Conditional Expectation and Variance

8 条件概率

离散、连续随机变量的条件期望、条件方差

每一种科学，只要达到一定程度的成熟，就会自动成为数学的一部分。

Every kind of science, if it has only reached a certain degree of maturity, automatically becomes a part of mathematics.

—— 大卫·希尔伯特 (David Hilbert) | 德国数学家 | 1862 ~ 1943

- ◀ `matplotlib.pyplot.errorbar()` 绘制误差棒
- ◀ `matplotlib.pyplot.stem()` 绘制火柴梗图
- ◀ `numpy.mean()` 计算均值
- ◀ `numpy.sqrt()` 计算平方根
- ◀ `numpy.std()` 计算标准差，默认分母为 n，不是 n - 1
- ◀ `numpy.var()` 计算方差，默认分母为 n，不是 n - 1
- ◀ `seaborn.heatmap()` 绘制热图

8.1 离散随机变量：条件期望

条件**期望** (conditional expectation 或 conditional expected value)，或条件**均值** (conditional mean)，是一个随机变量的相对于一个条件概率分布的**期望**。换句话说，这是给定的一个或多个其他随机变量值的条件下，某个特定随机变量的**期望**。

类似地，条件**方差** (conditional variance) 与通常的**方差**的定义完全一致，只不过将**期望**换成了条件**期望**，并将概率换成了条件概率而已。

条件**期望**和条件**方差**这两个概念在数据科学、机器学习算法中格外重要，本章分别讲解离散随机变量和随机随机变量的条件**期望**和条件**方差**。本书第 12 章则专门介绍高斯条件概率。

大家应该已经看到，本章**期望**、**方差**交替出现，为了帮助大家阅读，我们用给**期望**、**方差**涂了不同颜色。

什么是条件期望?

条件**期望**其实很好理解。比如，一个笼子里 10 只动物，其中 6 只鸡 (60%)、4 只兔 (40%)。如图 1 所示，分别只考虑鸡，只考虑兔，这就是“条件”。

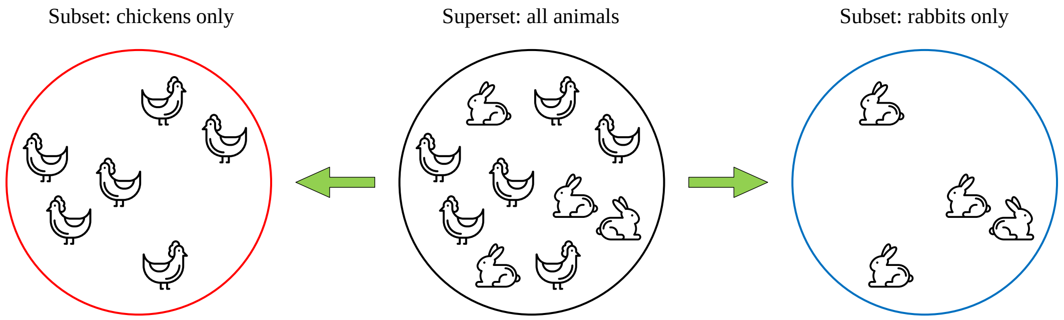

图 1. 解释条件

如图 2 所示，鸡的平均体重为 2 公斤，这个数值就是条件**期望**。再举个例子，兔子的平均体重为 4 公斤，这也是条件**期望**。

本书后续会用鸢尾花数据做例子给大家继续讲解条件**期望**。

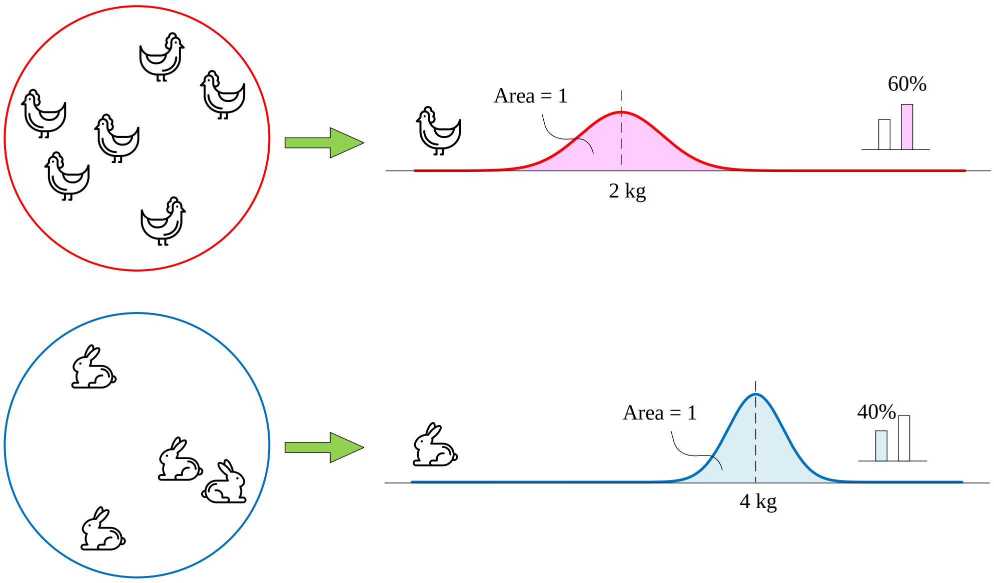

图 2. 解释条件**期望**

条件**期望** E(*Y*|*X* = *x*)

如果 *X* 和 *Y* 均为离散随机变量，给定 *X* = *x* 条件下，*Y* 的条件**期望** E(*Y*|*X* = *x*) (conditional mean of *Y* given *X* = *x*) 定义为：

$$\mathrm{E}\Big(Y \,\Big|\, \underbrace{X = x}_{\text{Given}}\Big) = \overbrace{\sum_{y} y \cdot \underbrace{p_{Y|X}\left(y|x\right)}_{\text{Conditional}}}^{\text{Expectation}} = \sum_{y} y \cdot \frac{\overbrace{p_{X,Y}\left(x,y\right)}^{\text{Joint}}}{\underbrace{p_X\left(x\right)}_{\text{Marginal}}} = \frac{1}{\underbrace{p_X\left(x\right)}_{\text{Marginal}}} \sum_{y} y \cdot \overbrace{p_{X,Y}\left(x,y\right)}^{\text{Joint}} \tag{1}$$

(1) 相当于求加权平均数。从几何角度来看，如图 3 所示，条件概率质量函数 $p_{Y|X}\left(y|x\right)$ 分别乘以对应 *y* 值 (黄色高亮)，然后求和，结果就是条件**期望** E(*Y*|*X* = *x*)。

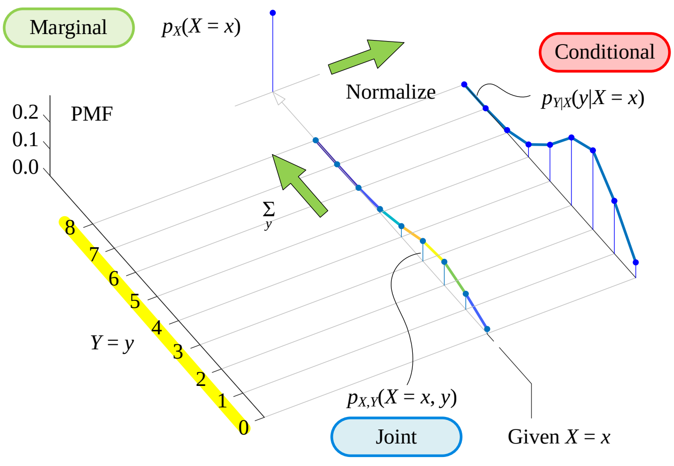

图 3. 条件概率 PMF $p_{Y|X}(y|x)$，X 和 Y 均为离散随机变量

解剖条件期望 $\mathrm{E}(Y|X = x)$

下面，我们进一步解剖 (1)。

给定 $X = x$ 条件下，也就是说离散随机变量 X 固定在 x，满足这个条件的样本构成了全新的“样本空间”。

$p_{Y|X}(y|x)$ 是给定 $X = x$ 条件下 Y 的概率质量函数，相当于 (1) 中加权平均数中的权重。

回忆本书第 4 章，利用贝叶斯定理，$p_X(x) > 0$，条件概率质量函数 $p_{Y|X}(y|x)$ 可以通过联合 PMF $p_{X,Y}(x,y)$ 和边际 PMF $p_X(x)$ 相除得到：

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{\underbrace{p_X(x)}_{\text{Normalize}}} \tag{2}$$

其中，分母中的边缘概率 $p_X(x)$ 起到归一化的效果。

(1) 中大西格玛求和 $\sum_y(\cdot)$ 代表“穷举”一切可能的 y 值，计算“$y \times$ 条件概率 $p_{Y|X}(y|x)$”之和，也就是“$y \times$ 权重”之和，即加权平均数。

比较期望 $\mathrm{E}(Y)$、条件期望 $\mathrm{E}(Y|X = x)$

对比离散随机变量 Y 的期望 $\mathrm{E}(Y)$、条件期望 $\mathrm{E}(Y|X = x)$：

$$\begin{aligned} \mathrm{E}(Y) &= \sum_y y \cdot \underbrace{p_Y(y)}_{\text{Weight}} \\ \mathrm{E}(Y \mid X = x) &= \sum_y y \cdot \underbrace{p_{Y|X}(y \mid x)}_{\text{Weight}} \end{aligned} \tag{3}$$

容易发现，我们不过是把求均值的权重从边缘 PMF $p_Y(y)$ 换成了条件 PMF $p_{Y|X}(y|x)$。$\sum_y y$ 都是遍历所有 y 的取值。

作为权重，$p_Y(y)$ 和 $p_{Y|X}(y|x)$ 的求和都为 1，即：

$$
\begin{aligned}
\sum_y \underbrace{p_Y(y)}_{\text{Marginal}} &= 1 \\
\sum_y \underbrace{p_{Y|X}(y|x)}_{\text{Conditional}} &= 1
\end{aligned} \tag{4}
$$

上两式实际上都是本书第 3 章介绍的全概率定理 (law of total probability) 的体现。

注意，**期望** E(Y) 是一个标量值。而 E($Y|X=x$) 在不同的 $X=x$ 条件下结果不同，即 E($Y|X$) 代表一组数。也就是说，E($Y|X$) 可以看做是个向量。

既然 E($Y|X$) 代表一组数，我们立刻就会想到 E($Y|X$) 肯定也有**期望**，即均值！

也就是说，笼子里的鸡的平均体重、兔子的平均体重，这两个均值还能再算一个均值，即笼子里所有动物的平均体重。

全期望定理

全**期望**定理 (law of total expectation)，又叫双重**期望**定理 (double expectation theorem)、重叠**期望**定理 (iterated total expectation)，指的是：

$$
\underbrace{\mathrm{E}(Y)}_{\text{Expectation}} = \overbrace{\mathrm{E}\left[\underbrace{\mathrm{E}(Y|X)}_{\text{Conditional expectation}}\right]}^{\text{Expectation of conditonal expectations}} = \sum_x \underbrace{\mathrm{E}(Y|X=x)}_{\text{Conditional expectation}} \cdot \underbrace{p_X(x)}_{\text{Marginal}} \tag{5}
$$

