数据科学基础知识

第一部分:概率论

ISEP第二年
2023-2024

基于 Nathalie Colin 和 Jean-Claude Guillerot 教授的课程

概率论

第五次会议（2023 年 10 月 27 日）：

第 7 章:期望值、特征
二维功能和时刻
随机变量

第 6 章（续）:高斯随机
变量

概率论

第五次会议（2023 年 10 月 27 日）：

第 7 章:期望值、特征
二维功能和时刻
随机变量

第 6 章（续）:高斯随机
变量

函数的期望值 (1/3)

提醒：随机变量函数的期望值

$$Y = g(X)$$

$$E(Y) = E(g(X)) = \int g(x) f_X(x)\, dx$$

定义： 2 个变量的函数的期望值

$$Z = g(X,Y)$$

连续案例

$$E(Z) = E(g(X,Y)) = \iint g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

离散案例

$$E(Z) = E(g(X,Y)) = \sum_{i,j} g(x_i, y_j) P_{i,j}$$

$P_{i,j}$: 二维分布

函数的期望值 (2/3)

备注:预期值的线性

克x, y( ) =å (我 kkk 克x, y)

例如g(x,y) 我 kkk 我 E g(x,y) kk(

总和的期望值=期望值的总和 =

例子: $$E(Z)= E(aX +bY)= aE(X)+bE(Y)$$

二维条件期望值:

到目前为止,我们已经看到: $$F_{和}\left(y \mid X = x\right) = \frac{F_{X,Y}(x,y)}{F_X(x)}$$

所以: $$E(Y|X = x) = \int y f_Y(y|X = x)dy = \frac{\int y f_{x,y}(x,y)dy}{f_X(x)}$$

函数的期望值 (3/3)

二维条件期望值（续）：

一般来说：

$$例如\left(g(X,Y)\mid X=x\right)=\int g(x,y)\,f(y\mid x)\,dy=\frac{\int g(x,y)\,f(x,y)\,dy}{F_X(x)}$$

相似地：

$$例如\left(g(X,Y)\mid Y=y\right)=\int g(x,y)\,f(x\mid y)\,dx=\frac{\int g(x,y)\,f(x,y)\,dx}{f_和(y)}$$

特征函数(1/2)

定义：

$$j_{XY,}\left(tt_{12}E\right) = \left(jt\,x\,jt\,y_{2}\right)$$

$$j_{XY,}\ _{12}(TT=)\ 不是e(x+y)dxdy\ XY$$

t1和t2 :确定性实变量

$j_{X}(t_1) = j_{XY,}(t_1, 0)$　　$j_{和}(t_2) = j_{XY,}(0, t_2)$　　$j_{坐标系}(0,0)=1$

定理：反演公式：

f和$_{X,Y}$　　$j_{X,Y}$

$$_{XY,}()\,fxy = \frac{1}{(2)p^{2}}\ 哦-+(jt\,x\,jt\,y_{2})\ j_{XY,}\left(tt_{12}\right)dt\,dt_{12}$$

特征函数 (2/2)

力矩生成函数：

$$\left.\frac{\partial^{在+} \varphi_{XY}(T_1,T_2)}{\partial T_1^{在} \partial T_2}\right|_{T_1=T_2=0} = j^{在+} E\left(XY^{在}\right)$$

证明类似于一维情况：特征函数的泰勒级数（在点（0,0)附近）。

$$\varphi_{坐标系}(t_1,t_2) = 1 + jE(X)\,t_1 + jE(Y)\,t_2 - \frac{1}{2}E\left(X^2\right)t_1^2 - \frac{1}{2}E\left(Y^2\right)t_2^2 - E(XY)\,t_1 t_2 + \cdots$$

一对随机变量的矩 (1/5)

定义

对于 1 个随机变量,阶矩为n：

$$= E_n X = x(f)x^n dx \quad ò \quad ^n \quad _X \; (\;)$$

对于 2 个随机变量， n阶矩使得n = l + k：

$$_{路克,} \quad ((\,,\,)^{路克}\, m) = E\text{不}XY^{路克} = xyf_{XY,}xy\, dxdy$$

对于 2 个离散随机变量：

$$(\,m_{l,k} = EXY^{路克}\,) = \text{å}_i\text{å}_j\, ij \quad \text{坐标}^{k\;我}_{j} \;,$$

