

## 10

## Bivariate Gaussian Distribution

## 二元高斯分布

椭圆的影子几乎无处不在



自然之书是用数学语言写成的，符号是三角形、圆形和其他几何图形；不理解几何图形，别想读懂自然之书；没有它们，我们只能在黑暗的迷宫中徘徊不前。

*The book of nature is written in mathematical language, and the symbols are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth.*

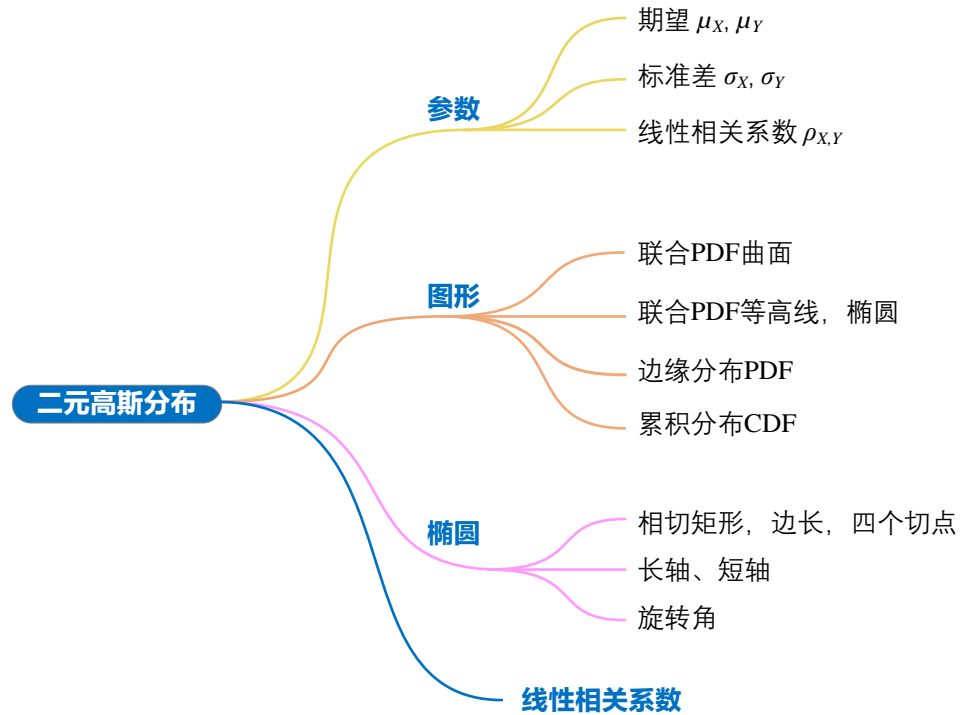
—— 伽利略·伽利莱 (Galilei Galileo) | 意大利物理学家、数学家及哲学家 | 1564 ~ 1642



```

< matplotlib.patches.Rectangle() 绘制长方形
< matplotlib.pyplot.axhline() 绘制水平线
< matplotlib.pyplot.axvline() 绘制垂直线
< matplotlib.pyplot.contour() 绘制等高线图
< matplotlib.pyplot.contourf() 绘制填充等高线图
< scipy.stats.multivariate_normal() 多元高斯分布
< scipy.stats.multivariate_normal.cdf() 多元高斯分布 CDF 函数
< scipy.stats.multivariate_normal.pdf() 多元高斯分布 PDF 函数

```



## 10.1 二元高斯分布：看见椭圆

### 概率密度函数

**二元高斯分布** (bivariate Gaussian distribution)，也叫**二元正态分布** (bivariate normal distribution)，它的概率密度函数  $f_{X,Y}(x,y)$  解析式如下：

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp \left( \underbrace{\frac{-1}{2(1-\rho_{X,Y}^2)} \left( \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{X,Y} \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right)}_{\text{Ellipse}} \right) \quad (1)$$

其中， $\mu_X$  和  $\mu_Y$  分别为随机变量  $X$ 、 $Y$  的期望值， $\sigma_X$  和  $\sigma_Y$  分别为随机变量  $X$ 、 $Y$  的标准差， $\rho_{X,Y}$  为  $X$  和  $Y$  线性相关系数。分母中，系数  $2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}$  完成归一化，也就是让  $f_{X,Y}(x,y)$  和水平面围成的体积为 1。

注意，观察 (1)，显然  $\rho_{X,Y}$  取值区间为  $(-1, 1)$ ，不能为  $\pm 1$ ；否则，分母为 0。

此外，丛书之前反复提到二元高斯分布和椭圆的关系。我们在 (1) 中已经看到了椭圆解析式。



(1) 中蕴含的椭圆解析式形式正是我们在《数学要素》第 9 章讲过的特殊类型。

### PDF 曲面形状

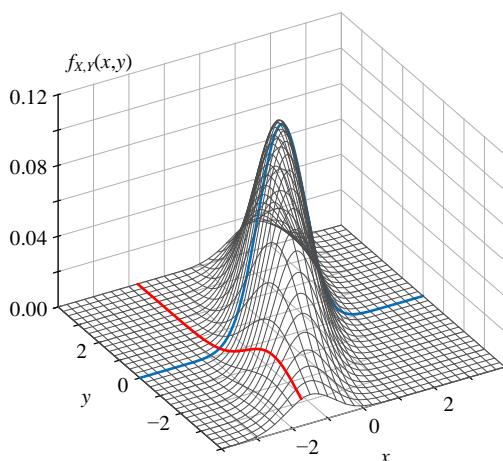
给定如下条件：

$$\mu_X = 0, \quad \mu_Y = 0, \quad \sigma_X = 1, \quad \sigma_Y = 2, \quad \rho_{X,Y} = 0.75 \quad (2)$$

绘制满足条件的二元正态分布密度函数曲面，具体如图 1 所示。

容易发现， $\mu_X$  和  $\mu_Y$  决定曲面中心所在位置； $\sigma_X$  和  $\sigma_Y$  影响曲面在  $x$  和  $y$  方向的形状。而  $\rho_{X,Y}$  似乎提供了曲面的扭曲。实际上， $\sigma_X$ 、 $\sigma_Y$ 、 $\rho_{X,Y}$  都影响了曲面的倾斜。

下面，我们从几个侧面来深入观察二元高斯分布 PDF  $f_{X,Y}(x,y)$  曲面。

图 1. 二元高斯分布 PDF 函数曲面  $f_{X,Y}(x,y)$ ,  $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$ 

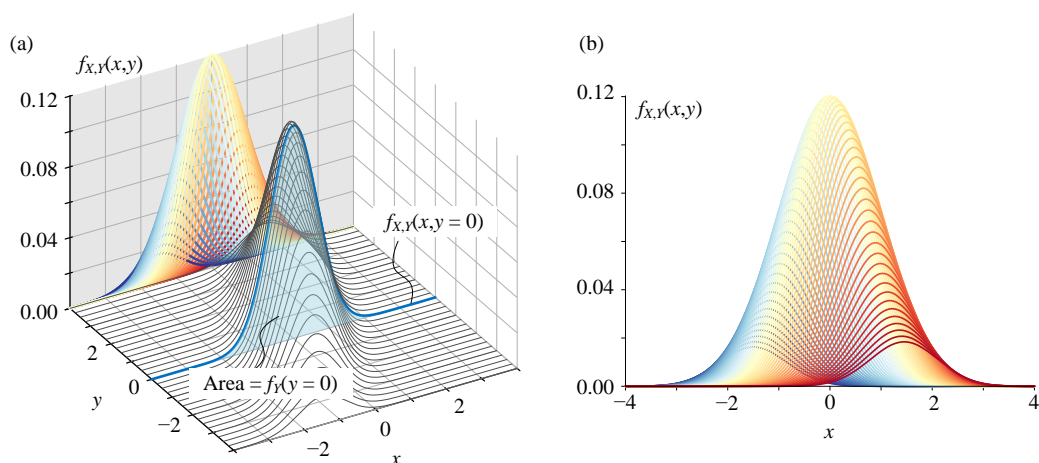
## 沿 $x$ 剖面线

图 2 所示为  $f_{X,Y}(x,y)$  曲面沿  $x$  方向的剖面线，以及这些曲线在  $xz$  平面上的投影。这些曲线，相当于是 (1) 中  $y$  取定值时 PDF 对应的曲线。比如  $y = 0$  时，曲线的解析式：

$$f_{X,Y}(x, y=0) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(-\frac{1}{2(1-\rho_{X,Y}^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \frac{2\rho_{X,Y}\mu_Y}{\sigma_Y}\left(\frac{x-\mu_X}{\sigma_X}\right) + \left(\frac{\mu_Y}{\sigma_Y}\right)^2\right)\right) \quad (3)$$

观察这条曲线，我们都能看到一元正态分布的影子。

注意，举个例子，图 2 (a) 中  $f_{X,Y}(x,y=0)$  这条曲线和横轴围成的图形面积并不为 1，面积对应边缘 PDF  $f_Y(y=0)$ 。因此图 2 (b) 中这些曲线虽然看起来像一元高斯分布 PDF，实际上并不是。但是经过一定的缩放，它们可以成为条件高斯分布的 PDF。

图 2. PDF 函数曲面  $f_{X,Y}(x,y)$ , 沿  $x$  方向的剖面线,  $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

大家试想一下，如果我们可以得到  $y = 0$  时边缘 PDF  $f_Y(y = 0)$  的具体值，就可以利用贝叶斯定理得到条件概率  $f_{X|Y}(x | y = 0)$ ：

$$f_{X|Y}(x | y = 0) = \frac{f_{X,Y}(x, y = 0)}{f_Y(y = 0)} \quad (4)$$

分母中的  $f_Y(y = 0)$  起到归一化的作用。而  $f_{X|Y}(x | y = 0)$  摇身一变成了条件高斯分布的 PDF。



这是本书第 12 章要讲解的内容。

## 沿 $y$ 剖面线

图 3 所示为  $f_{X,Y}(x,y)$  曲面沿  $y$  方向的剖面线，以及这些曲线在  $yz$  平面上的投影。曲线相当于  $x$  取定值，联合 PDF  $f_{X,Y}(x,y)$  随  $y$  变化。

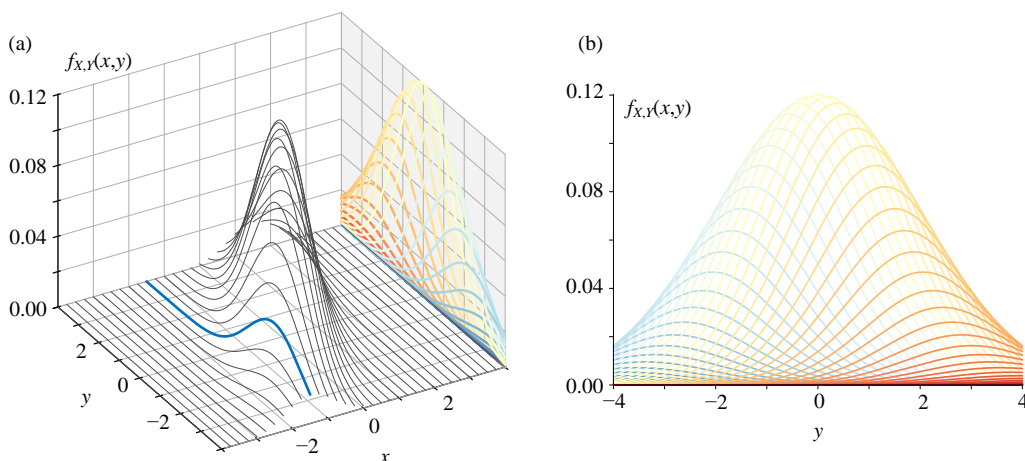
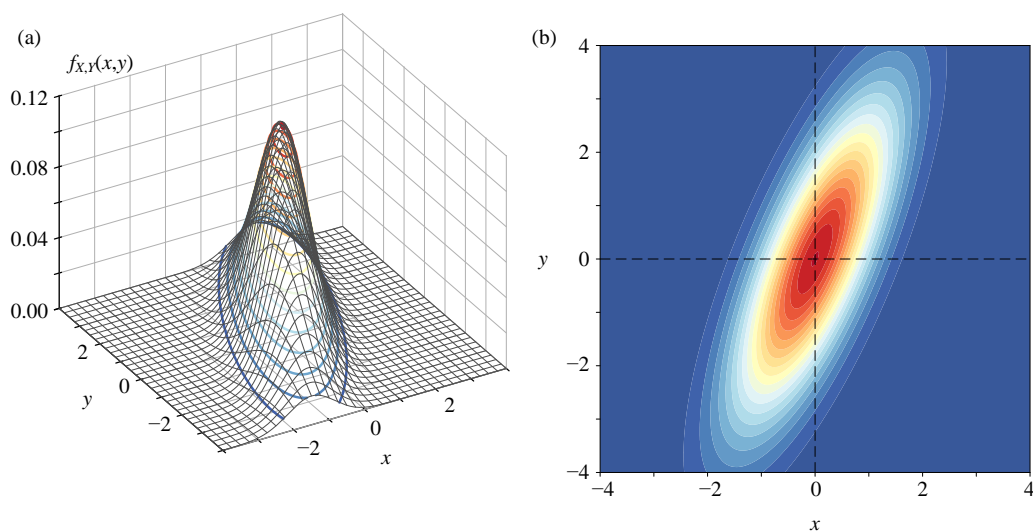


图 3. PDF 函数曲面  $f_{X,Y}(x,y)$ ，沿  $y$  方向的剖面线， $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

## 等高线

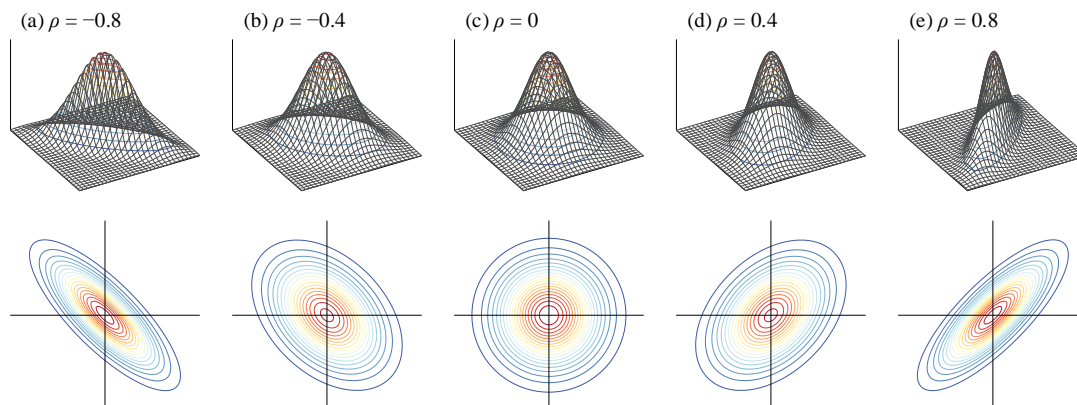
图 4 所示为  $f_{X,Y}(x,y)$  曲面等高线。很明显，我们已经从等高线中看到椭圆。特别是在图 4 (b) 中，我们看到一系列同心旋转椭圆。这并不奇怪，因为 (1) 中  $\exp()$  函数中蕴含着一个椭圆解析式。

