第 一 章 的重点

1、事件关系，运算规律，加法定理

2、互斥与独立事件

3、全概率及贝叶斯公式的应用：书21页例1.3.10

第 一 章 的核心（条件概率，乘法公式，加法公式）

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(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：$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{P(A)}$$

抽签公平性

例1

第 一 章 补充

1. 六个事件关系、运算规律(5-7页):

德·摩根律: $\overline{A \cup B} = \overline{A} \cap \overline{B}$ $\quad \overline{A \cap B} = \overline{A} \cup \overline{B}$

$$A - B = A - AB = (A \cup B) - B$$

2、概率的一些重要性质:

1) (概率单调性) 若事件 A 和 B 满足 $A \subset B$, 则有 $P(A) \le P(B)$, $P(B - A) = P(B) - P(A)$ 成立

P14例1.2.6: $\quad (1) P(A - B) = P(A) - P(AB);$

$\quad (2) P(A - B) = P(A \cup B) - P(B).$

2).概率加法定理: 对试验 E 的任意两个事件 A 和 B 有

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

第 一 章 补充

独立与互斥的概念

1. 独立事件和互不相容事件

1) 称A、B为<u>互不相容，</u>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：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：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。

若 $P(A)>0, P(B)>0$,当 A与B互不相容,有

$P(AB)=P(\varphi)=0$,当A与B相互独立,则有

$P(AB)=P(A)P(B)>0$,两者不能同时成立.

对于任意两个随机事件A和B，$0<P(B)<1$,以下结论中正确的有______.

(1)若$AB=\emptyset$, 则A与B一定相互独立;

(2)若 $AB\neq\phi$, 则$P(AB)>0$;

(3) 若$P(A|B)=P(A)$,则A与B相互独立;

(4) 若$P(AB)=0$,则A与B互不相容.

第 一 章 补充

试求A_1，A_2，…，A_n至少有一个发生的概率，其中$0 < P(A_i)=p_i < 1$，

若（1）A_1，A_2，…，A_n 互不相容，

$$P= P(A_1)+ P(A_2)+ \dots + P(A_n)=p_1+p_2+ \dots +p_n$$

（2）A_1，A_2，…，A_n相互独立。

$$P = P\{ A_1 \cup A_2 \cup \dots \cup A_n \}$$

$$= 1 - P\{ \overline{A}_1 \overline{A}_2 \dots \overline{A}_n \} = 1 - \prod_{i}^{n} (1 - p_i)$$

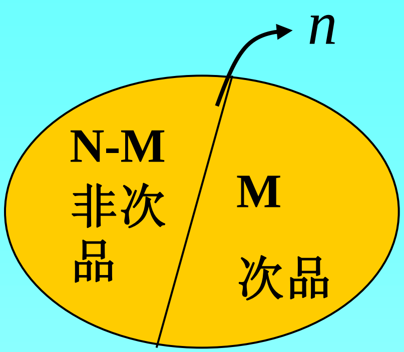
特别，当 $P(A_i)= p$, $i=1, 2, \dots, n$,有

$$P\{ A_1 \cup A_2 \cup \dots \cup A_n \} = 1 - (1 - p)^n$$

任意关系时：$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n}^{n} P(A_i A_j)$

$$+ \sum_{1 \le i < j < k \le n}^{n} P(A_i A_j A_k) - \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n)$$

推广一般模型：一批同类产品共N件，其中有M件次品，从中随机抽取n件，求恰有m件次品的概率。



B={有m件次品}

$$P(B)=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n},\ m=0,1,\cdots,l=\min\{M,n\}$$

超几何分布：不考虑顺序（组合角度）

(1)题目改为依次不放回抽取n件（考虑顺序）

$$P(B)=C_n^m\cdot\frac{P_M^m P_{N-M}^{n-m}}{P_N^n}=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$$

(2)题目改为有放回依次抽取n件

$$P(B)=\frac{M^m\bullet(N-M)^{n-m}}{N^n}\bullet C_n^m=C_n^m\left(\frac{M}{N}\right)^m\cdot\left(\frac{N-M}{N}\right)^{n-m}$$

——二项分布

第 二至三章的重点

1、分布函数的性质及分段表达

2、分布律与密度函数的非负, 归一性(五种常用分布)

