第四章 随机变量的数字特征

分布函数能完整地描述 随机变量的统计特性，但实际应用中并不都需要知道分布函数，而只需知道随机变量的某些特征.

例如判断棉花质量时，既看纤维的平均长度又要看纤维长度与平均长度的偏离程度，平均长度越长，偏离程度越小，质量就越好.

考察一射手的水平，既要看他的平均环数是否高，还要看他弹着点的范围是否小，即数据的波动是否小.

由上面例子看到，与 随机变量有关的某些数值，虽不能完整地描述 随机变量但能清晰地描述 随机变量在某些方面的重要特征，这些数字特征在理论和实践上都具有重要意义.

随机变量某一方面的概率特性都可用

数字来描写

本章内容

- ❑ 随机变量的平均取值 —— 数学期望
- ❑ 随机变量取值平均偏离均值的情况 —— 方差
- ❑ 描述两 随机变量间的某种关系的数 —— 协方差与相关系数

§ 4.1 数学期望

Mathematical Expectation

引例 甲乙两射手进行射击训练，**100**次射击命中环数与次数记录如下:

甲:

环数	8	9	10
次数	30	10	60

乙:

环数	8	9	10
次数	20	50	30

如何评价甲、乙射手的技术优劣?

解 计算平均环数

甲: $$(8\times30+9\times10+10\times60)\div100=8\times0.3+9\times0.1+10\times0.6=9.3$$

乙: $$(8\times20+9\times50+10\times30)\div100=8\times0.2+9\times0.5+10\times0.3=9.1$$

甲： $(8\times30+9\times10+10\times60)\div100 = \underline{8\times0.3}+\underline{9\times0.1}+\underline{10\times0.6}=9.3$

乙： $(8\times20+9\times50+10\times30)\div100=\underline{8\times0.2}+\underline{9\times0.5}+\underline{10\times0.3}=9.1$

从平均射中环数看，甲的技术优于乙

频率 $f_k \longrightarrow$ **概率** p_k $\quad \sum_{k=1}^{3} k f_k \longrightarrow \sum_{k=1}^{3} k p_k$

数学期望的定义

离散型 设 X 为离散随机变量， 分布 $P\{X=x_k\}=p_k, \quad k=1,2,\cdots$

若无穷级数 $\sum_{k=1}^{+\infty} x_k p_k$ 绝对收敛,则称其和为 X 的

数学期望， 记作 $E(X)$， 即

$$E(X)=\sum_{k=1}^{+\infty} x_k p_k$$

数学期望的计算

例　已知随机变量X的分布律：

X	4	5	6
P	1/4	1/2	1/4

求数学期望E（X）

解

$$E(X)=4\times\frac{1}{4}+5\times\frac{1}{2}+6\times\frac{1}{4}=5$$

$$\boxed{E(X)=p_1x_1+p_2x_2+p_3x_3}$$

例 一种赌博，如果净赢利的期望值为零，就称为公平赌博。张三以**1:30**的赌注与李四打赌：抛掷一对均匀的骰子一次，若掷出双**6**点，张三给李四**30**元，否则李四给张三**1**元。你让你为这种赌博公平吗？

解、抛掷一对均匀的骰子一次，抛出双**6**的概率为**1/36,** 设在一次抛掷中，张三的赢钱数为**X**，李四的赢钱数为**Y**，则**X**和**Y**的概率分布为

X	1	−30
P	35/36	1/36

Y	−1	30
P	35/36	1/36

$$E(X)=1\times(35/36)-30\times(1/36)=5/36$$

$$E(Y)=-1\times(35/36)+30\times(1/36)=-5/36$$

解、抛掷一对均匀的骰子一次，抛出双**6**的概率为**1/36,**
设在一次抛掷中，张三的赢钱数为**X**，李四的赢钱数
为**Y**，则**X**和**Y**的概率分布为

X	1	−30
P	35/36	1/36

Y	−1	30
P	35/36	1/36

$$E(X)=1\times(35/36)-30\times(1/36)=5/36$$

$$E(Y)=-1\times(35/36)+30\times(1/36)=-5/36$$

由于期望值不为**0**，故赌博不公平，长期赌下去，
对张三有利。

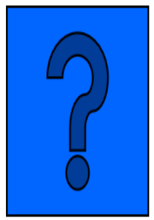
数学期望在医学上的一个应用

An application of Expected Value in Medicine

考虑用验血的方法在人群中普查某种疾病。集体做法是每10个人一组，把这10个人的血液样本混合起来进行化验。如果结果为阴性，则10个人只需化验1次；若结果为阳性，则需对10个人在逐个化验，总计化验11次。假定人群中这种病的患病率是10%，且每人患病与否是相互独立的。试问：这种分组化验的方法与通常的逐一化验方法相比，是否能减少化验次数？