评论：

Ø一阶矩有两个:E(X)和E(Y)

Ø二阶矩共有三个:E (X2)、 E (Y2)、 E(XY)

Ø一般来说， n阶矩有(n+ 1)个

一对随机变量的矩 (2/5)

定义（续）：

m10的核心时刻 = E(X) 和m01 =E（Y）：

米路克, = 前米米 ((- 1,0)(我 - 0,1)) k

米路克, == 不((xm)ymfxy dxdy - 1,0)(我 - 0,1) k XY,

$米_{0,0}$ = 1

$米_{1,0}$ = = $米_{0,1}$ 0

$米_{2,0}$ = s^2_{X} $_{0,2}$ 米 = s^2 和 稍等 $米_{1,1}$ XY = s , (X, Y)的协方差

·协方差也表示为cov (X, Y)

一对随机变量的矩 (3/5)

协方差：

$$, = ((\ - \quad ,)(\ - \quad ,)) = \quad ,$$

$$_{1,1} = \quad _{X,Y} = E(XY) - E(X)E(Y) = m1,1 - m1,0m0,1$$

其中m10 = E(X) 且m01 = E(Y)

备注： σX,Y可正可负。

~~施瓦茨不等式：二阶矩之间的关系~~

$$EXY^2 \pounds EXEY \quad (^2)(\,)^2$$

$m1,1^2$ £毫米2,0 0,2

对于零均值随机变量：

米^2 英镑1,1 米2,0 米0,2

一般来说：

$$s^2_{XY} \text{ 英镑 } s^2_X s^2 \text{ 和}$$

一对随机变量的矩(4/5)

相关系数:

$$= \frac{((\$ \quad ,)(\quad ,))}{\sqrt{((\$ \quad ,))(\$(\quad ,))}} = \frac{\quad , \quad}{, \quad ,} = \frac{,}{\quad}$$

特性:

相关系数衡量线性相关性
X 和 Y 之间

- $\boxed{-1 £ r £1}$ （使用柯西-施瓦茨不等式）

特殊值:

- r = ± 1 : X 和 X 之间存在完美的线性相关性
Y:　Y=aX+ b

- r = 0 : X 和 Y 不相关

$\boxed{E(XY) = E(X)E(Y)}$

2 个随机变量的矩 (5/5)

独立性和相关性：

独立性：

f (X, y) f (x) f (y)XY XY

m l,k E(XY) E(X)E(Y) 路克 = 我 k

毫米等克 l,k l,0 0,k ” l 和 k

对于 l = k = 1：

m1,1 = m1,0m0,1 => r = 0 所以 X 和 Y 去相关

独立性=> （意味着)去相关

去相关=>独立,除非 (X, Y) 是高斯分布

正交性 :E(XY) = 0

应用:线性回归（1/2）

目标:给定一对随机变量 (X,Y),且 X 固定,我们希望通过 X 的线性函数逼近Y ,表示为 g(X)： g(X) = aX+ b使得之间的均方误差Y和g(X)最小。

我们表示Z = Y − (aX + b)并且我们寻找a和b使得E(Z2) = E((Yg(X))2} = E((Y − (aX + b))2) (均方误差)最小。

通过应用期望值的一些属性,我们得到：

$$E(Z^2) = E((Y - aX)^2) + b^2 - 2bE(Y - aX)$$

应用:线性回归（2/2)

让我们表示ε = E(Z2) = E (Yg(X))2均方误差

我们计算导数以获得最小化ε的a和b值

和(!) ==E (!) + !E ! ( ) + ! − 2 ( ) − 2 ( ) + 2 ( )

衍生品: “# =E “$ (!) - ( ) + ( ) = 0

“# ”% = - ( ) + ( ) = 0

这使:

A = 埃西 (埃西) - () () 2 2 / 执行程序) - (( )) = s XY / s X 2 b E(Y)- aE(X) 和 r = s XY / s XYS

A = r s 和 / s X bE(Y)- r s 和 / s X 以前的) ® 这是 分钟 = s 2 和 (1 - ® CAS 2 2)号 情况r2=0;1

回归线

和 = r s 和 / s X (x - E(X)) E(Y)+

同样:如果Y=aX+b,我们毫不费力地发现， r=+-1。

结论,2个随机数的充要条件

变量之间呈线性相关,即|r|=1。

如果相关系数为零会发生什么?

