第五章　随机变量的数字特征与极限定理总结与例题

一、重点与难点

二、主要内容

三、典型例题

一、重点与难点

1.重点

数学期望的性质和计算

方差的性质和计算

相关系数的性质和计算

2.难点

数字特征的计算

二、主要内容

矩、协方差矩阵

协方差
相关系数

数字特征的引入

几种重要分布的期望和方差

数学期望
$E(X)$

方差
$D(X)=E(X-E(X))^2$

离散型
连续型

随机变量函数的数学期望

离散型
连续型

期望性质

多维随机变量函数的数学期望

方差性质

性质的应用

二、主要内容

极限定理

大数定律

客观背景

中心极限定理

切比雪夫大数定律

独立同分布下的中心极限定理

独立同分布下大数定律

辛钦大数定律

中心极限定理的应用

贝努里大数定律

大数定律的应用

德莫佛-拉普拉斯定理(二项分布的正态近似)

离散型随机变量的数学期望

定义1 设X是离散型随机变量，它的概率分布列是: $P(X=X_k)=p_k$，$k=1,2,\ldots$

如果$\sum_{k=1}^{\infty}|x_k|p_k$有限,定义$X$的数学期望

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

也就是说,离散型随机变量的数学期望是一个绝对收敛的级数的和.

如果$\sum_{i=1}^{\infty}|x_i|p_i$发散，则X的数学期望不存在。

连续型随机变量的数学期望

定义2 设X是连续型随机变量，其密度函数为$f(x)$,如果

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

有限,定义X的数学期望为

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

也就是说,连续型随机变量的数学期望是一个绝对收敛的积分.

随机变量函数的数学期望

设X是一个随机变量，$Y=g(X)$，则

$$E(Y)=E[g(X)]=\begin{cases}\sum\limits_{k=1}^{\infty} g(x_k)p_k, & X\text{离散型}\\ \int_{-\infty}^{\infty} g(x)f(x)dx, & X\text{连续型}\end{cases}$$

当X为离散型时,$P(X=x_k)=p_k$；

当X为连续型时,X的密度函数为$f(x)$.

随机变量函数的数学期望

设(X, Y)是二维随机变量, $Z=g(X, Y)$，则

$$EZ = E[g(X,Y)]$$

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

当(X, Y)是离散型时：分布列为

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

随机变量函数的数学期望

当(X, Y)是连续型时：联合概率密度为$f(x, y)$

由此可知：已知(X, Y)的联合概率密度$f(x, y)$，

可以求EX，EY

即

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

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

数学期望的性质

1. 设C是常数，则$E(C)=C$;

2. 若k是常数，则$E(kX)=kE(X)$;

3. $E(X_1+X_2) = E(X_1)+E(X_2)$;

注意:由$E(XY)=E(X)E(Y)$不一定能推出X,Y独立

推广：$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E(X_i)$

4. 设X、Y独立，则 $E(XY)=E(X)E(Y)$;

推广：$E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E(X_i)$（诸$X_i$独立时）

方差的定义

设X是一个随机变量，若$E[X-E(X)]^2<\infty$，则称

$$D(X)=E[X-E(X)]^2 \qquad (1)$$

为X的方差.

方差的算术平方根$\sqrt{D(X)}$称为标准差

由于它与X具有相同的度量单位，在实际问题中经常使用.

方差的计算

由定义知，方差是随机变量X的函数 $g(X)=[X-E(X)]^2$的数学期望．

$$D(X)=\begin{cases}\sum_{k=1}^{\infty}[x_k-E(X)]^2p_k, \\ \int_{-\infty}^{\infty}[x-E(X)]^2f(x)dx,\end{cases}$$

X为离散型，

$P(X=x_k)=p_k$

X为连续型，

$X\sim f(x)$

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

$E(X^2)\geq[E(X)]^2$

方差的性质

1. 设C是常数,则$D(C)=0$;

2. 若C是常数,则$D(CX)=C^2 D(X)$;

3. 若X_1与X_2 独立，则

$$D(X_1+X_2)= D(X_1)+D(X_2);$$

可推广为：若$X_1,X_2,\ldots,X_n$相互独立,则

$$D[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} D(X_i)$$

$$D[\sum_{i=1}^{n} C_i X_i] = \sum_{i=1}^{n} C_i^2 D(X_i)$$

方差的性质

4. $D(X)=0 \Leftrightarrow P(X=C)=1$， 这里$C=E(X)$

5. 若X与Y独立，则

$$D(XY)=D(X)D(Y)+D(X)[E(Y)]^2+D(Y)[E(X)]^2;$$

常用分布的期望和方差

若$X \sim B(1,p)$,则$E(X)=p, D(X)=p(1-p)$,

若$X \sim B(n,p)$,则$E(X)=np, D(X)=np(1-p)$,

若$X \sim P(\lambda)$,则$E(X)=D(X)=\lambda, \quad (\lambda>0)$,

若$X \sim G(p)$,则$E(X)=1/p, D(X)=(1-p)/p^2$,

若$X \sim U(a,b)$,则$E(X)=(a+b)/2$,

$$D(X)=(b-a)^2/12,$$

若$X \sim E(\lambda)$,则$E(X)=1/\lambda, D(X)=1/\lambda^2, (\lambda>0)$,

若$X \sim N(\mu,\sigma^2)$,则$E(X)=\mu, D(X)=\sigma^2, (\sigma>0)$.

