概率论与数理统计
第四章 随机变量的数字特征

金玲飞

复旦大学软件学院
Email: lfjin@fudan.edu.cn

2018.11.20

目录

- 数学期望的概念，数学期望的性质，随机变量的函数的数学期望。
- 方差的概念，方差的性质;
- 协方差与相关系数的概念及性质。
- 切比雪夫不等式及其应用。

4.1 数学期望

离散型随机变量的数学期望

定义 (4.1.1 Expectation)

设离散型随机变量X的分布律为

X	x_1	x_2	$\cdots$	x_k	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

则当

$$\sum_{k\geq 1}|x_k|p_k < \infty$$

时，称$E(X)=\sum_{k\geq 1} x_k p_k$为$X$的数学期望，简称期望，记为$E(X)$。如果

$$\sum_{k\geq 1}|x_k|p_k = \infty$$

则称X的数学期望不存在。

- 若c为常数，$E(c) = c$。

- 若c为常数，$E(c)=c$。

例子 (4.1.2)

设X是服从参数为p的伯努利分布：$P(X=0)=1-p$，$P(X=1)=p$。则X的期望为

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

- 若c为常数，$E(c) = c$。

例子 (4.1.2)

设X是服从参数为p的伯努利分布：$P(X = 0) = 1 - p$，$P(X = 1) = p$。则X的期望为

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

期望与概率的关系

设A为随机变量，令

$$I_A(\omega) = \begin{cases} 0, & \omega \notin A \\ 1, & \omega \in A \end{cases}$$

则I_A为随机变量，且它的期望是$E(I_A) = P(A)$。

几个常见的分布的期望

例子 (4.1.3)

设随机变量X服从参数为λ的泊松分布：
$p_k = \frac{\lambda^k}{k!}e^{-\lambda}$，$k = 0, 1, 2, \cdots$。求$E(X)$.

几个常见的分布的期望

例子 (4.1.3)

设随机变量X服从参数为λ的泊松分布：
$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$，$k = 0, 1, 2, \cdots$。求$E(X)$.

$$\sum_{k=0}^{\infty} |k| p_k = \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}$$

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

泊松分布的参数λ就是它的期望。

例子 (4.1.4)

设随机变量X服从二项分布$B(n,p)$，求$E(X)$。

例子 (4.1.4)

设随机变量X服从二项分布$B(n,p)$，求$E(X)$。

$$E(X) = np$$

例子 (4.1.5)

设随机变量X 取$x_k = (-1)^k \frac{2^k}{k}$的概率为$p_k = \frac{1}{2^k}, k = 1, 2, \ldots$，问$X$的数学期望存在不？

例子 (4.1.5)

设随机变量X 取$x_k=(-1)^k\frac{2^k}{k}$的概率为$p_k=\frac{1}{2^k}, k=1,2,\ldots$，问$X$的数学期望存在不？

例子 (4.1.6)

设随机变量X取非负整值，且$P(X=n)=\frac{AB^n}{n!}$。若已知$E(X)=a$，求常数A, B。

例子 (4.1.7)

袋中装有N只球，但其中白球数为随机变量，只知其期望为n。试求从该袋中任取一球得到白球的概率。

例子 (4.1.7)

袋中装有N只球，但其中白球数为随机变量，只知其期望为n。试求从该袋中任取一球得到白球的概率。

例子 (4.1.8)

设随机变量X取值$1, 2, 3, \cdots$，且$E(X)$存在，则

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

连续型随机变量的数学期望

定义 (4.1.2)

设X是概率密度为$f(x)$的连续型随机变量。如果

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

则称X的数学期望存在，并称$\int_{-\infty}^{\infty} x f(x) dx$为$X$的数学期望，简称期望，记为$E(X)$。如果

$$\int_{-\infty}^{\infty} |x| f(x) dx = \infty$$

则称X的数学期望不存在。

例子 (4.1.9)

设随机变量X服从均匀分布$U[a, b]$，求$E(X)$。

例子 (4.1.9)

设随机变量X服从均匀分布$U[a,b]$，求$E(X)$。

$$E(X)=\frac{a+b}{2}$$

例子 (4.1.9)

设随机变量X服从均匀分布$U[a, b]$，求$E(X)$。

$$E(X) = \frac{a+b}{2}$$

例子 (4.1.10)

设随机变量X服从参数为λ的指数分布，求$E(X)$。

例子 (4.1.9)

设随机变量X服从均匀分布$U[a,b]$，求$E(X)$。

$$E(X)=\frac{a+b}{2}$$

例子 (4.1.10)

设随机变量X服从参数为λ的指数分布，求$E(X)$。

$$E(X)=\frac{1}{\lambda}$$

平均寿命为$1/\lambda$。

例子 (4.1.11)

