

# CS7323 从数据学习因果关系

## 第一次作业

1. 请根据图 1 中所示表格的联合概率， 写出对应的条件概率的计算表达式（至少写出三个）， 并说明其意义、以及条件概率和联合概率的差异。

Y \ X	Time studied (minutes)				
		$x_1$ (0-20)	$x_2$ (21-40)	$x_3$ (41-60)	$x_4$ (>60)
% correct	$y_1$ (0-20)	2/200	0	0	8/200
	$y_2$ (21-40)	10/200	2/200	8/200	0
	$y_3$ (41-59)	2/200	4/200	32/200	32/200
	$y_4$ (60-79)	0	20/200	30/200	10/200
	$y_5$ (80-100)	0	4/200	16/200	20/200
	$p_X(x) \rightarrow$	14/200	30/200	86/200	70/200
	$p_Y(y) \downarrow$				

Table.3 Two-way table of dataset of the relationship in a classroom of 200 students between the amount of time studied and the percent correct

图 1

2. 图 2 中， 随机变量  $X_1$  和  $X_2$  服从正态分布， 分别在  $X_1$  和  $X_2$  中取两点 a 和 b (如图)， 在 a 和 b 的连线上， 以垂直于  $X_1 X_2$  平面对  $X_1$  和  $X_2$  的联合分布进行切片， 请写出所得到的切面对应的条件概率表达式。

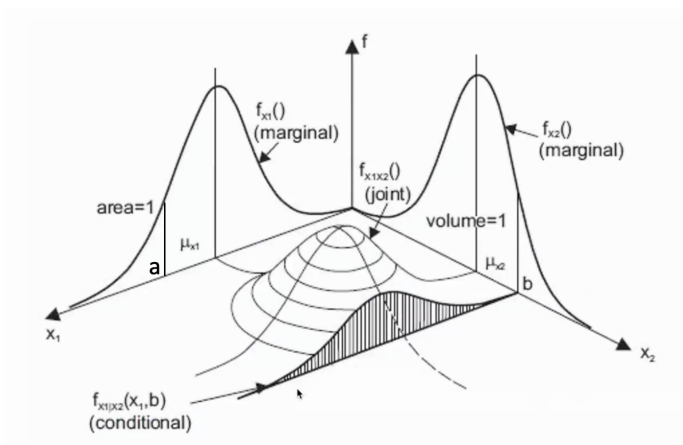


图 2

3. 图 3 所示公式中，  $X$  和  $Y$  为两个随机变量，  $a$ ，  $b$  为常数， 请解释图中公式的含义。

$$\begin{aligned}
\operatorname{cov}(X, a) &= 0 \\
\operatorname{cov}(X, X) &= \operatorname{var}(X) \\
\operatorname{cov}(X, Y) &= \operatorname{cov}(Y, X) \\
\operatorname{cov}(aX, bY) &= ab \operatorname{cov}(X, Y) \\
\operatorname{cov}(X + a, Y + b) &= \operatorname{cov}(X, Y)
\end{aligned}$$

图 3

4. 请证明以下概率论形式的柯西不等式。

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2] \cdot \mathbb{E}[Y^2]$$

5. (1) 两个变量  $X$  和  $Y$  的关系分布形式如图 4 所示, 请说明为什么这两变量的协方差  $\operatorname{cov}(X, Y)$  都为 0。

(2) 若已知变量  $X$  是取值范围为  $[-1, 1]$  的均匀分布,  $Y = X^2$  请计算  $X$  和  $Y$  的协方差  $\operatorname{cov}(X, Y)$ 。

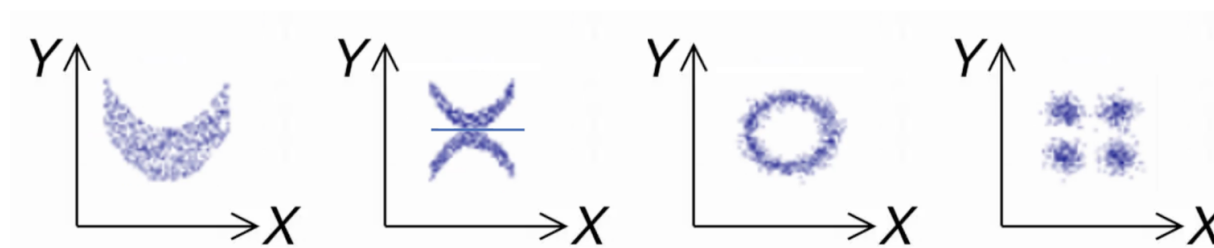


图 4