Probability & Statistics for DS & AI

Conditional Distribution

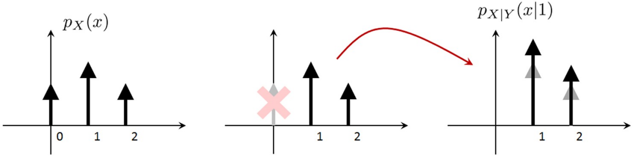
Michele Guindani

Summer

Conditional PMF

Let X and Y be two discrete random variables. The conditional PMF of X given Y is

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$



Suppose X is the sum of two coins with PMF $0.25, 0.5, 0.25$. Let Y be the first coin. When X is unconditioned, the PMF is just $[0.25, 0.5, 0.25]$. When X is conditioned on $Y = 1$, then "$X = 0$" cannot happen. Therefore, the resulting PMF $p_{X|Y}(x|1)$ only has two states. After normalization we obtain the conditional PMF $[0, 0.66, 0.33]$.

See examples 5.17; 5.7; 1.18 in your textbook

Conditional PDF

Let X and Y be two continuous random variables. The conditional PDF of X given Y is

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

Conditional Expectation

The conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \sum_x x \, p_{X|Y}(x \mid y)$$

for the discrete random variables, and

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

- **What is conditional expectation?**
- $\mathbb{E}[X \mid Y = y]$ is the expectation using $f_{X|Y}(x \mid y)$.
- The integration is taken w.r.t. x, because $Y = y$ is given and fixed.

Law of total expectation

Example

- Suppose there are two classes of cars. Let X be the speed and C be the class.
- When $C = 1$, we know that $X \sim$ Gaussian (μ_1, σ_1). We know that $\mathbb{P}[C = 1] = p$.
- When $C = 2, X \sim$ Gaussian (μ_2, σ_2).
- Also, $\mathbb{P}[C = 2] = 1 - p$.
- Suppose you see a car on the freeway, what is its average speed?

Law of Total expectation

The problem has given us everything we need. In particular, we know the conditional PDFs of X, and the marginal pmf of C:

$f_{X|C}(x \mid 1) =$

$f_{X|C}(x \mid 2) =$

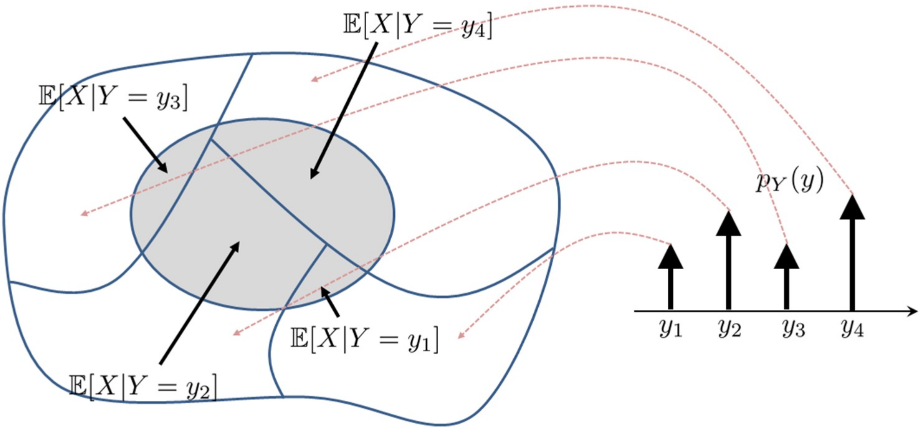
Conditioned on C, we have two expectations:

$\mathbb{E}[X \mid C = 1] =$

$\mathbb{E}[X \mid C = 2] =$

The overall expectation is:

Law of total expectation



Multivariate Gaussian

A d-dimensional joint Gaussian has a PDF

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \right\}$$

where d denotes the dimensionality of the vector $\boldsymbol{x}$.

