第四章 随机变量的数字特征

分布函数能够完整地描述随机变量的统计特性，但在一些实际问题中，只需知道随机变量的某些特征，因而不需要求出它的分布函数. 例如：

评定某企业的经营能力时，只要知道该企业人均赢利水平；

研究水稻品种优劣时，我们关心的是稻穗的平均粒数及每粒的平均重量；

检验棉花的质量时，既要注意纤维的平均长度，又要注意 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平均值的情况 —— 方差
- 描述两个随机变量之间的某种关系的数 —— 协方差与相关系数

§4.1 随机变量的数学期望

引例1 甲乙两学生参加数学竞赛, 观察其胜负

	初赛	复赛	决赛	总成绩	算术平均	加权平均		
						3:3:4	2:3:5	2:2:6
甲	90	85	53	228	76	73.7	70.0	66.8
乙	88	80	57	225	75	73.2	70.1	67.8
胜者	甲	甲	乙	甲	甲	甲	乙	乙

引例2 测量 50 个圆柱形零件直径（见下表）

尺寸（cm）	8	9	10	11	12	
数量（个）	8	7	15	10	10	50

则这 50 个零件的平均直径为

$$\frac{8\times8+9\times7+10\times15+11\times10+12\times10}{50}$$

$$=10.14cm$$

换一个角度看，从这50个零件中任取一个零件，它的尺寸为随机变量X，则X的分布律为

X	8	9	10	11	12
P	$\frac{8}{50}$	$\frac{7}{50}$	$\frac{15}{50}$	$\frac{10}{50}$	$\frac{10}{50}$

则这 50 个零件的平均直径为

$$\overline{D}=\sum_{k=8}^{12}k\times P(X=k)=\sum_{k=8}^{12}kp_k=10.14$$

称之为这 5 个数字的加权平均，数学期望的概念源于此

一、离散型随机变量数学期望

1、定义 设 X 为离散型随机变量，其分布律为

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

若无穷级数

$$\sum_{k=1}^{+\infty} x_k p_k$$

绝对收敛，则称其和为随机变量 X 的数学期望

记作 $E(X)$

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

例1　$X \sim B(n,\ p)$, 求 $E(X)$.

解　$$E(X)=\sum_{k=0}^{n} kC_n^k p^k(1-p)^{n-k}$$

$$=\sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{(n-1)-(k-1)}$$

$$=np\sum_{k=0}^{n-1} C_{n-1}^k p^k(1-p)^{n-k}$$

$$=np$$

例2　$X \sim P(\lambda)$，求 $E(X)$.

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

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

例3　设 $X \sim$ 参数为 p 的几何分布，求$E(X)$.

解　$$E(X)=\sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p\left(\sum_{k=1}^{+\infty} kx^{k-1}\right)\Bigg|_{x=1-p}$$

$$=p\left(\sum_{k=1}^{+\infty} x^k\right)'\Bigg|_{x=1-p} = p\frac{1}{(1-x)^2}\Bigg|_{x=1-p} = \frac{1}{p}$$

例4　某人的一串钥匙上有n把钥匙，其中只有一把能打开自己的家门，他随意地试用这串钥匙中的某一把去开门. 若每把钥匙试开一次后除去，求打开门时试开次数的数学期望.

解　设试开次数为X,

$$P\{X=k\}=1/n\ ,\ k=1,2,\ldots,n$$

于是　$$E(X)=\sum_{k=1}^{n}k\cdot\frac{1}{n}=\frac{1}{n}\cdot\frac{(1+n)n}{2}=\frac{n+1}{2}$$

2、几个常见的离散型随机变量的数学期望

分布	分布律	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
$P(\lambda)$	$P\{X=k\}=\dfrac{\lambda^k e^{-\lambda}}{k!}$ $k=0,1,2,\cdots$	λ

二、连续型随机变量数学期望

1、定义 设 X 为连续型随机变量, 其概率密度为

$$f(x)$$

若广义积分

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

绝对收敛，则称此积分为随机变量 X 的数学期望

记作 $E(X)$

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

随机变量的数学期望的本质 —— 加 权 平 均

它是一个数不再是随机变量

例5　$X \sim N(\mu, \sigma^2)$，求 $E(X)$.

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

$$\overset{令\frac{x-\mu}{\sigma}=u}{=} \int_{-\infty}^{+\infty}(u\sigma+\mu)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du = \mu$$

例6　设 $X \sim U(a,b)$，求 $E(X)$.

