EE2S31 Signal Processing – Stochastic Processes

Lecture 2: Random vectors & conditional probability models – Chs. 8 & 7

Alle-Jan van der Veen

–

25 Apr 2022

Today

- Extension of last week's multiple variables: **random vectors**
- **Conditional probability models:**
  - Conditioning a random variable by an event
  - Conditioning two random variables by an event
  - Conditioning by another random variable

(Ch. 8) Random vectors

Why random vectors?

- More concise representations.
- Allows to use principles from linear algebra.

Notation

- A random vector is the column vector
$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} = [X_1, \cdots, X_N]^T$$
- Transpose operator: $\cdot^T$ or $\cdot'$
- Sample (realization) of random vector: $\boldsymbol{x} = [x_1, \cdots, x_N]^T$
- CDF of a random vector $\boldsymbol{X}$: $F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1,\cdots,X_N}(x_1, \cdots, x_N)$
- PMF of a (discrete) random vector $\boldsymbol{X}$:
$P_{\boldsymbol{X}}(\boldsymbol{x}) = P_{X_1,\cdots,X_N}(x_1, \cdots, x_N)$
- PDF of a (continuous) random vector $\boldsymbol{X}$:
$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1,\cdots,X_N}(x_1, \cdots, x_N)$

Example

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{cases} 6e^{-\boldsymbol{a}^T\boldsymbol{x}} & \boldsymbol{x} \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{with} \quad \boldsymbol{a} = [1 \;\; 2 \;\; 3]^T.$$

What is the CDF $F_{\boldsymbol{X}}(\boldsymbol{x})$?

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{cases} 6e^{-\boldsymbol{a}^T\boldsymbol{x}} & \boldsymbol{x} \geq 0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 6e^{-x_1-2x_2-3x_3} & x_i \geq 0 \,\forall\, i \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} F_{\boldsymbol{X}}(\boldsymbol{x}) &= \begin{cases} \int_0^{x_1}\int_0^{x_2}\int_0^{x_3} 6e^{-u_1-2u_2-3u_3}\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3 & x_i \geq 0 \,\forall\, i \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} (1-e^{-x_1})(1-e^{-2x_2})(1-e^{-3x_3}) & x_i \geq 0 \,\forall\, i \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Pairs of random vectors

Joint CDF, PDF and PMF of two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$:

- **CDF** of random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$F_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = F_{X_1,\cdots,X_N,Y_1,\cdots,Y_N}(x_1,\cdots,x_N,y_1,\cdots,y_N)$$

- **PMF** of (discrete) random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = P_{X_1,\cdots,X_N,Y_1,\cdots,Y_N}(x_1,\cdots,x_N,y_1,\cdots,y_N)$$

- **PDF** of (continuous) random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = f_{X_1,\cdots,X_N,Y_1,\cdots,Y_N}(x_1,\cdots,x_N,y_1,\cdots,y_N)$$

Independent random vectors

Two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent if

- Discrete RVs: $P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = P_{\boldsymbol{X}}(\boldsymbol{x})P_{\boldsymbol{Y}}(\boldsymbol{y})$
- Continuous RVs: $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x})f_{\boldsymbol{Y}}(\boldsymbol{y})$

Expected values for random vectors

For a random matrix $\boldsymbol{A}$, with A_{ij} the (i,j)th element of $\boldsymbol{A}$, $\mathrm{E}[\boldsymbol{A}]$ is a matrix with $\mathrm{E}[A_{ij}]$ as its (i,j)th element.

The expected value of the random vector $\boldsymbol{X}$ therefore equals

$$\mathrm{E}[\boldsymbol{X}] = \begin{bmatrix} \mathrm{E}[X_1] \\ \vdots \\ \mathrm{E}[X_N] \end{bmatrix}$$

The correlation matrix

Now consider the vector $\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}$, shown for $N = 3$.

$$\boldsymbol{X}\boldsymbol{X}^T = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} [X_1,\, X_2,\, X_3] = \begin{bmatrix} X_1^2 & X_1X_2 & X_1X_3 \\ X_2X_1 & X_2^2 & X_2X_3 \\ X_3X_1 & X_3X_2 & X_3^2 \end{bmatrix}$$

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right] &= \begin{bmatrix} \mathrm{E}[X_1^2] & \mathrm{E}[X_1X_2] & \mathrm{E}[X_1X_3] \\ \mathrm{E}[X_2X_1] & \mathrm{E}[X_2^2] & \mathrm{E}[X_2X_3] \\ \mathrm{E}[X_3X_1] & \mathrm{E}[X_3X_2] & \mathrm{E}[X_3^2] \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{E}[X_1^2] & r_{X_1X_2} & r_{X_1X_3} \\ r_{X_2X_1} & \mathrm{E}[X_2^2] & r_{X_2X_3} \\ r_{X_3X_1} & r_{X_3X_2} & \mathrm{E}[X_3^2] \end{bmatrix} \end{aligned}$$

$\boldsymbol{R}_{\boldsymbol{X}} = \mathrm{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right]$ is known as the **correlation matrix** and extends the concept of the correlation $\mathrm{E}[XY]$ to vectors.

