第四章：随机变量的数字特征

已知X的分布函数，我们不仅可以通过其分布求任意一个随机事件的概率，还可以通过分布得知随机变量取值的**平均值**以及其取值**偏离平均值**的程度，这就是随机变量的**数学期望**和**方差**。

一维随机变量的数字特征 I

数学期望反映的是随机变量的**加权平均值**。

定义离散型随机变量X为某位选手射中的环数，已知其分布列为

$$X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 0.01 & 0.02 & 0.05 & 0.06 & 0.06 & 0.1 & 0.2 & 0.3 & 0.1 & 0.1 \end{pmatrix}$$

根据随机变量的分布，我们用下面的办法来计算这个选手的平均环数是合理的

$$\sum_{i=1}^{10} i \cdot \frac{n_i}{N} = \sum_{i=1}^{10} i \cdot P(X = i)$$

数学期望

设离散型随机变量X的概率分布列$P(X=x_i)=p_i$, $i=1,2,\cdots$，若级数$\sum_i |x_i|\, p_i$收敛，则称X的数学期望存在，并称

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

为X的**数学期望或均值**，简称X的**期望**

期望的存在性

要求$\sum_{i=1}^{\infty} p_i |x_i|$ 收敛，是为了说明无论怎么排列$\{x_i\}$，数项级数$\{S_n = \sum_{i=1}^{n} x_i p_i\}$将收敛到同样的极限。

下面几种情形不需要考虑$\sum_{i=1}^{\infty} p_i |x_i|$ 是否收敛，直接计算期望值即可。

1. 随机变量取值有限的情形
2. 随机变量取值非负的情形

离散型随机变量

对于离散随机变量函数$g(X)$，其期望如下计算

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

如果$P(X = x_i) = p_i$且$\sum_{i=1}^{\infty} |g(x_i)|\, p_i$收敛。

实例与练习

1. 已知

$$X \sim B(n,p) \quad X \sim Pois(\lambda) \quad X \sim Geo(p)$$

求$E[X]$.

2. 设X的分布列为

$$\begin{pmatrix} -1 & 0 & 1 & 3 \\ \frac{1}{8} & \frac{1}{4} & \frac{3}{8} & \frac{1}{4} \end{pmatrix}$$

计算$E[X]$, $E[-X+2]$, $E(X^2)$.

连续型随机变量的数学期望

设连续型随机变量X的分布密度函数为f_X，那么对于任意的x，有

$$P(x \le X \le x + \Delta x) = \int_x^{x+\Delta x} f_X(t)\,\mathrm{d}t \approx f_X(x)\Delta x$$

我们可以将$[-\infty, \infty]$划分为为$\bigcup_{i=-\infty}^{\infty}[x_{i-1}, x_i]$，那么根据Riemann积分的定义，$X$的期望可以表示为

$$E_X(x) = \lim_{\Delta x_i \to 0} \sum_{i=-\infty}^{\infty} x_i f_X(x_i)\Delta x_i = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$$

连续型随机变量的数学期望

定义4.1.2

设连续型随机变量X的分布密度函数为f_X,如果

$$\int_{-\infty}^{\infty} |x|\, f_X(x)\,\mathrm{d}x < \infty$$

那么称

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$$

为X的**数学期望(均值)**，简称为X的**期望**

随机变量函数的期望

已知连续型随机变量X的密度函数f_X，求$Y=g(X)$的数学期望，**不需要先求得**Y的分布密度函数$f_Y(y)$，再通过$\int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}x$计算

实际上，若

$$\int_{-\infty}^{\infty} |g(x)|\, f_X(x)\,\mathrm{d}x < \infty$$

那么$g(X)$的期望可以通过下面的方法计算

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$$

练习与思考

计算服从下面几种分布的连续型随机变量X的期望$E[X]$

1. $X \sim U[a,b]$
2. $X \sim N(\mu, \sigma^2)$
3. $X \sim \mathrm{E}xp(\lambda)$
4. X服从标准的柯西分布