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

$$=\frac{a+b}{2}$$

2、几个常见的连续型随机变量的数学期望

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

注意：不是所有的随机变量都有数学期望

例如：**Cauchy**分布的密度函数为

$$f(x)=\frac{1}{\pi(1+x^2)},\quad -\infty<x<+\infty$$

但 $\int_{-\infty}^{+\infty}|x|\,f(x)dx=\int_{-\infty}^{+\infty}\frac{|x|}{\pi(1+x^2)}dx$ 发散

它的数学期望不存在

三、随机变量函数的数学期望

1、 设X为离散型随机变量，分布律为

$$P\{X = x_i\} = p_i, \quad i = 1,2,\cdots$$

$Y = g(X)$, 若级数

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

绝对收敛，则

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

2、 设X为连续型随机变量，概率密度为$f(x)$

$Y = g(X)$， 若广义积分

$$\int_{-\infty}^{+\infty} g(x)f(x)dx$$

绝对收敛，则

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

3、 设(X,Y)为二维离散型随机变量，分布律为

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

$Z=g(X,Y)$，若级数

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)p_{ij}$$

绝对收敛，则

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

4、 设(X,Y)为二维连续型随机变量，概率密度函数为$f(x,y)$

$Z=g(X,Y)$， 若广义积分

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$$

绝对收敛，则

$$E(Z)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$$

例7　设二维连续随机变量(X,Y)的密度函数为

$$f(x,y)=\begin{cases}\frac{1}{4}x(1+3y^2), & 0<x<2, 0<y<1,\\ 0, & \text{其它}\end{cases}$$

求$E(X), E(Y), E(X+Y), E(XY), E(Y/X)$

解

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

$$=\int_0^2 x\cdot\frac{1}{4}xdx\int_0^1(1+3y^2)dy=\frac{4}{3}$$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy$$

$$=\int_0^2\frac{1}{4}xdx\int_0^1 y(1+3y^2)dy=\frac{5}{8}$$

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

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy$$

$$=E(X)+E(Y)$$ —— 数学期望的性质

$$=\frac{4}{3}+\frac{5}{8}=\frac{47}{24}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(xy)f(x,y)dxdy$$

$$=\int_0^2 x\cdot\frac{1}{2}xdx\int_0^1 y\cdot\frac{1}{2}(1+3y^2)dy$$

$$=\frac{4}{3}\cdot\frac{5}{8}=\frac{5}{6}$$

$$=E(X)\cdot E(Y)$$ —— 数学期望的性质

注意：X,Y 相互独立

$$E\left(\frac{Y}{X}\right)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(\frac{y}{x}\right)f(x,y)dxdy$$

$$=\int_0^2\frac{1}{2}dx\int_0^1 y\cdot\frac{1}{2}(1+3y^2)dy$$

$$=\frac{5}{8}$$

例8 设 $(X,Y)\sim N(0,1;0,1;0)$, 求$Z=\sqrt{X^2+Y^2}$ 的数学期望.

解 $$E(Z)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sqrt{x^2+y^2}\,f(x,y)dxdy$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sqrt{x^2+y^2}\,\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}dxdy$$

$$=\int_{0}^{2\pi}\left(\int_{0}^{+\infty}r\frac{1}{2\pi}e^{-\frac{r^2}{2}}rdr\right)d\theta$$

$$=\sqrt{\frac{\pi}{2}}$$

四、数学期望的性质

1、$E(C)=C$

2、$E(aX)=aE(X)$

3、$E(X+Y)=E(X)+E(Y)$

$$E\left(\sum_{i=1}^{n}a_iX_i+C\right)=\sum_{i=1}^{n}a_iE(X_i)+C$$

4、当X,Y相互独立时，$E(XY)=E(X)E(Y)$.

5、若存在常数a使$P\{X\geq a\}=1$，则$E(X)\geq a$

若存在常数b使$P\{X\leq b\}=1$，则$E(X)\leq b$

注 性质 4 的逆命题不成立，即

若$E(XY)=E(X)E(Y)$，X,Y不一定相互独立

反例 1

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

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

$$E(X)=E(Y)=0; \qquad E(XY)=0;$$

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

但 $$P\{X=-1,Y=-1\}=\frac{1}{8}$$

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

反例 2 $(X,Y)\sim U(D),\quad D=\{(x,y)\mid x^2+y^2\le 1\}$

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

$$f_X(x)=\begin{cases}\dfrac{2\sqrt{1-x^2}}{\pi}, & -1<x<1,\\ 0, & \text{其它}\end{cases}$$

$$f_Y(y)=\begin{cases}\dfrac{2\sqrt{1-y^2}}{\pi}, & -1<y<1,\\ 0, & \text{其它}\end{cases}$$

