

HW 1 Solutions

Shravan Vasishth

Exercise 1

The French mathematician Pierre-Simon Laplace (1749-1827) was the first person to show definitively that the proportion of female births in the French population was less than 0.5, in the late 18th century, using a Bayesian analysis based on a uniform prior distribution.

Suppose you were doing a similar analysis but you had more definite prior beliefs about the ratio of male to female births. In particular, if θ represents the proportion of female births in a given population, you are willing to place a $\text{Beta}(100,100)$ prior distribution on θ .

- Show that this means you are more than 95% sure that θ is between 0.4 and 0.6, although you are ambivalent as to whether it is greater or less than 0.5.
- Now you observe that out of a random sample of 1,000 births, 511 are boys. What is your posterior probability that $\theta > 0.5$?

Solution to exercise 1

The probability that θ lies between 0.6 and 0.4:

```
pbeta(0.6,shape1=100,shape2=100)-pbeta(0.4,shape1=100,shape2=100)
```

```
## [1] 0.9956798
```

The data: $n=1000$ people, 511 boys. (Assuming a binary gender here because back then they only recognized two genders.) So, $k = 1000-511 = 489$ girls.

The posterior will be $\text{Beta}(a=100+489, b=100+511)$. (See lecture slides and book if it is not clear why). That means that the proportion of female births is:

```
a<-100+489
b<-100+1000-489
## The mean of a Beta distribution:
a/(a+b)
```

```
## [1] 0.4908333
```

The probability that θ is larger than 0.5:

```
pbeta(0.5,shape1=100+489,shape2=100+1000-489,lower.tail=FALSE)
```

```
## [1] 0.2626087
```

Exercise 2

Suppose we are modeling the egocentricity of a person as measured by their speech. A researcher theorizes that one can get an idea of a speaker's egocentricity by measuring the number of times they produce a first-person pronoun in a sentence per day (e.g., starting a sentence with "I", or using a word like "me", "my", etc., anywhere in a sentence).

The number of times x that such a pronoun is used can be modeled by a Poisson distribution.

We are going to first define a Gamma prior on the λ parameter (representing the rate of occurrences of the pronouns) in the Poisson distribution. We have prior data suggesting that the mean number of occurrences of a first person personal pronoun in a population is 75 and the variance is 185. Using the mean and variance, figure out the hyperparameters of the Gamma prior, i.e., find out what a and b will be in $\text{Gamma}(a,b)$. Hint: a/b is the mean of a Gamma distribution, and a/b^2 is the variance. Once you have figured out the Gamma hyperparameters, plot the Gamma distribution representing the priors.

Now suppose that the observed number of pronoun utterances per day over four days is 200,87,99,121. What is the posterior distribution of the λ parameter? This distribution will be a Gamma distribution with updated hyperparameters a^* and b^* . Your task is to figure out what these updated values a^* and b^* are.

Now, suppose that you get additional data for *two* further days, and the number of produced utterances over the two days taken together is 300. [Hint: this means that the data 300 is from $n=2$ days].

- i. Use the posterior you just obtained ($\text{Gamma}(a^*, b^*)$) as a prior for this new data, and write down the posterior distribution for λ as a $\text{Gamma}(a^{**}, b^{**})$ distribution, where a^{**}, b^{**} are the updated hyperparameters.
- ii. Now, start with your original prior $\text{Gamma}(a,b)$, and use all the data available (200,87,99,121,300) and derive the posterior distribution of λ , expressed as a Gamma distribution with some hyperparameters a^{***} and b^{***} . (Hint: think carefully about what n is. Is it 5 or 6?)

Is there any difference in the posterior between the above two analyses (i) and (ii)?

Solution to exercise 2

The number of times x that the word of interest is uttered in one day can be modeled by a Poisson distribution:

$$f(x | \lambda) = \frac{\exp(-\lambda)\lambda^x}{x!} \quad (1)$$

where the rate λ is unknown, and the numbers of utterances of the target word on each day are independent given λ .

