# Exercise Sheet 2
## Continuous problems

**Solutions**. You do not need to submit your exercise solutions to us. Of course you are free to ask questions during tutorial hours! Exercise solutions will be published on Moodle at the same time with the exercise sheet. Use solutions responsibly, otherwise you risk not being able to solve the exam to a satisfactory level.

**Need more exercises?** Look at chapters 3-5 of the book *Bayes rules!* (https://www.bayesrulesbook.com)

## Exercise 1 Prior elicitation

(this is the slightly adapted exercise 3.1 from Bayes rules!)

In each situation below, tune a Beta($\alpha, \beta$) model that accurately reflects the given prior information and visualize it. In many cases, there's no single "right" answer, but rather multiple "reasonable" answers.

**Hint:** Often you can constrain the possible values for $\alpha$ and $\beta$ using a given expectation or variance:

$$E[\pi] = \frac{\alpha}{\alpha + \beta}, \quad \mathrm{Var}[\pi] = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}. \tag{1}$$

This might save you time when trying out different values for $\alpha$ and $\beta$. If you're not in the mood for calculations however, just try them out!

a) Your friend applied to a job and tells you: "I think I have a 40% chance of getting the job, but I'm pretty unsure." When pressed further, they put their chances between 20% and 60%.

b) A scientist has created a new test for a rare disease. They expect that the test is accurate 80% of the time with a variance of 0.005.

c) Another scientist in the same field claims that he expects the same accuracy, however a ten times larger variance of 0.05. Can the beta distribution model this prior expectation well?

d) Your Aunt Jo is a successful mushroom hunter. She boasts: "I expect to find enough mushrooms to feed myself and my co-workers at the auto-repair shop 90% of the time, but if I had to give you a likely range it would be between 85% and 100% of the time."

e) Sal (who is a touch hyperbolic) just interviewed for a job, and doesn't know how to describe their chances of getting an offer. They say, "I couldn't read my interviewer's expression! I either really impressed them and they are absolutely going to hire me, or I made a terrible impression and they are burning my resumé as we speak."

If you are in the mood for more prior elicitation, you may also complete exercise 3.2 in Bayes rules!

## Exercise 2 Älplermagronen

Heidi and Peter both like Älplermagronen very much (Swiss dish with maccheroni, potatos, onions, cream, a lot of cheese and apple puré). However, they strongly disagree and often argue about one critical point: whether the apple puré should be mixed with the rest or be served separately.

- Heidi who does not like mixing says: "I'm very sure that almost nobody mixes the apple puré with the rest! I would say that around 5% of the population are doing it, maybe up to 10%, but certainly not 20%!"
- Peter who likes it very much argues: "I do not believe this! I would guess that the ratio is about fifty-fifty!" (he seems however less sure in his statement than Heidi)

Soonafter, Peter participates in a dinner, where he is the only one out of 6 people who mixes their apple puré with the rest.

a) What conjugate family do you choose to model the proportion $q$ of the Swiss population that mixes their apple puré with the rest? Elicit priors that reflect the opinions of Heidi and Peter well and visualize them (e.g. with PreliZ).

b) Compute the parameters of Heidi and Peter's posterior distributions. How do their opinions change in the light of data? Compare to their prior distributions.

c) How much would it change for Heidi's posterior opinion if instead of just Peter all of the 6 participants would have mixed? Could Peter convince her with this data? If no, how many people would they have needed to be (all mixing) to increase the expectation of Heidi's opinion to at least 50%?

## Exercise 3 Maths of the beta-binomial family

In this exercise, you are going to explore some of the maths of the beta-binomial family, just to get to know its inner workings a bit better.

a) The beta-family of distributions parameterized by $\alpha$ and $\beta$ is given by

$$p(\pi; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\pi^{\alpha-1}\,(1-\pi)^{\beta-1}.$$

Compute its mean $E[\pi] = \int_0^1 \pi\; p(\pi; \alpha, \beta)\;\mathrm{d}\pi$.

**Hints:** Use that $\displaystyle\int_0^1 p(\pi; \alpha, \beta)\;\mathrm{d}\pi = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and that $\Gamma(z+1) = z\,\Gamma(z)$.

b) If a Beta$(\alpha, \beta)$ prior is used and $n$ experiments are run, resulting in $k$ positive outcomes, then we update our belief to a Beta$(\alpha+k, \beta+n-k)$ posterior.

Show that the posterior mean $E[\pi|Y=k]$ can be decomposed into a weighted sum of the prior mean $E[\pi]$ and the empirical mean $\bar{y} = \frac{k}{n}$:

$$E[\pi|Y=k] = \frac{\alpha+\beta}{\alpha+\beta+n}\;E[\pi] + \frac{n}{\alpha+\beta+n}\;\bar{y}.$$

**Hint:** Make the ansatz

$$\underbrace{\frac{\alpha+k}{\alpha+\beta+n}}_{E[\pi|Y=k]} = w\,\underbrace{\frac{\alpha}{\alpha+\beta}}_{E[\pi]} + (1-w)\,\underbrace{\frac{k}{n}}_{\bar{y}},$$

solve for $w$ and compute $1-w$.

c) What are the limits $\lim_{n\to 0} E[\pi|Y=k]$ and $\lim_{n\to\infty} E[\pi|Y=k]$? Give an interpretation. How do you need to choose $\alpha$ and $\beta$ such that data dominate less for large $n$?