$$f(x,y)\ne f_X(x)f_Y(y)$$

但

$$E(X)=\int_{-1}^{1} x\frac{2\sqrt{1-x^2}}{\pi}=0;$$

$$E(XY)=\iint_{x^2+y^2\le 1} xy\frac{1}{\pi}dxdy=0;$$

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

证 性质5

设 X 为连续型，密度函数为$f(x)$, 分布函数为 $F(x)$, 则

$$P\{X \geq a\} = 1 - P\{X < a\}$$

$$\longrightarrow \quad 1 - F(a) = 1 \quad \longrightarrow \quad F(a) = 0$$

$$\longrightarrow \quad F(x) = 0, \quad x \leq a$$

$$\longrightarrow \quad f(x) = 0, \quad x \leq a$$

故 $E(X) = \int_a^{+\infty} xf(x)dx \geq \int_a^{+\infty} af(x)dx = a$

例9 五个独立元件，寿命分别为 $X_1, X_2, \cdots, X_5$，都服从参数为 λ 的指数分布，若将它们

(1) 串联； (2) 并联

成整机，求整机寿命的均值.

解 (1) 设整机寿命为 N，$N = \min\limits_{k=1,2,\cdots,5}\{X_k\}$

$$F_N(x) = 1 - \prod_{k=1}^{5}(1 - F_k(x)),$$

$$= \begin{cases} 1 - e^{-5\lambda x}, & x > 0, \\ 0, & \text{其它}, \end{cases}$$

$$f_N(x)=\begin{cases}5\lambda e^{-5\lambda x}, & x>0,\\ 0, & \text{其它},\end{cases}$$

即 $N\sim E(5\lambda)$, $\quad E(N)=\dfrac{1}{5\lambda}$

(2) 设整机寿命为 $M=\max\limits_{k=1,2,\cdots,5}\{X_k\}$

$$F_M(x)=\prod_{k=1}^{5}F_k(x)=\begin{cases}(1-e^{-\lambda x})^5, & x>0,\\ 0, & \text{其它},\end{cases}$$

$$f_M(x)=\begin{cases}5\lambda e^{-\lambda x}(1-e^{-\lambda x})^4, & x>0,\\ 0, & \text{其它},\end{cases}$$

$$E(M)=\int_{-\infty}^{+\infty} x f_M(x)dx$$

$$=\int_0^{+\infty} 5\lambda x e^{-\lambda x}(1-e^{-\lambda x})^4 dx$$

$$=\frac{137}{60\lambda}$$

$$\frac{E(M)}{E(N)}=\frac{137/60\lambda}{1/5\lambda}>11$$

可见，并联组成整机的平均寿命比串联组成整机的平均寿命长11倍之多.

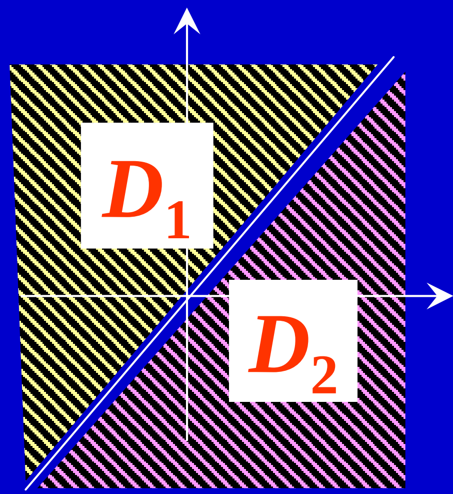
例10 设$X \sim N(0,1)$, $Y \sim N(0,1)$, X, Y 相互独立，求 $E(\max\{X,Y\})$.

解 $$f(x,y) = f_X(x) f_Y(y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

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

$$= \iint_{D_1} \max\{x,y\} f(x,y)dxdy$$

$$+ \iint_{D_2} \max\{x,y\} f(x,y)dxdy$$



$$= \iint_{D_1} y \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dxdy + \iint_{D_2} x \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dxdy$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \int_{x}^{+\infty} y e^{-\frac{y^2}{2}} dy + \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy \int_{y}^{+\infty} x e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\pi}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \int_{x}^{+\infty} y e^{-\frac{y^2}{2}} dy = \frac{1}{\pi}\int_{-\infty}^{+\infty} e^{-x^2} dx = \frac{1}{\sqrt{\pi}}$$