推导过程如下，不要求大家记忆：

$$
\begin{aligned}
\mathrm{E}\left[\underbrace{\mathrm{E}(Y|X)}_{\text{Conditional expectation}}\right] &= \sum_x \underbrace{\mathrm{E}(Y|X=x)}_{\text{Conditional expectation}} \cdot \underbrace{p_X(x)}_{\text{Marginal}} = \sum_x \underbrace{\left\{\sum_y y \cdot \underbrace{p_{Y|X}(y|x)}_{\text{Conditional}}\right\}}_{\text{Conditional expectation}} \cdot \underbrace{p_X(x)}_{\text{Marginal}} \\
&= \sum_x \sum_y y \cdot \overbrace{\underbrace{p_{Y|X}(y|x)}_{\text{Conditional}} \cdot \underbrace{p_X(x)}_{\text{Marginal}}}^{\text{Use Bayes' Rule}} = \sum_x \sum_y y \cdot \overbrace{p_{X,Y}(x,y)}^{\text{Joint}} \\
&= \sum_x \sum_y y \cdot \overbrace{\underbrace{p_{X|Y}(x|y)}_{\text{Conditional}} \cdot \underbrace{p_Y(y)}_{\text{Marginal}}}^{\text{Use Bayes' Rule}} = \sum_y y \cdot \underbrace{p_Y(y)}_{\text{Marginal}} \cdot \underbrace{\overbrace{\sum_x p_{X|Y}(x|y)}^{=1}}_{\text{Law of total probability}} \\
&= \sum_y y \cdot \underbrace{p_Y(y)}_{\text{Marginal}} = \mathrm{E}(Y)
\end{aligned} \tag{6}
$$

以上推导中，用到了双重求和调换变量顺序，这是因为 x 和 y 构成的网格“方方正正”；否则，不能轻易调换求和顺序。这个双重积分变量顺序类似。《数学要素》第 14 章探讨过这个问题，请大家回顾。

其实，全**期望**定理很好理解！

还是用本章前文的例子。前文提到，笼子里的鸡的平均体重 2 kg，兔子的平均体重为 4 kg。整个笼子里所有动物的平均体重就是 2 × 60% + 4 × 40% = 2.8 kg。2.8 kg 就是“条件**期望**的**期望**”。笼子里的鸡占比较高，因此整个笼子里动物的平均体重稍微“偏向”鸡体重的“条件**期望**”。

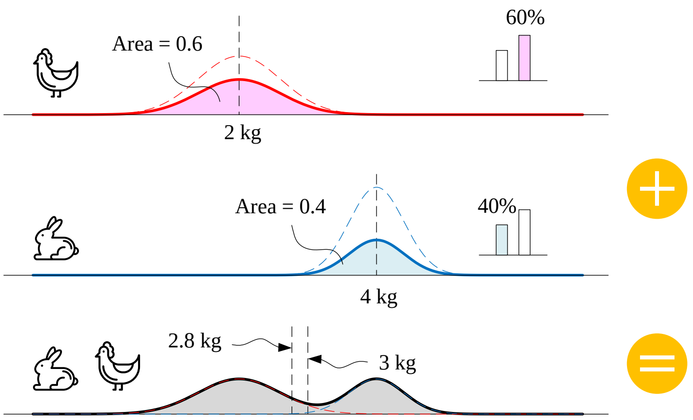

图 4. 解释全**期望**定理

大家如果要问，为什么求“条件**期望**的**期望**”要用加权平均？而不是用 (2 + 4) / 2 = 3 kg？

我们举个极端例子来解释。除了所有鸡之外，如果整个笼子里只有一只兔子，它的体重为 8 kg，也就是说“所有”兔子的平均体重也是 8 kg。假设所有鸡的平均体重还是 2 kg。大家自己思考，如果用 2 kg 和 8 kg 的平均值 5 kg 代表整个笼子里所有动物的平均体重，这是否合理？

条件**期望** $\mathrm{E}(X|Y=y)$

同理，如图 5 所示，给定 $Y = y$ 这个条件下，$p_Y(y) > 0$，X 的**期望** $\mathrm{E}(X|Y=y)$ 定义为：

$$
\begin{aligned}
\mathrm{E}\left(X \,\middle|\, \underbrace{Y=y}_{\text{Given}}\right) &= \overbrace{\sum_x x \cdot \underbrace{p_{X|Y}(x|y)}_{\text{Conditional}}}^{\text{Expectation}} \\
&= \sum_x x \cdot \frac{\overbrace{p_{X,Y}(x,y)}^{\text{Joint}}}{\underbrace{p_Y(y)}_{\text{Marginal}}} = \frac{1}{\underbrace{p_Y(y)}_{\text{Marginal}}} \sum_x x \cdot \overbrace{p_{X,Y}(x,y)}^{\text{Joint}}
\end{aligned}
\tag{7}
$$

请大家自行分析上式，并比较 $\mathrm{E}(X)$ 和 $\mathrm{E}(X|Y=y)$：

$$
\begin{gathered}
\mathrm{E}(X)=\sum_{x} x \cdot \underbrace{p_{X}(x)}_{\text{Weight}} \\
\mathrm{E}(X \mid Y=y)=\sum_{x} x \cdot \underbrace{p_{X \mid Y}(x \mid y)}_{\text{Weight}}
\end{gathered} \tag{8}
$$

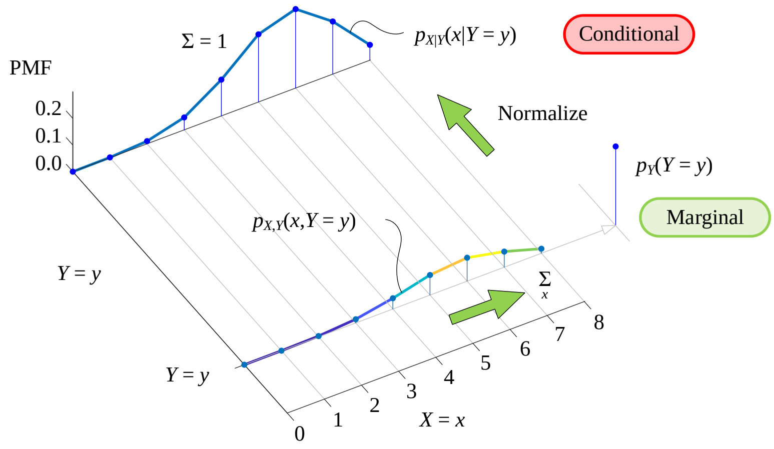

图 5. 条件概率 PMF $p_{X|Y}(x|y)$，X 和 Y 均为离散随机变量

对于条件**期望** $\mathrm{E}(X|Y)$，全**期望**定理为：

$$
\underbrace{\mathrm{E}(X)}_{\text{Expectation}}=\mathrm{E}\Bigg[\underbrace{\mathrm{E}(X \mid Y)}_{\text{Conditional expectation}}\Bigg] \tag{9}
$$

基于事件的条件**期望**

给定事件 C 发生的条件下 ($\Pr(C) > 0$)，随机变量 X 的条件**期望**为：

$$
\begin{aligned}
\mathrm{E}(X \mid C) & =\sum_{x} x \cdot \underbrace{p_{X \mid C}(x \mid C)}_{\text{Conditional}} \\
& =\sum_{x} x \cdot \frac{\overbrace{p_{X, C}(x, C)}^{\text{Joint}}}{\operatorname{Pr}(C)}
\end{aligned} \tag{10}
$$

这个式子类似前文的两个随机变量的条件**期望**，大家会在本章后续看到上式的用途。

独立

特别地，如果 X 和 Y 独立，则：

$$\begin{aligned} \mathrm{E}\left(Y \mid X=x\right) &= \mathrm{E}\left(Y\right) \\ \mathrm{E}\left(X \mid Y=y\right) &= \mathrm{E}\left(X\right) \end{aligned} \tag{11}$$

8.2 离散随机变量：条件方差

在上一节的基础上，本节介绍离散随机变量的条件方差。

条件方差 $\mathrm{var}(Y|X = x)$

给定 $X = x$ 条件下，Y 的条件方差 $\mathrm{var}(Y|X = x)$ (conditional variance of Y given $X = x$) 定义为：

$$\begin{aligned} \mathrm{var}\left(Y \mid X=x\right) &= \overbrace{\sum_{y} \Bigg(\underbrace{y - \overbrace{\mathrm{E}\left(Y \mid X=x\right)}^{\text{Expectation}}}_{\text{Deviation}} \Bigg)^2 \cdot \underbrace{p_{Y|X}\left(y \mid x\right)}_{\text{Conditional}}}^{\text{Expectation}} \\ &= \sum_{y} \left(y - \mathrm{E}\left(Y \mid X=x\right) \right)^2 \cdot \frac{\overbrace{p_{X,Y}\left(x,y\right)}^{\text{Joint}}}{\underbrace{p_X\left(x\right)}_{\text{Marginal}}} \\ &= \frac{1}{\underbrace{p_X\left(x\right)}_{\text{Marginal}}} \sum_{y} \Bigg(\underbrace{y - \mathrm{E}\left(Y \mid X=x\right)}_{\text{Deviation}} \Bigg)^2 \cdot \overbrace{p_{X,Y}\left(x,y\right)}^{\text{Joint}} \end{aligned} \tag{12}$$

下面解剖上式。

$\mathrm{E}\left(Y \mid X=x\right)$ 是 (1) 中求得的条件期望，也就是比较的基准。

$y - \mathrm{E}\left(Y \mid X=x\right)$ 代表偏差，即每个 y 和 $\mathrm{E}\left(Y \mid X=x\right)$ 之间的偏离。$y - \mathrm{E}\left(Y \mid X=x\right)$ 平方后，再以 $p_{Y|X}\left(y \mid x\right)$ 为权重，求平均值，结果就是条件方差。

对比离散随机变量 Y 的方差和条件方差：

$$\begin{aligned} \mathrm{var}\left(Y\right) &= \sum_{y} \Bigg(\underbrace{y - \mathrm{E}\left(Y\right)}_{\text{Deviation}} \Bigg)^2 \cdot \underbrace{p_Y\left(y\right)}_{\text{Weight}} \\ \mathrm{var}\left(Y\right) &= \sum_{y} \Bigg(\underbrace{y - \mathrm{E}\left(Y \mid X=x\right)}_{\text{Deviation}} \Bigg)^2 \cdot \underbrace{p_{Y|X}\left(y \mid x\right)}_{\text{Weight}} \end{aligned} \tag{13}$$

可以发现两处变差异，度量偏差的基准从 $\mathrm{E}\left(Y\right)$ 变成 $\mathrm{E}\left(Y \mid X=x\right)$。加权平均的权重从 $p_Y\left(y\right)$ 变成 $p_{Y|X}\left(y \mid x\right)$。

类似方差的简便计算技巧，条件方差 var($Y|X = x$) 也有如下计算技巧：

$$
\begin{aligned}
\operatorname{var}(Y) &= \mathrm{E}\left(Y^2\right) - \mathrm{E}(Y)^2 \\
\operatorname{var}(Y|X=x) &= \mathrm{E}\left(Y^2|X=x\right) - \mathrm{E}(Y|X=x)^2
\end{aligned} \tag{14}
$$

全方差定理

全方差定理 (law of total variance)，又叫重叠期望定理 (law of iterated variance)，指的是：

$$
\operatorname{var}(Y) = \underbrace{\mathrm{E}\left(\operatorname{var}(Y \mid X)\right)}_{\text{Expectation of conditional variance}} + \underbrace{\operatorname{var}\left(\mathrm{E}(Y \mid X)\right)}_{\text{Variance of conditional expectation}} \tag{15}
$$

