第四章　随机变量的数字特征

- 数 学 期 望
- 方 差
- 协 方 差 及 相 关 系 数
- 矩、协 方 差 矩 阵

随机变量及其分布能够完整地描述随机变量的统计规律。但在一些实际问题中，这样的全面描述有时并不方便。

例如

- 要比较两个品种的母鸡的年产蛋量，通常只需比较它们的年产蛋量的平均值就可以。这时若不比较平均值，而只看它们的分布律，虽然全面却使人难以掌握又不能迅速地作出判断。
- 评价棉花的质量时，即需要注意纤维的平均长度，又需要注意纤维长度与平均长度的偏离程度，平均长度越大，偏离程度越小，质量就越好。

这种由随机变量的分布所确定的，能刻画随机变量某一方面的特征的常数统称为**数字特征**。

Story

§1 数学期望

一、数学期望的概念

引例——分赌本问题

A, B 两人赌技相同, 各出赌金100元,并约定先胜三局者为胜, 取得全部 200 元.由于出现意外情况 ,在 A 胜 2 局 B 胜1 局时 ,不得不终止赌博, 如果要分赌金,该如何分配才算公平?

分析 假设继续赌两局,则结果有以下四种情况:

AA	AB	BA	BB
A胜B负	A胜B负	B胜A负	B胜A负
A胜B负	B胜A负	A胜B负	B胜A负

把已赌过的三局(A 胜2局B 胜1局)与上述结果相结合, 即 A、B 赌完五局

前三局	A 胜2局B 胜1局			
后两局	AA	AB	BA	BB
		A胜		B胜

一、数学期望的概念

引例——分赌本问题

故有，在赌技相同的情况下，A, B最终获胜的可能性大小之比为**3:1**。即A应获得赌金的3/4，而B只能获得赌金的1/4。

因此, A 能“**期望**”得到的数目应为

$$200\times\frac{3}{4}+0\times\frac{1}{4}=150(\text{元})$$

而B 能“**期望**”得到的数目应为

$$200\times\frac{1}{4}+0\times\frac{3}{4}=50(\text{元})$$

若设随机变量 X 为:在 A 胜2局B 胜1局的前提下, 继续赌下去 A 最终所得的赌金。

则X 所取可能值为: $200 \quad 50$

其概率分别为: $\frac{3}{4} \quad \frac{1}{4}$

一、数学期望的概念

引例——分赌本问题

因而A期望所得的赌金即为X的“**期望**”值，

$$200\times\frac{3}{4}+0\times\frac{1}{4}=150(\text{元})$$

即为X 的可能值与概率之积的累加

离散型r.v.的数学期望

设离散型r.v. X的分布律为

$$P\{X=x_k\}=p_k,\quad(k=1,2,\dots)$$

若级数$\sum_{k=1}^{\infty}x_k p_k$绝对收敛，则称级数$\sum_{k=1}^{\infty}x_k p_k$为随机变量$X$的**数学期望**，记为$E(X)$。即

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

§1 数学期望

一、数学期望的概念

关于定义的几点说明

（1）$E(X)$是一个实数,而非变量,它是一种加权平均,与一般的平均值不同，它从**本质上体现了随机变量 X 取可能值的真正的平均值,也称均值.**

（2）随机变量的数学期望与一般变量的算术平均值不同.

（3）**级数的绝对收敛性保证了级数的和不随级数各项次序的改变而改变**,因为数学期望是反映随机变量X取可能值的平均值,它不应随可能值的排列次序而改变.

假设

X	1	2
P	0.02	0.98

随机变量 X 的算术平均值为

$$\frac{1+2}{2}=1.5,$$

$$E(X)=1\times 0.02+2\times 0.98=1.98.$$

它从本质上体现了随机变量X取可能值的平均值.

当随机变量 X 取各个可能值是等概率分布时，X 的期望值与算术平均值相等.

§1　数学期望

一、数学期望的概念

几个常用离散型r.v.的数学期望

1. 两点(0-1)分布

r.v. X服从两点分布，其概率分布律为：$P\{X=1\}=p,\ P\{X=0\}=1-p$

则 $$E(X)=1\times p+0\times(1-p)=p$$

2. 二项分布

r.v. $X\backsim b(n,p)$，其分布律为：

$$P\{X=k\}=C_n^k p^k(1-p)^{n\ k},(k=0,1,\dots,n)$$

则 $$E(X)=\sum_{k=0}^{n}k\cdot C_n^k p^k(1-p)^{n-k}=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^k(1-p)^{n-k}$$

$$\overset{令l=k-1}{=}np\sum_{k=1}^{n}C_{n-1}^k p^l(1-p)^{(n-1)-l}=np(p+1-p)^{n-1}=np$$

3. 泊松分布

r.v. $X\backsim\pi(\lambda)$，其概率分布律为：

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{\ \lambda},(k=0,1,2,\dots)$$

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

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

一、数学期望的概念

典型例题

例 每张福利彩票售价5元，各有一个对奖号，每售出100万张设一个开奖组，用摇奖器当众摇出一个6位数的中奖号码(可以认为从000 000到999 999的每个数都有百万分之一的概率被摇到)，兑奖规则如下：

