第四章 随机变量的数字特征

• 中山大学 • 计算机学院 • 郑培嘉 • Email：zhpj@mail.sysu.edu.cn

目录

1. 数学期望
2. 方差
3. 协方差与相关系数
4. 矩、协方差矩阵

1. 数学期望

数学期望

例：一射手进行打靶练习，规定射入区域e_2得2分；射入区域e_1得1分；脱靶(射入区域e_0得0分). 射手一次射击所得分数X是随机变量。X的分布律为:

$$P\{X=k\}=p_k, \qquad k=0,1,2$$

现射击N次，其中得0分a_0次，得1分a_1次，得2分a_2次，$a_0+a_1+a_2=N$. 射击N次得分总和为$a_0\times 0+a_1\times 1+a_2\times 2$. 平均射击得分为:

$$\frac{a_0\times 0+a_1\times 1+a_2\times 2}{N}=\sum_{k=0}^{2}k\frac{a_k}{N}$$

这里$\frac{a_k}{N}$为事件$\{X=k\}$的频率. 当N很大时，$\frac{a_k}{N}$在一定意义下接近于事件$\{X=k\}$的概率p_k. 即随机变量X的观察值的算数平均$\sum_{k=0}^{2}k\frac{a_k}{N}$在一定意义下接近于$\sum_{k=0}^{2}kp_k$，我们称$\sum_{k=0}^{2}kp_k$为随机变量$X$的**数学期望**或**均值**。

数学期望

◆ **定义**：设离散型随机变量X的分布律为：

$$P\{X=x_k\}=p_k, \qquad k=1,2,\ldots.$$

若级数$\sum_{k=1}^{\infty}x_kp_k$绝对收敛，则称级数$\sum_{k=1}^{\infty}x_kp_k$的和为随机变量X的**数学期望**，记为$E(X)$. 即

$$E(X)=\sum_{k=1}^{\infty}x_kp_k$$

设连续型随机变量X的概率密度为$f(x)$, 若积分$\int_{-\infty}^{\infty}xf(x)\,dx$绝对收敛, 则积分$\int_{-\infty}^{\infty}xf(x)\,dx$的值为随机变量$X$的**数学期望**，记为$E(X)$. 即

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

数学期望简称**期望**，又称为**均值**。

数学期望

例：某医院当新生儿诞生时，医生需要对婴儿的各方面情况进行评分，设新生儿的得分X是一个随机变量，X的分布律如下表：

X	0	1	2	3	4	5	6	7	8	9	10
p_k	0.002	0.001	0.002	0.005	0.02	0.04	0.18	0.37	0.25	0.12	0.01

试求X的数学期望$E(x)$.

解：$E(x) = 0 \times 0.002 + 1 \times 0.001 + 2 \times 0.002 + 3 \times 0.005 + 4 \times 0.02 + 5 \times 0.04 + 6 \times 0.18 + 7 \times 0.37 + 8 \times 0.25 + 9 \times 0.12 + 10 \times 0.01 = 7.15$（分）

即若考察1000个新生儿，则一个新生儿的平均得分为7.15，1000个新生儿共得分7150分。

数学期望

例：有两个相互独立的电子装置，他们的寿命$X_k\ (k=1,2)$服从同一指数分布，其概率密度为：

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

若将这两个电子装置串联成整机，求整机寿命 N 的数学期望。

解： X_k 的分布函数为：

$$F(x)=\begin{cases}1-e^{-x/\theta}, x>0 \\ 0, x\le 0\end{cases}$$

而$N=\min\{X_1,X_2\}$的分布函数为

$$F_{min}(x)=1-[1-F(x)]^2=\begin{cases}1-e^{-\frac{2x}{\theta}}, & x>0 \\ 0, & x\le 0\end{cases}$$

数学期望

因而N的概率密度为：

$$f_{min}(x) = \begin{cases} \frac{2}{\theta} e^{-\frac{2x}{\theta}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

于是N的数学期望为：

$$E(N) = \int_{-\infty}^{\infty} x\, f_{min}(x) dx = \int_{0}^{\infty} \frac{2x}{\theta} e^{-\frac{2x}{\theta}} dx = \frac{\theta}{2}$$

数学期望

例：某车站每天8: 00~9: 00，9: 00~10: 00都恰有一辆客车到站，但到站时刻是随机的，且两者到站时间相互独立，其规律为:

到站时刻	8:10 9:10	8:30 9:30	8:50 9:50
概率	$\frac{1}{6}$	$\frac{3}{6}$	$\frac{2}{6}$

一旅客8：20到车站，求他候车时间的数学期望.