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

判断X的数学期望是否存在，如果存在，求出$E[X]$。

5. $X \sim U[0, 2\pi]$，$Y = \sin X$，求$E[Y]$

数学期望的运算性质 I

性质1 任意常数c的数学期望等于c

证明：随机变量X服从

$$P(X)=\begin{cases}1, & X=c\\ 0, & \text{else}\end{cases}$$

所以其期望就是

$$E[X]=c\times 1+0=c$$

数学期望的运算性质 II

性质2 (线性性) 设随机变量X和Y的期望都存在，a, b都是常数，那么

$$E[aX + bY] = aE[X] + bE[Y]$$

证明：只对连续型的随机变量证明。

先证明随机变量$Z = aX + bY$的期望存在。假设X, Y的联合密度函数为$f_{X,Y}(x, y)$，判断下面积分的收敛性

$$\begin{aligned}
&\int_{\mathbb{R}^2} |ax + by|\, f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y \\
&\le |a| \int_{\mathbb{R}^2} |x|\, f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y + |b| \int_{\mathbb{R}^2} |y|\, f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y \\
&= |a| \int_{-\infty}^{\infty} |x|\, f_X(x)\, \mathrm{d}x + |b| \int_{-\infty}^{\infty} |y|\, f_Y(y)\, \mathrm{d}y \\
&< \infty
\end{aligned}$$

上面的式子表明$E[Z]$存在，根据积分运算的线性性

$$
\begin{aligned}
E[aX+bY] &= \int_{\mathbb{R}^2} (ax+by) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= a\int_{\mathbb{R}^2} x f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y + b\int_{\mathbb{R}^2} y f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= a\int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x + b\int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y \\
&= aE[X] + bE[Y]
\end{aligned}
$$

数学期望的运算性质 IV

性质3 若随机变量X与Y独立，且X与Y的期望都存在，那么

$$E[XY] = E[X]E[Y].$$

证明留作练习

随机变量的方差 I

随机变量的均值只能反映随机变量的平均预期，但却无法从中看出其取值的波动性大小，即随机试验的结果偏离平均值的程度。

巧克力加工厂A，B分别包装一带巧克力的重量$X_A \sim U[50-0.5, 50+0.5]$，$X_B \sim U[50-1.5, 50+1.5]$，那么

$$E[X_A] = E[X_B] = 50,$$

从均值来看，这两个加工厂的包装水平是一样的，但是

$$P(X_A \in [50-1.5, 50-0.5] \cup [50+0.5, 50+1.5]) = 0$$

$$P(X_B \in [50-1.5, 50-0.5] \cup [50+0.5, 50+1.5]) = \frac{2}{103}$$

方差的定义

我们用**方差**来描述一个随机变量取值偏离平均值的程度。

定义1.4.3

设随机变量X有**有限**的数学期望，如果$E[(X-E[X])^2]<\infty$，则称

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

为X的**方差**，而称$\sqrt{Var[X]}$为X的**标准差**，记为$\sigma[X]$.

方差

1. 对于离散型的随机变量：

$$Var[X] = \sum_{i=1}^{\infty} (x_i - E[X])^2 p_i$$

2. 对于连续型的随机变量：

$$Var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)\, \mathrm{d}x$$

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

课堂练习

计算下列随机变量的方差

1. $X \sim B(1, p)$
2. $X \sim \text{Pois}(\lambda)$
3. $X \sim \text{Geo}(p)$
4. $X \sim U[a, b]$
5. $X \sim \text{Exp}(\lambda)$
6. $X \sim N(\mu, \sigma^2)$

随机变量的矩 I

定义4.1.4

设X为随机变量，c为常数，k为正整数，如果$E[|X-c|^k]<\infty$，那么称

$$E[(X-c)^k]$$

为X关于c的k阶矩。

1. $c=0$，称$E[X^k]$为X的k**阶原点矩**，例如**均值**。
2. $c=E[X]$，$E[(X-E[X])^k]$称为X的k**阶中心矩**，例如**方差**。

