Numerical Characteristics of Multi-dimensional Random Vectors

November 20, 2023

回顾: 多维正态分布的标准化

Focus: 设 n 维随机向量 $X=(X_1,X_2,\ldots,X_n)\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 以及正定矩阵 $\boldsymbol{\Sigma}$ 的特征值分解 $\boldsymbol{\Sigma}=\boldsymbol{U}^\top\boldsymbol{\Lambda}\boldsymbol{U}$, 则随机向量

$$Y=\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}(X-\boldsymbol{\mu})\sim\mathcal{N}(\mathbf{0}_n,\boldsymbol{I}_n)\ .$$

我们需要明确两件事情:

- $\mathcal{N}(\mathbf{0}_n,\boldsymbol{I}_n)$ 的分量之间是独立的. 进一步, 当 $\boldsymbol{\Sigma}$ 为对角阵的时候, 多元正态分布的随机变量之间是独立的.
- $\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}(X-\boldsymbol{\mu})$ 所带来的线性变换可以将 X 标准化.

线性运算的基本性质

- 回忆: 矩阵的特征值和特征向量. 以二维为例, 矩阵 $\mathbf{A}$ 拥有特征值 λ_1 和 λ_2, 分别对应特征向量 $\boldsymbol{u}_1, \boldsymbol{u}_2$. 则有

$$\mathbf{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i\ , \quad i = 1, 2$$

进而有 $\mathbf{A} = \mathbf{U}^\top \boldsymbol{\Lambda} \mathbf{U}$ with $\mathbf{U}^\top = \mathbf{U}^{-1}$, 即

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$$

- 几何性质. For the case of $\mathbf{A}\boldsymbol{v}$ given any $\boldsymbol{v}$,
  - 矩阵 $\mathbf{U}$ 负责对 $\boldsymbol{v}$ 进行旋转
  - 矩阵 $\boldsymbol{\Lambda}$ 负责对 $\boldsymbol{v}$ 进行放缩

线性运算的基本性质

举个栗子, $\mathbf{A} = [1, 2; 2, 1]$

- 求特征值. $|\lambda \mathbf{I} - \mathbf{A}| = 0$, 得 $\lambda_1 = -1$ 和 $\lambda_2 = 3$.
- 求特征向量. 当 $\lambda_1 = -1$ 时,

$$(\lambda_i \mathbf{I} - \mathbf{A})\boldsymbol{u}_1 = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix}$$

  所以, $\boldsymbol{u}_1 = (1/\sqrt{2}, -1/\sqrt{2})$. 同理, 有 $\boldsymbol{u}_2 = (1/\sqrt{2}, 1/\sqrt{2})$.
- 根据 $\mathbf{UA} = \mathbf{\Lambda U}$, 可理解为
  - 正交坐标轴 $\mathbf{U}$ 上的向量 = 正交坐标轴 $\mathbf{U}$ 的线性组合
  - 协交坐标轴 $\mathbf{A}$ 上的向量 = 协交坐标轴 $\mathbf{A}$ 的线性组合
- For the case of $\mathbf{A}\boldsymbol{v}$, $\mathbf{U}\boldsymbol{v}$ 实现了对 $\boldsymbol{v}$ 的旋转, 而 $\mathbf{\Lambda}$ 实现了对 $|\boldsymbol{v}|$ 的放缩.

为了授课方便, 这里只关心对称的方阵. 同理, 可以推至非方阵的情况.

线性运算的基本性质

如果我们面对一个运算 $\mathbf{A}\boldsymbol{x} + \boldsymbol{b}$,

- $\boldsymbol{b}$ 是平移
- $\mathbf{U}$ 是旋转, if $|\mathbf{U}| = 1$
- $\mathbf{\Lambda}$ 是放缩, as $|\mathbf{A}\boldsymbol{x}| = |\mathbf{\Lambda}\boldsymbol{x}|$

面对 $Y = \mathbf{\Lambda}^{-1/2}\boldsymbol{U}(X - \boldsymbol{\mu}) = \mathbf{\Sigma}^{-1/2}(X - \boldsymbol{\mu})$

