



[Course](#) > [Unit 5 Bayesian statistics](#) > [Homework 9: Bayesian Statistics](#) > 1. Practice with Priors

## 1. Practice with Priors

One side concept introduced in the second Bayesian lecture is the *conjugate prior*. Simply put, a prior distribution  $\pi(\theta)$  is called **conjugate** to the data model, given by the *likelihood function*  $L(X_i|\theta)$ , if the posterior distribution  $\pi(\theta|X_1, X_2, \dots, X_n)$  is part of the same distribution family as the prior.

This problem will give you some more practice on computing posterior distributions, where we make use of the proportionality notation. It would be helpful to try to think of computations in forms that are reduced as much as possible, as this will help with intuition towards assessing whether a prior is conjugate.

This problem makes use the Gamma distribution (written as **Gamma**  $(k, \theta)$ ) is a probability distribution with parameters  $k > 0$  and  $\theta > 0$ , has support on  $(0, \infty)$ , and whose density is given by  $f(x) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\Gamma(k) \theta^k}$ . Here,  $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$  is the Euler Gamma function.

(a)

3.0/3 points (graded)

Suppose we have the prior  $\pi(\lambda) \sim \text{Exp}(a)$  (where  $a > 0$ , and conditional on  $\lambda$ , we have observations  $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d}}{\sim} \text{Exp}(\lambda)$ ). Compute the posterior distribution  $\pi(\lambda|X_1, X_2, \dots, X_n)$ .

The posterior distribution for  $\lambda$  is a Gamma distribution. What are its parameters? Enter your answer in terms of  $a$ ,  $n$ , and  $\sum_{i=1}^n X_i$

Generating Speech Output

(Enter **Sigma\_i(X\_i)** for  $\sum_{i=1}^n X_i$  . Do not worry if the parser does not render properly; the grader works independently. If you wish to have proper rendering, enclose **Sigma\_i(X\_i)** by brackets. )

$\theta =$   ✓ Answer: 1/(a+Sigma\_i(X\_i))

$k =$   ✓ Answer: n+1

Now, we examine two properties of the prior. Is the prior proper/improper? Is it a conjugate prior?

- ☒ The prior is proper and is a conjugate prior.
- ☐ The prior is improper and is a conjugate prior.
- ☒ The prior is proper and is not a conjugate prior. ✓
- ☐ The prior is improper and is not a conjugate prior.



#### Solution:

**Course convention:** If the family of distributions that the prior belongs to is specified as a strict subset of the family of distributions that the posterior is specified to belong to, then the convention in the course is that the prior still does **not** count a conjugate prior. However, if the family of distributions that the prior belongs to is specified as the full family of distributions that the posterior belongs to, then we will call the prior a conjugate prior.

We write out the expressions for the prior and posterior distributions in proportionality notation. Take note that our desired posterior distribution has variable  $\lambda$ , so we only consider multiplicative terms that depend on  $\lambda$ , and ignore the rest.

$\pi(\lambda) \propto \text{Exp}(a)$  corresponds to

Generating Speech Output

$$\pi(\lambda) = ae^{-a\lambda} \propto e^{-a\lambda}.$$

This is our prior distribution. Next, we compute the conditional likelihood function  $L(X_1, \dots, X_n | \lambda)$ .

For each  $i$ ,  $X_i \propto \text{Exp}(\lambda)$  corresponds to

$$\pi(X_i | \lambda) \propto \lambda e^{-X_i \lambda},$$

which this time can no longer be simplified through proportionality. notation. Using the i.i.d assumption on observations gives the likelihood function

$$\begin{aligned} L(X_1, \dots, X_n | \lambda) &\propto \pi(X_1) \dots \pi(X_n) \\ &\propto (\lambda e^{-X_1 \lambda}) \dots (\lambda e^{-X_n \lambda}) \\ &= \lambda^n e^{-(\sum X_i) \lambda}. \end{aligned}$$

By Bayes' theorem, we can combine the prior distribution with the conditional likelihood function to get

$$\begin{aligned} \pi(\lambda | X_1, \dots, X_n) &\propto \pi(\lambda) L(X_1, \dots, X_n | \lambda) \\ &\propto (e^{-a\lambda}) (\lambda^n e^{-(\sum X_i) \lambda}) \\ &\propto e^{-(a + \sum X_i) \lambda} \lambda^n. \end{aligned}$$

Finally, we match our posterior distribution with the form for the Gamma distribution. The distribution is over  $\lambda$ , so we match  $\lambda$  in our posterior distribution to  $x$  in the Gamma distribution. Then we notice that the exponent of  $\lambda$  in the posterior is  $n$ , and in the distribution the exponent of  $x$  is  $k - 1$ , so equating gives  $k - 1 = n$ , or  $k = n + 1$ . Next, we notice that the exponent of  $e$  in the posterior is  $-(a + \sum X_i) \lambda$ , while in the Gamma distribution it is  $-\frac{x}{\theta}$ . Equating with  $\lambda = x$  here gives  $\theta = \frac{1}{a + \sum X_i}$ .