3、随机变量函数的密度函数：书85页例3.4.4和例3.4.5

基本的分布函数法:重点是一维

例2

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

套用（必写） 大写 小写 分段

注：极值分布(88页)

X+Y和的密度函数

(书89页例3.4.10及3.4.12)

例3（2006年半期

书99页32题）

4、随机变量的独立性，正态分布的可加性及二维正态的性质

例4 例5

第 二 章 补充

对于一个贝努里试验，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；

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

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

首次成功的试验次数

例1:某人有3发子弹，他射击空中气球的命中率为0.9，命中则停止射击. 他用去的子弹个数X的分布律为？

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；

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

负二项分布（或帕斯卡(Pascal)分布）

(3) n 次试验中事件A 发生的总次数X. ($B(n,p)$)

1、事件A 首次发生时的试验次数Z服从几何分布

$$P\{Z=k\}=p(1-p)^{k-1},k=1,2,\cdots$$ 第二章习题第2题(1)

2、对于一个n重贝努里试验，可以考察如下问题：

(1) n 次试验中事件A 发生的总次数X服从二项分布；

(2) 事件A 首次发生时的试验次数Z；

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

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

例1：某人有3发子弹，他射击空中气球的命中率为0.9，命中则停止射击。他用去的子弹个数X的分布律为？

首次成功的试验次数

例2 在一条汽车通道上有四盏信号灯，每盏灯各以1/2的概率允许或禁止汽车通行，试求汽车停止前进时已通通过的信号灯盏数X的分布律和分布函数。(信号灯独立）

解：设A_i＝{在第i盏灯前禁止汽车通过},i=1,2,3,4

首次成功的试验次数

$$P\{X=0\}=P(A_1)=p=\frac{1}{2}$$

$$P\{X=1\}=P(\overline{A}_1A_2)=(1-p)p=(\frac{1}{2})^2$$

$$P\{X=2\}=P(\overline{A}_1\overline{A}_2A_3)=(1-p)p^2=(\frac{1}{2})^3$$

$$P\{X=3\}=P(\overline{A}_1\overline{A}_2\overline{A}_3A_4)=(1-p)p^3=(\frac{1}{2})^4$$

$$P\{X=4\}=P(\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4)=(1-p)^4=(\frac{1}{2})^4$$

$$
F(x)=\begin{cases}
0, x<0 \\
0.5, 0\le x<1 \\
0.5+(0.5)^2=0.75, 1\le x<2 \\
0.5+(0.5)^2+(0.5)^3=0.875, 2\le x<3 \\
0.5+(0.5)^2+(0.5)^3+(0.5)^4 \\
=0.9375, \qquad 3\le x<4 \\
0.5+(0.5)^2+(0.5)^3+(0.5)^4+(0.5)^4 \\
=1, \qquad x\ge 4
\end{cases}
$$

1.分布函数: 设X是一个随机变量, x是任意实数,称函数

$$F(x) = P\{X \le x\} = P\{\omega: X(\omega) \le x\},$$

为随机变量X 的分布函数, $F(x)$ 也记为$F_X(x)$.

1) $P\{a < X \le b\} = P\{X \le b\} - P\{X \le a\} = F(b) - F(a)$

$P\{X = x\} = F(x) - F(x-)$

2) 分布函数的性质:

(1)若 $x_1 \le x_2$, 则有$F(x_1) \le F(x_2)$

（2） $0 \le F(x) \le 1$, 且$\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$

（3） $F(x)$ 是右连续函数, 即$F(x+0) = F(x)$

二、 离散型随机变量

$P\{X=x_i\}=p_i$:(1) $p_i \geqslant 0$;(2) $\sum_{i=1}^{\infty} p_i = 1$

分布函数与分布律的关系:

$$F(x)=P\{X \le x\} = P[\bigcup_{x_i \le x} \{X = x_i\}] = \sum_{x_i \le x} P\{X = x_i\}$$

$P\{X= x\} = F(x) - F(x-)$ (对所有满足 $x_i \le x$ 的 i 求和)

三、 连续型随机变量

（1）$f(x) \geqslant 0$; （非负性）

（2）$\int_{-\infty}^{+\infty} f(x)dt = 1.$ （概率曲线下面积为1）

4）能进行分布函数与概率密度的转换;