$\mathrm{E}\left(\operatorname{var}(Y \mid X)\right)$ 是条件方差的期望 (加权平均数)：

$$
\underbrace{\mathrm{E}\left(\operatorname{var}(Y \mid X)\right)}_{\text{Expectation of conditional variance}} = \sum_x \operatorname{var}(Y \mid X = x) \cdot p_X(x) \tag{16}
$$

条件方差的期望 $\mathrm{E}\left(\operatorname{var}(Y \mid X)\right)$ 还不够解释整体的方差。缺少的成分是条件期望的方差 $\operatorname{var}\left(\mathrm{E}(Y \mid X)\right)$：

$$
\underbrace{\operatorname{var}\left(\mathrm{E}(Y \mid X)\right)}_{\text{Variance of conditional expectation}} = \sum_x \left(\mathrm{E}(Y \mid X = x) - \mathrm{E}(Y)\right)^2 \cdot p_X(x) \tag{17}
$$

根据全期望定理，$\mathrm{E}(Y \mid X = x)$ 的期望为 $\mathrm{E}(Y)$。

换个方向思考，(15) 相当于对 var(Y) 的分解：

$$
\begin{aligned}
\operatorname{var}(Y) &= \underbrace{\mathrm{E}\left(\operatorname{var}(Y \mid X)\right)}_{\text{Expectation of conditional variance}} + \underbrace{\operatorname{var}\left(\mathrm{E}(Y \mid X)\right)}_{\text{Variance of conditional expectation}} \\
&= \sum_x \underbrace{\operatorname{var}(Y \mid X = x)}_{\text{Deviation within a subset}} \cdot \underbrace{p_X(x)}_{\text{Weight}} + \underbrace{\sum_x \left(\overbrace{\mathrm{E}(Y \mid X = x) - \mathrm{E}(Y)}^{\text{Deviation of a subset from superset}}\right)^2 \cdot \overbrace{p_X(x)}^{\text{Weight}}}_{\text{Deviation among all subsets}}
\end{aligned} \tag{18}
$$

这方便我们理解哪些成分 (子集内部、子集之间)，以多大的比例贡献了整体的方差。

如图 6 所示，条件方差的期望解释的是子集 (鸡子集、兔子集) 各自内部差异。

条件期望的方差解释的是子集 (鸡子集、兔子集) 和母集 (所有动物) 之间的差异。

而代表鸡子集、兔子集就是鸡、兔各自的平均体重 (条件期望)，代表母集就是笼子里所有动物的平均体重 (总体期望)。

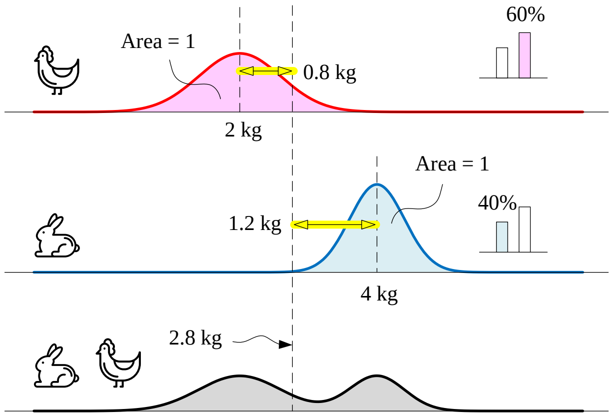

图 6. 解释全**方差**定理

如图 7 所示，子集内部差异 (**方差**) 不变，如果增大子集之间的差异，也就是增大了子集和母集的差异，这会导致整体的**方差**增大。

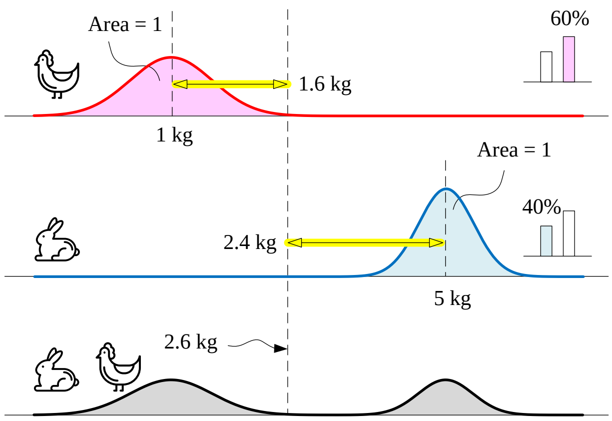

图 7. 解释全**方差**定理，增大子集之间差异，整体**方差**增大

条件**方差** $\operatorname{var}(X|Y=y)$

给定 $Y=y$ 条件下，X 的条件**方差** $E(X|Y=y)$ (conditional variance of X given $Y=y$) 定义为：

$$
\begin{aligned}
\operatorname{var}(X|Y=y) &= \overbrace{\sum_{x}\Big(\underbrace{x-\overbrace{E(X|Y=y)}^{\text{Expectation}}}_{\text{Deviation}}\Big)^2 \cdot \underbrace{p_{X|Y}(x|y)}_{\text{Conditional}}}^{\text{Expectation}} \\
&= \sum_{x}\big(x-E(X|Y=y)\big)^2 \cdot \frac{\overbrace{p_{X,Y}(x,y)}^{\text{Joint}}}{\underbrace{p_Y(y)}_{\text{Marginal}}} \\
&= \frac{1}{\underbrace{p_Y(y)}_{\text{Marginal}}}\sum_{x}\Big(\underbrace{x-E(X|Y=y)}_{\text{Deviation}}\Big)^2 \cdot \overbrace{p_{X,Y}(x,y)}^{\text{Joint}}
\end{aligned} \tag{19}
$$

条件**方差** $\mathrm{var}(X|Y=y)$ 也有如下计算技巧：

$$\mathrm{var}\left(X\,|Y=y\right)=\mathrm{E}\left(X^2\,|Y=y\right)-\mathrm{E}\left(X\,|Y=y\right)^2 \tag{20}$$

对于随机变量 X，它的全**方差**定理为：

$$\mathrm{var}\left(X\right)=\underbrace{\mathrm{E}\left(\mathrm{var}\left(X\mid Y\right)\right)}_{\text{Expectation of conditional variance}}+\underbrace{\mathrm{var}\left(\mathrm{E}\left(X\mid Y\right)\right)}_{\text{Variance of conditional expectation}} \tag{21}$$

8.3 离散随机变量条件期望、条件方差：以鸢尾花为例

给定花萼长度，条件期望 $\mathrm{E}(X_2 \mid X_1 = x_1)$

大家已经在本书第 4 章见过图 8 中左图。这幅图给出的是条件概率 $p_{X2 \mid X1}(x_2 \mid x_1)$。提醒大家回忆，图中 $p_{X2 \mid X1}(x_2 \mid x_1)$ 每列 PMF (即概率) 和为 1，即满足 (4)。

下面，我们试着利用图 8 中左图计算花萼长度 $X_1 = 6.5$ 为条件下，条件**期望** $\mathrm{E}(X_2 \mid X_1 = 6.5)$：

$$\begin{aligned}\mathrm{E}\left(X_2\,|X_1=6.5\right)&=\sum_{x_2}x_2\cdot p_{X2|X1}\left(x_2\mid 6.5\right)\\&=\underset{\text{cm}}{2.0}\times 0+\underset{\text{cm}}{2.5}\times 0.19+\underset{\text{cm}}{3.0}\times 0.65+\underset{\text{cm}}{3.5}\times 0.16+\underset{\text{cm}}{4.0}\times 0+\underset{\text{cm}}{4.5}\times 0\\&\approx 2.984\ \text{cm}\end{aligned} \tag{22}$$

注意，上式中条件概率的结果还是 cm。建议大家手算剩余所有 $\mathrm{E}(X_2 \mid X_1 = x_1)$。

图 8 中右上图给出的是热图 $x_2\cdot p_{X2|X1}\left(x_2\mid x_1\right)$，相当于一个二元函数。

图 9 所示为从矩阵乘法视角看条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$ 运算。

图 10 所示为条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$ 的火柴梗图。图 10 中还给出了鸢尾花花萼长度 X_1 的边缘 PMF $p_{X1}(x_1)$。

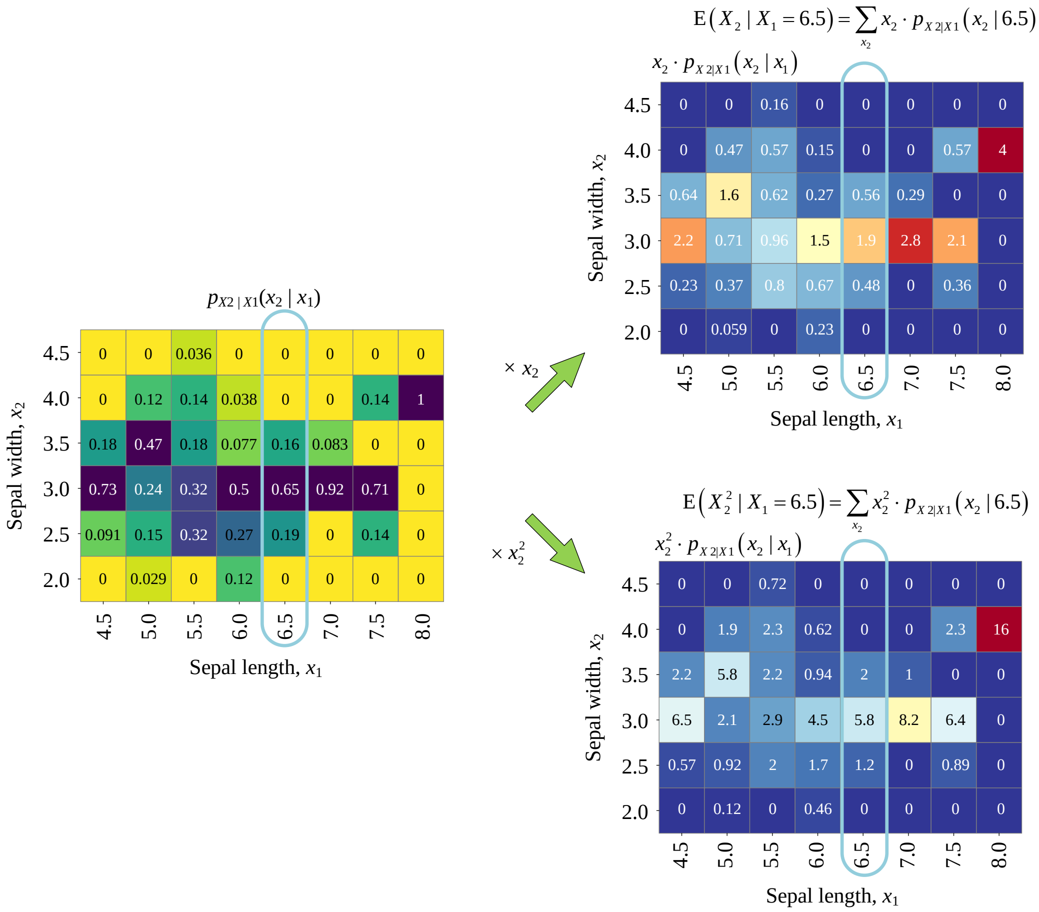

图 8. 给定花萼长度 X_1，花萼宽度 X_2 的条件概率 $p_{X2 \mid X1}(x_2 \mid x_1)$ 热图，$x_2 \times p_{X2 \mid X1}(x_2 \mid x_1)$ 热图，$x_2^2 \cdot p_{X2 \mid X1}\left(x_2 \mid x_1\right)$ 热图

