

## 基础部分

### 第四课 概率统计

#### 第 4 章 随机变量的数字特征

##### 本章内容

##### §4.1 数学期望与方差

##### §4.2 协方差与相关系数

##### 本章提要 (略, 请见大纲)

##### § 4.1 数学期望与方差

本节引入一个  $r.v.$  的数字特征的概念: 数学期望、矩和方差, 学习它们的性质以及它们间的关系. 熟练掌握它们的计算方法, 也应该熟知重要(常用)分布的期望和方差.

##### [内容精讲]

##### 4.1.1 数字特征的引入与定义

##### 引例

**定义 4.1.1** 1) 设离散型  $r.v.$   $X$  有分布  $\{p_k\}$  且下列级数绝对收敛, 即

$$\sum_k |x_k| p_k < \infty.$$

则称  $X$  的数学期望存在, 其值为  $EX := \sum_k x_k p_k$ .

2) 设连续型  $r.v.$   $X$  有密度分布且下列积分绝对可积, 即

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

则称  $X$  的数学期望存在, 其值为  $EX := \int_{-\infty}^{\infty} x f_X(x) dx$

可将上述两式合写为(实际上是一般化)

$$EX = \int_{-\infty}^{\infty} x dF_X(x). \quad (4.1.1)$$

**注意:** 一个  $r.v.$  的数学期望(只要存在)不再是  $r.v.$ , 而是一个确定的实数——随机因素已经加权平均掉了.

**定义 4.1.2** 设  $r.v.$   $X$  的  $df$  为  $F_X(x)$ ,  $y=g(x)$  是 Borel 可测函数. 如果  $\int_{-\infty}^{\infty} |g(x)| dF_X(x) < \infty$ , 则称  $Y=g(X)$  的数学期望存在, 其值为

$$Eg(X) := \int_{-\infty}^{\infty} g(x) dF_X(x). \quad (4.1.2)$$

一般地, 可建立如下关系.

$$Eg(X) = \int_{-\infty}^{\infty} g(x) dF_X(x) = \int_{-\infty}^{\infty} y dF_{g(X)}(y) = EY.$$

**定义 4.1.3** 设  $r.v.$   $X=(X_1, X_2, \dots, X_n)$  的各分量  $X_i$  的期望存在, 则定义  $X$  的数学期望为  $EX=(EX_1, EX_2, \dots, EX_n)$ . 又设  $Y=g(X_1, X_2, \dots, X_n)$  是  $r.v.$ , 如果  $Y$  积分绝对可积, 则称  $Y=g(X_1, X_2, \dots, X_n)$  的数学期望存在, 其值为

$$Eg(X_1, \dots, X_n) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) dF_X(x_1, \dots, x_n).$$

写为向量形式, 且可建立类似一维时的变换公式

$$\begin{aligned} Eg(X) &= \int_{R_n} g(x) dF_X(x), \quad x=(x_1, x_2, \dots, x_n); \\ &= \int_{-\infty}^{\infty} y dF_{g(X_1, \dots, X_n)}(y) = EY. \end{aligned} \quad (4.1.3)$$

**定义 4.1.4**  $k$  阶矩;  $k$  阶绝对矩;  $k$  阶中心矩和  $k$  阶绝对中心矩; 方差; 专记为  $DX$ , 或  $\text{Var} X$ . 标准差或均方差.

### 4.1.2 期望与方差的性质

以下恒设所涉及的期望存在.

#### 定理 4.1.1

E1)  $E c = c$ ; E2) 线性:  $E(aX + bY) = aEX + bEY$ .

E3) 保界: 设  $a \leq X \leq b$ ,  $a, b \in \mathbb{R}$ , 则  $a \leq EX \leq b$ .

E4) 设诸  $X_i$  独立, 则  $EX_1 X_2 \dots X_n = \prod_{i=1}^n EX_i$ .

E5) 记  $g(x) = E(X - x)^2$ , 则  $g(x)$  在  $x = EX$  取最小值, 即  
 $E(X - EX)^2 \leq E(X - x)^2, \forall x \in \mathbb{R}_1$ .

#### 定理 4.1.2

D1)  $DX = EX^2 - (EX)^2$ .

