# 6A Bayesian inference II

## Class problems

**6A1** (Predicting coin tosses) A coin has an unknown probability $P$ for turning up heads (1), and $1-P$ probability for tails (0), independently on each toss. A priori we assume the unknown probability parameter $P$ to have the uniform distribution over $[0,1]$.

(a) If the unknown probability has value $P = p$, what is then the probability that 10 consecutive tosses are all heads?

(b) Before seeing any data, working from the prior distribution of $P$, what is the (prior) predictive probability that the first 10 tosses will be all heads? Hint: By the law of total probability,

$$f(\vec{x}) = \int_0^1 f_{\vec{X}|P}(\vec{x} \mid p)\, f_P(p)\, dp.$$

One of the things in the integrand is the prior, and the other you get from (a). Then you need to do the integral.

(c) After 20 tosses, of which 20 were heads, what is the posterior distribution for $P$? You can report it either by its density function, or its name and parameters.

(d) After the 20 observed heads in (c), what is now the (posterior) predictive probability that the *next* 10 tosses will be all heads? Hint: Again use the law of total probability, but apply the posterior distribution of $P$. (Or consult slides of lecture 5B.)

(e) Compare the numerical results from (b) and (d) to the alternative scenario: A coin is *known* to be fair ($p = 0.5$), we toss it ten times, and we ask for the probability of obtaining ten heads. Can you explain, by common sense, the relative order of these three numbers?

**6A2** (Waiting times) In this exercise we practice working with a new distribution family. Keep calm, read carefully, look at the known facts and think which fact(s) could be useful in your task.

If a continuous random variable $\Lambda$ has density function

$$f_\Lambda(\lambda) = \begin{cases} c\,\lambda^{\alpha-1}\,e^{-\beta\lambda} & \text{when } \lambda > 0, \\ 0 & \text{otherwise,} \end{cases}$$

we say that $\Lambda$ has the **gamma distribution** with parameters $\alpha > 0$ and $\beta > 0$. We denote this by $\Lambda \sim \mathrm{Gam}(\alpha, \beta)$. The $c$ is a constant that ensures that $\int_0^\infty f(\lambda) = 1$, so that $f$ is indeed a density function. It is known that if $\Lambda \sim \mathrm{Gam}(\alpha, \beta)$, then

$$\mathrm{E}(\Lambda) = \alpha/\beta.$$

It is also known that if $\alpha$ is a positive integer, then $c = \beta^\alpha/(\alpha-1)!$, where $\alpha!$ is the factorial.

(a) Particle decays happen at random intervals, independently, at an unknown rate of $\Lambda$ decays per second. For the unknown parameter, we assume a prior density $\Lambda \sim \mathrm{Gam}(2, 10)$. Write down its density function. Calculate its mean (applying the known facts above it should be easy). Based on the mean, explain roughly what kind of decay rate we are expecting for the particles.

(b) If the decay rate is $\Lambda = \lambda$, then each decay interval $X_i$ has the exponential distribution with rate $\lambda$. Thus it has density

$$f_{X_i|\Lambda}(x_i \mid \lambda) = \lambda e^{-\lambda x_i},$$

for $x_i > 0$. This is the likelihood. Write down the unnormalized posterior density

$$f_\Lambda(\lambda)\, f_{\vec{X}|\Lambda}(\vec{x} \mid \lambda),$$

when three decay intervals $\vec{x} = (x_1, x_2, x_3) = (3.0, 12.2, 16.1)$ have been observed, measured in seconds. Apply the fact that the three observations are independent.

(c) Using the unnormalized posterior density, find the MAP estimate of $\Lambda$. Hint: Logarithm and derivate?

(d) From the form of the unnormalized posterior density, you should recognize that it is the density of a gamma distribution, up to a normalizing constant. What are its parameters $\alpha$ and $\beta$? Hint: Look separately at the exponents of $\lambda$ and $e$.

(e) Now that you know the posterior distribution of $\Lambda$, calculate its mean (again use the formula to make it easy). Based on the mean, explain roughly what kind of decay rate are we now expecting for the particles.

(f) (Optional, requires computer.) The $q$-quantile of the $\mathrm{Gam}(\alpha, \beta)$ distribution can be computed with the R command `qgamma(q, alpha, beta)`. Find a 95% credible interval for the parameter $\Lambda$, based on its posterior distribution.

## Home problems

**6A3** (Multinomial DNA model) The human DNA is can be modelled as a string of length $3 \times 10^9$, made of the four letters A, C, G, T. Only about 1.5% of the string consists of *protein-coding* regions (direct recipes for building proteins by concatenating amino acids). Other parts of the DNA have other duties (partially unknown).

