概率统计

第一章 随机事件与概率

频率

$$\text{频率} = \frac{\text{频数}}{\text{试验次数}}$$

概率

定义：频率具有稳定性的事件叫作随机事件，频率的稳定值叫作该随机事件的概率。

随机事件A在条件S下发生的概率为p，记作

$$P(A) = p$$

等概完备事件组

定义：称一个事件组$A_1, A_2, A_3, \cdots, A_n$为一个**等概完备事件组**，如果它具有下列三条性质：

1. **等可能性**：$A_1, A_2, A_3, \cdots, A_n$发生的机会相同
2. **完备性**：在人一次试验中，$A_1, A_2, A_3, \cdots, A_n$至少有一个发生（也就是所谓的“除此之外，不可能有别的结果”）
3. **互不相容性**：在任一次试验中，$A_1, A_2, A_3, \cdots, A_n$之多有一个发生（也就是所谓“他们是互相排斥的”）

等概完备事件组又称等概基本事件组，其中的任意事件$A_i(i = 1, 2, \cdots, n)$称为**基本事件。**

对于只满足条件2、3的事件组，称为**完备事件组。**

事件的运算

1. 必然事件表示为U，不可能事件表示为V
2. 包含：如果事件A发生，那么B必发生，就成事件B包含事件A，记作

$$A \subset B$$

3. 相等：如果事件A包含事件B，同时事件B包含事件A，那么就称事件A与B相等，或称等价，记作

$$A = B$$

4. 并：事件“A或B”称为事件A与事件B的并，记作

$$A \cup B \quad \text{或} \quad A + B$$

5. 交：事件“A且B”称为事件A和事件B的交，记作

$$A \cap B \quad \text{或} \quad AB \quad \text{或} \quad A \cdot B$$

6. 对立事件：事件“非A”称为A的对立事件，记作$\overline{A}$，有

$$A \cap \overline{A} = V$$

$$A \cup \overline{A} = U$$

7. 事件的差：事件A同B的差表示A发生而B不发生的事件，记作$A \setminus B$，由定义可知

$$A \setminus B = A \cap \overline{B}$$

事件的互不相容性

如果事件A与事件B不能都发生，即

$$AB = V(\text{不可能事件})$$

那么，称A与B是互不相容事件。

概率的加法公式

如果事件A，B互不相容，则

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

条件概率

如果A，B是条件S下的两个随机事件，$P(A) \neq 0$，则称在A发生的前提下B发生的概率为**条件概率**，记作$P(B \mid A)$

概率的乘法公式

$$P(AB) = P(A)P(B \mid A)$$

进一步地，

$$P(A)P(B \mid A) = P(B)P(A \mid B)$$

事件的独立性

事件A的发生并不影响事件B的发生，即

$$P(B \mid A)=P(B)$$

称两个事件A，B是**相互独立**的，如果

$$P(AB)=P(A)P(B)$$

全概公式

设事件组$A_1,A_2,A_3,\cdots,A_n$为完备事件组，则对任意一个事件B有

$$P(B)=\sum_{i=1}^{n} P(B \mid A_i)P(A_i)$$

考虑$i=2$时的简化情况，有

$$P(B)=P(B \mid A)P(A)+P(B \mid \overline{A})P(\overline{A})$$

逆概公式

设事件组$A_1,A_2,A_3,\cdots,A_n$为完备事件组，则对任意一个事件B有

$$P(A_j \mid B)=\frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}\ (j=1,\cdots,n)$$

逆概公式也称为**贝叶斯公式**，本质上是乘法公式与全概公式的结合，即

$$P(A_j \mid B)=\frac{P(A_jB)}{P(B)}=\frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}\ (j=1,\cdots,n)$$