D2) (切贝雪夫不等式) 设  $X$  的方差存在, 则  $\forall \varepsilon > 0$ ,

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2} \quad \text{或} \quad P(|X - EX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2}$$

D3)  $X = c$  (常数),  $a, b \in \mathbb{R}$ , 的充要条件是  $DX = 0$ .

D4) (线性组合的方差公式)

$$\begin{aligned} D\left(\sum_{i=1}^n c_i X_i\right) &= \sum_{i,j=1}^n c_i c_j E(X_i - EX_i)(X_j - EX_j) \\ &= \sum_{i=1}^n c_i^2 DX_i + 2 \sum_{i < j} c_i c_j E(X_i - EX_i)(X_j - EX_j) \end{aligned}$$

$$\text{当诸 } X_i \text{ 独立时, } D\sum_{i=1}^n c_i X_i = \sum_{i=1}^n c_i^2 DX_i. \quad (4.1.4)$$

特别地,  $D(X_1 \pm X_2) = DX_1 + DX_2$ .

### 4.1.3 重要分布的期望与方差表

表 4.1.1 重要(常用)分布的数学期望及方差简表

分布	记号	数学期望	方差
二项分布	$B(n, p)$	$np$	$npq$
几何分布	$Ge(p)$	$1/p$	$q/p^2$
Poisson 分布	$P(\lambda)$	$\lambda$	$\lambda$
均匀分布	$U(a, b)$	$(a+b)/2$	$(b-a)^2/12$
指数分布	$Ex(\lambda)$	$1/\lambda$	$1/\lambda^2$
正态分布	$N(\mu, \sigma^2)$	$\mu$	$\sigma^2$

#### [ 典型例题 ]

#### ● 矩的基本计算

例 4.1.1 设  $X \sim B(n, p)$ , 求  $EX$  和  $DX$ .

解 I 解 II

例 4.1.2 两名射手各向自己的靶独立射击, 直到有一次命中时该射手才(立即)停止射击. 如第  $i$  名射手每次命中概率  $p_i (0 < p_i < 1)$ ,  $i = 1, 2$ . 求两射手均停止射击时脱靶(未命中)总数的数学期望.

$$\left[ EX = \frac{1}{p_1} + \frac{1}{p_2} - 2 \right]$$

例 4.1.3 设  $X \sim N(\mu, \sigma^2)$ , 求  $EX$  和  $DX$ .

解 I (定义法) 由 (4.1.1) 及  $X \sim N(\mu, \sigma^2)$ ,

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

令  $(x-\mu)/\sigma = t$ , 得

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

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

令  $(x-\mu)/\sigma = t$ , 得

$$\begin{aligned} DX &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt = -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t de^{-t^2/2} \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = \sigma^2 \end{aligned}$$

解 II (标准化法) 因  $X$  的标准化  $X^* \sim N(0, 1)$ ,

$$EX^* = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

其密度关于  $y$  轴对称, 从而上一积分  $= 0$ ,

故由期望的线性性质  $EX = E(\sigma X^* + \mu) = \mu$ .

注意  $EX^* = 0$  而  $EX^{*2} = 1$ . 故  $DX^* = 1$ .

由  $r.v$  标准化的定义及方差性质

$$DX = D(\sigma X^* + \mu) = \sigma^2 DX^* + D\mu = \sigma^2. \quad \square$$

注 设  $Y \sim N(0, 1)$ , 则

$$\begin{cases} EY^{2n+1} = 0, \\ EY^{2n} = (2n-1)!! = (2n-1)(2n-3)\cdots 3 \cdot 1. \end{cases}$$

例 4.1.4 设  $X \sim P(\lambda)$ , 求  $DX$ .

● 综合题

例 4.1.5 设某网络服务器首次失效时间  $\sim \text{Ex}(\lambda)$ , 现随机购得 4 台中, 求下列事件概率:

1) 事件  $A$ : 至少有一台其寿命  $\geq$  此类服务器期望寿命.

2) 事件  $B$ : 有且仅有一台寿命  $<$  此类服务器期望寿命.

解 1)  $P(A) = 0$ .