对于三阶和四阶中心矩，概率学上有明确的定义

定义4.1.5

设X为随机变量，如果$E[X^4]<\infty$，则称

$$\frac{E[(X-E[X])^3]}{(Var[X])^{\frac{3}{2}}}$$

为X的**偏度**，偏度刻画X的分布的偏斜程度。而称

$$\frac{E[(X-E[X])^4]}{(Var[X])^2}$$

为X的**峰度**，峰度刻画随机变量的分布在其均值附近的陡峭程度。

随机变量的高阶矩 II

当密度函数$f_X(x)$关于$E[X]$对称时，偏度为零。

$$
\begin{aligned}
E[(X-E[X])^3] &= \int_{-\infty}^{\infty} (x-c)^3 f_X(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x^3 f_X(x+c)\,\mathrm{d}x
\end{aligned}
$$

记$g(x)=f_X(x+c)$，根据$f_X(x+c)=f_X(c-x)$，
知$g(x)=g(-x)$，即$g(x)$是偶函数，且$E[X^3]<\infty$

$$
E[(X-E[X])^3] = (\int_{-\infty}^{0} + \int_{0}^{\infty}) x^3 g(x)\,\mathrm{d}x = 0
$$

练习

计算下面分布的偏度和峰度

$$U(a,b),\quad Exp(\lambda),\quad N(\mu,\sigma)$$

随机向量函数的数字特征 I

对于二维的随机变量(X,Y)，假设$E[g(X,Y)]$存在，那么

1. 设离散型随机变量(X,Y)有概率分布$P(X=x_i,Y=y_i)=p_{ij}$，$i,j=1,2,\cdots$，那么

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

2. 设连续型随机变量(X,Y)有分布密度函数$f_{X,Y}$，那么

$$E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

思考与练习

1. 假设(X,Y)的联合分布列为

$$
\begin{aligned}
&P(X=0,Y=0)=P(X=1,Y=1)\\
&=P(X=0,Y=1)=P(X=1,Y=0)=0.25
\end{aligned}
$$

   求$E\left(\sin\left(\frac{\pi}{2}(X+Y)\right)\right)$

2. 在长度为a的线段上任取两个点X与Y，求这两个点之间的平均长度。
3. 设X_1与X_2是独立同分布的随机变量，它们均服从指数分布$Exp(\lambda)$，求$Y=\max\{X_1,X_2\}$和$Z=\min\{X_1,X_2\}$的数学期望。

协方差和相关系数 I

定义4.2.1

设(X,Y)为二维随机向量，且$Var[X]<\infty$，$Var[Y]<\infty$，

1. 称
$$Cov(X,Y)=E[(X-E[X])(Y-E[Y])]$$
为X与Y的**协方差**。
2. 称
$$r(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var[X]}\sqrt{Var[Y]}}$$
3. 若$r(X,Y)=0$，称X与Y**不相关**。

协方差的意义

1. $Cov(X,Y) > 0$，则X与Y有同增同减的倾向。
2. $Cov(X,Y) < 0$，则X与Y有此消彼涨的倾向。
3. $Cov(X,Y) = 0$，则要么X与Y独立，要么X与Y有某种非线性关系(见例题4.2.1)。

协方差的性质 I

1. $$Cov(X,Y) = E[XY] - E[X]E[Y]$$

   即

   $$E[XY] = E[X]E[Y] \Leftrightarrow Cov(X,Y) = 0$$

2. **独立与相关性**　X与Y相互独立，那么X与Y一定不相关，反之不成立。

3. **对称性**　设随机变量X和Y的方差都存在，那么

   $$Cov(X,Y) = Cov(Y,X)$$

协方差的性质 II

4 **双线性性** 设随机变量X，Y和Z的方差都存在，a和b为常数，那么

$$Cov(aX + bY, Z) = aCov(X, Z) + bCov(Y, Z)$$

以及

$$Cov(Z, aX + bY) = aCov(Z, X) + bCov(Z, Y)$$

方差的运算性质 I

性质1： 任意常数c的方差为0

$$Var[c] = E[(c - E[c])^2] = 0$$

性质2： 若随机变量X与Y**相互独立或者不相关**，并且X与Y的方差都存在，a，b为常数，那么

$$Var[aX + bY] = a^2Var[X] + b^2Var[Y]$$

对于任意的随机向量(X, Y)

$$Var(aX \pm bY) = a^2Var(X) + b^2Var(Y) \pm 2abCov(X, Y)$$

均值以及方差性质的运用 I

若X_1与X_2相互独立，且$X_1, X_2 \sim B(1,p)$，
则$X_1 + X_2 \sim B(2,p)$。进一步的，
若$X_1 \sim B(m,p)$，$X_2 \sim B(n,p)$，则$X_1 + X_2 \sim B(m+n,p)$。