## Exercise 4 Playing around with your magic coin

The posterior is a balance between the prior and the data (through likelihood). In this exercise you will run some simulations to demonstrate this.

You have bought a magic coin with inherent long-run probability for heads $\pi = 0.7$. This is indicated on the wrapping and you got assured by the magic store owner that this value can be trusted (in the long-run of course). You choose to play a bit around with the magic coin to explore the properties of the beta-binomial conjugated family.

a) Create three datasets by running three simulations with $n_1 = 10$, $n_3 = 30$ and $n_2 = 100$ coin tosses. Store the values of $k$ (that vary each time you re-run the code). Compute the frequentist estimate $\hat{\pi}_i = \frac{k_i}{n_i}$ for each dataset.

b) You want to try four different priors:

- Uninformed prior: Beta(1,1)
- Weakly-informed prior: Beta(10,4.5)
- Strongly-informed prior: Beta(184, 80)
- wrong, strongly-informed prior: Beta(80, 184)

Visualize these four priors using e.g. PreliZ!

c) Copy the function `plot_beta_binomial()` from `week2_tongue_rolling.ipynb` (week 2 materials) to your notebook and run the definition so that you have access to it later. Now create three plots using `plot_beta_binomial()` for the three different datasets using the uninformed Beta(1,1) prior. Can you reproduce some conclusions from the lecture? Do the same for the weakly-informed, the strongly-informed and the wrong, strongly-informed prior and provide comments to what you observe.

Since you sampled the coin tosses randomly, you may re-run the cells for this exercise in your notebook - you may be surprised how some results vary by quite a bit!

## Exercise 5 Exam absence

A lecturer wants to model the number of absences due to sickness at the final exam of a linear algebra module. He conducted the following measurements so far (HS=Herbstsemester, FS=Frühlingssemester):

| semester | HS 19 | FS 20 | HS 20 | FS 21 | HS 21 | FS 22 | HS 22 | FS 23 | HS 23 | FS 24 |
|---|---|---|---|---|---|---|---|---|---|---|
| absences | 0 | 2 | NA | 5 | 2 | 1 | 0 | 1 | 1 | 2 |

In HS 20, the students were relieved of the exam during the onset of the Covid-19 pandemic.

a) Due to previous experiences (however without recording data), he thinks that typically one absence needs to be expected and the he will probably not reach the maximum of 5 again (during the pandemic).

   Use a library such as PreliZ to devise an appropriate gamma distribution that reflects this prior opinion.

b) What's the lecturers posterior distribution for the number of absences? Use the gamma-Poisson conjugate family and its update rule. Compute also posterior mean, mode and standard deviation as summaries.

c) Unlike the lecturer, you do not trust closed-form maths and want to do a simulation to compute posterior mean and standard deviation. You plan to implement it in a similar way as you simulated a beta-binomial posterior in the previous exercise:

   1. Draw $N = 10'000$ samples $\lambda_i$ from the prior distribution.
   2. Draw 9 samples of absences $k$ for each $\lambda_i$ using a Poisson likelihood (same sample number as in provided data).
   3. Filter out only samples that are the same as recorded data (order doesn't matter, you may use sorting)

   Argue that this simulation is very inefficient and might not produce the results you desire, getting even worse if you collect more data. Code is not required, you might nevertheless try it!

## Exercise 6 Empirical Bayes

For 'original' Bayesians the case is clear: the prior comes before any data and it is not allowed to leak any information from the data into the prior! (you could compare this with leaking test data into a training set in machine learning)

However, there is not just one faction of Bayesians. *Empirical* Bayesians (calling their approach **empirical Bayes** argue that it does not matter so much if enough data are provided and only summaries of the data are used as input to the prior. In this exercise you will use such an ansatz to solve the previous exercise without the need to 'painfully' elicit a prior before (it will get worse, you just had to do it for one parameter - what about a regression model with 40 parameters?).

a) Expectation and variance of the gamma distribution are given as

$$\mu = E[\lambda] = \frac{s}{r}, \quad \sigma^2 = \mathrm{Var}[\lambda] = \frac{s}{r^2}.$$

Compute $s$ and $r$ as a function of $\mu$ and $\sigma$.

b) Compute $\mu$ and $\sigma$ from the data (see previous exercise) and create and visualize the resulting empirical prior and compare it to your prior distribution elicited in the previous exercise.

c) Use the data to update your prior and visualize the resulting posterior. Compare it to the posterior you got in the previous exercise. How big is the difference? Should you curse empirical Bayesians or do they maybe have a point to make a Bayesian analysis a bit less rigoros but more simple?

We will use empirical Bayes with Bayesian linear regression models starting from week 6!