To summarize, the parameters of our distribution are  $k = n + 1$  and  $\theta = \frac{1}{a + \sum X_i}$ . Our prior is proper as it's an exponential distribution which has a defined cdf (cumulative distribution function). It is not a conjugate prior because the posterior is not an exponential distribution due to the  $\lambda^n$  terms that makes it part of the more general Gamma distribution instead. (refer to the course convention).

Generating Speech Output

**i** Answers are displayed within the problem

(b)

4.0/4 points (graded)

Keep the observation model from the previous part, i.e. assume that our observations still satisfy  $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d}}{\sim} \text{Exp}(\lambda)$ . Now, consider a uniform prior  $\pi(\lambda) \propto 1$ . Compute the posterior distribution  $\pi(\lambda | X_1, X_2, \dots, X_n)$ .

Again, the posterior distribution for  $\lambda$  is an Gamma distribution. What are its parameters? Enter your answer in terms of  $n$  and  $\sum_{i=1}^n X_i$

(Enter **Sigma\_i(X\_i)** for  $\sum_{i=1}^n X_i$ . Do not worry if the parser does not render properly; the grader works independently. If you wish to have proper rendering, enclose **Sigma\_i(X\_i)** by brackets. )

$\theta =$

✓ Answer: 1/(Sigma\_i(X\_i))

$k =$

✓ Answer: n+1 Now, we examine two properties of the prior. Select the

correct answer choice.

☐ The prior is proper and is a conjugate prior.

☐ The prior is improper and is a conjugate prior.

☐ The prior is proper and is not a conjugate prior.

☒ The prior is improper and is not a conjugate prior.

Generating Speech Output



Which of the following statements describing the relationship between the prior/posterior distributions in the parts (a) and (b) are true?



For a fixed set of observations, taking the limit as  $a \rightarrow 0$  in the prior in (a) gives the same posterior as in (b).



For a fixed set of observations, taking the limit as  $a \rightarrow \infty$  in the prior in (a) gives the same posterior as in (b).



Assuming a true parameter  $\lambda$ , as we have more and more observations the two posterior distributions will converge to having the same mean.



### Solution:

We use a similar approach as we did in the previous subproblem. Note that our conditional distribution of  $X_i$  given  $\lambda$  is the same, so we have the same conditional likelihood function  $L(X_1, \dots, X_n | \lambda)$ . The only part that changes is that our posterior distribution here is  $\pi(\lambda) \propto 1$ , the uniform distribution.

Now, we apply Bayes' formula again to get that

$$\begin{aligned}\pi(\lambda | X_1, \dots, X_n) &\propto \pi(\lambda) L(X_1, \dots, X_n | \lambda) \\ &\propto (1) (\lambda^n e^{-(\sum X_i)\lambda}) \\ &\propto e^{-(\sum X_i)\lambda} \lambda^n.\end{aligned}$$

To write this in terms of the Gamma distribution, we use the same variable-matching process as we have done in the previous part. Here, we get that  $n$  (in the posterior) would correspond to  $k - 1$  (in the Gamma distribution), and  $-(a + \sum X_i) \lambda$  (in the posterior)  $-\frac{x}{\theta}$  (in the Gamma distribution). Equating then gives (taking into account that  $\lambda$  in the posterior corresponds to  $x$  in the Gamma distribution) that  $k - 1 = n$  and  $\frac{1}{\theta} = \sum X_i$ . Thus, we get the parameters  $k = n + 1$  and  $\theta = \frac{1}{\sum X_i}$ .

The prior  $\pi(\lambda)$  is not a proper prior because it's the uniform distribution, which is definitely not integrable, and it is also not a conjugate prior. The posterior is not a uniform distribution.

