第1讲 随机事件与概率

一、 基本观念

1. 随机试验

2. 随机事件

必然事件：Ω;

不可能事件：$\varnothing$

3. 样本空间

样本点：随机试验的每一个（不可再分）可能结果，记为ω，由一个样本点构成的事件称为基本事件

样本空间：$\Omega=\{\omega\}$

二、 事件的关系与运算

1. 关系

（1） 包含

（2） 相等

（3） 相容

（4） 对立

2. 运算

（1） 和（并）

（2） 差

（3） 积（交）

（4） 运算法则

① 吸收律

若$A \subset B$,则$A \cup B=B, A \cap B=A$.

② 交换律

$A \cup B=B \cup A, A \cap B=B \cap A$.

③ 结合律

$(A \cup B) \cup C=A \cup(B \cup C),(A \cap B) \cap C=A \cap(B \cap C)$.

④ 分配律

$A \cap(B \cup C)=(A \cap B) \cup(A \cap C), A \cup(B \cap C)=(A \cup B) \cap(A \cup C), A \cap(B-C)=(A \cap B)-(A \cap C)$.

⑤ 对偶律 (德・摩根律)

$\overline{A \cup B}=\bar{A} \cap \bar{B}, \overline{A \cap B}=\bar{A} \cup \bar{B}$.

三、 概率的定义

1. 描述性定义

2. 统计性定义

在相同条件下做重复试验,事件A出现的次数k和总的试验次数n之比$\frac{k}{n}$称为事件A在这n次试验中出现的频率,当试验次数n充分大时,频率将“稳定”于某常数$p.n$越大,频率偏离这个常数p的可能性越小.这个常数p就称为事件A的概率.（频率不是概率）

3. 公理化定义

设随机试验的样本空间为Ω,如果对每一个事件A都有一个确定的实数$P(A)$,且事件函数$P(\cdot)$满足:

（1） 非负性:$P(A) \geqslant 0$;

（2） 规范性:$P(\Omega)=1$;

（3） 可列可加性: 对任意可列个两两互不相容事件$A_1, A_2, \cdots, A_n, \cdots$(即$A_i A_j=\varnothing, i \neq j ; i, j=1,2, \cdots$),有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P\left(A_i\right), \tag{1}$$

则称$P(\cdot)$为概率, $P(A)$为事件A的概率.

四、 古典概型和几何概型

1. 古典概型

满足：

① 只有有限个样本点(基本事件);

②每个样本点(基本事件)发生的可能性都一样，

如果古典概型的基本事件总数为n,事件A包含k个基本事件,也叫作有利于A的基本事件为k个,则A的概率为

$$P(A)=\frac{k}{n} \tag{2}$$

由上式计算得出的概率称为A的古典概率.

(1) 数数法

① 列举法

② 集合对应法：

- 加法原理：完成一件事有n类办法, 第一类办法中有m_1种方法, 第二类办法中有m_2种方法, $\cdots\cdots$,第n类办法中有m_n种方法, 则完成此事共有$\sum_{i=1}^{n} m_i$种方法.
- 乘法原理：完成一件事有n个步骤. 第一步有m_1种方法, 第二步有m_2种方法, $\cdots\cdots$,第n步有m_n种方法,则完成此事共有$\prod_{i=1}^{n} m_i$种方法.
- 排列：从n个不同的元素中取出$m(\leqslant n)$个元素, 并按照一定顺序排成一列, 叫作排列. 所有排列的个数叫作排列数, 记作

$$\mathrm{P}_n^m=n(n-1)(n-2)\cdots(n-m+1)=\frac{n!}{(n-m)!}. \tag{3}$$

  当$m=n$时, $\mathrm{P}_m^n=\frac{n!}{0!}=n!$,叫作全排列.
- 组合—从n个不同的元素中取出$m(\leqslant n)$个元素, 并成一组, 叫作组合. 所有组合的个数叫作组合数, 记作

$$\mathrm{C}_n^m=\frac{\mathrm{P}_n^m}{m!}. \tag{4}$$

③ 逆数法: 先求$\bar{A}$中的基本事件数$n_{\bar{A}}$,将基本事件总数n减去$n_{\bar{A}}$便得A中的基本事件数, 这种方法常用于计算含有“至少”字样的事件的概率.

(2) 随机分配 【⭐⭐⭐】

将n个可辨质点随机地分配到N个盒子中

- 若每盒可容纳任意多个质点，则有N^n种不同分法
- 若每盒可容纳至多一个质点，则有P_N^n种不同分法

(3) 简单随机抽样

设$\Omega=\{\omega_1,\omega_2,\ldots,\omega_N\}$含$N$个元素,称$\Omega$为总体,如果各元素被抽到的可能性相同,自总体$\Omega$的抽样称作简单随机抽样,突出一个“取”字. 简单随机抽样分为先后有放回、先后无放回及任取这三种不同的方式,在每种抽样方式下各种不同抽法(基本事件)的总数列表如下.

抽样方式	抽样总数
先后有放回n次	N^n
先后无放回n次	P_N^n
任取n个	C_N^n

(5)

2. 几何概型

①样本空间(基本事件空间)Ω是一个可度量的有界区域;

②每个样本点(基本事件)发生的可能性都一样,即样本点落入Ω的某一可度量的子区域S的可能性大小与S的几何度量成正比,而与S的位置及形状无关.

在几何概型随机试验中,如果S是样本空间Ω的一个可度量的子区域,则事件$A=${样本点落人区域S}的概率为

$$P(A)=\frac{S_A\text{的几何度量}}{\Omega\text{的几何度量}} \tag{6}$$

由上式计算得出的概率称为A的几何概率.

五、 概率的基本性质与公式

1. 性质

(1) 有界性

(2) 单调性

2. 公式

(1) 加法公式

$$P(A\bigcup B)=P(A)+P(B)-P(AB).$$

$$P(A\bigcup B\bigcup C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC).$$

$P(A \bigcup B \bigcup C \bigcup D) = [P(A) + P(B) + P(C) + P(A)] - [P(AB) + P(AC) + P(AD) + P(BC) + P(BD) + P(CD)] + [P(ABC) + P(BCD)$

(2) 减法公式

$P(A - B) = P(A) - P(AB) = P(A\bar{B})$.

(3) 条件概率公式

设A, B为任意两个事件, 若$P(A) > 0$,我们称在已知事件A发生的条件下, 事件B发生的概率为条件概率, 记为$P(B \mid A)$,且

$$P(B \mid A) = \frac{P(AB)}{P(A)}. \tag{7}$$

(4) 乘法公式

如果$P(A) > 0$,则$P(AB) = P(A)P(B \mid A)$.

一般地, 对于$n > 2$,如果$P\left(A_1 A_2 \cdots A_{n-1}\right) > 0$,则

$$P\left(A_1 A_2 \cdots A_n\right) = P\left(A_1\right) P\left(A_2 \mid A_1\right) P\left(A_3 \mid A_1 A_2\right) \cdots P\left(A_n \mid A_1 A_2 \cdots A_{n-1}\right). \tag{8}$$

(5) 全概率公式

如果$\bigcup_{i=1}^n A_i = \Omega, A_i A_j = \varnothing (i \neq j; i, j = 1, 2, \cdots, n), P\left(A_i\right) > 0$,则对任事件$B$,有

$$B = \bigcup_{i=1}^n A_i B, P(B) = \sum_{i=1}^n P(A_i) P(B|A_i) \tag{9}$$

(6) 贝叶斯公式（逆概率公式）

如果$\bigcup_{i=1}^n A_i = \Omega, A_i A_j = \varnothing (i \neq j; i, j = 1, 2, \cdots, n), P\left(A_i\right) > 0$,则对任一事件$B$,只要$P(B) > 0$,就有

$$P(A_j|B) = \frac{P(A_j) P(B|A_j)}{\sum_{i=1}^n P(A_i) P(B|A_i)} (j = 1, 2, \ldots, n) \tag{10}$$

六、 独立性

1. 事件的独立性

- 直观性定义 设A, B为两个事件,如果其中任何一个事件发生的概率不受另外一个事件发生与否的影响,则称事件A与B相互独立.,...,A,相互独立.
- 数学定义设A, B为两个事件,如果$P(AB) = P(A)P(B)$,则称事件A与B相互独立,简称A与B独立.（如果$P(ABC) = P(A)P(B)P(C)$,则称事件A、B、C相互独立，且A、B、C两两独立；A、B、C两两独立不能推出A、B、C相互独立）

如果各个试验结果是相互独立的, 则称这些试验是相互独立的.