分析：　设随机抽取的**10**人组所需的化验次数为**X**

我们需要计算**X**的数学期望，然后与**10**比较



先求出化验次数**X**的分布律。

注意求 **X** 期望值的步骤！

化验次数**X**的可能取值为**1，11**

{X=1}=“10人都是阴性”

$$P\{X=1\}=(1-0.1)^{10}=0.9^{10}$$

{X=11}=“至少**1**人阳性”

$$P\{X=11\}=1-0.9^{10}$$

$$E(X)\ =\ 0.9^{10}\times 1+(1-0.9^{10})\times 11=7.513<10$$

结论：　分组化验法的次数少于逐一化验法的次数

问题的进一步讨论

1、概率**p**对是否分组的影响

若**p=0.2**，则

$$E(X) = 0.8^{10}\times 1+(1-0.8^{10})\times 11 = 9.9262$$

当**p>0.2057**时，**E(X)>10**

2、概率**p**对每组人数**n**的影响

当**p=0.1**时，为使

$$E(X) = 0.9^{n}\times 1+(1-0.9^{n})\times 11 < 10$$

$$\longrightarrow n < 21.86$$

当**p=0.2**时，可得出**n<10.32**，才能保证

E(X)<10.

问题的进一步讨论

在第二次世界大战期间，所有美国的士兵都要进行一次**Wassermann**检验（梅毒的一种间接检验）。真正患有梅毒的的士兵约占全部检验者的**0.2%**左右。由于检验方法的灵敏度高，为了减少检验计划的巨大开支，采用了上述的分组检验方法。每组人数为**8**，减少工作量近**73%**。

连续型随机变量的数学期望E(X)

◆ 连续型随机变量

定义　设连续型随机变量X的概率密度为 f (x), 则

若广义积分 $\int_{-\infty}^{+\infty} xf(x)dx$ 绝对收敛，则称此积分为 X的数学期望

即

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

数学期望的计算

例　已知随机变量**X**的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{\pi\sqrt{1-x^2}} & |x|<1\\ 0 & |x|\geq 1\end{cases}$$

求数学期望。

解

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

$$=\int_{-\infty}^{-1} x\cdot 0\cdot dx+\int_{-1}^{1} x\frac{1}{\pi\sqrt{1-x^2}}dx+\int_{1}^{+\infty} x\cdot 0\cdot dx$$

$$=0$$

数学期望的意义

E(X)反映了随机变量X取值的“概率平均”,是X的可能值以其相应概率的加权平均。

试验次数较大时，X的观测值的算术平均值 $\overline{x}$ 在E(X)附近摆动

$$\overline{x} \approx E(X)$$

数学期望又可以称为期望值(**Expected Value**)，均值(**Mean**)

例　某商店对某种家用电器的销售采用先使用后付款的方式，记使用寿命为**X**（以年计），且规定：**X≤1**，每台付款**1500**元；**1<X≤2**，每台付款**2000**元；**2<X≤3**，每台付款**2500**元；**X>3**，每台付款**3000**元。这种家用电器的寿命**X**服从指数分布，其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{10}e^{-\frac{x}{10}} & x>0\\ 0 & 其他\end{cases}$$

求该商店每台电器收费**Y**的数学期望。

解、先求**X**落在各个时间区间内的概率

$$P\{X\le 1\}=\int_0^1\frac{1}{10}e^{-\frac{x}{10}}dx=1-e^{-0.1}=0.0952$$

$$P\{1<X\le 2\}=\int_1^2 \frac{1}{10}e^{-\frac{x}{10}}dx=e^{-0.1}-e^{-0.2}=0.0861$$

$$P\{2<X\le 3\}=\int_2^3 \frac{1}{10}e^{-\frac{x}{10}}dx=e^{-0.2}-e^{-0.3}=0.0779$$

$$P\{X>3\}=\int_3^{+\infty} \frac{1}{10}e^{-\frac{x}{10}}dx=e^{-0.3}=0.7408$$

根据以上计算，每台电器收费的数学期望为

$$E(Y)$$
$$=1500\times 0.0952+2000\times 0.0861+2500\times 0.0779+3000\times 0.7408$$
$$=2732.15$$

