Tutoring Session 3

Parameter Inference

1 The importance of a good prior

We have seen that a prior can help mitigate overfitting of the maximum likelihood estimate. But setting a prior causes *inductive bias*: Certain solutions are preferred over others for subjective reasons. (*Subjective* means they are not motivated purely mathematically. They may be objective or reasonable from our intuitive understanding of the problem.)

Often, this is desired—certain model parameters indeed are more likely "from experience". In this exercise, however, we will see how a sloppy choice of a prior can impose a harmful inductive bias.

Problem 1: You are visiting Alice's casino. You have been gambling in this casino for years, and you have no doubt in Alice's integrity. Today when you arrive at the casino, she is in hospital and her son Bob has taken over the casino. You don't know Bob much, but him being Alice's offspring, you are sure you can trust him just as much. As an eager student of statistics, who just learned about priors, you decide to test him and walk up to your favourite game: Guess the flip! You like the elegant simplicity of the game: You place a bet on the outcome of a coin flip.

Taking into account Bob's splendid family background, you choose a centred Beta distribution as a prior, i.e., parameters $a = b = n > 0$. What you don't know is that Bob is trying to make most money out of his short intermission as the boss of the casino. Not being the most clever guy, he has decided to use coins that *always* show up tails.

Determine how long in terms of n and N it takes to recover from your overwhelming trust:

- Determine ML, MAP, and fully Bayesian estimates for θ, the probability of tails showing up next.
- Interpret these results. Which estimate takes longest to recover? Why? Is this expected from the results in the lecture?

2 Mark, you, and the coin

Once again, you're sitting in Mark's office. This time however, it is not Mark who sits in the nice leather chair, no, it is his speaking parrot Zucky. Mark himself is having a relaxing bath in his personal spa right next to the office. You can't see him, but because of a funny *bling* noise coming from the spa you know that Mark is engaged in his favourite past time, tossing gold coins. Mark shouts: "Dude, last time we talked, you seemed to have a knack for golden coin tossing, that's why you are here. I wonder if this coin here is biased. Let me flip it a couple of times for you and you tell me what you think!" He starts tossing (according to the *blings*) so you shout back: "Yo, Mark, you know, would be nice if I could *see* every toss... Maybe you can shout what every toss results in?" Mark starts: "Ok, ehrrr, no, sorry, need to call my friend Larry. Let's do it like this then: I toss, and Zucky will tell you the result. Oh, wait, right, Zucky finds it funny to tell sometimes, ehrrr, not the truth, don't know where he picked that habit. Check out the sample run I did with him yesterday, it is the paper lying right next to you. I'll start tossing when you're done with your math magic, just let me know, Zucky and I are waiting. Yo, Larry, ..."

You sort your thoughts and start modelling: Denote with f the result of a coin flip ($f = 0$ is heads, $f = 1$ is tails). Model the bias of the coin with θ_1 and use θ_2 for Zucky's *truthfulness*. Zucky's answer is denoted by z. Furthermore, assume that θ_2 is independent of f and θ_1. Thus, $p(z \mid f, \theta_2)$ is given as:

	$z = 0$	$z = 1$
$f = 0$	θ_2	$1 - \theta_2$
$f = 1$	$1 - \theta_2$	θ_2

Problem 2: Make a *similar* 2×2 table for the joint probability distribution $p(f, z \mid \boldsymbol{\theta})$ in terms of $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Show your work. Note that the likelihood function $p(f, z \mid \theta_1, \theta_2)$ factorises and simplifies under our independence assumptions, i.e.,

$$p(f, z \mid \theta_1, \theta_2) = p(z \mid f, \theta_2) p(f \mid \theta_1).$$

Problem 3: The sample run on the paper looks like this:

f	1	1	0	1	1	0	0
z	1	0	0	0	1	0	1

What are the maximum likelihood estimates for θ_1 and θ_2? Justify your answer.

3 The probabilistic coin game

In the following we are considering a more involved version of predicting coin tosses. Instead of one coin that we observe tosses from, two coins with different characteristics exist. At the beginning of a series of N coin flips, one of the two coins is drawn randomly and with this coin the observed tosses are performed. After N tosses the goal is to predict the outcome of the next flip with this coin.

One of the two coins is drawn randomly and 10 coin tosses are made: 7 heads and 3 tails.

Assume for coin number 1 a prior of $p(\theta \mid c = 1) = \text{Beta}(\theta \mid 4, 4)$ and for coin number 2 a prior of $p(\theta \mid c = 2) = \text{Beta}(\theta \mid 6, 2)$. The overall prior for a randomly drawn coin should be $p(\theta) = 0.5p(\theta \mid c = 1) + 0.5p(\theta \mid c = 2)$.

Problem 4: Why is this overall prior a valid assumption? Argue in 2–3 sentences.

Problem 5: Compute $p(\theta \mid \mathcal{D})$ where $\mathcal{D}$ denotes the observed data. Show your work! Use the following steps:

1. Write $p(\theta \mid \mathcal{D})$ in terms of $p(\theta, c \mid \mathcal{D})$ for $c = 1$ and $c = 2$.
2. Find an expression that involves the class-dependent posterior of θ, $p(\theta \mid c, \mathcal{D})$ for $c = 1, 2$. Why is this advantageous?
3. Compute an easier expression for this posterior via Bayes' Rule.
4. Why is $p(\mathcal{D} \mid \theta, c) \equiv p(\mathcal{D} \mid \theta)$, i.e., why is the likelihood independent of the class? What will be the posterior distribution?
5. Determine the missing components from step 2, i.e., the factors that are not the class-dependent posterior. If you get stuck, inspect your results from steps 3 and 4 closely to get to a solution.
6. Put the pieces together and determine the posterior distribution $p(\theta \mid \mathcal{D})$.

Problem 6: Sketch in one or two sentences how you then can use the computed posterior in this prediction game. (No computations are required!)