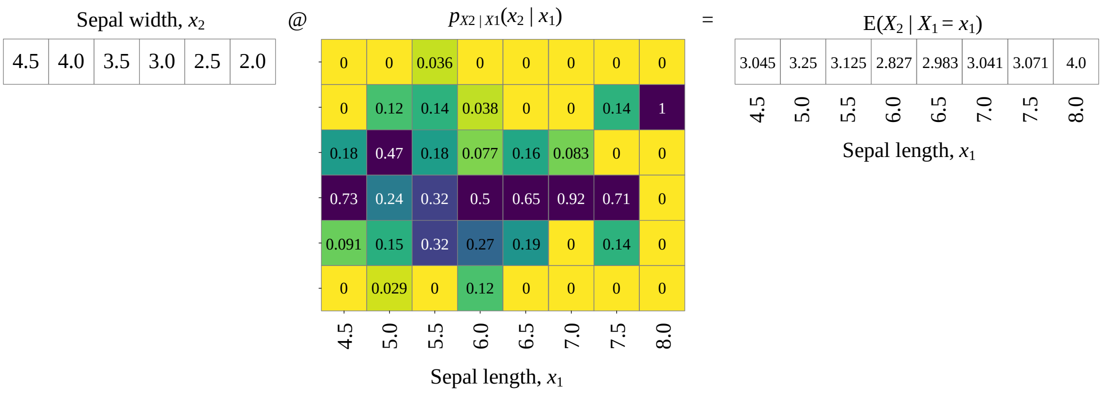

图 9. 矩阵乘法视角看条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$

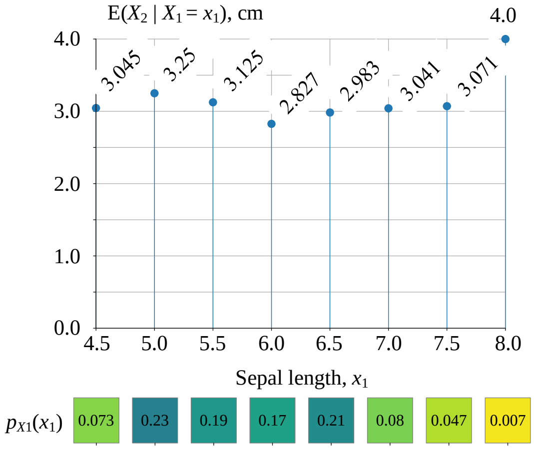

图 10. 给定花萼长度 X_1，花萼宽度 X_1 的条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$，和边缘 PMF $p_{X1}(x_1)$

根据 (5) 的全**期望**定理，我们可以利用条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$ 和边缘 PMF $p_{X1}(x_1)$ 计算**期望** $\mathrm{E}(X_2)$：

$$
\begin{aligned}
\mathrm{E}(X_2) &= \sum_{x_1} \mathrm{E}(X_2 \mid X_1 = x_1) \cdot p_{X1}(x_1) \\
&= \underbrace{3.045}_{\text{cm}} \times 0.073 + \underbrace{3.25}_{\text{cm}} \times 0.23 + \underbrace{3.125}_{\text{cm}} \times 0.19 + \underbrace{2.827}_{\text{cm}} \times 0.17 + \\
&\quad \underbrace{2.983}_{\text{cm}} \times 0.21 + \underbrace{3.041}_{\text{cm}} \times 0.08 + \underbrace{3.071}_{\text{cm}} \times 0.047 + \underbrace{4}_{\text{cm}} \times 0.007 \\
&\approx 3.063 \text{ cm}
\end{aligned}
\tag{23}
$$

给定花萼长度，条件方差 $\mathrm{var}(X_2 \mid X_1 = x_1)$

利用 (12) 计算花萼长度 $X_1 = 6.5$ 为条件下，条件**方差** $\mathrm{var}(X_2 \mid X_1 = 6.5)$：

$$
\begin{aligned}
\mathrm{var}(X_2 \mid X_1 = 6.5) &= \sum_{x_2} \left(x_2 - \mathrm{E}(X_2 \mid X_1 = 6.5)\right) \cdot p_{X2|X1}(x_2 \mid 6.5) \\
&= \underbrace{(2.0 - 2.985)^2}_{\text{cm}^2} \times 0 + \underbrace{(2.5 - 2.985)^2}_{\text{cm}^2} \times 0.19 + \underbrace{(3.0 - 2.985)^2}_{\text{cm}^2} \times 0.65 + \\
&\quad \underbrace{(3.5 - 2.985)^2}_{\text{cm}^2} \times 0.16 + \underbrace{(4.0 - 2.985)^2}_{\text{cm}^2} \times 0 + \underbrace{(4.0 - 2.985)^2}_{\text{cm}^2} \times 0 \\
&\approx 0.088 \text{ cm}^2
\end{aligned}
\tag{24}
$$

条件**方差** $\mathrm{var}(X_2 \mid X_1 = 6.5)$ 的单位为 cm^2。同样建议大家手算剩余条件**方差** $\mathrm{var}(X_2 \mid X_1 = x_1)$。

采用技巧计算，计算条件**期望**。首先计算花萼长度 $X_1 = 6.5$ 为条件下，花萼宽度平方的**期望**：

$$\begin{aligned}\mathrm{E}\left(X_1^2 \middle| X_1 = 6.5\right) &= \sum_{x_2} x_1^2 \cdot p_{X2|X1}\left(x_2 \mid 6.5\right) \\ &= \underbrace{2.0^2}_{\text{cm}^2} \times 0 + \underbrace{2.5^2}_{\text{cm}^2} \times 0.19 + \underbrace{3.0^2}_{\text{cm}^2} \times 0.65 + \underbrace{3.5^2}_{\text{cm}^2} \times 0.16 + \underbrace{4.0^2}_{\text{cm}^2} \times 0 + \underbrace{4.5^2}_{\text{cm}^2} \times 0 \\ &\approx 9 \ \text{cm}^2\end{aligned} \tag{25}$$

图 11 所示为花萼宽度平方值 X_2^2 的条件**期望** $\mathrm{E}\left(X_2^2 \mid X_1 = x_1\right)$ 的火柴梗图。

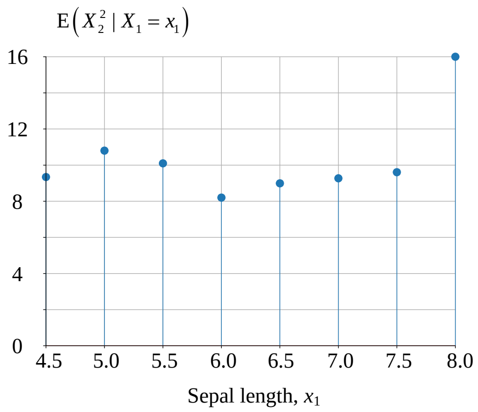

图 11. 给定花萼长度 X_1，花萼宽度平方值 X_2^2 的条件**期望** $\mathrm{E}\left(X_2^2 \mid X_1 = x_1\right)$

然后计算条件**方差**:

$$\mathrm{var}\left(X_2 \middle| X_1 = 6.5\right) = \mathrm{E}\left(X_1^2 \middle| X_1 = 6.5\right) - \mathrm{E}\left(X_2 \middle| X_1 = 6.5\right)^2 = 9 - 2.984^2 \approx 0.088 \tag{26}$$

图 12 所示为花萼长度取不同值时条件**方差** $\mathrm{var}(X_2 \mid X_1 = x_1)$ 的火柴梗图，请大家用作检查自己手算结果。

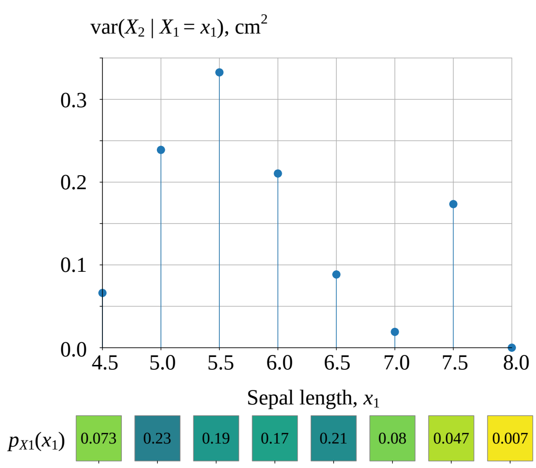

图 12. 给定花萼长度 X_1，花萼宽度的条件**方差** $\mathrm{var}(X_2 \mid X_1 = x_1)$

大家肯定早就发现，条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$、条件**方差** $\mathrm{var}(X_2 \mid X_1 = x_1)$ 都消去了 x_2 这个变量，两者仅仅随着 $X_1 = x_1$ 取值变化。这也不难理解，**期望**和**方差**代表“汇总”，本质上就是“降维”。某个维度上的信息细节不再重要，我们把这个“压扁”。

压扁过程中，不同的聚合方式得到不同的统计量，比如**期望**、**方差**等等。

全方差定理：还原方差 $\mathrm{var}(X_2)$

根据 (17) 中给出的全**方差**定理，下面我们利用条件**方差** $\mathrm{var}(X_2 \mid X_1)$ 和条件**期望** $\mathrm{E}(X_2 \mid X_1)$ 计算花萼宽度的**方差** $\mathrm{var}(X_2)$。$\mathrm{var}(X_2)$ 可以写成两部分之和：

$$\mathrm{var}(X_2) = \underbrace{\mathrm{E}\left(\mathrm{var}(X_2 \mid X_1)\right)}_{\text{Expectation of conditional variance}} + \underbrace{\mathrm{var}\left(\mathrm{E}(X_2 \mid X_1)\right)}_{\text{Variance of conditional expectation}} \tag{27}$$

第一部分是条件**方差**的**期望** $\mathrm{E}\left(\mathrm{var}(X_2 \mid X_1)\right)$：

$$\underbrace{\mathrm{E}\left(\mathrm{var}(X_2 \mid X_1)\right)}_{\text{Expectation of conditional variance}} = \sum_{x_1} \mathrm{var}(X_2 \mid X_1 = x_1) \cdot p_{X1}(x_1) \tag{28}$$

代入具体数值，我们可以计算得到 $\mathrm{E}\left(\mathrm{var}(X_2 \mid X_1)\right)$：

$$\begin{aligned}
\underbrace{\mathrm{E}\left(\mathrm{var}(X_2 \mid X_1)\right)}_{\text{Expectation of conditional variance}} &= \sum_{x_1} \mathrm{var}(X_2 \mid X_1 = x_1) \cdot p_{X1}(x_1) \\
&\approx \underbrace{0.066}_{\text{cm}^2} \times 0.073 + \underbrace{0.238}_{\text{cm}^2} \times 0.226 + \underbrace{0.332}_{\text{cm}^2} \times 0.186 + \underbrace{0.210}_{\text{cm}^2} \times 0.173 + \\
&\quad \underbrace{0.088}_{\text{cm}^2} \times 0.206 + \underbrace{0.019}_{\text{cm}^2} \times 0.08 + \underbrace{0.173}_{\text{cm}^2} \times 0.046 + \underbrace{0}_{\text{cm}^2} \times 0.006 \\
&\approx \underbrace{0.0048}_{X_1=4.5} + \underbrace{0.0541}_{X_1=5.0} + \underbrace{0.0620}_{X_1=5.5} + \underbrace{0.0364}_{X_1=6.0} + \underbrace{0.0182}_{X_1=6.5} + \underbrace{0.0015}_{X_1=7.0} + \underbrace{0.0080}_{X_1=7.5} + \underbrace{0}_{X_1=8.0} \\
&\approx 0.185\ \text{cm}^2
\end{aligned} \tag{29}$$