- $-\boldsymbol{\mu}$ 或者 $-\mathbf{\Sigma}^{-1/2}\boldsymbol{\mu}$ 是平移, 使得 $X - \boldsymbol{u}$ 以原点为中心 (X 以 $\boldsymbol{u}$ 为中心)
- $\mathbf{U}$ 是旋转
- $\mathbf{\Lambda}^{-1/2}$ 是放缩, 尤其是在 $\mathbf{\Lambda}^{-1/2}\bar{X}$ 的情形下

线性变换可以将 X 标准化

我们有

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{y}^\top \boldsymbol{y} = \boldsymbol{y}^\top \mathbf{I}_n \boldsymbol{y}$$

- 一方面,

$$\bar{\boldsymbol{x}}^\top \boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}^\top \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^\top \bar{\boldsymbol{x}} = \left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)^\top \boldsymbol{\Lambda}^{-1}\left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)$$

- 另一方面,

$$|\boldsymbol{\Sigma}^{-1/2}\bar{X}| = |\boldsymbol{\Lambda}^{-1/2}\bar{X}| = |Y|$$

所以,

$$\begin{aligned}
|\bar{\boldsymbol{x}}^\top \boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}| &= |\left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)^\top \boldsymbol{\Lambda}^{-1}\left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)| \\
&= |\left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)^\top|\,|\boldsymbol{\Lambda}^{-1}|\,|\left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)| \\
&= |\boldsymbol{x}^\top|\,|\boldsymbol{\Lambda}^{-1}|\,|\boldsymbol{x}| \\
&= |\boldsymbol{y}^\top|\,|\boldsymbol{y}|
\end{aligned}$$

线性变换可以将 X 标准化

根据

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\right)$$

和

$$Y = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}(X-\boldsymbol{\mu})$$

有

$$f_Y(\boldsymbol{y}) = (2\pi)^{-n/2}\exp\left(-\boldsymbol{y}^\top\boldsymbol{y}/2\right)$$

对角的协方差矩阵对应独立性

<u>求证:</u> 当 $\boldsymbol{\Sigma}$ 为对角阵的时候, 多元正态分布的随机变量之间是独立的.

<u>回顾:</u> 多维正态分布的密度函数

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\right)$$

如果 $\boldsymbol{\Sigma}$ 为对角阵, 假设为 $\boldsymbol{\Sigma}^{-1} = \mathrm{Diag}\{\lambda_1^{-1},\ldots,\lambda_n^{-1}\}$. 则指数项可以拆解为

$$\frac{-1}{2}\left[\lambda_1^{-1}(\bar{x}_1)^2+\cdots+\lambda_n^{-1}(\bar{x}_n)^2\right]$$

进而有

$$\begin{aligned} f(\boldsymbol{x}) &= (2\pi)^{-n/2}\left(\prod_{i=1}^n \lambda_i^{-1/2}\right)\exp\left(\frac{-1}{2}\left[\lambda_1^{-1}(\bar{x}_1)^2+\cdots+\lambda_n^{-1}(\bar{x}_n)^2\right]\right) \\ &= \prod_{i=1}^n\left(\frac{1}{\sqrt{2\pi}\sqrt{\lambda_i}}\right)\exp\left[-\frac{(\bar{x}_1)^2}{2\lambda_1}-\cdots-\frac{(\bar{x}_n)^2}{2\lambda_n}\right] \end{aligned}$$

对角的协方差矩阵对应着独立性

重新整理

$$
\begin{aligned}
f(\boldsymbol{x}) &= (2\pi)^{-n/2}\left(\prod_{i=1}^{n}\lambda_i^{-1/2}\right)\exp\left(\frac{-1}{2}\left[\lambda_1^{-1}(\bar{x}_1)^2+\cdots+\lambda_n^{-1}(\bar{x}_n)^2\right]\right)\\
&= \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sqrt{\lambda_i}}\right)\exp\left[-\frac{(\bar{x}_1)^2}{2\lambda_1}-\cdots-\frac{(\bar{x}_n)^2}{2\lambda_n}\right]\\
&= \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sqrt{\lambda_i}}\exp\left(-\frac{(\bar{x}_i)^2}{2\lambda_i}\right)\right)
\end{aligned}
$$

因此, 有

$$
f(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)
$$

基于上述运算, 我们可以推导多维正态分布其他性质 – 证明已留作思考题

设随机向量 $X = (X_1, X_2, \ldots, X_n)^\top$ 和 $Y = (Y_1, Y_2, \ldots, Y_m)^\top$, 以及