(1) 兑奖号与中奖号码的后两位相同者中四等奖，奖金20元；

(2) 兑奖号与中奖号码的后3位相同者中三等奖，奖金300元；

(3) 兑奖号与中奖号码的后4位相同者中二等奖，奖金5000元；

(4) 兑奖号与中奖号码后五位全同，而第一位有相同的奇偶性者中一等奖，奖金额为31.5万元；

(5) 由于中了高奖金额者必中低奖金额的奖，特规定，只领取所中最高额的奖金，试求每张彩票的期望所得

一、数学期望的概念

典型例题

例 在一个人数很多的团体中普查某种疾病，为此要抽检N个人的血，可以用两种方法进行。(*i*)将每个人的血分别去验，这就需验N次。(*ii*)按k个人一组进行分组，把从k个人抽来的血混合在一起进行检验，如果这混合血液呈阴性反应，就说明k个人的血都呈阴性反应，这样，这k个人的血就只需验一次。若呈阳性，则再对这k个人的血液分别进行化验。这样，k个人的血总共要化验k+1次。假设每个人化验呈阳性的概率为p，且这些人的试验反应试相互独立的。试说明当p较小时，选取适当的k，按第二种方法可以减少化验的次数。并说明k取什么值时最适宜。

k	2	3	4	5	8	10	30	33	34
E(X)	0.690	0.604	0.594	0.610	0.695	0.751	0.991	0.994	1.0016

p	0.1	0.08	0.04	0.02	0.01
k	4	4	6	8	11
E(X)	0.594	0.534	0.384	0.274	0.205

一、数学期望的概念

连续型r.v.的数学期望

设连续型r.v. X的概率密度为$f(x)$，若积分

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

绝对收敛，则称积分$\int_{-\infty}^{\infty} xf(x)dx$ 的值为随机变量X的**数学期望**，记为$E(X)$。即

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

数学期望$E(X)$完全由随机变量X 的概率分布所确定。若X服从某一分布，也称$E(X)$是这一分布的数学期望。

§1 数学期望

一、数学期望的概念

几个常用连续型r.v.的数学期望

1. 均匀分布

r.v. $X \backsim b(a, b)$，其概率密度为：$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b \\ 0, & else\end{cases}$

其中a, b为常数，且$a<b$

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_{a}^{b} x\frac{1}{b-a}dx=\frac{a+b}{2}$$

2. 指数分布

r.v. X服从参数为$\theta>0$的指数分布，其概率密度为：$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{1}{\theta}x} & x>0 \\ 0 & x\le 0\end{cases}$

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_{0}^{\infty} x\frac{1}{\theta}e^{-\frac{1}{\theta}x}dx=\theta$$

也可以记 $f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0 \\ 0 & x\le 0\end{cases}$ 则有 $E(X)=\dfrac{1}{\lambda}$

一、数学期望的概念

几个常用连续型r.v.的数学期望

3. 正态分布

r.v. $X \sim N(\mu, \sigma^2)$，其概率密度为：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

其中μ和σ都是常数, μ任意, $\sigma>0$

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{-\infty}^{\infty}x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \overset{令t=\frac{x-\mu}{\sigma}}{=} \int_{-\infty}^{+\infty}\frac{\sigma t+\mu}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

$$=\int_{-\infty}^{+\infty}\frac{\sigma t}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt+\int_{-\infty}^{+\infty}\frac{\mu}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\mu$$

一、数学期望的概念

典型例题

例 随机变量X的分布函数为$F(x)=\begin{cases}\frac{1}{2}e^{x} & x\le 0\\ \frac{1}{2} & 0<x\le 1\\ 1-\frac{1}{2}e^{-(x-1)} & x>1\end{cases}$，求$E(X)$

例 随机变量X的概率密度为 $f(x)=\begin{cases}\frac{1}{2}\cos\frac{x}{2} & 0\le x\le \pi\\ 0 & else\end{cases}$

对X独立地重复观察4次，用Y表示观察值大于$\pi/3$的次数，求Y^2的数学期望

二、随机变量函数的数学期望

定 理

设Y是r.v. X的函数： $Y=g(X)$ (g是连续函数).

- 如果X是离散型r.v.，它的分布律为$P\{X=x_k\}=p_k$, (k=1, 2, …)，若$\sum_{k=1}^{\infty} g(x_k)p_k$绝对收敛，则有

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

- 如果X是连续型r.v.，它的概率密度为$f(x)$，若$\int_{-\infty}^{\infty} xf(x)dx$绝对收敛，则有

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

二、随机变量函数的数学期望

定理的推广

设Z是r.v. X，Y的函数$Z=g(X, Y)$(g是连续函数)，那么，是Z一个一维r.v.。若二维随机变量(X, Y)的概率密度为$f(x, y)$，则有

$$E(Z)=E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

设上式右边的积分绝对收敛。若(X, Y)是离散r.v.，其分布律为$P\{X=x_i, Y=y_j\}=p_{ij}$，$(i,j=1, 2, \ldots)$，则有

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

二、随机变量函数的数学期望

典型例题

例 设X服从$(0,2\pi)$上的均匀分布，求E(sin$X$)。

解 这里g(X)=sin$X$，X的概率密度为：$f(x)=\begin{cases}\dfrac{1}{2\pi} & 0\le x\le 2\pi\\ 0 & else\end{cases}$

$$E(\sin X)=\int_{-\infty}^{+\infty}\sin x f(x)dx=\int_0^{2\pi}\sin x\times\frac{1}{2\pi}dx=0$$