如果r2 = 0 变量不相关,所以

E(XY)=E(X)E(Y) 应

用:当变量去相关时计算总和的方差。

(+=) ( ) + ( ) = (+)

备注:相关性是线性的度量

两个变量之间的依赖性。

即使两个变量非常密切相关,它们之间的相关性也可能很弱。这意味着,如果存在关系,则它不是线性的,例如,它可以是

多项式、指数等

总之

状况	然后，
如果	变量 X 和 Y 是:
$f_{XY}(x,y) = f_X(x)\, f_Y(y)$	独立且不相关。
$E(XY) = E(X)E(Y)$	不相关的
$E(XY) = 0$	正交
$r=\pm 1$	线性相关

方差公式

· Cov(X,Y) = E(XY) – E(X)E(Y) ·
Var(aX + bY) = a2Var(X) + b2Var(Y) + 2abCov(X,Y) ·
Cov(X, X) = Var(X) ·
Cov(aX + bY, cU + dV) = ac.Cov(X,U) + ad.Cov(X,V)
+ bc.Cov(Y,U) +
bd.Cov(Y,V) · Cov(aX + bY, cX + dY) = ac.Var(X) + (ad + bc)Cov(X,Y) +
bd.Var(Y)

随机向量或二维变量 (X,Y) 的协方差矩阵

$$V(X,Y) = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}$$

概率论

第五次会议（2023 年 10 月 27 日）：

第 7 章:期望值、特征
二维功能和时刻
随机变量

第 6 章（续）:高斯随机
变量

高斯随机变量

定义：

如果密度为,这对随机变量或二维向量 (X, Y) 是高斯分布的
形式：

$$f_{X,Y}(x,y) = e^{-\varphi(x,y)}$$

和

$$\varphi(x,y) = ax^2 + bxy + cy^2 + dx + ey + k$$

备注:仅取决于5个参数

k由下式确定： $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,dy = 1$

通常的介绍：

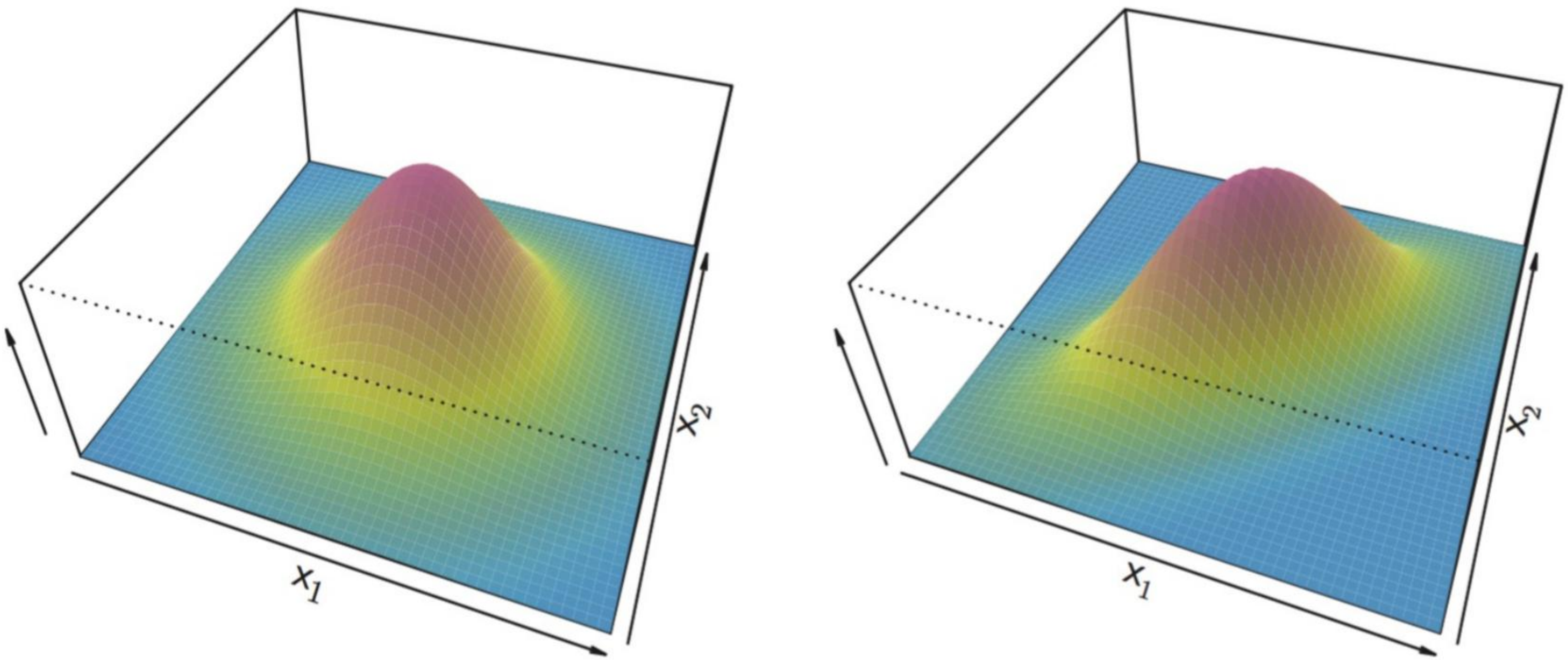
$$f_{X,Y}(x,y) = \frac{1}{2\pi s_X s_Y \sqrt{1-r^2}}\, e^{-\frac{1}{2(1-r^2)}\varphi(x,y)}$$

