5B Bayesian inference

About notation. The general and unambiguous notation for the density of whatever random variable X is $f_X(x)$. So if the random variable is Θ, we write $f_\Theta(\theta)$. Similarly, for a joint density we write $f_{X,Y}(x,y)$, and for a conditional density we write $f_{X \mid \Theta}(x \mid \theta)$.

The subscript clarifies which density function we are talking about. The arguments inside parentheses are where the function is being evaluated. So $f_X(5)$ is the density of X evaluated at point 5.

However, for brevity, we often leave out the subscripts if the argument makes it clear which f we are talking about. So we may write $f(\theta)$ or $f(\vec{x} \mid \theta)$ or $f(\theta \mid \vec{x})$. Do not be confused by this. The f's refer to different functions. This is the typical shorthand notation in Bayesian inference, because otherwise we would have so many subscripts. Look at the arguments to understand which density function it is.

Don't use the shorthand if it is not clear. For example, does $f(1)$ refer to the density of X or Θ?

Class problems

5B1 (Missing airplane) This problem is very loosely based on the (different but likewise Bayesian) method that was used to locate the airplane that disappeared during Malaysia Airlines flight MH370 in March 2014.

The search area is divided into four quadrants. From background information, we assess that the missing airplane has probabilities 0.5, 0.3, 0.1 and 0.1 for being in quadrants 1, 2, 3, 4 respectively. We also assess that whenever we search a quadrant, if the airplane really is there, we find it with probability 30% (and fail to find with probability 70%), independent of any previous search attempts. If the airplane is *not* in the quadrant we search, of course we do not find it.

(a) The airplane location is modelled as a random variable $\Theta \in \{1, 2, 3, 4\}$. Express its prior distribution as a table.

(b) The search crew first searches quadrant 1. Let X be an indicator variable: $X = 1$ if the airplane was now found, and $X = 0$ otherwise. Determine the likelihood function for Θ given the observation $X = 0$.

(c) The airplane was *not* found on the first attempt. Now determine the *posterior* distribution of the airplane location, based on this information. Do you think we should search quadrant 1 again?

(d) The search crew decides to search quadrant 1 again. Let Y be the indicator variable for the second search result. The result is again negative (airplane not found). Determine the posterior distribution of the airplane location, based on this information. Which quadrant do you think we should search next?

Solution.

(a) The **prior** distribution for the unknown location Θ of the airplane is

θ	1	2	3	4
$f_\Theta(\theta)$	0.5	0.3	0.1	0.1

(b) The likelihood, for the one data point $x = 0$, is $f(x \mid \theta) = \mathrm{P}(X = 0 \mid \Theta = \theta)$. This is the probability of *not* finding the airplane when searching quadrant 1. This is 0.7 if the airplane *is* in that quadrant (i.e. if $\Theta = 1$), and 1.0 if the airplane is not there (i.e. if $\Theta \neq 1$). So the **likelihood**, as a function of θ, is

θ	1	2	3	4
$f_{X\mid\Theta}(0 \mid \theta)$	0.7	1	1	1

(c) The posterior density is simply the product of the prior and the likelihood, times a normalizing constant $c > 0$, so

$$f_{\Theta\mid X}(\theta \mid 0) \;=\; c \cdot f_\Theta(\theta) f_{X\mid\Theta}(0 \mid \theta),$$

or in shorthand,

$$f(\theta \mid 0) \;=\; c \cdot f(\theta) f(0 \mid \theta).$$

The normalizing constant c is in fact $1/f_X(x) = 1/\big(\sum_\theta f(\theta) f(x \mid \theta)\big)$, but it is usually easiest to compute the thing in stages, as follows.

Let us first compute the *unnormalized* posterior, that is, the posterior density without the normalizing constant c. This is **prior times likelihood**. Computing this for all possible values of θ:

θ	1	2	3	4
$f_\Theta(\theta) f_{X\mid\Theta}(0 \mid \theta)$	0.35	0.30	0.10	0.10

The sum of the unnormalized posterior densities is

$$\sum_\theta f_\Theta(\theta) f_{X\mid\Theta}(0 \mid \theta) \;=\; 0.35 + 0.30 + 0.10 + 0.10 \;=\; 0.85.$$

Dividing by this constant, we obtain the true posterior density

θ	1	2	3	4
$f_{\Theta\mid X}(\theta \mid 0)$	0.412	0.353	0.118	0.118

Quadrant 1 still has the greatest probability of containing the airplane, so perhaps we should search there again. (That search has the highest probability of success.)

We can say that $\theta = 1$ is the *maximum a posteriori estimate* (MAP estimate) for θ, because it has the greatest posterior probability.