设随机变量X服从正态分布$N(\mu, \sigma^2)$，求$E(X)$。

例子 (4.1.11)

设随机变量X服从正态分布$N(\mu,\sigma^2)$，求$E(X)$。

例子 (4.1.12 Cauchy distribution)

若随机变量X的概率密度为

$$f(x)=\frac{1}{\pi}\frac{1}{1+x^2}$$

问$E(X)$是否存在？

随机变量函数的数学期望

定义 (4.1.1)

设X为随机变量，$g(x)$为（分段）连续函数或（分段）单调函数，令$Y = g(X)$。

(1) 若X为离散型随机变量，分布律为$P(X = x_k) = p_k$，$k = 1, 2, \ldots$，则当$\sum_{k \geq 1} |g(x_k)| p_k < \infty$时，$Y = g(X)$的数学期望存在，且有

$$E(Y) = E(g(X)) = \sum_{k \geq 1} g(x_k) p_k$$

定义 (4.1.1)

(2) 若X为连续型随机变量，概率密度为$f(x)$，则当$\int_{-\infty}^{\infty}|g(x)|f(x)dx < \infty$时，$Y = g(X)$的数学期望存在，且有

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

例子 (4.1.15)

设随机变量X的概率密度为

$$f(x)=\begin{cases}\frac{2x}{\pi^2}, & 0\le x\le\pi\\ 0, & \text{其他}\end{cases}$$

求$E(\sin X)$。

推论

设(X,Y)为二维离散型随机变量，$Z=g(x,y)$为（分块）二元连续函数。其中$P_{ij}, f(x,y)$分别为二维离散型随机变量和连续型随机变量的分布律和概率密度。则其数学期望为

$$E(Z)=E(g(X,Y))=\begin{cases}\sum_{i\geq 1}\sum_{j\geq 1} g(x_i,y_j)p_{ij}, & \text{离散型}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy & \text{连续型}\end{cases}$$

其中对应无穷级数和广义积分均绝对收敛，*i.e.*,

$$\sum_{i\geq 1}\sum_{j\geq 1}|g(x_i,y_j)|p_{ij}<\infty$$

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

例子 (4.1.16)

设随机变量(X, Y)的概率密度为

$$f(x, y) = \begin{cases} 15x^2y, & 0 < x < y < 1 \\ 0, & else \end{cases}$$

设$Z = XY$，求$E(Z)$。

例子 (4.1.16)

设随机变量(X, Y)的概率密度为

$$f(x, y) = \begin{cases} 15x^2y, & 0 < x < y < 1 \\ 0, & else \end{cases}$$

设$Z = XY$，求$E(Z)$。

例子 (4.1.17)

设随机变量X与Y相互独立，分别服从参数为4和2的指数分布，求$E(XY)$。

数学期望的性质

性质

(1) $E(kX) = kE(X)$

(2) $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

数学期望的性质

性质

(1) $E(kX) = kE(X)$

(2) $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

(3) 单调性：若$X \leq Y$，则$E(X) \leq E(Y)$

数学期望的性质

性质

(1) $E(kX) = kE(X)$

(2) $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

(3) 单调性：若$X \leq Y$，则$E(X) \leq E(Y)$

(4) 有界性：若$a \leq X \leq b$，则$a \leq E(X) \leq b$

性质

(5) 若$X_1,\ldots,X_n$相互独立，则$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n)$

性质

(5) 若$X_1, \ldots, X_n$相互独立，
则$E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2)\cdots E(X_n)$

(6) 收缩性：$|E(X)| \le E(|X|)$

性质

(5) 若$X_1, \ldots, X_n$相互独立，
则$E(X_1X_2\cdots X_n) = E(X_1)E(X_2)\cdots E(X_n)$

(6) 收缩性：$|E(X)| \le E(|X|)$

(7) 马尔科夫不等式：对任何$c > 0$，
有$P(|X| \ge c) \le \frac{E(|X|)}{c}$

性质

(5) 若$X_1,\ldots,X_n$相互独立，
则$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n)$

(6) 收缩性：$|E(X)|\le E(|X|)$

(7) 马尔科夫不等式：对任何$c>0$，
有$P(|X|\ge c)\le\frac{E(|X|)}{c}$

(8) 若$E(|X|)=0$，则$P(X=0)=1$

例子 (4.1.19)

设随机变量X服从二项分布$B(n, p)$，求$E(X)$。

例子 (4.1.19)

设随机变量X服从二项分布$B(n,p)$，求$E(X)$。

$$E(X)=np$$

例子 (4.1.20)

将n个球放入M个盒子中，设每个球落入各个盒子是等可能的，求有球的盒子数X的期望。

4.2 方差与标准差

方差的定义

定义 (4.2.1 Variance)