例 游客乘电梯从底层到电视塔顶层观光，电梯于每个整点的第5分钟，25分钟和55分钟从底层起行。假设一游客在早上8点的第X分钟到达底层的侯梯处，且X在[0,60]上均匀分布，求该游客等候时间的数学期望。

二、随机变量函数的数学期望

典型例题

例 设某种产品的市场需求量是随机变量X（单位：t），它服从[2000,4000]上的均匀分布。设每销售这种产品一吨，可得到利润3万元，但是若销售不出囤积仓库，每吨需保管费用1万元，应生产多少吨产品，使获利最大？

解 设产品生产量为a（单位：t），$a\in[2000,4000]$，获利数Y(单位：万元)是随机变量，Y是市场需求量X的函数

$$Y=g(X)=\begin{cases}3a & X\ge a\\ 3X-(a-X) & X<a\end{cases}$$

$$E(Y)=\int_{-\infty}^{+\infty}g(x)f(x)dx=\frac{1}{2000}\int_{2000}^{4000}g(x)dx$$

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

$$=\frac{1}{1000}\left(-a^2+7000a-4000000\right)$$

由 $\dfrac{dE(Y)}{da}=0$ 得到：a=3500，此时$E(Y)$最大

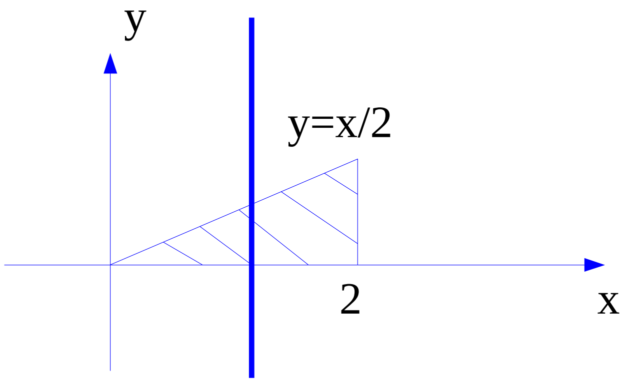
即得当生产3500吨产品时，平均获利最大

二、随机变量函数的数学期望

典型例题

例 设r.v. X与Y的联合概率密度为：

$$f(x,y)=\begin{cases}2xy & (x,y)\in G\\ 0 & else\end{cases}$$



解 $E(2XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}2xyf(x,y)dxdy$

$$=\int_0^2\left[\int_0^{\frac{x}{2}}2xy\cdot 2xydy\right]dx=4\int_0^2 x^2dx\int_0^{\frac{x}{2}}y^2dy=\frac{1}{6}\int_0^2 x^5dx=\frac{16}{9}$$

三、数学期望的性质

1. 设C是常数，则有$E(C)=C$

证明： $E(X)=E(C)=1\times C=C$

2. 设X是一个r.v.，C是常数，则有$E(CX)=CE(X)$

证明： $E(CX)=\sum_k Cx_k p_k=C\sum_k x_k p_k=CE(X)$

$$E(CX)=\int_{-\infty}^{\infty} Cxf(x)dx=C\int_{-\infty}^{\infty} xf(x)dx=CE(X)$$

3. 设X，Y是两个r.v.，则有$E(X+Y)=E(X)+E(Y)$

证明： $E(X+Y)=\sum_k (x_k+y_k)p_k$

$$=\sum_k x_k p_k+\sum_k y_k p_k=E(X)+E(Y)$$

三、数学期望的性质

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

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

$$=E(X)+E(Y)$$

4. 设X，Y是相互独立的随机变量，则有$E(XY)=E(X)E(Y)$

证明：

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

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_X(x)f_Y(y)dxdy=\int_{-\infty}^{\infty}xf_X(x)dx\int_{-\infty}^{\infty}yf_Y(y)dy$$

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

白猫黑猫问题

三、数学期望的性质

典型例题

例 设r.v. X与Y的概率密度分别为：

$$f_X(x)=\begin{cases}2x & 0\le x\le 1\\ 0 & else\end{cases}\qquad f_Y(y)=\begin{cases}e^{-(y-5)} & y>5\\ 0 & else\end{cases}$$

X和Y相互独立，求$E(XY)$

解 由于X和Y相互独立，故有：

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

$$=\int_0^1 2xdx\cdot\int_5^{+\infty} ye^{-(y-5)}dy=\frac{2}{3}e^5\int_5^{+\infty} ye^{-y}dy=4$$

§2 方差

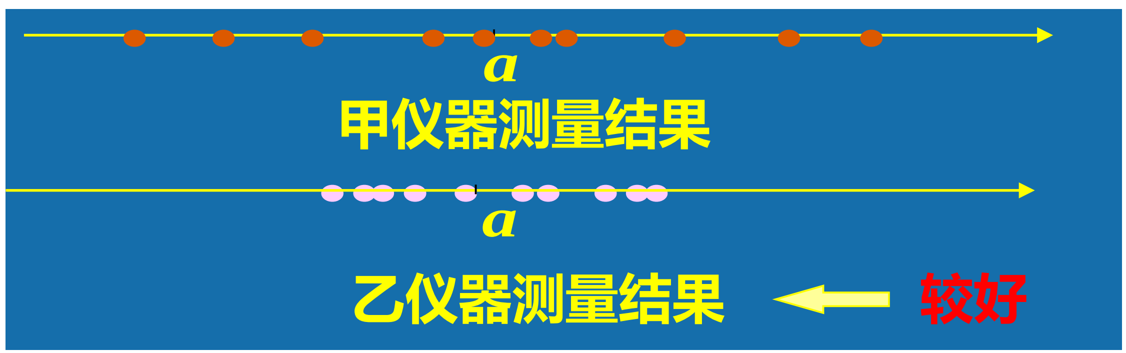
一、方差的概念

引例

随机变量的数学期望，它体现了随机变量取值的平均水平，是随机变量的一个重要的数字特征。但是在一些场合，仅仅知道平均值是不够的。

例 某零件的真实长度为a，现用甲、乙两台仪器各测量10次，将测量结果X用坐标上的点表示如图：

测量结果的均值都是 a



若要评价一下两台仪器的优劣，哪台仪器好一些呢?