这也就是为什么高斯分布被称作是一种**椭圆分布** (elliptical distribution)。本章后续将揭开高斯分布和椭圆的更多联系。

图 4. PDF 函数曲面  $f_{X,Y}(x,y)$ , 空间等高线和平面填充等高线,  $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$ 

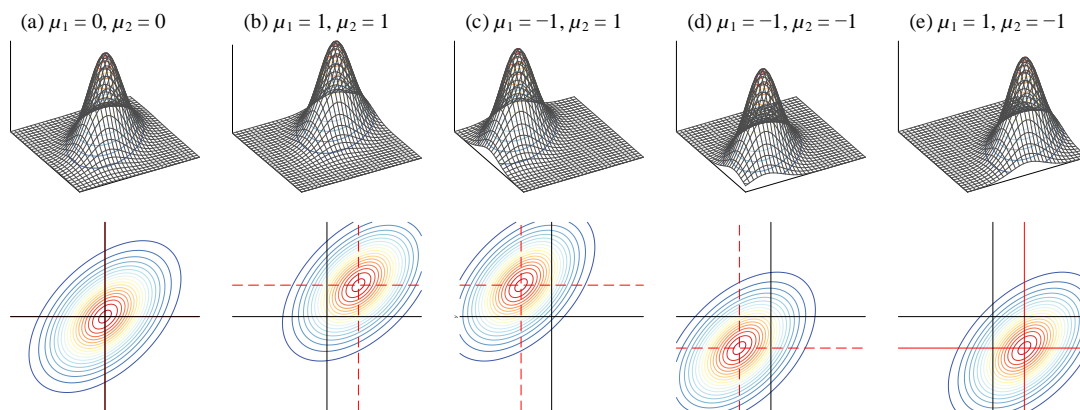
## 相关性系数

为了方便大家了解相关性系数对二元高斯分布 PDF 的影响, 设定  $\sigma_X = 1, \sigma_Y = 1$ 。如图 5 所示, 相关性系数对二元高斯分布 PDF 曲面和等高线形状的影响。

图 5. 不同相关性系数, 二元高斯分布 PDF 曲面和等高线,  $\sigma_X = 1, \sigma_Y = 1$ 

## 质心

如图 6 所示, 固定相关性系数和标准差, 改变质心仅仅影响曲面中心位置。

图 6. 不同质心位置，二元高斯分布 PDF 曲面和等高线， $\sigma_x = 1, \sigma_y = 1$ 

Bk5\_Ch10\_01.py 绘制本节图像。



在 Bk5\_Ch10\_01.py 基础上，我们用 Streamlit 制作了一个应用，大家可以改变  $\rho_{X,Y}$ 、 $\sigma_X$ 、 $\sigma_Y$  三个参数，观察二元高斯 PDF 曲面、等高线变化。请大家参考 Streamlit\_Bk5\_Ch10\_01.py。

## 10.2 边缘分布：一元高斯分布

### 边缘分布

大家可能也已经注意到，不考虑  $Y$  的时候， $X$  应该服从一元高斯分布。而  $\mu_X$  和  $\sigma_X$  是描述随机变量  $X$  的参数。也就是说，有了这两个参数，我们就可以写出  $X$  的边缘 PDF  $f_X(x)$  —— 一元高斯分布概率密度函数：

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu_X}{\sigma_X}\right)^2\right) \quad (5)$$

同理， $\mu_Y$  和  $\sigma_Y$  是描述随机变量  $Y$  的参数，对应写出  $Y$  的边缘 PDF  $f_Y(y)$ ：

$$f_Y(y) = \frac{1}{\sigma_Y \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2\right) \quad (6)$$

在图4平面等高线基础上添加  $f_X(x)$  和  $f_Y(y)$  边缘 PDF 图像子图，我们便得到图7。

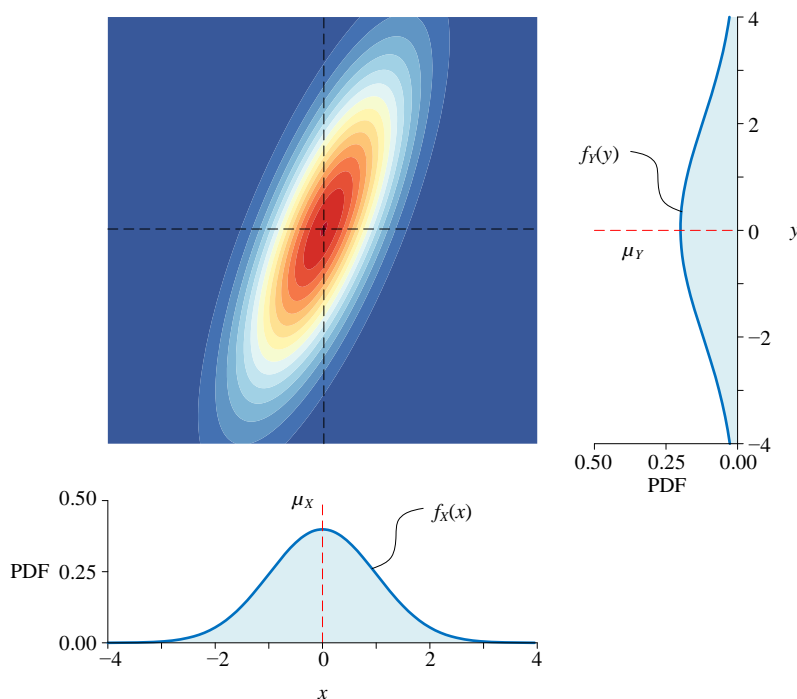


图7. 二元高斯分布 PDF 和边缘 PDF,  $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

## 偏积分求边缘分布 PDF

下面，以  $Y$  的边缘分布概率密度函数  $f_Y(y)$  为例证明二元高斯分布 PDF“偏积分”得到一元高斯分布 PDF。

连续随机变量  $Y$  的边缘分布概率密度函数  $f_Y(y)$  可以通过  $f_{X,Y}(x,y)$  对  $x$  偏积分得到，即：

$$f_Y(y) = \int_{-\infty}^{+\infty} \overbrace{f_{X,Y}(x,y)}^{\text{Eliminate } x} dx \quad (7)$$

令，

$$G(x,y) = \frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{X,Y}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{(1-\rho_{X,Y}^2)} \quad (8)$$

这样，二元高斯分布可以写成：

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(\frac{-1}{2}G(x,y)\right) \quad (9)$$



将 (8) 中  $G(x,y)$  写成：

$$G(x,y) = \frac{\left(\frac{x-\mu_X}{\sigma_X} - \rho_{X,Y} \frac{y-\mu_Y}{\sigma_Y}\right)^2}{(1-\rho_{X,Y}^2)} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$$

$$= \frac{\left(x - \left(\mu_X + \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} (y-\mu_Y)\right)\right)^2}{(1-\rho_{X,Y}^2)\sigma_X^2} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \quad (10)$$

令

$$t = t(y) = \mu_X + \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} (y - \mu_Y) \quad (11)$$

可以发现  $t$  仅仅是  $y$  的函数，与  $x$  无关，这样便于积分。

$G(x,y)$  进一步整理为：

$$G(x,y) = \frac{(x-t)^2}{(1-\rho_{X,Y}^2)\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \quad (12)$$

将 (12) 代入 (9) 得到：

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(\frac{-1}{2}\left(\frac{(x-t)^2}{(1-\rho_{X,Y}^2)\sigma_X^2}\right)\right) \times \exp\left(\frac{-1}{2}\left(\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right) \quad (13)$$

将 (13) 代入 (7) 得到：

$$f_Y(y) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(\frac{-1}{2}\left(\frac{(x-t)^2}{(1-\rho_{X,Y}^2)\sigma_X^2}\right)\right) \times \exp\left(\frac{-1}{2}\left(\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right) dx$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \cdot \exp\left(\frac{-1}{2}\left(\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right) \cdot \int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2}\left(\frac{(x-t)^2}{(\sqrt{1-\rho_{X,Y}^2})\sigma_X}\right)^2\right) dx \quad (14)$$

回忆，我们在《数学要素》讲解过高斯函数积分：

$$\int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2}\left(\frac{(x-t)^2}{(\sqrt{1-\rho_{X,Y}^2})\sigma_X}\right)^2\right) dx = \sqrt{2\pi}\sqrt{1-\rho_{X,Y}^2}\sigma_X \quad (15)$$

将 (15) 代入 (14)，得到：

$$\begin{aligned}
 f_Y(y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \cdot \exp\left(\frac{-1}{2} \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right) \sqrt{2\pi} \sqrt{(1-\rho_{X,Y}^2)} \sigma_X \\
 &= \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(\frac{-1}{2} \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)
 \end{aligned} \tag{16}$$

再次强调，联合 PDF  $f_{X,Y}(x,y)$  二重积分得到的是概率，也就是曲面体积代表概率。而  $f_{X,Y}(x,y)$  偏积分得到的还是概率密度，即边缘概率密度  $f_X(x)$  或  $f_Y(y)$ 。边缘 PDF  $f_X(x)$  和  $f_Y(y)$  进一步积分才得到概率。

## 独立

图 8 所示为二元高斯分布参数对 PDF 等高线影响。

特别地，当相关性系数  $\rho_{X,Y}$  为 0 时：

$$\begin{aligned}
 f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y} \times \exp\left(\frac{-1}{2} \left( \left( \frac{x-\mu_X}{\sigma_X} \right)^2 + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right)\right) \\
 &= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(\frac{-1}{2} \left( \frac{x-\mu_X}{\sigma_X} \right)^2\right) \times \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(\frac{-1}{2} \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2\right) \\
 &= f_X(x) f_Y(y)
 \end{aligned} \tag{17}$$

观察图 8 (b)、(e)、(h)，我们发现椭圆等高线为正椭圆。

注意，独立意味着两个变量的取值之间没有任何关系，即它们的联合概率分布等于它们的边缘概率分布的乘积。而相关则表示两个变量之间存在某种形式的关联关系，可以是线性的，也可以是非线性的。因此，线性相关系数为 0 只是说明两个变量之间不存在线性关系，但并不能推断它们是否独立。

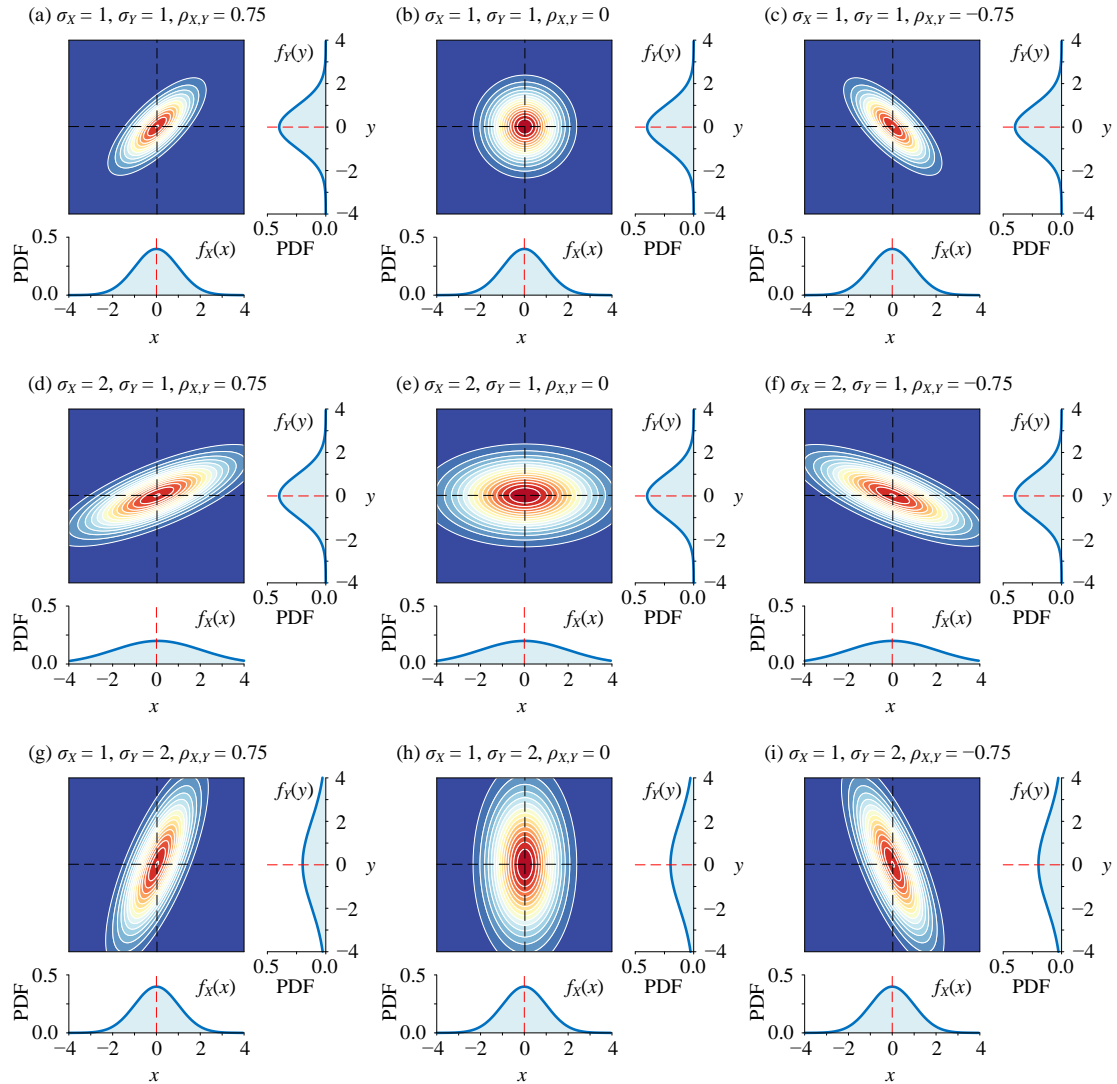
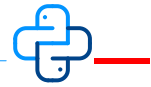


图 8. 二元高斯分布参数对 PDF 等高线影响



Bk5\_Ch10\_02.py 绘制本节图像。请大家自行调整分布参数。

## 10.3 累积分布函数：概率值

二元高斯分布的累积分布函数 CDF  $F_{X,Y}(x,y)$  是对 PDF  $f_{X,Y}(x,y)$  的二重积分：

$$F_{X,Y}(x,y) = \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(s,t) ds dt \quad (18)$$

图 9 所示为二元高斯分布累积分布函数 CDF 曲面。

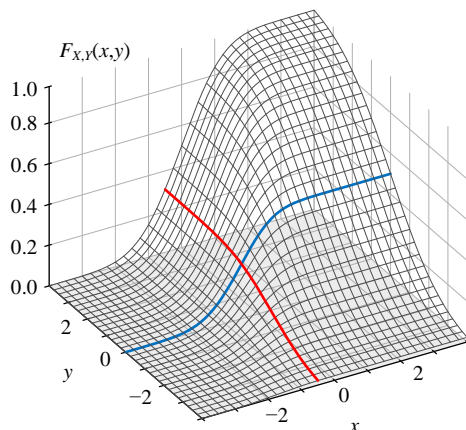
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