2. n重伯努利概型

① 每次试验只有A和$\overline{A}$两个结果

② 每次试验A发生的概率$p = P(A)$不变

③ 实验独立重复n次

如果用X表示n次独立试验事件A发生的次数，则$X \sim B(n, p)$，即

$$P(X = k) = C_n^k p^k (1 - p)^{n-k} (k = 0, 1, \ldots, n) \tag{11}$$

第2讲 一维随机变量及其分布

一、 一维随机变量

1. 概念

定义在Ω上的实值单值函数$X(\omega)$为随机变量，简记为随机变量X

（ω通过$X(\omega)$映射到实数轴上的X）

2. 分布函数

(1) 概念

设X是随机变量,x是任意实数, 称函数$F(x) = P\{X \leqslant x\} (x \in \mathbf{R})$为随机变量$X$的分布函数, 或称$X$服从$F(x)$分布, 记为$X \sim F(x)$.

(2) 性质

① $F(x)$是x的单调不减函数, 即对任意实数$x_1 < x_2$,有$F\left(x_1\right) \leqslant F\left(x_2\right)$;

② $F(x)$是x的右连续函数, 即对任意$x_0 \in \mathbf{R}$,有$\lim_{x \to x_0^*} F(x) = F\left(x_0 + 0\right) = F\left(x_0\right)$;

③ $F(-\infty) = \lim_{x \to -\infty} F(x) = 0, F(+\infty) = \lim_{x \to +\infty} F(x) = 1$.

满足以上三条性质的函数$F(x)$必是某个随机变量X的分布函数, 所以, 这三条性质也是判断函数$F(x)$是否为某一随机变量X 的分布函数的充要条件.

(3) 应用——求概率

$$\begin{array}{l}P\{X\leqslant a\}=F(a);\\P\{X<a\}=F(a-0);\\P\{X=a\}=F(a)-F(a-0).\end{array}\tag{12}$$

二、 一维离散型随机变量

1. 分布律

如果随机变量X只可能取有限个或可列个值$x_1,x_2,\cdots$,则称X为离散型随机

$$P\left\{X=x_i\right\}=p_i,i=1,2,\cdots\tag{13}$$

为X的分布列、分布律或概率分布, 记为$X\sim p_i$,概率分布常常用表格形式或矩阵形式表示, 即

$$\begin{array}{c|ccc}X&x_1&x_2&\cdots\\\hline P&p_1&p_2&\cdots\end{array}\text{或}X\sim\begin{pmatrix}x_1&x_2&\cdots\\p_1&p_2&\cdots\end{pmatrix}.\tag{14}$$

2. 性质

- 归一性：数列$\{p_i\}$是离散型随机变量的概率分布的充要条件: $p_i\geqslant 0(i=1,2,\cdots)$,且$\sum_i p_i=1$.
- 区间与点的概率：设离散型随机变量X的概率分布为$P\left\{X=x_i\right\}=p_i$,则$X$的分布函数

$$\begin{array}{c}F(x)=P\{X\leqslant x\}=\sum_{x_i\leqslant x}P\left\{X=x_i\right\},\\P\left\{X=x_i\right\}=P\left\{X\leqslant x_i\right\}-P\left\{X<x_i\right\}=F\left(x_i\right)-F\left(x_i-0\right),\end{array}\tag{15}$$

3. 应用——求概率

(1) 0—1分布$B(1,p)$

如果X的概率分布为$X\sim\begin{pmatrix}1&0\\p&1-p\end{pmatrix}$,即$P\{X=1\}=p,P\{X=0\}=1-p$,则称$X$服从参数为$p$的$\mathbf{0-1}$分布, 记为$X\sim B(1,p)(0<p<1)$.

(2) 二项分布$B(n,p)$

如果X的概率分布为$P\{X=k\}=C_n^kp^k(1-p)^{n-k}(k=0,1,\cdots,n;0<p<1)$,则称$X$服从参数为$(n,p)$的二项分布, 记为$X\sim B(n,p)$.

(3) 泊松分布$P(\lambda)$

如果X的概率分布为

$$P\{X=k\}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}(k=0,1,\cdots;\lambda>0),\tag{16}$$

则称X服从参数为λ的泊松分布, 记为$X\sim P(\lambda)$.

【注】某场合下，单位时间内源源不断的质点来流的个数，λ为强度（$E(X)=\lambda$）

当二项分布的n很大而p很小时，泊松分布可作为二项分布的近似，其中$\lambda=np$

(4) 几何分布$G(p)$

如果X的概率分布为

$$P\{X=k\}=(1-p)^{k-1}p(k=1,2,\cdots;0<p<1),\tag{17}$$

则称X服从参数为p的几何分布, 记为$X\sim G(p)$.

【注】设X表示伯努利试验中事件A首次发生所需做的试验次数, 则$X\sim G(p)$,其中$p=P(A)$.

几何分布称为离散型的等待分布

(4) 超几何分布$H(n,N,M)$

如果X的概率分布为$P\{X=k\}=\frac{\mathrm{C}_{\mathrm{M}}^k\mathrm{C}_{\mathrm{N}-k}^{n-k}}{\mathrm{C}_{\mathrm{N}}^n}(\max\{0,n-N+M\}\leqslant k\leqslant\min\{M,n\};M,N,n$为正整数且$M\leqslant N,n\leqslant N,k$为整数), 则称$X$服从参数为$(n,N,M)$的超几何分布, 记为$X\sim H(n,N,M)$.

三、 一维连续型随机变量

1. 概率密度

如果随机变量X的分布函数可以表示为

$$F(x)=\int_{-\infty}^x f(t)\mathrm{d}t(x\in\mathbf{R}),\tag{18}$$

其中$f(x)$是非负可积函数, 则称X为连续型随机变量, 称$f(x)$为X的概率密度函数, 简称概率密度, 记为$X\sim f(x)$.

$f(x)$为某一随机变量X的概率密度的充分必要条件:$f(x)\geqslant 0$,且$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$(由此可知, 改变$f(x)$有限个点的值,$f(x)$仍然是概率密度).

设X为连续型随机变量, $X\sim f(x)$,则对任意实数c,有$P\{X=c\}=0$;对实数轴上任一集合B,有

$$P\{X\in B\}=\int_B f(x)\mathrm{d}x,\tag{19}$$

特别地,

$$P\{a<X<b\}=P\{a\leqslant X<b\}=P\{a<X\leqslant b\}=P\{a\leqslant X\leqslant b\}=\int_a^b f(x)\mathrm{d}x=F(b)-F(a). \tag{20}$$

2. 性质

3. 应用——求概率

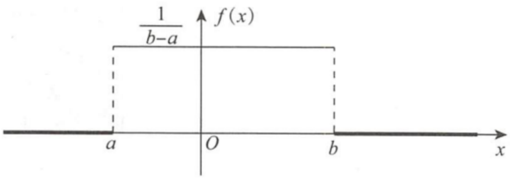
4. 三大分布

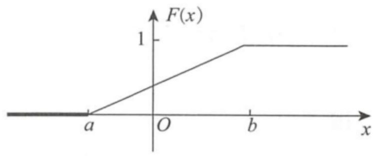
（1） 均匀分布

如果随机变量X的概率密度和分布函数分别为

$$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b,\\ 0, & \text{其他},\end{cases}\quad F(x)=\begin{cases}0, & x<a,\\ \frac{x-a}{b-a}, & a\le x\le b,\\ 1, & x\ge b\end{cases} \tag{21}$$

则称X在区间(a,b)上服从均匀分布, 记为$X\sim U(a,b)$.