独立试验序列概型

设每次射击打中目标的概率为p，连续射击n次，求恰好打中k次的概率。

计算公式：
设单次试验中，事件A发生的概率为$p(0<p<1)$，则在n次重复实验中，

$$P(A\text{发生}k\text{次})=C_n^k p^k q^{n-k} \quad (q=1-p)\,(k=0,1,2,\cdots,n)$$

第二章 随机变量与概率分布

随机变量

定义：对于条件组S下的每一个可能结果ω都唯一的对应到一个实数值$X(\omega)$，则称实值变量$X(\omega)$为一个随机变量，简记为X。

随机变量分为**离散型随机变量**和**连续型随机变量**。

举个例子：设盒中有5个球，其中2个白球、3个黑球，从中随便取3个球。则“抽得的白球数”X是一个随机变量。

一、离散型随机变量的概率分布

概率分布

将随机变量X的所有可能取值到其相应概率的映射称为X的概率分布，记为

$$p_k = P\{X = x_k\} \quad (k = 1, 2, \cdots)$$

常用的离散型随机变量的概率分布

1. 两点分布
   随机变量X仅取两个值：0或1，即

$$P\{X = 1\} = p \; (0 < p < 1)$$
$$P\{X = 0\} = q = 1 - p$$

2. 二项分布

$$P\{X = k\} = C_n^k P^k q^{n-k} \quad (k = 0, 1, 2, \cdots, n)(0 < p < 1)$$

   随机变量X满足二项分布可简记为：$X \sim B(n, p)$

3. 泊松分布

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \cdots, n)$$

   泊松分布是二项分布当

$$\lim_{n \to \infty} np = \lambda$$

   时的极限

4. 超几何分布

$$P\{X = m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} \quad (m = 0, 1, 2, \cdots, l) \text{ 其中}, l = min(M, n)$$

   举个例子：设一堆同类产品共N个，其中有M个次品。现从中任取n个（假定$n \leq N - M$），则这n个样品中所含次品个数X是一个离散型随机变量，其概率分布为超几何分布。

二、连续型随机变量

概率密度函数

定义：对于随机变量X，如果存在非负可积函数$p(x)(-\infty < x < \infty)$，使对任意的$a, b(a < b)$都有

$$P\{a < X < b\} = \int_a^b p(x)\mathrm{d}x$$

则称X为**连续性随机变量**；称$p(x)$为X的**概率密度函数**，简称概率密度或密度。

与离散型随机变量类比：将离散型随机变量X的离散值无限细分，则X的概率分布将变为概率密度函数。

显然，概率密度函数满足以下两条性质：

1. 对任何实数a，有

$$P\{X = a\} = 0$$

2. $$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$$

常见概率密度函数

1. 均匀分布
   如果随机变量X的概率密度为

$$p(x) = \begin{cases} \lambda & 当a \le x \le b \\ 0 & 其他 \end{cases} \quad (a < b)$$

   则称X服从$[a, b]$区间上的均匀分布

2. 指数分布

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & 当x \ge 0 \\ 0 & 当x < 0 \end{cases} \quad (\lambda > 0)$$

3. 正态分布

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad (-\infty < x < \infty) \ (\sigma > 0)$$

   变量X服从正态分布$N(\mu, \sigma^2)$可简记为$X \sim N(\mu, \sigma^2)$
   标准正态分布：参数$\mu = 0, \sigma = 1$时的正态分布，即$N(0, 1)$。它的密度函数为

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

   一个重要的积分：

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x = 1$$

通过正态分布的密度函数求某个区间的概率时，需要计算密度函数的积分，这种计算非常复杂，因此我们通过已经计算好数值的Φ函数来帮助求解：

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t$$

那么对于标准正态分布，有

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

对于一般正态分布$N(\mu, \sigma^2)$，常常使用**变量替换法**将其转化为标准正态分布，即令

$$t = \frac{x - \mu}{\sigma}$$

这时，$X \sim N(\mu, \sigma) \to T \sim N(0, 1)$。这样，对于一般正态分布也能轻易地计算其积分了

4. Γ分布

$$p(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & x > 0 \\ 0 & x \le 0 \end{cases} \quad (\alpha > 0, \beta > 0)$$

其中

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} \mathrm{d}x$$

变量X服从Γ分布可简记为$X \sim \Gamma(\alpha, \beta)$

5. 韦布尔分布

$$p(x) = \begin{cases} m \frac{x^{m-1}}{\eta^m} e^{-(\frac{x}{\eta})^m} & x > 0 \\ 0 & x \le 0 \end{cases}$$

