目 录

第三章　随机变量的数字特征

§3.2　方差、标准差和矩

§3.2.1　方差和标准差

现在我们转到本章开始时候提到的另一类数字特征, 即刻画随机变量在其中心位置附近散布程度的数字特征, 其中最重要的是方差. 在实际应用中, 方差不仅是信息度量的标准也是风险度量的标准.

定义 3.2.1. *设X 为随机变量, 分布为F, 则称*

$$Var(X) = E(X - EX)^2 = \sigma^2$$

为X (或分布F)的方差, 其平方根$\sqrt{Var(X)} = \sigma$ (取正值) 称为X (或分布F)的标准差.

显然有

$$Var(X) = EX^2 - (EX)^2.$$

对随机变量的方差, 我们可以得到

定理 3.2.1. *设c为常数. 则有*

1. $0 \le Var(X) = EX^2 - (EX)^2$, *因此*$Var(X) \le EX^2$.
2. $Var(cX) = c^2 Var(X)$
3. $Var(X) = 0$*当且仅当*$P(X = c) = 1$, *其中*$c = EX$. *此时,我们称*X*退化到常数*c.
4. *对任何常数*c*有*, $Var(X) \le E(X - c)^2$, *其中等号成立当且仅当*$c = EX$.
5. *如果随机变量*X*和*Y*相互独立*, a, b*为常数. 则*$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$.

证明上述定理，我们介绍一个引理。

引理 3.2.1. *如果ξ为退化于0的随机变量，则有$E\xi^2 = 0$；反之，如果随机变量ξ的2阶矩存在而且$E\xi^2 = 0$，则ξ必为退化于0的随机变量.*

Proof. 如果ξ为退化于0的随机变量，则有$P(\xi = 0) = 1$, 故有$E\xi^2 = 0$。反之，如果随机变量ξ平方可积，并且$E\xi^2 = 0$，但是ξ不退化于0，则有$P(\xi = 0) < 1$。那么就存在$\delta > 0$和$0 < \epsilon < 1$，使得$P(|\xi| > \delta) > \epsilon$，于是$E\xi^2 > \delta^2\epsilon$。导致矛盾，所以$\xi$必退化到0. □

常见分布的方差：

1. 二项分布$X \sim B(n, p)$:

$$VarX = np(1-p)$$

2. Poisson 分布$X \sim P(\lambda)$:

$$VarX = \lambda$$

3. 均匀分布$X \sim U[a, b]$:

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

4. 指数分布$X \sim Exp(\lambda)$:

$$VarX = 1/\lambda^2$$

5. 正态分布$X \sim N(\mu, \sigma^2)$:

$$VarX = \sigma^2$$

由此得到正态分布$N(\mu, \sigma^2)$ 中另一参数σ^2 的解释: 它就是分布的方差, 正态分布完全由其均值μ 和方差σ^2 决定, 故也常称为“均值为μ 方差为σ^2 的正态分布”. 方差σ^2 越小, 则X的取值以更大的概率集中在其均值μ附近.

定义 3.2.2. *我们称*

$$X^* = \frac{X - EX}{\sqrt{Var(X)}}$$

为X的标准化随机变量. 易见$EX^* = 0, Var(X^*) = 1$.

我们引入标准化随机变量是为了消除由于计量单位的不同而给随机变量带来的影响. 例如, 我们考察人的身高, 那么当然可以以米为单位, 得到X_1, 也可以以厘米为单位, 得到X_2. 于是就有得到$X_2 = 100X_1$. 那么这样一来, X_2与X_1 的分布就有所不同. 这当然是一个不合理的现象. 但是通过标准化, 就可以消除两者之间的差别, 因为我们有$X_2^* = X_1^*$. 对于正态分布, 我们经过标准化$Y = (X - \mu)/\sigma$, 就可以得出均值为0方差为1的正态分布, 即标准正态分布.

§3.2.2 矩

下面我们引入矩的概念，并将之与我们前面所说的期望、方差建立联系.

定义 3.2.3. 设X为随机变量, c 为常数, r为正整数, 则$E[(X-c)^r]$ 称为X 关于c 点的r 阶矩.

比较重要的有两个情况:

1. $c=0$. 这时$\alpha_k=EX^r$ 称为X 的r 阶原点矩.

2. $c=EX$. 这时$\mu_k=E[(X-EX)^r]$ 称为X 的r 阶中心矩.

容易看出, 一阶原点矩就是期望, 二阶中心矩就是X的方差$Var(X)$.

§3.3 协方差和相关系数

现在我们来考虑多维随机向量的数字特征, 以二维的情况为例, 设(X,Y) 为二维随机变量, X, Y本身都是一维随机变量, 那么它们相应的均值方差, 我们都在上两节中讨论过了, 我们更有兴趣的数字特征是反映分量之间关系的那种量, 其中最重要的, 是本节要讨论的协方差和相关系数.

