Multivariate Statistical Analysis

多元统计分析

2026 年 3 月 24 日

已学知识点 (Recap)

第 3 章 多元中的一些概念与模型

3.1 协方差

- 协方差是相关性的一种度量.
- 协方差仅度量线性相关程度.
- 协方差与度量单位有关.
- 协方差为零也可能存在非线性相关.
- 协方差为零不一定独立.
- 独立协方差一定为零.
- 协方差为负对应着向下倾斜的散点图. 协方差为正对应着向上倾斜的散点图.
- 变量与自身的协方差就是方差 $\mathrm{Cov}(X, X) = \sigma_{XX} = \sigma_X^2$.
- 对于较小的 n，在计算协方差时，我们用 $\frac{1}{n-1}$ 代替 $\frac{1}{n}$.

已学知识点 (Recap)

第 3 章 多元中的一些概念与模型

3.2 相关系数

- 相关系数是标准化之后对相关性的一种度量.
- 相关系数的绝对值不超过 1.
- 相关系数仅度量线性相关程度.
- 不相关亦可能存在非线性的相关.
- 不相关不一定独立. 但对两个正态分布的变量而言，不相关则意味着独立.
- 独立一定不相关.
- 负相关对应着向下倾斜的散点图. 正相关对应着向上倾斜的散点图.
- Fisher 的 Z 变换 $W = \frac{1}{2}\log\left(\frac{1+r_{XY}}{1-r_{XY}}\right)$ 可用于相关系数的假设检验.
- 对小样本，可用修正后的 Fisher 的 Z 变换 $W^* = W - \frac{3W+\tanh(W)}{4(n-1)}$.

已学知识点 (Recap)

第 3 章 多元中的一些概念与模型

3.3 描述性统计

- 数据矩阵的重心由其均值向量 $\bar{\boldsymbol{x}} = \frac{1}{n}\mathscr{X}^{\mathrm{T}}\mathbf{1}_n$ 给出.

- 数据矩阵观测值的离散程度由样本协方差矩阵给出:

$$\mathscr{S} = \frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{x}_k - \bar{\boldsymbol{x}}\right)\left(\boldsymbol{x}_k - \bar{\boldsymbol{x}}\right)^{\mathrm{T}} = \frac{1}{n}\mathscr{X}^{\mathrm{T}}\mathscr{X} - \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^{\mathrm{T}} = \frac{1}{n}\mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}$$

$\mathscr{S}_u = \frac{n}{n-1}\mathscr{S}$ 是总体协方差矩阵 Σ 的无偏估计

样本协方差矩阵 $\mathscr{S}$ 非负定: $\mathscr{S} \geq 0$

中心化矩阵: $\mathscr{H} = \mathscr{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$

$\mathscr{H}^{\mathrm{T}} = \mathscr{H}, \quad \mathscr{H}^2 = \mathscr{H}$

- 样本相关矩阵由 $\mathscr{R} = \mathscr{D}^{-1/2}\mathscr{S}\mathscr{D}^{-1/2}$ 给出.

对角阵: $\mathscr{D}^{-1/2} = \mathrm{diag}\left(s_{X_1X_1}^{-1/2}, s_{X_2X_2}^{-1/2}, \cdots, s_{X_pX_p}^{-1/2}\right)$

已学知识点 (Recap)

第 3 章 多元中的一些概念与模型

- 线性变换：$\boldsymbol{Y} = \mathscr{A}_{q\times p}\boldsymbol{X}$　　变换后的数据矩阵: $\mathscr{Y} = \mathscr{X}\mathscr{A}^{\mathrm{T}}$　（初始数据矩阵）

$$\bar{\boldsymbol{y}} = \mathscr{A}\,\bar{\boldsymbol{x}}$$

$$\mathscr{S}_{\mathscr{Y}} = \mathscr{A}\mathscr{S}_{\mathscr{X}}\mathscr{A}^{\mathrm{T}}$$

- Mahalanobis 变换：$z_i = \mathscr{S}^{-1/2}\left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}\right)$,　$i = 1, 2, \ldots, n$

变换后的数据矩阵: $\mathscr{Z} = \left(z_1, z_2, \ldots, z_n\right)^{\mathrm{T}} = \mathscr{H}\mathscr{X}\mathscr{S}^{-1/2}$

$\mathscr{S}_{\mathscr{Z}} = \mathscr{I}_p$,　消除了变量之间的相关性

已学知识点 (Recap)

第 3 章 多元中的一些概念与模型

- 简单线性回归模型

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

$\rightarrow \mathbb{E}(\varepsilon_i) = 0$, $\mathbb{V}\mathrm{ar}(\varepsilon_i) = \sigma^2$, or $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$

参数估计 $\Longrightarrow$ $\hat{\beta}_1 = \dfrac{s_{XY}}{s_{XX}}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$\widehat{\sigma^2} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

总平方和分解公式： $\displaystyle\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

决定系数： $r^2 = \dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \dfrac{\text{回归平方和}}{\text{总平方和}} = 1 - \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

```
> summary(wages.lm)

Call:
lm(formula = logwage ~ education, data = wages)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.78239 -0.25265  0.01636  0.27965  1.61101 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.239194   0.054974   22.54   <2e-16 ***
education   0.078600   0.004262   18.44   <2e-16 ***
---
Signif. codes  '**' 0.001 '**' 0.01

Residual standard error: 0.4038 on 2176 degrees of freedom
Multiple R-squared:  0.1351,    Adjusted R-squared:  0.1347 
F-statistic:   340 on 1 and 2176 DF,  p-value: < 2.2e-16
```

Model : $\texttt{logwage} = \beta_0 + \beta_1 \times \texttt{education} + \varepsilon\,, \qquad \varepsilon \sim N\left(0,\, \sigma^2\right)$

or $\qquad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,, \quad \varepsilon_i \overset{\text{i.i.d.}}{\sim} \left(0,\, \sigma^2\right), \quad i = 1,\, 2,\, \ldots,\, 2178$

$e_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right), \quad i = 1,\, 2,\, \ldots,\, 2178$

(Intercept): Estimate → $\hat{\beta}_0$; Std. Error → $\mathrm{SE}\left(\hat{\beta}_0\right)$; t value → $= \dfrac{\hat{\beta}_0}{\mathrm{SE}\left(\hat{\beta}_0\right)}$; Pr(>|t|) → $2 \times P\left\{t_{2176} > |22.54|\right\}$

education: Estimate → $\hat{\beta}_1$; Std. Error → $\mathrm{SE}\left(\hat{\beta}_1\right)$; t value → $= \dfrac{\hat{\beta}_1}{\mathrm{SE}\left(\hat{\beta}_1\right)}$; Pr(>|t|) → $2 \times P\left\{t_{2176} > |18.44|\right\}$

*** 显著性标志

$\hat{\sigma} \leftarrow$ Residual standard error: 0.4038

2176 degrees of freedom $\rightarrow n - 2$

Multiple R-squared: 0.1351 → $R^2 = \dfrac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \dfrac{\mathrm{SSE}}{\mathrm{SST}} = \widehat{\mathrm{Cor}}\,(\boldsymbol{X},\, \boldsymbol{Y})^2$

已学知识点 (Recap)

第 3 章 多元中的一些概念与模型

- 单因子方差分析 (ANOVA)

$$y_{ij} = \mu_j + \varepsilon_{ij} = \left(\mu + \alpha_j\right) + \varepsilon_{ij}, \quad i = 1, 2, \ldots, m\,; \; j = 1, 2, \ldots, p$$

第 j 个水平的均值（指 μ_j 与 $\mu + \alpha_j$）

$$\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N\left(0, \sigma^2\right)$$

问题

$$\begin{cases} \mathrm{H}_0 & : \quad \mu_1 = \mu_2 = \cdots = \mu_p = \mu \\ \mathrm{H}_1 & : \quad \mu_k \neq \mu_l \quad \text{for some } k \text{ and } l \end{cases}$$

- ▸ 单因子方差分析模型中的大部分信息都包含在方差分析表中

Source	SS	df	MS	E(MS)
Treatment	$\mathrm{SS}_{\text{treatment}} = m\sum_{j=1}^{p}\left(\bar{y}_j - \bar{y}\right)^2$	$p-1$	$\frac{\mathrm{SS}_{\text{treatment}}}{p-1}$	$\sigma^2 + \frac{m}{p-1}\sum_{j=1}^{p}\alpha_j^2$
Error	$\mathrm{SS}_{\text{full}} = \sum_{i=1}^{m}\sum_{j=1}^{p}\left(y_{ij} - \bar{y}_j\right)^2$	$p(m-1)$	$\frac{\mathrm{SS}_{\text{full}}}{p(m-1)}$	σ^2