其中 $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$ 称为 概率积分

$$\left(\int_{-\infty}^{+\infty} e^{-x^2} dx\right)^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+y^2)} dxdy$$

$$= 4\int_{0}^{+\infty}\int_{0}^{+\infty} e^{-(x^2+y^2)} dxdy$$

$$= 4\int_0^{+\infty}\int_0^{+\infty} e^{-(x^2+y^2)}dxdy = 4\int_0^{\frac{\pi}{2}} d\theta \int_0^{+\infty} e^{-r^2} rdr$$

$$= 4\times\frac{\pi}{2}\times\frac{1}{2} = \pi$$

所以 $\int_{-\infty}^{+\infty} e^{-x^2}dx = \sqrt{\pi}$

一般地，若 $X \sim N(\mu,\sigma^2), Y \sim N(\mu,\sigma^2)$,

X, Y 相互独立，则

$$E(\max\{X,Y\}) = \mu + \frac{\sigma}{\sqrt{\pi}}$$

$$E(\min\{X,Y\}) = \mu - \frac{\sigma}{\sqrt{\pi}}$$

例11　市场上对某种产品每年的需求量为X吨，$X \sim U[2000,4000]$，每出售一吨可赚3万元，售不出去，则每吨需仓库保管费1万元，问应该生产这中商品多少吨，才能使平均利润最大？

解

$$f_X(x)=\begin{cases}\dfrac{1}{2000}, & 2000 < x < 4000,\\ 0, & \text{其它}\end{cases}$$

设每年生产 y 吨的利润为 Y

显然，$2000 < y < 4000$

$$Y = g(X) = \begin{cases} 3y, & y \le X, \\ 3X - (y - X)\cdot 1, & y > X \end{cases}$$

$$g(x) = \begin{cases} 3y, & y \le x, \\ 4x - y, & y > x \end{cases}$$

$$E(Y) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

$$= \int_{2000}^{y} (4x - y)\frac{1}{2000}dx + \int_{y}^{4000} 3y \frac{1}{2000} dx$$

$$= \frac{1}{2000}(-2y^2 + 14000y - 8\times 10^6)$$

$$\frac{dE(Y)}{dy}=\frac{1}{2000}(-4y+14000)\overset{令}{=}0$$

显然， $$\frac{d^2E(Y)}{dy^2}=-\frac{4}{2000}<0$$

故 $y = 3500$ 时，$E(Y)$最大， $E(Y)= 8250$万元

例12 将 4 个可区分的球随机地放入 4 个盒子中，每盒容纳的球数无限，求空着的盒子数的数学期望.

解一 设 X 为空着的盒子数, 则 X 的概率分布为

X	0	1	2	3
P	$\frac{4!}{4^4}$	$\frac{C_4^1 C_3^1 P_4^2}{4^4}$	$\frac{C_4^2 (C_4^2 + C_2^1 C_4^3)}{4^4}$	$\frac{C_4^1}{4^4}$

$$E(X) = \frac{81}{64}$$

解二 再引入 $X_i, i = 1,2,3,4$

$$X_i = \begin{cases} 1, & 第i盒空, \\ 0, & 其它, \end{cases}$$

$$X = X_1 + X_2 + X_3 + X_4$$

X_i	1	0
P	$\left(\frac{3}{4}\right)^4$	$1-\left(\frac{3}{4}\right)^4$

$$E(X_i) = \left(\frac{3}{4}\right)^4$$

$$E(X) = 4 \cdot \left(\frac{3}{4}\right)^4 = \frac{81}{64}$$

验血方案的选择

为普查某种疾病, n 个人需验血, 可采用两种方法验血：

(1) 分别化验每个人的血, 共需化验 n 次；

(1) 将 k 个人的血混合在一起化验，若化验结果为阴性, 则此 k 个人的血只需化验一次；若为阳性, 则对 k 个人的血逐个化验，找出有病者, 这时 k 个人的血需化验 $k+1$ 次.

设某地区化验呈阳性的概率为 p，且每个人是否为阳性是相互独立的. 试说明选择哪一种方法可以减少化验次数.

解 为简单计，设 n 是 k 的倍数，设共分成 n/k 组

第 i 组需化验的次数为 X_i

X_i	1	$k+1$
P	$(1-p)^k$	$1-(1-p)^k$

$$E(X_i)=(1-p)^k+(k+1)[1-(1-p)^k]$$

$$=(k+1)-k(1-p)^k$$

$$E(X)=\sum_{i=1}^{n/k}E(X_i)=\frac{n}{k}\left[(k+1)-k(1-p)^k\right]$$

$$=n\left[1-\left((1-p)^k-\frac{1}{k}\right)\right]$$

若$\left((1-p)^k-\frac{1}{k}\right)>0$, 则$\boldsymbol{E(X)<n}$

例如，

$n=10000,\ \ p=0.001,\ \ k=10,$

$$E(X)=10000\left[1-\left(0.999^{10}-\frac{1}{10}\right)\right]\approx 1100<<10000$$