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu_x} \\ \boldsymbol{\mu_y} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma_{xx}} & \boldsymbol{\Sigma_{xy}} \\ \boldsymbol{\Sigma_{yx}} & \boldsymbol{\Sigma_{yy}} \end{pmatrix}\right),$$

则有

- 随机向量 X 和 Y 的边缘分布为 $X \sim \mathcal{N}(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_{xx}})$ 和 $Y \sim \mathcal{N}(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_{yy}})$
- 随机向量 X 和 Y 相互独立的充要条件是 $\boldsymbol{\Sigma_{xy}} = \mathbf{0}_{m\times n}$ (元素全为零的 $m \times n$ 矩阵)
- 在 $X = \boldsymbol{x}$ 的条件下, 随机向量 Y 的分布

$$Y \mid X = \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_y} + \boldsymbol{\Sigma_{yx}}\boldsymbol{\Sigma_{xx}^{-1}}(\boldsymbol{x} - \boldsymbol{\mu_x}), \boldsymbol{\Sigma_{yy}} - \boldsymbol{\Sigma_{yx}}\boldsymbol{\Sigma_{xx}^{-1}}\boldsymbol{\Sigma_{xy}})$$

- 在 $Y = \boldsymbol{y}$ 的条件下, 随机向量 X 的分布

$$X \mid Y = \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu_x} + \boldsymbol{\Sigma_{xy}}\boldsymbol{\Sigma_{yy}^{-1}}(\boldsymbol{y} - \boldsymbol{\mu_y}), \boldsymbol{\Sigma_{xx}} - \boldsymbol{\Sigma_{xy}}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\Sigma_{yx}})$$

数字特征的引言

类似于一维随机变量的数字特征, 多维随机变量也有数字特征

- 一方面是各分量自己的数字特征, 比如: 期望、方差、标准差等
- 另一方面是分量之间的关联程度, 反映随机变量间相依关系的数字特征, 即协方差与相关系数.

多维随机向量函数的期望

定义 0.54 [Informally, 期望 = < 取值, 概率 >]

离散随机变量 (X, Y) 的分布列为 $p_{ij} = P(X = x_i, Y = y_j)$, 则随机变量 $Z = g(X, Y)$ 的期望为

$$\mathbb{E}[Z] = \mathbb{E}[g(X, Y)] = \sum_{i,j} g(x_i, y_j) p_{ij}.$$

连续随机变量 (X, Y) 的概率密度为 $f(x, y)$, 则随机变量 $Z = g(X, Y)$ 的期望为

$$\mathbb{E}[Z] = \mathbb{E}[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy.$$

期望的性质

- 若随机向量 $X \geq Y$, 则 $\mathbb{E}[X] \geq \mathbb{E}[Y]$;
- 线性性. 对任意随机向量 X 和 Y, 有 $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$;
- 对任意随机向量 X 和 Y, 有 Cauchy-Schwartz 不等式:

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

- 独立可乘. 若随机变量 X 与 Y 相互独立, 有 $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$;
  - 独立方差. 若随机变量 X 与 Y 相互独立, 有 $\mathbb{VAR}(X+Y) = \mathbb{VAR}(X) + \mathbb{VAR}(Y)$.

多维随机向量函数的期望：例 0.96

例 0.96 设随机变量 X 与 Y 相互独立且都服从标准正态分布 $\mathcal{N}(0,1)$, 求 $\mathbb{E}[\max(X,Y)]$.

解答：例 0.96

题目: 设随机变量 X 与 Y 相互独立且都服从标准正态分布 $\mathcal{N}(0,1)$, 求 $\mathbb{E}[\max(X,Y)]$.