- ▸ 我们可以用 $\Re$ 中的 aov() 函数来作方差分析

```
sales.aov <- aov(Sales ~ Strategy, data = Pullover)
summary(sales.aov)
```

```
> summary(sales.aov)
            Df Sum Sq Mean Sq F value  Pr(>F)   
Strategy     2  102.6   51.30   8.783 0.00115 **
Residuals   27  157.7    5.84                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

已学知识点 (Recap)

第 3 章 多元中的一些概念与模型

- 多重线性回归模型： 无截距模型： $\boldsymbol{y} = \mathscr{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathscr{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

回归系数： $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$ 误差项： $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

响应变量 Y，解释变量 $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_p\right)^{\mathrm{T}}$.

参数估计 $\Longrightarrow$ $\hat{\boldsymbol{\beta}} = \left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}\boldsymbol{y}$

拟合值 $\Longrightarrow$ $\hat{\boldsymbol{y}} = \mathscr{X}\hat{\boldsymbol{\beta}} = \mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}\boldsymbol{y} = \mathscr{P}\boldsymbol{y}$ 向量 $\boldsymbol{y}$ 在列空间 $C\left(\mathscr{X}\right)$ 上的投影

残差 $\Longrightarrow$ $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \mathscr{P}\boldsymbol{y} = \left(\mathscr{I}_n - \mathscr{P}\right)\boldsymbol{y} = \mathscr{Q}\boldsymbol{y}$ 向量 $\boldsymbol{y}$ 在 $C\left(\mathscr{X}\right)$ 的正交补上的投影

决定系数：
$$r^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{\text{回归平方和}}{\text{总平方和}} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

调整后的决定系数：
$$r_{\text{adj}}^2 = r^2 - \frac{p\left(1 - r^2\right)}{n - p - 1}$$

已学知识点 (Recap)

第 3 章 多元中的一些概念与模型

- 多重线性回归模型：

```
> summary(pullover.lm2)

Call:
lm(formula = X1 ~ ., data = pullover)

Residuals:
    Min      1Q  Median      3Q     Max 
-13.369  -9.406   1.599   5.151  19.729 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 65.66956   57.12507   1.150  0.29407   
X2          -0.21578    0.32194  -0.670  0.52764   
X3           0.48519    0.08678   5.591  0.00139 **
X4           0.84373    0.37400   2.256  0.06491 . 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.7 on 6 degrees of freedom
Multiple R-squared:  0.9067,    Adjusted R-squared:  0.8601 
F-statistic: 19.44 on 3 and 6 DF,  p-value: 0.001713
```

$\mathrm{H}_0:\ \beta_1=0 \leftrightarrow \mathrm{H}_1:\ \beta_1\neq 0 \implies$ 接受 H_0

$\mathrm{H}_0:\ \beta_2=0 \leftrightarrow \mathrm{H}_1:\ \beta_2\neq 0 \implies$ 拒绝 H_0

$\mathrm{H}_0:\ \beta_3=0 \leftrightarrow \mathrm{H}_1:\ \beta_3\neq 0 \implies$ 拒绝 H_0

$$\begin{cases}\mathrm{H}_0:\ \beta_1=\beta_2=\beta_3=0\\ \mathrm{H}_1:\ \text{at least one } \beta_j\neq 0\end{cases} \implies \text{拒绝 } \mathrm{H}_0$$

Multivariate Distributions

多　元　分　布

概述

多元分布

- 分布函数与密度函数 (Distribution and Density Function)
- 矩与特征函数 (Moments and Characteristic Functions)
- 变量变换 (Transformations)
- 多元正态分布 (The Multinormal Distribution)
- 抽样分布与极限定理 (Sampling Distributions and Limit Theorems)
- 厚尾分布 (Heavy-Tailed Distributions)
- 自助法 (Bootstrap)

分布函数与密度函数

随机向量 $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$

$\boldsymbol{X}$ 的累积分布函数 (cumulative distribution function) 定义为

$$F(\boldsymbol{x}) = P(\boldsymbol{X} \leq \boldsymbol{x}) = P\left(X_1 \leq x_1,\, X_2 \leq x_2,\, \ldots,\, X_p \leq x_p\right)$$

- 对连续型随机变量 $\boldsymbol{X}$，存在一个非负的概率密度函数 (probability density function) f 使得

$$F(\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}} f(\boldsymbol{u}) d\boldsymbol{u} = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f\left(u_1,\, u_2,\, \ldots,\, u_p\right) du_1 \cdots du_p$$

$$\int_{-\mathbb{R}^p} f(\boldsymbol{u}) d\boldsymbol{u} = 1\ ,\quad f(\boldsymbol{x}) = \frac{\partial^p F(\boldsymbol{x})}{\partial x_1 \cdots \partial x_p}$$

分布函数与密度函数

随机向量 $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$

$\boldsymbol{X}$ 的累积分布函数 (cumulative distribution function) 定义为

$$F(\boldsymbol{x}) = P(\boldsymbol{X} \leq \boldsymbol{x}) = P\left(X_1 \leq x_1,\ X_2 \leq x_2,\ \ldots,\ X_p \leq x_p\right)$$

- 对离散型随机变量 $\boldsymbol{X}$，我们关注该变量取值在可列个或有限个点 $\left\{\boldsymbol{c}_j\right\}_{j\in J}$ 的概率，

$$P(\boldsymbol{X} \in D) = \sum_{\left\{j:\ \boldsymbol{c}_j \in D\right\}} P\left(\boldsymbol{X} = \boldsymbol{c}_j\right)$$

分布函数与密度函数

- 若将随机向量 $X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \\ X_{k+1} \\ \vdots \\ X_p \end{pmatrix}$ 分块为 $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$，其中 $\boldsymbol{X}_1 \in \mathbb{R}^k$，$\boldsymbol{X}_2 \in \mathbb{R}^{p-k}$，则称

$$F_{X_1}(\boldsymbol{x}_1) = P(\boldsymbol{X}_1 \le \boldsymbol{x}_1) = F(x_1, \ldots, x_k, \infty, \ldots, \infty)$$

为**边际** (marginal) **分布函数** (cdf). 称 $F = F(\boldsymbol{x})$ 为**联合** (joint) **分布函数** (cdf).

- 对连续型随机向量 $\boldsymbol{X}$，$f_{X_1}(\boldsymbol{x}_1) = \int_{-\infty}^{\infty} f(\boldsymbol{x}_1, \boldsymbol{x}_2)\, d\boldsymbol{x}_2$.

- 给定 $\boldsymbol{X}_1 = \boldsymbol{x}_1$ 时，$\boldsymbol{X}_2$ 的条件概率密度为 $f\left(\boldsymbol{x}_2 \middle| \boldsymbol{x}_1\right) = \dfrac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f_{X_1}(\boldsymbol{x}_1)}$.

分布函数与密度函数

- **例：** 考虑函数

$$f(x_1, x_2) = \begin{cases} \frac{1}{2}x_1 + \frac{3}{2}x_2, & 0 \le x_1, x_2 \le 1 \\ 0, & \text{其它} \end{cases}$$

概率密度函数

$$f(x_1, x_2) \ge 0, \quad \iint f(x_1, x_2)\, dx_1 dx_2 = \frac{1}{2} \cdot \frac{x_1^2}{2}\Big|_0^1 + \frac{3}{2} \cdot \frac{x_2^2}{2}\Big|_0^1 = \frac{1}{4} + \frac{3}{4} = 1$$

边际密度：

$$\begin{cases} f_{X_1}(x_1) = \int f(x_1, x_2)\, dx_2 = \int_0^1 \left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right) dx_2 = \frac{1}{2}x_1 + \frac{3}{4}, & 0 \le x_1 \le 1 \\ f_{X_2}(x_x) = \int f(x_1, x_2)\, dx_1 = \int_0^1 \left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right) dx_1 = \frac{3}{2}x_2 + \frac{1}{4}, & 0 \le x_2 \le 1 \end{cases}$$

条件密度：

$$\begin{cases} f(x_2 | x_1) = \frac{f(x_1, x_2)}{f(x_1)} = \frac{\frac{1}{2}x_1 + \frac{3}{2}x_2}{\frac{1}{2}x_1 + \frac{3}{4}}, & 0 \le x_2 \le 1, (0 \le x_1 \le 1) \\ f(x_1 | x_2) = \frac{f(x_1, x_2)}{f(x_2)} = \frac{\frac{1}{2}x_1 + \frac{3}{2}x_2}{\frac{3}{2}x_2 + \frac{1}{4}}, & 0 \le x_1 \le 1, (0 \le x_2 \le 1) \end{cases}$$

分布函数与密度函数

定义 4.1 X_1 与 X_2 相互独立的充分必要条件是 $f(\boldsymbol{x}) = f(\boldsymbol{x}_1, \boldsymbol{x}_2) = f_{X_1}(\boldsymbol{x}_1) f_{X_2}(\boldsymbol{x}_2)$.

- X_1 与 X_2 相互独立当且仅当 $f(\boldsymbol{x}_1 | \boldsymbol{x}_2) = f_{X_1}(\boldsymbol{x}_1)$，$f(\boldsymbol{x}_2 | \boldsymbol{x}_1) = f_{X_2}(\boldsymbol{x}_2)$.
- 独立性可解释如下：已知 $X_2 = x_2$ 不会改变 X_1 的概率，反之亦然.

! 不同的联合概率密度函数可能会有相同的边际概率密度函数.

分布函数与密度函数

- 例：考虑下述两个概率密度函数

$$f(x_1, x_2) = \begin{cases} 1, & 0 < x_1, x_2 < 1 \\ 0, & \text{其它} \end{cases}$$

$$f(x_1, x_2) = \begin{cases} 1 + \alpha(2x_1 - 1)(2x_2 - 1), & 0 < x_1, x_2 < 1 \\ 0, & \text{其它} \end{cases}, \quad (-1 \le \alpha \le 1)$$

$$f_{X_1}(x_1) = \begin{cases} 1, & 0 < x_1 < 1 \\ 0, & \text{其它} \end{cases}, \quad f_{X_2}(x_2) = \begin{cases} 1, & 0 < x_2 < 1 \\ 0, & \text{其它} \end{cases}$$

$$f_{X_1}(x_1) = \int_0^1 \left[1 + \alpha(2x_1 - 1)(2x_2 - 1)\right] dx_2 = 1, \quad 0 < x_1 < 1$$

$$f_{X_2}(x_2) = \int_0^1 \left[1 + \alpha(2x_1 - 1)(2x_2 - 1)\right] dx_1 = 1, \quad 0 < x_2 < 1$$

分布函数与密度函数

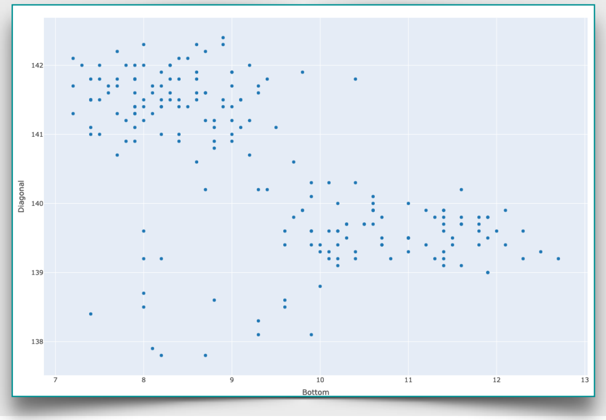
- 例：利用钞票数据集讨论独立性.