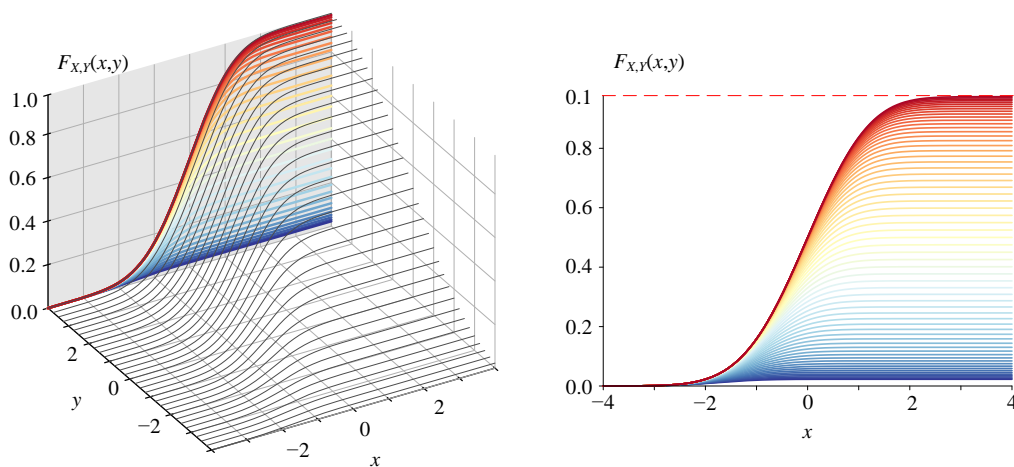
本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：[jiang.visualize.ml@gmail.com](mailto:jiang.visualize.ml@gmail.com)

图 9. 二元高斯分布累积函数 CDF 曲面,  $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$ 

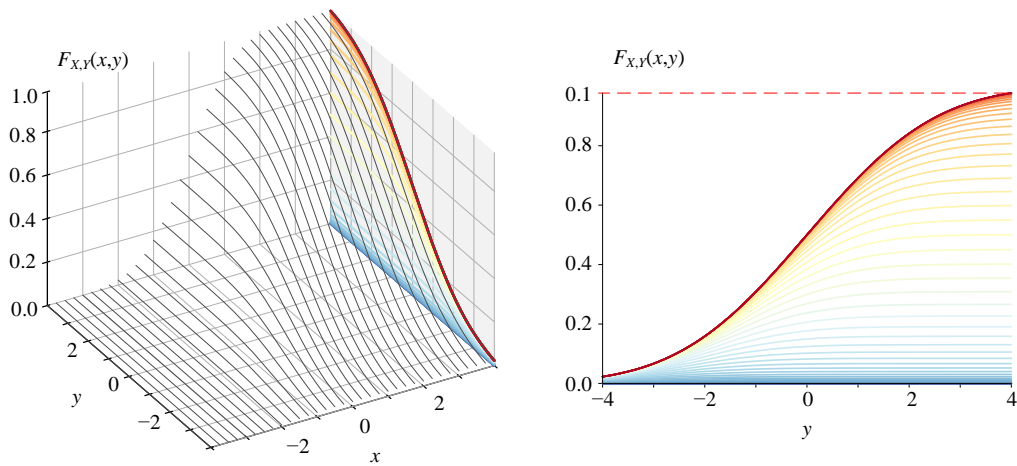
### 沿 $x$ 剖面线

和上一节一样，下面从几个侧面来观察二元高斯分布 CDF 曲面  $F_{X,Y}(x,y)$ 。图 10 所示为  $F_{X,Y}(x,y)$  曲面沿  $x$  方向的剖面线，以及这些曲线在  $xz$  平面上的投影。

图 10. CDF 曲面  $F_{X,Y}(x,y)$ , 沿  $x$  方向的剖面线,  $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$ 

### 沿 $y$ 剖面线

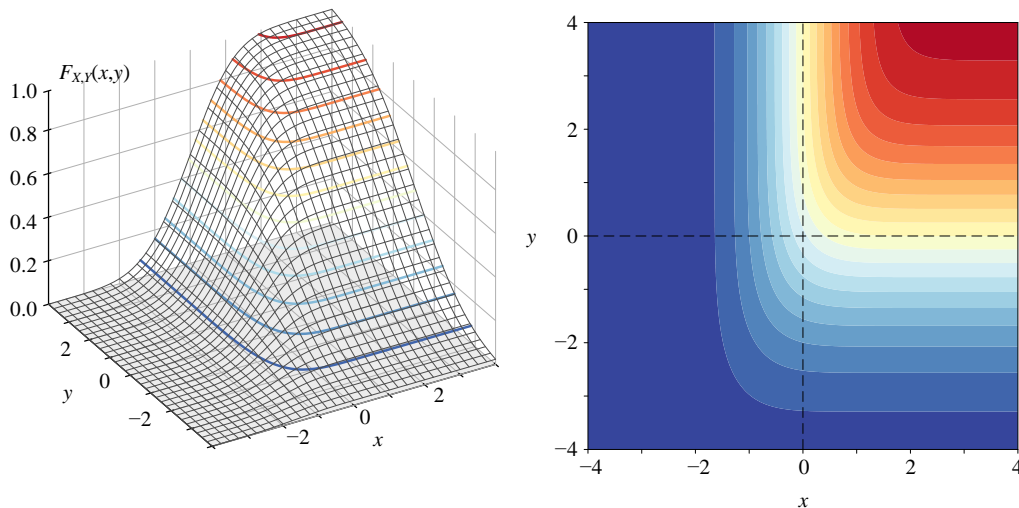
图 11 所示为  $F_{X,Y}(x,y)$  曲面沿  $y$  方向的剖面线，以及这些曲线在  $yz$  平面上的投影。

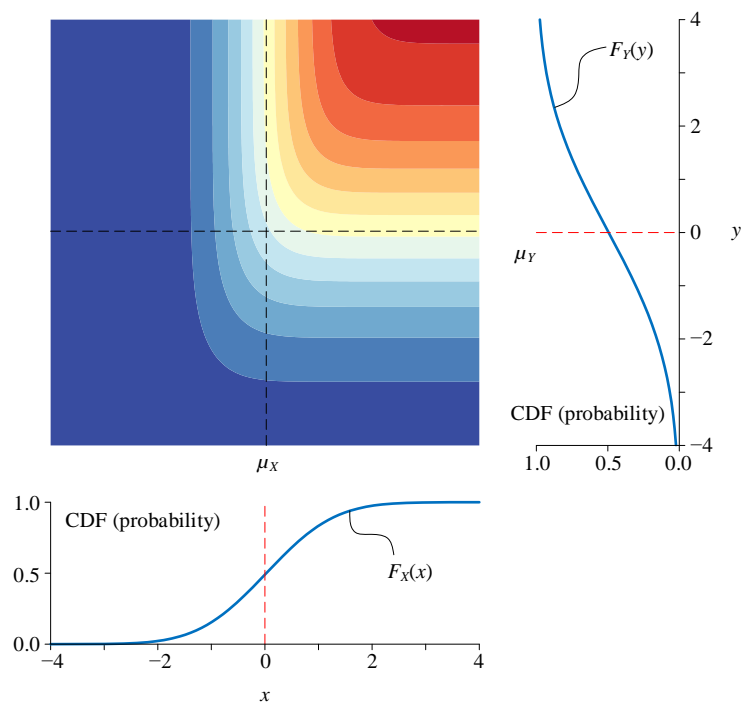
图 11. CDF 曲面  $F_{X,Y}(x,y)$ , 沿  $y$  方向的剖面线,  $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$ 

## 等高线

图 12 所示为 CDF 函数曲面  $F_{X,Y}(x,y)$  的等高线。图 13 所示为, 在  $F_{X,Y}(x,y)$  的平面填充等高线基础上, 又绘制了边缘 CDF  $F_X(x)$ 、 $F_Y(y)$  曲线。

请大家修改上一节代码绘制本节图像。只需要把 `scipy.stats.multivariate_normal.pdf()` 换成 `scipy.stats.multivariate_normal.cdf()` 函数。

图 12. CDF 函数曲面  $F_{X,Y}(x,y)$ , 空间等高线和平面填充等高线,  $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

图 13. CDF 函数曲面  $F_{X,Y}(x,y)$  平面填充等高线，边缘概率分布 CDF

## 10.4 用椭圆解剖二元高斯分布

大家已经在 (1) 看到了椭圆的解析式，这一节我们对二元高斯分布和椭圆的关系进行定量研究。

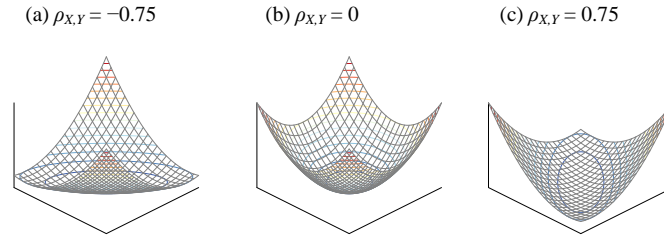
### 二次曲面

利用 (8) 中定义的  $G(x,y)$ 。将 (8) 代入 (1)，得到：

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(\frac{-1}{2}G(x,y)\right) \quad (19)$$

图 14 所示为  $G(x,y)$  代表的几种曲面。

但是，对于二元高斯分布来说，如果 PDF 解析式存在，相关性的取值范围为  $(-1, 1)$ ，此时协方差矩阵为正定。请大家思考如果，协方差矩阵为半正定， $G(x,y)$  曲面的形状是什么？

图 14.  $G(x, y)$  代表的几种曲面

## 椭圆

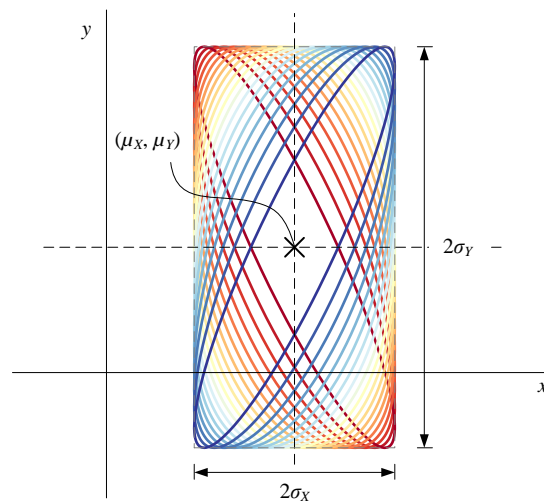
令  $G(x, y) = 1$ ，当  $\rho_{X,Y}$  在  $(-1, 1)$  变化时，我们便得到椭圆的解析式：

$$\frac{1}{(1-\rho_{X,Y}^2)} \left( \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{X,Y} \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right) = 1 \quad (20)$$

$(\mu_1, \mu_2)$  确定椭圆中心位置， $\sigma_1$ 、 $\sigma_2$  和  $\rho$  三者共同决定椭圆长短轴长度和旋转角度。

“鸢尾花书”《数学要素》第 9 章介绍过，形如 (20) 解析式的椭圆有重要的特点——椭圆和长  $2\sigma_X$ 、宽  $2\sigma_Y$  的矩形相切。

图 15 给出的矩形框中心位于  $(\mu_X, \mu_Y)$ ，矩形框长度为  $2\sigma_X = 2$ 、宽度为  $2\sigma_Y = 4$ 。图 15 中一系列椭圆对应的相关性系数  $\rho_{X,Y}$  的变化范围为  $[-0.9, 0.9]$ 。

图 15. 椭圆和中心在  $(\mu_X, \mu_Y)$  长  $2\sigma_X$ 、宽  $2\sigma_Y$  的矩形相切

相关性系数  $\rho$  大于 0 时，即线性正相关，椭圆长轴指向约东北方向。线性相关性系数  $\rho$  小于 0 时，即负相关，椭圆长轴指向约西北方向；特别提醒读者注意的是，当相关性系数  $\rho$  为 0 时，椭圆为正椭圆。

图 16 所示为三种标准差  $\sigma_X$ 、 $\sigma_Y$  大小不同的情况，和矩形相切的椭圆随着相关性系数变化情况。

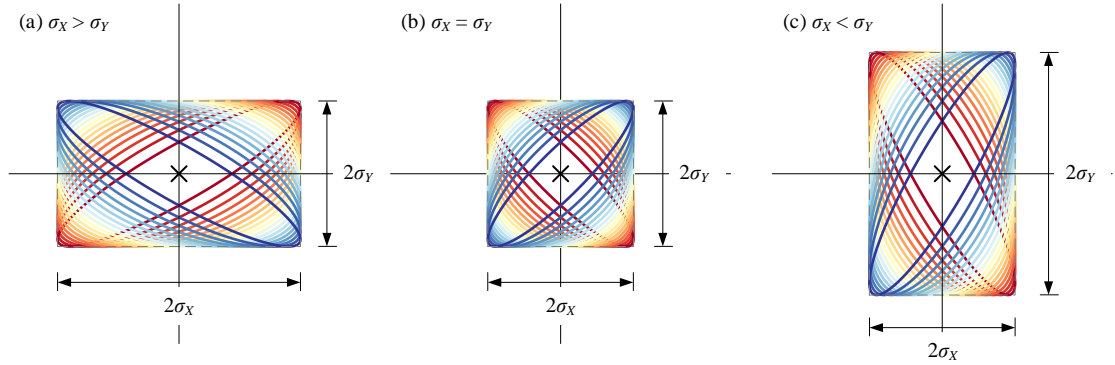


图 16. 三种标准差  $\sigma_X$ 、 $\sigma_Y$  大小不同的情况

#### 四个切点

椭圆和矩形有四个切点，下面我们来求解这四个切点的具体位置。考虑特殊情况  $\mu_X = 0, \mu_Y = 0$ ，(20) 可以简化为：

$$\frac{1}{(1-\rho_{X,Y}^2)} \left( \left( \frac{x}{\sigma_X} \right)^2 - \frac{2\rho_{X,Y}}{\sigma_X\sigma_Y} xy + \left( \frac{y}{\sigma_Y} \right)^2 \right) = 1 \quad (21)$$

将  $y = \sigma_Y$  代入 (21)，得到：

$$\left( \frac{x}{\sigma_X} - \rho_{X,Y} \right)^2 = 0 \quad (22)$$

这样我们便得到一个切点：

$$\begin{cases} x = \rho_{X,Y}\sigma_X \\ y = \sigma_Y \end{cases} \quad (23)$$

同理，获得所有四个切点  $A$ 、 $B$ 、 $C$ 、 $D$  的具体位置：

$$A(\rho_{X,Y}\sigma_X, \sigma_Y), B(\sigma_X, \rho_{X,Y}\sigma_Y), C(-\rho_{X,Y}\sigma_X, -\sigma_Y), D(-\sigma_X, -\rho_{X,Y}\sigma_Y) \quad (24)$$