随机变量的函数的数学期望

定理 1：一维情形

设 $Y = g(X)$ 是随机变量 X 的函数，

- 离散型 $P\{X = x_k\} = p_k, \quad k = 1, 2, \cdots$

$$E(Y) = E[g(X)] = \sum_{k=1}^{\infty} g(x_k) p_k$$

- 连续型 概率密度为 $f(x)$

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

例　已知随机变量X的分布律如下表：

X	-1	0	1
p_k	0.25	0.50	0.25

求$E(X^2+1)$和$E((\frac{X}{1+X^2})^2)$。

解

$$E(X^2+1)=((-1)^2+1)\times 0.25+((0)^2+1)\times 0.50+$$

$$(1^2+1)\times 0.25=\ 1.5$$

$$E((\frac{X}{1+X^2})^2)=(\frac{-1}{1+(-1)^2})^2\times 0.25+(\frac{0}{1+0^2})^2\times 0.50+$$

$$(\frac{1}{1+1^2})^2\times 0.25=0.125$$

例　已知 X 服从$[0,2\pi]$上的均匀分布，求 $Y=\sin X$的数学期望。

解
$$E(Y)=E(\sin X)=\int_{-\infty}^{+\infty}\sin x\, f(x)\,dx$$

因为
$$f(x)=\begin{cases}\dfrac{1}{2\pi}, & 0\le x\le 2\pi;\\ 0, & \text{其它。}\end{cases}$$

所以
$$E(\sin X)=\int_0^{2\pi}\frac{1}{2\pi}\sin x dx=0$$

随机变量的函数的数学期望

定理 **2**：二维情形

设 $Z = g(X,Y)$ 是随机变量 X， Y的函数，

- 离散型 $P\{X = x_i, Y = y_j\} = p_{ij}, \quad i, j = 1, 2, \cdots$

$$\boxed{E[g(X,Y)] = \sum_i \sum_j g(x_i, y_j) p_{ij}}$$

- 连续型 联合概率密度为 $f(x,y)$

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

例1　已知 X,Y 的联合分布为

$Y \backslash X$	-1	0	1
0	0.07	0.18	0.15
1	0.08	0.32	0.20

求 $E(X)$ $E(Y)$,$E(XY)$

例2　已知

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

求E(X),E(Y),E(XY)

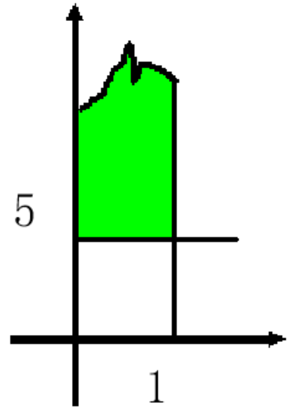
例　设相互独立的随机变量X，Y的密度函数分别为

$$f_1(x)=\begin{cases}2x, & (0\le x\le 1)\\ 0, & \text{其它}\end{cases}\qquad f_2(y)=\begin{cases}e^{-(y-5)}, & (y\ge 5)\\ 0, & \text{其它}\end{cases}$$

求E（XY）

解

$$E[XY]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy$$



$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_1(x)f_2(y)dxdy$$

$$=\int_0^1 dx\int_5^{+\infty}xy\cdot 2x\cdot e^{-(y-5)}dy$$

$$=\int_0^1 2x^2dx\cdot\int_5^{+\infty}ye^{-(y-5)}dy \quad =4$$

数学期望的性质

◆. $E(C)=C$ C 为常数

◆. $E(CX)=CE(X)$

◆. $E(X+Y)=E(X)+E(Y)$

◆ 当随机变量 X,Y 相互独立时

$$E(XY)=E(X)E(Y)$$

注 逆命题不成立，即

若 $E(XY)=E(X)E(Y)$，X,Y 不一定独立

反例

$Y \backslash X$ (p_{ij})	-1	0	1	$p_{\cdot j}$
-1	1/8	1/8	1/8	3/8
0	1/8	0	1/8	2/8
1	1/8	1/8	1/8	3/8
$p_{i\cdot}$	3/8	2/8	3/8	

XY	-1	0	1
P	2/8	4/8	2/8

$$E(X)=E(Y)=0;$$

$$E(XY)=0;$$

$$E(XY)=E(X)E(Y)$$

$$P\{X=-1,Y=-1\}=\frac{1}{8}\neq P\{X=-1\}P\{Y=-1\}=\left(\frac{3}{8}\right)^2$$

0-1分布的数学期望

分布律

X服从0-1分布，其概率分布为

$P\{X=1\}=p$

$P\{X=0\}=1-p$

X	0	1
P	1-p	p

数学期望

$$E(X)=0\times(1-p)+1\times p=p$$

若X 服从参数为 p 的0-1分布， 则E(X) = p

二项分布的数学期望

分布律　　X服从二项分布，其概率分布为

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}$$

数学期望

根据二项分布的定义，随机变量**X**是**n**重伯努利试验中事件**A**发生的次数，且在每次实验中事件**A**发生的概率为**p**。引进随机变量

$$X_i = \begin{cases} 0, & A\text{在第}i\text{次试验中不发生} \\ 1, & A\text{在第}i\text{次试验中发生} \end{cases}$$

