概率论

一、 随机事件与概率

1.1 随机事件及运算

1.1.1 随机现象与随机试验

1.1.2 样本空间与随机事件

子事件、并事件、交事件、互不相容与对立事件、差事件;

交换律、结合律、分配律、对偶律（德摩根）

1.2 概率的公理化定义与概率的性质

1.2.1 等可能概型

样本空间中的每一个样本点发生的可能性都相同，又可分为古典概型和几何概型。古典概型指样本空间只包含有限个不同的可能结果；几何概型是样本空间为某个区域（如一维区间，二维平面区域或三维空间区域）。

1.2.2 概率的统计定义

用事件发生的频率作为事件发生的概率

1.2.3 概率的公理化定义与性质

1、概率的公理化定义：

设E为随机试验，Ω为相应的样本空间，若对任意事件A，有唯一的实数$P(A)$与之对应，且满足下面的条件，则称数$P(A)$为事件A的概率：

(1) (非负性)对于任意事件A，总有$P(A)\geqslant 0$；

(2) (规范性)$P(\Omega)=0$；

(3) (可列可加性)若$A_1,A_2,\cdots,A_n,\cdots$为两两互不相容的事件组，即$A_iA_j=\varnothing$，$i\neq j$，$i,j=1,2,\cdots$，则有$P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)$

2、概率的性质：

(1)$P(\varnothing)=0$；

(2)有限可加性；

(3)$P\left(\bar{A}\right)=1-P(A)$；

(4)$P(B-A)=P(B)-P(AB)$；

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

1.3 条件概率与事件的独立性

1.3.1 条件概率

定义：设E是随机试验，Ω是样本空间，A,B是事件且$P(A)>0$，称$P(B|A)=\dfrac{P(AB)}{P(A)}$为事件$A$发生的条件下事件$B$发生的概率。

拓展：设$A_1,A_2,\cdots,A_n$为事件组，且$P(A_1A_2\cdots A_n)>0$，则：

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_n)$$

1.3.2 事件的独立性

设A,B为随机试验E的两个事件，若满足等式$P(AB)=P(A)P(B)$，则称两事件相互独立

1.4 全概率公式与贝叶斯公式

1、全概率公式

设$A_1,A_2\cdots,A_n$为样本空间Ω的一个完备事件组，且$P(A_i)>0(i=1,2,\cdots,n)$，B为任一事件，则有$P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$

2、贝叶斯公式

设$A_1,A_2\cdots,A_n$为样本空间Ω的一个完备事件组，$P(A_i)>0(i=1,2,\cdots,n)$，B为任一事件且$P(B)>0$，则$P(A_i|B)=\dfrac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(A_j)P(B|A_j)}$，$i=1,2,\cdots,n$

二、 离散型随机变量

2.1 随机变量

设随机试验的样本空间为Ω，若$X=X(\omega)$为定义在样本空间Ω上的实值单值函数，则称$X=X(\omega)$为(一维)随机变量。

2.2 一维离散型随机变量

如果一个随机变量所有可能取值为有限个元素，则为离散型随机变量。称X取值x_i的概率$P(X=x_i)\triangleq p_i$为离散型随机变量X的分布律。

2.3 离散型随机变量的数学期望与方差

2.3.1 数学期望

设离散型随机变量X的分布律为$P(X=x_i)=p_i$，$i=1,2,\cdots,n,\cdots$，若$\sum\limits_{i=1}^{\infty}|x_i|p_i$收敛，则称$E(X)=\sum\limits_{i=1}^{\infty}x_i p_i$为随机变量$X$的数学期望；若不收敛，则数学期望不存在。

另外，令随机变量$Y=g(X)$，若$\sum\limits_{i=1}^{\infty}|g(x_i)|p_i$收敛，则有$E(Y)=E(g(X))=\sum\limits_{i=1}^{\infty}g(x_i)p_i$

2.3.2 方差与标准差

设X是一个随机变量，若$E(X-E(X))^2$存在，则称$Var(X)=E(X-E(X))^2$为X的方差，称$\sqrt{Var(X)}$为X的标准差，记为$\sigma(X)$。