§3.3.1 协方差

定义 3.3.1. 我们称

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

为X与Y的协方差, 其中Cov是英文单词$Covariance$的缩写.

由协方差的定义, 我们立刻可以得到协方差具有如下性质:

1. $Cov(X,Y)=Cov(Y,X)$, $Cov(X,X)=Var(X)$

2. $Cov(X,Y)=EXY-EXEY$, 显然若X、Y 相互独立, 则$Cov(X,Y)=0$

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

4. 对任何实数a_1, a_2, b_1, b_2, 有

$$Cov(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_ib_jCov(X_i, Y_j)$$

如果$\xi_1, \cdots, \xi_n$是定义在同一概率空间下的随机变量，并且其中每个随机变量都是平方可积的。称矩阵

$$\begin{aligned}
\Sigma &= (b_{ij}) = (cov(\xi_i, \xi_j)) \\
&= \begin{pmatrix} D(\xi_1) & cov(\xi_1, \xi_2) & \cdots & cov(\xi_1, \xi_n) \\ cov(\xi_2, \xi_1) & D(\xi_2) & \cdots & cov(\xi_2, \xi_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\xi_n, \xi_1) & cov(\xi_n, \xi_2) & \cdots & D(\xi_n) \end{pmatrix}
\end{aligned}$$

为$\xi_1, \cdots, \xi_n$的协方差矩阵。显然$\Sigma \geq 0$。

例 3.3.1. 设$(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$，则$(X, Y)$的协方差矩阵为

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

§3.3.2 相关系数

定义 3.3.2. 设随机变量X, Y为随机变量, 称

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{VarX} \cdot \sqrt{VarY}},$$

为X与Y的相关系数. 当$\rho_{X,Y} = 0$时, 则称X与Y不相关.

由定义容易看出, 若令$X^* = (X - EX)/\sqrt{VarX}$和$Y^* = (Y - EY)/\sqrt{VarY}$ 分别为X和Y相应的标准化随机变量, 则$\rho_{X,Y} = Cov(X^*, Y^*)$. 因此, 形式上可以把相关系数视为"标准尺度下的协方差", 从这个角度上说, 相关系数可以更好的反映两个随机变量间的关系, 而不受它们各自所用度量单位的影响.

例 3.3.2. 设$(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$, 则$\rho_{X,Y} = \rho$.

相关系数有如下的性质:

1. 若X和Y相互独立, 则$\rho_{X,Y}=0$

2. $|\rho_{X,Y}|\le 1$, 等号成立当且仅当X,Y之间存在严格的线性关系, 即

$$\begin{aligned}\rho_{X,Y}=1, &\quad \text{则存在 } a>0, b\in\mathbb{R} \text{ 使得 } X=aY+b \quad (\text{正相关})\\ \rho_{X,Y}=-1, &\quad \text{则存在 } a<0, b\in\mathbb{R} \text{ 使得 } X=aY+b \quad (\text{负相关})\end{aligned}$$

[**注**]: $\rho_{X,Y}$ 也常称作X和Y线性相关系数, 只能刻画X和Y间的线性相依程度, $|\rho_{X,Y}|$ 越接近1, 就表示X, Y 间的线性相关程度越高; $|\rho_{X,Y}|=0$时, 只是表示X和Y间不存在线性相关, 但可以存在非线性的函数关系.

为证明2,我们看如下引理。

引理 3.3.1. *[$Cauchy-Schwarz\ Inequality$] 设$\xi,\eta$均平方可积，则有*

$$[E\xi\eta]^2\le E\xi^2E\eta^2$$

等号成立当且仅当$P(\xi=t_0\eta)=1$，其中t_0为一常数。

Proof. 易知, 对任何$t\in\mathcal{R}$, 都有

$$g(t):=E\eta^2\cdot t^2-2E\xi\eta\cdot t+E\xi^2=E(\xi-t\eta)^2\ge 0\ ,$$

所以二次函数$g(t)$的判别式

$$\Delta=4(E\xi\eta)^2-4E\xi^2\cdot E\eta^2\le 0,$$

故得不等式.

如果存在$t_0\in\mathcal{R}$, 使得$P(\xi=t_0\eta)=1$, 显然就有

$$(E\xi\eta)^2=E\xi^2E\eta^2.$$

反之, 如果不等式等号成立, 那么方程$g(t)=0$有唯一的实根t_0, 即有

$$E(\xi-t_0\eta)^2=g(t_0)=0,$$

于是由引理3.2.1知$\xi-t_0\eta$是退化于0的随机变量, 即有$P(\xi=t_0\eta)=1$. □

推论 3.3.1. *设随机变量ξ,η平方可积, 则有*

$$cov(\xi,\eta)\le\sqrt{D\xi}\cdot\sqrt{D\eta},$$

并且等号成立, 当且仅当存在$t_0\in\mathcal{R}$, 使得$P(\xi=t_0\eta)=1$.