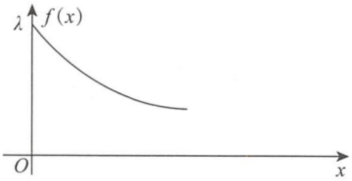


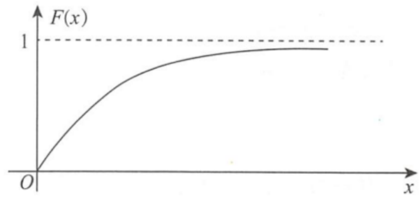
（2） 指数分布

如果X的概率密度和分布函数分别为

$$f(x)=\begin{cases}\lambda \mathrm{e}^{-\lambda x}, & x>0,\\ 0, & \text{其他},\end{cases}\quad F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0\end{cases}(\lambda>0), \tag{22}$$

则称X服从参数为λ的指数分布, 记为$X\sim E(\lambda)$.





指数分布称为连续型的等待分布，λ为失效率（$E(X)=\frac{1}{\lambda}$）

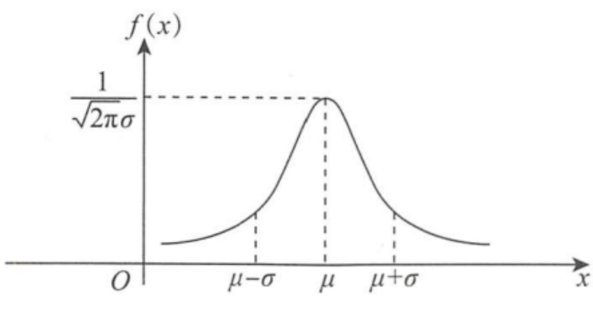
指数分布具有无记忆性（已工作s小时无误，再工作t小时无误的概率与只工作t小时无误的概率一样）

（3） 正态分布

如果X的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}(-\infty<x<+\infty), \tag{23}$$

其中$-\infty<\mu<+\infty,\sigma>0$,则称$X$服从参数为$\left(\mu,\sigma^2\right)$的正态分布或称$X$为正态变量, 记为$X\sim N\left(\mu,\sigma^2\right)$.此时$f(x)$的图形关于直线$x=\mu$对称, 即$f(\mu-x)=f(\mu+x)$,并在$x=\mu$处有唯一最大值



称$\mu=0,\sigma=1$时的正态分布$N(0,1)$为标准正态分布, 通常记标准正态分布的概率密度为$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}$，分布函数为$\phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{1}{2}t^2}dt$，显然$\varphi(x)$为偶函数，则

$$\Phi(0)=\frac{1}{2},\quad \Phi(-x)=1-\Phi(x). \tag{24}$$

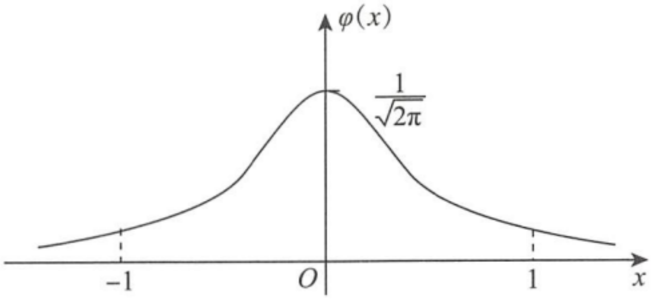


图 2-6

若$X \sim N(0,1), P\{X>\mu_a\}=\alpha$,则称$\mu_a$为标准正态分布的上侧α分位数(上α分位点).

如果$X \sim N\left(\mu, \sigma^2\right)$,则

$$\begin{gathered} F(x)=P\{X \leqslant x\}=\Phi\left(\frac{x-\mu}{\sigma}\right)(\text{标准化}), \\ F(\mu-x)+F(\mu+x)=1, \\ P\{a<X<b\}=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right), \\ aX+b \sim N\left(a\mu+b, a^2\sigma^2\right)(a \neq 0). \end{gathered} \tag{25}$$

四、 一维随机变量函数的分布

设X为随机变量, 函数$y=g(x)$,则以随机变量X作为自变量的函数$Y=g(X)$也是随机变量, 称为随机变量X的函数. 例如 : $Y=aX^2+bX+c, Y=|X-a|, Y=\begin{cases}X, & X \leqslant 1, \\ 1, & X>1,\end{cases}$等等.

1. 离散型→离散型

设X为离散型随机变量, 其概率分布为$p_i=P\{X=x_i\}(i=1,2,\cdots)$,则$X$的函数$Y=g(X)$也是离散型随机变量, 其概率分布为$P\{Y=g(x_i)\}=p_i(i=1,2,\cdots)$,即

$$Y \sim \begin{pmatrix} g(x_1) & g(x_2) & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}. \tag{26}$$

如果有若干个$g(x_i)$值相同, 则合并诸项为一项$g(x_k)$,并将相应概率相加作为Y取$g(x_k)$值的概率.

2. 连续型→连续型（或混合型）

设X为连续型随机变量, 其分布函数、概率密度分别为$F_X(x)$与$f_X(x)$,随机变量$Y=g(X)$是X的函数, 则Y的分布函数和概率密度可用分布函数法求得.

$$F_Y(y)=P\{Y \leqslant y\}=P\{g(X) \leqslant y\}=\int_{\bar{g}(x) \leqslant y} f_X(x)\mathrm{d}x. \tag{27}$$

如果$F_Y(y)$连续, 且除有限个点外, $F_Y'(y)$存在且连续, 则Y的概率密度$f_Y(y)=F_Y'(y)$.

第3讲 多维随机变量及其分布

一、 二维（n维）随机变量

1. 概念

如果$X_1, X_2, \cdots, X_n$是定义在同一个样本空间Ω上的n个随机变量, 则称$(X_1, X_2, \cdots, X_n)$为n维随机变量或n维随机向量 , $X_i(i=1,2,\cdots,n)$称为第i个分量.
当$n=2$时, 记(X,Y)为二维随机变量或二维随机向量.

2. 联合分布函数

（1） 概念

对任意的n个实数$x_1, x_2, \cdots, x_n$,称n元函数

$$F(x_1, x_2, \cdots, x_n)=P\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\} \tag{28}$$

为n维随机变量$(X_1, X_2, \cdots, X_n)$的联合分布函数.

当$n=2$时, 则对任意的实数x, y,称二元函数

$$F(x,y)=P\{X \leqslant x, Y \leqslant y\} \tag{29}$$

为二维随机变量(X,Y)的联合分布函数, 简称分布函数, 记为$(X,Y) \sim F(x,y)$.

（2） 性质

① **单调性** $F(x,y)$是x, y的单调不减函数:

对任意固定的y,当$x_1<x_2$时, $F(x_1,y) \leqslant F(x_2,y)$;

对任意固定的x,当$y_1<y_2$时, $F(x,y_1) \leqslant F(x,y_2)$.

② **右连续性** $F(x,y)$是x,y的右连续函数:

$$\begin{aligned}&\lim_{x\to x^+}F(x,y)=F(x_0+0,y)=F(x_0,y);\\&\lim_{y\to y^+}F(x,y)=F(x,y_0+0)=F(x,y_0).\end{aligned}\tag{30}$$

③ **有界性** $F(-\infty,y)=F(x,-\infty)=F(-\infty,-\infty)=0,F(+\infty,+\infty)=1.$

④ **非负性** 对任意$x_1<x_2,y_1<y_2$,有

$$P\{x_1<X\leqslant x_2,y_1<Y\leqslant y_2\}=F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\geqslant 0.\tag{31}$$

3. 边缘分布函数

设二维随机变量(X,Y)的联合分布函数为$F(x,y)$,随机变量X与Y的分布函数$F_X(x)$与$F_Y(y)$分别称为(X,Y)关于X和关于Y的边缘分布函数. 由概率性质得

$$\begin{aligned}F_X(x)&=P\{X\leqslant x\}=P\{X\leqslant x,Y\leqslant +\infty\}\\&=\lim_{y\to+\infty}P\{X\leqslant x,Y\leqslant y\}\\&=\lim_{y\to+\infty}F(x,y)=F(x,+\infty).\end{aligned}\tag{32}$$

同理，有$F_Y(y)=F(+\infty,y)$

二、 二维离散型随机变量

如果二维随机变量(X,Y)只能取有限对值或可列对值$(x_1,y_1),(x_1,y_2),\cdots,(x_n,y_n),\cdots$,则称$(X,Y)$为二维离散型随机变量.

