第 4 章教学要求：

1. 理解随机变量的数学期望与方差的概念，掌握其性质与计算方法.

2. 了解 0-1 分布、二项分布、泊松分布、几何分布、超几何分布、均匀分布、指数分布和正态分布的数学期望与方差.

3. 了解协方差、相关系数与矩的概念及性质，并会计算.

4. 了解契比雪夫不等式.

第 4 章　随机变量的数字特征

什么是数字特征？

数字特征是与随机变量有关的一些数值，它们虽然不能完整地描述随机变量(完整地描述随机变量的统计特性的是分布律和概率密度或分布函数)，但可以从总体上概括地反映出随机变量在某些方面的重要特征. 例如，从总体上说，随机变量取值的平均值是多少？随机变量的取值与其平均值的差的情况？随机变量与随机变量的关系如何？

描述随机变量这些总体情况的数值即为随机变量的数字特征.

随机变量的数字特征包括数学期望、方差、相关系数和矩等.

§4.1 数学期望

数学期望是随机变量的取值的平均值.

4.1.1 离散型的随机变量的数学期望

例 4.1 (P_{113})

例 4.2 (P_{113})

定义 4.1　设 X 是离散型随机变量，分布律为 $P\{X=x_i\}=p_i\ (i=1,2,\cdots)$. 若级数 $\sum_i x_i p_i$ 绝对收敛，则称该级数的和为 X 的数学期望，记为 E(X)，即

$$E(X)=\sum_i x_i p_i\ . \qquad (4.1)$$

数学期望简称**期望**或**均值**.

例 4.3（P_{114}） 如果 $X \sim B(1,p)$，则 $E(X)=p$.

例 4.4（P_{114}） 如果 $X \sim P(\lambda)$，则 $E(X)=\lambda$.

例 4.5（P_{114}） 如果 $X \sim G(p)$，则 $E(X)=\dfrac{1}{p}$.

离散型随机变量的函数的期望有如下定理.

定理 4.1 （1）设 X 是离散型随机变量，分布律为 $P\{X=x_i\}=p_i\ (i=1,2,\cdots)$， $Y=g(X)$（g 是连续函数），若级数 $\sum_i g(x_i)p_i$ 绝对收敛，则有

$$E(Y)=\sum_i g(x_i)p_i\ . \tag{4.2}$$

（2）设(X,Y)是二维离散型随机变量，分布律为 $P\{X=x_i,Y=y_j\}=p_{ij},i,j=1,2,\cdots$，$Z=g(X,Y)$（g 是连续函数），若级数 $\sum_i\sum_j g(x_i,y_j)p_{ij}$ 绝对收敛，则有

$$E(Z)=\sum_i\sum_j g(x_i,y_j)p_{ij}\ . \tag{4.3}$$

这个定理的结论可以推广到两个以上随机变量的函数的情形.

例 4.6（P_{115}）

例 4.7（P_{115}）

4.1.2 连续型随机变量的数学期望

定义 4.2 设 X 是连续型随机变量，密度函数为 $f(x)$. 若积分 $\int_{-\infty}^{+\infty}xf(x)dx$ 绝对收敛，则称该积分的值为 X 的数学期望，记为 $E(X)$，即

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx\ . \tag{4.4}$$

例 4.8（P_{116}） 如果 $X \sim U(a,b)$，则 $E(X)=\dfrac{a+b}{2}$.

例 4. 9 (P_{116})　如果 $X \sim E(\lambda)$，则 $E(X) = \frac{1}{\lambda}$.

例 4. 10 (P_{116})　如果 $X \sim N(0,1)$，则 $E(X) = 0$.

连续型随机变量的函数的期望有如下定理.

定理 4. 2　(1) 设 X 是连续型随机变量，密度函数为 f(x)，$Y = g(X)$ (g 是连续函数). 若积分 $\int_{-\infty}^{+\infty} g(x)f(x)dx$ 绝对收敛，则有

$$E(Y) = \int_{-\infty}^{+\infty} g(x)f(x)dx . \tag{4. 5}$$

(2) 设(X,Y)是二维连续型随机变量，密度函数为 f(x,y)，$Z = g(X,Y)$ (g 是连续函数). 若积分 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$ 绝对收敛，则有

$$E(Z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy . \tag{4. 6}$$

这个定理的结论可以推广到两个以上随机变量的函数的情形.

