# MAD 2018-19, Assignment 5

Jonas Peters, Fabian Gieseke, Aasa Feragen

hand in until: 09.01.2019 at 23:59

**General comments:** The assignments in MAD must be completed and written individually. You are allowed (and encouraged) to discuss the exercises in small groups. If you do so, you are required to list your group partners in the submission. The report must be written completely by yourself. In order to pass the assignment, you will need to get at least 40% of the available points. The data needed for the assignment can be found in the assignment folder that you download from Absalon.

**Submission instructions:** Submit your report as a PDF, not zipped up with the rest. Please add your source code to the submission, both as executable files and as part of your report. To include it in your report, you can use the `lstlisting` environment in LaTeX, or you can include a "print to pdf" output in your pdf report.

## Coin Game in Python

**Exercise 1** (**Based on Exercises 3.1, 3.2 and 3.3 in the book, 4 points**). *Consider the coin game discussed in L10. You will compute the posterior distribution $p(r|y_N)$ for three different priors:*

*a) For $\alpha = \beta = 1$, the beta distribution becomes uniform between 0 and 1. In particular, if the probability of a coin landing heads is given by $r$ and a beta prior is placed over $r$, with parameters $\alpha = 1 = \beta$, then this prior can be written as:*

$$p(r) = 1 \quad (0 \leq r \leq 1).$$

*Using this prior, compute the posterior density for $r$ if $y$ heads are observed in $N$ tosses (i.e. multiply this prior by th ebinomial likelihood and manipulate the result to obtain something that looks like a beta density).*

*b) Repeat exercise a) for the following prior, also a particular form of the beta density:*

$$p(r) = \begin{cases} 2r & 0 \leq r \leq 1 \\ 0 & o.w. \end{cases}$$

*What are the values of the prior parameters $\alpha$ and $\beta$ that result in $p(r) = 2r$?*

*c) Repeat exercise a) for the following prior, again a particular form of the beta density:*

$$p(r) = \begin{cases} 3r^2 & 0 \leq r \leq 1 \\ 0 & o.w. \end{cases}$$

*What are the prior parameters here?*

Deliverables. *a) The posterior expression, b) the posterior expression and the parameter values, c) the posterior expression and the parameter values.*

**Exercise 2** (**Based on Exercise 3.8 in the book, 4 points**). *Use Python to generate coin tosses where the probability of heads is 0.7, and use it to generate 1000 tosses. For each of the three priors in the previous exercise, plot the prior distribution as well as every hundredth posterior distribution (i.e., $p(r|y_{100})$, $p(r|y_{200})$, $\ldots$, $p(r|y_{1000})$).*

Deliverables. *Your source code, and the three figures generated.*

## Probabilistic regression

**Exercise 3** (**Bayesian Regression, 4 points**). *In this exercise, we will revisit linear regression from a Bayesian perspective. You will imitate the Bayesian regression from Lecture 11, but apply it to the Olympic 100m dataset used in the book (found in the file `men-olympics-100.txt`). Here, you should use the years (found in the first column) as your input $x_n$ and the winner times (second column) as your output $t_n$.*

a) *Assume that the noise in your model is normally distributed with zero mean and variance* $\sigma^2 = 10$. *What is the likelihood* $p(\mathbf{t}|\mathbf{w}, \mathbf{X}, \sigma^2)$ *of observing the given output* $\mathbf{t}$ *given the model defined by* $\mathbf{w}$ *when the input data matrix is* $\mathbf{X}$*?*

b) *Assume that the prior distribution for your model parameters* $\mathbf{w}$ *is* $p(\mathbf{w}) = \mathcal{N}(\mu_0, \mathbf{\Sigma}_0)$, *and let the likelihood* $p(\mathbf{t}|\mathbf{w}, \mathbf{X}, \sigma^2)$ *of observing the data matrix* $\mathbf{X}$ *given model parameters* $\mathbf{w}$ *be given by* $p(\mathbf{t}|\mathbf{w}, \mathbf{X}, \sigma^2) = \mathcal{N}(\mu_l, \mathbf{\Sigma}_l)$. *What is the corresponding posterior distribution* $p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \sigma^2)$*?*

c) *Set your prior parameters to be* $\mu_0 = [0, 0]^T$ *and* $\mathbf{\Sigma}_0 = \begin{bmatrix} 100 & 0 \\ 0 & 5 \end{bmatrix}$. *Assume that your prior and likelihood are both normally distributed as in b). Implement a function that computes the corresponding posterior probability density.*

d) *What is the mean and covariance of the posterior probability density after seeing the entire dataset?*

Deliverables. *a) A formula, b) a formula, c) your source code, d) the mean and covariance.*

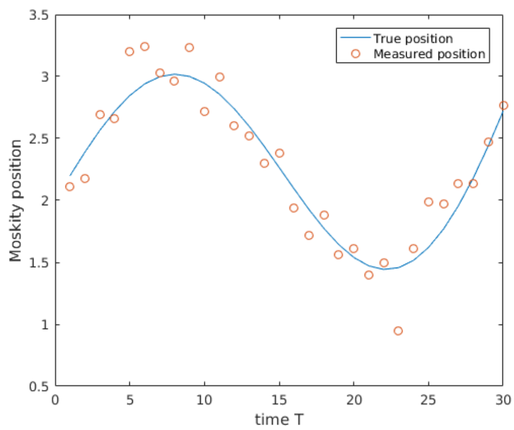
## Generalization to a new problem

In the final exercise, you will apply what you have learned about Bayesian analysis to a new problem: Tracking an object over time.