称

$$p_{ij}=P\{X=x_i,Y=y_j\},i,j=1,2,\cdots\tag{33}$$

为(X,Y)的分布律或随机变量X和Y的联合分布律,记为$(X,Y)\sim p_{ij}$.联合分布律常用表格形式表示.

$X \backslash Y$	y_1	$\cdots$	y_j	$\cdots$	$P\{X=x_i\}$
x_1	p_{11}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$P\{Y=y_j\}$	$p_{\cdot 1}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

数列$\{p_{ij}\},i,j=1,2,\cdots$是某一二维离散型随机变量的概率分布的充分必要条件为

$$p_{ij}\geqslant 0,\quad \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}p_{ij}=1.\tag{34}$$

1. 联合分布律

设(X,Y)的概率分布为$p_{ij},i,j=1,2,\cdots$,则(X,Y)的联合分布函数为

$$F(x,y)=P\{X\leqslant x,Y\leqslant y\}=\sum_{x_i\leqslant x y_j\leqslant y}p_{ij}.\tag{35}$$

设G是平面上的某个区域,则

$$P\{(X,Y)\in G\}=\sum_{(x_i,y_j)\in G}p_{ij}.\tag{36}$$

2. 边缘分布律

X,Y的边缘分布分别为

$$\begin{aligned}p_{i\cdot}&=P\{X=x_i\}=\sum_{j=1}^{\infty}P\{X=x_i,Y=y_j\}=\sum_{j=1}^{\infty}p_{ij}(i=1,2,\cdots);\\p_{\cdot j}&=P\{Y=y_j\}=\sum_{i=1}^{\infty}P\{X=x_i,Y=y_j\}=\sum_{i=1}^{\infty}p_{ij}(j=1,2,\cdots).\end{aligned}\tag{37}$$

3. 条件分布律

如果$(X,Y)\sim p_{ij}(i,j=1,2,\cdots)$,对固定的$j$,如果$p_{\cdot j}=P\{Y=y_j\}>0$,则称

$$P\{X=x_i\mid Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}}(i=1,2,\cdots)\tag{38}$$

为X在“$Y=y_j$”条件下的条件分布.

同理, 对固定的i,如果$p_{i\cdot}>0$,可定义Y在“$X=x_i$”条件下的条件分布

$$P\{Y=y_j\mid X=x_i\}=\frac{p_{ij}}{p_{i\cdot}}(j=1,2,\cdots).\tag{39}$$

三、 二维连续型随机变量

如果二维随机变量(X, Y)的联合分布函数$F(x, y)$可以表示为

$$F(x, y)=\int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v) \mathrm{d} u \mathrm{~d} v,(x, y) \in \mathbf{R}^{2}, \tag{40}$$

其中$f(x, y)$是非负可积函数, 则称(X, Y)为二维连续型随机变量, 称$f(x, y)$为(X, Y)的概率密度, 记为$(X, Y) \sim f(x, y)$.

二元函数$f(x, y)$是概率密度的充分必要条件为

$$f(x, y) \geqslant 0, \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \mathrm{d} x \mathrm{~d} y=1. \tag{41}$$

1. 联合概率密度

设(X, Y)的联合分布函数为$F(x, y)$,概率密度为$f(x, y)$,则

① $F(x, y)$为(x, y)的二元连续函数, 且

$$F(x, y)=P\{X \leqslant x, Y \leqslant y\}=\int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v) \mathrm{d} u \mathrm{~d} v; \tag{42}$$

② 设G为平面上某个区域,则【⭐⭐⭐】

$$P\{(X, Y) \in G\}=\iint_{G} f(x, y) \mathrm{d} x \mathrm{~d} y; \tag{43}$$

③ 若$f(x, y)$在点(x, y)处连续，则$\frac{\partial^{2} F(x, y)}{\partial x \partial y}=f(x, y)$

④ 若$F(x, y)$连续且可导, 则(X, Y)是连续型随机变量, 且$\frac{\partial^{2} F(x, y)}{\partial x \partial y}$是它的概率密度.

2. 边缘概率密度

设$(X, Y) \sim f(x, y)$,则X的边缘分布函数为

$$F_{X}(x)=F(x,+\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(u, v) \mathrm{d} v\right] \mathrm{d} u, \tag{44}$$

所以X是连续型随机变量, 其概率密度

$$f_{X}(x)=\int_{-\infty}^{+\infty} f(x, y) \mathrm{d} y, \tag{45}$$

称$f_{X}(x)$为(X, Y)关于X的边缘概率密度.

同理, Y 也是连续型随机变量, 其概率密度为$f_{Y}(y)=\int_{-\infty}^{+\infty} f(x, y) \mathrm{d} x$.

3. 条件概率密度

设$(X, Y) \sim f(x, y)$,边缘概率密度$f_{X}(x)>0$,则称

$$f_{Y \mid X}(y \mid x)=\frac{f(x, y)}{f_{X}(x)} \tag{46}$$

为Y在 “$X=x$”条件下的条件概率密度.

同理可定义X在“$Y=y$”条件下的条件概率密度

$$f_{Y \mid X}(y \mid x)=\frac{f(x, y)}{f_{X}(x)} \tag{47}$$

若$f_{X}(x)>0, f_{Y}(y)>0$,则有概率密度乘法公式

$$f(x, y)=f_{X}(x) f_{Y \mid X}(y \mid x)=f_{Y}(y) f_{X \mid Y}(x \mid y). \tag{48}$$

4. 二维均匀分布【⭐】

称(X, Y)在平面有界区域D上服从均匀分布, 如果(X, Y)的概率密度为

$$f(x, y)=\begin{cases}\frac{1}{S_{D}}, & (x, y) \in D, \\ 0, & \text{其他}, \end{cases} \tag{49}$$

其中S_{D}为区域D的面积.

5. 二维正态分布

如果(X, Y)的概率密度为

$$f(x, y)=\frac{1}{2 \pi \sigma_{1} \sigma_{2} \sqrt{1-\rho^{2}}} \exp \left\{-\frac{1}{2\left(1-\rho^{2}\right)}\left[\left(\frac{x+\mu_{1}}{\sigma_{1}}\right)^{2}-2 \rho\left(\frac{x-\mu_{1}}{\sigma_{1}}\right)\left(\frac{y-\mu_{2}}{\sigma_{2}}\right)+\left(\frac{y-\mu_{2}}{\sigma_{2}}\right)^{2}\right]\right\}, \tag{50}$$

其中$\mu_{1} \in \mathbf{R}, \mu_{2} \in \mathbf{R}, \sigma_{1}>0, \sigma_{2}>0,-1<\rho<1$,则称$(X, Y)$服众参数为$\mu_{1}, \mu_{2}, \sigma_{1}^{2}, \sigma_{2}^{2}, \rho$的二维正态分布, 记为$(X, Y) \sim N\left(\mu_{1}, \mu_{2} ; \sigma_{1}^{2}, \sigma_{2}^{2} ; \rho\right)$.此时有

① $X \sim N\left(\mu_{1}, \sigma_{1}^{2}\right), Y \sim N\left(\mu_{2}, \sigma_{2}^{2}\right)$, ρ为X与Y的相关系数, 即

$$\rho=\frac{\operatorname{Cov}(X, Y)}{\sqrt{D X} \sqrt{D Y}}=\frac{\operatorname{Cov}(X, Y)}{\sigma_{1} \sigma_{2}}. \tag{51}$$

② X,Y的条件分布都是正态分布.

③ $aX+bY(a\neq 0$或$b\neq 0)$服从正态分布.

④ X与Y相互独立的充分必要条件是X与Y不相关, 即$\rho=0$.