```
library(mclust)
data(banknote)
# 钞票数据的相关矩阵
round(cor(banknote[, 2:7]), digits = 2)
```

```
> # 钞票数据的相关矩阵
> round(cor(banknote[, 2:7]), digits = 2)
         Length  Left Right Bottom   Top Diagonal
Length     1.00  0.23  0.15  -0.19 -0.06     0.19
Left       0.23  1.00  0.74   0.41  0.36    -0.50
Right      0.15  0.74  1.00   0.49  0.40    -0.52
Bottom    -0.19  0.41  0.49   1.00  0.14    -0.62
Top       -0.06  0.36  0.40   0.14  1.00    -0.59
Diagonal   0.19 -0.50 -0.52  -0.62 -0.59     1.00
```

X_4 与 X_6 几乎不独立.

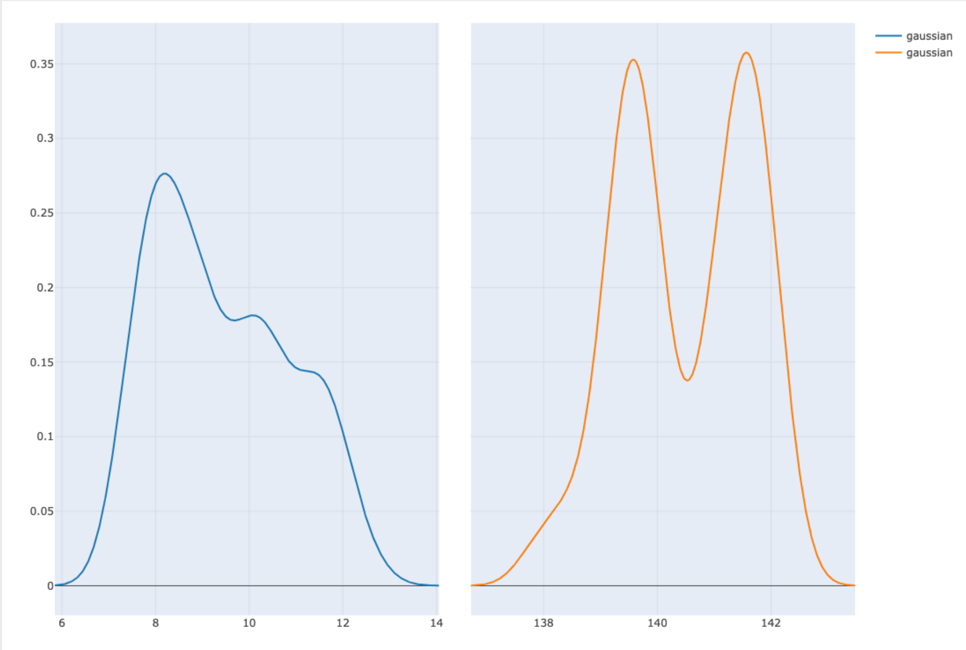
```
# X4与X6的散点图
library(ggplot2)
library(plotly)
fig = plot_ly(banknote, x = ~Bottom, y = ~Diagonal, type = "scatter", mode = "markers") %>%
  layout(plot_bgcolor = '#e5ecf6',
         xaxis = list(zerolinecolor = '#ffff', zerolinewidth = 2, gridcolor = 'ffff'),
         yaxis = list(zerolinecolor = '#ffff', zerolinewidth = 2, gridcolor = 'ffff'))
fig
```



分布函数与密度函数

- 例：利用钞票数据集讨论独立性.

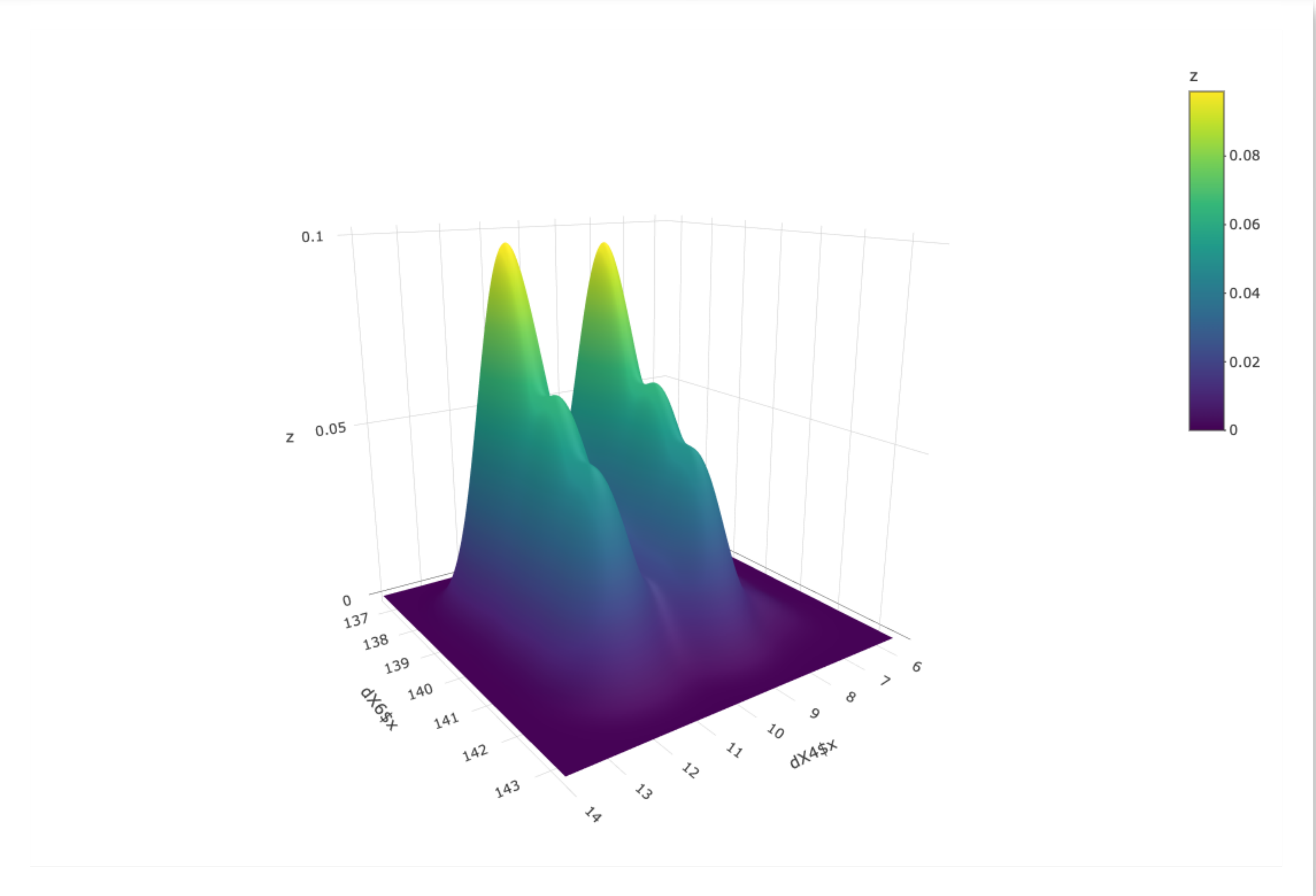
```
# X4与X6的核密度图
fig1 = plot_ly()
dX4 = density(banknote$Bottom, kernel = "gaussian", na.rm = TRUE)
fig1 = add_lines(fig1, x = dX4$x, y = dX4$y, name = "gaussian") %>%
  layout(plot_bgcolor='#e5ecf6',
         xaxis = list(zerolinecolor = '#ffff', zerolinewidth = 2, gridcolor = 'ffff'),
         yaxis = list(zerolinecolor = '#ffff', zerolinewidth = 2, gridcolor = 'ffff'))

