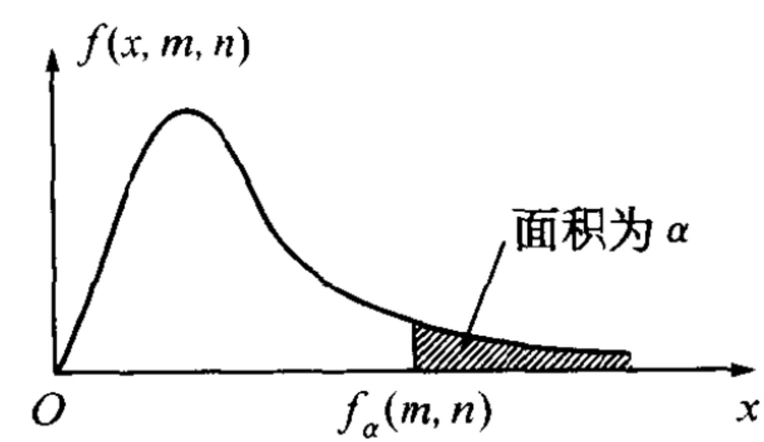
$F \sim F(n_1, n_2)$

$P\{F > F_{\alpha}(n_1, n_2)\} = \alpha$

若 $F \sim F(n_1, n_2) \Rightarrow \dfrac{1}{F} \sim F(n_2, n_1)$

若 $F \sim F(n_1, n_2)$

$$\Rightarrow F_{\alpha}(n_2, n_1) = \frac{1}{F_{1-\alpha}(n_1, n_2)}$$



二、抽样分布定理(正态总体的统计量分布)

定理6.2.4 设 $X_1, X_2, \ldots, X_n$ 是正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，$\overline{X}, S^2$ 分别是样本均值和样本方差，则

(1) $\overline{X}$ 与 S^2 相互独立; $\Leftrightarrow \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$; (2) $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$;

(3) $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$; $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$; (4) $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

定理6.2.5: 设正态总体 X 与 Y 相互独立，

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

(2) 当 $\sigma_1^2 = \sigma_2^2$ 时，

$$T = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

$$S_w = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

注意方差不等时，双样本的情况：

例1：在两个工厂生产的蓄电池中，分别取10个蓄电池测得其电容量（单位：安培小时）如下：

甲厂:140,141,135,142,140,143,138,137,142,137

乙厂:141,143,139,139,140,141,138,140,142,138

试检验两厂蓄电池的性能有无显著差异（取显著水平α=0.0.5）。

解：①设X为甲厂的蓄电池电容量，Y为乙厂的蓄电池电容量

$$\because X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)\ \mu_1,\sigma_1,\mu_2,\sigma_2\text{均未知}$$

方法二：由于$n_1=n_2=10$,可成对抽取$Z_i=X_i-Y_i,i=1,\cdots,10$

$$\therefore \overline{Z}=\overline{X}-\overline{Y}\sim N(\mu_1-\mu_2,\frac{1}{10}(\sigma_1^2+\sigma_2^2))$$

检验：$\mathrm{H}_0:\mu_1-\mu_2=0$，$\mathrm{H}_1:\mu_1-\mu_2\neq 0$

在$\mu_1=\mu_2$时,$T=\dfrac{\overline{Z}}{S_z/\sqrt{n}}\sim t(n-1),S_z=\dfrac{1}{10-1}\sum_{1}^{n}(z_i-\overline{z})^2$

代入数据$|t|\approx 0.0426<t_{0.025}(9)=2.2622$

所以接受$\mathrm{H}_0:\mu_1-\mu_2=0$。

1)矩估计　基本思想 替换原则:用样本矩替换相应的总体矩(LLN)

矩估计法前提: 总体X 的m阶原点矩存在，即$E(X^m)$

替换原则:　令 $\frac{1}{n}\sum_{i=1}^{n} X_i^K = E(X^K)$, $K=1, 2, \ldots, m$

$m=1$, 令 $\overline{X}=E(X)$　$m=2$, 令 $\begin{cases}\overline{X}=E(X)\\ M_2=D(X)\end{cases}$

注意：像均值和方差相等情况,一般选用低阶矩.