The covariance matrix

Similarly, we can define the **covariance matrix**

$$\boldsymbol{C_X} = \mathrm{E}\left[(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^T\right] = \boldsymbol{R_X} - \mathrm{E}[\boldsymbol{X}]\mathrm{E}[\boldsymbol{X}]^T.$$

For the vector $\boldsymbol{X} = [X_1, X_2, X_3]^T$ we get

$$\boldsymbol{C_X} = \mathrm{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right] - \mathrm{E}[\boldsymbol{X}]E[\boldsymbol{X}]^T = \begin{bmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & \mathrm{cov}(X_1, X_3) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \mathrm{cov}(X_2, X_3) \\ \mathrm{cov}(X_3, X_1) & \mathrm{cov}(X_3, X_2) & \mathrm{var}(X_3) \end{bmatrix}.$$

If the X_i are uncorrelated ($\mathrm{cov}(X_i, X_j) = 0$), then $\boldsymbol{C_X}$ is diagonal.

If the random variables $\{X_i\}$ are **independent, identically distributed** (i.i.d.), then $\boldsymbol{C_X} = \sigma^2 \boldsymbol{I}$.

Cross-covariance & cross-correlation matrix

For two random vectors, their **cross-correlation matrix** is defined as

$$\boldsymbol{R_{XY}} = \mathrm{E}\left[\boldsymbol{X}\boldsymbol{Y}^T\right]$$

and their **cross-covariance matrix** is

$$\boldsymbol{C_{XY}} = \mathrm{E}\left[\boldsymbol{X}\boldsymbol{Y}^T\right] - \mathrm{E}\left[\boldsymbol{X}\right]\mathrm{E}[\boldsymbol{Y}^T]$$

Linear transformations

If $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ is a linear transformation of a random vector $\boldsymbol{X}$, then

$$\begin{aligned} \mathrm{E}[\boldsymbol{Y}] &= \boldsymbol{A}\,\mathrm{E}[\boldsymbol{X}] + \boldsymbol{b} \\ \boldsymbol{C_Y} &= \boldsymbol{A}\boldsymbol{C_X}\boldsymbol{A}^T \\ \boldsymbol{C_{YX}} &= \boldsymbol{A}\boldsymbol{C_X} \end{aligned}$$

Exercise 8.5.2

$\boldsymbol{X} = [X_1, X_2]^T$ is the Gaussian random vector with $\mathrm{E}[\boldsymbol{X}] = [0, 0]^T$ and covariance matrix
$$\boldsymbol{C_X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

What is the PDF of $Y = [2, 1]\boldsymbol{X}$?

Exercise 8.5.2

$\boldsymbol{X} = [X_1, X_2]^T$ is the Gaussian random vector with $\mathrm{E}[\boldsymbol{X}] = [0, 0]^T$ and covariance matrix

$$\boldsymbol{C_X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

What is the PDF of $Y = [2, 1]\boldsymbol{X}$?

Y is the sum of two Gaussians, is therefore Gaussian with mean

$$\mathrm{E}[Y] = \mathrm{E}[2X_1 + X_2] = 0$$

and variance

$$\begin{aligned}
\mathrm{var}[Y] &= \mathrm{E}[Y^2] - 0 = \mathrm{E}[YY^T] \\
&= \mathrm{E}\left\{[2 \;\; 1]\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}[X_1 \;\; X_2]\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\} \\
&= [2 \;\; 1]\mathrm{E}\left\{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}[X_1 \;\; X_2]\right\}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \\
&= [2 \;\; 1]\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 10
\end{aligned}$$

Gaussian variables

In Ch. 5 we saw

$$f_{X,Y}(x,y) = \frac{\exp\left[-\frac{\left(\frac{x-\mathrm{E}[X]}{\sigma_X}\right)^2 - \frac{2\rho(x-\mathrm{E}[X])(y-\mathrm{E}[Y])}{\sigma_X\sigma_Y} + \left(\frac{y-\mathrm{E}[Y]}{\sigma_Y}\right)^2}{2(1-\rho^2)}\right]}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

- Extending this to higher dimensions is rather impractical.
- Using vector notation a very concise and useful expression can be obtained.

Gaussian random vectors

Let $\boldsymbol{X}$ be a vector of correlated Gaussian RVs: $\boldsymbol{X} = [X_1, X_2, \cdots, X_N]^T$.

The PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$ is then given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\exp\left[-\frac{1}{2}\left(\boldsymbol{x} - \mathrm{E}[\boldsymbol{X}]\right)^T \boldsymbol{C}_{\boldsymbol{X}}^{-1} \left(\boldsymbol{x} - \mathrm{E}[\boldsymbol{X}]\right)\right]}{(2\pi)^{N/2} \det(\boldsymbol{C}_{\boldsymbol{X}})^{1/2}}$$

Special case: $N = 2$

$$\begin{aligned}
\boldsymbol{C}_{\boldsymbol{X}} &= \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \\
\det(\boldsymbol{C}_{\boldsymbol{X}}) &= \sigma_X^2\sigma_Y^2(1-\rho^2) \\
\boldsymbol{C}_{\boldsymbol{X}}^{-1} &= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)} \begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}
\end{aligned}$$

Verify that this leads to the expression on the previous slide!

Uncorrelated Gaussian random vectors

PDF of Gaussian random vector:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\exp\left[-\frac{1}{2}\left(\boldsymbol{x} - \mathrm{E}[\boldsymbol{X}]\right)^T \boldsymbol{C}_{\boldsymbol{X}}^{-1} \left(\boldsymbol{x} - \mathrm{E}[\boldsymbol{X}]\right)\right]}{(2\pi)^{N/2} \det(\boldsymbol{C}_{\boldsymbol{X}})^{1/2}}$$

Let $\boldsymbol{X}$ be a vector of *uncorrelated* Gaussian RVs: $\boldsymbol{X} = [X_1, \cdots, X_N]^T$.

- $\boldsymbol{C}_{\boldsymbol{X}} = \mathrm{diag}(\sigma^2_{X_1}, \sigma^2_{X_2}, \cdots, \sigma^2_{X_N})$
- $\det(\boldsymbol{C}_{\boldsymbol{X}}) = \prod_{i=1}^{N} \sigma^2_{X_i}$
- $\left(\boldsymbol{x} - \mathrm{E}[\boldsymbol{X}]\right)^T \boldsymbol{C}_{\boldsymbol{X}}^{-1} \left(\boldsymbol{x} - \mathrm{E}[\boldsymbol{X}]\right) = \sum_{i=1}^{N} \frac{(x_i - \mathrm{E}[X_i])^2}{\sigma^2_{X_i}}$

The PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$ is then given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{N} \frac{\exp[-(x_i - \mathrm{E}[X_i])^2/2\sigma^2_{X_i}]}{\sqrt{2\pi\sigma^2_{X_i}}} = \prod_{i=1}^{N} f_{X_i}(x_i)$$

Hence, the variables $X_1, \cdots, X_N$ are independent.

Linear transformation of random vectors

Let $\boldsymbol{X}$ be a continuous random vector and $\boldsymbol{A}$ an invertible matrix. Then, $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ has the PDF

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|\det(\boldsymbol{A})|} f_{\boldsymbol{X}}\left(\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b})\right)$$

Derivation:

$$\begin{aligned} F_{\boldsymbol{Y}}(\boldsymbol{y}) &= \mathrm{P}[\boldsymbol{Y} \leq \boldsymbol{y}] = \mathrm{P}[\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b} \leq \boldsymbol{y}] = \mathrm{P}[\boldsymbol{X} \leq \boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b})] \\ &= F_{\boldsymbol{X}}(\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b})) \end{aligned}$$

Next, take derivatives to find $f_{\boldsymbol{Y}}(\boldsymbol{y})$.

Transformation of Gaussian random vectors

Let $\boldsymbol{X}$ be a Gaussian random vector and $\boldsymbol{A}$ an invertible matrix.

What is the PDF of $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$?

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{1}{|\det(\boldsymbol{A})|} f_{\boldsymbol{X}}\left(\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{b})\right) \\ &= \frac{\exp\left[-\frac{1}{2}\left(\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{b}) - E[\boldsymbol{X}]\right)^T \boldsymbol{C}_{\boldsymbol{X}}^{-1}\left(\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{b}) - E[\boldsymbol{X}]\right)\right]}{(2\pi)^{N/2}|\det(\boldsymbol{A})|\det(\boldsymbol{C}_{\boldsymbol{X}})^{1/2}}. \end{aligned}$$

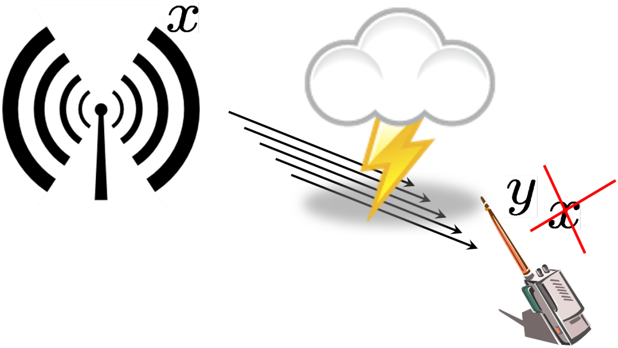
Using some manipulations, this can be rewritten as