$F(x) = \int_{-\infty}^{x} f(t)\,dt$ $\qquad F'(x) = f(x)$

一、二维随机变量的联合分布

分布函数的定义：$F(x,y)=P\{X\le x, Y\le y\}$

四条性质：$P\left\{x_1 < X \le x_2, y_1 < Y \le y_2\right\}$

$= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$

联合分布函数及性质与边缘分布函数的关系

$$F_X(x) = P\{X \le x\} = P\{X \le x, Y < +\infty\} = \lim_{y\to+\infty} F(x,y)$$

$$F_Y(y) = P\{Y \le y\} = P\{X \le +\infty, Y < y\} = \lim_{x\to+\infty} F(x,y)$$

二、离散型：联合分布律

$$P\{X=x_i, Y=y_j\}=p_{ij} \quad i,j=1,2,\ldots. \qquad (*)$$

若 1) $p_{ij}\geq 0 \qquad i,j=1,2,\ldots..$

2) $\sum_i \sum_j p_{ij}=1$

$$F(x,y)=P\{X\leq x, Y\leq y\}=\sum_{x_i\leq x}\sum_{y_j\leq y} p_{ij}$$

联合分布律及性质与边缘分布律关系

$$P\{X=x_i\}=\sum_{j=1}^{+\infty} p_{ij}=p_{i\cdot} \quad i=1,2,\ldots$$

$$P\{Y=y_j\}=\sum_{i=1}^{+\infty} p_{ij}=p_{\cdot j} \quad j=1,2,\ldots$$

三、连续型：联合概率密度

1.联合概率密度及性质

$$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)dudv$$

2.边缘概率密度

注意带参变量积分的计算

X,Y的边缘概率密度为

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

正态分布(GAUSS分布) $N(\mu,\sigma^2)$（$\sigma>0$）

$$\varphi(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad x\in R$$

有 $\Phi(-x)=1-\Phi(x)$

分位数　　$X\sim N(0,1)$，若实数u_α使

$$P\{X>u_\alpha\}=\alpha$$

则称u_α为标准正态分布的对应于α的上侧分位数.

$$-u_\alpha=u_{1-\alpha}$$

$$\Phi(u_\alpha)=1-\alpha$$

第 三 章 补充

重要结论：

P87 例3.4.7:正态随机变量的线性函数也服从正态分布

$$X \sim N(\mu, \sigma^2), Y = aX + b(a \neq \mathbf{0})$$

$$\Rightarrow Y \sim N(a\mu + b, (a\sigma)^2)$$

P_{90}页例3.4.11 正态分布具有可加性：

X与Y相互独立，$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2),$

$$\Rightarrow X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

若Y服从正态分布, 而Y表成两个独立随机变量X_1, X_1之和, 则X_1, X_1必服从正态分布. 这称为正态分布的“再生性”.

二维正态分布　　　第 三 章 补充

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

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$\Leftrightarrow\rho=0$

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

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)

第 三 章 补充

1.设随机变量X与Y相互独立，$X \sim N(0,\frac{1}{2}), Y \sim N(0,\frac{1}{2})$,

试求概率$P\{X-Y\geq 0\}$。

解：由正态分布随机变量的可加性有：

$$Z = X - Y \sim N(0,1)$$

正态随机变量Z关于纵轴对称，

故$P\{X-Y\geq 0\}=\frac{1}{2}$

第 三 章的补充

1.概率密度函数与条件概率密度函数

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

当$y>0$时，$\boldsymbol{f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}=\begin{cases} \dfrac{1}{y}, & y>x>0 \\ 0, & \text{其他} \end{cases}}$