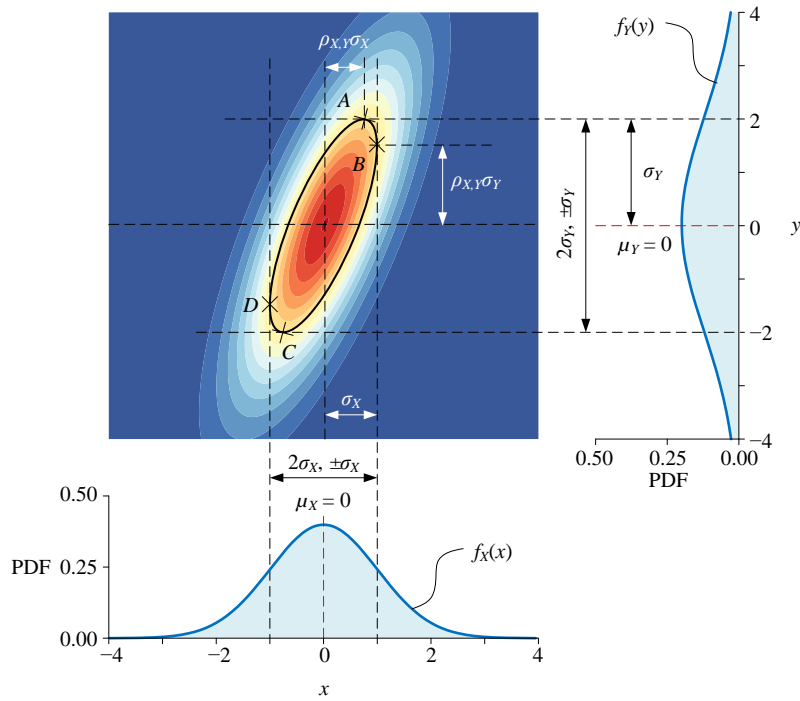


图 17. 二元高斯分布 PDF 和边缘 PDF,  $\sigma_x = 1, \sigma_y = 2, \rho_{x,y} = 0.75$

## 椭圆和矩形

$\mu_x$  和  $\mu_y$  均不为 0 的一般情况，四个切点的位置平移  $(\mu_x, \mu_y)$ ，为：

$$\begin{aligned} A(\mu_x + \rho_{x,y}\sigma_x, \mu_y + \sigma_y), \quad B(\mu_x + \sigma_x, \mu_y + \rho_{x,y}\sigma_y), \\ C(\mu_x - \rho_{x,y}\sigma_x, \mu_y - \sigma_y), \quad D(\mu_x - \sigma_x, \mu_y - \rho_{x,y}\sigma_y) \end{aligned} \quad (25)$$

图 18 所示为椭圆和矩形切点位置随  $\sigma_x$ 、 $\sigma_y$ 、 $\rho_{x,y}$  变化关系，请大家自行总结规律。

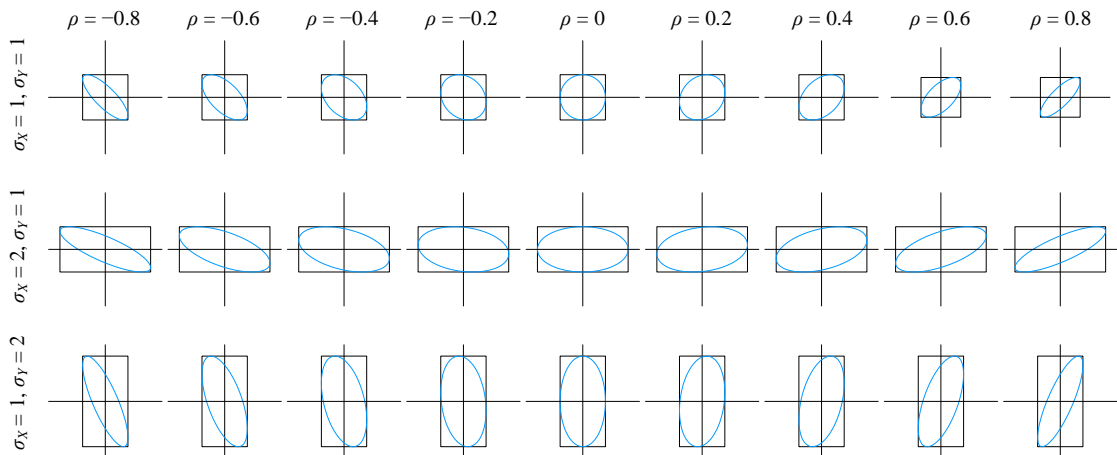
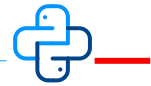


图 18. 椭圆和矩形切点随  $\sigma_x$ 、 $\sigma_y$ 、 $\rho_{x,y}$  变化关系



Bk5\_Ch10\_03.py 绘制图 18。

## 椭圆形状

再怎么强调椭圆和高斯分布的紧密联系也不为过。图 19 这个旋转椭圆的位置、形状、旋转角度等信息，蕴含着高斯分布的中心  $(\mu_X, \mu_Y)$ 、标准差  $\sigma_X$  和  $\sigma_Y$ 、相关性系数  $\rho_{X,Y}$ 。也就是说，某个二元高斯分布可以用特定椭圆来代表。

图 19 中还有很多椭圆相关的性质值得我们挖掘。

图 19 所示两个椭圆，蓝色椭圆上所有点到代入 (8) 都等于 1，类似一元高斯分布中的  $\mu \pm \sigma$ 。而更大一点的红色椭圆所有点代入 (8) 都等于 4，平方根为 2，类似一元高斯分布中的  $\mu \pm 2\sigma$ 。

上面所述的平方根 (1、2) 正是《矩阵力量》第 20 章讲过马氏距离。本章后文将稍微回顾马氏距离，本书第 23 章还要深入讲解马氏距离。

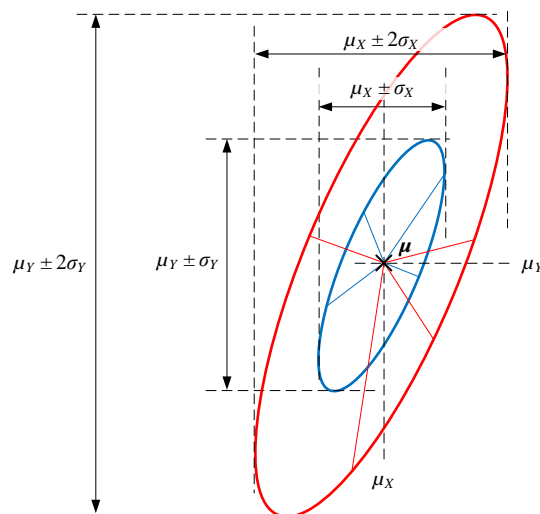


图 19. 两个椭圆

图 20 中的浅蓝色直角三角形的两条直角边长度分别是  $\rho_{X,Y}\sigma_Y$ 、 $\sigma_X$ ，其中  $\theta$  角的正切值为：

$$\tan \theta = \frac{\rho_{X,Y}\sigma_Y}{\sigma_X} \quad (26)$$

图 20 所示  $AC$  线段、 $BD$  线段和条件概率、线性回归有着直接联系。本书第 12 章将专门讲解高斯分布条件概率。



图 20 中两条红色线为椭圆的长轴和短轴所在方向，这两条直线又和主成分分析有着密切的关系。这是本书第 14、25 章要探讨的内容。

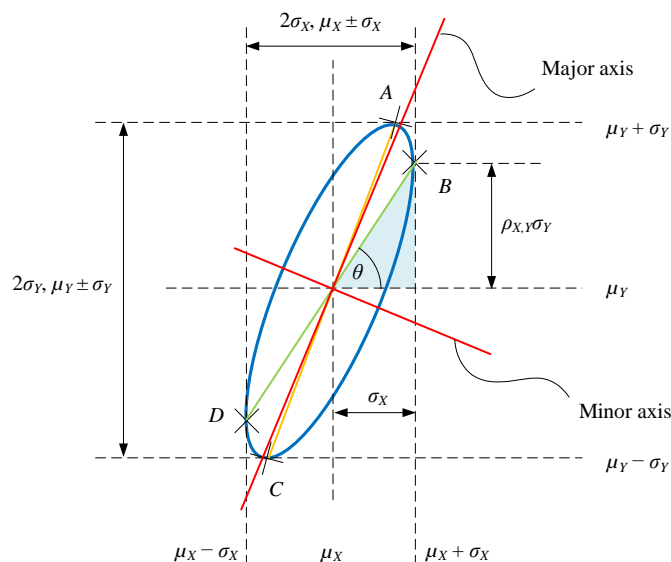


图 20. 椭圆中的四条直线

## 10.5 聊聊线性相关性系数

### 几种可视化方案

图 21 所示为相关性系数的几种可视化方案，比如散点图、二元高斯 PDF 曲面、PDF 等高线、条件概率直线、向量夹角。



大家应该在《矩阵力量》第 23 章见过图 21，当时我们特别讨论了利用向量可视化线性相关性系数。

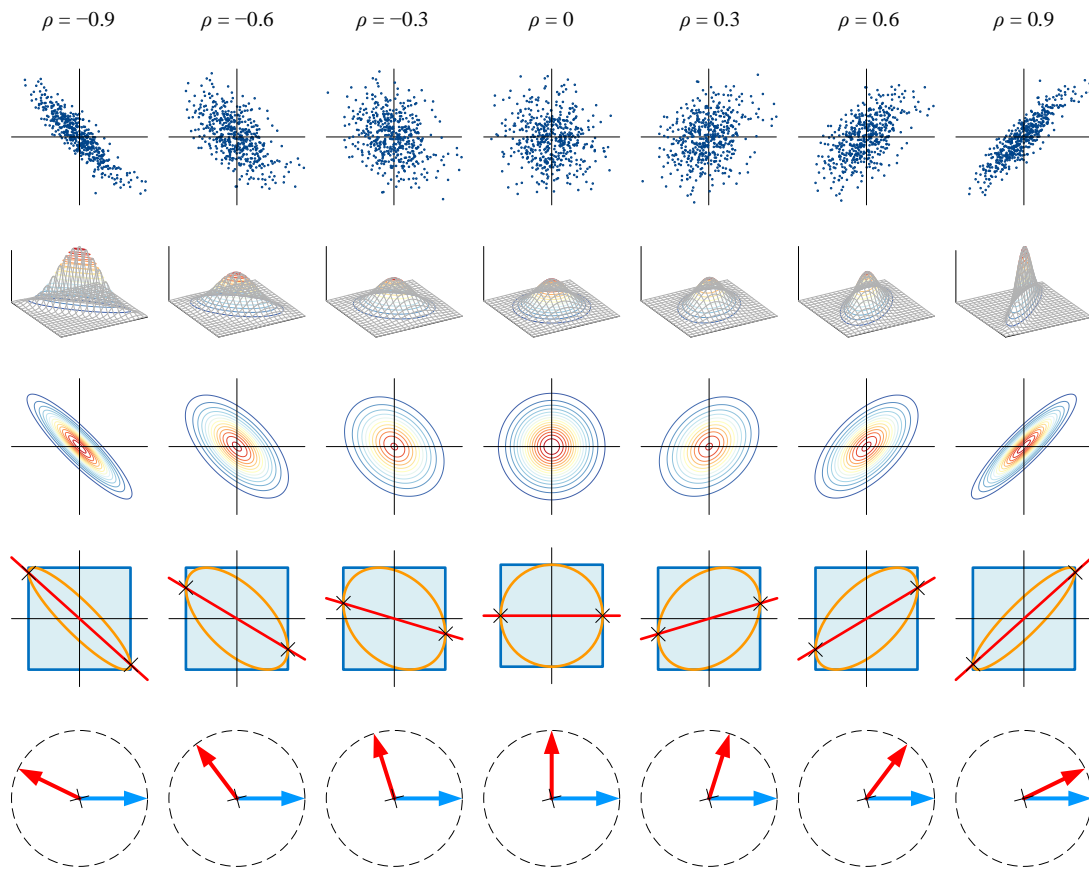
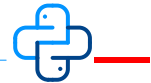


图 21. 相关性系数的几种可视化方案



Bk5\_Ch10\_04.py 可以绘制图 21 大部分图像。请大家自行修改参数。

### 独立 vs 线性相关性系数为 0

本章前文提过，线性相关系数反映的是两个随机变量间的线性关系，但是随机变量之间除了线性关系还可能存在其它关系。

举个例子，随机变量  $X$  在  $[-1, 1]$  连续均匀分布。令  $Y = X^2$ ，显然， $X$  和  $Y$  存在二次关系，并不独立。但是两者的协方差为 0：

$$\begin{aligned}
 \text{cov}(X, Y) &= \text{cov}(X, X^2) \\
 &= E[X \cdot X^2] - E[X] \cdot E[X^2] \\
 &= E[X^3] - E[X]E[X^2] \\
 &= 0 - 0 \cdot E[X^2] = 0
 \end{aligned}
 \tag{27}$$

这意味着的线性相关性系数为 0。

### 安斯库姆四重奏

图 22 是**安斯库姆四重奏** (Anscombe's quartet) 的四组散点图。观察图中四组散点图，我们可以发现数据的关系完全不同。但是，它们的相关性系数几乎完全一致。

这幅图告诉我们，线性相关性系数不是万能的，它只适合度量随机变量之间的“线性关系”。此外，线性相关性系数特别容易受到**离群值** (outlier) 的影响，这一点可以从图 22 (c) 看出来。

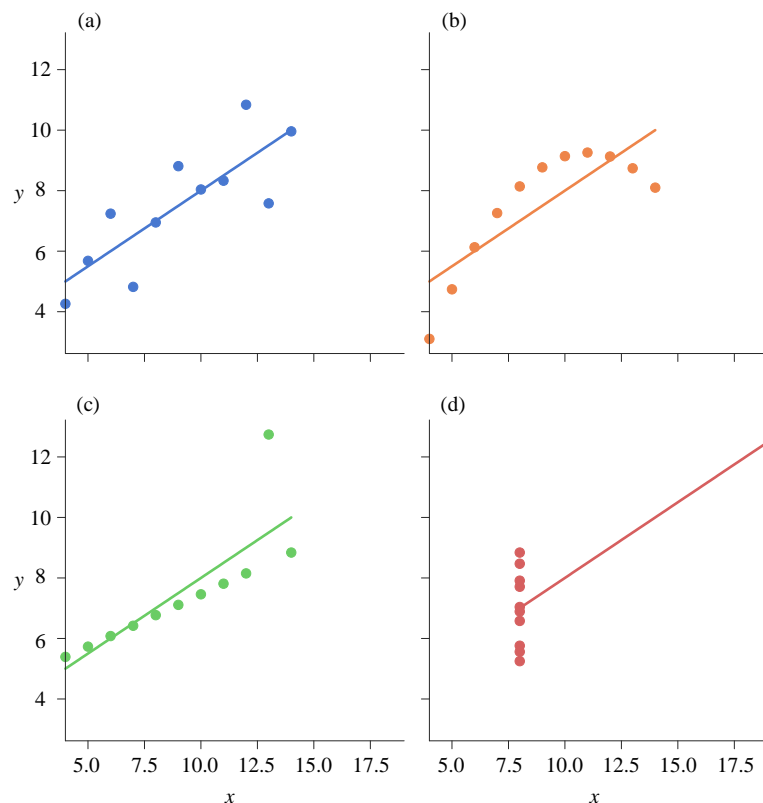


图 22. 安斯库姆四重奏

### 向量空间：线性无关

我们在《矩阵力量》第 7 章中介绍过，给定向量组  $V = [v_1, v_2, \dots, v_D]$ ，如果存在不全为零  $\alpha_1, \alpha_2, \dots, \alpha_D$  使得下式成立。