例 4. 11 (P_{117})

例 4. 12 (P_{117})

4. 1. 3 数学期望的性质

性质 1　若 c 是常数，则 $E(c) = c$.

性质 2　若 $a < X < b$，则 $a < E(X) < b$.

性质 3　设 X 是随机变量，c 是常数，则

$$E(cX) = cE(X). \qquad \text{(线性性)}$$

性质 4　设 X,Y 是随机变量，则

$$E(X+Y) = E(X) + E(Y). \qquad \text{(线性性)}$$

推论 设 $X_1,X_2,\cdots,X_n$ 是 n 个随机变量， $c_1,c_2,\cdots,c_n$ 为任意常数，则

$$E(c_1X_1+c_2X_2+\cdots+c_nX_n)=c_1E(X_1)+c_2E(X_2)+\cdots+c_nE(X_n).$$

性质 5 设 X,Y 是随机变量，若 X,Y 相互独立，则

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

性质 5 可以推广到多个随机变量的情形：若随机变量 $X_1,X_2,\cdots,X_n$ 相互独立，则

$$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n).$$

例 4.13 (P_{118}) 如果 $X\sim N(\mu,\sigma^2)$，则 $E(X)=\mu$.

例 4.14 (P_{118}) 如果 $X\sim B(n,p)$，则 $E(X)=np$.

例 4.15 (P_{118}) 如果 $X\sim H(N,M,n)$，则 $E(X)=\dfrac{nM}{N}$.

例 4.16 (P_{119})

4.1.4 数学期望的应用

例 4.17 (P_{119})

例 4.18 (P_{120})

例 4.19 (P_{121})

作业 (P_{122})：2.-3. 5. 7. 9.-13. 1. 6. 11. 14.-15.

§4.2 方差和标准差

4.2.1 方差

方差是随机变量的取值的稳定性的度量值，它从另一个角度来刻画和区别随机变量.

看两个例子.

例 4. 20 (P_{124})

例 4. 21 (P_{124})

总结：这两个例子的一个共同点是，在对两个随机变量作比较时，一旦它们的均值相等，那就比较两者的稳定性.

“稳定性”是对随机变量的取值集中(或分散)的一种度量. 取值的集中(或分散)是相对均值而言的，用|X-E(X)|表达. 集中(或分散)程度的整体反映用平均值 E(|X-E(X)|)来表达. E(|X-E(X)|)大，则说 X 的稳定性差；反之，则说 X 的稳定性好.

考虑到带绝对值运算不方便，以及 $(X-E(X))^2$ 也是对随机变量的取值集中(或分散)的一种度量，所以一般用 $E[(X-E(X))^2]$ 来度量随机变量 X 的稳定性.

定义 4. 3 设 X 是一个随机变量，若 $E[(X-E(X))^2]$ 存在，则称为 X 的**方差**，记为 D(X)，即

$$D(X)=E[(X-E(X))^2]. \tag{4. 7}$$

而 $\sqrt{D(X)}$ 称为 X 的**标准差**或**均方差**，记为 $\sigma(X)$ 或 σ_X.

D(X)小，表示 X 的取值比较集中在 E(X)的附近；D(X)大，表示 X 的取值相对 E(X)比较分散.

根据期望的性质，可以导出：

$$D(X)=E(X^2)-E^2(X). \tag{4. 8}$$

* ①该公式中 $E(X),E(X^2),E[(X-E(X))^2]$ 的计算，须依 X 的类型分别按离散型的或连续型的随机变量的期望公式或变量函数的期望公式计算.

②已知 $D(X),E(X),E(X^2)$ 中的任意两个就可以求出另一个.

例 4. 22 (P_{126})

例 4. 23 (P_{126}) 如果 $X\sim B(1,p)$，则 $D(X)=p(1-p)$.

例 4. 24 (P_{126})　如果 $X \sim P(\lambda)$，则 $D(X) = \lambda$.

例 4. 25 (P_{127})　如果 $X \sim G(p)$，则 $D(X) = \frac{1-p}{p^2}$.

例 4. 26 (P_{127})　如果 $X \sim U(a,b)$，则 $D(X) = \frac{1}{12}(b-a)^2$.

例 4. 27 (P_{128})　如果 $X \sim E(\lambda)$，则 $D(X) = \frac{1}{\lambda^2}$.

例 4. 28 (P_{128})　如果 $X \sim N(0,1)$，则 $D(X) = 1$.