解答:

- 由题易知 X 与 Y 的联合密度函数为

$$f(x,y)=f_X(x)f_Y(y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}.$$

- 将二维平面区域分为两部分 $D_1=\{(x,y):x\geq y\}$, $D_2=\{(x,y):x<y\}$, 于是得到

$$\begin{aligned}
\mathbb{E}[\max(X,Y)]&=\int\int_{D_1}xf(x,y)dxdy+\int\int_{D_2}yf(x,y)dxdy\\
&=\int_{-\infty}^{+\infty}dy\int_{y}^{+\infty}xf(x,y)dx+\int_{-\infty}^{+\infty}dx\int_{x}^{+\infty}yf(x,y)dy\\
&=2\int_{-\infty}^{+\infty}dy\int_{y}^{+\infty}xf(x,y)dx=\frac{1}{\pi}\int_{-\infty}^{+\infty}dy\int_{y}^{+\infty}xe^{-\frac{x^2+y^2}{2}}dx\\
&=\frac{1}{\pi}\int_{-\infty}^{+\infty}e^{-y^2}dy=\frac{1}{\sqrt{\pi}}
\end{aligned}$$

多维随机向量函数的期望：例 0.97

例 0.97 某水果超市在每星期一进货一定数量的新鲜水果, 假设一周内出售水果的件数 $X \sim U(10, 20)$. 若这一周内出售一件水果获利 10 元, 若不能出售则因为水果过期而每件亏损 4 元, 求期望意义下水果超市的最优进货策略.

解答：例 0.97

题目: 如上所述.

解答:

- 不妨设水果超市每周进货 n 件 $(10 \leq n \leq 20)$, 则它的周利润为

$$Y = \begin{cases} 10n, & X \geq n \\ 10X - 4(n - X), & X < n \end{cases}$$

- 周利润 Y 是关于 X 的随机变量, 考虑在期望下的最优策略

$$\begin{aligned} \mathbb{E}[Y] &= \sum_{i=10}^{n-1}(10i - 4(n-i))P(X=i) + \sum_{i=n}^{20} 10nP(X=i) \\ &= \sum_{i=10}^{n-1}\frac{14i - 4n}{10} + \sum_{i=n}^{20} n = \frac{-7n^2 + 243n + 630}{10}. \end{aligned}$$

上式对 n 求导并令导数为零, 求解可得 $n = 17.36$. 则 n 可能取 17 或者 18, 经验证, 最后取 $n = 17$.

多维随机向量函数的条件期望

回顾: 对二维随机向量 (X, Y) 而言, 随机变量 X 的条件分布, 即给定随机变量 Y 取值的条件下 X 的概率分布. 而条件期望是条件分布的数学期望, 具体定义如下:

定义 0.55 设 (X, Y) 为离散型随机向量, 在 $Y = y$ 的条件下 X 的条件分布列为 $P(X = x_i|Y = y)$, 称

$$\mathbb{E}(X|y) = \mathbb{E}(X|Y = y) = \sum_i x_i P(X = x_i|Y = y)$$

为在 $Y = y$ 的条件下 X 的条件期望. 设 (X, Y) 为连续型随机向量, 在 $Y = y$ 的条件下 X 的条件密度函数为 $f_{X|Y}(x|y)$, 称

$$\mathbb{E}(X|y) = \mathbb{E}(X|Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx$$

为在 $Y = y$ 的条件下 X 的条件期望.

无条件期望 $\mathbb{E}(X)$ 是一个常数, 而条件期望 $\mathbb{E}(X|Y = y)$ 是 y 的函数; 例如用 X 表示中国成年人的身高, Y 表示中国成年人的足长, 我国公安部门研究得足长为 y 的中国成年人平均身高为 $\mathbb{E}(X|Y = y) = 6.876y$, 此公式常用于公安痕迹侦查中.

条件期望的性质

- 线性性. 对任意常数 a, b 有 $\mathbb{E}(aX_1 + bX_2|Y) = a\mathbb{E}(X_1|Y) + b\mathbb{E}(X_2|Y)$;

- 函数型. 对离散型随机向量 (X, Y) 和函数 $g(X)$, 有
$$\mathbb{E}(g(X)|Y) = \sum_i g(x_i)P(X = x_i|Y = y)$$
对连续型随机向量 (X, Y) 和函数 $g(X)$, 有
$$\mathbb{E}(g(X)|Y) = \int_{-\infty}^{+\infty} g(x)f(x|Y = y)dx$$

- 若随机向量 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则在 $Y = y$ 的条件下随机变量 X 服从正态分布 $\mathcal{N}(\mu_x - \rho\sigma_x(y - \mu_y)/\sigma_y, (1 - \rho^2)\sigma_x^2)$, 由此可得
$$\mathbb{E}(X|y) = \mu_x - \frac{\rho\sigma_x(y - \mu_y)}{\sigma_y}$$

多维随机向量函数的条件期望：例 0.98

例 0.98 设随机向量 (X, Y) 的联合概率密度为

$$f(x, y) = \begin{cases} \exp(-y), & 0 < x < y < +\infty \\ 0, & \text{其它} \end{cases}$$

求条件期望 $\mathbb{E}(X|y)$.

