

Advanced probabilistic machine learning

Linus Falk

August 28, 2023

Contents

| | | |
|----------|--|----------|
| 1 | Introduction | 2 |
| 2 | Probability review | 2 |
| 2.1 | Conditional probability | 2 |
| 2.2 | Bayes theorem | 2 |
| 2.3 | Random variables | 3 |
| 2.4 | Joint probability | 4 |
| 2.5 | Marginalization and conditioning | 4 |
| 2.6 | Bayes theorem: random variables | 4 |
| 2.7 | Probabilistic modelling | 4 |
| 2.8 | Bayes theorem in Probabilistic modelling | 4 |
| 2.9 | Concluding remarks | 5 |
| 2.10 | Summary | 5 |

1 Introduction

This course bla bla bla

2 Probability review

Let first take a look at the *Sample space*:

- Dice: $\Omega = \{1,2,3,4,5,6\}$
- Coin flip: $\Omega = \{\text{heads, tails}\}$

There are also *Events* which is a subset of the sample space:

- Dice: $A = \{1,2\}$
- Coin flip = $\{\text{tails}\}$

Three important axioms to remember are.

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. For disjoint sets, A_1, A_2, \dots
 $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$

To sum it up here are some examples:

1. $A \subseteq B \Rightarrow P(A) \leq P(B)$
2. $P(A^c) = 1 - P(A)$
3. $P(\emptyset) = 0$
4. $0 \leq P(A) \leq 1$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

2.1 Conditional probability

The **definition** of conditional probability for two events are the following:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (1)$$

Provided that the probability of event A is greater than 0, $P(A) > 0$. From this we can derive the **product rule**: $P(A \cap B) = P(B|A)P(A)$

2.2 Bayes theorem

Given two events called A and B:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (2)$$

Example: medical test paradox

Let's characterize a medical test:

- Sensitivity (True positive rate) = $P(+|Disease)$
- Specificity (True negative rate) = $P(-|No Disease)$

Then characterize a medical condition in a population:

- Prevalence = "*proportion of a particular population found to be affected by a medical condition.*"

Question: Assume that for a certain disease the Sensitivity is 0.99, Specificity is 0.99 and Prevalence = 0.01. If you pick someone randomly from the population and test the person for this disease, obtaining a positive result. What is the probability of the person actually having the disease?

Using Bayes theorem:

$$P(Disease|+) = \frac{P(+|Disease)P(Disease)}{P(+)} = \frac{tpr \cdot p}{P(+)} \quad (3)$$

and by the law of total probability:

$$P(+) = P(+ \cap Disease) + P(+ \cap No disease)$$

$$P(+) = P(+|disease)P(disease) + (1 - P(-|No disease))(1 - P(disease))$$

$$P(+) = tpr \cdot p + (1 - tnr)(1 - p)$$

Hence we have:

$$P(disease|+) = \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.01 \cdot 0.99} = 0.5 \quad (4)$$

2.3 Random variables

The definition of a **random variable** \mathcal{X} is a quantity that depends on a random event. Where a **discrete random variable** is a countable number of values. Described by the probability mass function $p(x) = P(X = x)$. An **example:** 6-sided dice with $p(x) = 1/6, x = 1, \dots, 6$. A **continuous random variable** is an uncountable number of values. Ex: any value in \mathbb{R} . **Example:** \mathcal{X} is the height of an adult randomly sampled from the population. It is described by the probability density function: $p(x)$.

A **probability density function** $p(x)$ describes the probability for a *continuous* random variable x falling into a given interval:

$$P(a < X < b) = \int_a^b p(x) dx \quad (5)$$

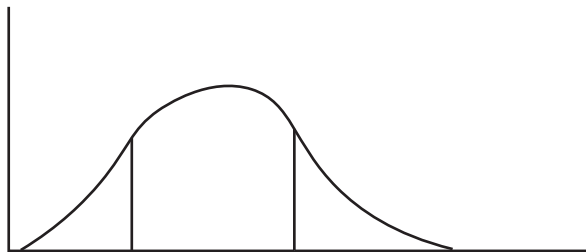


Figure 1: Example of caption

2.4 Joint probability

Given two random variables: X and Y :

- Discrete: probability mass function
 $p(x, y) = P(X = x, Y = y)$
- Continuous: (probability density function)
 $P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d p(x, y) dy dx$