2)  $X \sim \text{Ex}(\lambda)$ , 则  $EX = 1/\lambda$ ,

一台服务器的寿命  $X$  小于  $1/\lambda$  的概率为

$$p_0 = \int_0^{1/\lambda} \lambda e^{-\lambda x} dx = 1 - e^{-1} \approx 1 - 0.3679 = 0.6321.$$

4 台服务器中寿命小于  $1/\lambda$  的台数  $Y \sim B(4, p_0)$ , 故所求概率为  $C_4^1 p_0 (1-p_0)^3 = 4e^{-3}(1-e^{-1}) \approx 0.1259$ .

例 4.1.6 将  $n$  只球 (1~ $n$  号) 随机地放进  $n$  只盒子 (1~ $n$ ) 中去, 一只盒子装一只球. 若一只球装入与球同号的盒子中, 称为一个配对. 记  $X$  为总的配对数, 求  $EX$ .

例 4.1.7 设  $X_1, X_2, \dots, X_n \text{ iid}, \sim N(\mu, \sigma^2)$ , 记  $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ . 令  $Y_k := X_k - \bar{X}$ ,

求  $DY_k, k=1, 2, \dots$ .

解  $(1-\frac{1}{n})X_1, \frac{1}{n}X_2, \dots, \frac{1}{n}X_n$  独立且都有正态分布. 由和的方差性质,

$$DY_1 = D(X_1 - \bar{X}) = D\left(\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{i=2}^n X_i\right)$$

$$= \left(\frac{n-1}{n}\right)^2 \sigma^2 + \left(-\frac{1}{n}\right)^2 (n-1)\sigma^2 = \frac{n-1}{n}\sigma^2$$

故  $DY_k = \frac{n-1}{n}\sigma^2$ ,  $k=1, 2, \dots$ .  $\square$

### ● 函数的期望

**例 4.1.8** 设风速  $V \sim U_{(0, a)}$ , 飞机机翼受到的正压力  $W$  是  $V$  的函数:  $W = kV^2$  ( $k>0$ , 常数), 求  $EW$ . 【 $ka^2/3$ 】

**例 4.1.9** 设  $r.v. (X, Y)$  的 pdf 为  $f(x, y) = \begin{cases} x+y, & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{otherwise.} \end{cases}$  试求  $EXY$ .

$$EXY = \iint_{R_2} xyf(x, y)dxdy = \int_0^1 \int_0^1 xy(x+y)dxdy$$

【  $EXY=1/3$ 】

$$= 2 \int_0^1 x^2 dx \int_0^1 y dy = 2 \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{3}$$

### ● 应用

**例 4.1.10** 按季节出售的某种应时商品, 每售出一公斤获利润  $b$  元. 如到季末尚有剩余商品, 则每公斤净亏损  $c$  元. 设某商店在季度内这种商品的销售量  $X$  (以公斤计) 是一  $r.v.$ , 在区间  $[s_1, s_2]$  上服从均匀分布. 为使商店所获得利润的数学期望最大, 问商店应进多少货?

**解** 以进货  $s$  时所得利润记为  $g_s(X)$ , 则

$$g_s(X) = \begin{cases} bX - c(s - X) & s_1 < X \leq s \\ sb & s < X \leq s_2 \end{cases}$$

$X$  的 pdf 为  $f_X(x) = \frac{1}{s_2 - s_1} I(s_1 < x \leq s_2)$

于是

$$E[g_s(X)] = \int_{s_1}^{s_2} g_s(x) \cdot \frac{1}{s_2 - s_1} dx$$

$$= \int_{s_1}^s [bx - c(s - x)] \frac{1}{s_2 - s_1} dx + \int_s^{s_2} sb \frac{1}{s_2 - s_1} dx$$

$$= \left[ -\frac{b+c}{2} s^2 + (cs_1 + bs_2)s - \frac{b+c}{2} s_1^2 \right] / (s_2 - s_1).$$

令  $(E[g_s(X)])' = 0$ , 解得  $s = (cs_1 + bs_2) / (b+c)$ .

**例 4.1.11 (不等式应用)** 设 Ber-试验的参数  $p=0.75$ . 问至少需要进行多少次这种试验, 才能使频率在 0.74 到 0.76 之间的概率至少为 0.90? (解略: 18750)

## s3.2 协方差与相关系数

本节讨论刻画两个  $r.v.$  之间相互关系的数字特征: 协方差和相关系数. 正确理解这两个数字特征的概念、相互关系以及性质. 熟练掌握它们的计算方法.