协方差的定义与性质

1.定义　任意两个随机变量X和Y的协方差，记为$Cov(X,Y)$，定义为

$$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$

2.简单性质

（1）　$Cov(X,Y)=Cov(Y,X)$　$Cov(X,a)=0$

（2）　$Cov(aX,bY)=ab\,Cov(X,Y)$　a,b是常数

（3）　$Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

（4）若X与Y独立，　$Cov(X,Y)=0$.

协方差的定义与性质

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

随机变量和的方差与协方差的关系

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

$$D\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} D(X_i)+2\sum_{i<j}\sum Cov(X_i,X_j)$$

若$X_1,X_2,\ldots,X_n$两两独立,，上式化为

$$D\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} D(X_i)$$

相关系数

定义：设$D(X)>0, D(Y)>0$，称

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

在不致引起混淆时，记ρ_{XY}为 ρ .

若 $\rho=0$ 则称X与Y不相关

- 相关系数是表示两个随机变量之间线性相关程度的一个数字特征.（无量纲）

相关系数的性质

1. $|\rho| \le 1$

2. $|\rho| = 1$ $\Longleftrightarrow$ 存在常数$a,b(b\neq 0)$，

使$P\{Y=a+bX\}=1$,

即X和Y以概率1线性相关.

3. X和Y独立时，$\rho=0$，但其逆不真.

$|\rho|$的值越接近于1, Y与X的线性相关程度越高;

$|\rho|$的值越接近于0, Y与X的线性相关程度越弱.

相关系数的性质

若X与Y独立，则X与Y不相关，

但由X与Y不相关，不一定能推出X与Y独立.

下面四个是等价的：

$$\rho = 0 \Leftrightarrow Cov(X,Y) = 0 \Leftrightarrow E(XY) = EXEX$$

$$\Leftrightarrow D(X \pm Y) = DX + DY.$$

二维正态

设$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$,

二维正态分布的两个边缘密度仍是正态分布 .

$X\sim N(\mu_1,\sigma_1{}^2), Y\sim N(\mu_2,\sigma_2{}^2), \rho=\rho_{XY}$

若(X,Y)服从二维正态分布，则

X与Y独立 $\Leftrightarrow$ X与Y不相关

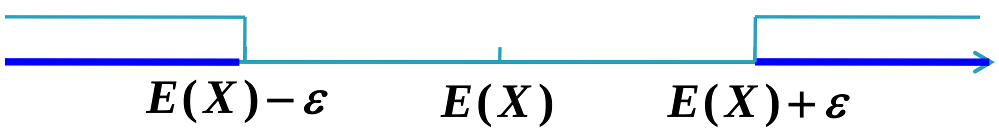
切比雪夫不等式

定理1 对任意随机变量X, 若$D(X)$存在，则对任意$\varepsilon > 0$有

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

或

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



三、典型例题

例1 设*r.v X*服从几何分布，概率分布列为

$$P(X=k)=p(1-p)^{k-1},\ k=1,2,\ldots,\ n$$

其中$0<p<1$,求$D(X)$

无穷递缩等比级数求和公式

解： 记$q=1-p$

$$E(X)=\sum_{k=1}^{\infty}kpq^{k-1}=p\sum_{k=1}^{\infty}(q^{k})'$$

求和与求导交换次序

$$=p\Big(\sum_{k=1}^{\infty}q^{k}\Big)'=p\Big(\frac{q}{1-q}\Big)'=\frac{1}{p}$$

$$= p[\sum_{k=1}^{\infty} k(k-1)q^{k-1} + \sum_{k=1}^{\infty} kq^{k-1}]$$

$$= qp(\sum_{k=1}^{\infty} q^k)'' + E(X) = qp(\frac{q}{1-q})'' + \frac{1}{p}$$

$$= qp\frac{2}{(1-q)^3} + \frac{1}{p} = \frac{2q}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$$

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

$$= \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

例2 设随机变量X和Y相互独立且$X\sim N(1,2)$, $Y\sim N(0,1)$. 试求$Z=2X-Y+3$的概率密度.

解: $X\sim N(1,2), Y\sim N(0,1)$，且$X$与$Y$独立,

故X和Y的联合分布为正态分布，X和Y的任意线性组合是正态分布.