若$X \sim B(n,p)$，$X_i \sim B(1,p)$，$i = 1, 2, \cdots, n$，相互独立，那么

$$X = X_1 + X_2 + \cdots + X_n$$

根据期望的线性性，

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

均值以及方差性质的运用 II

$E[X_i] = p$，故

$$E[X] = np$$

再由方差的性质

$$\begin{aligned} Var[X] &= Var[X_1 + X_2 + \cdots + X_n] \\ &= Var[X_1] + Var[X_2] + \cdots + Var[X_n] \\ &= np(1-p) \end{aligned}$$

数字特征的性质 I

定理4.2.1

设(X,Y)为二维随机向量，且$Var[X]<\infty$，$Var[Y]<\infty$

1. 若X与Y独立，则$r(X,Y)=0$，即X与Y不相关。
2. $$|r(X,Y)|\le 1$$

考虑下面这三个随机变量

$$\bar{X}=\frac{X-E[X]}{\sqrt{Var[X]}}\quad \bar{Y}=\frac{Y-E[Y]}{\sqrt{Var[Y]}}\quad Z=\bar{X}\pm\bar{Y}$$

首先

$$Var[aX]=a^2Var[X]$$

数字特征的性质 II

那么

$$Var[\bar{X}] = Var[\bar{Y}] = 1$$

其次

$$Cov(aX, bY) = E[abXY] - E[aX]E[bY] = abCov(X, Y)$$

计算下

$$\begin{aligned} Var[\bar{X} \pm \bar{Y}] &= Var[\bar{X}] + Var[\bar{Y}] \pm 2Cov(\bar{X}, \bar{Y}) \\ &= 2 \pm 2r(X, Y) \geq 0 \end{aligned}$$

所以$|r(X, Y)| \leq 1$

正相关　若$r(X,Y)=1$，则存在常数$a>0$和b，使得

$$P(Y=aX+b)=1$$

负相关　若$r(X,Y)=-1$，则存在常数$a<0$和b，使得

$$P(Y=aX+b)=1$$

只讨论第一种情况，第二种类似讨论
若$r(X,Y)=1$时，根据上面第二条，

$$Var[\bar{Y}-\bar{X}]=0,$$

即

$$\bar{Y}-\bar{X}=c, \quad a.e.$$

其中c是一个常数。变换上式得

$$Y = \sqrt{\frac{Var[Y]}{Var[X]}}X + c\sqrt{Var[Y]} - \sqrt{\frac{Var[Y]}{Var[X]}}\sqrt{E[X]} + E[Y]$$

令

$$a = \sqrt{\frac{Var[Y]}{Var[X]}} > 0,$$
$$b = c\sqrt{Var[Y]} - \sqrt{\frac{Var[Y]}{Var[X]}}\sqrt{E[X]} + E[Y]$$

相关系数只能反映X与Y的线性相关程度，而不能刻画X与Y的非线性关系。

二维正态分布的相关性

对于$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，$r(X,Y)=\rho$。

对于二维联合正态分布，

$$\rho=0\Leftrightarrow X,Y\ \text{独立且不相关}$$

即X与Y独立与X与Y不相关是等价的，这个性质是联合正态分布特有的。

例题 4.2.2 设(X,Y)的联合分布密度函数为

$$f_{X,Y}(x,y)=\begin{cases}8xy,\ 0\le y\le x,0\le x\le 1\\ 0,\ \text{其他}\end{cases}$$