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \dots + \alpha_D v_D = \mathbf{0} \quad (28)$$

则称向量组  $V$  **线性相关** (linear dependence)；否则， $V$  **线性无关** (linear independence)。请大家注意区分。

## 正交 vs 线性相关性系数为 0

随机变量  $X$  和  $Y$  的协方差可以通过下式计算得到：

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) \quad (29)$$

如果  $X$  和  $Y$  独立，则

$$\text{cov}(X, Y) = 0 \quad (30)$$

这意味着：

$$E(XY) = E(X)E(Y) \quad (31)$$

本书前文提过，随机变量  $X$  和  $Y$  的有序样本集合看做是向量  $\mathbf{x}$  和  $\mathbf{y}$ 。如果向量  $\mathbf{x}$  和  $\mathbf{y}$  内积为 0，这意味着  $\mathbf{x}$  和  $\mathbf{y}$  **正交** (orthogonal)，这对应  $E(XY) = 0$ 。

## 相关性系数的变化

线性相关性系数受到具体样本数据选取的影响。如图 23 所示，对于鸢尾花所有 150 个样本点，花萼长度、花萼宽度的线性相关性系数小于 0。但是，分别计算三个不同标签的数据的花萼长度、花萼宽度的线性相关性系数，发现这三个值都显著大于 0。

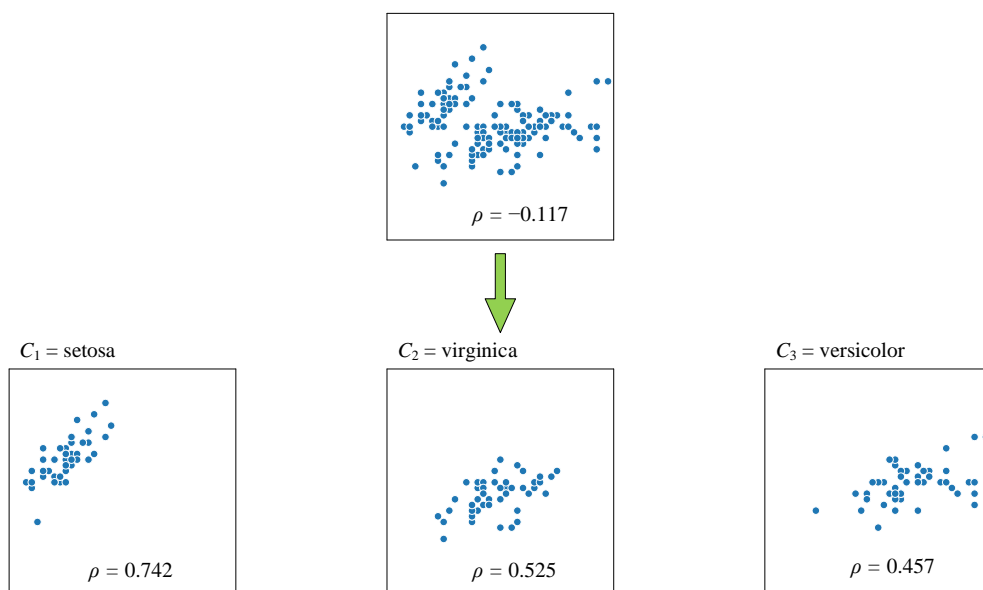


图 23. 鸢尾花不同分类的线性相关性系数

大家将会在《数据有道》一册中看到，如图 24 所示，时间序列数据的相关性系数还会随时间窗口变化。图 24 中，大家看到相关性系数出现陡然上升或下降 (高亮) 的情况，这可能都是由几个样本点带来的影响，值得深入研究。

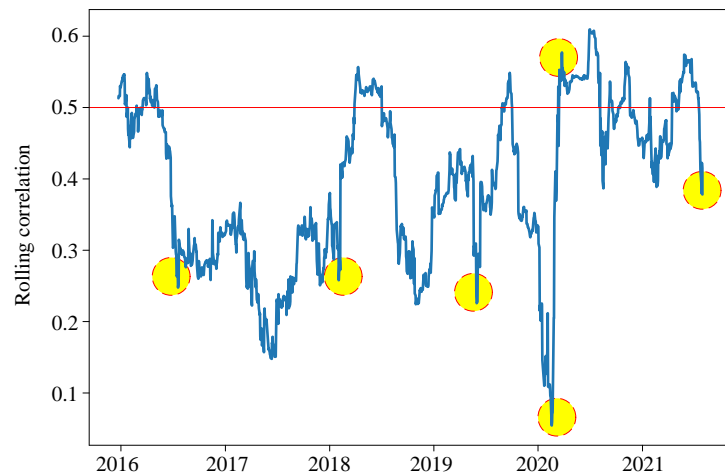


图 24. 移动线性相关性系数

## 10.6 以鸢尾花数据为例：不考虑分类标签

本节和下一节用二元高斯分布估计鸢尾花花萼长度  $X_1$ 、花萼宽度  $X_2$  的联合概率密度函数  $f_{X_1, X_2}(x_1, x_2)$ 。相信大家还记得我们在本书第 7 章采用 KDE 估计联合概率密度函数  $f_{X_1, X_2}(x_1, x_2)$ 。这两节采用本书和 7 章类似的结构，方便大家比较阅读。

### 二元高斯分布 → 联合概率密度函数 $f_{X_1, X_2}(x_1, x_2)$

假设  $(X_1, X_2)$  服从二元高斯分布：

$$(X_1, X_2) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (32)$$

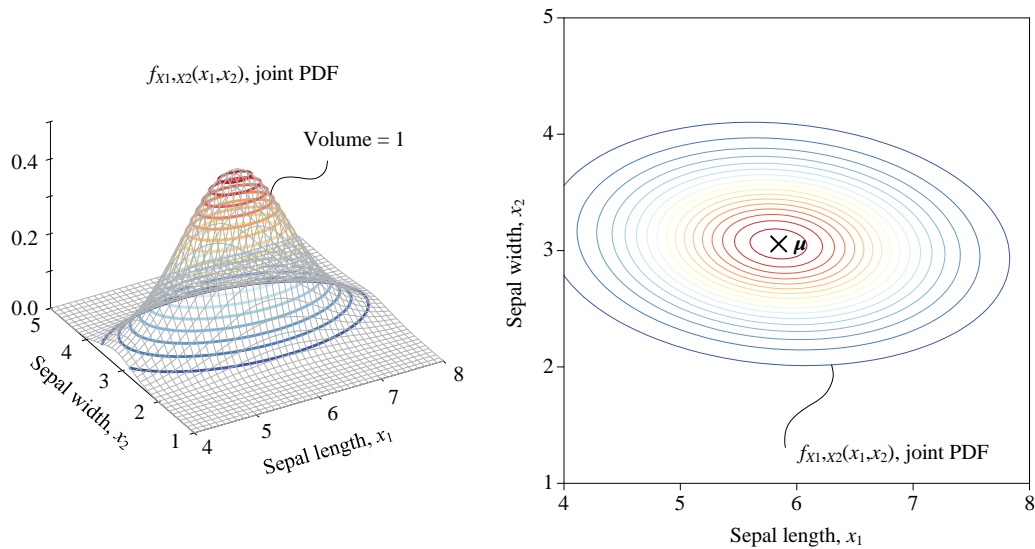
利用鸢尾花 150 个样本数据，我们可以估算得到  $(X_1, X_2)$  的质心和协方差矩阵分别为：

$$\boldsymbol{\mu} = \begin{bmatrix} 5.843 \\ 3.057 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.685 & -0.042 \\ -0.042 & 0.189 \end{bmatrix} \quad (33)$$

$(X_1, X_2)$  的联合概率密度函数  $f_{X_1, X_2}(x_1, x_2)$  解析式则为：

$$f_{X_1, X_2}(x_1, x_2) \approx \frac{\exp \left[ \frac{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{-0.739x_1^2 - 0.33x_1x_2 - 2.668x_2^2 + 9.651x_1 + 18.248x_2 - 56.093} \right]}{2\pi \times 0.358 \sqrt{|\boldsymbol{\Sigma}|}} \quad (34)$$

图 25 所示为假设  $(X_1, X_2)$  服从二元高斯分布时，联合概率密度函数  $f_{X_1, X_2}(x_1, x_2)$  的三维等高线和平面等高线。

图 25.  $f_{X1,X2}(x_1, x_2)$  联合概率密度三维等高线和平面等高线，不考虑分类

举个例子，花萼长度 ( $X_1$ ) 为 6.5、花萼宽度 ( $X_2$ ) 为 2.0 时，利用 (34) 估计得到联合概率密度值为：

$$f_{X1,X2}(x_1 = 6.5, x_2 = 2.0) \approx 0.0205 \quad (35)$$

注意，这个数值是概率密度，不是概率。但是这个值某种程度上也代表可能性。

### 马氏距离椭圆的性质

《矩阵力量》第 20 章介绍过**马氏距离** (Mahalanobis distance 或 Mahal distance)，具体定义为：

$$d = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \quad (36)$$

图 26 所示为基于鸢尾花花萼长度、花萼宽度样本数据的马氏距离椭圆。图中，黑色旋转椭圆分别代表马氏距离为 1、2、3、4。图中，还有一个  $\boldsymbol{\mu}_1 \pm \sigma_1$  和  $\boldsymbol{\mu}_2 \pm \sigma_2$  构成的矩形。根据本章前文所学，我们知道马氏距离为 1 椭圆和矩形相切于四个点。



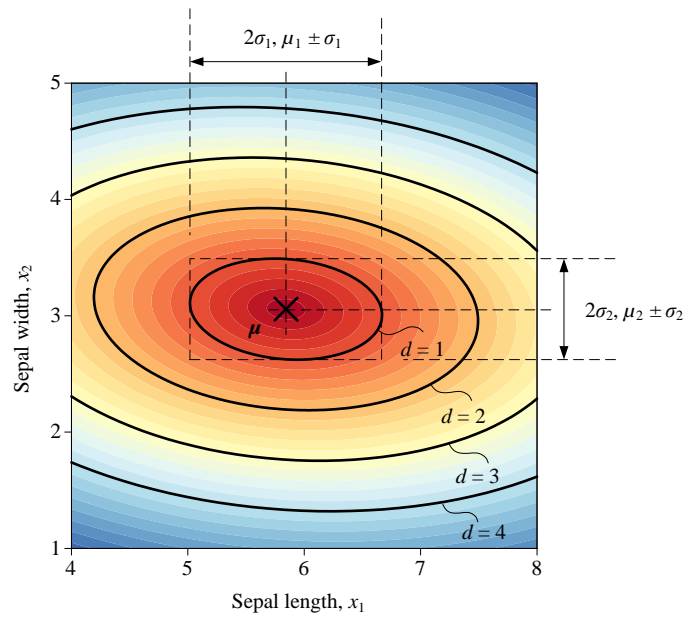


图 26. 马氏距离的椭圆，鸢尾花花萼长度、花萼宽度样本数据

还有一个需要大家注意的矩形。如图 27 所示，这个矩形和马氏距离为 1 椭圆同样相切，但是它的长边平行于椭圆的长轴。请大家自行计算椭圆长轴倾斜角。

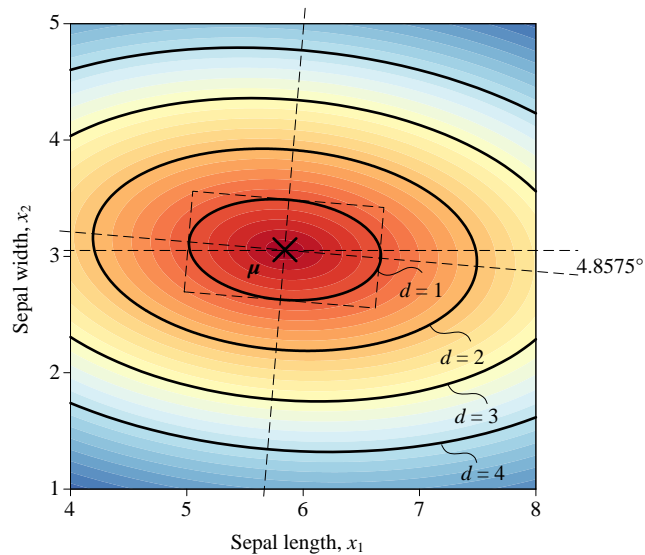


图 27. 马氏距离的椭圆的长轴、短轴，以及对应矩形

我们已经知道马氏距离和概率密度之间的关系为：

$$f_{X_1, X_2}(x_1, x_2) = \frac{\exp\left(-\frac{1}{2}d^2\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \quad (37)$$

对于 (34)，当  $d = 1$  时：

$$f_{X_1, X_2}(x_1, x_2)|_{d=1} = \frac{\exp\left(-\frac{1}{2} \times 1^2\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \approx 0.2693 \quad (38)$$

当  $d = 2$  时：

$$f_{X_1, X_2}(x_1, x_2)|_{d=2} = \frac{\exp\left(-\frac{1}{2} \times 2^2\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \approx 0.0601 \quad (39)$$

如图 28 所示，利用二重积分，我们可以计算两幅子图中阴影区域对应的概率：

$$\iint_D f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \quad (40)$$

从概率统计角度来看，阴影区域有什么意义？这个问题的答案留到本书第 23 章回答。

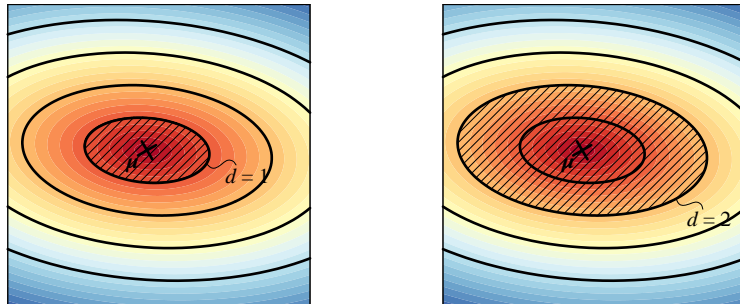
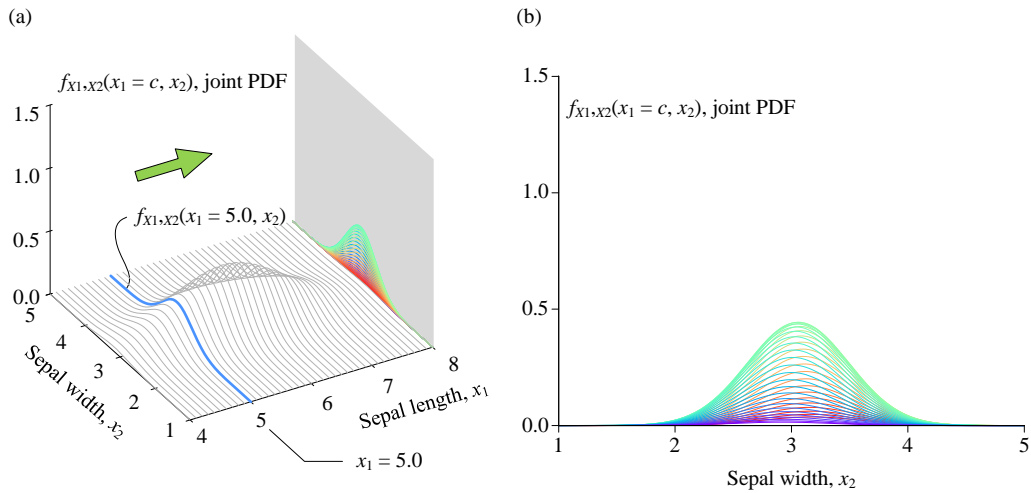
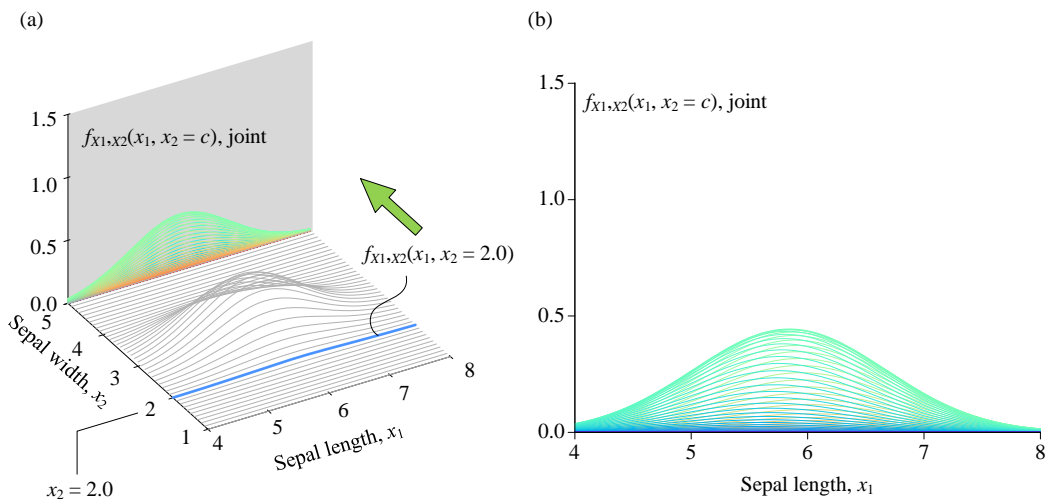