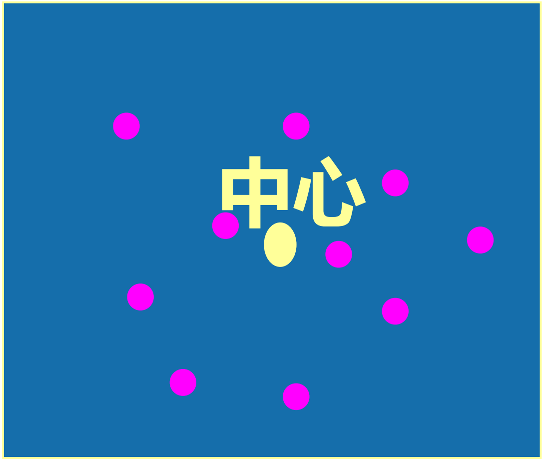
因为乙仪器的测量结果集中在均值附近

§2 方差

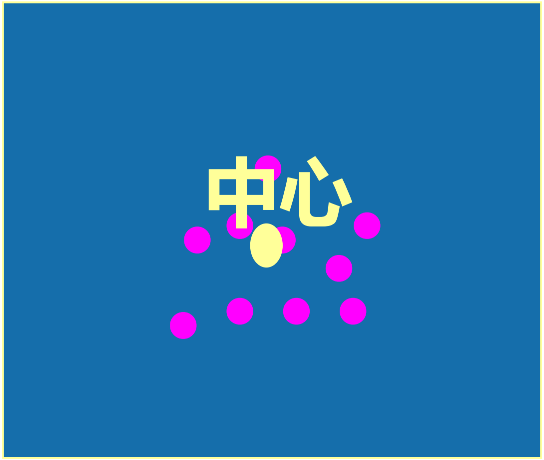
一、方差的概念

引例

又如 甲、乙两门炮同时向一目标射击10发炮弹, 其落点距目标的位置如图:



甲炮射击结果



乙炮射击结果

哪门炮射击效果好一些呢? **乙较好**

因为乙炮的弹着点较集中在中心附近。

§2 方差

一、方差的概念

在实际问题中常关心随机变量与均值的偏离程度，为此需要引进另一个数字特征，这个数字特征就是我们这一讲要介绍的**方差**(Variance)

可用E{|X-E(X)|}，但不方便；所以通常用$E\{[X-E(X)]^2\}$来**度量随机变量X与其均值E(X)的偏离程度。**

方差的定义

设X是一个r.v.，若概率密度为$f(x)$，若$E\{[X-E(X)]^2\}$存在，则称$E\{[X-E(X)]^2\}$为X的方差，即为$D(X)$或$Var(X)$，即

$$D(X)=Var(X)=E\{[X-E(X)]^2\}$$

在应用上还引入量$\sqrt{D(X)}$，记为$\sigma(X)$，称为标准差或均方差

若X的取值比较集中,则D(X)较小；若X的取值比较分散,则$D(X)$较大。**方差是刻画随机变量取值离散程度的一个量。**

二、方差的计算

由定义知，方差实际是r.v.X的函数$g(X)=(X-E(X))^2$的数学期望。

于是，对于离散型r.v.，有

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

其中$P\{X=x_k\}=p_k$, $(k=1, 2, \ldots)$是X的分布律

对于连续型r.v.，有 $D(X)=\int_{-\infty}^{\infty}[x-E(X)]^2 f(x)dx$

其中$f(x)$是X的概率密度函数

r.v.X的方差可按下式公式计算$D(X)=E(X^2)-[E(X)]^2$

证明： $$D(X)=E\{[X-E(X)]^2\}=E\{X^2-2XE(X)+[E(X)]^2\}$$

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

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

二、方差的计算

常见分布的方差

1. 两点(0-1)分布 (或$X\backsim b(1,p)$)

由$E(X)=p$，$E(X^2)=0^2\cdot(1-p)+1^2\cdot p=p$

得到 $D(X)=E(X^2)-\{E(X)\}^2=p(1-p)$

2. 二项分布

r.v. $X\backsim b(n,p)$，其分布律为：$P\{X=k\}=C_n^k p^k(1-p)^{n\ k},(k=0,1,\ldots,n)$

则 $E(X)=np$，$E(X^2)=E(X+X(X-1))=E(X)+E(X(X-1))$

$$=np+\sum_{k=0}^{n}k(k-1)\cdot C_n^k p^k(1-p)^{n-k}=np+\sum_{k=2}^{n}k(k-1)\cdot C_n^k p^k(1-p)^{n-k}$$

$$\overset{令l=k-2}{=}np+n(n-1)p^2\sum_{k=0}^{n-2}\frac{(n-2)!}{l!(n-2-l)!}p^k(1-p)^{n-2-k}=np+n(n-1)p^2$$

$$D(X)=E(X^2)-\{E(X)\}^2=np-n(n-1)p^2-(np)^2=np(1-p)$$

二、方差的计算

常见分布的方差

3. 泊松分布

r.v. $X \backsim \pi(\lambda)$，其概率分布律为：$P\{X=k\}=\dfrac{\lambda^k}{k!}e^{\lambda},(k=0,1,2,\dots)$