满足非负、归一性

2.分布函数与条件分布函数：79页例3.3.3

满足三条性质

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

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

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

为0

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

不为0

例6

第 三 章的补充

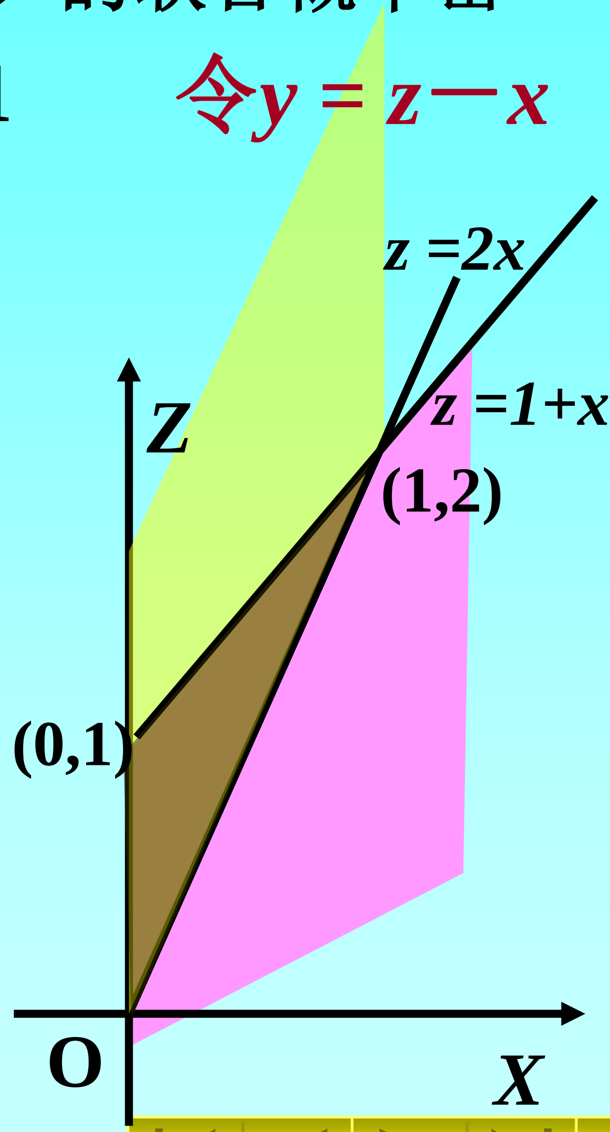
例3.4.6：已知二维随机变量（*X,Y*）的联合概率密度为

$$f(x,y)=\begin{cases}2(x+y) & 0\le x\le y\le 1\\ 0 & 其他\end{cases}$$

令$y=z-x$

求：*Z=X+Y* 的概率密度。

解：在*XOZ*平面上作出区域



$$G=\{(x,z)|0\le x\le z-x\le 1\}$$

$$=\left\{(x,z)\middle|0\le x\le z/2,\ 2x\le z\le 1+x\right\}$$

$$f(x,z-x)=\begin{cases}2z & (x,z)\in G\\ 0 & 其他\end{cases}$$

第 三 章的补充

当 $z\le 0$ 或 $z\geq 2$ 时 $f_Z(z)=0$

当 $0<z\le 1$ 时 $f_Z(z)=\int_0^{z/2} 2z\,dx = z^2$

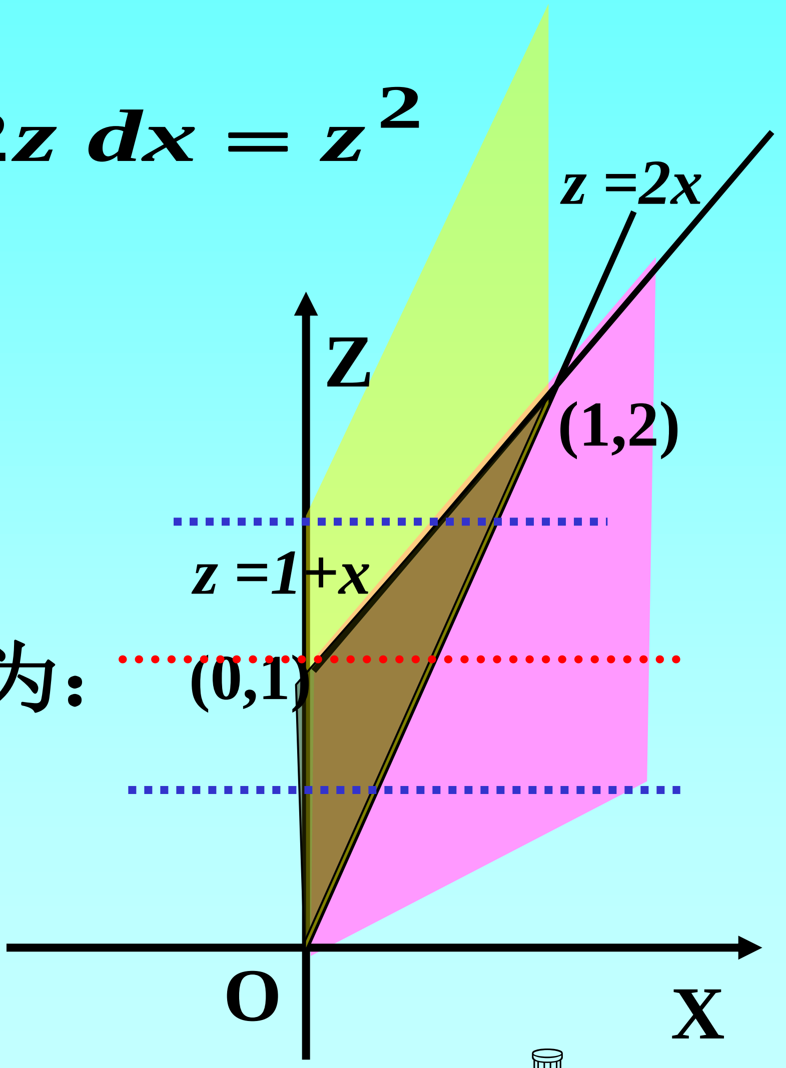
当 $1<z<2$ 时

$$f_z(z)=\int_{z-1}^{z/2} 2z\;dx$$

$$=2z-z^2$$

综上得 $Z=X+Y$ 的概率密度为：

$$f_z(z)=\begin{cases} z^2 & 0<z\le 1 \\ 2z-z^2 & 1<z\le 2 \\ 0 & \text{其他} \end{cases}$$



第四章随机变量的数字特征

都是实数

1.数学期望

离散

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

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

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

习题1

连续

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

习题5

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

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

性质：4条

1）$E(c)=c$

2）$E(cX)=cE(X)$ $\Longleftrightarrow$ 2）$E(cX+b)=cE(X)+b$

3）$E\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} E(X_i)$

4）若$X_1,X_2,\ldots, X_n$相互独立，则

$$E\left(\prod_{i=1}^{n} X_i\right)=\prod_{i=1}^{n} E(X_i)$$

习题11,13

2.方差：$D(X)= E\{[X-E(X)]^2\}=D(X)=E(X^2)-E(X)^2$

性质：3条

1) $D(c)=0$　　2) $D(cX)=c^2D(X)$

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

常用分布的期望和方差表（记住）：

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

二者无线性关系

$\rho_{XY}=0$, 称X与Y不相关

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 $\Longleftrightarrow$ $\rho=0$ （不相关）

二、数字特征计算

1. 利用数字特征的性质；

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

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

$$=3E(X^2)=3$$

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

2. 利用特殊分布的可加性.

