Freie Universität Berlin
Liu-Wei Wang, Max von Kleist

9. Homework
Foundations of Mathematics and Statistics

Deadline: January 29, 10:00 (**before** the lecture)

The homework should be worked out in groups. Pen & paper exercises will be discussed on the board. Python programs must be submitted via Whiteboard, plots printed and handed in (write the names of the group members and their student numbers on the sheet).

Homework 1 (Bayesian Optimization, Programming counts $1/6 + 1/6$)
Consider the structural model $\mathcal{M}$:

$$x(t_i) = \frac{\lambda}{\delta}\left(1 - e^{-\delta t_i}\right).$$

The likelihood of observing y for a parameter choice $\lambda, \delta > 0$, is given by,

$$\mathcal{L}(y|\lambda, \delta) = \prod_i \frac{x_i^{y_i} e^{-x_i}}{y_i!}.$$

Given the measurements are $y = \{10, 10, 11, 8, 9, 7, 8, 9, 9, 10\}$ all taken at the same time $t = 100$.
a) (**to be uploaded via Whiteboard**) write a program "Exc9Task1.py" and submit it via Whiteboard. The program should do the following: It reads an Input file provided in whiteboard ("Input.txt"). The input file contains two numbers, which denote parameter values for λ and δ.
(a.i) Using these parameters, compute the negative log likelihood of the data.
(a.ii) Consider the following *gaussian* priors (assuming the two parameters are uncorrelated)

$$p(\theta_j) = \frac{1}{\sqrt{(2\sigma_j^2 \cdot \pi)}} e^{-\frac{(\theta_j - \mu_j)^2}{2 \cdot \sigma_j^2}}$$

with $\mu_1 = 15$ and $\sigma_1^2 = 5$ for $\theta_1 = \lambda$ and $\mu_2 = 1$ and $\sigma_2^2 = 10$ for $\theta_2 = \delta$. Compute the negative log prior probability of the parameters.
(a.iii) Compute the unnormalized negative log posterior $p(\lambda) \cdot p(\delta) \cdot \mathcal{L}_y(\theta) = p(\theta) \cdot \mathcal{L}_y(\theta)$.
Save the **negative (natural) log** of these three numbers (tasks a.i-iii) into a file "Exc11Task1.txt" in the comma-separated text format using two digits after the comma (format '%1.2f'), e.g.

$$23.31 \tag{1}$$
$$56.34 \tag{2}$$
$$107.43 \tag{3}$$

where the first number corresponds to the negative log-likelihood, the second to the negative log prior, the third is the negative log (unnormalized) posterior.
b) (**to be printed and discussed**) Make a contour plot of the **negative (natural) log** likelihood, for the given data, with axis $\lambda \in [5, 30]$ and $\delta \in [0.5, 3]$ plotting the contour levels from 20 to 50 (e.g. levels = np.arange(20, 50, 2)). On the contour plot, plot the line $(\lambda, 0.1 \times \lambda)$ for values $\lambda \in [5, 30]$. Now, compute the **negative (natural) log** of the unnormalized posterior $p(\lambda) \cdot p(\delta) \cdot \mathcal{L}_y(\theta) = p(\theta) \cdot \mathcal{L}_y(\theta)$ and make a contour plot of it, for the given data.
What can you say about the region of minimum, or about the posedness of the problem?

Homework 2 (MCMC, Programming counts 1/3 + 1/3)
Implement a Metropolis-Hastings algorithm to sample from the posterior probability using the
model, data, likelihood function and prior from task 2.
a) (**to be uploaded via Whiteboard**) Write a program "Exc9Task2.py" that samples from
the posterior probability. Set the seed to the number contained in Input2.txt. I.e. load "In =
np.loadtxt('Input2.txt')" and set "np.random.seed(seed=int(In[0]))". Perform 10 iterations and
use a Gaussian proposal function with standard deviation $\sigma = 3$. I.e.: $\theta' = \mathcal{N}(\theta., \sigma)$. Use start
parameters $\theta = (30, 3)$. Reject all negative parameters and write derived parameters into file
"Exc9Task2.txt" using two digits after the comma.
b) (**to be printed and discussed**) Perform > 100000 iterations using your posterior sampler.
Disregard the first 90% of the derived parameters as a 'burn-in'.
Make a histogram of the marginal empirical distributions of derived parameters for λ and δ
respectively and make a contour or kernel density plot of the joint distribution of the parameters.
Randomize your initial parameters. Do you get the same posterior parameter distributions? Why,
or why not?
How are the parameter distributions related to task 1?

Homework 3 (KL divergence; Bernulli trial, pen & paper)
You repeated a single coin flip 20 times and computed the frequency of 'heads', $f(\text{heads}) = e^{-1}$.
a) For parameter $\theta = p(\text{heads}) = 0.5$, compute the Kullbach-Leibler divergence by hand.
b) Using KL divergence, what is the maximum likelihood parameter estimate θ^*?

Homework 4 (Maximum a posteriori, pen & paper)
Given some likelihood function $\mathcal{L}_{y|\theta}$ and a multidimensional La Place distributed prior $v(\theta)$ that
is centered on $\mu_j = 0 \quad \forall\, \theta_j$. Write down the maximum a posteriori (MAP) problem and explain,
which regularization of a maximum likelihood (ML) problem this corresponds to.