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\exp\left[-\frac{1}{2}\left(\boldsymbol{y} - E[\boldsymbol{Y}]\right)^T \boldsymbol{A}^{-T}\boldsymbol{C}_{\boldsymbol{X}}^{-1}\boldsymbol{A}^{-1}\left(\boldsymbol{y} - \mathrm{E}[\boldsymbol{Y}]\right)\right]}{(2\pi)^{N/2}\det(\boldsymbol{A}\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^T)^{1/2}}.$$

$\boldsymbol{Y}$ is thus also Gaussian with $\mathrm{E}[\boldsymbol{Y}] = \boldsymbol{A}\,\mathrm{E}[\boldsymbol{X}] + \boldsymbol{b}$ and $\boldsymbol{C}_{\boldsymbol{Y}} = \boldsymbol{A}\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{A}^T$

(But, we already knew this: sum of Gaussians is Gaussian.)

Ch.7 Conditional probability models



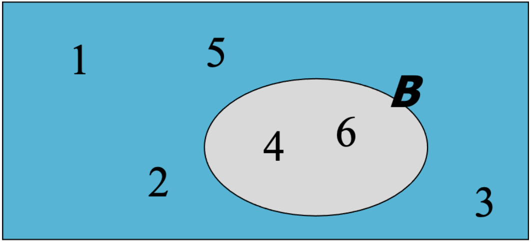
Model: $Y = X + N$

Imagine we observe realizations of Y, while our interest is X.

- Derive $P_X(x)$?
- Probability of X given an observation y: $P_{X|Y}(x|y)$?

Conditional probability

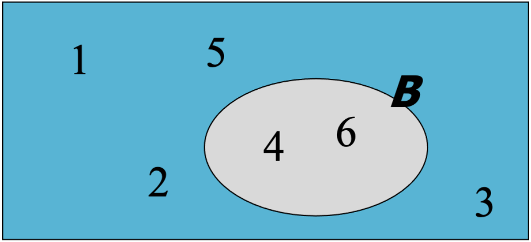
Sometimes the occurrence of one event influences the probability of occurrence of other events.



- P[odd number] ?

Conditional probability

Sometimes the occurrence of one event influences the probability of occurrence of other events.



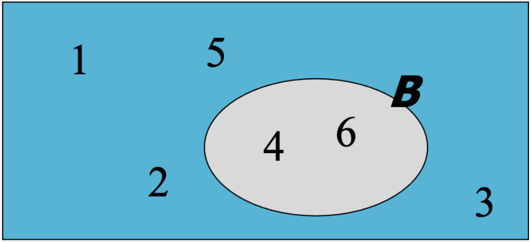
- P[odd number] ?
- P[odd number if we know that the outcome is in event B] ?

Conditional probability

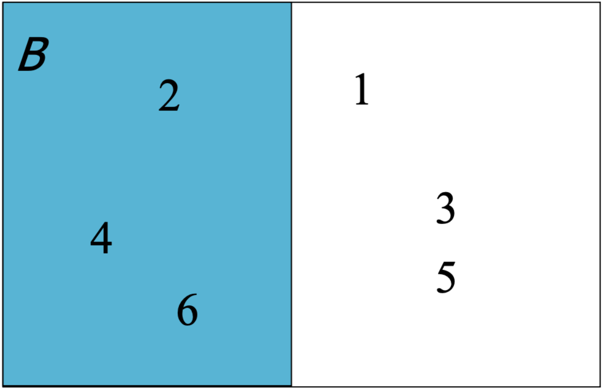
Interpretation: $\mathrm{P}[A|B]$ is the probability of A, given that the event B has already occurred.

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]} = \frac{\mathrm{P}[A, B]}{\mathrm{P}[B]} \qquad \text{(Bayes' theorem)}$$

$$\mathrm{P}[A, B] = \mathrm{P}[A|B]\mathrm{P}[B]$$

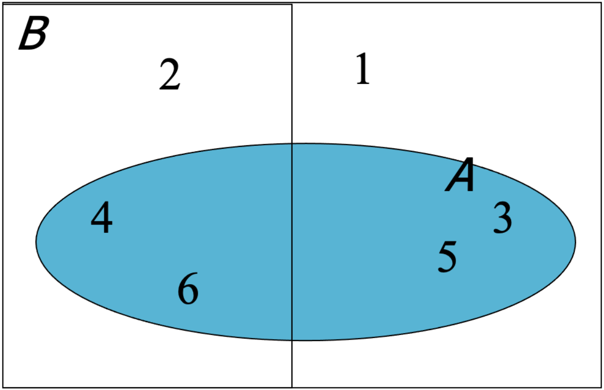


Example (1)