第二部分是条件**期望**的**方差** $\mathrm{var}\left(\mathrm{E}(X_2 \mid X_1)\right)$。代入具体值计算得到：

$$\begin{aligned}
\underbrace{\mathrm{var}\left(\mathrm{E}(X_2 \mid X_1)\right)}_{\text{Variance of conditional expectation}} &= \sum_{x_1} \left(\mathrm{E}(X_2 \mid X_1 = x_1) - \mathrm{E}(X_2)\right)^2 \cdot p_{X_1}(x_1) \\
&\approx 0.025\ \text{cm}^2
\end{aligned} \tag{30}$$

如果大家看到这还会犯糊涂，不理解为什么 $\sum_{x_1}$ 求和遍历的是 x_1？我告诉大家一个小技巧，因为 X_2 已经被“折叠”！不管是条件期望 $\mathrm{E}(X_2 \mid X = x_1)$、还是期望 $\mathrm{E}(X_2)$，都已经将 X_2 折叠成一个具体的数值，因此无法遍历。

这样 X_2 的方差约为：

$$\operatorname{var}(X_2) = \underbrace{\mathrm{E}\left(\operatorname{var}(X_2 \mid X_1)\right)}_{\text{Expectation of conditional variance}} + \underbrace{\operatorname{var}\left(\mathrm{E}(X_2 \mid X_1)\right)}_{\text{Variance of conditional expectation}} \approx 0.185 + 0.025 = 0.211 \text{cm}^2 \tag{31}$$

在 $\operatorname{var}(X_2)$ 中，第一部分 $\mathrm{E}\left(\operatorname{var}(X_2 \mid X_1)\right)$ 贡献超过 85%。而 $\mathrm{E}\left(\operatorname{var}(X_2 \mid X_1)\right)$ 可以进一步展开，图 13 所示为各个不同成分对花萼宽度 X_2 的方差 $\operatorname{var}(X_2)$ 的贡献，这也可以叫做钻取 (drill down)。

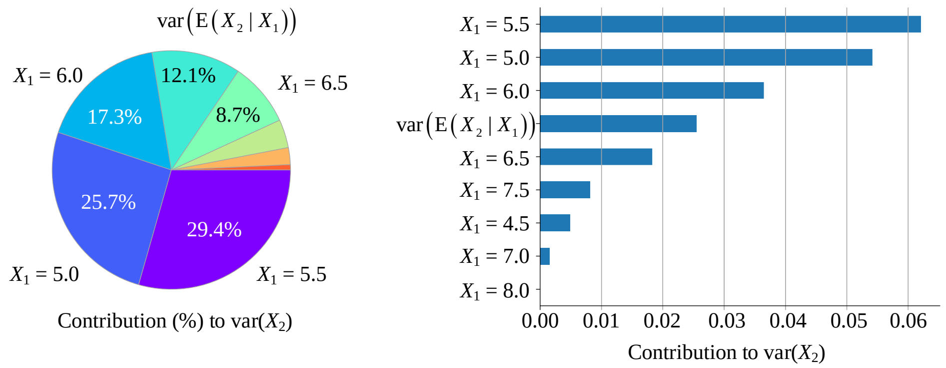

图 13. 各个不同成分对花萼宽度 X_2 的方差 $\operatorname{var}(X_2)$ 的贡献

给定花萼长度，条件标准差 $\operatorname{std}(X_2 \mid X_1 = x_1)$

(24) 开方便获得条件**标准差** $\operatorname{std}(X_2 \mid X_1 = 6.5)$：

$$\sigma_{X_2 \mid X_1 = 6.5} = \operatorname{std}\left(X_2 \mid X_1 = 6.5\right) = 0.295 \text{ cm} \tag{32}$$

上式的单位和鸢尾花宽度单位一致，我们便可以把条件**标准差**和图 10 画在一起，得到图 14。这幅图给出的是 $\mathrm{E}(X_2 \mid X_1 = x_1) \pm \operatorname{std}(X_2 \mid X_1 = x_1)$。

圆点 ● 展示的是 $\mathrm{E}(X_2 \mid X_1 = x_1)$，即条件**期望**，代表给定 $X_1 = x_1$ 条件下，鸢尾花数据在花萼宽度上的一种“预测”！这和我们讲过的回归思想本质上相同。$\mathrm{E}(X_2 \mid X_1 = x_1)$ 代表当 $X_1 = x_1$ 时鸢尾花花萼宽度最合适的“预测”。也就是说，回归可以看成是条件概率！本书后续还会沿着这个思路展开讨论。

而我们用误差棒 (error bar) 展示± $\operatorname{std}(X_2 \mid X_1 = x_1)$，代表给定 $X_1 = x_1$ 条件下，鸢尾花数据在花萼宽上的“波动”。误差棒的宽度越大，说明波动越大；反之，则说明波动越小。

特别地，当花萼长度 X_1 为 8.0 cm 时，条件均**方差** $\operatorname{std}(X_2 \mid X_1 = 8.0)$ 为 0。这是因为，这一处只有一个样本点。

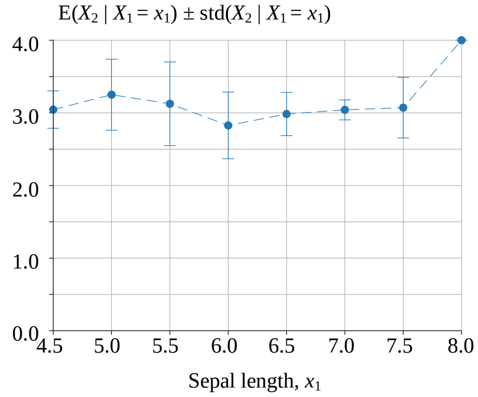

图 14. 给定花萼长度 X_1，花萼宽度 X_1 的条件**期望** $E(X_2 | X_1 = x_1) \pm std(X_2 | X_1 = x_1)$

给定花萼宽度，条件期望 $E(X_1 | X_2 = x_2)$

图 15 给出的是条件概率 $p_{X1 | X2}(x_1 | x_2)$。同样提醒大家注意图中 $p_{X1 | X2}(x_1 | x_2)$ 每行 PMF (即概率)和为 1。

利用 图 15 计算花萼宽度 $X_2 = 2.0$ 为条件下，条件**期望** $E(X_1 | X_2 = 2.0)$：

$$
\begin{aligned}
E\left(X_1 | X_2 = 2.0\right) &= \sum_{x_1} x_1 \cdot p_{X1|X2}\left(x_2 | 2.0\right) \\
&= \underset{\text{cm}}{4.5} \times 0 + \underset{\text{cm}}{5.0} \times 0.25 + \underset{\text{cm}}{5.5} \times 0 + \underset{\text{cm}}{6.0} \times 0.75 + \\
&\quad \underset{\text{cm}}{6.5} \times 0 + \underset{\text{cm}}{7.0} \times 0 + \underset{\text{cm}}{7.5} \times 0 + \underset{\text{cm}}{8.0} \times 0 \\
&\approx 5.7 \text{ cm}
\end{aligned} \tag{33}
$$

条件概率的结果还是 cm。同样建议大家手算剩余所有 $E(X_1 | X_2 = x_2)$。

此外，请大家也根据全**期望**定理，利用 $E(X_1 | X_2 = x_2)$ 计算 $E(X_1)$。并用条件**方差** $var(X_1 | X_2)$ 和条件**期望** $E(X_1 | X_2)$ 计算花萼长度的**方差** $var(X_1)$。

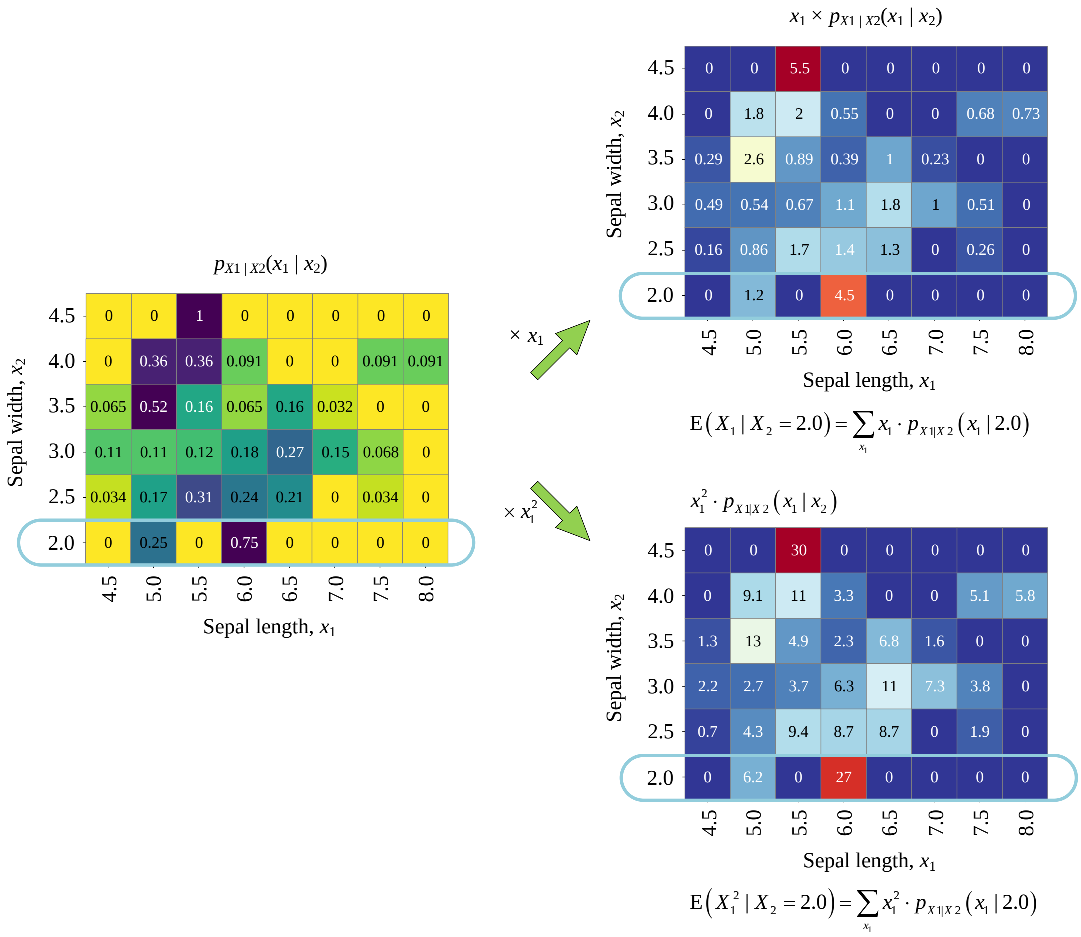

图 15. 给定花萼宽度，花萼长度的条件概率 $p_{X1 | X2}(x_1 | x_2)$

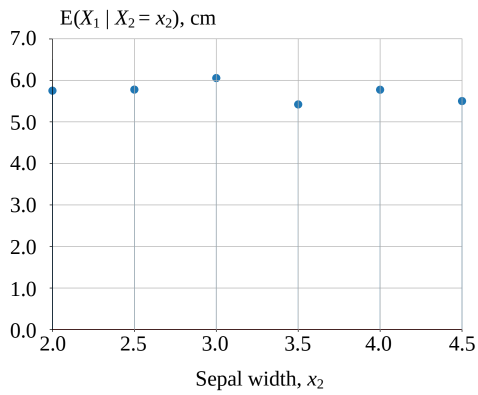

图 16. 给定花萼宽度 X_2，花萼宽度的条件**期望** $\mathrm{E}(X_1 | X_2 = x_2)$

条件方差 $\mathrm{var}(X_1 | X_2 = x_2)$

在花萼宽度 $X_2 = 2.0$ 为条件下，条件**方差** $\text{var}(X_1 \mid X_2 = 2.0)$：

$$\begin{aligned}
\text{var}\left(X_1 \middle| X_2 = 2.0\right) &= \sum_{x_1}\left(x_1 - \text{E}\left(X_1 \middle| X_2 = 2.0\right)\right)\cdot p_{X1|X2}\left(x_1 \mid 2.0\right) \\
&= \underbrace{\left(4.5-5.75\right)^2}_{\text{cm}^2}\times 0 + \underbrace{\left(5.0-5.75\right)^2}_{\text{cm}^2}\times 0.25 + \underbrace{\left(5.5-5.75\right)^2}_{\text{cm}^2}\times 0 + \underbrace{\left(6.0-5.75\right)^2}_{\text{cm}^2}\times 0.75 + \\
&\quad \underbrace{\left(6.5-5.75\right)^2}_{\text{cm}^2}\times 0 + \underbrace{\left(7.0-5.75\right)^2}_{\text{cm}^2}\times 0 + \underbrace{\left(7.5-5.75\right)^2}_{\text{cm}^2}\times 0 + \underbrace{\left(8.0-5.75\right)^2}_{\text{cm}^2}\times 0 \\
&= 0.1875\ \text{cm}^2
\end{aligned} \tag{34}$$