已知 $E(X)=\lambda$，$E(X^2)=\sum\limits_{k=0}^{\infty}k^2\cdot e^{-\lambda}\dfrac{\lambda^k}{k!}=\sum\limits_{k=0}^{\infty}(k-1+1)k\cdot e^{-\lambda}\dfrac{\lambda^k}{k!}$

$$=\sum_{k=0}^{\infty}(k-1)k\cdot e^{-\lambda}\frac{\lambda^k}{k!}+\sum_{k=0}^{\infty}k\cdot e^{-\lambda}\frac{\lambda^k}{k!}=\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}e^{-\lambda}}{(k-2)!}+E(X)=\lambda^2+\lambda$$

$$D(X)=E(X^2)-\{E(X)\}^2=\lambda^2+\lambda-\lambda^2=\lambda$$

4. 指数分布

r.v. X服从参数为$\lambda>0$的指数分布，其概率密度为：$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$

$E(X)=\dfrac{1}{\lambda}$，$E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)dx=\int_{-\infty}^{+\infty}x^2\lambda e^{-\lambda x}dx=\dfrac{2}{\lambda^2}$

$$D(X)=E(X^2)-\{E(X)\}^2=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}$$

二、方差的计算

常见分布的方差

5. 正态分布

r.v. $X \sim N(\mu, \sigma^2)$ ，$E(X)=\mu$

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

$$\overset{令 t=\frac{x-\mu}{\sigma}}{=} \sigma^2 \int_{-\infty}^{+\infty} \frac{t^2}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}dt=\sigma^2$$

6. Γ分布

$$f(x)=\begin{cases} \dfrac{\beta}{\Gamma(\alpha)}(\beta x)^{\alpha-1}e^{-\beta x} & x>0 \\ 0 & x\le 0 \end{cases}$$

$$E(X)=\alpha/\beta \qquad D(X)=\alpha/\beta^2$$

$$E(X^k)=\frac{\alpha(\alpha+1)(\alpha+1)\cdots(\alpha+k-1)}{\beta^k}$$

二、方差的计算

常见随机分布的期望和方差

分 布	参 数	数学期望	方 差
两点分布	$0<p<1$	p	$p(1-p)$
二项分布	$n\geqslant 1$, $0<p<1$	np	$np(1-p)$
泊松分布	$\lambda>0$	λ	λ
均匀分布	$a<b$	$(a+b)/2$	$(a-b)^2/12$
指数分布	$\theta>0$	θ	θ^2
正态分布	$\mu,\sigma>0$	μ	σ^2

二、方差的计算

典型例题

例 袋中有n张卡片，号码分别为1,2,...,n,从中有放回地取k张卡片，令X表示所抽取的k张卡片的号码之和，试求$E(X)$及$D(X)$。

解 令X_i表示第i张卡片的号码(i=1,2,...,n)，则$X=X_1+X_2+\dots+X_k$ 因为是有放回抽取，所以X_i之间相互独立。且$P\{X_i=j\}=\frac{1}{n}(j=1,2,\dots,n)$，$i=1,2,\dots,k$

$$E(X_i)=\sum_{j=1}^{n} j\cdot\frac{1}{n}=\frac{1}{n}\frac{n(n+1)}{2}=\frac{(n+1)}{2}$$

$$E(X_i^2)=\sum_{j=1}^{n} j^2\cdot\frac{1}{n}=\frac{1}{n}\frac{n(n+1)(2n+1)}{6}=\frac{(n+1)(2n+1)}{6}$$

$$D(X_i)=E(X_i^2)-\{E(X_i)\}^2=\frac{(n+1)(2n+1)}{6}-\left(\frac{n+1}{2}\right)^2=\frac{n^2-1}{12}$$

从而 $E(X)=\sum_{i=1}^{n}E(X_i)=\frac{k(n+1)}{2}$，$D(X)=\sum_{i=1}^{n}D(X_i)=\frac{k(n^2-1)}{12}$

二、方差的计算

典型例题

例 某人有一笔资金，可投入两个项目：房地产和商业，其收益都与市场状态有关。若把未来市场划分为好、中、差三个等级，其发生的概率分别为0.2、0.7、0.1。通过调查，该投资者认为投资于房地产的收益X(万元)和投资于商业的收益Y(万元)的分布分别如下，问如何投资好？