Event B: "Even outcome" when rolling the dice

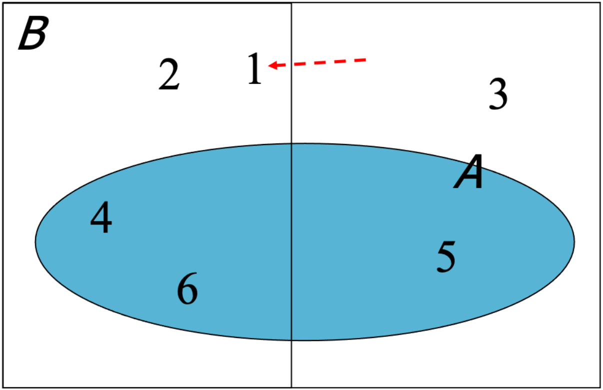
Example (2)



Event A: "3 or more" when rolling the dice.

How large is $\mathrm{P}[A|B]$?

Example (3)



Different B! How large is $\mathrm{P}[A|B]$ now?

Ch.7 Conditional probability models

Starting from the conditional probability, we can also define the conditional CDF:

- **Conditional probability** (Bayes' theorem)

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[A, B]}{\mathrm{P}[B]} = \frac{\mathrm{P}[B|A]\mathrm{P}[A]}{\mathrm{P}[B]}$$

- **Conditional CDF**
  Let event $A = \{X \leq x\}$. Then

$$\mathrm{P}[A|B] = \mathrm{P}[X \leq x|B] \,.$$

Conditioning the CDF, PMF and PDF by an event

CDF, PMF and PDF conditioned by an event:

- **Conditional CDF:** $F_{X|B}(x) = \mathrm{P}[X \leq x|B]$
- **Conditional PMF:** $P_{X|B}(x) = \mathrm{P}[X = x|B]$
- **Conditional PDF:** $f_{X|B}(x) = \frac{\mathrm{d}F_{X|B}(x)}{\mathrm{d}x}$

Conditioning by an event changes the probabilities:

$$P_{X|B}(x) = \begin{cases} \dfrac{P_X(x)}{\mathrm{P}[B]} & x \in B \\ 0 & \text{otherwise} \end{cases} \qquad f_{X|B}(x) = \begin{cases} \dfrac{f_X(x)}{\mathrm{P}[B]} & x \in B \\ 0 & \text{otherwise} \end{cases}$$

Those outcomes x where $x \notin B$ will get zero probability, while those outcomes x where $x \in B$ will get proportionally higher.

Example: calculating the conditional PMF

Let X be the time in integer minutes one waits for a bus:

$$P_X(x) = \begin{cases} \frac{1}{20} & x = 1, 2, ..., 20 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose the bus has not arrived by the 6th minute. What is the conditional PMF of the waiting time?

Let A be the event that the bus has not yet arrived after 6 minutes: $\mathrm{P}[A] = 14/20$.

$$P_{X|X>6}(x) = P_{X|A}(x) = \begin{cases} \frac{1/20}{14/20} = \frac{1}{14} & x = 7, 8, ..., 20 \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 7.1.1

Discrete random variable X has CDF

$$F_X(x) = \begin{cases} 0 & x < -3, \\ 0.4 & -3 \le x < 5, \\ 0.8 & 5 \le x < 7, \\ 1 & x \ge 7. \end{cases}$$

Find the conditional CDF $F_{X|X>0}(x)$ and PMF $P_{X|X>0}(x)$.

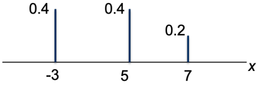
Exercise 7.1.1

Discrete random variable X has CDF

$$F_X(x) = \begin{cases} 0 & x < -3, \\ 0.4 & -3 \le x < 5, \\ 0.8 & 5 \le x < 7, \\ 1 & x \ge 7. \end{cases}$$

Find the conditional CDF $F_{X|X>0}(x)$ and PMF $P_{X|X>0}(x)$.

$$P_X(x) = \begin{cases} 0.4 & x = -3, \\ 0.4 & x = 5, \\ 0.2 & x = 7, \\ 0 & \text{otherwise} \end{cases}$$



Event $B = \{X > 0\}$ has probability $\mathrm{P}[X > 0] = P_X(5) + P_X(7) = 0.6$.

$$P_{X|X>0}(x) = \begin{cases} \frac{P_X(x)}{\mathrm{P}[X>0]} & x \in B, \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 2/3 & x = 5, \\ 1/3 & x = 7 \\ 0 & \text{otherwise} \end{cases}$$

PDF & PMF with a partition

Let $B_1, B_2, \cdots, B_M$ be M *different* (non-overlapping) events, together covering all possible outcomes S_X: a *partition*.