求$E[X]$，$E[Y]$，$Var[X]$，$Var[Y]$，$Cov(X,Y)$和$r(X,Y)$

定义4.2.2

设随机向量$\boldsymbol{X}=(X_1,X_2,\cdots,X_n)'$的每个分量都有有限的方差，定义

$$E[\boldsymbol{X}]=(E[X_1],E[X_2],\cdots,E[X_n])'$$

和

$$Var[\boldsymbol{X}]=E[(\boldsymbol{X}-E[\boldsymbol{X}])(\boldsymbol{X}-E[\boldsymbol{X}])']=(Cov(X_i,X_j))_{n\times n}$$

条件数学期望 I

条件数学期望是在条件分布的数学期望。

离散型 设(X,Y)是二维离散型随机向量，有有限的数学期望，在$\{Y=b_j\}$发生的条件下，X的**条件数学期望(条件期望)**，就是对条件分布列

$$P(X=a_i\,|Y=b_j\,),\ i=1,2,\cdots$$

求数学期望，即

$$E[X\,|Y=b_j\,]=\sum_{i=1}^{\infty}a_iP(X=a_i\,|Y=b_j\,)$$

条件数学期望 II

同理，在$X = a_i$发生的条件下，Y的数学期望，就是在条件分布列

$$P(Y = b_j \,|X = a_i\,),\ j = 1, 2, \cdots$$

下求Y的数学期望

$$E[Y\,|X = a_i\,] = \sum_{j=1}^{\infty} b_j P(Y = b_j\,|X = a_i\,)$$

条件数学期望 III

连续型 若(X,Y)是连续型的随机向量，并且有有限的数学期望。在$\{Y=y\}$发生的情况下，X的条件数学期望，就是基于条件分布密度函数$f_{X|Y}(x|y)$求X数学期望

$$E[X|Y=y]=\int_{-\infty}^{\infty} x f_{X|Y}(x\mid y)\,\mathrm{d}x$$

在$\{X=x\}$发生的情况下，Y的条件数学期望，就是在条件分布密度函数$f_{Y|X}(y|x)$下求Y的数学期望

$$E[Y|X=x]=\int_{-\infty}^{\infty} y f_{Y|X}(y\mid x)\,\mathrm{d}y$$

条件期望的性质

1. 条件期望是条件的函数。比如$E[X|Y=y]$就是y的函数。
2. 当X和Y相互独立的时候，由于

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)$$

于是$f_{X|Y}(x\mid y)=f_X(x)$，于是

$$E[X|Y]=E[X]$$

同理

$$E[Y|X]=E[Y]$$

期望公式

$$E[E[X|Y]]=E[X],\ E[E[Y|X]]=E[Y]$$

重期望公式

1. 若Y是离散型随机变量，则

$$E[X] = \sum_i E[X \mid Y = y_i] P(Y = y_i)$$

2. 若Y是连续型随机变量，则

$$E[X] = \int_{\mathbb{R}} E[X \mid Y = y] f_Y(y)\,\mathrm{d}y$$

条件期望的运用 I

假设$E[X_i]=\mu$，$i=1,\cdots,N$，N是随机变量，且与X_i都独立，求$E[\sum_{k=1}^{N}X_k]$。

解：

$$
\begin{aligned}
E[\sum_{k=1}^{N}X_k] &= E\left(E[\sum_{k=1}^{N}X_k\,|N]\right)\\
&= \sum_{n=1}^{\infty}E[\sum_{k=1}^{N}X_k\,|N=n]\cdot P(N=n)\\
&= \sum_{n=1}^{\infty}E[\sum_{k=1}^{n}X_k]\cdot P(N=n)\\
&= \sum_{n=1}^{\infty}n\mu\cdot P(N=n)=\mu\sum_{n=1}^{\infty}nP(N=n)=\mu E[N]
\end{aligned}
$$

条件期望的运用 II

若(X,Y)服从二维正态分布$(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

$$E_{X|Y}(x\,|y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$$

$$E_{Y|X}(y\,|x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$$

即$E_{X|Y}$和$E_{Y|X}$分别是Y和X的线性函数，这是正态分布的特征。