We are told that the prior mean of λ is 75 and prior variance for λ is 185.

In order to visualize the prior, we first fit a Gamma density prior for λ based on the above information.

As explained in the assignment (the hint), we know that for a Gamma density with hyperparameters a, b , the mean is $\frac{a}{b}$ and the variance is $\frac{a}{b^2}$. Since we are given values for the mean and variance, we can solve for a, b , which gives us the Gamma density.

If $\frac{a}{b} = 75$ and $\frac{a}{b^2} = 185$, it follows that

- $a = 75 \times b$
- $a = 185 \times b^2$

So, $185 \times b^2 = 75 \times b$. Simplifying: $185 \times b = 75$. So $b = \frac{75}{185}$.

This means that $a = 75 \times \frac{75}{185} = \frac{5625}{185}$.

The Gamma distribution for the prior on λ is as shown below.

$$\lambda \sim \text{Gamma}\left(\frac{5625}{185}, \frac{75}{185}\right) \quad (2)$$

```
a<-5625/185
b<-75/185
```

```
lambda<-0:200
plot(lambda,dgamma(lambda,shape=a,rate=b),type="l",lty=1,main="Gamma prior",ylab="density",cex.lab=2,cex
```

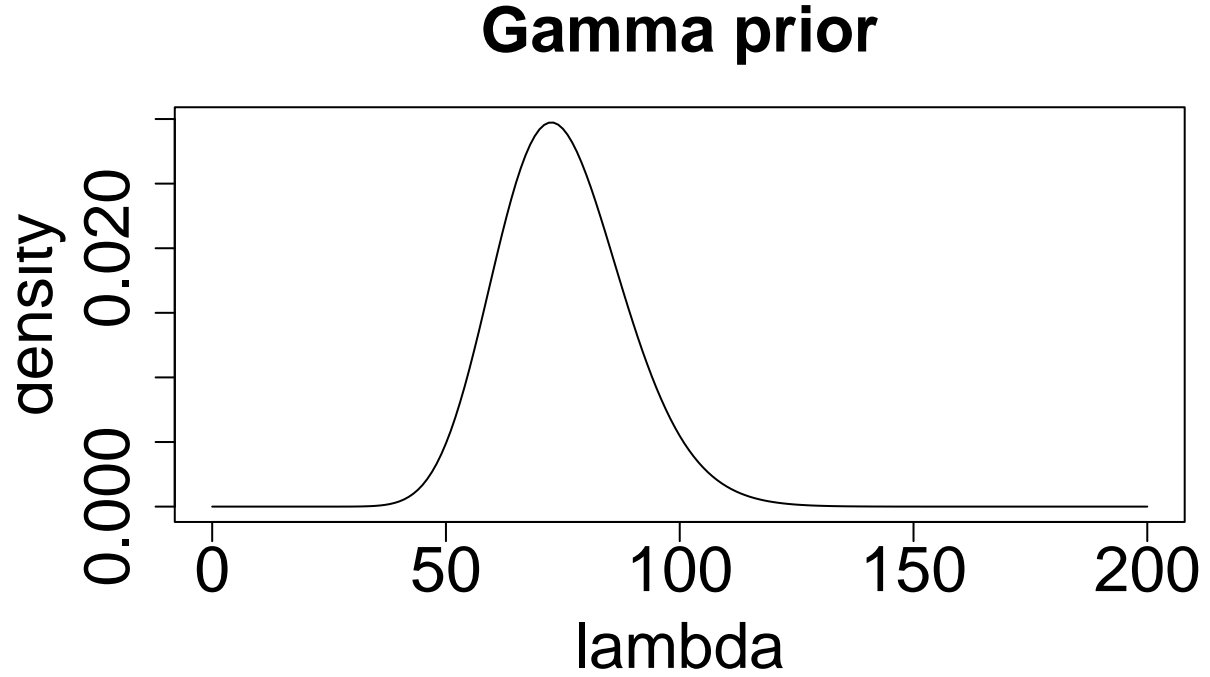


Figure 1: The Gamma prior for the parameter lambda.