Within protein-coding regions, based on some studies, the four letters occur at relative frequencies $(0.17, 0.29, 0.33, 0.21)$ respectively. Outside these regions, the letters occur at relative frequencies $(0.25, 0.25, 0.25, 0.25)$. Our simplified model states that each letter occurs randomly with these probabilities, independent of the other letters.

We are looking at a randomly chosen location $i$ in the string, and are trying to find whether it is *inside* a protein-coding region ($\Theta = 1$), or *outside* them ($\Theta = 0$). We look at a sequence of 100 letters around the location $i$, and observe a string AACTG. . . TGA, that contains 16 A's, 26 C's, 38 G's and 20 T's, in some order. We assume, for simplicity, that the whole 100 letters is either completely within a protein-coding region, our completely outside them.

Since $\Theta$, the indicator for our location being inside a protein-coding region, is unknown, we treat it as the unknown parameter, whose value determines the letter probabilities. Note that here we are not estimating the multinomial probability parameters; they are given, the only unknown thing is which of the two distributions applies.

(a) What is the prior distribution of $\Theta$? What does it mean?

(b) If $\Theta = 0$, what is the probability of observing exactly the 100-letter sequence we observed? Hint: Think of the ordered three-party samples in Lecture 5B. Do not worry that the probability is very small. Express it with at least three significant digits. *Alternatively*, you can use the likelihoods of the *counts*.

(c) If $\Theta = 1$, what is the probability of observing exactly the 100-letter sequence we observed?

(d) Calculate the posterior distribution of $\Theta$, and explain what it means.

This is a grossly simplified model of the DNA, and the numbers are somewhat made up. However, similar methods are have been used for finding which regions of the DNA are protein-coding.

**6A4** (Coin shaking) Professor Abel has made a coin-tossing machine. Initially, the coin is placed tails up (0) on the bottom of a drinking glass. Then a machine shakes the glass for a while, and the coin is inspected to see whether the top face is tails (0) or heads (1). This is done repeatedly to generate a sequence of random numbers $X_1, X_2, X_3, \ldots$, where $X_i \in \{0, 1\}$ is the coin position after $i$ shaking rounds. The initial position $X_0 = 0$ is known.

After 50 shaking rounds, the coin positions were

$$\vec{x} = (x_1, x_2, \ldots, x_{50}) = (0000110111\ 1111100000\ 0000000100\ 0111110011\ 1111111111)$$

(We have dropped the commas between digits for brevity, and added a small space after each ten digits, but really this is a sequence of 50 integers, each zero or one.)

By our physical understanding, we assume that on each round, the coin will *flip* with probability $\theta$, independent of its current position and of what happened before. So to each round we associate a random variable $F_i \in \{0, 1\}$, where 0 indicates no flip, and 1 indicates flip. If the coin does not flip, then $X_i = X_{i-1}$. If it flips, then $X_i = 1 - X_{i-1}$.

(a) Look at the data $\vec{x}$. Does it seem like the coin is flipping with probability $\theta = 0.5$ every time?

(b) We observe that the coin flipped 9 times in our experiment. Treat the unknown flipping probability as a continuous random variable $\Theta$, with uniform prior over the interval $[0, 1]$. Find the posterior distribution of $\Theta$, and also a posterior point estimate of your choice. (Posterior mean and posterior mode would be easy choices.)

Hint. Think of the sequence of *flips*, not of the sequence of *coin positions*. The flips are independent. You can use either the actual flip sequence, or the number of flips as your data; the result will be the same. Recall Lectures 5A and 5B. Apply the fact that in the binary model, the prior $\mathrm{Beta}(1, 1)$ becomes the posterior $\mathrm{Beta}(1+a, 1+b)$ when $a$ ones and $b$ zeros are observed. You do *not* need to find the normalization constant manually. You can also use the fact that the mean of $\mathrm{Beta}(a, b)$ is $a/(a + b)$.

(c) Find a 95% credible interval for $\Theta$, that is, an interval that contains $\Theta$ with 95% probability (conditional on the observed flips). Hint: use a computer. For example, the R command `qbeta(q, a, b)` gives the $q$-quantile of the $\mathrm{Beta}(a, b)$ distribution.

(d) Professor Abel observes that the coin landed heads 28 times, and tails 22 times. Because the observed relative frequencies were approximately half and half, he claims that his machine manages to toss coins very randomly, with each result being heads with 50% probability, independent of other results. What do you think of his reasoning?

(e) Suppose that after a long series of experiments, we have become convinced that $\theta = 0.2$ (to a high precision). Consider a series of ten shakings. What is the probability that after ten shakings, the coin is in the same position as before them? Report with three decimals. Hint: Consider the number of flips that occurred. What is the probability that it is even?