**Exercise 4 (Moskito killing, 4 points).** *Moskitos are very annoying.*

*You are going to design a moskito killing device that tracks the location of a moskito over time* $t = [0, 1, \ldots, T]$ *from noisy measurements, and tries to kill the moskito by building a predictive model for the moskito's location* $P_t$ *at any time step* $t \leq T$. *An explosive missile will be fired to explode at your predicted location* $P_t$ *no later than time* $t = T$, *so it is essential that you correctly estimate the moskito's location.*

*Your moskito killer utilizes a computer vision-based measurement device that feeds you a measurement* $x_t$ *of the moskito's position in mm at any time* $t \leq T$. *You know that the measurement has normally distributed noise with zero mean and standard deviation* $\sigma = 0.2$ *mm, which means that firing at the (incorrectly) measured location will sometimes cause you to miss the moskito by a lot – and potentially harm something non-moskito.*




*Throughout the exercise, we assume that the world is* 1*-dimensional, and that you are tracking the moskito's location on the real line:* $P_t \in \mathbb{R}$.

***Defining a model:*** *Our experimental setup is modelled through the equation*

$$x_t = P_t + \varepsilon_t, \text{ where } \varepsilon_t \sim \mathcal{N}(0, \sigma^2),$$

*which means that the observation* $x_t$ *is the sum of the location predicted by our model* $P_t$ *and the known noise from the measurement device. This gives us a likelihood for the observation* $x_t$*:*

$$p(x_t|P_t) = \mathcal{N}(P_t, \sigma^2).$$

*We assume, moreover, that the* $\varepsilon_t$ *are IID.*

*We will be interested in the distribution over the location* $P_t$ *at time* $t$, *given our measurement* $x_t$ *at time* $t$ *and our distribution over the location* $P_{t-1}$ *at the previous time step* $t-1$. *Using Bayes rule, we obtain:*

$$p(P_t|P_{t-1}, x_t) \propto p(x_t|P_t, P_{t-1})p(P_t|P_{t-1}) = p(x_t|P_t)p(P_t|P_{t-1}), \tag{1}$$

*where the first factor is the likelihood that we just obtained, and the second factor is a prior.*

***Defining a prior:*** *We expect the fly to move in a relatively smooth way. Thus, at each time step $t$, we will use a prior distribution centered at the predicted location $P_{t-1}$ at time $t-1$. For the variance, we want to retain the uncertainty of our prediction at time $t-1$, which is given by the posterior variance $\sigma^2_{posterior}(t-1)$ at the previous time step $t-1$. However, the fly does not stand still, and we thus want to add a "drift" to the variance, which captures the expected change in location. Summarizing, we will use the prior*

$$p(P_t) = \mathcal{N}(P_{t-1}, \sigma^2_{prior}), \quad \sigma^2_{prior} = \sigma^2_{posterior}(t-1) + \sigma^2_{drift}.$$

*A good choice for the parameter $\sigma_{drift}$ relates to the derivative of the curve; in this exercise we will start out by setting $\sigma_{drift} = 0.1$ mm.*

***Computing a posterior:*** *Let's denote by $f_{\mu,\sigma^2}$ the PDF of $\mathcal{N}(\mu, \sigma^2)$. Now, using Eq. (1) and our defined likelihood and prior from above, we obtain*

$$p(P_t|P_{t-1}, x_t) \propto f_{P_t,\sigma^2}(x_t) f_{P_{t-1},\sigma^2_{prior}}(P_t) = f_{x_t,\sigma^2}(P_t) f_{P_{t-1},\sigma^2_{prior}}(P_t) = f_{\mu_{post},\sigma^2_{post}}(P_t),$$

*where, by the formula for the product of two normal distributions, we have*

$$\mu_{post} = \frac{x_t \sigma^2_{prior} + P_{t-1}\sigma^2}{\sigma^2 + \sigma^2_{prior}},$$

*and*

$$\sigma^2_{post} = \frac{\sigma^2 \sigma^2_{prior}}{\sigma^2 + \sigma^2_{prior}}.$$

*Your task is to implement this model, and to analyze how it is affected by different parameter choices.*

a) *Your first task is to estimate the moskito location at time 0. As we have no idea where the moskito is before seeing any data, we make an initial guess $P_{-1} = 0$ for the moskito's location, with initial variance $\sigma^2_{posterior}(-1) = 100$. Please compute the posterior distribution over the location $P_0$ and include its mean and variance in your report. Please comment on what you see.*
b) *Implement an iterative algorithm that computes the posterior distribution over the location $P_t$ as explained above, for all $t = 0, \ldots, T$ for a given time $T$. Apply this to the moskito data found in `moskito_measurements.txt` for $T = 29$, and plot a curve consisting of your posterior mean at any time $t = 0, \ldots, T$. In the plot, also include a scatter plot of the measurements as well as a plot of the true locations from `moskito_true.txt`. What do you see?*
c) *With these settings, how long do you have to track the moskito before you are able to kill it, if the reach of the moskito missile is 0.1 mm?*
d) *What happens if you increase the prior variance by setting the drift $\sigma^2_{drift} = 0.3^2$? What do you see? How long does it now take you to kill the moskito?*
e) *What is the effect of increasing the standard deviation on your likelihood?*
f) *Although the use of the previous measurements of the moskito locations helps smoothen your estimate of its current location, this tracking algorithm is still relatively naïve. What might improve the algorithm?*

Deliverables. *a) The mean and standard deviation of the posterior distribution of the moskito's location at time $t = 0$, and a comment; b) your source code (snippet in report and source code submitted separately), a plot and a comment; c) a killing time point; d) a comment and a new killing time point; e) a comment and explanation; f) a description of your improvement idea.*