The law of total probability says

$$\text{(discrete)} \quad P_X(x) = \sum_{i=1}^{M} P_{X|B_i}(x)\mathrm{P}(B_i)$$

$$\text{(continuous)} \quad f_X(x) = \sum_{i=1}^{M} f_{X|B_i}(x)\mathrm{P}(B_i)$$

$$\mathrm{E}[X] = \sum_{i=1}^{M} \mathrm{E}[X|B_i]\mathrm{P}[B_i]$$

PDF & PMF with a partition

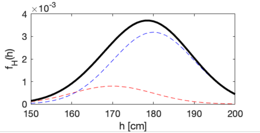
Example The height H of a Male is Gaussian(180,10). The height H of a Female is Gaussian(170,10). There are 4 times more Males than Females in class.

$$\mathrm{P}(M) = 4/5\,, \qquad \mathrm{P}(F) = 1/5$$

$$f_{H|M}(h) = \frac{1}{100\sqrt{2\pi}} e^{-(h-180)^2/200}\,, \quad f_{H|F}(h) = \frac{1}{100\sqrt{2\pi}} e^{-(h-170)^2/200}\,.$$

Then

$$f_H(h) = f_{H|M}(h)\mathrm{P}(M) + f_{H|F}(h)\mathrm{P}(F)$$



$$\mathrm{E}[H] = \mathrm{E}[H|M]\mathrm{P}[M] + \mathrm{E}[H|F]\mathrm{P}[F] = 180 \cdot \frac{4}{5} + 170 \cdot \frac{1}{5}$$

Conditioning multiple RVs by an event

For RVs X and Y and event B, the joint conditional PMF and PDF are given by

$$P_{X,Y|B}(x,y) = \mathrm{P}[X=x, Y=y|B] = \begin{cases} \frac{P_{X,Y}(x,y)}{\mathrm{P}[B]} & (x,y) \in B \\ 0 & \text{otherwise} \end{cases}$$

$$f_{X,Y|B}(x,y) = \begin{cases} \frac{f_{X,Y}(x,y)}{\mathrm{P}[B]} & (x,y) \in B \\ 0 & \text{otherwise} \end{cases}$$

Those outcomes x and y where $(x,y) \notin B$ will get zero probability, while those outcomes x and y where $(x,y) \in B$ will get proportionally higher.

Exercise – Conditional PDF

X and Y are RVs with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{15} & 0 \leq x \leq 5, 0 \leq y \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate the conditional PDF $f_{X,Y|B}(x,y)$ with $B = \{X + Y \geq 4\}$.

Exercise – Conditional PDF

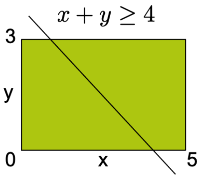
X and Y are RVs with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{15} & 0 \leq x \leq 5, 0 \leq y \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate the conditional PDF $f_{X,Y|B}(x,y)$ with $B = \{X + Y \geq 4\}$.

$\mathrm{P}[B] = 1/2$

$$f_{X,Y|B}(x,y) = \begin{cases} \frac{2}{15} & 0 \leq x \leq 5,\ 0 \leq y \leq 3,\ x + y \geq 4 \\ 0 & \text{otherwise} \end{cases}$$



Conditional expectations

For RVs X and Y and event B, the conditional expected value of $g(X, Y)$ given event B is given by

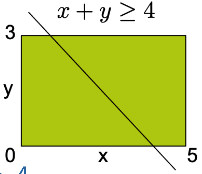
$$\mathrm{E}[g(X, Y)|B] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x, y) P_{X,Y|B}(x, y)$$

$$\mathrm{E}[g(X, Y)|B] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y|B}(x) \mathrm{d}x \mathrm{d}y$$

Exercise – Conditional PDF

X and Y are RVs with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{15} & 0 \leq x \leq 5, 0 \leq y \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$



The conditional PDF $f_{X,Y|B}(x,y)$ with $B = \{X + Y \geq 4\}$ is

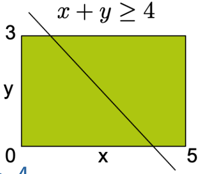
$$f_{X,Y|B}(x,y) = \begin{cases} \frac{2}{15} & 0 \leq x \leq 5,\ 0 \leq y \leq 3,\ x + y \geq 4 \\ 0 & \text{otherwise} \end{cases}$$

Calculate $\mathrm{E}[XY|B]$

Exercise – Conditional PDF

X and Y are RVs with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{15} & 0 \le x \le 5, 0 \le y \le 3 \\ 0 & \text{otherwise.} \end{cases}$$



The conditional PDF $f_{X,Y|B}(x,y)$ with $B = \{X + Y \ge 4\}$ is

$$f_{X,Y|B}(x,y) = \begin{cases} \frac{2}{15} & 0 \le x \le 5,\ 0 \le y \le 3,\ x + y \ge 4 \\ 0 & \text{otherwise} \end{cases}$$

Calculate $\mathrm{E}[XY|B]$

$$\mathrm{E}[XY|B] = \int_0^3 \int_{4-y}^5 xy \, \frac{2}{15} \, \mathrm{d}x\mathrm{d}y = \cdots$$

Conditioning by a random variable

So far we conditioned on an event $(x, y) \in B$.