(d) **Method 1: All data at once.** The likelihood for the full data $(x, y) = (0, 0)$ is

$$f_{X,Y \mid \Theta}(0, 0 \mid \theta) \;=\; \mathrm{P}(X = 0, Y = 0 \mid \Theta = \theta).$$

If the airplane is in quadrant 1, then the probability of obtaining data $(0, 0)$ is $0.7 \cdot 0.7 = 0.49$. If the airplane is not there, then certainly nothing is found so the probability of obtaining $(0, 0)$ is $1 \cdot 1 = 1.00$. Thus the likelihood function is

θ	1	2	3	4
$f(0, 0 \mid \theta)$	0.49	1	1	1

The posterior density $f(\theta \mid 0, 0)$ is obtained by computing, for all values of θ, the products $f(\theta) \cdot f(0, 0 \mid \theta)$, and then normalizing so that the numbers sum to one. We get the posterior density

θ	1	2	3	4
$f(\theta \mid 0, 0)$	0.329	0.403	0.134	0.134

After two failed searches in quadrant 1, it seems our next search should go to quadrant 2, which now has the maximum posterior probability.

Method 2. Incremental update. Alternatively, we can take the posterior for Θ that we computed after first observation, and use it as our *prior* for Θ before the second observation. Then calculate the new posterior as usual, by using only the second observation $Y = 0$. The numerical result is the same.

(It is a good idea for you to carry out the calculation in both ways and actually observe that the result is the same.)

5B2 (Metro intervals.) The metro runs regularly with an unknown interval θ (minutes). As Thomas arrives at the station at random times each day, his waiting time X (in minutes) has uniform distribution over $[0, \theta]$, with density

$$f(x \mid \theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta, \\ 0, & \text{otherwise.} \end{cases}$$

The interval is unknown to Thomas, so he models it as a random variable Θ. By his prior knowledge he assumes the interval is at least 1 minute. He takes the prior distribution of Θ to have density

$$f_\Theta(\theta) = \begin{cases} 0.2\,\theta^{-1.2}, & \theta \ge 1, \\ 0, & \text{otherwise.} \end{cases}$$

Over five days Thomas has observed waiting times $x = (7, 3, 2, 9, 6)$.

(a) **Draw** Thomas's *prior* density function over the interval $[0, 20]$ (by hand or computer). **Verify** that it is is indeed a density function, by calculating its integral over the range of possible values. Hint: You can integrate it just like a polynomial, even though the exponent is not an integer. For drawing, it is enough to calculate the function at a few points to get the rough shape.

(b) **Calculate** the *posterior* density of Θ and **draw** it over $[0, 20]$ by hand or computer. Hint: prior times likelihood gives the unnormalized posterior, but you need to normalize it.

(c) Thomas decides to use the *posterior mean* (i.e. expected value of the posterior distribution) as a point estimate for Θ. **Calculate** his estimate.

(d) Using the posterior distribution, **calculate** the probability that $\Theta < 15$.

Solution.

(a) For the picture, see below. The integral over the possible values $[1, \infty)$ is

$$\int_1^\infty 0.2\,\theta^{-1.2} d\theta = \left[\frac{0.2}{-0.2}\,\theta^{-0.2} \right]_{\theta=1}^\infty = 0 - \frac{0.2}{-0.2}\,1^{-0.2} = 1.$$

(b) The likelihood for data $\vec{x} = (x_1, \ldots, x_5)$ is

$$f(\vec{x} \mid \theta) = \prod_{i=1}^5 f(x_i \mid \theta) = \begin{cases} \theta^{-5}, & \theta \ge M, \\ 0, & \text{otherwise,} \end{cases}$$

where $M = \max(x_1, \ldots, x_5)$. For the data set $\vec{x} = (7, 3, 2, 9, 6)$ we have $M = 9$. The posterior density for Θ is

$$f(\theta \mid \vec{x}) = c \cdot f(\theta) f(\vec{x} \mid \theta) = \begin{cases} c \cdot 0.2\,\theta^{-6.2}, & \theta \ge 9, \\ 0, & \text{otherwise,} \end{cases}$$

