05. Bayesian Inference

To be completed Friday 25th of November, by 12 noon sharp. You are encouraged to submit in pairs, but are also allowed to submit alone. **One person** must submit the solution as **a single pdf file**, to the folder on Ilias, by naming *exercise_03_name1_matriculationnumber1_name2_matriculationnumber2.pdf*, with the obvious replacements in the strings. **Not adhering to the formatting requirements may result in the submission not being graded.**
Exercises are only graded in a binary fashion as sufficient or insufficient. To be graded as sufficient, you do not necessarily have to have correct solutions to every sub-question, but you must have made a clear and earnest effort to solve the entire exercise. Ultimately, what constitutes sufficient is at the discretion of the tutors. To be admitted to the exam, you must have submitted sufficient answers to at least 5 of the 6 (maybe 7) exercise sheets.

1. **EXAMple Question — Roll the dice**

   You find a strangely shaped six-sided dice and wonder what the probability of throwing each side is. You throw the dice N times and count the occurrences of each side. For inference, we can thus use a Multinomial likelihood, which generalizes the Bernoulli likelihood to non-binary events

   $$p(x|\boldsymbol{\theta}) = Mul(x; N, \boldsymbol{\theta}) = \frac{N!}{\prod_{i=1}^{6} x_i!} \prod_{i=1}^{6} \theta_i^{x_i} \text{ with } \sum_{i=1}^{6} x_i = N \text{ and } \sum_{i=1}^{6} \theta_i = 1$$

   Equivalently the Dirichlet distribution is a generalization of the Beta distribution and thus well suited as a prior.

   $$p(\theta) = Dir(\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{6} \theta_i^{\alpha_i - 1}$$

   here $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_6)$ and $B(\boldsymbol{\alpha})$ is the beta function. Calculate the posterior distribution.

2. **Theory Question — Gaussian posterior**

   Assume we want to measure some quantity x^*. Unfortunately, your sensor is imprecise, and we observe only noisy measurements x_i

   $$p(x_i|x^*) = \mathcal{N}(x; x^*, 1.) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-x^*)^2}$$

   We want to infer x^* and thus impose a prior $p(x^*) = \mathcal{N}(x^*; 0, \sigma_0^2) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2} x^{*2}}$. This is a conjugate prior for the Gaussian likelihood, and thus the posterior will also be Gaussian with parameters

   $$p(x^*|x_1, \ldots, x_N) = \mathcal{N}(x^*; \mu_p, \sigma_p) \qquad \mu_p = \sigma_p^2 \left(\sum_{i=1}^{N} x_i \right) \quad \text{and} \quad \sigma_p^2 = \frac{1}{\frac{1}{\sigma_0^2} + N}$$

   (a) Verify that this is the correct posterior by calculating $p(x^*|x_1, \ldots, x_N)$.
   Hint: Showing it up to proportionality is sufficient.

   (b) Consider the posterior mean μ_p as an estimator for x^*. Compute the bias and variance of this estimator. When is the bias/variance large, and when is it small?
   Hint: You can use: $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ and $Var[aX + b] = a^2 Var[X]$ for any non random a, b.

3. **Practical Question — Tweets of Donald Trump**
   Today we will look at the tweets of ex-president Donald Trump. We will perform Bayesian inference on a very simplified generative model for this data. Open the notebook, to continue.