特别注意相关系数对刻画两个  $r.v.$  间线性相依程度的作用, 以及不相关和独立性之间的联系和差别, 掌握不相关的充要条件.

恒设所涉及的矩存在, 也即相应的积分是绝对收敛的.

### 4.2.1 协方差和相关系数的定义与性质

#### 1. 协方差和相关系数的定义

**定义 4.2.1** 称  $E(X - EX)(Y - EY)$  为  $r.v. X$  与  $Y$  的协方差, 记为  $\text{Cov}(X, Y)$ , 即

$$\text{Cov}(X, Y) = E(X - EX)(Y - EY). \quad (4.2.1)$$

设  $DX>0$ 、 $DY>0$ ，并令  $X^* = \frac{X - EX}{\sqrt{DX}}$ ， $Y^* = \frac{Y - EY}{\sqrt{DY}}$ ，称

$$r_{XY} = EX^*Y^* = \frac{\text{Cov}(X, Y)}{\sqrt{DX}\sqrt{DY}} \quad (4.2.2)$$

为  $rvX$  与  $rvY$  的相关系数。如  $r_{XY}=0$ ，则称  $X$  与  $Y$  不相关。

注意： $\text{Cov}(X, Y) = r_{XY} \sigma_X \sigma_Y$

## 2. 协方差和相关系数的性质

### 定理 4.2.1

1)  $\text{Cov}(X, Y) = EXY - EXEY$ 。

2) 对称性  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

3) 对单个  $rv$  的线性性，例如  $\text{Cov}(aX_1 + bX_2, Y) = a \text{Cov}(X_1, Y) + b \text{Cov}(X_2, Y)$ ， $a, b$  是常数；

4) 和的方差

$$\begin{aligned} D\left(\sum_{i=1}^n c_i X_i\right) &= \sum_{i,j=1}^n c_i c_j \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n c_i^2 DX_i + 2 \sum_{i<j}^n c_i c_j \text{Cov}(X_i, X_j) \end{aligned}$$

特别，诸  $X_i$  不相关时， $D\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 DX_i$ 。

### 定理 4.2.2

1)  $|r_{XY}| \leq 1$ 。  $r_{XY} = r_{X^*Y^*}$

2)  $|r_{XY}| = 1 \Leftrightarrow Y$  是  $X$  的线性函数 ( $a.e$ )，即存在常数  $a$  ( $\neq 0$ ) 和  $b$  使  $P(Y=aX+b) = 1$ 。

相关系数的一个极端值  $|r_{XY}| = 1$  表明  $Y$  是  $X$  的线性函数 (可进一步得到：相关系数的绝对值越接近于 1，则  $Y$  越接近于是  $X$  的线性函数)。那么，另外一个极端值  $r_{XY}=0$ 。

### 定理 4.2.3 下列命题等价

- 1)  $X$  与  $Y$  不相关；
- 2)  $\text{Cov}(X, Y) = 0$ ；
- 3)  $EXY = EXEY$ ；
- 4)  $D(X \pm Y) = DX + DY$ 。

独立与不相关的关系。

### [ 典型例题 ]

## ● 协方差与相关系数的基本计算

例 4.2.1 设  $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ ，求相关系数。

例 4.2.2 某箱装有 100 件产品，其中一、二和三等品分别为 80、10 和 10 件，现在随机抽取一件，令

$$X_i = \begin{cases} 1, & \text{如若抽到 } i \text{ 等品}, \quad i=1, 2, 3. \\ 0, & \text{其它} \end{cases}$$

试求 1)  $X_1$  和  $X_2$  的联合分布。 2)  $X_1$  和  $X_2$  的相关系数。

解 1) 设  $A_i =$  “抽到  $i$  等品”， $i=1, 2, 3$ 。

由题知  $A_1, A_2, A_3$  不相容，且  $P(A_1) = 0.8$ ， $P(A_2) = P(A_3) = 0.1$

易见

$$\begin{aligned} P(X_1=0, X_2=0) &= P(A_3) = 0.1, & P(X_1=0, X_2=1) &= P(A_2) = 0.1, \\ P(X_1=1, X_2=0) &= P(A_1) = 0.8, & P(X_1=1, X_2=1) &= P(\emptyset) = 0. \end{aligned}$$