Given that

$$\text{Posterior} \propto \text{Prior Likelihood} \quad (3)$$

and given that the likelihood is:

$$\begin{aligned} L(\mathbf{x} \mid \lambda) &= \prod_{i=1}^n \frac{\exp(-\lambda) \lambda^{x_i}}{x_i!} \\ &= \frac{\exp(-n\lambda) \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \end{aligned} \quad (4)$$

we can compute the posterior as follows:

$$\text{Posterior} = \left[\frac{\exp(-n\lambda) \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \right] \left[\frac{b^a \lambda^{a-1} \exp(-b\lambda)}{\Gamma(a)} \right] \quad (5)$$

Disregarding the terms $x_i!$, $\Gamma(a)$, b^a , which do not involve λ , we have

$$\begin{aligned} \text{Posterior} &\propto \exp(-n\lambda) \lambda^{\sum_{i=1}^n x_i} \lambda^{a-1} \exp(-b\lambda) \\ &= \lambda^{a-1+\sum_{i=1}^n x_i} \exp(-\lambda(b+n)) \end{aligned} \quad (6)$$

First, note that the Gamma distribution in general is $\text{Gamma}(a, b) \propto \lambda^{a-1} \exp(-\lambda b)$. So it's enough to state the above as a Gamma distribution with some parameters a^* , b^* .

If we equate $a^* - 1 = a - 1 + \sum_i^n x_i$ and $b^* = b + n$, we can rewrite the above as:

$$\lambda^{a^*-1} \exp(-\lambda b^*) \quad (7)$$

This means that $a^* = a + \sum_i^n x_i$ and $b^* = b + n$.

Thus, the posterior has the form of a Gamma distribution with parameters $a^* = a + \sum_i^n x_i$, $b^* = b + n$.

The data:

```
x<-c(200,87,99,121)
n<-length(x)
sumx<-sum(x)
sumx
```

```
## [1] 507
```

The posterior's hyperparameters are:

```
(astar<-a+sumx)
```

```
## [1] 537.4054
```

```
(bstar<-b+n)
```

```
## [1] 4.405405
```

The posterior for λ is $\text{Gamma}(537.41, 4.41)$.

The new data we get is 300 utterances of the target word, but these are over two days. So our n is:

```
n<-2
```

We are adding two, not one, to our number of data points because our unit of measurement was number of utterances per day. If we get 300 utterances over two days, we have to use 2 as the sample size.

The summed values of the new observed utterances is:

```
sumx<-300
```

The new posterior is:

```
(astarstar<-astar+sumx)
```

```
## [1] 837.4054
```

```
(bstarstar<-bstar+n)
```

```
## [1] 6.405405
```

The updated posterior (rounded to two decimal places again) is $\text{Gamma}(837.41, 6.41)$.

Notice an important point that I mentioned in class relating to a question on updating the posterior given new data: the above calculation is exactly the same as pooling all the data, i.e, 200,87,99,121, 300, setting n=6, and then figuring out the posterior using the original prior that we had for the λ parameter.

```
x<-c(200,87,99,121,300)
n<-length(x) + 1 # the 300 is for two days
sumx<-sum(x)
sumx
```

```
## [1] 807
```

```
(astarstarstar<-a+sumx)
```

```
## [1] 837.4054
```

```
(bstarstarstar<-b+n)
```

```
## [1] 6.405405
```

Notice that the two sets of hyperparameters are the same:

Generally, it is the case that successively using the posterior from a previous study as a prior in a future study is the same thing as analyzing all the data from the previous and current study together, using the prior from the original study.

```
astarstar; bstarstar
```

```
## [1] 837.4054
```

```
## [1] 6.405405
```

```
astarstarstar; bstarstarstar
```

```
## [1] 837.4054
```

```
## [1] 6.405405
```

Bonus Exercise 3

Suppose that 1 in 1000 people in a population is expected to get HIV. Suppose a test is administered on a suspected HIV case, where the test has a true positive rate (the proportion of positives that are actually HIV positive) of 95% and true negative rate (the proportion of negatives are actually HIV negative) 98%. Use Bayes' theorem to find out the probability that a patient testing positive actually has HIV.