X	11	3	-3
P	0.2	0.7	0.1

Y	6	4	-1
P	0.2	0.7	0.1

三、方差的性质

- 设C是常数，则有$D(C)=0$

证明： $D(C)=E\{[C-E(C)]^2\}=0$

- 设X是一个r.v.，C是常数，则有

$$D(CX)=C^2D(X) \qquad D(X+C)=D(X)$$

证明： $$D(CX)=E\{[CX-E(CX)]^2\}$$
$$=C^2E\{[X-E(X)]^2\}=C^2D(X)$$
$$D(X+C)=E\{[X+C-E(X+C)]^2\}$$
$$=E\{[X-E(X)]^2\}=D(X)$$

三、方差的性质

设X，Y是两个r.v.，则有

$$D(X+Y)=D(X)+D(Y)+2E\{(X-E(X))(Y-E(Y))\}$$

特别，若X，Y相互独立，则有$D(X+Y)=D(X)+D(Y)$

证明: $D(X+Y)=E\{[(X+Y)-E(X+Y)]^2\}$

$$=E\{[(X-E(X))+(Y-E(Y))]^2\}$$

$$=E\{(X-E(X))^2\}+E\{(Y-E(Y))^2\}+2E\{[X-E(X)][Y-E(Y)]\}$$

$$=D(X)+D(Y)+2E\{(X-E(X))(Y-E(Y))\}$$

上式右端第三项:$2E\{(X-E(X))(Y-E(Y))\}$

$$=2E\{XY-XE(Y)-YE(X)+E(X)E(Y)\}$$

$$=2\{E(XY)-E(X)E(Y)-E(Y)E(X)+E(X)E(Y)\}$$

$$=2\{E(XY)-E(X)E(Y)\}$$

三、方差的性质

若X，Y相互独立，由数学期望的性质知上式右端为0，于是

$$D(X+Y)=D(X)+D(Y)$$

推广 若 $X_1, X_2, \cdots, X_n$ 相互独立,则有

$$D(X_1 \pm X_2 \pm \cdots \pm X_n) = D(X_1) + D(X_2) + \cdots + D(X_n).$$

$D(X)=0$的充要条件是X以概率1取常数$E(X)$，即 $P\{X=E(X)\}=1$

证明：[充分性]

设$P\{X=E(X)\}=1$，则有$P\{X^2=[E(X)]^2\}=1$，于是

$$D(X)=E(X^2)-[E(X)]^2=0$$

[证明必要性之前需先了解什么是切比雪夫不等式]

三、方差的性质

切比雪夫(Chebyshev)不等式

定 理

设r.v. X具有数学期望$E(X)=\mu$，方差$D(X)=\sigma^2$，则对于任意正数ε，不等式

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}\Leftrightarrow P\{|X-\mu|<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}.$$

证明： 我们只就连续型随机变量的情况来证明。

设r.v. X的概率密度为$f(x)$，则有

$$P\{|X-\mu|\geq\varepsilon\}=\int_{|x-\mu|\geq\varepsilon}f(x)dx\leq\int_{|x-\mu|\geq\varepsilon}\frac{|x-\mu|^2}{\varepsilon^2}f(x)dx$$

$$\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-\mu)^2f(x)dx=\frac{\sigma^2}{\varepsilon^2}$$

三、方差的性质

切比雪夫(Chebyshev)不等式

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}\Leftrightarrow P\{|X-\mu|<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}.$$

这个不等式给出了在r.v. X的分布未知的情况下，事件$\{|X-\mu|<\varepsilon\}$概率的下限的估计

例如：

$$P\{|X-\mu|<3\sigma\}\geq 0.8889$$

$$P\{|X-\mu|<4\sigma\}\geq 0.9375$$

随机调查中的频率估计概率问题

三、方差的性质

$D(X)=0$的充要条件是X以概率1取常数$E(X)$，即 $P\{X=E(X)\}=1$

证明：[必要性]

设$D(X)=0$，要证$P\{X=E(X)\}=1$

反证法

假设$P\{X=E(X)\}<1$，则对于某一个数$\varepsilon>0$，有

$$P\{|X-E(X)|\geq\varepsilon\}>0,$$

但由切比雪夫不等式，对于任意$\varepsilon>0$，因$\sigma^2=0$，有

$$P\{|X-E(X)|\geq\varepsilon\}=0$$

矛盾，于是

$$P\{X=E(X)\}=1$$

§2 方差

例：设随机变量(X,Y)服从二元正态分布，其密度函数为$f(x,y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$，$-\infty<x,y<\infty$, 令$Z=\sqrt{X^2+Y^2}$，求$Z$的方差。

例：设随机变量$X_1,X_2,\cdots X_n$相互独立，且都在区间[0，θ]上服从均匀分布，求：Y=max{$X_1,X_2,\cdots X_n$}, Z=min{$X_1,X_2,\cdots X_n$}的数学期望和方差。

三、协方差及相关系数的概念

若X，Y相互独立，则有$D(X+Y)=D(X)+D(Y)$

然而，若X，Y不相互独立，则有

$$D(X+Y)=D(X)+D(Y)+2\underbrace{\boxed{E\{(X-E(X))(Y-E(Y))\}}}_{\text{协方差}}$$

定义

量$E\{[X-E(X)][Y-E(Y)]\}$称为r.v. X与Y的协方差。记为$Cov(X,Y)$

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

而

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

称为r.v. X与Y的相关系数

三、协方差及相关系数的概念

定义说明

- X，Y的相关系数又称为标准协方差，它是一个无量纲的量
- 若X，Y相互独立，则有

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

- 若X，Y相互独立，则有

$$D(X+Y)=D(X)+D(Y)$$
$$+2E\{[X-E(X)][Y-E(Y)]\}$$
$$=D(X)+D(Y)+2Cov(X,Y)$$
$$=D(X)+D(Y)$$

- $Cov(X,Y)=Cov(Y,X)$ $\quad Cov(X,X)=D(X)$

三、协方差及相关系数的概念

典型例题

例 二维连续型r.v.(X,Y)的联合密度为：$f(x,y)=\begin{cases}\frac{1}{4}\left(1-x^3y+xy^3\right) & |x|<1,|y|<1\\ 0 & else\end{cases}$