fig2 = plot_ly()
dX6 = density(banknote$Diagonal, kernel = "gaussian")
fig2 = add_lines(fig2, x = dX6$x, y = dX6$y, name = "gaussian") %>%
  layout(plot_bgcolor='#e5ecf6',
         xaxis = list(zerolinecolor = '#ffff', zerolinewidth = 2, gridcolor = 'ffff'),
         yaxis = list(zerolinecolor = '#ffff', zerolinewidth = 2, gridcolor = 'ffff'))

subplot(fig1, fig2, shareY = TRUE)
```



分布函数与密度函数

- 例：利用钞票数据集讨论独立性.
  - 如果 X_4 与 X_6 相互独立，则 $\hat{f}(x_4, x_6) = \hat{f}_{X_4} \cdot \hat{f}_{X_6}$. 我们先作 $\hat{f}_{X_4} \cdot \hat{f}_{X_6}$ 的图形.

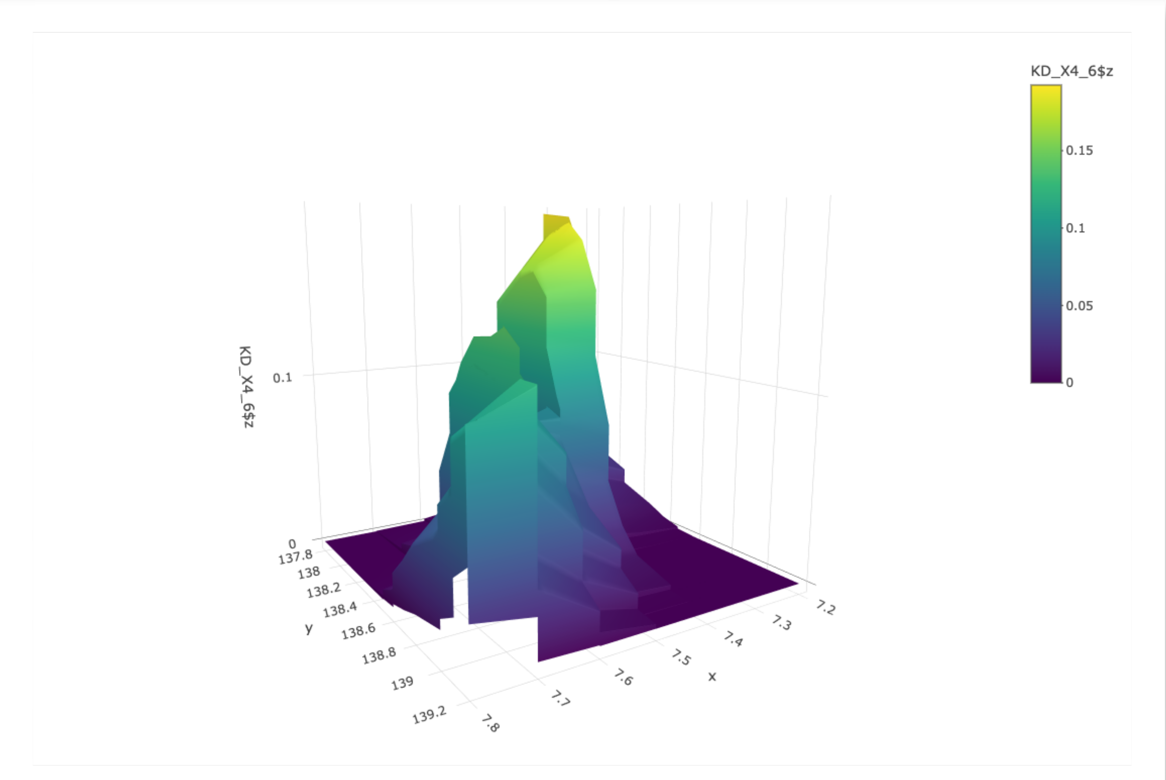
```
# f̂_X4 · f̂_X6 的图形
z = outer(dX4$y, dX6$y, function(a, b) a * b)
fig3 = plot_ly(x = ~dX4$x, y = ~dX6$x, z = ~z, type = "surface")
fig3
```



分布函数与密度函数

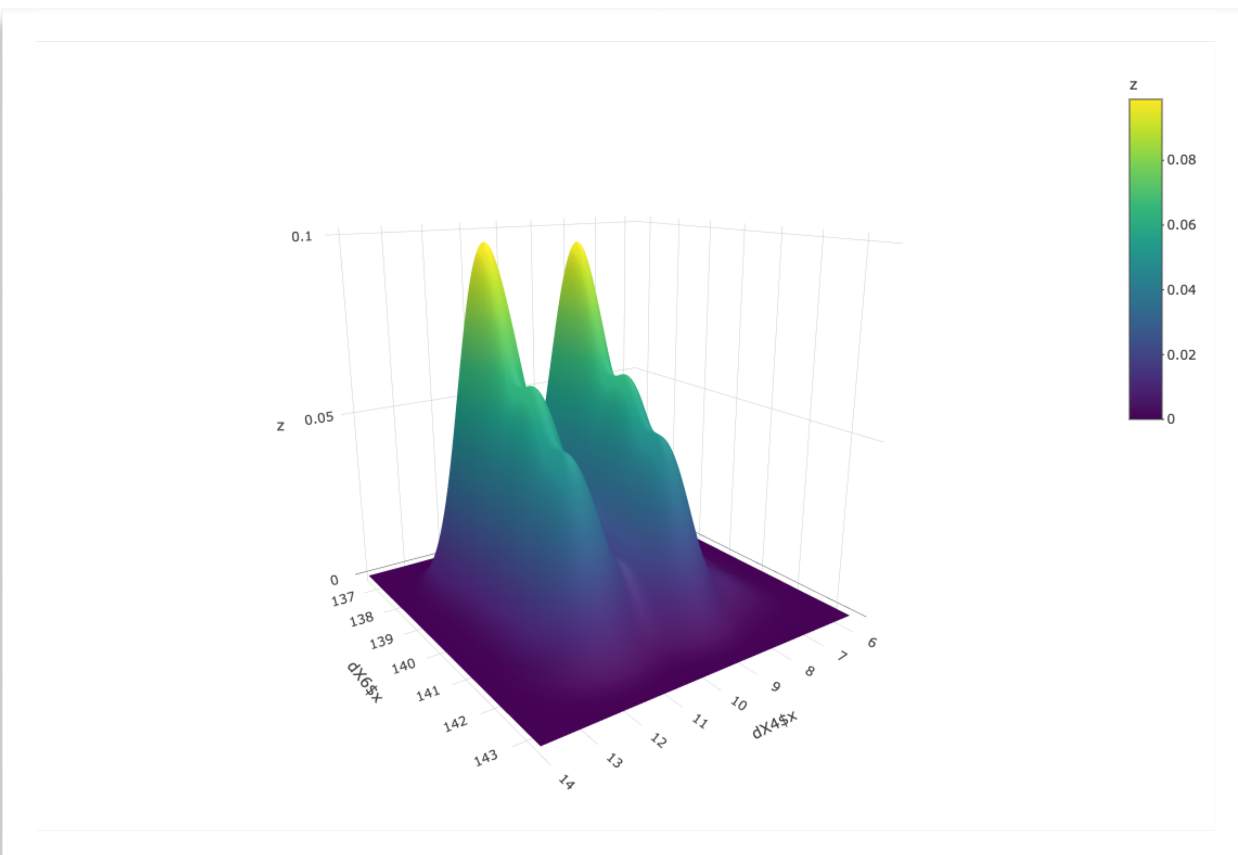
- 例：利用钞票数据集讨论独立性.
  - 再作 (X_4, X_6) 的联合密度函数的核密度图.

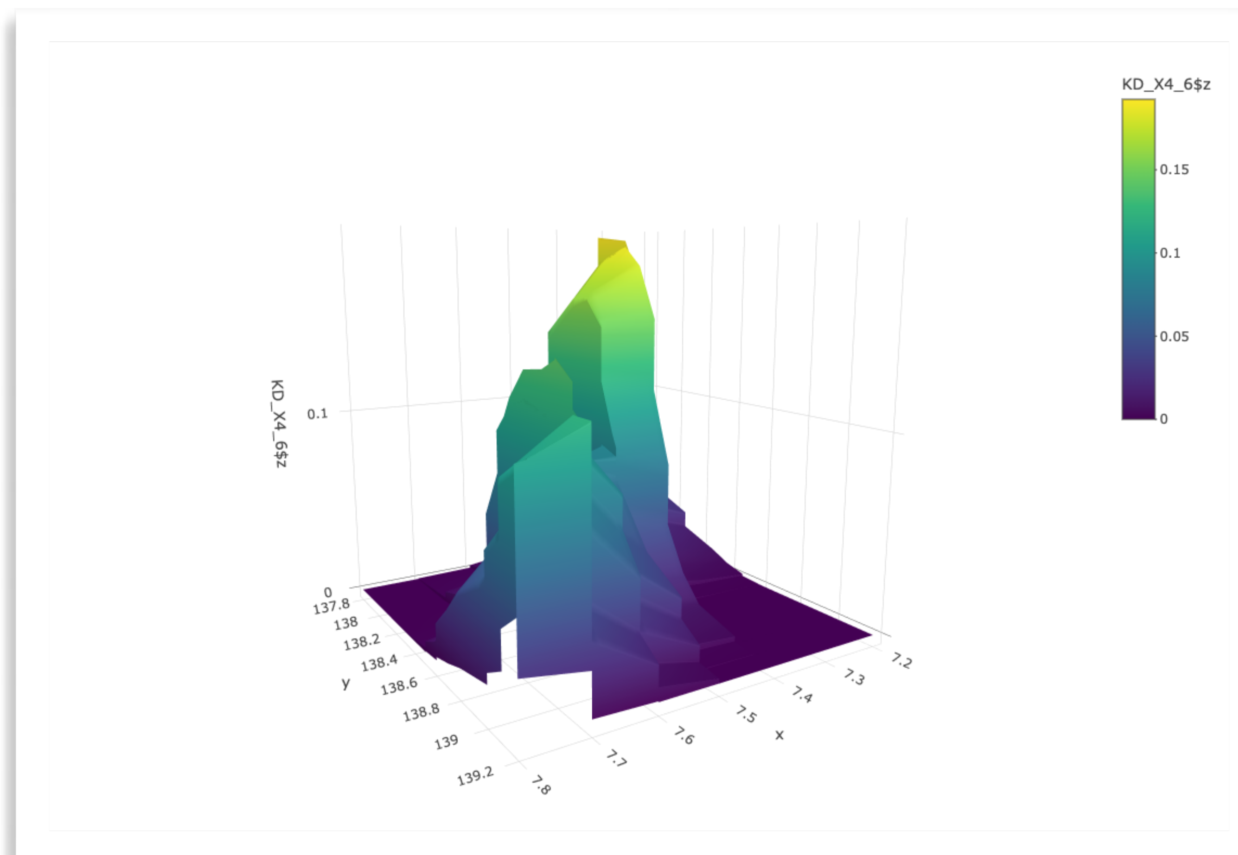
```
# X4与X6的联合概率密度函数的核密度图图
library(MASS)
x = sort(banknote$Bottom)
y = sort(banknote$Diagonal)
KD_X4_6 = kde2d(x, y, n = 25)
fig4 = plot_ly(x = ~x, y = ~y, z = ~KD_X4_6$z, type = "surface")
fig4
```



分布函数与密度函数

- 例：利用钞票数据集讨论独立性.
  - 如果 X_4 与 X_6 相互独立，则 $\hat{f}_{X_4} \cdot \hat{f}_{X_6}$ 与 (X_4, X_6) 的图形应该差异不大.





$\Longrightarrow$ X_4 与 X_6 不独立

矩与特征函数 (Moments and Characteristic Functions)

- 矩 (moment) - 期望 (expectation) 与协方差矩阵 (covariance matrix)

  - 期望：设随机向量 $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$ 的概率密度函数为 $f(\boldsymbol{x})$,

$$\mathrm{E}(\boldsymbol{X}) = \begin{pmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{pmatrix} = \int \cdots \int \boldsymbol{x} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \begin{pmatrix} \int \cdots \int x_1 f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\ \int \cdots \int x_2 f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\ \vdots \\ \int \cdots \int x_p f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \end{pmatrix} \triangleq \boldsymbol{\mu}$$

$$\Longrightarrow \quad \mathrm{E}(\alpha \boldsymbol{X} + \beta \boldsymbol{Y}) = \alpha \mathrm{E}(\boldsymbol{X}) + \beta \mathrm{E}(\boldsymbol{Y})$$

$$\mathrm{E}(\mathscr{A}\boldsymbol{X}) = \mathscr{A}\mathrm{E}(\boldsymbol{X}) \quad \longleftarrow \quad \mathscr{A}_{q\times p} \text{ 为实矩阵}$$

$$X \text{ 与 } Y \text{ 相互独立} \Longrightarrow \mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}) = \mathrm{E}(\boldsymbol{X})\,\mathrm{E}(\boldsymbol{Y}^{\mathrm{T}})$$

矩与特征函数 (Moments and Characteristic Functions)

- 矩 (moment) - 期望 (expectation) 与协方差矩阵 (covariance matrix)

  - 协方差矩阵： $\mathrm{Var}(\boldsymbol{X}) = \Sigma = \mathrm{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\right]$ ，$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$
  - 若随机向量 $\boldsymbol{X}$ 的均值向量为 $\boldsymbol{\mu}$，协方差矩阵为 Σ，则可以将其表示为 $\boldsymbol{X} \sim (\boldsymbol{\mu}, \Sigma)$.
  - 两个随机向量 $\boldsymbol{X} \sim (\boldsymbol{\mu}, \Sigma_{XX})$ 与 $\boldsymbol{Y} \sim (\boldsymbol{\nu}, \Sigma_{YY})$ 的协方差矩阵：

$$\Sigma_{XY} = \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\nu})^{\mathrm{T}}\right]$$

    - $\Sigma_{XY} = \Sigma_{YX}^{\mathrm{T}}$
    - $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立 $\Longrightarrow$ $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = 0$.
    - $\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \Longrightarrow \Sigma_{ZZ} = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}$

矩与特征函数 (Moments and Characteristic Functions)

- 矩 (moment) - 期望 (expectation) 与协方差矩阵 (covariance matrix)
  - $\boldsymbol{X}$ 的一阶矩：$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$.
  - $\boldsymbol{X}$ 的二阶矩：

$$\mathrm{E}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\right) = \left\{\mathrm{E}\left(X_i X_j\right)\right\}_{p\times p}, \quad i, j = 1, 2, \ldots, p$$