Special case: conditioning on partial knowledge on one of the variables: $B = \{X = x\}$ or $B = \{Y = y\}$.

For example: knowing $Y = y$ completely determines RV Y, and changes the knowledge we have about X (assuming Y and X are not independent).

Conditional PMF: $P_{X|Y}(x|y) = \dfrac{P_{X,Y}(x, y)}{P_Y(y)}$

Conditional PDF: $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x, y)}{f_Y(y)}$

Exercise 7.4.4

Z is a Gaussian(0,1) noise random variable that is independent of X, and $Y = X + Z$ is a noisy observation of X. What is the conditional PDF $f_{Y|X}(y|x)$?

Exercise 7.4.4

Z is a Gaussian(0,1) noise random variable that is independent of X, and $Y = X + Z$ is a noisy observation of X. What is the conditional PDF $f_{Y|X}(y|x)$?

Given $X = x$, we know that $Y = x + Z$.

Z is Gaussian(0,1). Adding x will shift the mean to x.

Thus, Y is Gaussian(x,1):

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}} e^{-(y-x)^2/2}$$

Exercise 7.4.4

Z is a Gaussian(0,1) noise random variable that is independent of X, and $Y = X + Z$ is a noisy observation of X. What is the conditional PDF $f_{Y|X}(y|x)$?

More "systematic" approach:

$$\begin{aligned} F_{Y|X}(y|x) &= \mathrm{P}[Y \leq y | X = x] \\ &= \mathrm{P}[x + Z \leq y | X = x] \\ &= \mathrm{P}[x + Z \leq y] \qquad (Z \text{ independent of } X) \\ &= \mathrm{P}[Z \leq y - x] \\ &= F_Z(y - x) \end{aligned}$$

$$f_{Y|X}(y|x) = \frac{\mathrm{d}F_{Y|X}(y|x)}{\mathrm{d}y} = \frac{\mathrm{d}F_Z(y-x)}{\mathrm{d}y} = f_Z(y-x)\,.$$

Conditional expectation

Discrete random variables:

$$\mathrm{E}\left[g(X,Y)|Y=y\right] \;=\; \sum_{x\in S_X} g(x,y)P_{X|Y}(x|y)$$

If X and Y are independent, then

$$P_{X|Y}(x|y) \;=\; P_X(x), \text{and } P_{Y|X}(y|x) = P_Y(y)$$

$$\mathrm{E}\left[X|Y=y\right] \;=\; \sum_{x\in S_X} x\,P_{X|Y}(x|y) = \sum_{x\in S_X} x\,P_X(x) = \mathrm{E}\left[X\right]$$

Continuous random variables: similarly,

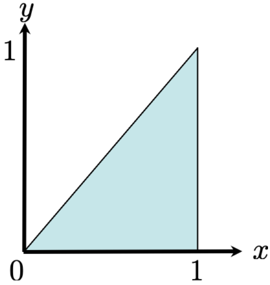
$$\mathrm{E}\left[g(X,Y)|Y=y\right] \;=\; \int_{-\infty}^{\infty} g(x,y)f_{X|Y}(x|y)\mathrm{d}x$$

If X and Y are independent, then

$$\mathrm{E}\left[X|Y=y\right] \;=\; \int_{-\infty}^{\infty} x\,f_{X|Y}(x|y)\mathrm{d}x = \int_{-\infty}^{\infty} x\,f_X(x)\mathrm{d}x = \mathrm{E}\left[X\right]$$

Example – conditional PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



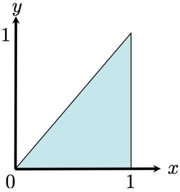
Find the conditional PDFs $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$.

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x = \int_y^1 2\,\mathrm{d}x = 2(1-y), \text{ for } 0 \le y \le 1$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y = \int_0^x 2\,\mathrm{d}y = 2x, \text{ for } 0 \le x \le 1.$$

Example – conditional PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \leq y \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



$$\begin{aligned} f_Y(y) &= 2(1-y)\,, \text{ for } 0 \leq y \leq 1 \\ f_X(x) &= 2x\,, \text{ for } 0 \leq x \leq 1 \end{aligned}$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{1}{x} & 0 \leq y \leq x \cancel{\leq 1} \\ 0 & \text{otherwise} \end{cases}$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{1-y} & \cancel{0 \leq} y \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

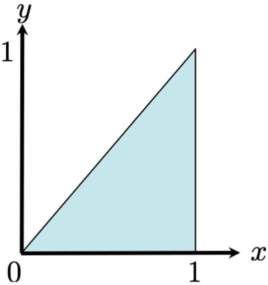
(uniform PDFs!)