分布函数

定义：设X是一随机变量（可以是连续型的，也可以是离散型的，甚至更一般的），称函数

$$F(x) = P(X \le x) \quad (-\infty < x < +\infty)$$

为X的分布函数。

连续型随机变量的分布函数事实上是其概率密度函数在区间$(-\infty, x)$上的不定上限积分。

随机变量函数的分布

随机变量函数：设$f(x)$是一个函数，所谓随机变量X的函数$f(X)$就是这样一个随机变量Y：当X取x时，它取值$y=f(x)$。记作

$$Y=f(X)$$

举个例子：设X是分子的速率，而Y是分子的动能，则Y是X的函数：$Y=\frac{1}{2}mX^2$（m是分子质量）

我们的目的是，根据已知的X的分布来寻求$Y=f(X)$的分布。

离散型随机变量函数的分布

假设离散型随机变量X,Y有如下关系：$Y=f(X)$。要得到$P\{Y=y_i\}$，只需求出$Y=y_i$时对应的x_i（可能有0个或多个对应值），将这些x_i对应的概率相加即可。

连续型随机变量函数的分布

分布函数法：已知X的分布，通过建立Y与X的分布函数之间的关系来求得Y的分布。

举个例子：已知$X\sim N(\mu,\sigma^2)$，求$Y=\frac{X-\mu}{\sigma}$的概率密度。
解：设Y的分布函数为$F_Y(y)$，于是

$$\begin{aligned}
F_Y(y)&=P(Y\le y)\quad(\text{根据分布函数的定义})\\
&=P(\frac{X-\mu}{\sigma}\le y)\quad(\text{因为}Y=\frac{X-\mu}{\sigma})\\
&=P(X\le\sigma y+\mu)\quad(\text{不等式变形})\\
&=F_X(\sigma y+\mu)\quad(\text{根据分布函数的定义})
\end{aligned}$$

其中$F_X(x)$为X的分布函数。那么，我们有

$$F_Y(y)=F_X(\sigma y+\mu)$$

将上式两边对y求微商，利用密度函数是分布函数的导数的关系，我们得到

$$p_Y(y)=p_X(\sigma y+\mu)\sigma$$

再将

$$p_X(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

代入，有

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

这表明$Y \sim N(0,1)$。

第三章 随机变量的数字特征

随机变量的期望

随机变量的期望$E(X)$是一个实数，它形式上是X所有可能取值的加权平均，代表了随机变量X的平均值。因此，也称期望为**均值**或**分布的均值**。

离散型随机变量的期望

$$E(X)=\sum_{k} x_k p_k \quad (=x_1p_1+x_2p_2+\cdots+x_kp_k+\cdots)$$

几个常用分布的期望

1. 两点分布

$$E(X)=1\cdot p+0\cdot q=p$$

2. 二项分布

$$E(X)=\sum_{k=1}^{n} kC_n^k p^k q^{n-k}=np$$

3. 泊松分布

$$\begin{aligned}
E(X)&=\sum_{k=0}^{\infty} k\cdot\frac{\lambda^k}{k!}e^{-\lambda}\\
&=\lambda e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!} \quad (\text{令}m=k-1)\\
&=\lambda e^{-\lambda}e^{\lambda} \quad (\text{根据泊松分布的密度之和为}1)\\
&=\lambda
\end{aligned}$$

4. 超几何分布

$$E(X)=\frac{nM}{N}$$

连续型随机变量的期望

定义：设连续型随机变量X的密度函数为$p(x)$，称

$$\int_{-\infty}^{+\infty} xp(x)\mathrm{d}x$$

为X的**期望**（或均值），记作$E(X)$。

本定义要求$\int_{-\infty}^{+\infty} |x|p(x)\mathrm{d}x$ 收敛

> 💡 在定性认识上，均值就是密度函数横坐标的中间值

几个常用分布的期望

1. 均匀分布

$$E(X) = \frac{1}{2}(b+a)$$

2. 指数分布

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xp(x)\mathrm{d}x \\ &= \int_{0}^{+\infty} \lambda x e^{-\lambda x}\mathrm{d}x \\ &= \frac{1}{\lambda}\int_{0}^{+\infty} te^{-t}\mathrm{d}t \quad (\text{令}t=\lambda x) \\ &= -\frac{1}{\lambda}\int_{0}^{+\infty} t\mathrm{d}e^{-t} \\ &= -\frac{1}{\lambda}\left[(te^{-t})\Big|_{0}^{+\infty} - \int_{0}^{+\infty} e^{-t}\mathrm{d}t\right] \\ &= \frac{1}{\lambda} \end{aligned}$$