例 3.3.3. 设$X \sim U(-\frac{1}{2}, \frac{1}{2})$, 而$Y = cosX$, 则

$$Cov(X,Y) = EXY = \int_{-1/2}^{1/2} xcosxdx = 0$$

所以X, Y不相关. 但是X, Y之间存在着非线性的函数关系.

定理 3.3.1. 对任何非退化的随机变量ξ, η平方可积, 如下四个命题相互等价:

(1) ξ与η不相关; (2) $cov(\xi, \eta) = 0$;

(3) $E\xi\eta = E\xi E\eta$; (4) $D(\xi + \eta) = D\xi + D\eta$.

下面我们来讨论不相关与独立性之间的关系.

定理 3.3.2. 对随机变量X, Y, 如果X与Y相互独立, 那么它们一定不相关; 但是如果它们不相关却未必相互独立.

例 3.3.4. 试证明若(X, Y)服从单位圆内的均匀分布, 则X, Y不相关但不独立.

解: 由(X, Y)服从单位圆内的均匀分布, 则(X, Y) 的联合密度函数

$$f(x,y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \le 1; \\ 0, & \text{其他}. \end{cases}$$

由此, 可得X和Y的边缘密度函数为

$$f_X(x) = f_Y(x) = \frac{2}{\pi}\sqrt{1 - x^2}, \quad -1 \le x \le 1.$$

因此, $EX = EY = 0$, 又

$$EXY = \int_{-1}^{1} x. \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y.\frac{1}{\pi} dydx = 0.$$

所以, $Cov(X, Y) = 0$, 从而$\rho_{X,Y} = 0$, 即X和Y不相关. 但由$f(x, y) \ne f_X(x).f_Y(y)$, 知$X$和$Y$显然不独立.

例 3.3.5. 设随机变量X和Y的分布律分别为

$$X \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}, \qquad Y \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

并且$P(X \cdot Y = 0) = 1$. 则X与Y不独立, 也不相关.

[注]: 只在正态情形下, 不相关与独立等价. 我们举二维正态的例子来说明, 不妨设$(X,Y)\sim N(a,b,\sigma_1^2,\sigma_2^2,\rho)$, 则$X$和$Y$独立等价于$\rho=\rho_{X,Y}=0$, 从而等价于$X$和$Y$不相关.

表 3.3.1 常见分布表

分布名称	参数	概率密度	期望	方差	特征函数
退化分布	c	$\binom{c}{1}$	c	0	e^{ict}
二点分布	p $(0<p<1)$	$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$	p	pq	$q+pe^{it}$
二项分布 $B(n,p)$	$n\geq 1$ $0<p<1$	$\binom{n}{k}p^k q^{n-k}$ $k=0,\cdots,n$	np	npq	$(q+pe^{it})^n$
几何分布	p $(0<p<1)$	$q^{k-1}p, k=1,2,\cdots$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1-qe^{it}}$
巴斯卡分布	r,p $r\in\mathbb{N}$ $0<p<1$	$\binom{k-1}{r-1}p^r q^{k-r}$, $k=r,r+1,\cdots$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$(\frac{pe^{it}}{1-qe^{it}})^r$
波松分布$P(\lambda)$	$\lambda(\lambda>0)$	$\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,\cdots$	λ	λ	$e^{\lambda(e^{it}-1)}$
超几何分布	$M,N,n\in\mathbb{N}$	$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	$\frac{nM}{N}$	$\frac{nM}{N}\frac{(N-M)}{N}\frac{N-n}{N-1}$	
均匀分布 $U(a,b)$	$a,b(a<b)$	$\frac{1}{b-a}I_{a<x<b}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
正态分布 $N(a,\sigma^2)$	a,σ^2	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$	a	σ^2	$e^{iat-\frac{1}{2}\sigma^2t^2}$
指数分布	$\lambda(\lambda>0)$	$\lambda e^{-\lambda x}I_{x>0}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$(1-\frac{it}{\lambda})^{-1}$
χ^2分布	$n(n\geq 1)$	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$ $x>0$	n	$2n$	$(1-2it)^{-n/2}$

§3.4 其他一些数字特征与相关函数

- 平均绝对差$E|X-EX|$
- 矩母函数Ee^{tX}, 其中$t\in\mathbb{R}$.
- 特征函数Ee^{itX}, 其中$t\in\mathbb{R}$, i 为虚数.

定义 3.4.1. 如果离散型随机变量X的分布律为$P(X = a_i) = p_i,\ i \in \mathbb{N}$, 那么

$$Ee^{itX} = \sum_{i=1}^{\infty} e^{ita_i} p_i.$$

如果连续型随机变量X的密度函数为$f(x)$, 那么

$$Ee^{itX} = \int_{-\infty}^{\infty} e^{itx} f(x) dx.$$