$$Var(X)=E(X^2)-[E(X)]^2$$

另外，概率论中经常对随机变量X作以下的线性变换：

$$\tilde{X}=X-E(X) \quad 或 \quad X^*=\frac{X-E(X)}{\sqrt{Var(X)}}$$

则有：

$$E\left(\tilde{X}\right)=0,\ Var\left(\tilde{X}\right)=Var(X);$$

$$E(X^*)=0,\ Var(X^*)=1$$

我们称$\tilde{X}$为随机变量X的中心化随机变量，X^*为标准化随机变量。

2.4 常用离散型随机变量及其分布

2.4.1 二项分布

1、二项分布

将伯努利试验重复n次，我们称随机变量X服从参数为n,p的二项分布，记为$X\sim B(n,p)$

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,2,\cdots,n$$

2、$0-1$分布

在二项分布$B(n,p)$中取$n=1$，称X为服从参数为p的$0-1$分布或两点分布，此时分布律为：

$$P(X=k)=\binom{1}{k}p^k(1-p)^{1-k}=p^k(1-p)^{1-k}\ \ (k=0,1)$$

3、二项分布的数学期望与方差

$$E(X)=np,\quad Var(X)=np(1-p)$$

2.4.2 泊松分布

1、泊松分布

若随机变量X的分布律为$P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$，$k=0,1,2,\cdots$，其中$\lambda>0$，则称随机变量$X$服从参数为$\lambda$的泊松分布，记为$X\sim P(\lambda)$。

2、二项分布的泊松近似

设$\lambda>0$为一个常数，n为任意正整数，令$\lambda=np_n$，则对任一固定的非负整数k有：

$$\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$$

3、泊松分布的数学期望与方差

$$E(X)=\lambda,\ Var(X)=\lambda$$

2.4.3 几何分布

在伯努利试验中，每次试验中事件A发生的概率为$P(A)=p(0<p<1)$，令随机变量X为事件A首次发生时试验的次数，则X的分布律为：

$$P(X=k)=(1-p)^{k-1}p,\ \ k=1,2,\cdots,$$

称随机变量X服从参数为p的几何分布，记为$X\sim Ge(p)$。另外$E(X)=\dfrac{1}{p}$，$Var(X)=\dfrac{1-p}{p^2}$

(几何分布的无记忆性)设随机变量$X\sim Ge(p)$，则对于任意正整数m和n有：

$$P(X>m+n|X>m)=P(X>n)$$

2.4.4 超几何分布

从有限个总体中进行不放回抽样。假设现有N件产品，其中M件不合格品$(M \leqslant N)$，若从中不放回抽样随机抽取n件$(n \leqslant N)$，令随机变量X为抽取的n件产品中不合格品的数量，则X的分布律为：

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=0,1,\cdots,m$$

其中$m=\min\{M,n\}$，且N,M均为正整数，此时称随机变量X服从参数为N,M,n的超几何分布，记为$X\sim h(N,M,n)$。
特别的，当抽取件数n远小于产品总数N时，可用二项分布近似代替。

2.5 离散型随机变量函数的分布律

三、连续型随机变量

3.1 随机变量的分布函数

设X为随机变量，对任意的$x \in R$，称$F(x)=P(X \leqslant x)$为随机变量X的累积分布函数，也记为$F_X(x)$，当$a<b$时，$\{X \leqslant a\} \subset \{X \leqslant b\}$，因此：

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

(离散型随机变量分布函数与分布律的关系)
(1) 设离散型随机变量X的分布律为$P(X=x_k)=p_k$，$k=1,2,\cdots$，则对任意$-\infty<x<+\infty$，有$F_X(x)=\sum\limits_{k:x_k \leqslant x} p_k$；
(2) 设离散型随机变量X的分布函数为$F(x)$，若x_k，$k=1,2,\cdots$为$F(x)$的不连续点，则X的取值范围为$\Omega_X=\{x_1,x_2,\cdots\}$，且$X$的分布律为：

$$P(X=x_k)=p_k=F(x_k)-F(x_k-0),\quad k=1,2,\cdots$$

3.2 连续型随机变量及其密度函数

若存在非负可积函数$f(x)$，使得对任意实数x，随机变量X的分布函数都有：

$$F(x)=\int_{-\infty}^{x} f(t)\,dt$$

则称随机变量X为连续型随机变量，称$f(x)$为连续型随机变量X的概率密度函数。

3.3 连续型随机变量的数学期望与方差

1、设连续型随机变量X的密度函数为$f(x)$，如果$\int_{-\infty}^{+\infty}|x|f(x)dx$存在，则称

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

为随机变量X的数学期望，或称为随机变量X所服从分布的数学期望，若$\int_{-\infty}^{+\infty}|x|f(x)dx$不存在，则称随机变量$X$的数学期望不存在。

2、设连续型随机变量X的密度函数为$f(x)$，$g(x)$为R上的连续函数，若$\int_{-\infty}^{+\infty}|g(x)|f(x)dx$存在，则$g(X)$的数学期望为：