$$= \begin{pmatrix} \mathrm{E}\left(X_1^2\right) & \mathrm{E}\left(X_1 X_2\right) & \cdots & \mathrm{E}\left(X_1 X_p\right) \\ \mathrm{E}\left(X_2 X_1\right) & \mathrm{E}\left(X_2^2\right) & \cdots & \mathrm{E}\left(X_2 X_p\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}\left(X_p X_1\right) & \mathrm{E}\left(X_p X_2\right) & \cdots & \mathrm{E}\left(X_p^2\right) \end{pmatrix}$$

矩与特征函数 (Moments and Characteristic Functions)

- 协方差矩阵 $\Sigma = \mathrm{Var}(\boldsymbol{X})$ 的性质

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

$$\Sigma = \left(\sigma_{X_iX_j}\right)_{p\times p}, \quad \sigma_{X_iX_j} = \mathrm{Cov}\left(X_i, X_j\right), \quad \sigma_{X_iX_i} = \mathrm{Var}\left(X_i\right)$$

$$\Sigma = \mathrm{E}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\right) - \boldsymbol{\mu}\,\boldsymbol{\mu}^{\mathrm{T}}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mathrm{E}\left(X_1\right) \\ \mathrm{E}\left(X_2\right) \\ \vdots \\ \mathrm{E}\left(X_p\right) \end{pmatrix} = \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

$$\Sigma \geq 0$$

- 方差与协方差的性质

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

$$\mathrm{Var}\left(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\right) = \boldsymbol{a}^{\mathrm{T}}\,\mathrm{Var}\,(\boldsymbol{X})\;\boldsymbol{a} = \sum_{i,\,j} a_i\,a_j\,\sigma_{X_iX_j}$$

$$\mathrm{Var}\left(\mathscr{A}\,\boldsymbol{X} + \boldsymbol{b}\right) = \mathscr{A}\;\mathrm{Var}\,(\boldsymbol{X})\;\mathscr{A}^{\mathrm{T}}$$

$$\mathrm{Cov}\,(\boldsymbol{X} + \boldsymbol{Y},\, \boldsymbol{Z}) = \mathrm{Cov}\,(\boldsymbol{X},\, \boldsymbol{Z}) + \mathrm{Cov}\,(\boldsymbol{Y},\, \boldsymbol{Z})$$

$$\mathrm{Var}\,(\boldsymbol{X} + \boldsymbol{Y}) = \mathrm{Var}\,(\boldsymbol{X}) + \mathrm{Cov}\,(\boldsymbol{X},\, \boldsymbol{Y}) + \mathrm{Cov}\,(\boldsymbol{Y},\, \boldsymbol{X}) + \mathrm{Var}\,(\boldsymbol{Y})$$

$$\mathrm{Cov}\left(\mathscr{A}\,\boldsymbol{X},\, \mathscr{B}\,\boldsymbol{Y}\right) = \mathscr{A}\;\mathrm{Cov}\,(\boldsymbol{X},\, \boldsymbol{Y})\;\mathscr{B}^{\mathrm{T}}$$

矩与特征函数 (Moments and Characteristic Functions)

- 例：设有概率密度函数 $f(x_1, x_2) = \begin{cases} \frac{1}{2}x_1 + \frac{3}{2}x_2, & 0 \le x_1, x_2 \le 1 \\ 0, & \text{otherwise} \end{cases}$

  - 则其均值向量 $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \frac{13}{24} \\ \frac{5}{8} \end{pmatrix}$

$$\mu_1 = \iint x_1 f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_0^1 \int_0^1 x_1 \left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right) \mathrm{d}x_1\,\mathrm{d}x_2 = \frac{13}{24}$$

$$\mu_2 = \iint x_2 f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_0^1 \int_0^1 x_2 \left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right) \mathrm{d}x_1\,\mathrm{d}x_2 = \frac{5}{8}$$

矩与特征函数 (Moments and Characteristic Functions)

- 例：设有概率密度函数 $f(x_1, x_2) = \begin{cases} \frac{1}{2}x_1 + \frac{3}{2}x_2, & 0 \le x_1, x_2 \le 1 \\ 0, & \text{其它} \end{cases}$
  - 则其均值向量 $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \frac{13}{24} \\ \frac{5}{8} \end{pmatrix}$
  - 其协方差矩阵 $\Sigma = \begin{pmatrix} \sigma_{X_1X_1} & \sigma_{X_1X_2} \\ \sigma_{X_2X_1} & \sigma_{X_2X_2} \end{pmatrix} = \begin{pmatrix} 0.0816 & -0.0052 \\ -0.0052 & 0.0677 \end{pmatrix}$

$$\sigma_{X_1X_1} = \mathrm{E}\left(X_1^2\right) - \mu_1^2 = \int_0^1\int_0^1 x_1^2\left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right)\mathrm{d}x_1\,\mathrm{d}x_2 - \left(\frac{13}{24}\right)^2 = \frac{3}{8} - \left(\frac{13}{24}\right)^2 \approx 0.0816$$

$$\sigma_{X_2X_2} = \mathrm{E}\left(X_2^2\right) - \mu_2^2 = \int_0^1\int_0^1 x_2^2\left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right)\mathrm{d}x_1\,\mathrm{d}x_2 - \left(\frac{5}{8}\right)^2 = \frac{11}{24} - \left(\frac{5}{8}\right)^2 \approx 0.0677$$

$$\sigma_{X_1X_2} = \mathrm{E}\left(X_1X_2\right) - \mu_1\mu_2 = \int_0^1\int_0^1 x_1x_2\left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right)\mathrm{d}x_1\,\mathrm{d}x_2 - \frac{13}{24}\times\frac{5}{8} = \frac{1}{3} - \frac{13}{24}\times\frac{5}{8} \approx -0.0052$$

