

3-2: 方差与矩

金百锁

课件 <http://bb.ustc.edu.cn/>

第三章随机变量的数字特征

4.2	方差和矩	1
4.2.1	方差 (Variance)	1
4.2.2	矩	10
4.3	协方差和相关系数	13
4.3.1	协方差	14
4.3.2	相关系数	17
4.4	其他一些数字特征与相关函数	25

3.2 方差和矩

3.2.1 方差 (Variance)

方差是刻画随机变量**在其中心位置附近的散布程度**. 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**.

甲乙两人射击水平如下所示

甲:	击中环数	8	9	10
	概率	0.3	0.1	0.6

乙:

击中环数	8	9	10
概率	0.2	0.5	0.3

试问两人谁的水平更稳定?

↑Example

↓Example

由上一讲知道, 两人每次射击的期望击中环数分别为 9.3 和 9.1, 因此, 两人射击 N 次的方差为

$$0.3N * (8 - 9.3)^2 + 0.1N * (9 - 9.3)^2 + 0.6N * (10 - 9.3)^2 = 0.81N$$

$$0.2N * (8 - 9.1)^2 + 0.5N * (9 - 9.1)^2 + 0.3N * (10 - 9.1)^2 = 0.49N$$

所以乙的水平更稳定。

设 X 为随机变量, 分布为 F , 若 X 平方可积, 则称

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

Definition

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

显然有

$$\text{Var}(X) = EX^2 - (EX)^2.$$

对随机变量的方差, 我们可以得到

定理 1. 设 c 为常数. 则有

1. $0 \leq \text{Var}(X) = EX^2 - (EX)^2$, 因此 $\text{Var}(X) \leq EX^2$.
2. $\text{Var}(cX) = c^2 \text{Var}(X)$
3. $\text{Var}(X) = 0$ 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, $\text{Var}(X) \leq E(X - c)^2$, 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 $\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)$.

证明上述定理, 我们介绍一个引理。

引理 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.

证明. 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 ξ 必退化到 0. □

常见分布的方差：

1. 二项分布 $X \sim B(n, p)$:

$$\text{Var}(X) = np(1 - p)$$

2. Poisson 分布 $X \sim P(\lambda)$:

$$\text{Var}(X) = \lambda$$

3. 均匀分布 $X \sim U[a, b]$:

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

4. 指数分布 $X \sim \text{Exp}(\lambda)$:

$$\text{Var}(X) = 1/\lambda^2$$

5. 正态分布 $X \sim N(\mu, \sigma^2)$:

$$\text{Var}(X) = \sigma^2$$

我们称

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

Definition

为 X 的**标准化随机变量**. 易见 $EX^* = 0, \text{Var}(X^*) = 1$.

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.

↑Example

某场考试中, 有一大题是选择题。每一小题的 4 个答案中只有一个是正确的, 若规定选对得 2 分, 不选 0 分, 选错扣 1 分, 当有一题你没有把握选择时, 你应该不选还是任选一个答案? 我们把答对的得分改为 3 分, 其余不变。问在此规则下你答还是不答?

↓Example

解: 令 X 表示随机选择答案后的得分, 则 X 的分布为

X	-1	2
P	$3/4$	$1/4$

所以由期望的定义,

$$EX = (-1) \times 3/4 + 2 \times 1/4 = -1/4$$

不选得 0 分。所以在这个规则下, 没有把握答题时, 以不选为最好策略。

仍以 X 表示回答的得分, 由于 $P(X = 3) = 1/4, P(X = -1) = 3/4$, 故 $EX = 0$, 答题平均得分和不答得分都是 0 分。但是两者的方差是不同的, 容易算出, $\text{Var}(X) = 3$, 得分会围绕 0 点有波动, 即答题是有风险的。如果你是一个冒险的人 (比如你答对, 你的成绩会在 85 分以上, 不答就是 84 分或以下了), 你就选择答题, 如果你是一个避险的人 (比如你估计已有 85 分或以上, 答错扣分后不能保证有 85 分), 你就选择不答。

3.2.2 矩

下面我们引入矩 (Moments) 的概念, 并将之与我们前面所说的期望、方差建立联系.

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.

Definition

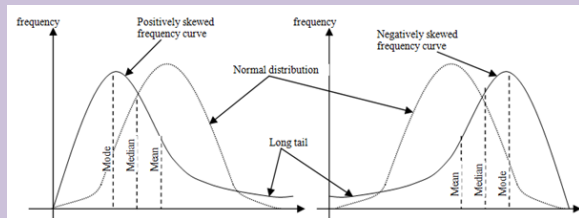
比较重要的有两个情况:

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 $Var(X)$.

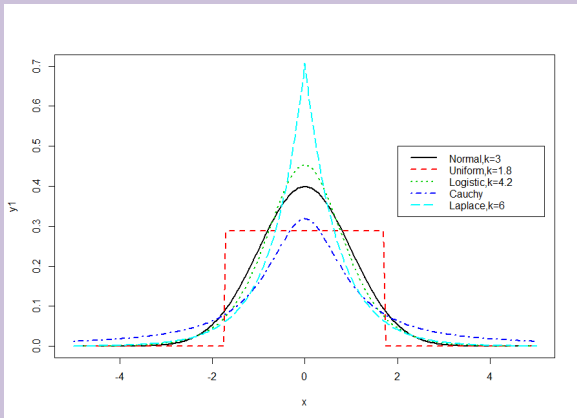
- 偏度系数

$$\gamma_1 = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] = \frac{\mu_3}{\sigma^3} = \frac{E [(X - \mu)^3]}{(E [(X - \mu)^2])^{3/2}}$$



- 峰度系数

$$\gamma_2 = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] = \frac{\mu_4}{\sigma^4} = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2}$$



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.