记住：正态、二项、泊松、均匀、指数的数字特征.

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:118 (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

$\not\Longrightarrow$ $l_1Z_1+l_2Z_2$ (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)

第 五 章

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.

（第七章）

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

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

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。

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \varepsilon\} = 1 \quad 即 \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i) \xrightarrow{P} 0$$

2. 掌握
- 切比雪夫大数定律；
- 独立同分布大数定律；
- <u>贝努里(***Bernulli***)大数</u>大数定律；

3. 重要结论：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
- 与泊松逼近定理区别 习题4—6,10

3. 作用
- 概率近似计算；
- 确定大样本估计量的分布.（第七和八章）

170页例7.3.7

第 五章的重点

1. 切比雪夫不等式

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

2.中心极限定理

1)独立同分布中心极限定理

2)棣莫佛—拉普拉斯中心极限定理

注：(1)中心极限定理中 “≈” $\longrightarrow$ (1~3分)

(2)与泊松逼近定理区别

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

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

第六 章的重点

1. 总体、个体、样本、样本值、统计量(统计值);

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

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。

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

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

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.

第七 章的重点

参数估计: 点估计和区间估计(枢轴变量)

1)矩估计 基本思想 替换原则:用样本矩替换相应的总体矩

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.

三个优良性准则中: 无偏和有效 (与第四章结合,n固定)

相合性 (n不固定, 与第五章*chebyshev*不等式结合)

注: (重点)书写规范:

156页例7.1.6和例7.1.7:

估计量 → 大写

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

估计值 → 小写

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

例4

如矩估计中: 令或替换原则 $\overline{X}=E(X)$

第八章的重点

1.正态总体的参数假设检验(单侧和双侧);

2. 单侧和双侧: 原假设和备择假设的提法:

新事物出现的情况作为备择假设

(183页例8.2.1和184页例8.2.2)

3.双侧: 186页例8.2.3

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 $$T = \frac{\overline{X} - \overline{Y}}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

不可少 不可少

(末尾) 在显著性水平$\alpha = 0.05$下,

不可少 认为这两个厂的产品质量没显著差别.