Generating Speech Output

Now, we examine the relationship between the prior and the posterior distributions in parts (a) and (b). Gamma distributions are distinct, so if the prior in (a) tends to (b) as a limit is taken, then it has to take the Gamma distribution parameters to the corresponding ones. The only different prior is  $\theta$ , which is  $\frac{1}{a+\sum X_i}$  in (a) and  $\frac{1}{\sum X_i}$  in (b), and it clear that taking  $a$  to 0 in  $\frac{1}{a+\sum X_i}$  would give  $\frac{1}{\sum X_i}$ ; on the other hand, taking  $a$  to infinity would take it to 0, not  $\frac{1}{\sum X_i}$ .

It is also true that given more and more observations, the two posterior distributions will converge to the same mean. Indeed, the mean of a Gamma distribution is  $k\theta$ , so the distribution in (a) has mean  $\frac{n}{a+\sum X_i}$  and the distribution in (b) has mean  $\frac{n}{\sum X_i}$ . Writing the mean of  $X_1, \dots, X_n$  as  $\bar{X}_n$  allows us to restate the posterior mean in (a) as  $\frac{1}{\frac{a}{n} + \bar{X}_n}$  and the posterior mean in (b) as  $\frac{1}{\bar{X}_n}$ .  $\bar{X}_n$  converges in probability to the theoretical mean of  $X_i$ , which is  $\frac{1}{\lambda}$  as  $X_i$  is drawn from the distribution  $\text{Exp}(\lambda)$ .

Thus, as  $n$  gets large,  $\frac{1}{\frac{a}{n} + \bar{X}_n}$  converges to

$$\lim_{n \rightarrow \infty} \frac{1}{\frac{a}{n} + \bar{X}_n} = \frac{1}{\frac{1}{\lambda}} = \lambda$$

, and so does  $\frac{1}{\bar{X}_n}$  (as we can plug in  $\frac{1}{\lambda}$  into  $\bar{X}_n$  by Slutsky's Theorem to get a limit of  $\frac{1}{\frac{1}{\lambda}} = \lambda$ ).

Submit

You have used 3 of 3 attempts

**i** Answers are displayed within the problem

(c)

3.0/3 points (graded)

Suppose we have the prior  $\pi(\lambda) \sim \text{Gamma}(a, b)$  (where  $a, b > 0$ , and conditional on  $\lambda$ , we have observations  $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$ ). Compute the posterior distribution  $\pi(\lambda | X_1, X_2, \dots, X_n)$ .

Generating Speech Output

the posterior distribution for  $\lambda$  is an Gamma distribution. What are its parameters? Enter your answer in terms of  $a, b, n$ , and  $\sum_{i=1}^n X_i$ .

(Enter **Sigma\_i(X\_i)** for  $\sum_{i=1}^n X_i$ . Do not worry if the parser does not render properly; the grader works independently. If you wish to have proper rendering, enclose **Sigma\_i(X\_i)** by brackets. )

$\theta =$   ✓ Answer: 1/(1/b+Sigma\_i(X\_i))

$k =$   ✓ Answer: a+n

Now, we examine two properties of the prior. Select the correct answer choice.

- ☒ The prior is proper and is a conjugate prior.
- ☐ The prior is improper and is a conjugate prior.
- ☐ The prior is proper and is not a conjugate prior.
- ☐ The prior is improper and is not a conjugate prior.



#### Solution:

This time, the prior distribution is a Gamma distribution. We try to simplify this using proportionality notation. With parameters  $a$  and  $b$  for  $k$  and  $\theta$ , respectively, we get that the pdf is equal to

$$\pi(\lambda) = \frac{\lambda^{a-1} e^{-\frac{\lambda}{b}}}{\Gamma(a) b^a}.$$

We are, however, only interested in terms that vary with  $\lambda$ , so we are able to simplify it to

Generating Speech Output

$$\pi(\lambda) \propto \lambda^{a-1} e^{-\frac{\lambda}{b}},$$

treating  $\Gamma(a) b^k$  in the denominator as a constant.