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

3、设连续型随机变量X的密度函数为$f(x)$，若$\int_{-\infty}^{+\infty}x^2f(x)dx$存在，则称

$$Var(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2f(x)dx$$

为随机变量X的方差，称$\sqrt{Var(X)}$为X的标准差，记为$\sigma(X)$。

另外，在实际运算中，$Var(X)=E(X^2)-[E(X)]^2$也同样适用。

3.4 常用连续型随机变量及其分布

3.4.1 均匀分布

若随机变量X的密度函数为$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b \\ 0, & 其他\end{cases}$，其中$a<b$，则称$X$服从区间$(a,b)$ (也可以包含等于a,b)上的均匀分布，记作$X\sim U(a,b)$，可得分布函数：

$$F(x)=\begin{cases}0, & x<a \\ \dfrac{x-a}{b-1}, & a\leqslant x<b \\ 1, & x\geqslant b\end{cases}$$

另外，$E(X)=\dfrac{a+b}{2}$，$Var(X)=\dfrac{(b-a)^2}{12}$。

3.4.2 正态分布

若随机变量X的密度函数为$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$，$-\infty<x<+\infty$，其中$-\infty<\mu<+\infty$，$\sigma>0$，则称$X$服从参数为$\mu,\sigma^2$的正态分布，记作$X\sim N(\mu,\sigma^2)$，称$X$为正态随机变量。$X$的分布函数为：

$$F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt,\quad -\infty<x<+\infty$$

在图像上，μ改变，图像平移；σ变小，图像变陡；σ变大，图像平缓。

(标准正态分布)　$X\sim N(0,1)$，密度函数$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，$-\infty<x<+\infty$

另外，$E(X)=\mu$，$Var(X)=\sigma^2$。

(正态分布的概率计算公式)　$P(a<X<b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$。

(正态分布的3σ准则)　当随机变量X在区间$(\mu-\sigma,\mu+\sigma),(\mu-2\sigma,\mu+2\sigma),(\mu-3\sigma,\mu+3\sigma)$内取值的概率分别为68.26%，95.44%，99.73%，X在之外取值的概率小于3%。

(标准正态分布的分位数)　设随机变量$X\sim N(0,1)$，若$P(X\leqslant u_p)=\Phi(u_p)=p$，称$u_p$为标准正态分布的$p$分位数。

(正态分布的线性不变性)正态分布经过线性变换后仍然服从正态分布。

3.4.3 指数分布

若随机变量X的密度函数为$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & \text{其他}\end{cases}$，其中$\lambda>0$，则称$X$服从参数为$\lambda$的指数分布，记作$X\sim E(\lambda)$。$X$的分布函数$F(x)=\begin{cases}1-e^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}$。

另外，$E(X)=\frac{1}{\lambda}$，$Var(X)=\frac{1}{\lambda^2}$，$P(X>s+t|X>s)=P(X>t)$，指数分布是连续型随机变量分布中唯一具有无记忆性的。

3.4.4 对数正态分布

若随机变量X的密度函数$f(x)=\begin{cases}\frac{1}{\sqrt{2\pi}\,\sigma x}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, & x>0\\ 0, & \text{其他}\end{cases}$，其中$-\infty<\mu<+\infty$，$\sigma>0$

则称X服从参数为μ，σ^2的对数正态分布，记作$X\sim LN(\mu,\sigma^2)$。X的分布函数为:

$$F(x)=\begin{cases}\int_{-\infty}^{\frac{\ln x-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt, & x>0\\ 0, & \text{其他}\end{cases}$$

若一个随机变量的对数服从正态分布，则称该随机变量服从对数正态分布，即：

若有$X\sim N(\mu,\sigma^2)$，那么$Y=e^X\sim LN(\mu,\sigma^2)$，$E(Y)=e^{\mu+\frac{\sigma^2}{2}}$,$Var(Y)=e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$

3.5 连续型随机变量函数的分布

分步函数法：

(1)有X的值域Ω_X得到Y的值域Ω_Y；

(2)在Y的值域上，计算$F_Y(y)=P(Y\leq y)=P(g(X)\leq y)=P(X\in D_y)=\int_{D_y}f(x)dx$，其中$D_y=\{x:g(x)\leq y\}$；

(3)由分布函数的性质写出完整的$F_Y(y)$表达式；

(4)求导得到密度函数$f_Y(y)$

3.6 其他常用数字特征

3.6.1 k 阶矩

1、设有随机变量X，对正整数k，若$E(|X|^k)$存在，则称$E(X^k)$为X的k阶原点矩。

2、设有随机变量X，对正整数k，若$E(|X-E(X)|^k)$存在，则称$E\{[X-E(X)]^k\}$为X的k阶中心矩。

3.6.2 分位数与中位数

设连续型随机变量X的分布函数为$F_X(x)$，则称满足$P(X\leq x_p)=F_X(x_p)=p$的实数x_p为X的p分位数。特别的，当$p=0.5$，称$x_{0.5}$为X的中位数。