图 28. 求阴影区域对应的概率

### 联合概率密度函数 $f_{X_1, X_2}(x_1, x_2)$ 的剖面线

$f_{X_1, X_2}(x_1, x_2)$  本质上是个二元函数，因此我们还可以使用“剖面线”分析二元函数。

当固定  $x_1$  取值时， $f_{X_1, X_2}(x_1 = c, x_2)$  代表一条曲线。将一系列类似曲线投影到竖直平面得到图 29 (b)。观察图 29 (b)，我们容易发现这些曲线都类似一元高斯分布。图 30 所示为固定  $x_2$  时，概率密度函数  $f_{X_1, X_2}(x_1, x_2)$  随  $x_1$  变化。

图 29. 固定  $x_1$  时，概率密度函数  $f_{X1,X2}(x_1, x_2)$  随  $x_2$  变化图 30. 固定  $x_2$  时，概率密度函数  $f_{X1,X2}(x_1, x_2)$  随  $x_1$  变化

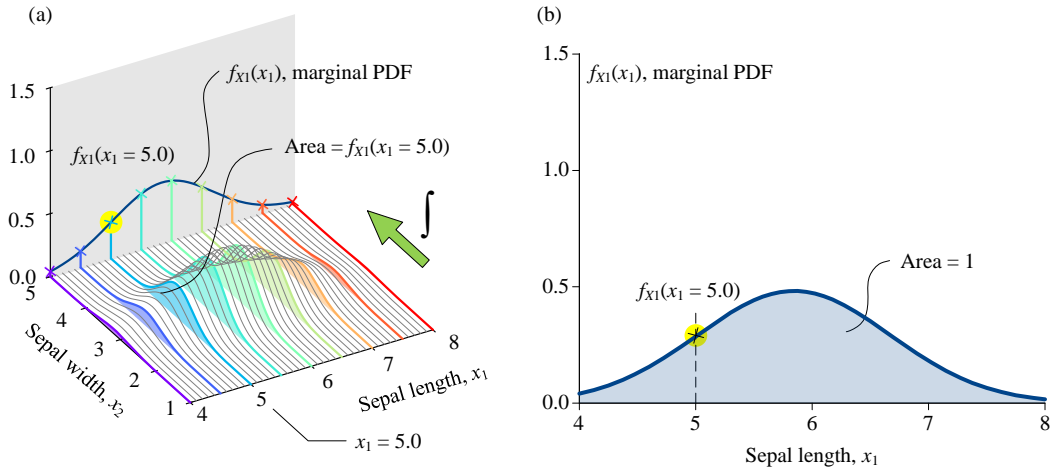
### 花萼长度边缘概率密度函数 $f_{X1}(x_1)$ ：偏积分

图 31 所示为求解花萼长度边缘概率  $f_{X1}(x_1)$  的过程：

$$\underbrace{f_{X1}(x_1)}_{\text{Marginal}} = \int_{-\infty}^{+\infty} \underbrace{f_{X1,X2}(x_1, x_2)}_{\text{Joint}} dx_2 \quad (41)$$

图 31 中彩色阴影面积对应边缘概率，即  $f_{X1}(x_1)$  曲线高度。 $f_{X1}(x_1)$  本身也是概率密度，不是概率值。 $f_{X1}(x_1)$  再积分可以得到概率。

如图 31 (b) 所示， $f_{X1}(x_1)$  曲线和整个横轴围成图形的面积为 1。通过本章前文学习，我们知道  $f_{X1}(x_1)$  也是一元高斯分布 PDF。


图 31. 偏积分求解边缘概率  $f_{X1}(x_1)$ 

### 花萼宽度边缘概率 $f_{X2}(x_2)$ ：偏求和

图 32 所示为求解花萼宽度边缘概率的过程：

$$f_{X2}(x_2) = \int_{-\infty}^{+\infty} f_{X1,X2}(x_1, x_2) dx_1 \quad (42)$$

如所示， $f_{X2}(x_2)$  为一元高斯分布 PDF。

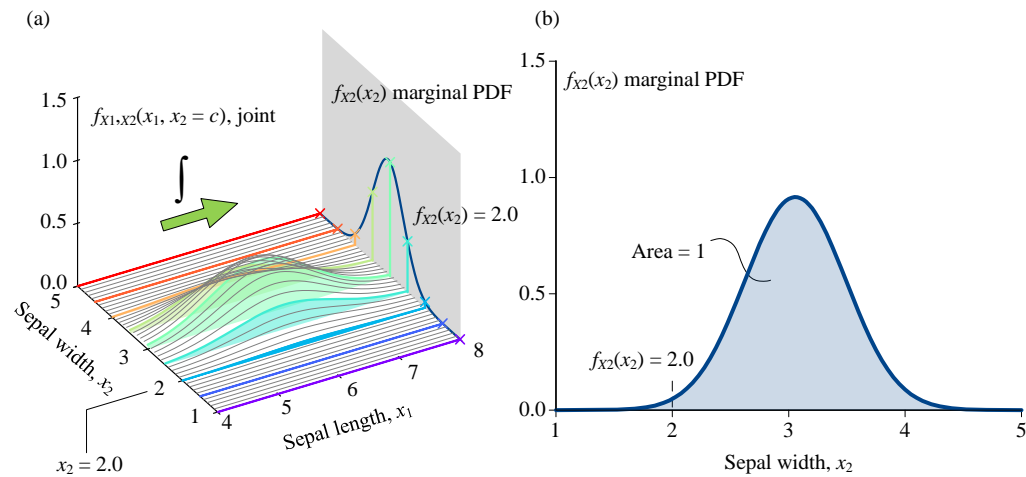

图 32. 偏积分求解边缘概率  $f_{X2}(x_2)$ 

图 33 所示为联合概率和边缘概率之间关系。图中联合概率密度  $f_{X1,X2}(x_1, x_2)$  采用二元高斯分布估计得到。图 33 中  $f_{X1,X2}(x_1, x_2)$  等高线并没有特别准确捕捉到鸢尾花花萼长度、花萼宽度样本散点分布细节。

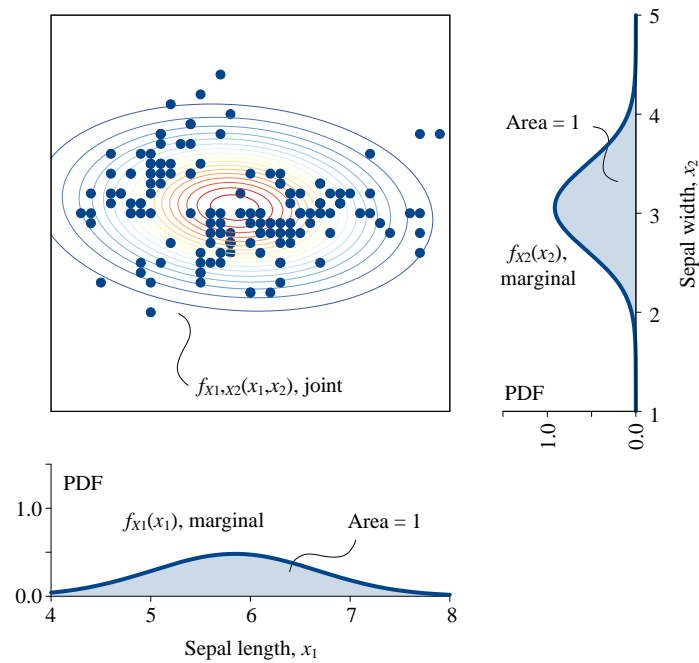


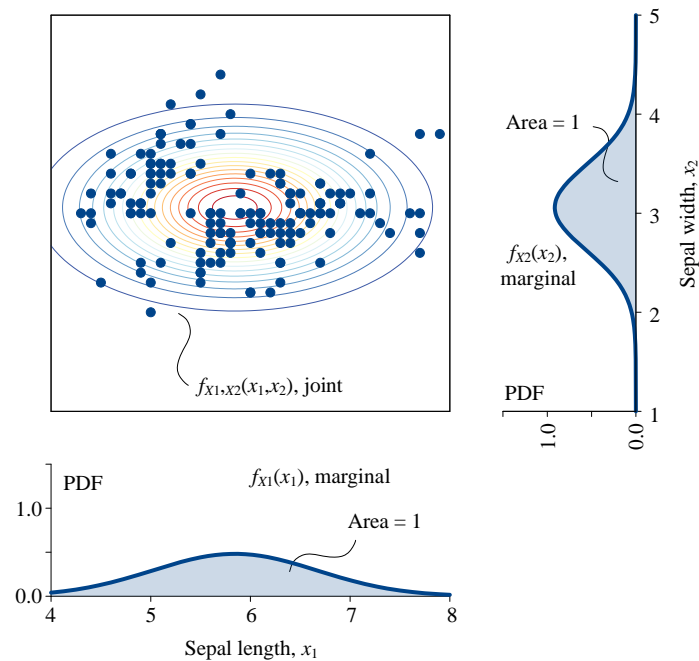
图 33. 联合概率和边缘概率之间关系

## 假设独立

如果假设  $X_1$  和  $X_2$  独立，则联合概率密度  $f_{X1,X2}(x1,x2)$  可通过下式计算得到：

$$f_{X1,X2}(x1,x2) = f_{X1}(x1) \cdot f_{X2}(x2) \quad (43)$$

图 34 所示为假设  $X_1$  和  $X_2$  独立时  $f_{X1,X2}(x1,x2)$  的平面等高线和边缘概率之间关系。椭圆等高线为正椭圆，而非旋转椭圆 (图 33)。

图 34. 联合概率，假设  $X_1$  和  $X_2$  独立

### 给定花萼长度，花萼宽度的条件概率密度 $f_{X_2|X_1}(x_2|x_1)$

如图 35 所示，利用贝叶斯定理，条件概率密度  $f_{X_2|X_1}(x_2|x_1)$  可以通过下式计算：

$$\underbrace{f_{X_2|X_1}(x_2|x_1)}_{\text{Conditional}} = \frac{\overbrace{f_{X_1, X_2}(x_1, x_2)}^{\text{Joint}}}{\underbrace{f_{X_1}(x_1)}_{\text{Marginal}}} \quad (44)$$

分母中的边缘概率  $f_{X_1}(x_1)$  ( $>0$ ) 起到归一化作用。如图 35 (b) 所示，经过归一化的条件概率曲线围成的面积变为 1。

将不同位置的条件概率密度  $f_{X_2|X_1}(x_2|x_1)$  曲线投影到平面得到图 36。我们隐约发现图 36 (b) 中每条曲线看上去都是一元高斯分布。

这难道是个巧合？

我们将在本书第 13 章揭晓答案。

$f_{X_2|X_1}(x_2|x_1)$  本身也是一个二元函数。图 37 所示为  $f_{X_2|X_1}(x_2|x_1)$  三维等高线和平面等高线。从平面等高线中，我们看到一系列直线。