As our conditional likelihood function is the same for each  $X_i$ , we can use the same  $\pi(X_1, \dots, X_n | \lambda)$  from the previous parts, which is

$$\pi(X_1, \dots, X_n | \lambda) \propto \lambda^n e^{-(\sum X_i)\lambda}.$$

Now, we apply Bayes' formula to get that

$$\begin{aligned} \pi(\lambda | X_1, \dots, X_n) &\propto \pi(\lambda) L(X_1, \dots, X_n | \lambda) \\ &\propto (\lambda^{a-1} e^{-\frac{\lambda}{b}}) (\lambda^n e^{-(\sum X_i)\lambda}) \\ &\propto e^{-(\sum X_i + \frac{1}{b})\lambda} \lambda^{a+n-1}. \end{aligned}$$

Similarly as in the previous parts, equating parameters with the Gamma distribution gives that  $k - 1 = a + n - 1$  and  $\frac{1}{\theta} = (\sum X_i) + \frac{1}{b}$ , so this converts to

$$k = a + n$$

and

$$\theta = \frac{1}{(\sum X_i) + \frac{1}{b}}.$$



The Gamma distribution is always a proper prior as it has a well-defined mean. This time, we finally have a conjugate prior because both the prior and the posterior are Gamma distributions.

Submit

You have used 2 of 3 attempts

**i** Answers are displayed within the problem

## Specific vs General Distributions

1/1 point (graded)

Consider what you've learned so far, and then try to devise a pattern relating the generality of the prior distribution with the possibility of it being a conjugate prior. Select the choice that correctly fills the blanks in the following sentence.

"Having a \_\_\_\_ prior distribution parametrization will make it more likely to have a conjugate prior, and having a \_\_\_\_ likelihood function distribution will make it more likely to have a conjugate prior."

☐ more general, more general

☒ more general, more specific

☐ more specific, more general

☐ more specific, more specific



Submit

You have used 2 of 3 attempts

## Discussion

Generating Speech Output

Bayesian statistics:Homework 9: Bayesian Statistics / 1. Practice with Priors

Hide Discussion

[◀ All Posts](#)

## [STAFF] Review grader answers for parameter K in parts (b) and (c)

question posted 6 days ago by [corderfj](#)

I got every answer for this section correct except for the parameter K in parts (b) and (c). I have reviewed my calculations and am fairly certain I got the right answers. Can you please make sure the grader is working correctly? I'm sorry if I'm missing something

This post is visible to everyone.

[Add a Response](#)

4 responses

**[arshad\\_javeed](#)**

5 days ago

+1

Add a comment

**[stelstoy](#)**

5 days ago

grader seems to be working fine - mine were marked correct. check your calculations.

Add a comment

**[sandipan\\_dey](#)**

Generating Speech Output 5 days ago

For (c), I am getting either of the following responses from the grader:

Invalid Input: a not permitted in answer as a variable

Invalid Input: b not permitted in answer as a variable



Recheck your computations.  $a$  and  $b$  are indeed parts of the answer; that does not mean that expressions for  $\theta$  and  $k$  require both  $a$  and  $b$  simultaneously though.

posted 4 days ago by [Hryhorchuk](#)



@sandipan\_dey, should be the other way round. You might have confused  $\theta$  with  $k$ .

posted 4 days ago by [Hryhorchuk](#)



For me  $\theta$  is a function of  $\sum_i X_i$  and  $a$ , whereas  $k$  is a function of  $n$  and  $b$  (also I am using proportionality to ignore expressions not containing  $\lambda$ ), i.e.,  $a$  and  $b$  are simultaneously not there, but i am getting the same invalid input warning @Hryhorchuk. Any clue? thanks in advance.

posted 4 days ago by [sandipan\\_dey](#)



Double check the order of the parameters of the Gamma distribution.

posted 4 days ago by [Erocha](#) (Community TA)



@sandipan\_dey, note the order of parameters in general form of  $\text{Gamma}(k, \theta)$ , and then in the prior  $\text{Gamma}(a, b)$ . If you start with the prior and "mutate" it by the likelihood, then it is quite unlikely that you get  $\text{Gamma}(f(b), g(a))$ , and not  $\text{Gamma}(f(a), g(b))$ ...

posted 4 days ago by [Hryhorchuk](#)



Thank you very much @Erocha, @Hryhorchuk - the order of the parameters was wrong indeed.

Generating Speech Output d 3 days ago by [sandipan\\_dey](#)