条件**方差** $\text{var}(X_1 \mid X_2 = 2.0)$ 的单位为 cm^2。同样建议大家手算剩余条件**方差** $\text{var}(X_1 \mid X_2 = x_2)$。

利用条件**方差**计算技巧，首先计算花萼宽度 $X_2 = 2.0$ 为条件下，花萼长度平方的**期望**：

$$\begin{aligned}
\text{E}\left(X_1^2 \middle| X_2 = 2.0\right) &= \sum_{x_1} x_1^2 \cdot p_{X1|X2}\left(x_2 \mid 2.0\right) \\
&= \underbrace{4.5^2}_{\text{cm}^2}\times 0 + \underbrace{5.0^2}_{\text{cm}^2}\times 0.25 + \underbrace{5.5^2}_{\text{cm}^2}\times 0 + \underbrace{6.0^2}_{\text{cm}^2}\times 0.75 + \\
&\quad \underbrace{6.5^2}_{\text{cm}^2}\times 0 + \underbrace{7.0^2}_{\text{cm}}\times 0 + \underbrace{7.5^2}_{\text{cm}^2}\times 0 + \underbrace{8.0^2}_{\text{cm}^2}\times 0 \\
&= 33.25\ \text{cm}^2
\end{aligned} \tag{35}$$

图 17 所示为给定花萼长度 X_2，花萼宽度平方值 X_1^2 的条件**期望** $\text{E}\left(X_1^2 \mid X_2 = x_2\right)$。请大家自行代入计算条件**方差** $\text{var}(X_1 \mid X_2 = 2.0)$。

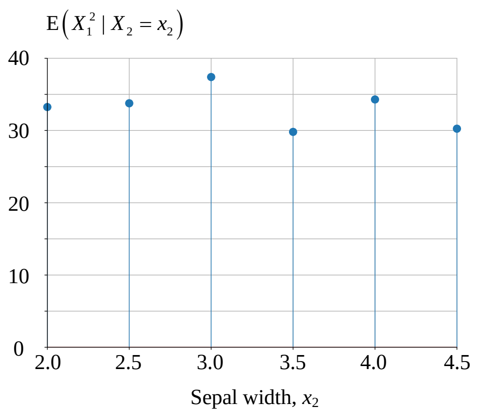

图 17. 给定花萼长度 X_2，花萼宽度平方值 X_1^2 的条件**期望** $\text{E}\left(X_1^2 \mid X_2 = x_2\right)$

图 18 所示为条件**方差** $\text{var}(X_1 \mid X_2 = x_2)$ 的火柴梗图。同样，条件**期望** $\text{E}(X_1 \mid X_2 = x_2)$、条件**方差** $\text{var}(X_1 \mid X_2 = x_2)$ 都"折叠"了 x_1 这个维度，两者仅仅随着 $X_2 = x_2$ 取值变化。

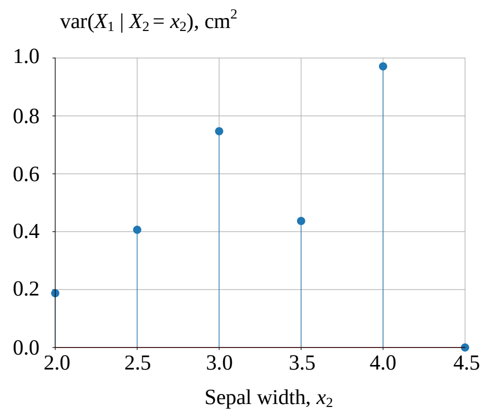

图 18. 给定花萼宽度 X_2，花萼宽度的条件**方差** $\mathrm{var}(X_1 \mid X_2 = x_2)$

给定花萼长度，条件标准差 $\mathrm{std}(X_2 \mid X_1 = x_1)$

(24) 开方便获得条件标准差 $\mathrm{std}(X_2 \mid X_1 = 6.5)$：

$$\sigma_{X_2|X_1=6.5} = \mathrm{std}\left(X_2 \middle| X_1 = 6.5\right) = 0.295 \text{ cm} \tag{36}$$

上式的单位和鸢尾花宽度单位一致。类似图 14，我们也绘制给定花萼宽度 X_2，花萼长度 X_1 的条件**期望** $\mathrm{E}(X_1 \mid X_2 = x_2) \pm \mathrm{std}(X_1 \mid X_2 = x_2)$。请大家自行分析这幅图像。

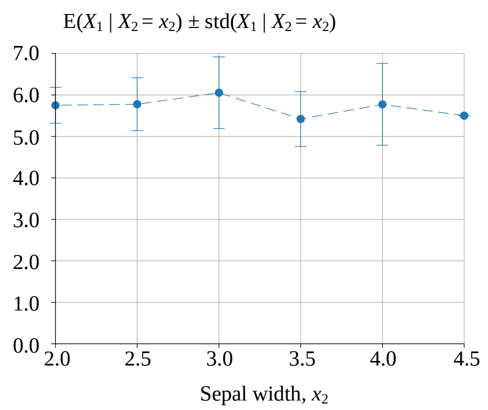

图 19. 给定花萼宽度 X_2，花萼长度 X_1 的条件**期望** $\mathrm{E}(X_1 \mid X_2 = x_2) \pm \mathrm{std}(X_1 \mid X_2 = x_2)$

考虑标签：花萼长度

给定鸢尾花分类标签 $Y = C_1$，花萼长度 X_1 的条件**期望**：

$$
\begin{aligned}
\mathrm{E}\left(X_1 \middle| Y=C_1\right) &= \sum_{x_1} x_1 \cdot p_{X1|Y}\left(x_1 \mid C_1\right) \\
&= \underbrace{4.5}_{\text{cm}} \times 0.22 + \underbrace{5.0}_{\text{cm}} \times 0.56 + \underbrace{5.5}_{\text{cm}} \times 0.2 + \underbrace{6.0}_{\text{cm}} \times 0.02 + \\
&\quad \underbrace{6.5}_{\text{cm}} \times 0 + \underbrace{7.0}_{\text{cm}} \times 0 + \underbrace{7.5}_{\text{cm}} \times 0 + \underbrace{8.0}_{\text{cm}} \times 0 \\
&= 5.01 \text{ cm}
\end{aligned}
\tag{37}
$$

给定鸢尾花分类标签 $Y = C_1$，花萼长度 X_1 平方**期望**：

$$
\begin{aligned}
\mathrm{E}\left(X_1^2 \middle| Y=C_1\right) &= \sum_{x_1} x_1^2 \cdot p_{X1|Y}\left(x_1 \mid C_1\right) \\
&= \underbrace{4.5^2}_{\text{cm}^2} \times 0.22 + \underbrace{5.0^2}_{\text{cm}^2} \times 0.56 + \underbrace{5.5^2}_{\text{cm}^2} \times 0.2 + \underbrace{6.0^2}_{\text{cm}^2} \times 0.02 + \\
&\quad \underbrace{6.5^2}_{\text{cm}^2} \times 0 + \underbrace{7.0^2}_{\text{cm}^2} \times 0 + \underbrace{7.5^2}_{\text{cm}^2} \times 0 + \underbrace{8.0^2}_{\text{cm}^2} \times 0 \\
&= 25.225 \text{ cm}^2
\end{aligned}
\tag{38}
$$

给定鸢尾花分类标签 $Y = C_1$，花萼长度 X_1 条件**方差**：

$$
\begin{aligned}
\operatorname{var}\left(X_1 \middle| Y=C_1\right) &= \mathrm{E}\left(X_1^2 \middle| Y=C_1\right) - \mathrm{E}\left(X_1 \middle| Y=C_1\right)^2 \\
&= 25.225 - 5.01^2 \\
&= 0.1249 \text{ cm}^2
\end{aligned}
\tag{39}
$$

给定鸢尾花分类标签 $Y = C_1$，花萼长度 X_1 条件**标准差**：

$$
\sigma_{X_1|Y=C_1} = \sqrt{\operatorname{var}\left(X_1 \middle| Y=C_1\right)} = \sqrt{0.1249} = 0.353 \text{ cm}
\tag{40}
$$

请大家自行计算剩余两种情况 ($Y = C_2, C_3$)。并利用全**期望**定理，计算 $\mathrm{E}(X_1)$。

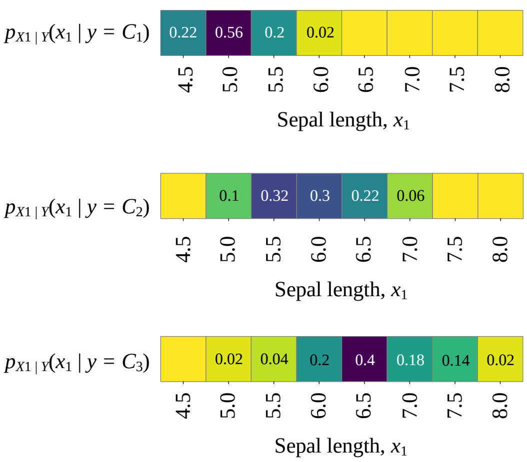

图 20. 给定鸢尾花标签 Y，花萼长度的条件 PMF

考虑标签：花萼宽度

给定鸢尾花分类标签 $Y = C_1$，花萼宽度 X_2 的条件**期望**：

$$\begin{aligned}\mathrm{E}\left(X_2 \mid Y=C_1\right) &= \sum_{x_2} x_2 \cdot p_{X2 \mid Y}\left(x_2 \mid C_1\right) \\ &= \underset{\text{cm}}{4.5} \times 0.07 + \underset{\text{cm}}{4.0} \times 0.18 + \underset{\text{cm}}{3.5} \times 0.46 + \underset{\text{cm}}{3.0} \times 0.32 + \underset{\text{cm}}{2.5} \times 0.02 + \underset{\text{cm}}{2.0} \times 0 \\ &= 3.43 \text{ cm}\end{aligned} \tag{41}$$