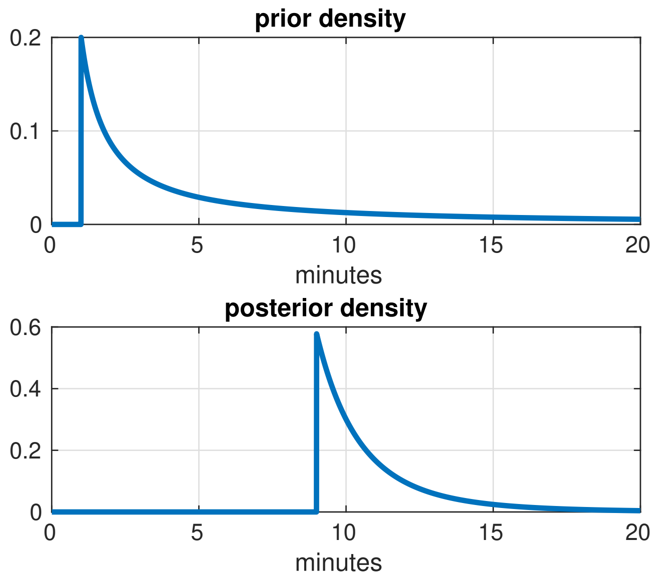
where $c > 0$ is a normalizing constant. We can simplify a little by writing $c \cdot 0.2 = C$. We can find its value by making sure that $\int f(\theta \mid \vec{x}) d\theta = 1$. Because

$$\int_9^\infty \theta^{-6.2} d\theta = \left[\frac{1}{-5.2}\theta^{-5.2}\right]_9^\infty = \frac{1}{5.2 \cdot 9^{5.2}},$$

we must take the $C = 5.2 \cdot 9^{5.2}$, so the posterior density is

$$f(\theta \mid \vec{x}) = (5.2 \cdot 9^{5.2}) \cdot \theta^{-6.2}, \qquad \text{when } \theta > 9$$

and zero otherwise.



You may notice that both the prior and the posterior distribution are *Pareto* distributions, which you have seen in earlier exercises. This is computationally nice: if the prior is Pareto and the likelihood is uniform, then the posterior is also Pareto.

Similar nice things can happen with other distributions; for example, in the coin tossing experiment, if the prior is Beta (or uniform), and the likelihood is binary, then the posterior is also Beta. For the general phenomenon, look up "conjugate prior".

(c) Using the posterior density from (b), we find the posterior mean, as usual, by integrating

$$\int_{-\infty}^\infty \theta f(\theta \mid \vec{x})\, d\theta = \int_9^\infty \theta\, C\theta^{-6.2}\, d\theta = C \int_9^\infty \theta^{-5.2}\, d\theta$$
$$= C \cdot \frac{9^{-4.2}}{4.2} = \frac{5.2 \cdot 9^{5.2}}{4.2 \cdot 9^{4.2}} = \frac{5.2 \cdot 9}{4.2} \approx 11.143.$$

(d) As usual, the probability of Θ being in an interval is found by integrating its density over that interval. Here we use the posterior density of Θ.

$$
\begin{aligned}
\mathrm{P}(\Theta < 15 \mid \vec{X} = \vec{x}) &= \int_9^{15} f(\theta \mid \vec{x})\, d\theta = \int_9^{15} (5.2 \cdot 9^{5.2}) \cdot \theta^{-6.2}\, d\theta \\
&= \frac{5.2 \cdot 9^{5.2}}{-5.2} \cdot \Big[\theta^{-5.2}\Big]_{\theta=9}^{15} = 9^{5.2} \cdot \big(9^{-5.2} - 15^{-5.2}\big) \approx 0.930.
\end{aligned}
$$

Home problems

5B3 (Leaf lengths) A botanist assumes that the leaf lengths (in cm) of a certain rose species have the normal distribution with an unknown mean Θ and a known standard deviation $\sigma = 2$. Each leaf length is independent from others (but they have this same distribution). Further, his prior distribution for Θ is normal with mean $\mu_0 = 10$ and standard deviation $\sigma_0 = 1$. The botanist has measured five leaf lengths as $\vec{x} = (9, 13, 14, 12, 17)$. After these observations, find:

(a) The posterior mean of Θ.

(b) An interval that contains Θ with probability 90%.

Hint. Under the given assumptions, the posterior distribution of Θ is also a normal distribution, whose mean and standard deviation have known formulas. See Lecture 5A and/or Ross's Section 7.8. Once you know the posterior distribution of Θ, you can use its CDF to find a suitable interval. You can use tables or a calculator. Check the R commands `qnorm` and `pnorm`.

Grading.
1 p for (a),(b) each. No penalty for small errors in calculation or rounding.

Solution.