四、 独立性

1. 概念

设二维随机变量(X,Y)的联合分布函数为$F(x,y)$，边缘分布函数分别为$F_X(x),F_Y(y)$如果对任意实数x,y都有

$$F(x,y)=F_X(x)\cdot F_Y(y)\ (\text{即事件}\ \{X\leqslant x\}\ \text{与}\ \{Y\leqslant y\}\ \text{相互独立}\ ), \tag{52}$$

则称X和Y相互独立，否则不相互独立。

n维随机变量的独立性同理

2. 相互独立的充要条件

(1) $F(x,y)=F_X(x)\cdot F_Y(y)$

n个随机变量$X_1,X_2,\cdots,X_n$相互独立

$\Leftrightarrow$对任意的n个实数$x_i(i=1,2,\cdots,n)$, n个事件$\{X_1\leqslant x_1\},\{X_2\leqslant x_2\},\cdots,\{X_n\leqslant x_n\}$相互独立.

(2) $p_{ij}=p_i\cdot p_j$

① 设(X,Y)为二维离散型随机变量, 则X与Y相互独立

$\Leftrightarrow$联合分布等于边缘分布相乘, 即

$$P\{X=x_i,Y=y_j\}=P\{X=x_i\}\cdot P\{Y=y_j\}(i,j=1,2,\cdots). \tag{53}$$

② n个离散型随机变量$X_1,X_2,\cdots,X_n$相互独立

$\Leftrightarrow$对任意的$x_i\in D_i=\{X_i$的一切可能值$\}(i=1,2,\cdots,n)$,有

$$P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n}P\{X_i=x_i\}. \tag{54}$$

(3) $f(x,y)=f_X(x)f_Y(y)$

① 设(X,Y)为二维连续型随机变量, 则X与Y相互独立

$\Leftrightarrow$概率密度等于边缘概率密度相乘, 即

$$f(x,y)=f_X(x)\cdot f_Y(y). \tag{55}$$

② 设$(X_1,X_2,\cdots,X_n)$为n维连续型随机变量, 则$X_1,X_2,\cdots,X_n$相互独立

$\Leftrightarrow$概率密度等于边缘概率密度相乘, 即

$$f(x_1,x_2,\cdots,x_n)=f_1(x_1)\cdot f_2(x_2)\cdot\cdots\cdot f_n(x_n), \tag{56}$$

其中$f_i(x_i)$为X_i的边缘概率密度.

3. 相互独立的性质

(1) 设$X_1,X_2,\cdots,X_n$相互独立, 则其中任意$k(2\leqslant k\leqslant n)$个随机变量也相互独立.

(2) ① 设(X,Y)为二维离散型随机变量, X与Y独立, 则条件分布等于边缘分布:

$$\begin{aligned}P\{X=x_i\mid Y=y_j\}=P\{X=x_i\}\,(P\{Y=y_j\}>0),\\ P\{Y=y_j\mid X=x_i\}=P\{Y=y_j\}\,(P\{X=x_i\}>0).\end{aligned} \tag{57}$$

② 设(X,Y)为二维连续型随机变量, X与Y独立, 则条件概率密度等于边缘概率密度:

$$\begin{aligned}f_{X|Y}(x\mid y)=\tfrac{f(x,y)}{f_Y(y)}=f_X(x)\,(f_Y(y)>0),\\ f_{Y|X}(y\mid x)=\tfrac{f(x,y)}{f_X(x)}=f_Y(y)\,(f_X(x)>0).\end{aligned} \tag{58}$$

(3) 若$X_1,X_2,\cdots,X_n$相互独立, $g_1(x),g_2(x),\cdots,g_n(x)$为一元连续函数, 则$g_1(X_1),g_2(X_2),\cdots,g_n(X_n)$相互独立.

一般地, 若$X_{11},\cdots,X_{1t_1},X_{21},\cdots,X_{2t_t},\cdots,X_{n1},\cdots,X_{nt}$相互独立, g_i是$t_i(i=1,2,\cdots,n)$元连续函数, 则$g_1(X_{11},\cdots,X_{1t_1}),g_2(X_{21},\cdots,X_{2t_t}),\cdots,g_n(X_{n1},\cdots,X_{nt_+})$也相互独立.

五、 函数的分布

1. （离散型，离散型）$\Rightarrow$离散型

若$Z=g(X,Y)$，先确定Z的值，而后求其相对应的概率。一般使用矩阵法

2. （连续型，连续型）$\Rightarrow$连续型

(1) 分布函数法

若$(X,Y)\sim f(x,y)$，则$Z=g(X,Y)$的分布函数

$$F(z)=P\{g(X,Y)\leq z\}=\iint_{g(x,y)\leq z}f(x,y)dxdy \tag{59}$$

(2) 卷积公式【⭐⭐⭐】

积谁不换谁，换完求偏导（加绝对值）

设$(X,Y)\sim f(x,y)$，则$Z=X+Y$的概率密度

$$f(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx=\int_{-\infty}^{+\infty}f(z-y,y)dy \tag{60}$$

设$(X,Y)\sim f(x,y)$，则$Z=X-Y$的概率密度

$$f(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx=\int_{-\infty}^{+\infty}f(z-y,y)dy \tag{61}$$

设$(X,Y)\sim f(x,y)$，则$Z=XY$的概率密度

$$f(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f(x,\frac{z}{x})dx=\int_{-\infty}^{+\infty}\frac{1}{|y|}f(\frac{z}{y},y)dy \tag{62}$$

设$(X,Y)\sim f(x,y)$，则$Z=\frac{X}{Y}$的概率密度

$$f(z)=\int_{-\infty}^{+\infty}|y|f(yz,y)dy \tag{63}$$

3. （离散型，连续型）$\Rightarrow$连续型

(1) 全概率公式

将事件对离散型的一切可能值进行全集分解，而后应用全概率公式求得Z的分布

4. 相互独立随机变量函数的分布

设$(X,Y)\sim F(x,y)$,则$Z=\max\{X,Y\}$的分布函数为

$$F_{\max}(z)=P\{\max\{X,Y\}\leqslant z\}=P\{X\leqslant z,Y\leqslant z\}=F(z,z). \tag{64}$$

当X与Y独立时,

$$F_{\max}(z)=F_X(z)\cdot F_Y(z). \tag{65}$$

设$(X,Y)\sim F(x,y)$,则$Z=\min\{X,Y\}$的分布函数为

$$\begin{aligned}F_{\min}(z)&=P\{\min\{X,Y\}\leqslant z\}=P\{\{X\leqslant z\}\cup\{Y\leqslant z\}\}\\&=P\{X\leqslant z\}+P\{Y\leqslant z\}-P\{X\leqslant z,Y\leqslant z\}\\&=F_X(z)+F_Y(z)-F(z,z).\end{aligned} \tag{66}$$

当X与Y独立时,

$$\begin{aligned}F_{\min}(z)&=F_X(z)+F_Y(z)-F_X(z)F_Y(z)\\&=1-[1-F_X(z)]\,[1-F_Y(z)].\end{aligned} \tag{67}$$

有些相互独立且服从同类型分布的随机变量, 其和的分布也是同类型的, 它们是二项分布、泊松分布、正态分布与χ^2分布.

设随机变量X与Y相互独立, 则:

若$X\sim B(n,p),Y\sim B(m,p)$,则$X+Y\sim B(n+m,p)$(注意$p$相同);

若$X\sim P(\lambda_1),Y\sim P(\lambda_2)$,则$X+Y\sim P(\lambda_1+\lambda_2)$;

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

若$X\sim\chi^2(n),Y\sim\chi^2(m)$,则$X+Y\sim\chi^2(n+m)$.