例5

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.

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

2. 单侧和双侧：新事物出现的情况作为备择假设

(183页例8.2.1和184页例8.2.2)

3. 拒绝域的确定：确定H_0的拒绝域时应遵循有利准则

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：$X \sim N(\mu,\sigma_0^2)$,

$$\beta(\mu)=\Phi\left(u_{\alpha/2}+\frac{\mu-\mu_0}{\sigma_0/\sqrt{n}}\right)-\Phi\left(-u_{\alpha/2}-\frac{\mu-\mu_0}{\sigma_0/\sqrt{n}}\right)$$

第九章

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)

$$Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\hat{a}-\hat{b}x_i)^2=\sum_{i=1}^{n}\varepsilon_i^2$$

(回归参数b,回归常数a的估计:书205页例9.2.1和212例9.2.2)

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

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(书223页）

考试题型：简答题、综合和计算题(与半期一致).

考试必胜宝典：

复习原则：

认真看书,

掌握基本理论、基本方法、基本思想.

1. 一电子信号在[0,T]时间内随机出现，设信号在[0,T/4)内出现，[T/4,T/2)内出现，[T/2,T]内出现的情况下被截获的概率分别是 1/3, 2/3, 1/6；现信号已被截获，求它是在[T/2,T]内出现的概率。

设电子信号出现的时间为 X，则 X~U(0,T)

袋中有四枚硬币：甲硬币两面都是花；乙硬币两面都是字；丙硬币一面是花一面是字，且质地均匀；丁硬币一面是花一面是 字，但质地不均匀导致出现花的可能性是2/3。现从袋中任取一枚硬币抛掷一次，出现了花，求这枚硬币是丙硬币的概率。

设随机试验为在[0,1]区间上任取一点,
则不可能事件，小概率事件(如概率为1%的事件),
概率为0但可能发生的事件？

设二维随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}6xy^2 & 0<x<1,0<y<1\\ 0 & \text{其它}\end{cases}$$

判断 X 与 Y 的独立性，并求相关系数 ρ_{XY}.

2. 设二维随机向量（X，Y）的联合概率密度为 $f(x,y)=\begin{cases}e^{-y}, & 0<x<y;\\ 0, & \text{其它}.\end{cases}$，求

（1）$P\{X+Y<1\}$； (2) $f_{X|Y}(x\mid y)$.

5. 设随机变量 X_1, X_2 相互独立，都服从区间 $(0,2)$ 上的均匀分布，求 $Y=\max\{X_1, X_2\}$ 的数学期望。

解： $E(Y)=E(\max(X_1, X_2))$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{\infty}\max(x,y)f(x,y)dxdy$$

$$=\frac{1}{4}\int_0^2\int_0^2\max(x,y)dxdy$$

$$=\frac{1}{4}\int_0^2\int_0^x xdydx+\frac{1}{4}\int_0^2\int_0^y ydxdy=\frac{2}{3}+\frac{2}{3}=\frac{4}{3}$$

3. 加法器在做加法运算时,根据四舍五入原则先对每个数取整后再运算，这样产生的误差服从区间[-0.5,0.5]上的均匀分布。问：要使误差总和的绝对值不超过 10 的概率大于 0.95，最多能有多少个数相加?（$\Phi(0.95)=0.829$，$\Phi(0.05)=0.52$，$\Phi(1.645)=0.95$，$\Phi(1.96)=0.975$）

解：设 X_i 为第 i 个数四舍五入产生的误差，

则 $X_i \sim U(-0.5,0.5)$，$i=1,2,\cdots,n$ 相互独立。

误差总和为 $\sum_{i=1}^{n} X_i$，$E\left(\sum_{i=1}^{n} X_i\right)=0, D\left(\sum_{i=1}^{n} X_i\right)=\frac{n}{12}$，