(a) The posterior distribution of Θ is normal, with mean

$$\mu_1 = \left(\frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \right) \mu_0 + \left(\frac{\frac{n}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \right) m(\vec{x}),$$

and standard deviation

$$\sigma_1 = \frac{1}{\sqrt{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}},$$

where $n = 5$ is the sample size, and $m(\vec{x}) = \frac{1}{5}\sum_{i=1}^{5} x_i = 13$ is the sample mean. Plugging in these values, we have the posterior mean

$$\mu_1 = \left(\frac{\frac{1}{1^2}}{\frac{1}{1^2} + \frac{5}{2^2}} \right) 10 + \left(\frac{\frac{5}{2^2}}{\frac{1}{1^2} + \frac{5}{2^2}} \right) 13 = \frac{4}{9} \times 10 + \frac{5}{9} \times 13 = 11\frac{2}{3} \approx 11.67$$

(b) Similarly, from the formula given, we get the posterior standard deviation

$$\sigma_1 = \frac{1}{\sqrt{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}} = \frac{1}{\sqrt{\frac{1}{1^2} + \frac{5}{2^2}}} = \frac{2}{3}.$$

Conditional on the observed data $\vec{x}$, the random variable $Z = \frac{\Theta - \mu_1}{\sigma_1}$ follows the standard normal distribution. Find a point $z > 0$ such that both tails of Z have probability 0.05, by finding where $\Phi(z) = 0.95$. From tables (or computer) we find $z \approx 1.64$. Thus

$$\mathrm{P}\left(-1.64 \le \frac{\Theta - \mu_1}{\sigma_1} \le 1.64 \;\middle|\; \vec{x} \right) = 0.9,$$

that is

$$P(\mu_1 - 1.64\sigma_1 \le \Theta \le \mu_1 + 1.64\sigma_1 \mid \vec{x}) \;=\; 0.9.$$

So an interval that contains Θ with 90% probability is

$$\Theta \;=\; \mu_1 \pm 1.64\sigma_1 \;=\; 11.67 \pm 1.09.$$

5B4 (Dangerous road) The number of accidents in a month, on a certain strip of road, is assumed to follow the Poisson distribution

$$f(x \mid \theta) \;=\; e^{-\theta}\frac{\theta^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

with an unknown mean parameter $\theta > 0$. The numbers of different months are assumed independent and follow the same distribution. From previous experience, a bayesian engineer has estimated the unknown parameter Θ to have probabilities 0.25, 0.50 and 0.25 for the values 1, 2, 3 respectively.

(a) During the first month, there are two accidents on the road. Find the posterior distribution of Θ after these observations. In other words, find the posterior probabilities for $\Theta = 1, 2, 3$.

(b) During the second month, there are no accidents on the road. Find the posterior distribution of Θ, using the observations of both months.

Grading.
(a) 0.5 p for correct likelihood function
(a) 0.5 p for correct posterior distribution
(b) 0.5 p for correct likelihood function
(b) 0.5 p for correct posterior distribution
No penalty for small errors in calculation or rounding.

Solution. Let X_i = number of accidents during month i, for $i = 1, 2$.

(a) The prior distribution for the (discrete) Θ is

θ	1	2	3
$f(\theta)$	0.25	0.50	0.25

If the mean parameter is θ, then the probability for 2 accidents in a month is

$$f(2 \mid \theta) = \frac{\theta^2}{2}e^{-\theta}.$$

So having the observation $X_1 = 2$, the likelihood function is on siis

θ	1	2	3
$f(2 \mid \theta)$	$\frac{1}{2}e^{-1}$	$2e^{-2}$	$\frac{9}{2}e^{-3}$

Multiplying prior by likelihood, we get the unnormalized posterior

θ	1	2	3
$f(\theta)f(2 \mid \theta)$	0.0460	0.1353	0.0560

Because the sum of the unnormalized values is 0.2373, by dividing by this value we obtain the posterior

θ	1	2	3
$p_1(\theta \mid 2)$	0.1938	0.5702	0.2360

These R commands might be useful:

```
p0 <- c(0.25, 0.5, 0.25)
th <- c(1, 2, 3)
f <- dpois(2,th)
p1 <- p0*f/sum(p0*f)
```

(b) Let's use the "all data at once" method. The prior is as before. If the mean parameter is θ, then the probability for *two* accidents during first month and *zero* accidents during second month is

$$f(2, 0 \mid \theta) = \left(\frac{\theta^2}{2!}e^{-\theta}\right)\left(\frac{\theta^0}{0!}e^{-\theta}\right) = \frac{\theta^2}{2}e^{-2\theta}.$$

So after the observations $X_1 = 2, X_2 = 0$, the likelihood function is

θ	1	2	3
$f(2, 0 \mid \theta)$	0.0677	0.0366	0.0112

As before, multiply prior with likelihood to obtain the unnormalized posterior, and then normalize to get the true posterior. The result is

θ	1	2	3
$f(\theta \mid 2, 0)$	0.4449	0.4817	0.0733

The following R commands may be useful.

```
p0 <- c(0.25, 0.5, 0.25)
th <- c(1, 2, 3)
f <- dpois(2, th)*dpois(0, th)
p1 <- p0*f/sum(p0*f)
```