Example – conditional PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{x} & 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{1-y} & y \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$



Interpretation:

- $x = 0.5$. Most likely value? $f_{Y|X}(y)$: any value $0 \le y \le 0.5$
- $x = 0.01$. Most likely value? $f_{Y|X}(y)$: any value $0 \le y \le 0.01$

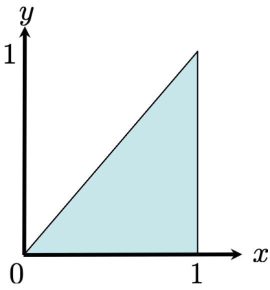
For dependent X and Y, knowledge of X changes knowledge on Y.

Example – conditional expected value

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{x} & 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{1-y} & y \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$



$$\begin{aligned} \mathrm{E}[X|Y=y] &= \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\mathrm{d}x = \int_y^1 \frac{x}{1-y}\mathrm{d}x = \left[\frac{x^2}{2(1-y)}\right]_{x=y}^{x=1} \\ &= \frac{1+y}{2} \end{aligned}$$

Conditional expectation

Notice the difference between

$$E[X|Y=y] = \frac{1+y}{2}$$

and

$$E[X|Y] = \frac{1+Y}{2}.$$

- $E[X|Y=y]$ is written in terms of the realization y, as the conditional information says $Y=y$.
- $E[X|Y]$ is still a RV because of the conditioning on Y: the PDF is $f_Y(y)$.

Theorem (iterated expectation): $E[E[X|Y]] = E[X]$.

Example – iterated expectation

Using the previous example,

$$\mathrm{E}[X|Y] = \frac{1+Y}{2}\,; \qquad f_Y(y) = 2(1-y)\,, \quad 0 \le y \le 1$$

- Iterated expectations gives

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X|Y]] = \int_{-\infty}^{\infty} \mathrm{E}[X|Y] f_Y(y)\mathrm{d}y \\ &= \int_0^1 \frac{1+y}{2} 2(1-y)\mathrm{d}y = \int_0^1 (1-y^2)\mathrm{d}y = \frac{2}{3} \end{aligned}$$

- Direct: with $f_X(x) = 2x$ $(0 \le x \le 1)$

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\mathrm{d}x = \int_0^1 2x^2 \mathrm{d}x = \frac{2}{3}$$

Bivariate Gaussian

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

$$= \cdots \text{eq.}(5.69) \cdots = \underbrace{\frac{e^{-(x-\mu_X)^2/\sigma_X^2}}{\sigma_X\sqrt{2\pi}}}_{f_X(x)} \cdot \underbrace{\frac{e^{-(y-\tilde{\mu}_Y)^2/\tilde{\sigma}_Y^2}}{\tilde{\sigma}_Y\sqrt{2\pi}}}_{f_{Y|X}(y|x)}$$

with $\tilde{\mu}_Y = \mu_Y + \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$, $\tilde{\sigma}_Y = \sigma_Y\sqrt{1-\rho_{X,Y}^2}$.

Given $X = x$, the conditional probability model of Y is Gaussian, with $\mathrm{E}[Y|X=x] = \tilde{\mu}_Y$ and $\mathrm{var}[Y|X=x] = \tilde{\sigma}_Y$.

Exercise 7.6.2

X and Y are jointly Gaussian random variables with $\mathrm{E}[X] = \mathrm{E}[Y] = 0$ and $\mathrm{var}[X] = \mathrm{var}[Y] = 1$. Furthermore, $\mathrm{E}[Y|X] = X/2$. Find $f_{X,Y}(x, y)$.

From the problem statement, we learn that

$$\mu_X = \mu_Y = 0\,, \qquad \sigma_X^2 = \sigma_Y^2 = 1\,.$$

From Theorem 7.16, the conditional expectation of Y given X is

$$\mathrm{E}[Y|X] = \tilde{\mu}_Y(X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) = \rho X$$

In the problem statement, we learn that $\mathrm{E}[Y|X] = X/2$. Hence $\rho = 1/2$. From the expression of the PDF of a bivariate Gaussian, the joint PDF is

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{3\pi^2}} e^{-2(x^2 - xy + y^2)/3}\,.$$

To do for this week:

- Read chapter 7, 8
- Make (some of) the indicated exercises:
  7.1.1, 7.2.3, 7.2.9, 7.3.1, 7.3.3, 7.3.5, 7.3.9, 7.5.1, 7.5.3, 7.5.5
  8.1.3, 8.2.3, 8.4.1, 8.4.3, 8.4.5