3.6.3 偏度与峰度

设随机变量X的前三阶矩存在，则称$\beta_S(X)=\dfrac{E\{[X-E(X)]^3\}}{[Var(X)]^{\frac{3}{2}}}$为$X$的偏度，$\beta_S(X)<0$为左偏，$\beta_S(X)>0$为右偏。

设随机变量X的前四阶矩存在，则称$\beta_k(X)=\dfrac{E\{[X-E(X)]^4\}}{[Var(X)]^2}-3$为$X$的峰度。

3.6.4 变异系数

设随机变量X的数学期望$E(X)$和方差$Var(X)$存在，且$E(X)\neq 0$，则称

$$C_\nu(X)=\frac{\sqrt{Var(X)}}{|E(X)|}=\frac{\sigma(X)}{|E(X)|}$$

为X的变异系数。

四、随机向量

4.1 二维随机变量及其联合分布

4.1.1 二维离散型随机变量及其联合分布

1、如果一个二维随机变量的取值范围是有限数组(有限个点)，则称其为二维离散型随机变量。

2、设有二维随机变量(X,Y)，对任意的实数对(x,y)定义实值函数$F(x,y)=P(X\leqslant x,Y\leqslant y)$ $x\in(-\infty,+\infty)$，$y\in(-\infty,+\infty)$，则称$F(x,y)$为二维随机变量(X,Y)的联合分布函数。

4.1.2 二维离散型随机变量的联合分布律

4.1.3 二维连续型随机变量及联合概率密度函数

设有二维随机变量(X,Y)，其联合分布函数为$F(x,y)$，若存在一个定义域为$(-\infty,+\infty)$的非负实值函数$f(x,y)$，使得$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$，$-\infty<x,y<+\infty$，则称$(X,Y)$为二维连续型随机变量，并称二元函数$f(x,y)$为$(X,Y)$的联合概率密度函数。

4.2 边缘分布、随机变量的独立性和条件分布

4.2.1 边缘分布函数

设二维随机变量(X,Y)的联合分布函数为$F(x,y)$，称$F_X(x)=P(X\leqslant x)$，$x\in(-\infty,+\infty)$为随机变量X的边缘分布函数；同理，称$F_Y(y)=P(Y\leqslant y)$，$y\in(-\infty,+\infty)$为随机变量Y的边缘分布函数。

可以证明$F_X(x)=\lim\limits_{y\to+\infty}F(x,y)\triangleq F(x,+\infty)$，$F_Y(y)=\lim\limits_{x\to+\infty}F(x,y)\triangleq F(+\infty,y)$

4.2.2 边缘分布律和边缘密度函数

1、设二维离散型随机变量(X,Y)的联合分布律为：$P(X=a_i,Y=b_i)=p_{ij}$，$i,j=1,2,\cdots$，

由于$\bigcup_j\{Y=b_j\}=\Omega$，所以$P(X=a_i)=\sum_j p_{ij}=p_{i1}+p_{i2}+\cdots+p_{in}+\cdots\triangleq p_{i\cdot}$，$i=1,2,\cdots$ 为随机变量X的边缘分布律；同理可得Y的边缘分布律为

$P(Y=b_j)=\sum_i p_{ij}=p_{1j}+p_{2j}+\cdots+p_{nj}+\cdots\triangleq p_{\cdot j}$，$j=1,2,\cdots$

2、设二维连续型随机变量(X,Y)的联合密度函数为$f(x,y)$，$x,y\in(-\infty,+\infty)$，则称

$f_X(x)\triangleq\int_{-\infty}^{+\infty}f(x,y)dy$，$x\in(-\infty,+\infty)$为随机变量$X$的边缘密度函数；同理，称

$f_Y(y)\triangleq\int_{-\infty}^{+\infty}f(x,y)dx$，$y\in(-\infty,+\infty)$为随机变量$Y$的边缘密度函数。

4.2.3 随机变量的相互独立性

设二维离散型随机变量(X,Y)的联合分布律为$P(X=a_i,Y=b_j)=p_{ij}$，$i,j=1,2,\cdots$，若 $P(X=a_i,Y=b_j)=P(X=a_i)P(Y=b_j)$，对一切$i,j-1,2,\cdots$成立，则称$X$与$Y$相互独立。联合分布函数同理。

4.2.4 条件分布和条件数学期望