求证：X,Y不相关，但不相互独立

证明 当$|x|<1$，$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-1}^{1}\frac{1}{4}\left(1-x^3y+xy^3\right)dy=\frac{1}{2}$

当$|x|\geqslant 1$，$f_X(x)=0$；故 $f_X(x)=\begin{cases}\frac{1}{2} & |x|<1\\ 0 & else\end{cases}$ 则 $E(X)=\int_{-1}^{1}\frac{1}{2}dx=0$

同理可以得到：$f_Y(x)=\begin{cases}\frac{1}{2} & |y|<1\\ 0 & else\end{cases}$ 则 $E(Y)=0$

$$E(XY)=\int_{-1}^{1}\int_{-1}^{1}xy\frac{1}{4}\left(1-x^3y+xy^3\right)dxdy=0$$

$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)=0\Rightarrow X,Y$不相关

在区域$|X|<1$, $|Y|<1$中，当 $x,y\neq 0$, $f(x,y)\neq f_X(x)f_Y(y)$ $\Rightarrow X,Y$不相互独立

三、协方差的计算

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

证明： 由协方差的定义及期望的性质，可得

$$\begin{aligned}Cov(X,Y)&=E\{[X-E(X)][Y-E(Y)]\}\\&=E[XY-YE(X)-XE(Y)+E(X)E(Y)]\\&=E(XY)-2E(X)E(Y)+E(X)E(Y)\\&=E(XY)-E(X)E(Y).\end{aligned}$$

三、协方差的性质

- (1) $Cov(X, Y)=Cov(Y, X)$；
- (2) $Cov(X, X)=D(X)$， $Cov(X, c)=0$；
- (3) $Cov(aX, bY)= abCov(X, Y)$， 其中a, b为 常数;
- (4) $Cov(X+Y, Z)=Cov(X, Z)+Cov(Y, Z)$；
- (5) 若X,Y相互独立，则$Cov(X, Y)=0$；
- (6) $D(X\pm Y)=D(X)+D(Y) \pm 2Cov(X, Y)$。

四、相关系数的性质

(1) $|\rho_{XY}| \le 1$;

(2) $|\rho_{XY}|=1 \Leftrightarrow$ 存在常数a, b 使$P\{Y= a+bX\}=1$。

证明： 考虑以$a+bX$来近似表示Y，以均方误差

$$e=E\{[Y-(a+bX)]^2\}=E(Y^2)+b^2E(X^2)+a^2-2bE(XY)$$
$$+2abE(X)-2aE(Y)。$$

来衡量$a+bX$近似Y的好坏。e 越小则$a+bX$与Y近似程度越好。为此求a， b使e达到最小值，那么

$$\begin{cases} \dfrac{\partial e}{\partial a}=2a+2bE(X)-2E(Y)=0, \\ \dfrac{\partial e}{\partial b}=2bE(X^2)-2E(XY)+2aE(X)=0. \end{cases}$$

四、相关系数的性质

解得 $b_0 = \dfrac{Cov(X,Y)}{D(X)}$, $a_0 = E(Y) - E(X)\dfrac{Cov(X,Y)}{D(X)}$.

将a_0，b_0代入$e = E[(Y-(a+bX))^2]$

得 $\min\limits_{a,b} E\{[Y-(a+bX)]^2\} = E\{[Y-(a_0+b_0X)]^2\} = (1-\rho_{XY}^2)D(Y)$

(1) 由 $E\{[Y-(a_0+b_0X)]^2\}$ 及$D(Y)$的非负性，得知$1-\rho_{XY}^2 \geq 0$，亦即 $|\rho_{XY}| \leq 1$。

(2) 若 ρ_{XY}=1，有 $E\{[Y-(a_0+b_0X)]^2\} = 0$

从而 $0 = E\{[Y-(a_0+b_0X)]^2\} = D[Y-(a_0+b_0X)] + [E(Y-(a_0+b_0X))]^2$

从而 $D[Y-(a_0+b_0X)] = 0$, $E[Y-(a_0+b_0X)] = 0$

由方差的性质知 $P\{Y-(a_0+b_0X)=0\}=1$, 即$P\{Y=a_0+b_0X\}=1$

四、相关系数的性质

反之，若存在常数a^*，b^*使

$$P\{Y=a^*+b^*X\}=1,\text{即}P\{Y-(a^*+b^*X)=0\}=1$$

于是，$P\left\{[Y-(a^*+b^*X)]^2=0\right\}=1$

即得，$E\left\{[Y-(a^*+b^*X)]^2\right\}=0$

故有，$0=E\left\{[Y-(a^*+b^*X)]^2\right\}\geq\min\limits_{a,b}E\left\{[Y-(a+bX)]^2\right\}$