矩与特征函数 (Moments and Characteristic Functions)

- 条件期望 (conditional expectation)

给定 $\boldsymbol{X}_1 = \boldsymbol{x}_1$ 时，$\boldsymbol{X}_2$ 的条件位置参数

给定 $\boldsymbol{X}_2 = \boldsymbol{x}_2$ 时，$\boldsymbol{X}_1$ 的条件位置参数

$$\mathrm{E}\left(\boldsymbol{X}_2 \middle| \boldsymbol{x}_1\right) \triangleq \int \boldsymbol{x}_2 f\left(\boldsymbol{x}_2 \middle| \boldsymbol{x}_1\right) \mathrm{d}\boldsymbol{x}_2, \quad \mathrm{E}\left(\boldsymbol{X}_1 \middle| \boldsymbol{x}_2\right) \triangleq \int \boldsymbol{x}_1 f\left(\boldsymbol{x}_1 \middle| \boldsymbol{x}_2\right) \mathrm{d}\boldsymbol{x}_1$$

$$\mathrm{Var}\left(\boldsymbol{X}_2 \middle| \boldsymbol{X}_1 = \boldsymbol{x}_1\right) = \mathrm{E}\left(\boldsymbol{X}_2 \boldsymbol{X}_2^{\mathrm{T}} \middle| \boldsymbol{X}_1 = \boldsymbol{x}_1\right) - \mathrm{E}\left(\boldsymbol{X}_2 \middle| \boldsymbol{X}_1 = \boldsymbol{x}_1\right) \mathrm{E}\left(\boldsymbol{X}_2^{\mathrm{T}} \middle| \boldsymbol{X}_1 = \boldsymbol{x}_1\right)$$

$$\mathrm{Var}\,(\boldsymbol{X}) = \mathrm{E}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\right) - \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}$$

$$\mathrm{Var}\left(\boldsymbol{X}_1 \middle| \boldsymbol{X}_2 = \boldsymbol{x}_2\right) = \mathrm{E}\left(\boldsymbol{X}_1 \boldsymbol{X}_1^{\mathrm{T}} \middle| \boldsymbol{X}_2 = \boldsymbol{x}_2\right) - \mathrm{E}\left(\boldsymbol{X}_1 \middle| \boldsymbol{X}_2 = \boldsymbol{x}_2\right) \mathrm{E}\left(\boldsymbol{X}_1^{\mathrm{T}} \middle| \boldsymbol{X}_2 = \boldsymbol{x}_2\right)$$

- 条件相关系数 (conditional correlations)

$$\rho_{X_2,\, X_3 | X_1 = x_1} = \frac{\mathrm{Cov}\left(X_2,\, X_3 \middle| X_1 = x_1\right)}{\sqrt{\mathrm{Var}\left(X_2 \middle| X_1 = x_1\right) \mathrm{Var}\left(X_3 \middle| X_1 = x_1\right)}}$$

给定 $X_1 = x_1$ 时，X_2 与 X_3 的偏相关系数 (partial correlation).

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑概率密度函数 $f(x_1, x_2, x_3) = \begin{cases} \frac{2}{3}(x_1 + x_2 + x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

关于 x_1，x_2，x_3 对称

$$\Longrightarrow \quad f(x_1, x_2) = \int f(x_1, x_2, x_3)\, \mathrm{d}x_3 = \begin{cases} \frac{2}{3}\left(x_1 + x_2 + \frac{1}{2}\right), & 0 < x_1, x_2 < 1 \\ 0, & \text{其它} \end{cases}$$

$$f(x_1, x_3) = \begin{cases} \frac{2}{3}\left(x_1 + x_3 + \frac{1}{2}\right), & 0 < x_1, x_3 < 1 \\ 0, & \text{其它} \end{cases}, \quad f(x_2, x_3) = \begin{cases} \frac{2}{3}\left(x_2 + x_3 + \frac{1}{2}\right), & 0 < x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$$

$$\Longrightarrow \quad f(x_1) = \int f(x_1, x_2)\, \mathrm{d}x_2 = \begin{cases} \frac{2}{3}(x_1 + 1), & 0 < x_1 < 1 \\ 0, & \text{其它} \end{cases}$$

$$f(x_2) = \begin{cases} \frac{2}{3}(x_2 + 1), & 0 < x_2 < 1 \\ 0, & \text{其它} \end{cases}, \quad f(x_3) = \begin{cases} \frac{2}{3}(x_3 + 1), & 0 < x_3 < 1 \\ 0, & \text{其它} \end{cases}$$

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑概率密度函数 $f(x_1, x_2, x_3) = \begin{cases} \frac{2}{3}(x_1+x_2+x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

关于 x_1，x_2，x_3 对称

$$\Rightarrow f(x_1, x_2 \mid x_3) = \frac{f(x_1, x_2, x_3)}{f(x_3)} = \begin{cases} \frac{x_1+x_2+x_3}{x_3+1}, & 0 < x_1, x_2 < 1, \\ 0, & \text{其它}, \end{cases} \quad (0 < x_3 < 1)$$

$$\Rightarrow f(x_1 \mid x_3) = \frac{f(x_1, x_3)}{f(x_3)} = \begin{cases} \frac{x_1+x_3+\frac{1}{2}}{x_3+1}, 0 < x_1 < 1 \\ 0, & \text{其它} \end{cases}, \quad (0 < x_3 < 1)$$

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑概率密度函数 $f(x_1, x_2, x_3) = \begin{cases} \frac{2}{3}(x_1 + x_2 + x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

  关于 x_1，x_2，x_3 对称

  - 计算可得以下各阶矩

$$E(X_i) = \frac{5}{9}, \quad E(X_i^2) = \frac{7}{18}; \quad E\left(X_i X_j\right) = \frac{11}{36}, (i \neq j)$$

$$\operatorname{Var}(X_i) = E(X_i^2) - \left[E(X_i)\right]^2 = \frac{7}{18} - \left(\frac{5}{9}\right)^2 = \frac{13}{162}$$

$$\operatorname{Cov}\left(X_i, X_j\right) = E\left(X_i X_j\right) - E(X_i) E\left(X_j\right) = \frac{11}{36} - \left(\frac{5}{9}\right)^2 = -\frac{1}{324}$$