注意到

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E(X - EX)(Y - EY)$$

即

$X + Y$ 的波动性 = X 的波动性 + Y 的波动性 + X 和 Y 的相关性

3.3.1 协方差

如果随机变量 X 和 Y 平方可积, 我们称

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

Definition

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 $Cov(X, Y) = 0$

$$3. \text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$$

4. 对任何实数 a_1, a_2, b_1, b_2 , 有

$$\text{Cov}(a_1 X_1 + a_2 X_2, b_1 Y_1 + b_2 Y_2) = \sum_{i=1}^2 \sum_{j=1}^2 a_i b_j \text{Cov}(X_i, Y_j)$$

如果 ξ_1, \dots, ξ_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

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

为 ξ_1, \dots, ξ_n 的协方差矩阵。显然 $\Sigma \geq 0$ 。

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

[↑Example](#)

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

[↓Example](#)

3.3.2 相关系数

设随机变量 X, Y 为随机变量, 称

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

Definition

为 X 与 Y 的相关系数 (Coefficient of correlation).

当 $\rho_{X,Y} = 0$ 时, 则称 X 与 Y 不相关 (uncorrelated).

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.

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

[↑Example](#)

[↓Example](#)

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

$\rho_{X,Y} = 1$, 则存在 $a > 0, b \in \mathbb{R}$ 使得 $X = aY + b$ (正相关)

$\rho_{X,Y} = -1$, 则存在 $a < 0, b \in \mathbb{R}$ 使得 $X = aY + b$ (负相关)

[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 2. [*Cauchy – Schwarz Inequality*] 设 ξ, η 均平方可积, 则有

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

等号成立当且仅当 $P(\xi = t_0\eta) = 1$, 其中 t_0 为一常数。

证明. 易知, 对任何 $t \in \mathcal{R}$, 都有

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

所以二次函数 $g(t)$ 的判别式

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

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

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

反之, 如果等式等号成立, 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 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 1. 设随机变量 ξ, η 平方可积, 则有

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

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

↑Example

设 $X \sim U(-\frac{1}{2}, \frac{1}{2})$, 而 $Y = \cos X$, 证明 X, Y 不相关.
但是 X, Y 之间存在着非线性的函数关系.

↓Example

证: 由于 $EX = 0$,

$$E(XY) = E(X \cos(X)) = \int_{-1/2}^{1/2} x \cos x dx = 0$$

所以

$$\text{Cov}(X, Y) = EXY - EXEY = 0$$

即 X, Y 不相关. 但是 X, Y 之间存在着非线性的函数关系.

定理 2. 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) $\text{cov}(\xi, \eta) = 0$;
- (3) $E\xi\eta = E\xi E\eta$; (4) $\text{Var}(\xi + \eta) = \text{Var}(\xi) + \text{Var}(\eta)$.

下面我们来讨论不相关与独立性之间的关系.

定理 3. 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.

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.

[↑Example](#)

[↓Example](#)

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

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

由此, 可得 X 和 Y 的边缘密度函数为

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

因此, $EX = EY = 0$, 又

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

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

设随机变量 X 和 Y 的分布律分别为

↑Example

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

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

↓Example

[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.

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 $g(t) = Ee^{tX}$, 其中 $t \in \mathbb{R}$.
- 特征函数 $\phi(t) = Ee^{itX}$, 其中 $t \in \mathbb{R}$, i 为虚数.

定理 4. 对任何随机变量 X, Y , 分别有分布函数 F_X, F_Y 和特征函数 ϕ_X, ϕ_Y , 则

$$F_X = F_Y \leftrightarrow \phi_X = \phi_Y$$

表 3.1: 常见分布表

分布名称	参数	概率密度	期望	方差
退化分布	c	$\begin{pmatrix} c \\ 1 \end{pmatrix}$	c	0
二点分布	$p(0 < p < 1)$	$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$	p	pq
二项分布 $B(n, p)$	$n \geq 1, 0 < p < 1$	$\binom{n}{k} p^k q^{n-k}, k = 0, \dots, n$	np	npq
几何分布	$p(0 < p < 1)$	$q^{k-1} p, k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{q}{p^2}$
巴斯卡分布	r, p $r \in \mathbb{N}$ $0 < p < 1$	$\binom{k-1}{r-1} p^r q^{k-r},$ $k = r, r+1, \dots$	$\frac{r}{p}$	$\frac{rq}{p^2}$
泊松分布 $P(\lambda)$	$\lambda(\lambda > 0)$	$\frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \dots$	λ	λ
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 $U(a, b)$	$a, b(a < b)$	$\frac{1}{b-a} I_{a < x < b}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
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
指数分布	$\lambda(\lambda > 0)$	$\lambda e^{-\lambda x} I_{x>0}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
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	$2n$