3. 正态分布

$$E(X) = \mu$$

证明略。正态分布密度函数以$x=\mu$为对称轴，这就是其含义所在。

期望的简单性质

$$\begin{aligned} &E(c) = c; \\ &E(kX) = kE(X); \\ &E(X+b) = E(X)+b; \\ &E(kX+b) = kE(X)+b; \end{aligned}$$

随机变量函数的期望公式

对于离散型随机变量

$$E[f(X)] = \sum_i f(x_i)p_i$$

对于连续型随机变量

$$E[f(X)] = \int_{-\infty}^{+\infty} f(x)p(x)\mathrm{d}x$$

求随机变量函数的期望有如下两种方法：

1. 利用上述随机变量函数的期望公式直接求解；
2. 首先通过X的分布推出$f(X)$的分布，然后通过期望的定义求出$f(X)$的期望。

一般来说，第一种方法较为简单，是我们的首选方法。

随机变量的方差

统一定义

$$D(X) = E\left(\left[X - E(X)\right]^2\right)$$

这表明X的方差，就是$D = (X - c)^2$的均值（其中c为X的均值，即$c = E(X)$）

> 💡 定性认识，$D(X)$越小，则X取值越集中在$E(X)$附近。方差刻画了随机变量取值的分散程度。

方差简化计算公式：

$$D(X) = E(X^2) - E^2(X)$$

推导如下：

$$\begin{aligned}
D(X) &= \int_{-\infty}^{+\infty} \left[x - E(X)\right]^2 p(x)\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \left[x^2 - 2xE(X) + E^2(X)\right] p(x)\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} x^2 p(x)\mathrm{d}x - 2E(X)\int_{-\infty}^{+\infty} xp(x)\mathrm{d}x + E^2(X)\int_{-\infty}^{+\infty} p(x)\mathrm{d}x \\
&= E(X^2) - 2E(X)\cdot E(X) + E^2(X)\cdot 1 \\
&= E(X^2) - E^2(X)
\end{aligned}$$

离散型随机变量的方差

定义：设离散型随机变量的概率分布为

$$P(X=x_k)=P_k,\ k=1,2,\cdots$$

则称和数

$$\sum_k\left[x_k-E(X)\right]^2p_k$$

为X的**方差**，记作$D(X)$。

连续型随机变量的方差

定义：设连续型随机变量的密度为$p(x)$，则称

$$\int_{-\infty}^{+\infty}\left[x-E(X)\right]^2p(x)\mathrm{d}x$$

为X的**方差**，记作$D(X)$。

常用分布的方差

1. 两点分布

$$\begin{aligned}D(X)&=E(X^2)-E^2(X)\\&=(1^2\cdot p+0^2\cdot q)-p^2\\&=pq\end{aligned}$$

2. 二项分布

$$D(X)=npq$$

   证明略。

3. 泊松分布
   已知$E(X)=\lambda$，

$$\begin{aligned}E(X^2)&=\sum_{k=0}^{\infty}K^2\cdot\frac{\lambda^k}{k!}e^{-\lambda}\\&=\sum_{k=1}^{\infty}(k-1+1)\frac{\lambda^k}{(k-1)!}e^{-\lambda}\\&=\lambda^2\cdot\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda}+\lambda\cdot\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}\\&=\lambda^2+\lambda\end{aligned}$$

   则

$$D(X)=(\lambda^2+\lambda)-\lambda^2=\lambda$$

4. 均匀分布

$$D(X)=\frac{1}{12}(b-a)^2$$

5. 指数分布

$$D(X)=\frac{1}{\lambda^2}$$

6. 正态分布

$$D(X)=\sigma^2$$

方差的简单性质

$$\begin{aligned}&D(c)=0;\\&D(kX)=k^2D(X);\\&D(X+b)=D(X);\\&D(kX+b)=k^2D(X);\end{aligned}$$