2)  $EX_1 = 0.8$ ， $EX_2 = 0.1$ ， $DX_1 = 0.16$ ， $DX_2 = 0.09$ 。

$$EX_1 X_2 = P(X_1 X_2 = 1) = P(X_1 = 1, X_2 = 1) = 0,$$

$$\text{cov}(X_1, X_2) = EX_1 X_2 - EX_1 EX_2 = -0.08.$$

$$\rho = \frac{\text{cov}(X_1, X_2)}{\sqrt{DX_1} \cdot \sqrt{DX_2}} = \frac{-0.08}{\sqrt{0.16 \times 0.09}} = -\frac{2}{3}.$$

## ● 协方差与相关系数的性质

**例 4.2.3** 将一枚硬币重复掷  $n$  次, 以  $X$  和  $Y$  分别表示正面向上和反面向上的次数, 则  $X$  和  $Y$  的相关系数等于

- A)  $-1$ . B)  $0$ . C)  $1/2$ . D)  $1$ .

**例 4.2.4** 设  $X$  和  $Y$  的方差存在且大于  $0$ , 则  $D(X+Y)=D(X)+D(Y)$  是  $X$  和  $Y$

- A) 不相关的充分条件, 但不是必要条件;  
B) 独立的必要条件, 但不是充分条件;  
C) 不相关的充分必要条件;  
D) 独立的充分必要条件.

● **综合题**

**例 4.2.5** 设  $A, B$  是二随机事件;  $r_{XY}$

$$X = \begin{cases} 1 & \text{若 } A \text{ 出现} \\ -1 & \text{若 } A \text{ 不出现} \end{cases}, \quad Y = \begin{cases} 1 & \text{若 } B \text{ 出现} \\ -1 & \text{若 } B \text{ 不出现} \end{cases}.$$

试证明  $r_{XY}$  和  $Y$  不相关的充分必要条件是  $A$  与  $B$  独立.

**解 I** (略. 见 (00-3-12[8]) = (00-4-12[8]) 解)

**解 II** 引入 0-1 变量 (Ber-计数变量)

**例 4.2.6** 设  $r_{\vec{V}}(X, Y) \sim U_0, G = \{(x, y) \mid 0 \leq x \leq 2, 0 \leq y \leq 1\}$ ,

$$\text{记 } U = \begin{cases} 0, & \text{若 } X \leq Y \\ 1, & \text{若 } X > Y \end{cases}, \quad V = \begin{cases} 0, & \text{若 } X \leq 2Y \\ 1, & \text{若 } X > 2Y \end{cases}. \quad \text{求 } U \text{ 和 } V \text{ 的相关系数 } r.$$

**解** 由题设可得

$$P\{X \leq Y\} = \frac{1}{4}, \quad P\{X > 2Y\} = \frac{1}{2}, \quad P\{Y < X \leq 2Y\} = \frac{1}{4}.$$

(1)  $(U, V)$  有四个可能取值:  $(0, 0), (0, 1), (1, 0), (1, 1)$ .

$$P\{U=0, V=0\} = P\{X \leq Y, X \leq 2Y\} = P\{X \leq Y\} = \frac{1}{4};$$

$$P\{U=0, V=1\} = P\{X \leq Y, X > 2Y\} = 0;$$

$$P\{U=1, V=0\} = P\{X > Y, X \leq 2Y\} = P\{Y < X \leq 2Y\} = \frac{1}{4};$$

$$P\{U=1, V=1\} = 1 - \frac{1}{4} - \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

(2) 以上可见  $UV$  以及  $U$  和  $V$  的分布为

$$UV \sim \begin{bmatrix} 0 & 1 \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}; \quad U \sim \begin{bmatrix} 0 & 1 \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}; \quad V \sim \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

于是  $EU = 3/4, DV = 3/16; EV = 1/2, DV = 1/4; EUV = 1/2;$

$$\text{cov}(U, V) = E(UV) - EU \cdot EV = 1/8;$$

$$r = \frac{\text{cov}(U, V)}{\sqrt{DU \cdot DV}} = \frac{1}{\sqrt{3}}.$$