根据独立同分布中心极限定理得

$$P\{|\sum_{i=1}^{n} X_i| \le 10\} > 0.95 = P\{-\frac{10}{\sqrt{n/12}} \le \frac{\sum_{i=1}^{n} X_i}{\sqrt{n/12}} \le \frac{10}{\sqrt{n/12}}\}$$

$$\approx 2\Phi\left(\frac{10}{\sqrt{n/12}}\right)-1 \ge 0.95 \Rightarrow \frac{10}{\sqrt{n/12}} \ge 1.96 \Rightarrow n \le 312$$

•**已知某型号的二极管寿命服从参数为0.05的指数分布，现有盒装该型号二极管100个，问有多大把握能够保证这盒二极管总的使用时间不少于1800小时？**

2. 设总体 $X \sim N(0,1), Y \sim N(0,4)$ 相互独立， X_1, X_2, X_3, X_4 和 $Y_1, Y_2, Y_3 \cdots, Y_9$ 分别为来自总体 X 和 Y 的样本，确定统计量 $Z = \dfrac{3(X_1 + X_2 + X_3 + X_4)}{\sqrt{\sum_{i=1}^{9} Y_i^2}}$ 的分布.

σ

3. 设测量误差服从零均值的正态分布，进行 4 次独立测量，各次误差为 0.1, -0.08, 0.04, -0.07。求测量误差方差的极大似然估计值。 (1)矩估计量；

解：设测量误差 $X \sim N(0,\sigma^2)$， x_1,x_2,x_3,x_4 为样本观测值。

似然函数为

$$L(\sigma)=\prod_{i=1}^{4}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x_i^2}{2\sigma^2}}=\frac{1}{4\pi^2\sigma^4}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{4}x_i^2}\text{（5分）}=\frac{1}{4\pi^2\sigma^4}e^{-\frac{0.0229}{\sigma^2}}$$

令 $\dfrac{d\ln L}{d\sigma}=0$，得极大似然估计值 $\hat{\sigma}^2=0.005725$。

令 $E(X^2)=A_2$ 或 $D(X)=M_2$，可得

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2\cdots$$

矩估计量 $\sigma^2=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2$

4. 设 $\hat{\mu}_1$ 和 $\hat{\mu}_2$ 为总体均值 μ 的两个估计量，已知 $E(\hat{\mu}_1)=E(\hat{\mu}_2)=\mu$，$D(\hat{\mu}_1)=1, D(\hat{\mu}_2)=4$, $\rho_{\hat{\mu}_1\hat{\mu}_2}=1/2$, 作统计量 $\hat{\mu}_3=c_1\hat{\mu}_1+c_2\hat{\mu}_2(c_1>0, c_2>0)$，问：要使形如 $\hat{\mu}_3$ 的统计量为 μ 的无偏估计量且最有效，常数 c_1, c_2 应满足什么条件？（写出式子即可，不必解出具体结果）

解：$E(\hat{\mu}_3)=c_1\mu+c_2\mu=\mu \Rightarrow c_1+c_2=1$

应在 $c_1+c_2=1$ 的条件下

使 $D(\hat{\mu}_3)=c_1^{\ 2}D(\hat{\mu}_1)+c_2^{\ 2}D(\hat{\mu}_2)+2c_1c_2\operatorname{cov}(\hat{\mu}_1,\hat{\mu}_2)=c_1^{\ 2}+4c_2^{\ 2}+2c_1c_2$

达到最小值，以此确定 c_1, c_2

4. 设一种电子元件的寿命 X（以小时计）服从正态分布，μ,σ^2 均未知。现测得 n=16 只元件的寿命，计算得：$\bar{x}=241.5, s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}=98.73$。问：是否有理由认为元件的平均寿命为 225（小时）？（取 $\alpha=0.1$，$u_{0.05}=1.645$，$t_{0.05}(15)=1.7531$）

4. 设学校校车在两校区间固定路线上的运行时间 $X\sim N(\mu,\sigma^2)$（单位：分钟）。现测得校车 16 次运行时间，计算得 $\bar{x}=41.5, s^2=7.8$。在 0. 1 的显著性水平下, 能否认为校车的平均运行时间为 40 分钟？（$u_{0.05}=1.645, u_{0.1}=1.28,$，$t_{0.05}(15)=1.7531, t_{0.05}(16)=1.7459$）