$$=E\left\{[Y-(a_0+b_0X)]^2\right\}=(1-\rho_{XY}^2)D(XY)$$

即得，$\rho_{XY}=1$

五、相关系数的意义

相关系数是表征X, Y线性关系紧密程度的一个量。

当$|\rho_{XY}| = 1$时， X与Y之间以概率1存在着线性关系。

当$|\rho_{XY}|$越接近0时， X与Y之间的线性关系越弱。

当$|\rho_{XY}| =0$时， X与Y之间不存在线性关系(**不相关**)。

X与Y之间没有线性关系并不表示它们之间没有关系。

X与Y相互独立$\Rightarrow$不相关；但不相关$\not\Rightarrow$相互独立

五、相关系数的意义

典型例题

例 随机变量(X,Y)服从区域$D=\{(x,y)|0<x<1,0<x<y<1\}$上的均匀分布，试求相关系数$\rho_{XY}$。

例 掷一枚均匀的骰子两次，设X表示出现的点数之和，Y表示第一次出现的点数减去第二次出现的点数。求$D(X)$，$D(Y)$及ρ_{XY}，X与Y是否独立?

正态分布独立和不相关等价

§4 矩、协方差矩阵

一、基本概念

定义

设X和Y是随机变量，

- 若 $E(X^k)$, $k=1,2,\cdots$ 存在，称它为X的k阶原点矩，简称k阶矩
- 若$E\{[X-E(X)]^k\}$, $k=2,3,\cdots$ 存在，称它为X的k阶中心矩
- 若 $E(X^kY^l)$, $k,l=1,2,\cdots$ 存在，称它为X和Y的$k+l$阶混合矩
- 若 $E\{[X-E(X)]^k[Y-E(Y)]^l\}$, $k,l=1,2,\cdots$ 存在，称它为X和Y的$k+l$阶混合中心矩

§4 矩、协方差矩阵

一、基本概念

定义的说明

- 以上数字特征都是随机变量函数的数学期望

- 随机变量X 的数学期望$E(X)$是X的一阶原点矩；方差$D(X)$是二阶中心距；协方差$Cov(X,Y)$是X与Y的二阶混合中心距

- 在实际应用中，高于4阶的矩很少用。
  三阶中心距$E\{[X-E(X)]^3\}$主要用来衡量随机变量的分布是否有偏
  四阶中心距$E\{[X-E(X)]^4\}$主要用来衡量随机变量的分布在均值附近的陡峭程度如何

正态分布的n阶中心矩

一、基本概念

协方差矩阵

设$X_1, \dots, X_n$ 为n个随机变量，记

$$c_{ij}=Cov(X_i, X_j)=E\{[X_i-E(X_i)][X_j-E(X_j)]\}, \quad i, j=1, 2, \dots, n。$$

则称由c_{ij} 组成的矩阵为随机变量 $X_1, \dots, X_n$的**协方差矩阵**C。即

$$C=\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

由于$c_{ij}=c_{ji}\ (i\neq j,\ i, j=1,2,\dots,n)$， 故上述矩阵是**对称矩阵**。

一、基本概念

典型例题

例 设随机变量X的分布密度为

$$p(x)=\begin{cases}0.5x & 0<x<2\\ 0 & esle\end{cases}$$

求随机变量X的一至四阶原点矩和中心距

二、n维正态分布

n维正态随机变量概率密度的矩阵表示

二维情形

由于 $(X-\mu)^{\mathrm{T}}C^{-1}(X-\mu)$

$$=\frac{1}{\det C}(x_1-\mu_1,x_2-\mu_2)\begin{pmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}\begin{pmatrix}x_1-\mu_1\\ x_2-\mu_2\end{pmatrix}$$

$$=\frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

于是(X_1,X_2)的概率密度可写成

$$f(x_1,x_2)=\frac{1}{(2\pi)^{2/2}(\det C)^{1/2}}\exp\left\{-\frac{1}{2}(X-\mu)^T C^{-1}(X-\mu)\right\}.$$

二、n维正态分布

n维正态随机变量概率密度的矩阵表示

n维情形

对于n维正态随机变量$(X_1, X_2, \dots, X_n)$的情况，引入矩阵

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix}, \quad C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}.$$

n维正态随机变量$(X_1, X_2, \dots, X_n)$的情况，引入矩阵

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}(\det C)^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu)^T C^{-1}(X-\mu)\right\}.$$

二、n维正态分布

n维正态随机变量的重要性质

(1) $(X_1,X_2,\ldots,X_n)$的每一个分量X_i，$i=1,2,\ldots,n$ 也是正态变量。反之若X_i，$i=1, 2, \ldots, n$是正态变量且相互独立，则$(X_1, X_2, \ldots, X_n)$是n维正态变量。

(2) $(X_1,X_2,\ldots,X_n)$服从n 维正态分布的充要条件是$X_1,X_2,\ldots,X_n$的任意线性组合$k_1X_1+k_2X_2+\ldots+k_nX_n$（$k_1$，$k_2,\ldots$，$k_n$不全为零)服从一维正态分布。

(3) 若$(X_1,X_2,\ldots,X_n)$服从n 维正态分布， 设$Y_1,Y_2,\ldots,Y_k$是X_j，$j=1,2,\ldots,n$的线性函数，则$(Y_1,Y_2,\ldots,Y_k)$也服从多维正态分布。

(4) 设$(X_1,X_2,\ldots,X_n)$服从n 维正态分布，则$X_1,X_2,\ldots,X_n$相互独立与$X_1,X_2,\ldots,X_n$两两不相关等价。

谢　　谢!