给定鸢尾花分类标签 $Y = C_1$，花萼宽度 X_2 平方**期望**：

$$\begin{aligned}\mathrm{E}\left(X_2^2 \mid Y=C_1\right) &= \sum_{x_2} x_2^2 \cdot p_{X2 \mid Y}\left(x_2 \mid C_1\right) \\ &= \underset{\text{cm}^2}{4.5^2} \times 0.07 + \underset{\text{cm}^2}{4.0^2} \times 0.18 + \underset{\text{cm}^2}{3.5^2} \times 0.46 + \underset{\text{cm}^2}{3.0^2} \times 0.32 + \underset{\text{cm}^2}{2.5^2} \times 0.02 + \underset{\text{cm}^2}{2.0^2} \times 0 \\ &= 11.925 \text{ cm}^2\end{aligned} \tag{42}$$

给定鸢尾花分类标签 $Y = C_1$，花萼宽度 X_2 条件**方差**：

$$\begin{aligned}\operatorname{var}\left(X_2 \mid Y=C_1\right) &= \mathrm{E}\left(X_2^2 \mid Y=C_1\right) - \mathrm{E}\left(X_2 \mid Y=C_1\right)^2 \\ &= 11.925 - 3.43^2 \\ &= 0.1601 \text{ cm}^2\end{aligned} \tag{43}$$

给定鸢尾花分类标签 $Y = C_1$，花萼宽度 X_2 条件**标准差**：

$$\sigma_{X_2 \mid Y=C_1} = \sqrt{\operatorname{var}\left(X_2 \mid Y=C_1\right)} = \sqrt{0.1601} \approx 0.4 \text{ cm} \tag{44}$$

请大家自行计算鸢尾花其他标签条件下花萼长度、花萼宽度的条件**期望**、条件**方差**、条件**标准差**。

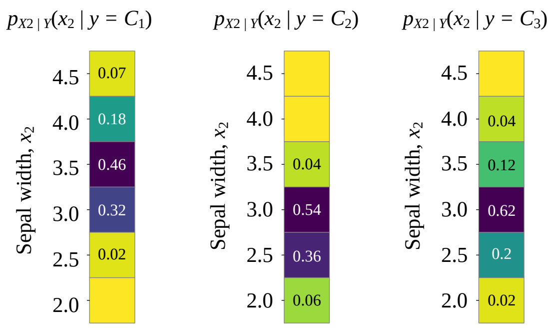

图 21. 给定鸢尾花标签 Y，花萼宽度的条件**期望** $\mathrm{E}(X_2 \mid Y = C_k)$、条件**方差** $\operatorname{var}(X_2 \mid Y = C_k)$，离散随机变量

Bk5_Ch08_01.py 代码绘制本节大部分图像。代码中用到了矩阵乘法和广播原则，请大家注意区分。

8.4 连续随机变量：条件期望

本节介绍如何计算连续随机变量的条件期望。

条件期望 E(Y|X = x)

如果 X 和 Y 均为连续随机变量，如图 22 所示，在给定 $X = x$ 条件下，条件期望 $\mathrm{E}(Y|X = x)$ 定义为：

$$\begin{aligned}\mathrm{E}\big(Y \,\big|\, \underbrace{X=x}_{\text{Given}}\big) &= \overbrace{\int_{-\infty}^{+\infty} y \cdot \underbrace{f_{Y|X}\left(y|x\right)}_{\text{Conditional}} \mathrm{d}\,y}^{\text{Expectation}} \\ &= \int_{-\infty}^{+\infty} y \cdot \frac{\overbrace{f_{X,Y}\left(x,y\right)}^{\text{Joint}}}{\underbrace{f_X\left(x\right)}_{\text{Marginal}}} \mathrm{d}\,y = \frac{1}{\underbrace{f_X\left(x\right)}_{\text{Marginal}}} \int_{-\infty}^{+\infty} y \cdot \overbrace{f_{X,Y}\left(x,y\right)}^{\text{Joint}} \mathrm{d}\,y\end{aligned} \tag{45}$$

上式中，边缘概率 $f_X(x)$ 可以通过下式得到：

$$f_X\left(x\right) = \int_{-\infty}^{+\infty} f_{X,Y}\left(x,y\right) \mathrm{d}\,y \tag{46}$$

(46) 代入 (45) 得到：

$$\mathrm{E}\left(Y|X=x\right) = \frac{1}{\int_{-\infty}^{+\infty} f_{X,Y}\left(x,y\right) \mathrm{d}\,y} \int_{-\infty}^{+\infty} y \cdot f_{X,Y}\left(x,y\right) \mathrm{d}\,y \tag{47}$$

上式，相当于消去了 y，这和本章前文提到的“降维”、“折叠”本质上没有任何区别。对于离散随机变量，折叠用的数学工具为求和符号 Σ；连续随机变量则用积分符号 ∫。

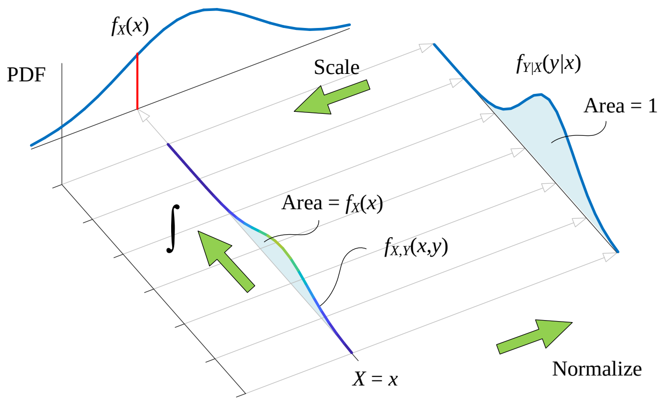

图 22. 联合概率 PDF $f_{X,Y}(x,y)$ 和条件概率 PDF $f_{Y|X}(y|x)$ 的关系，X 和 Y 均为连续随机变量

条件期望 $\mathrm{E}(X|Y = y)$

同理，如图 23 所示，条件期望 $\mathrm{E}(X|Y = y)$ 定义为：

$$\mathrm{E}\left(X\,|Y=y\right)=\frac{1}{\int_{-\infty}^{+\infty} f_{X,Y}\left(x,y\right)\mathrm{d}\,x}\int_{-\infty}^{+\infty} x\cdot f_{X,Y}\left(x,y\right)\mathrm{d}\,x \tag{48}$$

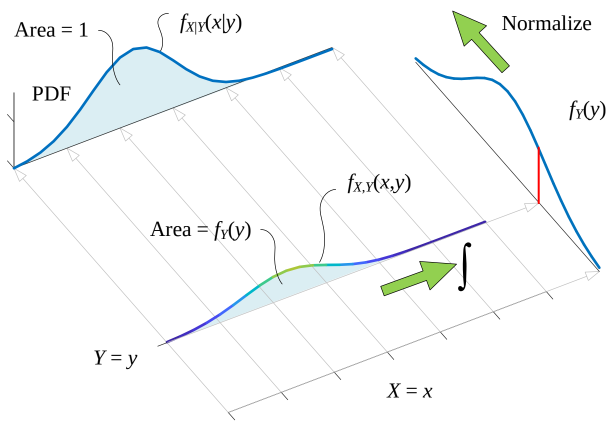

图 23. 联合概率 PDF $f_{X,Y}(x,y)$ 和条件概率 PDF $f_{X|Y}(x|y)$ 的关系，X 和 Y 均为连续随机变量

8.5 连续随机变量：条件方差

本节介绍如何求连续随机变量的条件方差。

条件方差 $\mathrm{var}(Y|X = x)$

在给定 $X = x$ 条件下，条件方差 $\mathrm{var}(Y|X = x)$ (conditional variance of Y given $X = x$) 定义为：

$$\begin{aligned}\mathrm{var}\left(Y\,|X=x\right)&=\mathrm{E}\left\{\left(Y-\mathrm{E}\left(Y\,|X=x\right)\right)^2|x\right\}\\&=\int_{-\infty}^{+\infty}\left(y-\mathrm{E}\left(Y\,|X=x\right)\right)^2\cdot f_{Y|X}\left(y|x\right)\mathrm{d}\,y\end{aligned} \tag{49}$$

对于连续随机变量，求条件方差也可以用 (14) 这个技巧。

条件方差 $\mathrm{var}(X|Y = y)$

条件方差 $\mathrm{var}(X|Y = y)$ 定义为：

$$\begin{aligned}\operatorname{var}(X|Y=y)&=\mathrm{E}\left\{\left(X-\mathrm{E}(X|Y=y)\right)^2|y\right\}\\&=\int_{-\infty}^{+\infty}\left(X-\mathrm{E}(X|Y=y)\right)^2\cdot f_{X|Y}(x|y)\,\mathrm{d}x\end{aligned}\tag{50}$$

有了以上理论基础，本书第 12 章将以二元高斯分布为例，继续深入讲解条件**期望**和条件**方差**。

8.6 连续随机变量：以鸢尾花为例

以鸢尾花为例：条件期望 $\mathrm{E}(X_2 \mid X_1 = x_1)$、条件方差 $\operatorname{var}(X_2 \mid X_1 = x_1)$

图 24 (a) 所示为条件概率 PDF $f_{X2 \mid X1}(x_2 \mid x_1)$ 随花萼长度、花萼宽度变化曲面。本书前文提过 $f_{X2 \mid X1}(x_2 \mid x_1)$ 也是一个二元函数。这个二元函数的重要特点有两个：

$$\begin{aligned}&f_{X2|X1}(x_2 \mid x_1)\geq 0\\&\int_{x_2} f_{X2|X1}(x_2 \mid x_1)\,\mathrm{d}x_2=1\end{aligned}\tag{51}$$

正如图 24 (a) 所示，阴影区域的面积为 1。

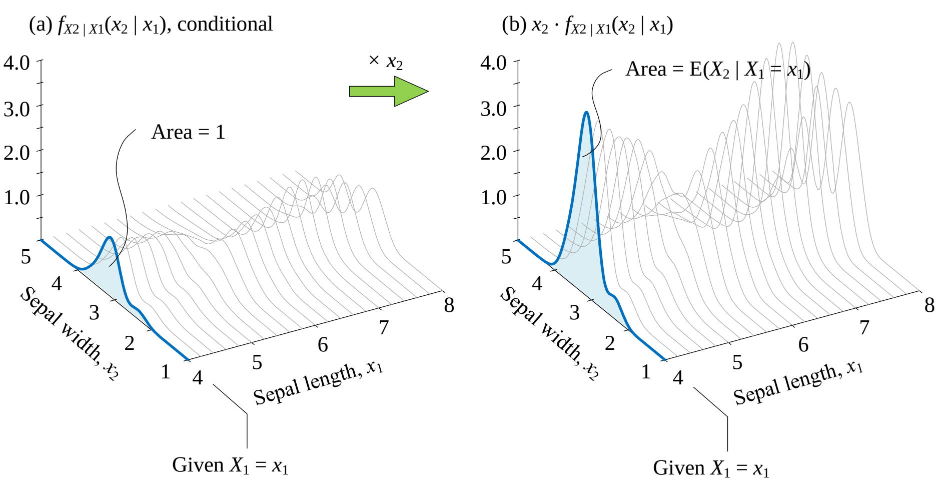

图 24. $f_{X2 \mid X1}(x_2 \mid x_1)$ 条件概率密度三维等高线和平面等高线，不考虑分类

根据 (47)，为了计算条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$，我们需要计算 $x_2 \cdot f_{X2 \mid X1}(x_2 \mid x_1)$ 和 x_2 围成图像的面积，即图 24 (b) 阴影部分面积：

$$\mathrm{E}\left(X_2 \mid X_1 = x_1\right) = \int_{-\infty}^{+\infty} x_2 \cdot \underbrace{f_{X2|X1}\left(x_2 \middle| x_1\right)}_{\text{Conditional}} \mathrm{d}\, x_2 \tag{52}$$

