P. Koumoutsakos
ETH Zentrum, CLT E 13
CH-8092 Zürich

HW 1 - Probability & Bayesian Inference

Issued: March 1, 2021
Due Date: March 15, 2021, 10:00am

1-Week Milestone: Solve tasks 1 and 2

Task 1: Probability Theory Reminders

In this exercise we fix the notation we will use during this course and refresh our memory on basic properties of random variables. Present your answers *in detail*.

a) [10pts] A random variable with normal (or Gaussian) distribution $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ has probability density function (pdf) given by

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

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

b) [10pts] The probability that a random variable X with pdf f_X is less or equal than any $x \in \mathbb{R}$ is given by,

$$P(X \leq x) \;=\; F_X(x) \;=\; \int_{-\infty}^{x} f_X(z)\, \mathrm{d}z\,. \tag{2}$$

The function F_X is called the cumulative distribution function (cdf).
The Laplace distribution with parameters μ and β has pdf,

$$f(x) = \frac{1}{2\beta} \exp\left(-\frac{|x-\mu|}{\beta}\right). \tag{3}$$

i) Find the cdf of the Laplace distribution.

ii) Use the cdf to find the median of the Laplace distribution.

c) [10pts] The pdf of the quotient $Q = X/Y$ of two random variables X, Y is given by,

$$f_Q(q) \;=\; \int_{-\infty}^{\infty} |x|\, f_{X,Y}(qx, x)\, \mathrm{d}x\,, \tag{4}$$

where $f_{X,Y}$ is the joint pdf of X and Y.
Assume that X and Y are independent random variables with pdfs $f_X(x) = \mathcal{N}\left(x|0, \sigma_X^2\right)$ and $f_Y(y) = \mathcal{N}\left(y|0, \sigma_Y^2\right)$.

i) Find the joint pdf of X and Y.

ii) Show that $Q = X/Y$ follows a Cauchy distribution with zero location parameter and scale $\gamma = \sigma_X/\sigma_Y$. The pdf of a Cauchy distribution with location parameter x_0 and scale γ is given by,

$$f(x) = \frac{1}{\pi} \frac{\gamma}{(x - x_0)^2 + \gamma^2} \,. \tag{5}$$

Task 2: Bayesian Inference

You are given a set of points $\boldsymbol{d} = \{d_i\}_{i=1}^N$ with $d_i \in \mathbb{R}$. You make the *modelling assumption* that the points come from N realisations of N *independent* random variables X_i, $i = 1, \ldots, N$, that follow normal distribution with unknown parameter μ and known parameter $\sigma = 1$.

a) [10pts] Formulate the *log-likelihood function* of μ,

$$\log \mathcal{L}(\mu) := \log p(\boldsymbol{d}|\mu), \tag{6}$$

where p is the conditional pdf of $\boldsymbol{d}$ conditioned on μ.

b) [15pts] Find the *maximum likelihood estimate* (MLE) of μ, i.e.,

$$\hat{\mu} = \arg\max_{\mu} \ \mathcal{L}(\mu). \tag{7}$$

You may find useful that $\arg\max_\mu \mathcal{L}(\mu) = \arg\max_\mu \log \mathcal{L}(\mu)$.

c) [30pts] Before observing any data $\boldsymbol{d}$ you had the belief that μ follows a normal distribution with mean μ_0 and variance σ_0^2. After observing the dataset $\boldsymbol{d}$ you *update your belief* by using Bayes' theorem. Identify the *posterior distribution* $p(\mu|\boldsymbol{d})$ of μ conditioned on $\boldsymbol{d}$. Calculate the mean and the variance of $p(\mu|\boldsymbol{d})$.

d) [5pts] Find the *maximum a posteriori* (MAP) estimate of μ, i.e.,

$$\hat{\mu} = \arg\max_{\mu} \ p(\mu|\boldsymbol{d}). \tag{8}$$

e) [10pts] Perform (c) and (d) using as prior an uninformative distribution, i.e. a uniform distribution in $\mathbb{R}$, and compare the MAP with the MLE. Although this is not a distribution, since it is not integrable over $\mathbb{R}$, we are allowed to use it in Bayes' theorem as long as the posterior is a distribution. These priors are called *improper priors* and a common choice in practical applications when there is no prior information on the parameters.

Task 3: Bayesian Inference: Linear Model

You are given the linear regression model that describes the relation between variables x and y,

$$y = \beta x + \epsilon,$$

where β is the regression parameter, y is the output quantity of interest (QoI) of the system, x is the input variable and ϵ is the random variable accounting for model and measurement errors. The error is modeled by a Gaussian distribution $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

You are given one measurement data point, $D = \{x_0, y_0\}$.

a) [20pts] Consider an uninformative prior for β and identify the posterior distribution of β after observing D. Calculate the MAP and the standard deviation of $p(\beta|D)$.

b) [20pts] Now, consider a Gaussian prior for β with mean 0 and variance τ^2, i.e. $\beta \sim \mathcal{N}(0, \tau^2)$, identify the posterior distribution $p(\beta|D)$. This form of regression is also known as Bayesian linear regression.

Guidelines for reports submissions:

- Submit a pdf file of your solution via Moodle until March 15, 2021, 10:00am.