第4讲 随机变量的数字特征

一、 一维随机变量的数字特征

1. 数学期望

(1) 概念

$X\sim p_i$

$$\begin{aligned}EX&=\sum_{i=1}^{\infty}x_ip_i\\E[g(X)]&=\sum_{i=1}^{\infty}g(x_i)p_i\end{aligned} \tag{68}$$

$X\sim f(x)$

$$\begin{aligned}EX&=\int_{-\infty}^{+\infty}xf(x)dx\\E[g(X)]&=\int_{-\infty}^{+\infty}g(x)f(x)dx\end{aligned} \tag{69}$$

(2) 性质

① 对任意常数a_i和随机变量$X_i(i=1,2,\cdots,n)$有

$$E\left(\sum_{i=1}^{n} a_i X_i\right)=\sum_{i=1}^{n} a_i E X_i \tag{70}$$

特别地,

$$Ec=c, \quad E(aX+c)=aEX+c, \quad E(X \pm Y)=EX \pm EY. \tag{71}$$

② 设X与Y相互独立, 则

$$E(XY)=EX \cdot EY, \quad E\left[g_1(X) \cdot g_2(Y)\right]=E\left[g_1(X)\right] \cdot E\left[g_2(Y)\right]. \tag{72}$$

一般地, 设$X_1, X_2, \cdots, X_n$相互独立, 则

$$E\left(\prod_{i=1}^{n} X_i\right)=\prod_{i=1}^{n} E X_i, \quad E\left[\prod_{i=1}^{n} g_i\left(X_i\right)\right]=\prod_{i=1}^{n} E\left[g_i\left(X_i\right)\right]. \tag{73}$$

2. 方差、标准差

(1) 概念

设X是随机变量, 如果$E\left[(X-EX)^2\right]$存在, 则称$E\left[(X-EX)^2\right]$为X的方差, 记为DX,即

$$DX=E\left[(X-EX)^2\right]=E\left(X^2\right)-(EX)^2. \tag{74}$$

称$\sqrt{DX}$为X的标准差或均方差, 记为$\sigma(X)$,称随机变量$X^*=\frac{X-EX}{\sqrt{DX}}$为$X$的标准化随机变量, 此时$EX^*=0, DX^*=1$.

(2) 性质

① $DX \geqslant 0, E\left(X^2\right)=DX+(EX)^2 \geqslant (EX)^2$.

② $Dc=0$.

③ $D(aX+b)=a^2 DX$.

④ $D(X \pm Y)=DX+DY \pm 2\operatorname{Cov}(X, Y)$.

⑤ 如果X与Y相互独立, 则

$$D(aX+bY)=a^2 DX+b^2 DY. \tag{75}$$

一般地, 如果$X_1, X_2, \cdots, X_n$相互独立, $g_i(x)$为x的连续函数, 则

$$\begin{gathered}D\left(\sum_{i=1}^{n} a_i X_i\right)=\sum_{i=1}^{n} a_i^2 DX_i, \\ D\left[\sum_{i=1}^{n} g_i\left(X_i\right)\right]=\sum_{i=1}^{n} D\left[g_i\left(X_i\right)\right].\end{gathered} \tag{76}$$

3. 切比雪夫不等式

如果随机变量X的期望EX和方差DX存在, 则对任意$\varepsilon>0$,有

$$P\{|X-EX| \geqslant \varepsilon\} \leqslant \frac{DX}{\varepsilon^2} \text{ 或 } P\{|X-EX|<\varepsilon\} \geqslant 1-\frac{DX}{\varepsilon^2}. \tag{77}$$

我们将常用分布的期望和方差列表如下.

分布	分布列p_k或概率密度$f(x)$	期望	方差
$0-1$分布	$P\{X=k\}=p^k(1-p)^{1-k}, k=0,1$	p	$p(1-p)$
二项分布 $B(n,p)$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k}, k=0,1,\cdots,n$	np	$np(1-p)$
泊松分布 $P(\lambda)$ $P(\lambda)$	$P\{X=k\}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, k=0,1,\cdots$	λ	λ
几何分布 $G(p)$	$P\{X=k\}=(1-p)^{k-1}p, k=1,2,\cdots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
正态分布 $N\left(\mu,\sigma^2\right)$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, -\infty<x<+\infty$	μ	σ^2
均匀分布 $U(a,b)$	$f(x)=\frac{1}{b-a}, a<x<b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布 $E(\lambda)$	$f(x)=\lambda\mathrm{e}^{-\lambda x}, x>0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

(78)

注意: 表中仅列出各分布概率密度的非零区域.

二、 二维随机变量的数字特征

1. 数学期望

设X, Y为随机变量, $g(X, Y)$为X, Y的函数(g是连续函数).

(1) 如果(X, Y)为离散型随机变量, 其联合分布为

$$p_{ij}=P\left\{X=x_i, Y=y_j\right\}(i, j=1,2,\cdots), \tag{79}$$

若级数$\sum_i \sum_j g\left(x_i, y_j\right) p_{ij}$绝对收敛, 则定义

$$E[g(X,Y)]=\sum_i\sum_j g\left(x_i,y_j\right)p_{ij}. \tag{80}$$

（2）如果(X,Y)为连续型随机变量, 其概率密度为$f(x,y)$,若积分$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\ \mathrm{d}y$绝对收敛, 则定义

$$E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\ \mathrm{d}y. \tag{81}$$

2. 协方差与相关系数

（1） 概念

如果随机变量X与Y的方差存在且$DX>0,DY>0$,则称$E[(X-EX)(Y-EY)]$为随机变量X与Y的协方差, 并记为$\operatorname{Cov}(X,Y)$,即

$$\operatorname{Cov}(X,Y)=E[(X-EX)(Y-EY)]=E(XY)-EX\cdot EY, \tag{82}$$

其中

$$E(XY)=\begin{cases}\sum_i\sum_j x_iy_jP\left\{X=x_i,Y=y_j\right\} \text{(离散型)},\\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)\mathrm{d}x\ \mathrm{d}y \text{(连续型)}.\end{cases} \tag{83}$$

称$\rho_{XY}=\frac{\operatorname{Cov}(X,Y)}{\sqrt{DX}\sqrt{DY}}$为随机变量$X$与$Y$的相关系数. 如果$\rho_{XY}=0$,则称$X$与$Y$不相关; 如果$\rho_{XY}\neq 0$,则称$X$与$Y$相关.

相关系数ρ_{XY}描述随机变量X与Y之间的线性相依性$\rho_{XY}=0$表示X与Y 之间不存在线性关系，但它们之间还可能存在某种非线性关系.

（2） 性质

① 对称性 $\operatorname{Cov}(X,Y)=\operatorname{Cov}(Y,X),\quad \rho_{XY}=\rho_{XX},$

$\operatorname{Cov}(X,X)=DX,\quad \rho_{XX}=1.$

② 线性性 $\operatorname{Cov}(X,c)=0,\quad \operatorname{Cov}(aX+b,Y)=a\operatorname{Cov}(X,Y),$

$\operatorname{Cov}\left(X_1+X_2,Y\right)=\operatorname{Cov}\left(X_1,Y\right)+\operatorname{Cov}\left(X_2,Y\right).$

一般地

$$\operatorname{Cov}\left(\sum_{i=1}^n a_iX_i,Y\right)=\sum_{i=1}^n a_i\operatorname{Cov}\left(X_i,Y\right). \tag{84}$$

③ 相关系数有界性 $|\rho_{XY}|\leqslant 1.$

④ 线性关系下的相关系数

如果$Y=aX+b$,则$\rho_{XY}=\begin{cases}1, & a>0,\\ -1, & a<0.\end{cases}$

三、 独立性与相关性的判定

1. 用分布判定独立性

随机变量X与Y相互独立, 意指对任意实数x,y,事件$\{X\leqslant x\}$与$\{Y\leqslant y\}$相互独立, 即(X,Y)的分布等于边缘分布相乘$:F(x,y)=F_X(x)\cdot F_Y(y).$

若(X,Y)是连续型的, 则X与Y独立的充要条件是

$$f(x,y)=f_X(x)\cdot f_Y(y) \tag{85}$$

若(X,Y)是离散型的, 则X与Y独立的充要条件是

$$P\left\{X=x_i,Y=y_j\right\}=P\left\{X=x_i\right\}\cdot P\left\{Y=y_j\right\}. \tag{86}$$

2. 用数字特征判定相关性

随机变量X与Y不相关, 意指X与Y之间不存在线性相依性, 即$\rho_{XY}=0$,其充要条件是

$$\rho_{XY}=0\Leftrightarrow \operatorname{Cov}(X,Y)=0\Leftrightarrow E(XY)=EX\cdot EY\Leftrightarrow D(X\pm Y)=DX+DY. \tag{87}$$

3. 几个重要结论

① 如果X与Y独立, 则X,Y不相关, 反之不然;

② 如果(X,Y)服从二维正态分布,则X,Y独立$\Leftrightarrow X,Y$不相关;

③ 由①知, 如果X与Y相关, 则X,Y不独立.