$$\Longrightarrow \quad \Sigma = \begin{pmatrix} \frac{13}{162} & -\frac{1}{324} & -\frac{1}{324} \\ -\frac{1}{324} & \frac{13}{162} & -\frac{1}{324} \\ -\frac{1}{324} & -\frac{1}{324} & \frac{13}{162} \end{pmatrix} \quad \Longrightarrow \quad \rho_{X_1 X_2} = \frac{-\frac{1}{324}}{\sqrt{\frac{13}{162} \cdot \frac{13}{162}}} = -\frac{1}{26} \approx -0.0385$$

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑概率密度函数 $f(x_1, x_2, x_3) = \begin{cases} \frac{2}{3}(x_1 + x_2 + x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

  关于 x_1，x_2，x_3 对称

  - 计算可得以下各阶矩

$$E(X_1 | X_3 = x_3) = E(X_2 | X_3 = x_3) = \frac{1}{12}\left(\frac{6x_3 + 7}{x_3 + 1}\right)$$

$$E(X_1^2 | X_3 = x_3) = E(X_2^2 | X_3 = x_3) = \frac{1}{12}\left(\frac{4x_3 + 5}{x_3 + 1}\right)$$

$$E(X_1 X_2 | X_3 = x_3) = \frac{1}{12}\left(\frac{3x_3 + 4}{x_3 + 1}\right)$$

不是 x_3 的线性函数

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑概率密度函数 $f(x_1, x_2, x_3) = \begin{cases} \frac{2}{3}(x_1 + x_2 + x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

关于 x_1，x_2，x_3 对称

$$\operatorname{Var}\left(\boldsymbol{X}_1 \middle| \boldsymbol{X}_2 = \boldsymbol{x}_2\right) = \operatorname{E}\left(\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{T}} \middle| \boldsymbol{X}_2 = \boldsymbol{x}_2\right) - \operatorname{E}\left(\boldsymbol{X}_1 \middle| \boldsymbol{X}_2 = \boldsymbol{x}_2\right)\operatorname{E}\left(\boldsymbol{X}_1^{\mathrm{T}} \middle| \boldsymbol{X}_2 = \boldsymbol{x}_2\right)$$

$$\Longrightarrow \operatorname{Var}\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \middle| X_3 = x_3\right) = \begin{pmatrix} \dfrac{12x_3^2 + 24x_3 + 11}{144(x_3+1)^2} & \dfrac{-1}{144(x_3+1)^2} \\ \dfrac{-1}{144(x_3+1)^2} & \dfrac{12x_3^2 + 24x_3 + 11}{144(x_3+1)^2} \end{pmatrix}$$

$$\rho_{X_2, X_3 | X_1 = x_1} = \frac{\operatorname{Cov}\left(X_2, X_3 \middle| X_1 = x_1\right)}{\sqrt{\operatorname{Var}\left(X_2 \middle| X_1 = x_1\right)\operatorname{Var}\left(X_3 \middle| X_1 = x_1\right)}}$$

$$\Longrightarrow \rho_{X_1X_2 | X_3 = x_3} = \frac{\dfrac{-1}{144(x_3+1)^2}}{\sqrt{\dfrac{12x_3^2 + 24x_3 + 11}{144(x_3+1)^2} \cdot \dfrac{12x_3^2 + 24x_3 + 11}{144(x_3+1)^2}}}$$

$x_3 \in (0, 1)$

$$\rho_{X_1X_2 | X_3 = x_3} \in \left(-\frac{1}{11}, -\frac{1}{47}\right) \approx (-0.0909, -0.0213)$$

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑密度函数 $f(x_1, x_2, x_3) = \begin{cases} 2x_2(x_1 + x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

X_2 与 (X_1, X_3) 独立

关于 x_1 与 x_3 对称

$$\Longrightarrow \quad f(x_1, x_3) = \int f(x_1, x_2, x_3)\,\mathrm{d}x_2 = \begin{cases} x_1 + x_3, & 0 < x_1, x_3 < 1 \\ 0, & \text{其它} \end{cases}$$

$$f(x_2) = \begin{cases} 2x_2, & 0 < x_1 < 1 \\ 0, & \text{其它} \end{cases}$$

$$f(x_1) = \int f(x_1, x_3)\,\mathrm{d}x_3 = \begin{cases} x_1 + \frac{1}{2}, & 0 < x_1 < 1 \\ 0, \text{其它} \end{cases}, \quad f(x_3) = \int f(x_1, x_3)\,\mathrm{d}x_1 = \begin{cases} x_3 + \frac{1}{2}, & 0 < x_3 < 1 \\ 0, \text{其它} \end{cases}$$

$$\Longrightarrow \quad \mathrm{E}(\boldsymbol{X}) = \begin{pmatrix} \frac{7}{12} \\ \frac{2}{3} \\ \frac{7}{12} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \frac{1}{144} & 0 & -\frac{1}{144} \\ 0 & \frac{1}{18} & 0 \\ -\frac{1}{144} & 0 & \frac{11}{144} \end{pmatrix}$$

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑密度函数 $f(x_1, x_2, x_3) = \begin{cases} 2x_2(x_1 + x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

关于 x_1 与 x_3 对称

X_2 与 (X_1, X_3) 独立

  - 讨论给定 $X_3 = x_3$ 时 (X_1, X_2) 的条件分布.

$f(x_3) = \begin{cases} x_3 + \frac{1}{2}, & 0 < x_3 < 1 \\ 0, & \text{其它} \end{cases}$

$$f(x_1, x_2 | x_3) = \frac{f(x_1, x_2, x_3)}{f(x_3)} = \begin{cases} \dfrac{4(x_1 + x_3)x_2}{2x_3 + 1}, & 0 < x_1, x_2 < 1 \\ 0, & \text{其它} \end{cases}, \quad (0 < x_3 < 1)$$

$f(x_1, x_3) = \begin{cases} x_1 + x_3, & 0 < x_1, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