切比雪夫不等式

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

第四章 随机向量

定义：我们称n个随机变量$X_1,X_2,\cdots,X_n$的整体$\xi=(X_1,X_2,\cdots,X_n)$为n维随机向量。

我们重点研究二维随机向量。

(二维)随机向量的联合分布与边缘分布

离散型随机向量

$\xi=(X,Y)$为二维离散型随机向量，当且仅当X,Y都是离散型随机变量。

一般称

$$P\{(X,Y)=(x_i,y_j)\}=p_{ij}\quad(i=1,2,\cdots;j=1,2,\cdots)$$

为$\xi=(X,Y)$的概率分布，也称为(X,Y)的**联合分布**。常采用**概率分布表**来表示离散型随机向量

的概率分布。这些p_{ij}具有2条基本性质：

1. 非负

$$p_{ij} \geq 0$$

2. 概率总和为1

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

三项分布：

$$P\{(X,Y)=(k_1,k_2)\} = \frac{n!}{k_1!k_2!(n-k_1-k_2)!} p_1^{k_1} p_2^{k_2} (1-p_1-p_2)^{n-k_1-k_2}$$

边缘分布与联合分布的关系

边缘分布：对于二维随机向量(X,Y)，分量X的概率分布称为(X,Y)的关于X的边缘分布。

$$P\{X=x_i\} = \sum_j p_{ij}$$

$$P\{Y=y_j\} = \sum_i p_{ij}$$

如果将(X,Y)的概率分布写在概率分布表中（i为行数，j为列数），则关于X的边缘分布为“将每行加和得到的一列”；关于Y的边缘分布为“将每列加和得到的一行”。

连续型随机向量

概念：对于二维随机向量$\xi=(X,Y)$，如果存在非负函数$p(x,y)(-\infty < x < +\infty, -\infty < y < +\infty)$，使对于任意一个邻边分别平行于坐标轴的矩形区域$D$：“即由不等式$a < x < b, c < y < d$确定的区域”，有

$$P\{(X,Y) \in D\} = \iint_D p(x,y) \mathrm{d}x\mathrm{d}y$$

则称随机向量$\xi=(X,Y)$为**连续型**的；并称$p(x,y)$为ξ的**分布密度**，也称$p(x,y)$为(X,Y)的**联合分布密度。**

由定义式容易得到

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x,y) \mathrm{d}x\mathrm{d}y = 1$$

> 💡 二维随机向量(X,Y)落在平面上任意区域D的概率，就等于联合密度$p(x,y)$在D上的积分。这就吧概率的计算转化为一个二重积分的计算。

> 💡 几何意义：$\{(X,Y)\in D\}$的概率，数值上就等于以曲面$z=p(x,y)$为顶、以平面区域D为底的曲顶柱体的体积。

边缘分布密度

定义：对于随机向量(X,Y)，作为其分量的随机变量X（或Y）的密度函数$p_X(x)$（或$p_Y(y)$），称为(X,Y)的关于X（或Y）的**边缘分布密度。**

当(X,Y)的联合密度$p(x,y)$已知是，可通过以下方法求得边缘密度

$$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y$$
$$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}x$$

随机变量的独立性

定义：设X,Y是两个随机变量，如果对任意的$a<b,c<d$，事件$\{a<X<b\}$与$\{c<Y<d\}$相互独立，则称X与Y是**相互独立**的。

重要定理：设X,Y分别有分布密度$p_X(x),p_Y(y)$，则X与Y相互独立的**充要条件**是：二元函数

$$p_X(x)p_Y(y)$$

是随机向量(X,Y)的联合密度。

二维正态分布

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}$$

两个边缘密度分别为两个一维正态分布：

$$P_X(x)=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$
$$P_Y(y)=\frac{1}{\sqrt{2\pi}\,\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

对于二维正态分布，**两个分量X与Y独立**的充要条件是$\rho=0$。

二维随机向量的分布函数

定义：设$\xi=(X,Y)$是二维随机向量，称函数

$$F(x,y)=P\{X\le x,Y\le y\}$$

为它的**分布函数。**

若$\xi = (X, Y)$的分布函数有二阶连续偏微商，则

$$\frac{\partial^2 F(x, y)}{\partial x \partial y}$$

就是ξ的**分布密度。**