2) 极大似然估计

基本思想：根据小概率事件原理, 按照最大可能性准则进行推断.

基本方法：求参数θ的估计值, 使似然函数达到极大值.

似然函数: 连续为联合概率密度, 离散为联合分布律.

优良性准则: 无偏性(没有系统误差,消除随机误差)和有效性(无偏性下, 控制随机误差大小)都是固定样本容量n:

相合性 (n较大, 与第五章*chebyshev*不等式结合)

$\overline{X}$ 和S^2 分别是μ和σ^2 的最小方差无偏和相合估计。

均值或者方差已知？双侧还是单侧？

区间估计(双侧)

$X \sim N(\mu, \sigma^2)$ $P\{w_{1-\alpha/2} \le W \le w_{\alpha/2}\} = 1-\alpha$

原则:选取最简单的优良估计量

原则:无偏、有效、相合

被估参数	条件	枢轴变量	优良估计量
μ	已知 σ^2	$U = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\bar{X}$
μ	未知 σ^2	$T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$\bar{X}$
σ^2	已知 μ	$\dfrac{n}{\sigma^2}\cdot\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$	$\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$
σ^2	未知 μ	$\dfrac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\dfrac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$	S^2

被估参数	条件	枢轴变量	优良估计量
$\mu_1-\mu_2$	已知σ_1^2与σ_2^2	$U=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\bar{X}-\bar{Y}$
$\mu_1-\mu_2$	未知σ_1^2和$\sigma_2^2(\sigma_1^2=\sigma_2^2)$	$\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$	$\bar{X}-\bar{Y}$
$\dfrac{\sigma_2^2}{\sigma_1^2}$	未知μ_1和μ_2	$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\dfrac{\sum_{i=1}^{n_1}\left(\dfrac{X_i-\bar{X}}{\sigma_1}\right)^2\Big/n_1-1}{\sum_{j=1}^{n_2}\left(\dfrac{Y_j-\bar{Y}}{\sigma_2}\right)^2\Big/n_2-1}\sim F(n_1-1,n_2-1)$	$\dfrac{S_2^2}{S_1^2}$
$\dfrac{\sigma_2^2}{\sigma_1^2}$	已知μ_1和μ_2	$\dfrac{\sum_{i=1}^{n_1}\left(\dfrac{X_i-\mu_1}{\sigma_1}\right)^2\Big/n_1}{\sum_{j=1}^{n_2}\left(\dfrac{Y_j-\mu_2}{\sigma_2}\right)^2\Big/n_2}\sim F(n_1,n_2)$	$\dfrac{\dfrac{1}{n_2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2}{\dfrac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}$

单侧置信区间

习题19，20

$$P\{\hat{\theta}_1(X_1, X_2, \cdots, X_n) \le \theta\} = 1-\alpha$$

则称随机区间[$\hat{\theta}_1$，$+\infty$]为参数θ的置信度为1-α的单侧置信区间. $\hat{\theta}_1$称为单侧置信下限

$$P\{\hat{\theta}_2(X_1, X_2, \cdots, X_n) \ge \theta\} = 1-\alpha$$

则称随机区间[$-\infty$，$\hat{\theta}_2$]为参数θ的置信度为1-α的单侧置信区间. $\hat{\theta}_2$称为单侧置信上限

第八章 假设检验

假设检验的基本思想：提出统计假设, 根据小概率事件原理对其进行检验. 具有概率性质的反证法.

假设检验目的：根据样本去推断是否拒绝原假设H_0

1. 检验统计量确定：与枢轴变量形式一致　均值或者方差已知?

2. 单侧和双侧:常规假设或保护原来情况为原假设;　双侧还是单侧?
新事物或新情况作为备择假设(教材例8.2.1和例8.2.2)

3. 拒绝域的确定(遵循有利准则):对H_1成立有利的区域作为拒绝域.

4.两类错误原因：样本随机性和推导的原理(小概率事件实际不发生)

5. 两类错误:第一类:弃真(落在拒绝域);第二类:纳伪(落在接受域)

不可能使两类错误同时都尽可能小！减小一类错误，必然使另一类错误增大.

先控制犯第一类错误的概率α，然后再使犯第二类错误的概率尽可能地小$\beta(\mu)$.