解：设旅客的候车时间为X，X的分布律为:

X	10	30	50	70	90
p_k	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}\times\frac{1}{6}$	$\frac{1}{6}\times\frac{3}{6}$	$\frac{1}{6}\times\frac{2}{6}$

则候车时间的数学期望为:

$$E(X)=10\times\frac{3}{6}+30\times\frac{2}{6}+50\times\frac{1}{36}+70\times\frac{3}{36}+90\times\frac{2}{36}=27.22$$

数学期望

例：某商店对某种家电销售采取先使用后付款的方式，记使用寿命为X，规定：

$X \le 1$, 一台付款1500元；　　$1 < X \le 2$, 一台付款2000元；

$2 < X \le 3$, 一台付款2500元；　　$X > 3$, 一台付款3000元；

设寿命X服从指数分布，其概率密度如下：

$$f(x) = \begin{cases} \frac{1}{10} e^{-x/10}, x > 0 \\ 0, x \le 0 \end{cases}$$

试求该商店一台这种家电收费Y的数学期望。

解：先求寿命X落在各个时间区间的概率，有

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

数学期望

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

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

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

一台家用电器收费Y的分布律为：

Y	1 500	2 000	2 500	3 000
p_k	0.095 2	0.086 1	0.077 9	0.740 8

则$E(Y) = 2732.15$，即平均一台收费2732.15元。

数学期望

例：在一个人数很多的团体中普查某种疾病，为此要抽验N个人的血，可以用两种方法进行.（i）将每个人的血分别去验，这就需验N次.（ii）按k个人一组进行分组，把从k个人抽来的血混合在一起进行检验. 如果这混合血液呈阴性反应，就说明k个人的血都呈阴性反应，这样，这k个人的血就只需验一次；若呈阳性，则再对这k个人的血液分别进行化验，这样，k个人的血总共要化验$k+1$次. 假设每个人化验呈阳性的概率为p，且这些人的试验反应是相互独立的. 试说明当p较小时，选取适当的k，按第二种方法可以减少化验的次数. 并说明k取什么值时最适宜.

解：各人的血呈阴性反应的概率为$q=1-p$.
因而k个人的混合血呈阴性反应的概率为q^k，k个人的混合血呈阳性反应的概率为$1-\mathrm{q}^k$。

数学期望

设以k个人为一组时，组内每人化验次数为X，则X是一个随机变量，其分布律为：

X	$\frac{1}{k}$	$\frac{k+1}{k}$
p_k	q^k	$1-q^k$

X的数学期望为$E(x) = \frac{1}{k}\mathrm{q}^k + \left(1+\frac{1}{k}\right)1-\mathrm{q}^k = 1-\mathrm{q}^k+\frac{1}{k}$.

N个人平均需化验次数为$N(1-\mathrm{q}^k+\frac{1}{k})$

因此只要选择k使$1-\mathrm{q}^k+\frac{1}{k}<1$

则N个人平均需化验次数$<\mathrm{N}$.当p固定时，选取k使得$\mathrm{L}=1-\mathrm{q}^k+\frac{1}{k}$小于1且取到最小值，此时为最好分组法。

如：$p=0.1; q=0.9; k=4$时， $\mathrm{L}=1-\mathrm{q}^k+\frac{1}{k}$取到最小值.

若$N=1000$, 以$k=4$分组，按第二种方法化验只需

$1000\left(1-0.9^4+\frac{1}{4}\right)=594$（次）。 **减少了40%工作量**

数学期望

例：设随机变量$X\sim\pi(\lambda)$,求$E(X)$。

X的分布律为$P\{X=k\}=\dfrac{\lambda^k e^{-\lambda}}{k!}$， $k=0,1,2,\dots,\ \lambda>0$

X的数学期望为:

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

例：设随机变量$X\sim\mathrm{U}(\mathrm{a},\mathrm{b})$,求$E(X)$。

X的概率密度为

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

X的数学期望为: $E(X)=\int_{-\infty}^{\infty}x\,f(x)dx=\int_a^b\frac{x}{b-a}dx=\frac{a+b}{2}$

数学期望

◆ **定理**：设Y是随机变量X的函数：$Y = g(X), g$是连续函数

➤ 如果X是离散型随机变量，分布律为$P\{X = x_k\} = p_k, k = 1,2, \dots,$若$\sum_{k=1}^{\infty} g(x_k)\, p_k$绝对收敛，则有

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

➤ 如果X是连续型随机变量，密度函数为$f(x)$,若$\int_{-\infty}^{\infty} g(x)\, f(x) dx$绝对收敛，则有

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

注：定理意义在于我们求$E(Y)$时，不需要求Y的分布律或密度函数，只需要知道X的分布律或密度函数即可