综上所述, 我们在讨论随机变量X与Y的相关性、独立性时, 总是先计算$\operatorname{Cov}(X,Y)$,而后按下列程序进行判断或再计算:

$$\operatorname{Cov}(X,Y)=E(XY)-EXEY\begin{cases}\neq 0\Leftrightarrow X\text{ 与 }Y\text{ 相关 }\Rightarrow X\text{ 与 }Y\text{ 不独立},\\ =0\Leftrightarrow X\text{ 与 }Y\text{ 不相关, 通过分布推断 }\begin{cases}X,Y\text{ 独立},\\ X,Y\text{ 不独立}.\end{cases}\end{cases} \tag{88}$$

【注】上述讨论均假设方差存在且不为零.

第6讲 数理统计

一、 总体与样本

1. 总体定义

研究对象的全体称为总体, 组成总体的每一个元素称为个体

2. 样本

（1） 定义

n个相互独立且与总体X具有相同概率分布的随机变量$X_1, X_2, \cdots, X_n$所组成的整体$(X_1, X_2, \cdots, X_n)$称为来自总体X,容量为n的一个简单随机样本, 简称样本. 一次抽样结果的n个具体数值$(x_1, x_2, \cdots, x_n)$称为样本$X_1, X_2, \cdots, X_n$的一个观测值 (或样本值).

（2） 分布

对于容量为n的样本$X_1, X_2, \cdots, X_n$,有如下定理:

假设总体X的分布函数为$F(x)$(概率密度为$f(x)$,或概率分布为$p_i = P\{X = x_i\}$),则$(X_1, X_2, \cdots, X_n)$的分布函数为

$$F(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F(x_i) \tag{89}$$

相应地:

① 对于离散型随机变量的样本$X_1, X_2, \cdots, X_n$,联合分布为

$$P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} = \prod_{i=1}^{n} P\{X_i = x_i\}; \tag{90}$$

② 对于连续型随机变量的样本$X_1, X_2, \cdots, X_n$,联合概率密度为

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i). \tag{91}$$

二、 统计量及其分布

1. 统计量

（1） 定义

设$X_1, X_2, \cdots, X_n$为来自总体X的一个样本, $g(x_1, x_2, \cdots, x_n)$为n元函数, 如果g中不含任何未知参数, 则称$g(X_1, X_2, \cdots, X_n)$为样本$X_1, X_2, \cdots, X_n$的一个统计量. 若$(x_1, x_2, \cdots, x_n)$为样本值, 则称$g(x_1, x_2, \cdots, x_n)$为$g(X_1, X_2, \cdots, X_n)$的观测值.

（2） 常用统计量

① 样本均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i; \tag{92}$$

② 样本方差

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 (= \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)); \tag{93}$$

③ 样本标准差

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}; \tag{94}$$

④ 样本k阶（原点）矩

$$A_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k (k = 1, 2, \cdots); \tag{95}$$

⑤ 样本k阶中心距

$$B_k = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^k (k = 2, 3, \cdots). \tag{96}$$

⑥ 顺序统计量

将样本$X_1, X_2, \cdots, X_n$的n个观测量按其取值从小到大的顺序排列, 得

$$X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}. \tag{97}$$

随机变量$X_{(k)}(k = 1, 2, \cdots, n)$称作第$k$顺序统计量, 其中$X_{(1)}$是最小顺序统计量, 而$X_{(n)}$是最大顺序统计量:

$$X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}, \quad X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}. \tag{98}$$

⑦ 性质

设总体X的期望$EX = \mu$,方差$DX = \sigma^2$, $X_1, X_2, \cdots, X_n$是取自总体X,容量为n的一个样本, $\bar{X}, S^2$分别为样本的均值和方差, 则

$$EX_i = \mu, \quad DX_i = \sigma^2 (i = 1, 2, \cdots, n), \quad E\bar{X} = EX = \mu, \quad D\bar{X} = \frac{1}{n} DX = \frac{\sigma^2}{n}, \quad E\left(S^2\right) = DX = \sigma^2. \tag{99}$$

2. 三大分布

(1) χ^2分布

若随机变量$X_1, X_2, \cdots, X_n$相互独立, 且都服从标准正态分布, 则随机变量$X=\sum_{i=1}^n X_i^2$服从自由度为n的χ^2分布, 记为$X \sim \chi^2(n)$.其概率密度$f(x)$的图形如图所示. 特别地, $X_i^2 \sim \chi^2(1)$.对给定的$\alpha(0<\alpha<1)$,称满足

$$P\left\{\chi^2>\chi_\alpha^2(n)\right\}=\int_{\chi_\alpha^2(n)}^{+\infty} f(x) \mathrm{d} x=\alpha \tag{100}$$

的$\chi_\alpha^2(n)$为$\chi^2(n)$分布的上α分位点.对于不同的α, n, $\chi^2(n)$分布上α分位点可通过查表求得.

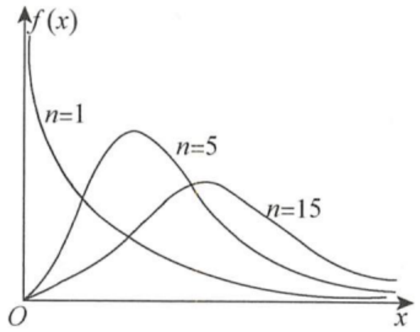


图 6-1

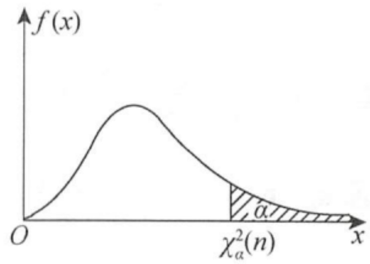


图 6-2

- χ^2分布的性质.
  ① 若$X_1 \sim \chi^2\left(n_1\right), X_2 \sim \chi^2\left(n_2\right)$, X_1与X_2相互独立, 则$X_1+X_2 \sim \chi^2\left(n_1+n_2\right)$.
  一般地, 若$X_i \sim \chi^2\left(n_i\right)(i=1,2, \cdots, m)$, $X_1, X_2, \cdots, X_m$相互独立, 则$\sum_{i=1}^m X_i \sim \chi^2\left(\sum_{i=1}^m n_i\right)$.
  ② 若$X \sim \chi^2(n)$,则$E X=n, D X=2 n$.

(2) t分布

设随机变量$X \sim N(0,1), Y \sim \chi^2(n)$, X与Y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, 记为$t \sim t(n)$.

t分布的概率密度$f(x)$的图形关于$x=0$对称, 因此$E t=0(n \geqslant 2)$.

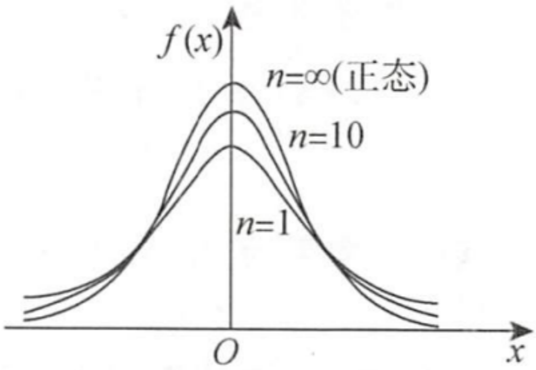


图 6-3

- t分布的性质.
  由t分布的概率密度$f(x)$图形的对称性知

$$P\left\{t>-t_\alpha(n)\right\}=P\left\{t>t_{1-\alpha}(n)\right\}, \tag{101}$$

  故$t_{1-\alpha}(n)=-t_\alpha(n)$.