解答：例 0.98

题目: 如上所述.

解答:

- 根据条件期望的定义 $\mathbb{E}(X|y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx$, 计算 Y 的边缘密度函数, 当 $y > 0$ 时
$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \int_0^y \exp(-y)dx = y\exp(-y),$$
由此得到条件分布
$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{y} \qquad (0 < x < y < +\infty),$$
由此可得
$$\mathbb{E}(X|y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\,dx = \int_0^y \frac{x}{y}\,dx = \frac{y}{2}.$$

二重期望公式

定理 0.30 [二重期望公式] 设 (X, Y) 是二维随机向量, 且 $\mathbb{E}(X)$ 存在, 则

$$
\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|Y)) \\
&= \begin{cases} \sum_i \mathbb{E}(X|Y=y_i)P(Y=y_i), & \text{若 } Y \text{ 是一个离散型随机变量} \\ \int_{-\infty}^{+\infty} \mathbb{E}(X|Y=y_i) f_Y(y)\, dy, & \text{若 } Y \text{ 是一个连续型随机变量} \end{cases}
\end{aligned}
$$

二重期望公式在实际中很有用, 譬如, 在计算取值范围很大的 X 的期望 $\mathbb{E}(X)$ 时, 可以通过构建与 X 有关的量 Y, 通过 Y 的不同取值将大范围划分成若干小区域. 先在小区域上求 X 的平均, 再对此类平均求加权平均, 即可得到大范围上 X 的期望 $\mathbb{E}(X)$.

推论: 全期望公式

回顾: 对任意事件 A 而言, 根据全概率公式有 $P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$; 即通过对样本空间的切割划分, 将一个复杂事件化成相对简单的事件来求其概率; 借鉴这种化繁为简的思想, 也可以通过全期望公式计算复杂事件的期望, 具体如下:

定理 0.31 设 $A_1, A_2, \ldots, A_n$ 是样本空间 Ω 一个分割, $A_iA_j = \emptyset$ 和 $\Omega = \bigcup_{i=1}^{n} A_i$. 对任意随机变量 X 有

$$\mathbb{E}(X) = \mathbb{E}(X|A_1)P(A_1) + \mathbb{E}(X|A_2)P(A_2) + \cdots + \mathbb{E}(X|A_n)P(A_n),$$

特别的, 随机事件 A 及其对立事件 $\bar{A}$ 构成空间 Ω 一个分割, 对任意随机变量 X 有

$$\mathbb{E}(X) = \mathbb{E}(X|A)P(A) + \mathbb{E}(X|\bar{A})P(\bar{A}).$$

可证.

证明: 全期望公式

根据二重期望公式定理 0.30, 假设 Y 有 m 个取值

$$
\begin{aligned}
\mathbb{E}(X) = \mathbb{E}_Y\left[\mathbb{E}(X \mid Y)\right] &= \sum_{i=1}^{m} \mathbb{E}(X \mid Y = y_i) P(Y = y_i) \\
&= \sum_{i=1}^{m} \left[\sum_{j} x_j P(x_j \mid Y = y_i)\right] P(Y = y_i) \\
&= \sum_{i=1}^{m} \sum_{j} x_j P(x_j \mid Y = y_i) P(Y = y_i) \\
&= \sum_{k=1}^{n} \sum_{j} x_j P(x_j \mid A_k) P(A_k)
\end{aligned}
$$

最后一步成立于

$$\begin{aligned}\sum_{k=1}^{n} P(x_j \mid A_k)P(A_k) &= \sum_{k=1}^{n} \sum_{k_i : y_{k_i} \in A_k} P(x_j \mid y_{k_i})P(y_{k_i} \mid A_k)P(A_k) \\ &= \sum_{i=1}^{m} P(x_j \mid Y = y_i)P(Y = y_i)\end{aligned}$$