这5个参数是

$$\varphi(x,y) = \frac{(x-m_x)^2}{s_X^2} - \frac{2r(x-m_x)(y-m_y)}{s_X s_Y} + \frac{(y-m_y)^2}{s_Y^2}$$

mx、我的、 r、 sx和sy

高斯随机变量



左:等方差和零相关性,右:不同方差和现有相关性。

对于p维向量,密度为: ( ) = * ! #„ 0#„(102)$ %#(102)

(,) ” | |

其中 με p是期望值向量(均值向量), V是协方差矩阵。

边际密度

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{坐标系}(x, y)\, dx$$

Y 的边际：

$$(\ ) = \frac{\$(4)}{\sqrt{\ }}$$

对于 X 也类似：

$$(\ ) = \frac{\$(4)}{\sqrt{\ }}$$

MX，s_X ： X 的期望值和标准差

我的，$s_{和}$:Y 的期望值和标准差

二维高斯密度⇒高斯边缘密度

反之则不一定成立

r 是X 和 Y 之间的“相关系数”，

独立性：

X 和 Y 均为高斯分布，并且彼此独立。 ⟷ $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ 和 (X,Y) 高斯分布，r = 0

重要提示：两个随机变量之间的不相关性（即 r = 0 ）意味着仅在高斯向量的情况下独立。

高斯向量属性汇总表

一般来说

- X 和 Y 独立 ⇒ cov(X,Y)=0
- cov(X,Y)=0 X 和 Y 独立

定义:(X,Y) 是高斯向量 ⇔ ∀ , ε , X+ Y 是高斯随机变量。

- 如果 (X,Y) 是高斯向量且 X 和 Y 独立 ⇔ cov(X,Y)=0。
- (X,Y) 高斯向量 ⇒ X 高斯和 Y 高斯（特别是分别为 =0 和 =0 的情况）。
- (X 高斯和 Y 高斯) (X,Y) 高斯向量。
- (X 高斯、Y 高斯且 (X,Y) 独立)⇒ (X,Y) 高斯向量。 （当且仅当 cov(X,Y)=0 时相反）
- (X 高斯, Y 高斯且 (X,Y) 独立) ⇒ ∀ , ε , X+ Y 是高斯随机变量。

关于高斯向量的定理

定理:设 X为 p中的高斯随机向量，

服从 p 维高斯分布

参数（均值向量)和V （协方差矩阵）。如果A是确定性dxp 矩阵和向量 U ∈ d

, 然后：

$$E(AX + U) = A\,E(X) + U$$

$$Var(AX+U)=AVAT$$