$$f(x_1 | x_3) = \frac{f(x_1, x_3)}{f(x_3)} = \begin{cases} \dfrac{2(x_1 + x_3)}{2x_3 + 1}, & 0 < x_1 < 1 \\ 0, & \text{其它} \end{cases}, \quad (0 < x_3 < 1)$$

$$f(x_2 | x_3) = f(x_2) = 2x_2, \quad 0 < x_2 < 1 \quad (0 < x_3 < 1)$$

$$\Longrightarrow \quad f(x_1, x_2 | x_3) = f(x_1 | x_3) \cdot f(x_2 | x_3)$$

给定 $X_3 = x_3$ 时，X_1 与 X_2 的条件分布相互独立

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑密度函数 $f(x_1, x_2, x_3) = \begin{cases} 2x_2(x_1+x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases}$

X_2 与 (X_1, X_3) 独立

关于 x_1 与 x_3 对称

  - 讨论给定 $X_3 = x_3$ 时 (X_1, X_2) 的条件分布.

$f(x_3) = \begin{cases} x_3 + \frac{1}{2}, & 0 < x_3 < 1 \\ 0, \text{其它} \end{cases}$

$$f(x_1, x_2 \mid x_3) = \frac{f(x_1, x_2, x_3)}{f(x_3)} = \begin{cases} \frac{4(x_1+x_3)x_2}{2x_3+1}, & 0 < x_1, x_2 < 1 \\ 0, & \text{其它} \end{cases}, \quad (0 < x_3 < 1)$$

$$\Longrightarrow \quad \mathrm{E}\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \middle| X_3 = x_3\right] = \begin{pmatrix} \frac{1}{3}\left(\frac{2+3x_3}{1+2x_3}\right) \\ \frac{2}{3} \end{pmatrix}, \quad (0 < x_3 < 1)$$

$$\Longrightarrow \quad \mathrm{Var}\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \middle| X_3 = x_3\right] = \begin{pmatrix} \frac{1}{18}\left(\frac{6x_3^2+6x_3+1}{(2x_3+1)^2}\right) & 0 \\ 0 & \frac{1}{18} \end{pmatrix}, \quad (0 < x_3 < 1)$$

矩与特征函数 (Moments and Characteristic Functions)

- 条件期望的性质

$$\mathrm{E}\left(X_2 \middle| X_1 = x_1\right) = h\left(x_1\right), \quad \mathrm{Var}\left(X_2 \middle| X_1 = x_1\right) = g\left(x_1\right)$$

  - 利用条件期望定义随机变量

$$h\left(X_1\right) \triangleq \mathrm{E}\left(X_2 \middle| X_1\right), \qquad g\left(X_1\right) \triangleq \mathrm{Var}\left(X_2 \middle| X_1\right)$$

  - 有趣的性质

$$\mathrm{E}\left(X_2\right) = \mathrm{E}\left[\mathrm{E}\left(X_2 \middle| X_1\right)\right]$$

$$\mathrm{Var}\left(X_2\right) = \mathrm{E}\left[\mathrm{Var}\left(X_2 \middle| X_1\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(X_2 \middle| X_1\right)\right]$$

矩与特征函数 (Moments and Characteristic Functions)

- 例：考虑密度函数 $f(x_1, x_2) = \begin{cases} 2e^{-\frac{x_2}{x_1}}, & 0 < x_1 < 1,\ x_2 > 0 \\ 0, & \text{其它} \end{cases}$

$$\Longrightarrow\quad f(x_1) = \int f(x_1, x_2)\,\mathrm{d}x_2 = \begin{cases} 2x_1, & 0 < x_1 < 1 \\ 0, & \text{其它} \end{cases} \quad\Longrightarrow\quad \mathrm{E}(X_1) = \frac{2}{3},\quad \mathrm{Var}(X_1) = \frac{1}{18}$$

$$f(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f(x_1)} = \begin{cases} \frac{1}{x_1}e^{-\frac{x_2}{x_1}}, & x_2 > 0 \\ 0, & \text{其它} \end{cases},\quad (0 < x_1 < 1)$$

$$\Longrightarrow\quad \mathrm{E}(X_2 \mid X_1) = X_1,\quad \mathrm{Var}(X_2 \mid X_1) = X_1^2$$

- 无须计算 $f(x_2)$ 的表达式，我们就可以得到

$$\mathrm{E}(X_2) = \mathrm{E}\left[\mathrm{E}(X_2 \mid X_1)\right] = \mathrm{E}(X_1) = \frac{2}{3}$$

$$\mathrm{Var}(X_2) = \mathrm{E}\left[\mathrm{Var}(X_2 \mid X_1)\right] + \mathrm{Var}\left[\mathrm{E}(X_2 \mid X_1)\right] = \mathrm{E}(X_1^2) + \mathrm{Var}(X_1) = \frac{2}{4} + \frac{1}{18} = \frac{10}{18}$$

矩与特征函数 (Moments and Characteristic Functions)

- 条件期望 $\mathrm{E}\left(X_2 \mid X_1\right)$ 看作 X_1 的一个函数 $h\left(X_1\right)$，可以解释为通过 X_1 的一个函数对 X_2 的一种条件近似.

  称之为 X_2 关于 X_1 的回归函数

  - 该近似的误差项为 $U = X_2 - \mathrm{E}\left(X_2 \mid X_1\right)$

定理 4.3 设 $\boldsymbol{X}_1 \in \mathbb{R}^k$，$\boldsymbol{X}_2 \in \mathbb{R}^{p-k}$，$\boldsymbol{U} = \boldsymbol{X}_2 - \mathrm{E}\left(\boldsymbol{X}_2 \mid \boldsymbol{X}_1\right)$. 则

1. $\mathrm{E}(\boldsymbol{U}) = \mathbf{0}$.
2. 用 $\boldsymbol{X}_1$ 的一个函数 $h\left(\boldsymbol{X}_1\right)$ 对 $\boldsymbol{X}_2$ 作近似，则 $\mathrm{E}\left(\boldsymbol{X}_2 \mid \boldsymbol{X}_1\right)$ 是一个“最佳”逼近，其中 $h: \mathbb{R}^k \to \mathbb{R}^{p-k}$. 所谓“最佳”是指均方误差 (MSE) 最小的意义下，其中

$$\mathrm{MSE}(h) = \mathrm{E}\left\{\left[\boldsymbol{X}_2 - h\left(\boldsymbol{X}_1\right)\right]^{\mathrm{T}}\left[\boldsymbol{X}_2 - h\left(\boldsymbol{X}_1\right)\right]\right\}.$$