Multivariate Gaussian

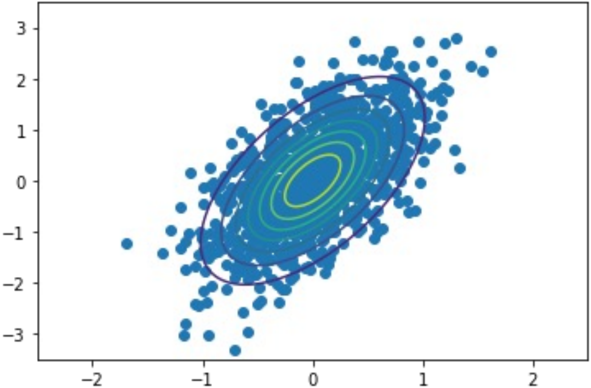
- Random vector: $\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix}$, and $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$

- Mean Vector:

$$\boldsymbol{\mu} \stackrel{\text{def}}{=} \mathbb{E}[\boldsymbol{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_d] \end{bmatrix}$$

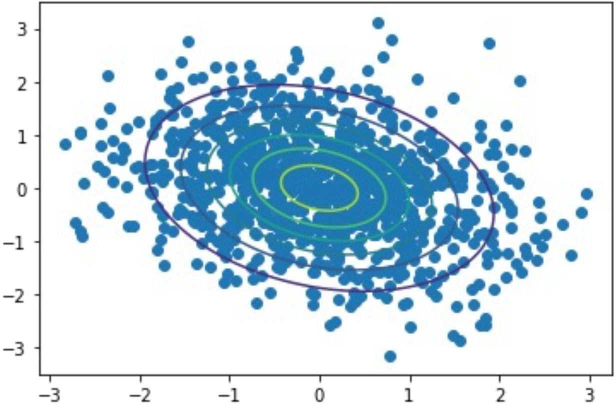
- Covariance:

$$\boldsymbol{\Sigma} \stackrel{\text{def}}{=} \mathrm{Cov}(\boldsymbol{X}) = \begin{bmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}(X_1, X_2) & \dots & \mathrm{Cov}(X_1, X_d) \\ \mathrm{Cov}[X_2, X_1] & \mathrm{Var}[X_2] & \dots & \mathrm{Cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_d, X_1) & \mathrm{Cov}(X_d, X_2) & \dots & \mathrm{Var}[X_d] \end{bmatrix}$$

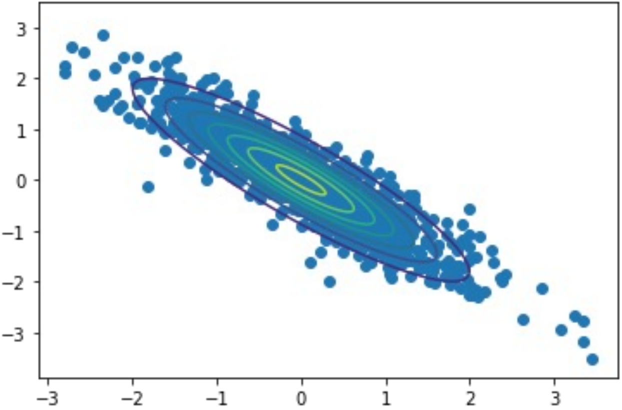


$$(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.25 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}$$

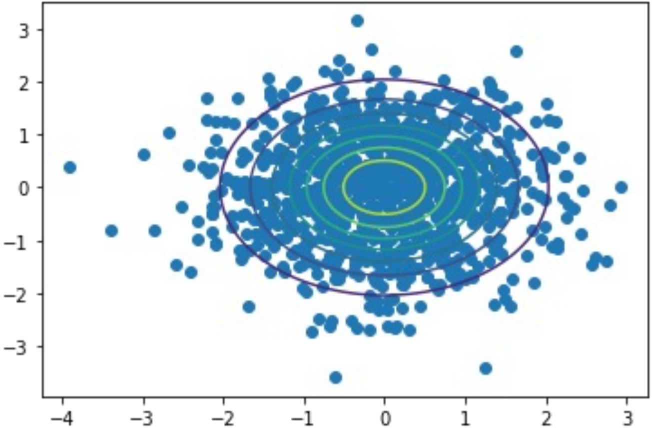
```python
# Python code: Overlay random numbers with the Gaussian contour.
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

X = stats.multivariate_normal.rvs([0,0],[[0.25,0.3],[0.3,1.0]],1000)
x1 = np.arange(-2.5, 2.5, 0.01)
x2 = np.arange(-3.5, 3.5, 0.01)
X1, X2 = np.meshgrid(x1,x2)
Xpos = np.empty(X1.shape + (2,))
Xpos[:,:,0] = X1
Xpos[:,:,1] = X2

F = stats.multivariate_normal.pdf(Xpos,[0,0],[[0.25,0.3],[0.3,1.0]])

plt.scatter(X[:,0],X[:,1])
plt.contour(x1,x2,F)
```



$$(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.25 \\ -0.25 & 1 \end{bmatrix}$$



$$(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}$$



$$(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$