4. 2. 2 方差的性质

性质 0　$D(X) \ge 0$.

性质 1　设 c 是常数，则 $D(c) = 0$.

性质 2　设 X 是随机变量，c 是常数，则 $D(cX) = c^2 D(X)$.

性质 3　设 X,Y 是随机变量，则

$$D(X+Y) = D(X) + D(Y) + 2E[(X-E(X))(Y-E(Y))].$$

特别地，当 $Y = c$（c 是常数）时，$D(X+c) = D(X)$；当 X,Y 相互独立时，$D(X+Y) = D(X) + D(Y)$.

推论　若 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量，$c_1, c_2, \cdots, c_n, b$ 是任意常数，则

$$D(c_1X_1 + c_2X_2 + \cdots + c_nX_n + b) = c_1^2 D(X_1) + c_2^2 D(X_2) + \cdots + c_n^2 D(X_n).$$

性质 4　设 X 是随机变量，则 $D(X) = 0$ 的充分必要条件是存在常数 c，使得 $P\{X = C\} = 1$，而且 $C = E(X)$.

例 4. 29 (P_{129})　如果 $X \sim B(n,p)$，则 $D(X) = np(1-p)$.

例 4. 30 (P_{129})　如果 $X \sim N(\mu, \sigma^2)$，则 $D(X) = \sigma^2$.

例 4. 31 (P_{130})

例 4. 32 (P_{130})

例 4. 33 (P_{130})　**标准化随机变量**

4.2.3　切比雪夫不等式

定理 4.3　若随机变量 X 具有数学期望 $E(X)=\mu$，方差 $D(X)=\sigma^2$，则对于任意正数 ε，有

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}. \tag{4.9}$$

＊　(4.9)式给出了对事件 $P\{|X-E(X)|\geq\varepsilon\}$ 概率的上限估计.

＊　由(4.9)易得

$$P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2}.$$

该式给出了对事件 $P\{|X-E(X)|<\varepsilon\}$ 概率的下限估计.

例 4.34(P_{132})

例 4.35(P_{132})

作业(P_{132})：2.　4.　5.-7.　3.　8.

§4.3 协方差和相关系数

4.3.1 协方差及其性质

我们有必要讨论两个随机变量之间的关系与关联度. 例如，身高 H 和体重 W 是有关联的，代表吸烟与健康程度的两个随机变量 S,K 也是有关联的，我们需要知道 H 与 W、S 与 K 的关联度，这很有意义.

反映两个随机变量有关联的一个量是**协方差**.

定义 4.4　设(X,Y)为二维随机变量，称 $E[(X-E(X))(Y-E(Y))]$ 为 X 与 Y 的**协方差**，记为

Cov(X,Y)，即

$$\mathrm{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))]\,. \tag{4.10}$$

根据期望的性质，可以导出：

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)\,. \tag{4.11}$$

如果令 Y=X，则

$$\mathrm{Cov}(X,X)=D(X)\,.$$

例 4.36 (P_{133})
例 4.37 (P_{134})

设 X,Y,Z 为随机变量，a,b,c 为常数，则

性质 1 $\mathrm{Cov}(X,Y)=\mathrm{Cov}(X,Y)$. 特别地，$\mathrm{Cov}(X,c)=0$.

性质 2 $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$ （a,b 是常数）.

性质 3 $\mathrm{Cov}(X+Y,Z)=\mathrm{Cov}(X,Z)+\mathrm{Cov}(Y,Z)$.

这一性质可以推广到多个随机变量的情形：

$$\mathrm{Cov}(\sum_{i=1}^{n}X_i,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)+\cdots+\mathrm{Cov}(X_n,Y)\,.$$

4.3.2 相关系数及其性质

相关系数是两个随机变量关联程度的度量值.

协方差 Cov(X,Y)的值的大小会受 X,Y 量纲的影响，所以它不能很好地反映两个随机变量的关联程度. 但如果把 X,Y 标准化，则可以消除这个影响.

设

$$X^* = \frac{X-E(X)}{\sqrt{D(X)}},\quad Y^* = \frac{Y-E(Y)}{\sqrt{D(Y)}},$$

则 $Cov(X^*,Y^*) = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 的值与 X,Y 的量纲无关，因而能准确地反映 X,Y 之间的关联程度.

