ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Models, Algorithms and Data (MAD): Introduction to Computing

Spring semester 2019

Prof. Dr. Jens Honore Walther
Dr. Julija Zavadlav
ETH Zentrum, CLT
CH-8092 Zürich

Set 13

Issued: 24.05.2019; Due: 02.06.2019

In this exercise we will start applying some of the probability concepts introduced in the lecture to the Gaussian distribution. This distribution is of great importance, as many of the processes observed in nature follow it. The second part of the exercise introduces Bayes rule and applies it to an example of current relevance. The very last exercise will cover uncertainty quantification where we will assess the probability that experimental values obtained in a coin-toss experiment are consistent with our assumption that the coin is fair. To do this, let us start with some notation

Joint Distributions Until now we only regarded distributions of single random variables. If we now want to formalize the probability that several random variables take certain values we have to introduce the notation of joint distributions

$$\text{Prob. that } X_1, \ldots, X_n \text{ take the realisation } x_1, \ldots, x_n \equiv P(X_1 = x_1, \ldots, X_n = x_n) \quad (1)$$

Where depending whether we want to know the probability for a specific realization or for all realizations we usually neglect the "$X_i =$" or "$= x_i$" for notational convenience. We call random variables independent $X \perp Y$ if their joint probability distribution factors into the individual ones for all realizations

$$P(X, Y) = P(X)P(Y) \quad (2)$$

This can be readily be generalized to more than 2 random variables.

Marginalisation Rule Indeed it is often the case that we are given the joint distributions, but do not know the individual ones. Luckily they can be reconstructed from the joint distribution via the marginalization rule

$$P(X_i) = \sum_{j \neq i} P(x_1, \ldots, x_{i-1}, X_i, x_{i+1}, \ldots, x_n) \quad (3)$$

Where the sum over $j \in \{1, \ldots, i-1, i+1, \ldots, n\}$ is over all realizations of the respective random variables.

Conditional Probabilities As we are now having probabilities that several random variables take a specific realization we can also introduce the probability that a certain value x is taken, given that the another random variable took the value y. This probability is defined by

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} \quad (4)$$

Also here we have a notation of independence, two random variables are conditionally independent $X \perp Y | Z$, if their conditioned joint probability distribution factors

$$P(X, Y|Z) = P(X|Z)P(Y|Z) \tag{5}$$

Chain Rule The chain rule is a direct consequence of the definition of conditional probabilities. It allows us to factor joint probabilities into their conditionals

$$\begin{aligned} P(X_1, \ldots, X_n) &= P(X_1|X_2, \ldots, X_n)P(X_2, \ldots, X_n) \\ &= \ldots \\ &= P(X_1|X_2, \ldots, X_n)P(X_2|X_3, \ldots, X_n) \cdots P(X_n) \end{aligned} \tag{6}$$

We remark that this contains a certain ambiguity, as it is not a-priori clear how to factor the joint distribution. Another valid possibility would for example be

$$P(X_1, \ldots, X_n) = P(X_n|X_{n-1}, \ldots, X_1)P(X_{n-1}|X_{n-1}, \ldots, X_1) \cdots P(X_1) \tag{7}$$

Question 1: Gaussian Distribution

a) A random variable with normal (or Gaussian) distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ has probability density function (pdf) given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{8}$$

Show that the mean and the variance of X are given by $\mathbb{E}[X] = \mu$ and $\mathbb{E}[(X-\mu)^2] = \sigma^2$, respectively.

b) Consider the random variable

$$X = \sqrt{2}\,\mathrm{erf}^{-1}(2U - 1), \text{ where } U \sim \mathcal{U}(0,1).$$

The error function is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\mathrm{d}t.$$

Use the inversion method to show that $X \sim \mathcal{N}(0,1)$. Elaborate on why this method is not used in practice, but rather the Box Muller transform introduced in the script is used.

Question 2: Bayesian Inference

Bayes' rule reads as follows:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Here $P(X)$ is the *prior*, $P(Y|X)$ the *likelihood*, $P(Y)$ the *evidence*, and $P(X|Y)$ the *posterior*.

a) Given the fundamental properties in the preamble show that Bayes' rule holds.

Let us now apply this to the following hypothetical case where we are supposed to evaluate the safety of self-driving cars in a city. You are provided with statistics that are based on 100'000 cars, of which 10'000 are self-driving. The probability that a car hits a pedestrian given it is self-driving is 1/20'000, opposed to 1/15'000 for a human-driven car.

b) Now invoke Bayes' rule to calculate the probability that a car is self driving given a pedestrian is been git.
Hint: Consider the boolean (i.e. $\in \{0,1\}$*) variables* $X \sim \mathrm{Hit}$ *and* $Y \sim \mathrm{Self\text{-}Driving}$

c) Looking at Bayes' rule, and the more general case of continuous spaces where only the joint distribution over may random variables is known, where do you see an application for Monte Carlo quadrature?

Question 3: Uncertainty Quantification

In this exercise we will visit the coin-toss problem. We will denote the probability of obtaining head as q and resulting find the probability of obtaining tail as $1-q$

$$P(X = H) = q\,, \qquad P(X = T) = 1 - q \tag{9}$$

We now want to use our obtained knowledge on Bayes rule to estimate the probability q, thus allowing us to detect whether we have a fair coin. To do so we collect measurements $D = \{X_1, \ldots, X_n\}$ of coin tosses X_i and compute expected probability of obtaining head based on these measurements. To do so we invoke Bayes rule

$$P(q|D) = \frac{P(D|q)P(q)}{P(D)} \tag{10}$$

From this we can calculate the most probable q, which is simply the maximum of $P(q|D)$. As $P(D)$ does not depend on q it can be neglected in the following considerations. Thus in order to compute the most probable q we have to understand the other elements of the right hand side of the above equation. The prior $P(q)$ is our assumption of what q is. In the following we will assume no prior knowledge, thus assume that $P(q)$ is the uniform distribution (i.e. every value of q is equally likely)

$$P(q) = 1 \quad \text{for} \quad q \in [0,1] \tag{11}$$

The second term is the probability that we obtain our set of measurements assuming that q takes a certain value. We know that the probability distribution governing the coin-toss problem is the Binomial distribution, i.e. when tossing the coin n times the probability of obtaining k times head is given by

$$P(D|q) = \binom{n}{k} q^k (1-q)^{n-k} \tag{12}$$

Where the prefactor is the binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ giving the number of ways such an event can happen serves as normalization factor.

a) Calculate the most probable q analytically, i.e. calculate

$$q^* = \arg\max P(q|D) \tag{13}$$

Regard the following measurement sequence $D = \{H, T, H, H, T, H, H, H, T, H\}$. What is q^* in this case? Is the coin fair?
Hint: You might want to compose $P(q|D)$ *with some function before calculating the maximum.*

b) As we are regarding a probability distribution, we can also estimate the uncertainty about this found value q^*. Apply the Laplace approximation introduced in the lecture and estimate the standard deviation of the estimate based on the same dataset as in the previous question. *Hint: You might want to compose $P(q|D)$ with some function before computing the result.*

c) For the measurement series given above you can think of an experiment with a "real" biased coin. If we regard a case where we generate the samples on the computer, where do you see an application of the Monte Carlo methods?