这难道也是个巧合？

答案同样在本书第 13 章给出。

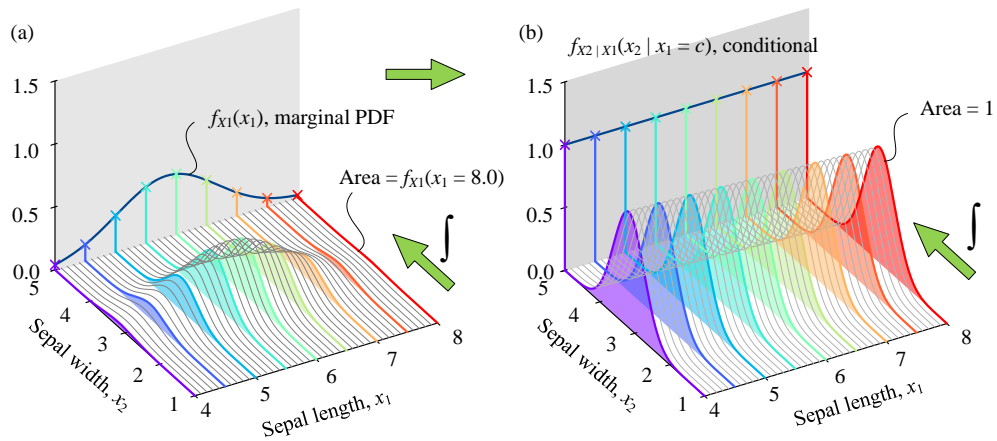


图 35. 计算条件概率  $f_{x2|x1}(x2|x1)$  原理

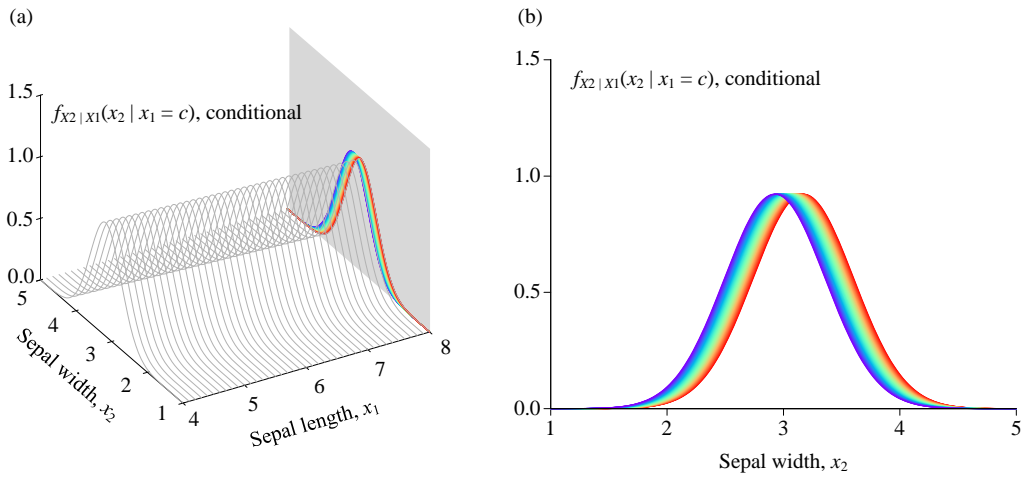


图 36.  $f_{x2|x1}(x2|x1)$  曲线投影到平面

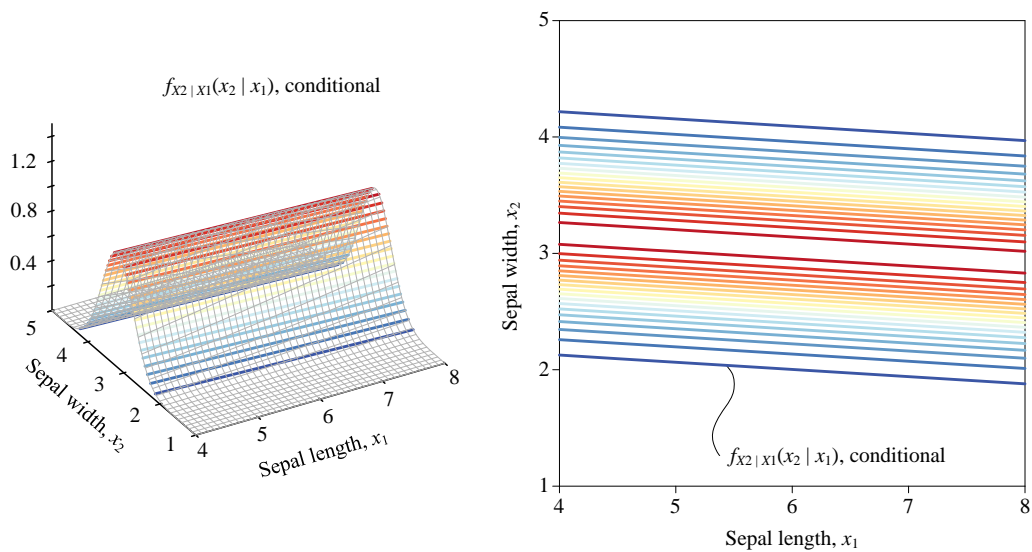


图 37.  $f_{x2|x1}(x2|x1)$  条件概率密度三维等高线和平面等高线，不考虑分类

### 给定花萼宽度，花萼长度的条件概率密度函数 $f_{x_1|x_2}(x_1|x_2)$

如图 38 所示，同样利用贝叶斯定理，条件概率密度  $f_{x_1|x_2}(x_1|x_2)$  可以通过下式计算：

$$\underbrace{f_{x_1|x_2}(x_1|x_2)}_{\text{Conditional}} = \frac{\overbrace{f_{x_1,x_2}(x_1,x_2)}^{\text{Joint}}}{\underbrace{f_{x_2}(x_2)}_{\text{Marginal}}} \quad (45)$$

类似前文，上式中分母中  $f_{x_2}(x_2) (> 0)$  起到归一化作用。

将不同位置的条件概率密度  $f_{x_1|x_2}(x_1|x_2)$  曲线投影到平面得到图 39。图 39 (b) 中每条曲线也都类似一元高斯分布曲线。

$f_{x_1|x_2}(x_1|x_2)$  同样也是一个二元函数，如图 40 所示的  $f_{x_1|x_2}(x_1|x_2)$  三维等高线和平面等高线。

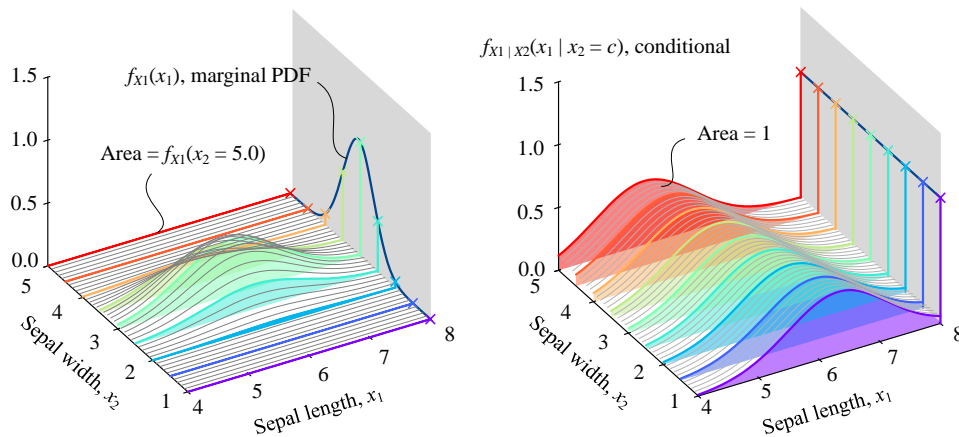


图 38. 计算条件概率  $f_{x_1|x_2}(x_1|x_2)$  原理

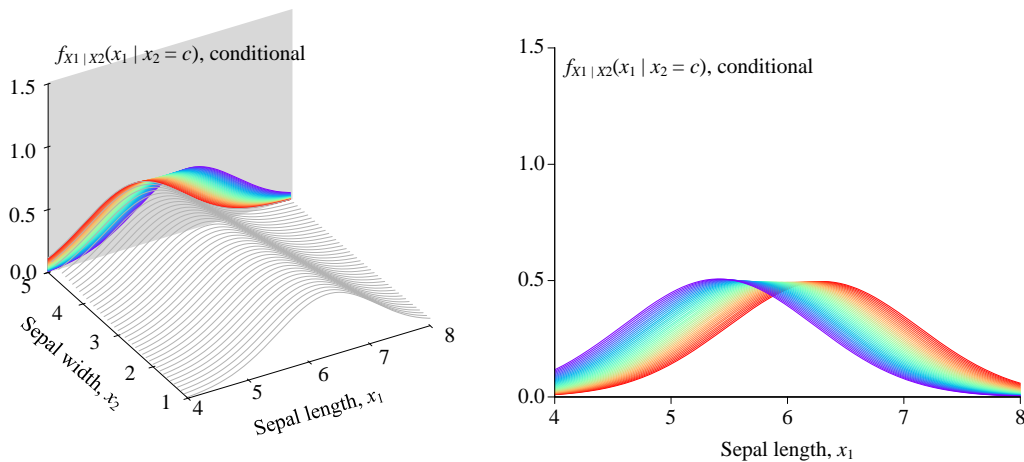
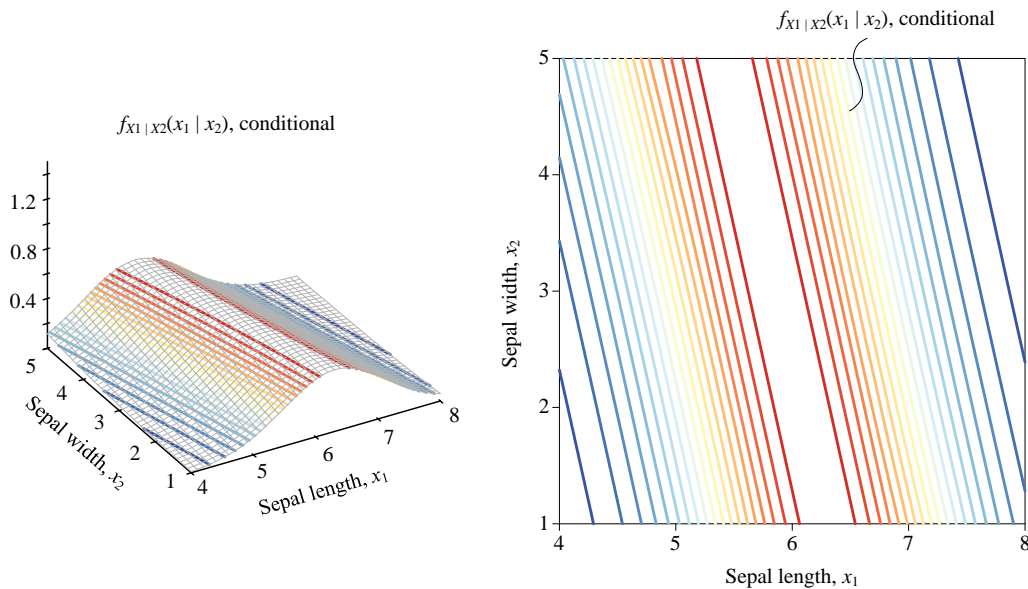


图 39.  $f_{x_1|x_2}(x_1|x_2)$  曲线投影到平面



图 40.  $f_{X1|X2}(x1|x2)$  条件概率密度三维等高线和平面等高线，不考虑分类

## 10.7 以鸢尾花数据为例：考虑分类标签

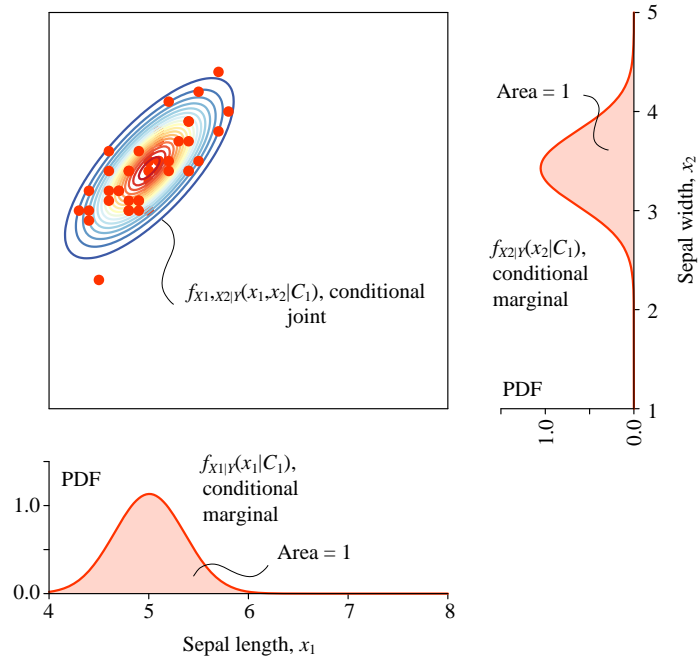
本节讨论考虑鸢尾花分类条件下的条件概率 PDF。

**给定分类标签  $Y = C_1$  (setosa)**

给定分类标签  $Y = C_1$  (setosa) 条件下，假设鸢尾花花萼长度、花萼宽度同样服从二元高斯分布。

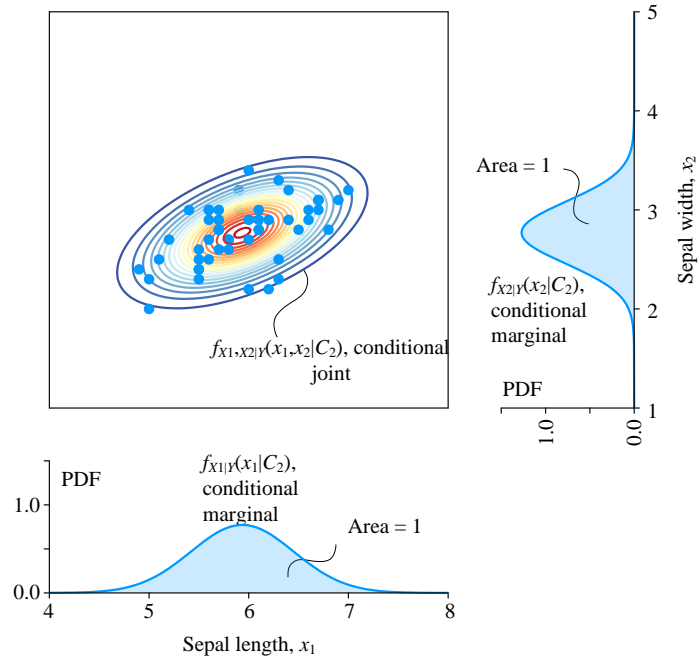
图 41 所示为，给定分类标签  $Y = C_1$  (setosa)，条件概率  $f_{X1,X2|Y}(x1, x2 | y = C1)$  平面等高线和条件边缘概率密度曲线。  $f_{X1,X2|Y}(x1, x2 | y = C1)$  曲面和整个水平面围成体积为 1。

图 41 中  $f_{X1|Y}(x1 | y = C1)$ 、 $f_{X2|Y}(x2 | y = C1)$  分别和  $x1$ 、 $x2$  围成的面积也是 1。

图 41. 条件概率  $f_{X1,X2|Y}(x1, x2 | y = C1)$  平面等高线和条件边缘概率密度曲线，给定分类标签  $Y = C1$  (setosa)

### 给定分类标签 $Y = C2$ (versicolor)

图 42 所示为，给定分类标签  $Y = C2$  (versicolor)，条件概率  $f_{X1,X2|Y}(x1, x2 | y = C2)$  平面等高线和条件边缘概率密度曲线。

图 42. 条件概率  $f_{X1,X2|Y}(x1, x2 | y = C2)$  平面等高线和条件边缘概率密度曲线，给定分类标签  $Y = C2$  (versicolor)

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：[jiang.visualize.ml@gmail.com](mailto:jiang.visualize.ml@gmail.com)

### 给定分类标签 $Y = C_3$ (virginica)

图 43 所示为，给定分类标签  $Y = C_3$  (virginica)，条件概率  $f_{X_1, X_2 | Y}(x_1, x_2 | y = C_3)$  平面等高线和条件边缘概率密度曲线。

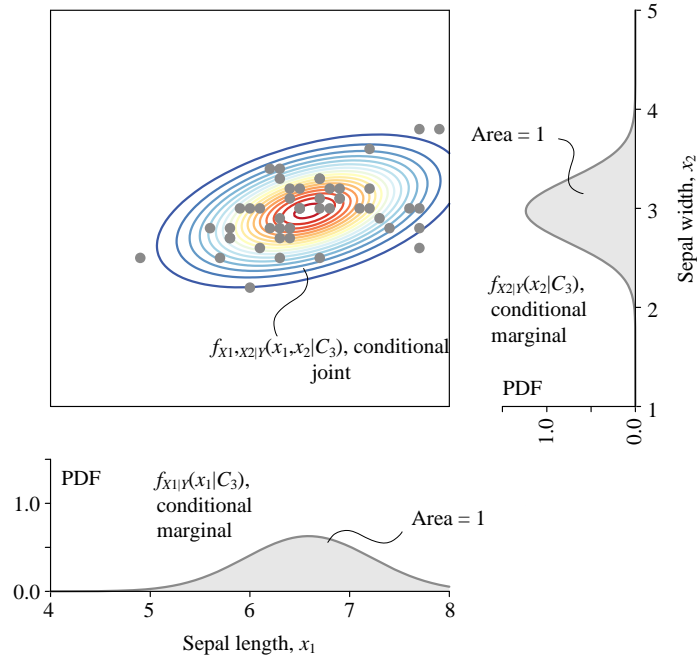


图 43. 条件概率  $p_{X_1, X_2 | Y}(x_1, x_2 | y = C_3)$  平面等高线和条件边缘概率密度曲线，给定分类标签  $Y = C_3$  (virginica)