即 $Z\sim N(E(Z), D(Z))$

$$E(Z)=2E(X)-E(Y)+3=2+3=5$$

$$D(Z)=4D(X)+D(Y)=8+1=9 \qquad Z\sim N(5, 3^2)$$

$$f_Z(z)=\frac{1}{3\sqrt{2\pi}}e^{-\frac{(z-5)^2}{18}}, \qquad -\infty<z<\infty$$

例3 设随机变量 (X,Y) 服从 $D=\{(x,y): y\ge 0, x^2+y^2\le 1\}$ 上的均匀分布, 定义随机变量 U,V 如下

$$U=\begin{cases}0, & X<0,\\ 1, & 0\le X<Y,\\ 2, & X\ge Y.\end{cases}\quad V=\begin{cases}0, & X\ge\sqrt{3}Y,\\ 1, & X<\sqrt{3}Y.\end{cases}$$

求 (U,V) 的联合概率分布,并计算 $P\{UV\ne 0\}$. ρ_{UV}

[思路] 写出(U,V)的所有可能取值,并利用均匀分布的特征计算其取值的概率.

解　由题设知 (X,Y) 的联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{2}{\pi}, & (x,y)\in D,\\ 0, & (x,y)\notin D.\end{cases}$$

(U,V) 有6个可能取值：

$(0,0)\quad(0,1)\quad(1,0)\quad(1,1)\quad(2,0)\quad(2,1)$

$$P=\{U=0,V=0\}=P(\varnothing)=0,$$

$$P=\{U=1,V=0\}=P(\varnothing)=0,$$

$$P=\{U=1,V=1\}=P\{0\le X<Y,X<\sqrt{3}Y\}$$

$$= P\{0 \le X < Y\} = \iint\limits_{0 \le x < y} f(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

$$= \iint\limits_{0 \le x < y} \frac{2}{\pi}\,\mathrm{d}x\,\mathrm{d}y = \frac{S_{扇AOC}}{S_{BCE}} = \frac{1}{4}.$$

$$P = \{U = 0, V = 1\} = P\{X < 0, X < \sqrt{3}Y\}$$

$$= P\{X < 0\} = \frac{S_{扇COE}}{S_{BCE}} = \frac{1}{2},$$

$$P = \{U = 2, V = 0\} = P\{Y \le X, X \ge \sqrt{3}Y\}$$

$$= P\{X \ge \sqrt{3}Y\} = \frac{S_{扇BOF}}{S_{BCE}} = \frac{1}{6},$$

$$P=\{U=2,V=1\}=P\{Y\le X,X<\sqrt{3}Y\}$$

$$=P\{Y\le X<\sqrt{3}Y\}=\frac{S_{扇AOF}}{S_{BCE}}=\frac{1}{12}.$$

所以(U,V)的联合概率分布为

$V \backslash U$	0	1	2
0	0	0	$\frac{1}{6}$
1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{12}$

$E(U)=3/4, D(U)=11/16$

$E(V)=5/6, D(V)=5/36$

$E(UV)=5/12,$

$$\rho_{UV}=\frac{Cov(U,V)}{\sqrt{D(U)D(V)}}$$

$$=-\sqrt{\frac{5}{11}}$$

从而

$$P\{UV \neq 0\}$$

$$= P\{U = 1, V = 1\} + P\{U = 2, V = 1\}$$

$$= \frac{1}{4} + \frac{1}{12} = \frac{1}{3}.$$

例4. 设r,v X，Y相的相关系数为 $1/\sqrt{3}$ ，X在[0，1]上服从均匀分布，Y 服从参数λ=1的指数分布，Z=2X+Y ，求$E(Z)$,$D(Z)$, ρ_{YZ}。

解

$$X \sim U[0,1], E(X) = 1/2, D(X) = 1/12$$

$$Y \sim E(1), E(X) = D(X) = 1$$

$$E(Z) = 2E(X) + E(Y) = 2,$$

$$D(Z) = 4D(X) + D(Y) + 4\rho\sqrt{D(X)D(Y)} = 2,$$

$$\rho_{YZ} = \frac{Cov(Y,Z)}{\sqrt{D(Y)D(Z)}} = \frac{Cov(Y,2X+Y)}{\sqrt{D(Y)D(Z)}}$$

$$= \frac{2Cov(Y,X) + D(Y)}{\sqrt{D(Y)D(Z)}} = \frac{2\sqrt{2}}{3}$$

例5 已知正常男性成人血液中，每一毫升白细胞数平均是7300，均方差是700 . 利用切比雪夫不等式估计每毫升白细胞数在5200~9400之间的概率 .

解： 设每毫升白细胞数为X

依题意，$E(X)=7300, D(X)=700^2$

所求为 $P(5200 \le X \le 9400)$

$$P(5200 \le X \le 9400)$$

$$=P(5200-7300 \le X-7300 \le 9400-7300)$$

$$=P(-2100 \le X-E(X) \le 2100)$$

$$=P\{|X-E(X)| \le 2100\}$$

由切比雪夫不等式

$$P\{|X-E(X)| \le 2100\} \ge 1-\frac{D(X)}{(2100)^2}$$

$$=1-\left(\frac{700}{2100}\right)^2=1-\frac{1}{9}=\frac{8}{9}$$

即估计每毫升白细胞数在5200~9400之间的概率不小于8/9 .