5. 观察落叶松的树龄 X 和平均高度 Y 有如下资料，

x_i	2	3	4	5	6	7	8
y_i	5.6	8	10.4	12.8	15.3	17.8	19.9

(1) 检验 Y 和 X 的线性相关显著性。（$\alpha=0.01, R_{0.01}(5)=0.8745, R_{0.01}(6)=0.8343$）

(2) 求平均高度 Y 关于树龄 X 的经验线性回归方程；

7. 有一种新安眠药，据说在一定剂量下，能比某种旧安眠药平均增加睡眠时间3小时，根据资料用某种旧安眠药时，平均睡眠时间为20.8小时。标准差为1.6小时，为了检验这个说法是否正确，收集到一组使用新安眠药的睡眠时间为26.7，22.0，24.1，21.0，27.2，25.0，23.4。试问：从这组数据能否说明新安眠药已达到新的疗效(假定睡眠时间服从正态分布，$\alpha=0.05$)。

解：$H_0: \mu_1 \geq 23.8$；$H_1: \mu_1 < 23.8$.

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

$$\cancel{t_{0.05}}\quad t \doteq 0.461 > -t_{0.05}(6) = -1.9432$$

∴拒绝H_1，接受H_0.

注意：新药的方差未知，用t分布， 1.6是旧药的标准差，1.6这个数据用不上，迷惑人。

宝典:

第七和八章，一定要区别总体均值，方差是否是已？是否变化?

4. 如果随机变量 Y 的分布函数 F(y)连续且严格单调增加，而随机变量 X～U(0,1) 令 $Z=F^{-1}(X)$，那么 Z 与 Y 同分布，请说明理由。

4. 理由：因为$X \sim U(0,1)$，所以

$$F_X(x)=\begin{cases}0, & x\le 0\\ x, & 0<x\le 1\\ 1, & x>1\end{cases}$$

所以：

$$\begin{aligned}F_Z(y)&=P\{Z\le y\}\\&=P\{F^{-1}(X)\le y\}\\&=P\{X\le F(y)\}\\&=F_X(F(y))\\&=F(y)\end{aligned}$$

备注：也可以不用分布函数，直接由$\frac{F(y)-0}{1-0}$算得结果$F(y)$

设随机变量X的分布函数F(x)连续且单调增加，求证: 随机变量Y=F(x) ~U(0,1) .

证明：由于$F(\cdot)\uparrow$，$F^{-1}(\cdot)$存在，

$$Y=F(X):\Omega\to[0,1]$$

$$\text{当}0<y<1\text{时}, F_Y(y)=P\{Y\le y\}=P\{F(X)\le y\}$$
$$=P\{X\le F^{-1}(y)\}$$
$$=F_X(F^{-1}(y))$$
$$=y$$

$$f_Y(y)=\begin{cases}1, & 0<y<1\\ 0, & else.\end{cases}\quad \text{或者}\quad F_Y(y)=\begin{cases}0, & y\le 0,\\ y, & 0<y<1,\\ 1, & y\ge 1\end{cases}$$

5. 一个病人可能得了甲、乙、丙三种病之中的一种，已知得这三种病的概率依次为 a_1,a_2,a_3（满足 $a_1+a_2+a_3=1$），为确定诊断，现在决定再做一种检验确认一下。1 次检验中，这种检验对疾病甲给出肯定结果的概率为 b_1，对于病症乙，则概率是 b_2，对于疾病丙，则概率是 b_3，一共进行了 n 次检验，有 m 次的结果是肯定的，根据该检验结果推断得疾病甲的概率。

设事件 $A=\{$病人得甲病$\}$，$B=\{$病人得乙病$\}$

$C=\{$病人得丙病$\}$，$D=\{$n次检验结果中m次肯定$\}$，由已知条件：

$$P(A)=a_1, P(D|A)=C_n^m b_1^m(1-b_1)^{n-m}$$

$$P(B)=a_2, P(D|B)=C_n^m b_2^m(1-b_2)^{n-m}$$

$$P(C)=a_3, P(D|C)=C_n^m b_3^m(1-b_3)^{n-m}$$

由贝叶斯公式：

$$P(A|D)=\frac{P(A)P(D|A)}{P(A)P(D|A)+P(B)P(D|B)+P(C)P(D|C)}$$

$$=\frac{a_1 b_1^m(1-b_1)^{n-m}}{\sum_{k=1}^{3} a_k b_k^m(1-b_k)^{n-m}}$$