2.5 Marginalization and conditioning

Marginalization (also called the **sum rule**) is defined as

$$\begin{aligned} p(\textcolor{red}{x}) &= \int_y p(\textcolor{red}{z}, \textcolor{blue}{y}) d\textcolor{blue}{y} \quad \text{if } \textcolor{blue}{y} \text{ is continuous} \\ p(\textcolor{red}{x}) &= \sum_{\textcolor{blue}{y}} p(\textcolor{red}{x}, \textcolor{blue}{y}) \quad \text{if } \textcolor{blue}{y} \text{ is discrete} \end{aligned} \tag{6}$$

Conditional probability also called the **product rule** is defined as:

$$\begin{aligned} p(\textcolor{red}{x}, \textcolor{blue}{y}) &= \frac{p(\textcolor{red}{x}, \textcolor{blue}{y})}{p(\textcolor{blue}{y})} \quad \text{where } p(\textcolor{blue}{y}) \neq 0 \\ &\Rightarrow p(\textcolor{red}{x}, \textcolor{blue}{y}) = p(\textcolor{red}{x}|\textcolor{blue}{y})p(\textcolor{blue}{y}) \end{aligned} \tag{7}$$

2.6 Bayes theorem: random variables

Both for probability mass functions and probability density functions:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{8}$$

2.7 Probabilistic modelling

In probabilistic modelling we consider two types of variables:

$$\begin{aligned} \mathcal{D} &= \{x_1, x_2, \dots, x_N\} : \quad \text{observed variables} \\ \Theta &= \{x_1, x_2, \dots, x_N\} : \quad \text{latent variables} \end{aligned} \tag{9}$$

In probabilistic modelling we treat *both* the **observed variables** and the **latent variables** as **random variables**

s0

We model this relationship between \mathcal{D} and Θ with its **Joint distribution**

$$p(\mathcal{D}, \Theta) = p(x_1, \dots, x_N, z_1, \dots, z_M) \tag{10}$$

2.8 Bayes theorem in Probabilistic modelling

The joint distribution factorizes into **likelihood** and a **prior**

$$p(\mathcal{D}, \Theta) = p(\mathcal{D}|\Theta)p(\Theta) \tag{11}$$

We can now find $p(\Theta|\mathcal{D})$ using **Bayes' theorem**

$$p(\Theta|\mathcal{D}) = \frac{p(\mathcal{D}|p(\Theta))}{p(\mathcal{D})} \quad (12)$$

- \mathcal{D} : observed data
- Θ : latent variables explaining the data
- $p(\Theta)$: **prior** belief of latent variables before seeing data
- $p(\mathcal{D}|\Theta)$: **likelihood** of the data in view of the latent variables
- $p(\Theta|\mathcal{D})$: **posterior** belief of latent variables in view of data
- $p(\mathcal{D})$: The **marginal likelihood** (presented in lecture 3)

Example: Coin flip (Bayesian viewpoint)

- $x \in \{0, 1\}$ represent the outcome of flipping a damaged coin
- $p(x = 1|\mu) = \mu$, which is also a random variable.
- Probability distribution $p(\mu)$: our *prior belief*

Question: After we observe N coin flips $\{x_1, \dots, x_N\}$, what is our belief $p(\mu|x_1, \dots, x_N)$?

Solution: Bayes theorem states that:

$$p(\mu|x_1, \dots, x_N) \propto p(x_1, \dots, x_N|\mu)p(\mu) \quad (13)$$

We will continue with this example next lecture...

2.9 Concluding remarks

Probabilistic/Bay inference is a flexible way of dealing the machine learning problems. Some properties to remember:

- Treat not only the data, but also the model and its parameters (if they are parametric) as random variables
- After learning you not only get a single model, you get a distribution of likely models.
- You can also encode prior knowledge you might have about the model and its parameters.

2.10 Summary

Probability distribution is a function that describes the likelihood of obtaining the possible values that a random variable can assume. **Conditional and marginalization** are two basic rules for manipulating probability distributions. **Frequentist vs Bayesian:** the first assume true values underlying some experiment and the second require some initial belief to be set on possible values.

- Bayes theorem: $p(x|y) = p(y|x)p(x)/p(y)$
- **Prior**, belief of parameters before we have seen any data
- **Likelihood**, belief of data in view of the parameters
- **Posterior**, belief of parameters after inferring data