(3) F分布

设随机变量$X \sim \chi^2\left(n_1\right), Y \sim \chi^2\left(n_2\right)$,且$X$与$Y$相互独立, 则$F=\frac{X / n_1}{Y / n_2}$服从自由度为$\left(n_1, n_2\right)$的$F$分布, 记为$F \sim F\left(n_1, n_2\right)$,其中$n_1$称为第一自由度, n_2称为第二自由度. F分布的概率密度$f(x)$的图形如图

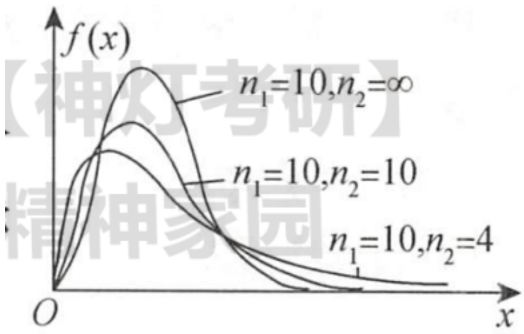


图 6-4

- F 分布的性质.
  ① 若$F \sim F\left(n_1, n_2\right)$,则$\frac{1}{F} \sim F\left(n_2, n_1\right)$.
  ② $F_{1-\alpha}\left(n_1, n_2\right)=\frac{1}{F_\alpha\left(n_2, n_1\right)}$.

3. 正态总体下的常用结论【⭐⭐⭐】

设$X_1, X_2, \cdots, X_n$是来自正态总体$N\left(\mu, \sigma^2\right)$的一个样本,$\bar{X}, S^2$分别是样本的均值和方差, 则

① $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$,即$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$;

② $\frac{1}{\sigma^2} \sum_{i=1}^n\left(X_i-\mu\right)^2 \sim \chi^2(n)$

③ $\frac{(n-1) S^2}{\sigma^2}=\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$($\mu$未知时, 在②中用$\bar{X}$替代$\mu$);

④ $\bar{X}$与S^2相互独立, $\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1)$($\sigma$未知时, 在①中用$S$替代$\sigma$).进一步有

$$\frac{n(\bar{X}-\mu)^2}{S^2} \sim F(1, n-1) \tag{102}$$

三、 参数的点估计

1. 概念

2. 方法

(1) 矩估计法

设总体X分布中有k个未知参数$\theta_1, \theta_2, \cdots$ 令样本矩=总体矩, 即

$$\frac{1}{n} \sum_{i=1}^n X_i^l=E(X^l)(l=1,2 \ldots, k), \tag{103}$$

这是包含k个未知参数$\theta_1, \theta_2, \cdots, \theta_k$的$k$个联立方程组 (称为矩法方程), 由此解得答案

$$\hat{\theta}_l=\hat{\theta}_l\left(X_1, X_2, \cdots, X_n\right)(l=1,2, \cdots, k), \tag{104}$$

由此解$\theta_1, \theta_2, \cdots, \theta_k$的矩估计值$\hat{\theta}_l\left(x_1, x_2, \cdots, x_n\right)$

(2) 最大似然估计法

对未知参数 θ 进行估计时, 在该参数可能的取值范围 I 内选取, 用使"样本获此观测值 $x_1, x_2, \cdots, x_n$ "的概率最大的参数值 $\hat{\theta}$ 作为 θ 的估计, 这样选定的 $\hat{\theta}$ 最有利于 $x_1, x_2, \cdots, x_n$ 的出现.

则令样本的似然函数

$$L(\theta)=L\left(x_1, x_2, \cdots, x_n ; \theta_1, \theta_2, \cdots, \theta_k\right)=\prod_{i=1}^n p\left(x_i ; \theta_1, \theta_2, \cdots, \theta_k\right) \text { 或 } \prod_{i=1}^n f\left(x_i ; \theta_1, \theta_2, \cdots, \theta_k\right). \tag{105}$$

如果 $p\left(x ; \theta_1, \theta_2, \cdots, \theta_k\right)$ 或 $f\left(x ; \theta_1, \theta_2, \cdots, \theta_k\right)$ 关于 $\theta_i(i=1,2, \cdots, k)$ 可微,则令

$$\frac{\partial L(\theta)}{\partial \theta_i}=0 \text { 或 } \frac{\partial \ln L(\theta)}{\partial \theta_i}=0. \tag{106}$$

若不可微，则似然方程组无解，应使用其他方法求解$\hat{\theta}_i$

3. 估计量的评价标准

(1) 无偏性

若参数 θ 的估计量 $\hat{\theta}=\hat{\theta}\left(X_1, X_2, \cdots, X_n\right)$ 对一切 n 及 $\theta \in I$, 有 $E \hat{\theta}=\theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估计量.

（2）有效性（最小方差性）

设 $\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$ 与 $\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 都是 θ 的无偏估计量, 如果 $D\left(\hat{\theta}_1\right)<D\left(\hat{\theta}_2\right)$, 则称 $\hat{\theta}_1$比 $\hat{\theta}_2$ 有效.

（3）一致性（相合性）

设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为未知参数 θ 的估计量, 如果对任意 $\varepsilon>0$, 有 $\lim_{n\to\infty}P\{|\hat{\theta}-\theta|<\varepsilon\}=1$, 即 $\hat{\theta}\xrightarrow{P}\theta\ (n\to\infty)$, 则称 $\hat{\theta}$ 为 θ 的一致估计量 (或相合估计量).

四、 参数的区间估计与假设检验

1. 区间估计

（1）概念

（2）正态总体均值的置信区间（置信水平为$1-\alpha$）

待估参数	其他参数	枢轴量的分布	置信区间
μ	σ^2已知	$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$
μ	σ^2未知	$t=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$	$\left(\bar{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\bar{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)$

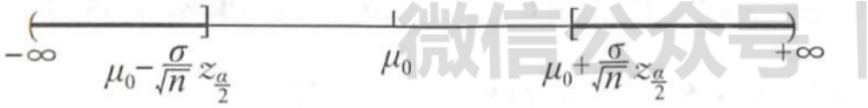
(107)

2. 假设检验

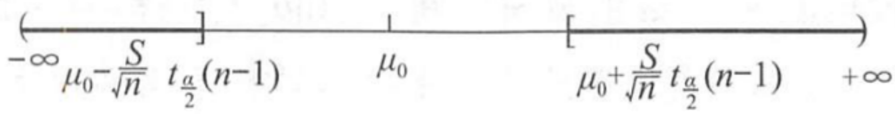
（1）思想方法

（2）正态总体下的六大检验及拒绝域

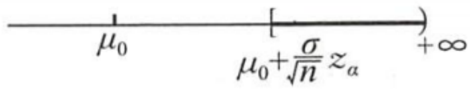
① σ^2 已知, μ 未知. $H_0:\mu=\mu_0,H_1:\mu\neq\mu_0$.



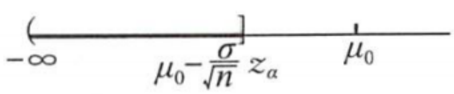
② σ^2 未知, μ 未知. $H_0:\mu=\mu_0,H_1:\mu\neq\mu_0$.



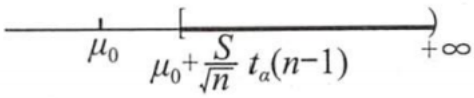
③ σ^2 已知, μ 未知. $H_0:\mu\leqslant\mu_0,H_1:\mu>\mu_0$.(或写 $\mu=\mu_0$)



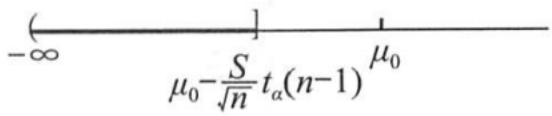
④ σ^2 已知, μ 未知. $H_0:\mu\geqslant\mu_0,H_1:\mu<\mu_0$.(或写 $\mu=\mu_0$)



⑤σ^2 未知, μ 未知. $H_0:\mu\leqslant\mu_0,H_1:\mu>\mu_0$.(或写 $\mu=\mu_0$)



⑥ σ^2 未知, μ 未知. $H_0:\mu\geqslant\mu_0,H_1:\mu<\mu_0$.(或写 $\mu=\mu_0$)



3. 两类错误

第一类错误（弃真）的概率$\alpha = P\{拒绝H_0|H_0为真\}$

第二类错误（取伪）的概率$\beta = P\{接受H_0|H_0为假\}$