例: $X\sim N(\mu,\sigma_0^2)$, $\beta(\mu)=\Phi\left(u_{\alpha/2}-\dfrac{\mu-\mu_0}{\sigma_0/\sqrt{n}}\right)-\Phi\left(-u_{\alpha/2}-\dfrac{\mu-\mu_0}{\sigma_0/\sqrt{n}}\right)$

七和八章主要区别：随机区间以较大概率包含待估参数;
该事件对应确定假设检验的接受域.其对立事件就能确定假设检验的拒绝域.

第六-八章 注意事项(丢分或掉坑的地方)

注: (重点)书写规范:

$\overline{X}, S^2, A_k, M_k$ 统计量: 大写

统计值:小写 $\overline{x}, s^2, a_k, m_k$

矩估计法先求出的是：估计量 → 大写

例4

极大似然估计法先求出的是：估计值 → 小写

例5

教材例7.1.6和例7.1.7:

$$\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$$ 估计量

$$\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}$$ 估计值

如矩估计中：令或替换原则 $\overline{X}=E(X)$

由题意检验假设 $H_0:\mu_1-\mu_2=0, H_1:\mu_1-\mu_2\neq 0$

当H_0成立时， 检验统计量为 $T=\dfrac{\overline{X}-\overline{Y}}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$

均值或者方差已知?
双侧还是单侧?

不可少 不可少

(末尾) 在显著性水平$\alpha=0.05$下, 认为这两个厂的产品质量没显著差别.

例2.(1)问该日生产的保险丝熔化时间的方差时否不超过400（α=0.01).

(2)检验甲枪弹的平均速度比乙枪弹的平均速度是否显著的大（α=0.05)

⑶ 能否认为灯泡的平均寿命在采用新工艺后有明显提高（α=0.01）

第九章 回归分析(包含丢分或掉坑的地方)

基本思想: 根据自变量X_1, X_2, $\cdots$ X_k与因变量Y的观察值去估计回归函数.

1. 一元线性回归模型(标准形式): $Y=a+bx+\varepsilon$, ε~$N(0,\sigma^2)$

(最小二乘法思想: 误差或残差平方和的极小值点)

$$Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\hat{a}-\hat{b}x_i)^2=\sum_{i=1}^{n}\varepsilon_i^{\ 2}$$ (注意:书后的第1题)

(回归参数b,回归常数a的估计:重点看教材例9.2.1和例9.2.2)

一元经验线性回归方程: $\hat{y}=\hat{a}+\hat{b}\,x$ $\qquad \hat{b}=\dfrac{l_{xy}}{l_{xx}} \quad \hat{a}=\bar{y}-\hat{b}\,\bar{x}$

不可少

2. (重点)相关系数的显著性检验法. $\hat{\sigma}^2=\dfrac{1}{n-2}(l_{yy}-\hat{b}^2 l_{xx})$

拒绝域: $|R|=|\hat{\rho}_{XY}|=\left|\dfrac{l_{XY}}{\sqrt{l_{XX}}\sqrt{l_{YY}}}\right|>R_\alpha(n-2)$ 例6

3.非线性回归问题的线性化处理.例9.4.3(教材)

第九章 回归分析(包含丢分或掉坑的地方)

基本思想: 根据自变量X_1, X_2, $\cdots$ X_k与因变量Y的观察值去估计回归函数.

1. 一元线性回归模型(标准形式): $Y=a+bx+\varepsilon$, $\varepsilon \sim N(0,\sigma^2)$
(最小二乘法思想: 误差或残差平方和的极小值点)

$$Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\hat{a}-\hat{b}x_i)^2=\sum_{i=1}^{n}\varepsilon_i^2$$ (注意:书后的第1题)

一元经验线性回归方程: $\hat{y}=\hat{a}+\hat{b}x$ $\qquad \hat{b}=\dfrac{l_{xy}}{l_{xx}}$ $\qquad \hat{a}=\bar{y}-\hat{b}\bar{x}$

不可少

$$\hat{\sigma}^2=\frac{1}{n-2}(l_{yy}-\hat{b}^2 l_{xx})$$

祝 大 家 考 试 顺 利！