设X为随机变量，若$E((X-E(X))^2)$存在，则X的方差为

$$D(X) = E((X-E(X))^2).$$

同时，X的标准差（或均方差）为

$$\sigma_X = \sqrt{D(X)}.$$

方差的定义

定义 (4.2.1 Variance)

设X为随机变量，若$E((X-E(X))^2)$存在，则X的方差为

$$D(X)=E((X-E(X))^2).$$

同时，X的标准差（或均方差）为

$$\sigma_X=\sqrt{D(X)}.$$

- $D(X), \sigma_X$ 均度量了X与$E(X)$的偏离程度。实际中常采用σ_X，但$D(X)$数学性质较好。
- $D(X), \sigma_X$的值较小，表示随机变量的取值较集中；反之，表示随机变量的取值较分散。

方差的计算

- 离散型随机变量的方差

$$D(X) = \sum_{k\geq 1}(x_k - E(X))^2 p_k$$

- 连续型随机变量的方差

$$D(X) = \int_{-\infty}^{\infty}(x - E(X))^2 f(x)dx$$

方差的计算

- 离散型随机变量的方差

$$D(X)=\sum_{k\geq 1}(x_k-E(X))^2 p_k$$

- 连续型随机变量的方差

$$D(X)=\int_{-\infty}^{\infty}(x-E(X))^2 f(x)dx$$

- 常用公式

$$D(X)=E(X^2)-(E(X))^2$$

例子 (4.2.3)

设随机变量X的分布律为

X	0	1
P	$1-p$	p

求$D(X)$。

几种常见的随机变量的方差

例子 (4.2.3)

设随机变量X的分布律为

X	0	1
P	$1-p$	p

求$D(X)$。

$$D(X)=p(1-p)$$

例子 (4.2.4)

设随机变量X服从参数为λ的泊松分布，求$D(X)$。

例子 (4.2.4)

设随机变量X服从参数为λ的泊松分布，求$D(X)$。

$$D(X) = \lambda$$

例子 (4.2.5)

设随机变量X服从均匀分布$U[a, b]$，求$D(X)$。

例子 (4.2.4)

设随机变量X服从参数为λ的泊松分布，求$D(X)$。

$$D(X) = \lambda$$

例子 (4.2.5)

设随机变量X服从均匀分布$U[a, b]$，求$D(X)$。

$$D(X) = \frac{(b-a)^2}{12}$$

例子 (4.2.6)

设随机变量X服从参数λ的指数分布，求$D(X)$。

例子 (4.2.6)

设随机变量X服从参数λ的指数分布，求$D(X)$。

$$D(X) = \frac{1}{\lambda^2}$$

例子 (4.2.7)

设随机变量X服从正态分布$N(\mu, \sigma^2)$，求$D(X)$。

例子 (4.2.6)

设随机变量X服从参数λ的指数分布，求$D(X)$。

$$D(X) = \frac{1}{\lambda^2}$$

例子 (4.2.7)

设随机变量X服从正态分布$N(\mu, \sigma^2)$，求$D(X)$。

$$D(X) = \sigma^2$$

方差的性质

性质 (1)

若c为常数，则 $D(c) = 0$。

方差的性质

性质 (1)

若c为常数，则 $D(c) = 0$。

性质 (2)

$D(aX + c) = a^2 D(X)$

方差的性质

性质 (1)

若c为常数，则 $D(c) = 0$。

性质 (2)

$D(aX + c) = a^2 D(X)$

性质 (3)

$D(X) \leq E((X - C)^2)$

性质 (4)

若$X_1, \ldots, X_n$相互独立，
则$D(X_1 + \cdots + X_n) = D(X_1) + \cdots + D(X_n)$

性质 (4)

若$X_1, \ldots, X_n$相互独立，

则$D(X_1 + \cdots + X_n) = D(X_1) + \cdots + D(X_n)$

性质 (5)

若$D(X) = 0$，则$P(X = E(X)) = 1$

例子 (4.2.10)

设随机变量X服从二项分布$B(n, p)$，求$D(X)$。

例子 (4.2.10)

设随机变量X服从二项分布$B(n,p)$，求$D(X)$。

$$D(X) = np(1-p)$$

例子 (4.2.11)

设随机变量$X_1,\ldots,X_n$独立同分布，
且$E(X_k)=\mu, D(X_k)=\sigma^2, k=1,\ldots,n$。
令$\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$，求$E(\overline{X}), D(\overline{X})$。

切比雪夫不等式

考虑事件$|X - E(X)| \geq \epsilon$发生的概率。

定理 (4.2.2 Chebyshev inequality)

对任意随机变量X，若$D(X)$存在，则对任意$\epsilon > 0$，有

$$P(|X - E(X)| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}$$

或等价地，

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

切比雪夫不等式

考虑事件$|X - E(X)| \geq \epsilon$发生的概率。

定理 (4.2.2 Chebyshev inequality)