### 全概率

如图 44 所示，利用全概率定理，三个条件概率等高线叠加可以得到联合概率密度，即：

$$\begin{aligned}
 f_{X_1, X_2}(x_1, x_2) = & f_{X_1, X_2 | Y}(x_1, x_2 | y = C_1) p_Y(C_1) + \\
 & f_{X_1, X_2 | Y}(x_1, x_2 | y = C_2) p_Y(C_2) + \\
 & f_{X_1, X_2 | Y}(x_1, x_2 | y = C_3) p_Y(C_3)
 \end{aligned} \tag{46}$$

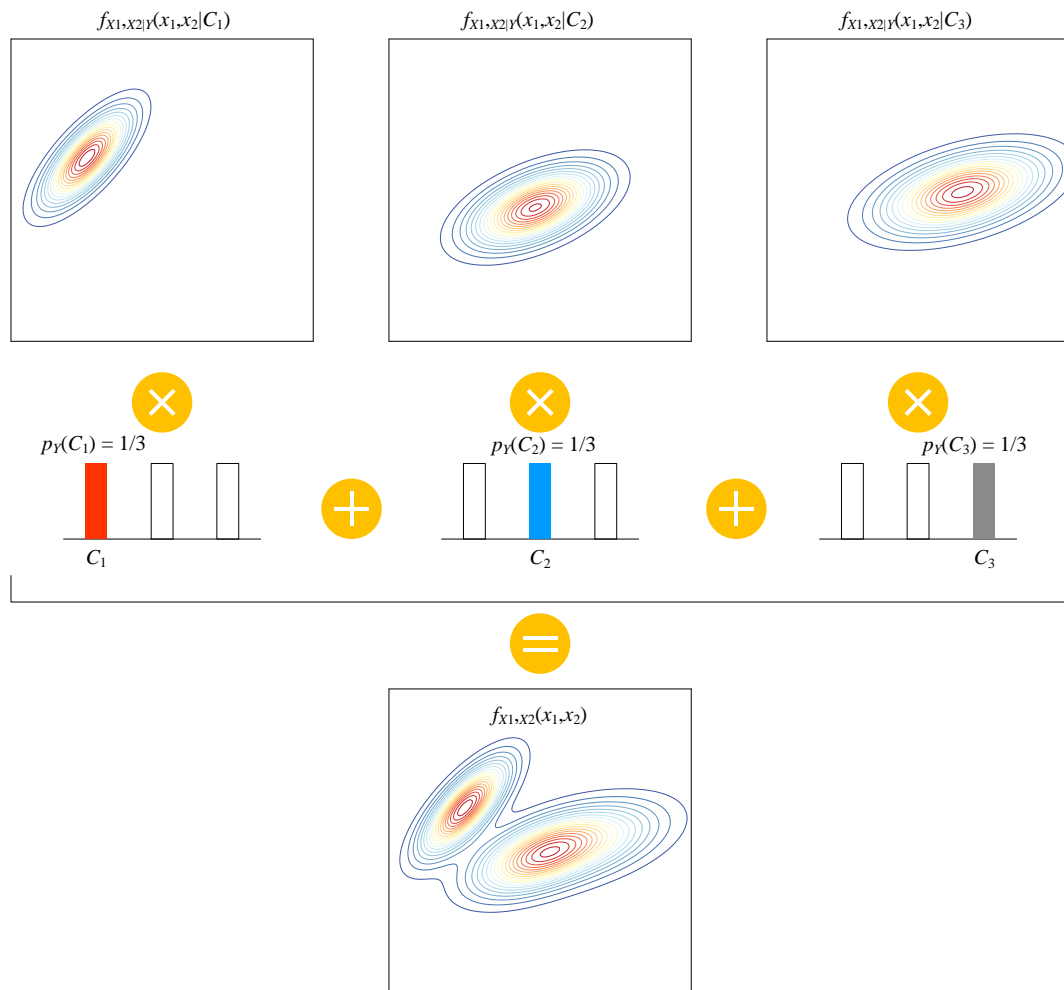


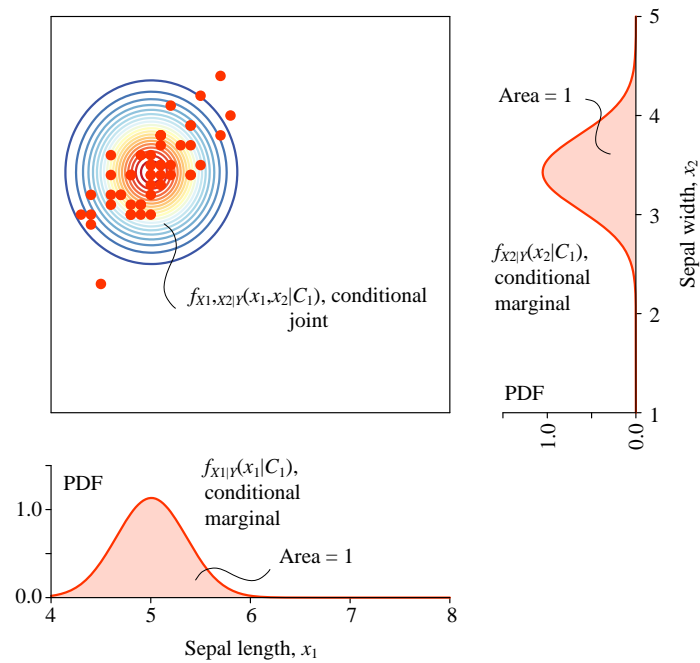
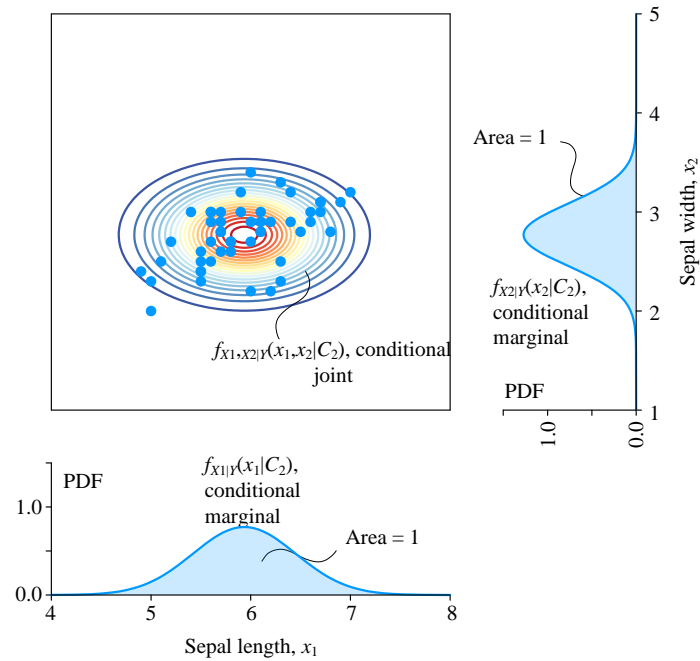
图 44. 估算联合概率密度，假设条件概率服从二元高斯分布

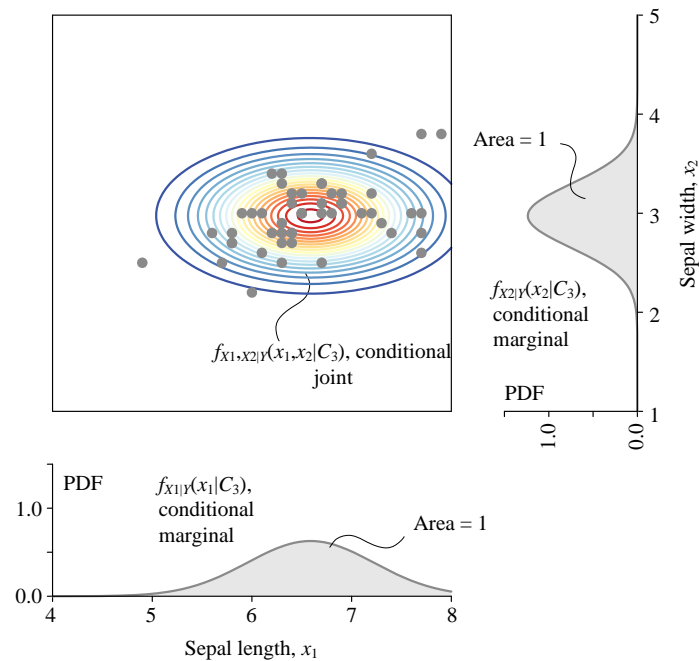
## 假设条件独立

如图 45 所示，如果假设条件独立， $f_{X_1, X_2|Y}(x_1, x_2 | y = C_1)$  可以通过下式计算得到：

$$\underbrace{f_{X_1, X_2|Y}(x_1, x_2 | y = C_1)}_{\text{Conditional joint}} = \underbrace{f_{X_1|Y}(x_1 | y = C_1)}_{\text{Conditional marginal}} \cdot \underbrace{f_{X_2|Y}(x_2 | y = C_1)}_{\text{Conditional marginal}} \quad (47)$$

同理我们可以计算得到  $f_{X_1, X_2|Y}(x_1, x_2 | y = C_2)$ 、 $f_{X_1, X_2|Y}(x_1, x_2 | y = C_3)$ ，具体如图 46、图 47 所示。

图 45. 给定  $Y = C_1$ ,  $X_1$  和  $X_2$  条件独立, 估算条件概率  $f_{X1,X2|Y}(x_1, x_2 | y = C_1)$ 图 46. 给定  $Y = C_2$ ,  $X_1$  和  $X_2$  条件独立, 估算条件概率  $f_{X1,X2|Y}(x_1, x_2 | y = C_2)$

图 47. 给定  $Y = C_3$ ,  $X_1$  和  $X_2$  条件独立, 估算条件概率  $f_{X1, X2|Y}(x_1, x_2 | y = C_3)$ 

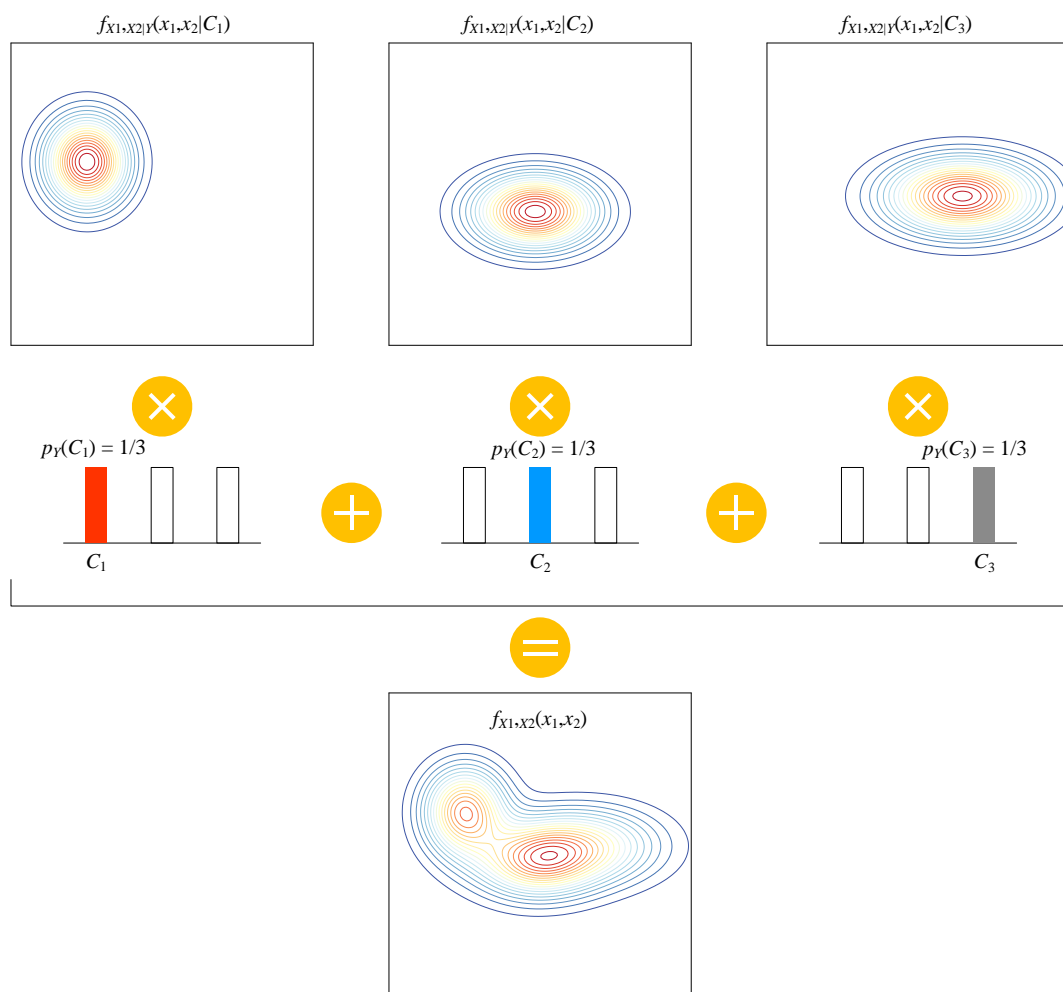
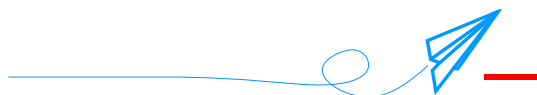
### 估计联合概率

如图 48 所示, 在假设条件独立情况下, 利用全概率定理估算  $f_{X1, X2}(x_1, x_2)$ :

$$\begin{aligned}
 f_{X1, X2}(x_1, x_2) = & f_{X1|Y}(x_1 | y = C_1) f_{X2|Y}(x_2 | y = C_1) p_Y(C_1) + \\
 & f_{X1|Y}(x_1 | y = C_2) f_{X2|Y}(x_2 | y = C_2) p_Y(C_2) + \\
 & f_{X1|Y}(x_1 | y = C_3) f_{X2|Y}(x_2 | y = C_3) p_Y(C_3) +
 \end{aligned} \tag{48}$$

图 44 和图 48 涉及的这些技术细节对于理解贝叶斯分类器原理有很重要意义。

➡ 本书第 19、20 章将从贝叶斯定理视角简单介绍分类原理, 《机器学习》一册将专门讲解朴素贝叶斯分类器。

图 48. 利用全概率定理，估算  $f_{X1,X2}(x1, x2)$ ，假设条件独立

二元高斯分布的概率密度函数的等高线呈现出椭圆形状，这一点极其重要。这个椭圆将把协方差矩阵、特征值分解、Cholesky 分解、条件概率、马氏距离、线性回归、主成分分析、高斯混合模型、高斯过程等一系列概念紧密联系起来。