Solution to bonus exercise 3

Need to find: $\text{Prob}(\text{HasHIV} \mid \text{TestedPositive})$

We know:

- $\text{Prob}(\text{HasHIV}) = 1/1000$. Therefore, $\text{Prob}(\neg \text{HasHIV}) = 999/1000$.
- $\text{Prob}(\text{TestedPositive} \mid \text{HasHIV}) = 0.95$
- $\text{Prob}(\neg \text{TestedPositive} \mid \neg \text{HasHIV}) = 0.98$. This implies that $\text{Prob}(\text{TestedPositive} \mid \neg \text{HasHIV}) = 1 - 0.98 = 0.02$

Notice that, from the law of total probability (section 1.3 of book draft),

$$\text{Prob}(\text{TestPos}) = \text{Prob}(\text{TestPos} \mid \text{HasHIV})\text{Prob}(\text{HasHIV}) + \text{Prob}(\text{TestPos} \mid \neg \text{HasHIV})\text{Prob}(\neg \text{HasHIV})$$

Here, $\text{TestPos} = \text{TestedPositive}$.

So,

$$\text{Prob}(\text{TestedPositive}) = 0.95 \times 1/1000 + 0.02 \times 999/1000 = 0.021$$

Now, Bayes' rule is:

$$\text{Prob}(A \mid B) = \frac{\text{Prob}(B \mid A)\text{Prob}(A)}{\text{Prob}(B)}$$

Let $A=HasHIV$, $B=TestedPositive$. We can rewrite the above in terms of our research question:

$$Prob(HasHIV|TestedPositive) = \frac{Prob(TestedPositive|HasHIV)Prob(HasHIV)}{Prob(TestedPositive)}$$

Plugging in the probabilities:

$$Prob(HasHIV|TestedPositive) = \frac{0.95 \times (1/1000)}{0.03} = 0.045$$

So, the probability of having HIV given that one has tested positive in this situation is, surprisingly, 0.045.

This has to do with the low base rate of HIV (1/1000). Suppose that HIV was rampant, and the probability of having HIV was 999/1000 (obviously a ridiculously unrealistic example, but just to illustrate the effect of base rates).

Now,

$$Prob(TestedPositive) = 0.95 \times 999/1000 + 0.02 \times 1/1000 = 0.949$$

This implies that

$$Prob(HasHIV|TestedPositive) = \frac{0.95 \times (999/1000)}{0.949} = 1$$

Now, the probability of having HIV given that one has tested positive is 1.

In other words, the base rate (the prior probability of something happening) matters a lot in interpreting the posterior probability given some data. This makes intuitive sense: if some event is highly unlikely a priori, then seeing any evidence that event has happened should not shift your prior belief much that the event happened. By contrast, if something is highly likely to happen a priori, then finding evidence that the event happened should lead you to believe that that event actually did happen.

One practical example of this is the infamous Bem paper I mentioned in class:

Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of personality and social psychology*, 100(3), 407.

It doesn't really matter that he got significant effects of pre-cognition; the prior probability of people being able to tell the future accurately is near 0. So any evidence he reports of pre-cognition doesn't shift my belief at all.

Here is another example, looking at frequentist null hypothesis testing and how it shifts our beliefs about the null hypothesis:

Daniel J. Schad and Shravan Vasishth. The posterior probability of a null hypothesis given a statistically significant result. *Quantitative Methods for Psychology*, 2022.

<https://danielschad.shinyapps.io/probnull/>