定义 4.5 设(X,Y)为二维随机变量，当 $D(X)>0, D(Y)>0$ 时，数 $\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 称为 X 与 Y 的**相关系数**，记为 ρ(X,Y)或 ρ_{XY}，即

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}. \tag{4.13}$$

引理（柯西-许瓦兹不等式） 设 X,Y 是随机变量，$E(X^2)<+\infty, E(Y^2)<+\infty$，则有

$$E^2(XY) \le E(X^2)\cdot E(Y^2).$$

相关系数有如下基本性质：

性质 1 $\rho(X,Y)=\rho(Y,X)$.

性质 2 $|\rho(X,Y)| \le 1$.

性质 3 |ρ(X,Y)|=1 的充分必要条件是存在常数 a,b，使得 $P\{Y=aX+b\}=1$. 且当 a>0 时，ρ(X,Y)=1；当 a<0 时，ρ(X,Y)=−1.

性质 3 说明 ρ(X,Y)是对随机变量 X,Y 之间线性关系程度的度量．当|ρ(X,Y)|=1 时，X,Y **以概率 1 成立线性关系**．ρ(X,Y)=1，称 X,Y **正线性相关**；ρ(X,Y)=−1，称 X,Y **负线性相关**．当 ρ(X,Y)=0 时，称 X,Y **不（线性）相关**．|ρ(X,Y)|接近于 1，表示 X,Y 有较紧密的线性关系；|ρ(X,Y)|接近于 0，表示 X,Y 间的线性关系程度很弱.

独立与不相关都是随机变量间联系“薄弱”的一种反映. 随机变量间不存在任何关系则独立，不相关则指随机变量间不存在线性关系.

显然，若随机变量 X 与 Y 独立，则 ρ(X,Y)=0；反之，不成立.

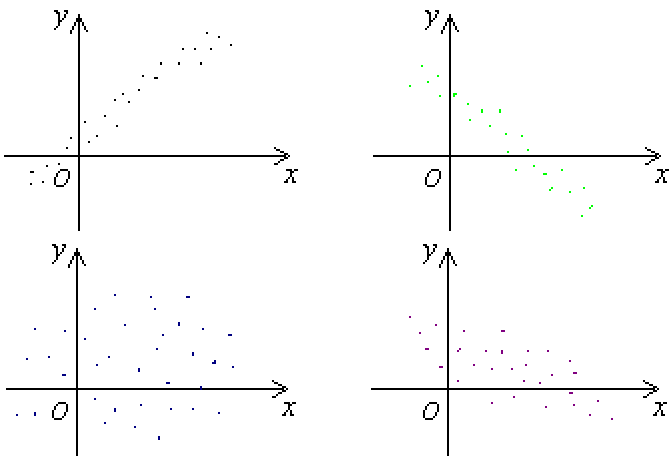
当 $D(X)>0, D(Y)>0$ 时，

$$
\begin{aligned}
& X\text{ 与 }Y\text{ 不相关} \\
\Leftrightarrow\ & \rho(X,Y)=0 \\
\Leftrightarrow\ & \sigma_{XY}=0 \\
\Leftrightarrow\ & E(UV)=E(X)E(Y) \\
\Leftrightarrow\ & D(X+Y)=D(X)+D(Y) \\
\Leftrightarrow\ & D(X-Y)=D(X)+D(Y).
\end{aligned}
$$

** 若 X 与 Y 相关，则 X 与 Y 肯定不相互独立；若 X 与 Y 不相关，则不能保证 X 与 Y 相互独立.

例如，在例 4.36 中，ρ(X,Y)=0，但 X,Y 不相互独立. 事实上，$Y=X^2$.

例 如图所示，试判断随机变量 X 与 Y 的线性关系程度.



例 抛掷一枚硬币 n 次，其中出现正面朝上的次数为 X，出现反面朝上的次数为 Y，求 X 与 Y 的相关系数 ρ(X,Y).

例 4.39（P_{138}）

例 4.40（P_{139}）

作业（P_{141}）：1.　3.−4.　6.　5.　7.−9.

§4.4 矩

设 X 是随机变量，称 $E(X^k)(k=1,2,\cdots)$ 为 X 的 k **阶原点矩**；称 $E[(X-E(X))^k](k=1,2,\cdots)$ 为 X 的 **k 阶中心矩**，其中 k 为正整数.

例 4.41（P_{142}）

作业（P_{143}）：1.　2.