Add a comment

**markweitzman** (Community TA)

4 days ago

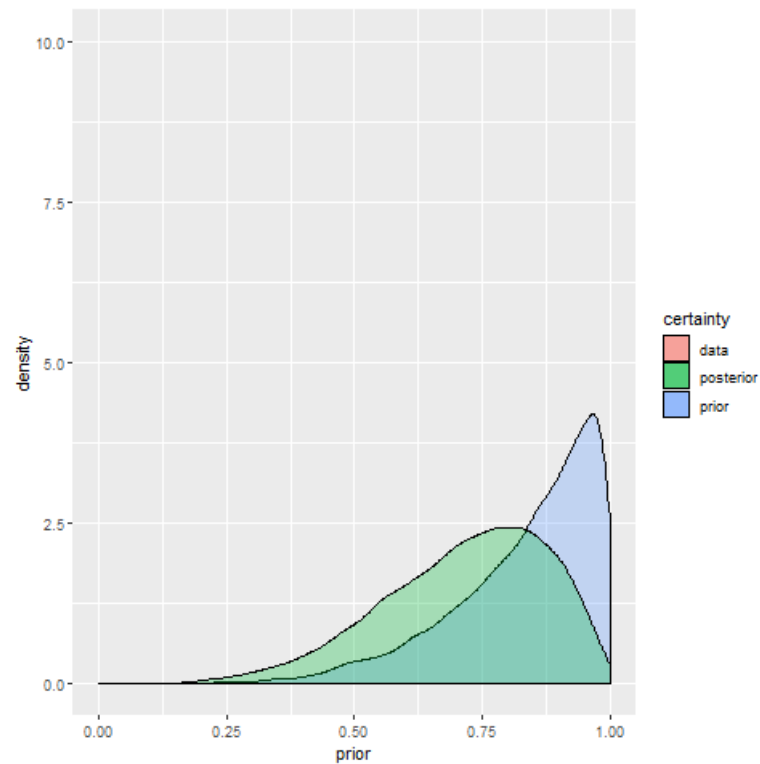


How certain is "fairly certain". Did you use the Bayesian approach and ask whether red X changed your assessment? How many previous assertions of "fairly certain" turned out to be correct v. grader error. Did you take this into account when forming a prior?

As others have noted grader is fine. You should update your "fairly certain" with this additional information :-)

Generating Speech Output

@markweitzman my implementation of your idea :, assuming online data (the x axis should be labeled 'probability' though, a typo) - with beta-Bernoulli conjugate prior.



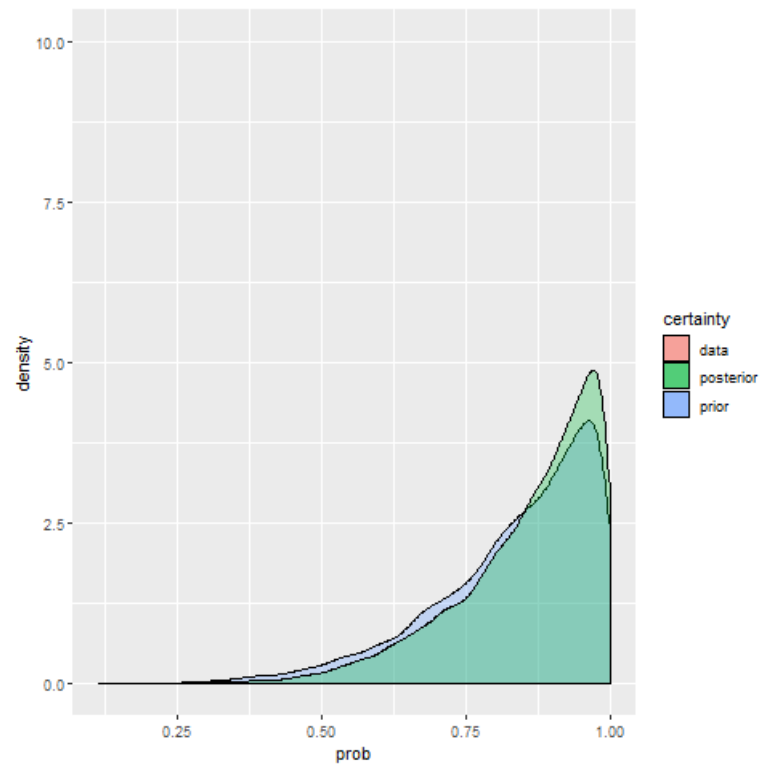
posted 3 days ago by [sandipan dey](#)

Wow, this is nice! How is the "data" plot done?

posted 3 days ago by [ya mukhin](#) (Staff)

Generating Speech Output

Thank you, I generated the data by sampling from  $X \sim \text{Ber}(p)$  online and by gradually decreasing  $p$  (probability of being certain) from 0.8 to 0.1 (supporting the decrease in certainty over time) @ya\_mukhin.



posted 2 days ago by [sandipan dey](#)

Add a comment

Showing all responses

Add a response:

Generating Speech Output

Preview

Submit

© All Rights Reserved

Generating Speech Output