然后，我们可以计算鸢尾花宽度平方的条件**期望** $\mathrm{E}\left(X_2^2 \mid X_1 = x_1\right)$：

$$\mathrm{E}\left(X_2^2 \mid X_1 = x_1\right) = \int_{-\infty}^{+\infty} x_2^2 \cdot \underbrace{f_{X2|X1}\left(x_2 \middle| x_1\right)}_{\text{Conditional}} \mathrm{d}\, x_2 \tag{53}$$

然后，可以利用技巧求得条件**方差** $\mathrm{var}(X_2 \mid X_1 = x_1)$：

$$\mathrm{var}\left(X_2 \mid X_1 = x_1\right) = \mathrm{E}\left(X_2^2 \mid X_1 = x_1\right) - \mathrm{E}\left(X_2 \mid X_1 = x_1\right)^2 \tag{54}$$

上式开平方得到，条件均**方差** $\mathrm{std}(X_2 \mid X_1 = x_1)$。

我们知道条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$、条件均**方差** $\mathrm{std}(X_2 \mid X_1 = x_1)$ 都随着 $X_1 = x_1$ 取值变化，而且它们两个单位都是 cm。我们想办法把它们画在一幅图上，具体如图 25 所示。

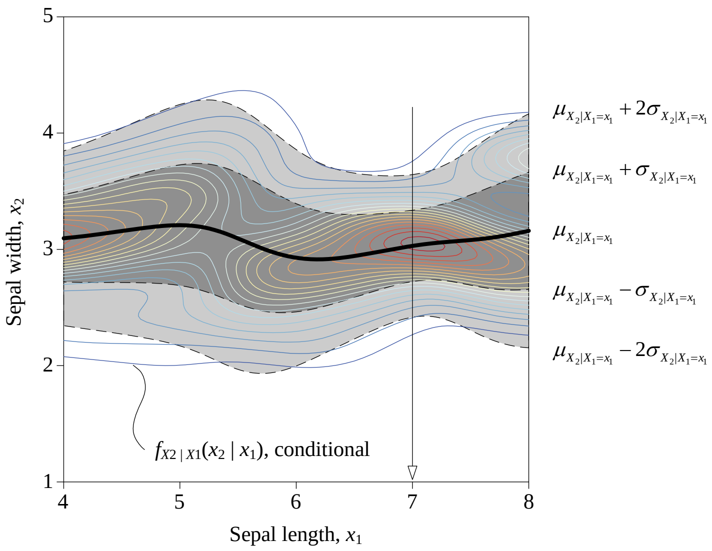

图 25. 条件**期望** $\mathrm{E}(X_2 \mid X_1 = x_1)$、条件均方差 $\mathrm{std}(X_2 \mid X_1 = x_1)$ 之间的关系

以鸢尾花为例：条件期望 $\mathrm{E}(X_1 \mid X_2 = x_2)$、条件方差 $\mathrm{var}(X_1 \mid X_2 = x_2)$

为了计算条件**期望** $\mathrm{E}(X_1 \mid X_2 = x_2)$，我们需要计算 $x_1 \cdot f_{X1 \mid X2}(x_1 \mid x_2)$ 和 x_1 围成图像的面积，即图 26 (b) 阴影部分面积：

$$\mathrm{E}\left(X_1 \mid X_2 = x_2\right) = \int_{-\infty}^{+\infty} x_1 \cdot \underbrace{f_{X1|X2}\left(x_1 \middle| x_2\right)}_{\text{Conditional}} \mathrm{d}\, x_1 \tag{55}$$

然后，我们可以计算鸢尾长度度平方的条件**期望** $\mathrm{E}\left(X_1^2 \mid X_2 = x_2\right)$：

$$\mathrm{E}\left(X_1^2 \mid X_2 = x_2\right) = \int_{-\infty}^{+\infty} x_1^2 \cdot \underbrace{f_{X1|X2}\left(x_1 \middle| x_2\right)}_{\text{Conditional}} \mathrm{d}\, x_1 \tag{56}$$

然后，可以利用技巧求得条件**方差** $\mathrm{var}(X_1 \mid X_2 = x_2)$：

$$\mathrm{var}\left(X_1 \mid X_2 = x_2\right) = \mathrm{E}\left(X_1^2 \mid X_2 = x_2\right) - \mathrm{E}\left(X_1 \mid X_2 = x_2\right)^2 \tag{57}$$

上式开平方得到条件均**方差** $\mathrm{std}(X_1 \mid X_2 = x_2)$。

我们知道条件**期望** $\mathrm{E}(X_1 \mid X_2 = x_2)$、条件**标准差** $\mathrm{std}(X_1 \mid X_2 = x_2)$ 都随着 $X_2 = x_2$ 取值变化，而且它们两个单位都是 cm。我们想办法把它们画在一幅图上，具体如图 27 所示。

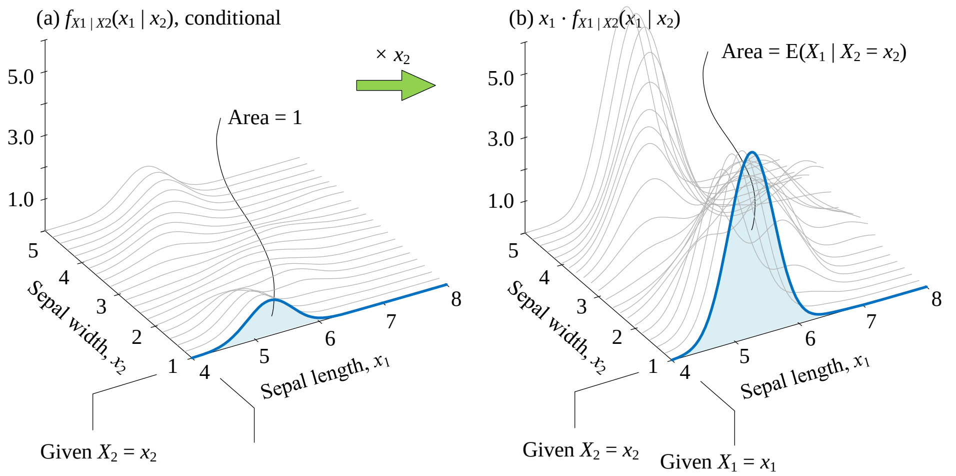

图 26. $f_{X1 | X2}(x_1 | x_2)$ 条件概率密度三维等高线和平面等高线

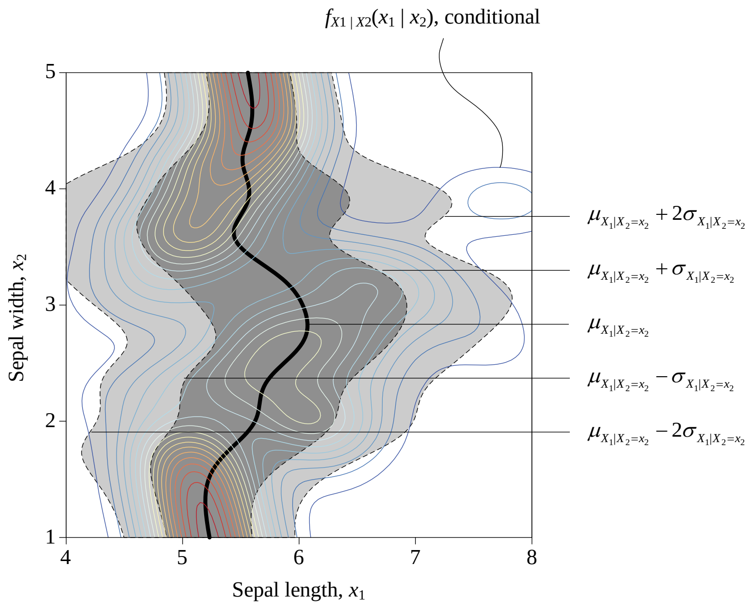

图 27. 条件**期望** $\mathrm{E}(X_1 \mid X_2 = x_2)$、条件**标准差** $\mathrm{std}(X_1 \mid X_2 = x_2)$ 之间的关系

以鸢尾花为例，考虑标签

同理，我们可以计算给定标签条件下，鸢尾花花萼长度 (图 28)、花萼宽度 (图 29) 的条件**期望**、条件**方差**等。请大家自己完成这几个数值计算。

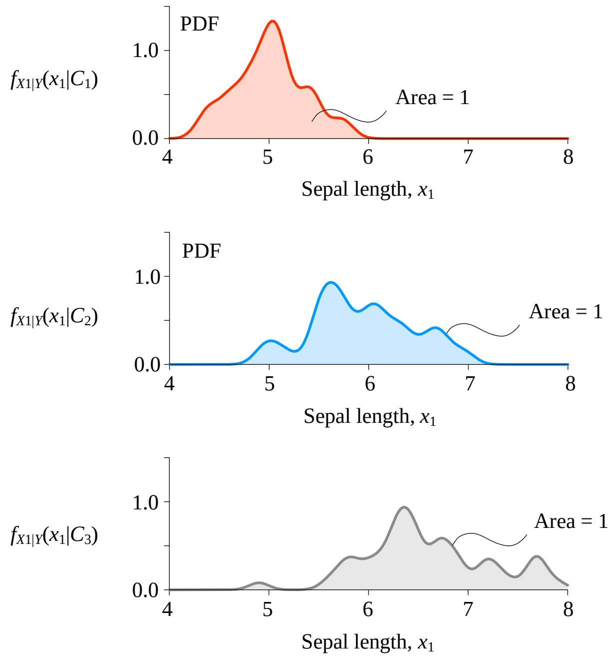

图 28. 给定鸢尾花标签 Y，花萼长度的条件**期望** $\mathrm{E}(X_1 \mid Y = C_k)$、条件**方差** $\mathrm{var}(X_1 \mid Y = C_k)$，连续随机变量

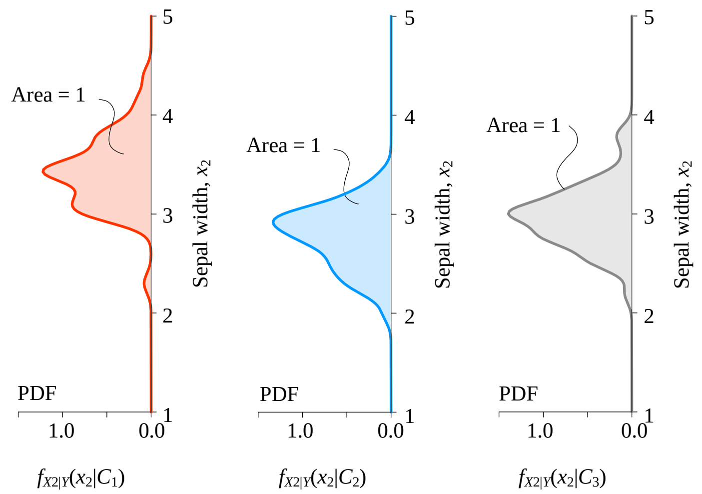

图 29. 给定鸢尾花标签 Y，花萼宽度的条件**期望** $\mathrm{E}(X_2 \mid Y = C_k)$、条件**方差** $\mathrm{var}(X_2 \mid Y = C_k)$，连续随机变量