对任意随机变量X，若$D(X)$存在，则对任意$\epsilon > 0$，有

$$P(|X - E(X)| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}$$

或等价地，

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

- 可利用期望和方差对X的概率进行估计。

分位数

定义 (4.2.2)

设连续型随机变量X的分布函数为$F(X)$。称满足条件

$$F(x_\alpha) = P(X \le x_\alpha) = \alpha, \forall\, 0 < \alpha < 1$$

的数α为X的下α分位数（或下α分位点）。称满足条件

$$P(X \ge \mu_\alpha) = \alpha$$

的数μ_α为X的上α分位数（或上α分位点）。
当$\alpha = 0.5$，上*0.5*分位点与下*0.5*分位点相等，统称为中位点。

矩

定义 (4.2.3)

设X为随机变量，k为正整数。
若$E(X^k)$存在，则称$E(X^k)$为X的k阶原点矩；
若$E((X-E(X))^k)$存在，则称$E((X-E(X))^k)$为X的k阶中心矩。

矩

定义 (4.2.3)

设X为随机变量，k为正整数。
若$E(X^k)$存在，则称$E(X^k)$为X的k阶原点矩；
若$E((X-E(X))^k)$存在，则称$E((X-E(X))^k)$为X的k阶中心矩。

- 期望是一阶原点矩
- 方差是二阶中心矩
- 高阶矩存在，则低阶矩存在。

定义 (4.2.3)

设X为随机变量，k, ℓ为正整数。
若$E((X-E(X))^k(Y-E(Y))^\ell)$存在，则
称$E((X-E(X))^k(Y-E(Y))^\ell)$为$X$的$k+\ell$阶混合中心矩。

4.3 协方差与相关系数

协方差

定义 (4.3.1 Co-variance)

设(X, Y)为二维随机变量，若

$$E(|(X-E(X))(Y-E(Y))|) < \infty,$$

则称$E((X-E(X))(Y-E(Y)))$为X与Y的协方差，记

$$Cov(X, Y) = E((X-E(X))(Y-E(Y)))$$

协方差

定义 (4.3.1 Co-variance)

设(X, Y)为二维随机变量，若

$$E(|(X-E(X))(Y-E(Y))|) < \infty,$$

则称$E((X-E(X))(Y-E(Y)))$为X与Y的协方差，记

$$Cov(X, Y) = E((X-E(X))(Y-E(Y)))$$

协方差的计算式

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

协方差

定义 (4.3.1 Co-variance)

设(X, Y)为二维随机变量，若

$$E(|(X - E(X))(Y - E(Y))|) < \infty,$$

则称$E((X - E(X))(Y - E(Y)))$为X与Y的协方差，记

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

协方差的计算式

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

$$D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X, Y)$$

协方差的性质

性质

- X, Y 独立且协方差存在时，$Cov(X,Y) = 0$。
- $Cov(X,Y) = Cov(Y,X)$（对称性）

协方差的性质

性质

- X, Y 独立且协方差存在时，$Cov(X, Y) = 0$。
- $Cov(X, Y) = Cov(Y, X)$（对称性）
- $Cov(aX + b, cY + d) = acCov(X, Y)$

协方差的性质

性质

- X,Y 独立且协方差存在时，$Cov(X,Y)=0$。
- $Cov(X,Y)=Cov(Y,X)$（对称性）
- $Cov(aX+b,cY+d)=acCov(X,Y)$
- $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

协方差的性质

性质

- X,Y 独立且协方差存在时，$Cov(X,Y)=0$。
- $Cov(X,Y)=Cov(Y,X)$（对称性）
- $Cov(aX+b,cY+d)=acCov(X,Y)$
- $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

$$Cov(kX,kY)=k^2Cov(X,Y)$$

相关系数

标准化

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

相关系数

标准化

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

此时，$E(X^*) = E(Y^*) = 0$，$D(X^*) = D(Y^*) = 1$。

相关系数

标准化

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

此时，$E(X^*) = E(Y^*) = 0$，$D(X^*) = D(Y^*) = 1$。

$$\text{Cov}(X^*, Y^*) = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

定义 (4.3.2 correlation coefficient)

设(X, Y)为二维随机变量。若$D(X) > 0, D(Y) > 0$，则X与Y的相关系数记为

$$\rho = \rho_{XY} = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

因此，相关系数就是随机变量标准化后的协方差。

定义 (4.3.2 correlation coefficient)

设(X,Y)为二维随机变量。若$D(X)>0, D(Y)>0$，则X与Y的相关系数记为

$$\rho=\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

因此，相关系数就是随机变量标准化后的协方差。

- 相关系数在线性变换下保持不变

$$\rho_{aX+b,cY+d}=\pm\rho_{XY}$$