则 $X = \sum_{i=1}^{n} X_i$，即二项分布可表示为 n 个**0-1**分布的和

则 $E(X) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = np$

If X~B(n, p), then E(X)= np

泊松分布的数学期望

分布律 $$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$$

数学期望

$$E(X)=\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\quad (k-1=t)$$

$$=\lambda e^{-\lambda}\sum_{t=0}^{\infty}\frac{\lambda^t}{t!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

If $X\sim P(\lambda)$, then $E(X)=\lambda$

均匀分布的期望

分布密度

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & 其它\end{cases}$$

数学期望

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_a^b \frac{x}{b-a}dx=\frac{a+b}{2}$$

正态分布的期望

分布密度

X~ N (μ， σ²)

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

数学期望

$$E(X)=\int_{-\infty}^{+\infty}x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$\xlongequal{t=\frac{x-\mu}{\sigma}}\int_{-\infty}^{+\infty}(\sigma t+\mu)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{t^2}{2}}\cdot\sigma dt$$

$$=\sigma\int_{-\infty}^{+\infty}t\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt+\mu\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

$$=\mu$$

指数分布的期望

分布密度

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & x\le 0\end{cases}$$

数学期望

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{+\infty}x\frac{1}{\theta}e^{-\frac{x}{\theta}}dx$$

$$=\left[-xe^{-\frac{x}{\theta}}\right]_0^{+\infty}+\int_0^{+\infty}e^{-\frac{x}{\theta}}dx=\left[-\theta e^{-\frac{x}{\theta}}\right]_0^{+\infty}$$

$$=\theta$$

常见 **r.v.** 的数学期望（熟记）

记住哦

分布	概率分布	期望
参数为p 的0-1分布	$P\{X=1\}=p$ $P\{X=0\}=1-p$	p
$B(n,p)$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k}$ $k=0,1,2,\cdots,n$	np
$P(\lambda)$	$P\{X=k\}=\dfrac{\lambda^k}{k!}e^{-\lambda}$ $k=0,1,2,\cdots$	λ

分布	概率密度	期望
区间(a,b)上的均匀分布	$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其它}\end{cases}$	$\dfrac{a+b}{2}$
$E(\theta)$	$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0,\\ 0, & \text{其它}\end{cases}$	θ
$N(\mu,\sigma^2)$	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ

例、一个民航客车载有20位旅客自机场开出，旅客有10个车站可以下车。如到达一个车站没有旅客下车就不停车，以X表示停车的次数，设每位旅客在各个车站下车是等可能的，并设各旅客是否下车相互独立，求E（X）。

解、引进随机变量

$$X_i=\begin{cases}0,\text{第i站没人下车}\\1,\text{第i站有人下车}\end{cases}$$

易知，$X=X_1+X_2+\cdots+X_{10}$

由题意，任意一个旅客在第i站不下车的概率为 9/10

因此，20位旅客都不在第i站下车的概率为 $(9/10)^{20}$

第i站有人下车（即至少有一人下车）的概率为 $1-(9/10)^{20}$

$$P\{X_i=0\}=\left(\frac{9}{10}\right)^{20}\qquad P\{X_i=1\}=1-\left(\frac{9}{10}\right)^{20}$$

$$\therefore E(X_i) = 1 - \left(\frac{9}{10}\right)^{20}, i = 1, 2, \cdots, 10$$

$$\begin{aligned}\therefore E(X) &= E(X_1 + X_2 + \cdots + X_{10}) \\ &= E(X_1) + E(X_2) + \cdots + E(X_{10}) \\ &= 10\left[1 - \left(\frac{9}{10}\right)^{20}\right]\end{aligned}$$

例、设一电路中电流 $I(A)$ 与电阻 $R(\Omega)$ 是两个相互独立的随机变量，其概率密度分别为

$$g(i)=\begin{cases}2i, 0\le i\le 1\\ 0, \quad 其他\end{cases} \qquad h(r)=\begin{cases}\dfrac{r^2}{9}, 0\le r\le 3\\ 0, \quad 其他\end{cases}$$

试求电压的均值。