edX

1. Practice with Priors

One side concept introduced 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 $\pi(\theta|X_1, X_2, \ldots, X_n)$ 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, θ)) is a probability distribution with parameters $k > 0$ and $\theta > 0$, has support on $(0, \infty)$, and whose density is given by $f(x) = \dfrac{x^{k-1}e^{-\frac{x}{\theta}}}{\Gamma(k)\,\theta^k}$. 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 $\pi(\lambda) \sim$ **Exp** (a) (where $a > 0$, and conditional on λ, we have observations $X_1, X_2, \cdots, X_n \overset{\text{i.i.d}}{\sim}$ **Exp** (λ). Compute the posterior distribution $\pi(\lambda|X_1, X_2, \ldots, X_n)$.

The posterior distribution for λ is a Gamma distribution. What are its parameters? Enter your answer in terms of a, 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.)

$\theta =$ 1/(Sigma_i(X_i)+$\boxed{1}$) 这里我把答案抄错了，但是也被判对了 ✔ **Answer:** 1/($\boxed{a}$+Sigma_i(X_i))

$k =$ n+1 ✔ **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 λ, so we only consider multiplicative terms that depend on λ, and ignore the rest.

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

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

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

For each i, $X_i \propto \mathbf{Exp}(\lambda)$ 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

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

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

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

Finally, we match our posterior distribution with the form for the Gamma distribution. The distribution is over λ, so we match λ in our posterior distribution to x in the Gamma distribution. Then we notice that the exponent of λ 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 additional λ^n terms that makes it part of the more general Gamma distribution instead.(refer to the course convention).

Submit You have used 2 of 3 attempts

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 $X_1, X_2, \cdots, X_n \overset{\text{i.i.d}}{\sim} \mathbf{Exp}(\lambda)$. Now, consider a uniform prior $\pi(\lambda) \propto 1$. Compute the posterior distribution $\pi(\lambda|X_1, X_2, \ldots, X_n)$.

Again, the posterior distribution for λ 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.)

$\theta =$ 1/(Sigma_i(X_i)) **Answer:** 1/(Sigma_i(X_i))

$k =$ n+1 **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. ✔

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

- [x] For a fixed set of observations, taking the limit as $a \to 0$ in the prior in (a) gives the same posterior as in (b). ✔
- [ ] For a fixed set of observations, taking the limit as $a \to \infty$ in the prior in (a) gives the same posterior as in (b).
- [x] Assuming a true parameter λ, 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 λ is the same, so we have the same conditional likelihood function $L\left(X_1, \ldots, X_n | \lambda\right)$. The only part that changes is that our posterior distribution here is $\pi\left(lambda\right) \propto 1$, the uniform distribution.

Now, we apply Bayes' formula again to get that

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

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 $-\left(a+\sum X_i\right)\lambda$ (in the posterior) $-\frac{x}{\theta}$ (in the Gamma distribution). Equating then gives (taking into account that λ 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 $\pi\left(\lambda\right)$ is not a proper prior because it's the uniform distribution, which is definitely not integrable, and it is also not a conjugate prior because the posterior is not a uniform distribution.

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 θ, 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 $X_1, \ldots X_n$ 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 $\textsf{Exp}\left(\lambda\right)$.

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

$$\lim_{n\to\infty} \frac{1}{\frac{a}{n}+\lambda} = \frac{1}{\frac{1}{\lambda}} = \lambda$$

, and so does $\frac{1}{\frac{1}{\bar{X}_n}}$ (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

Answers are displayed within the problem

(c)

3.0/3 points (graded)

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

Yet again, the posterior distribution for λ 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.)

$\theta =$ 1/(Sigma_i(X_i)+1/b) ✔ **Answer:** 1/(1/b+Sigma_i(X_i))

$k =$ n+a ✔ **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 θ, respectively, we get that the pdf is equal to

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

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

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

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 $\pi(X_1, \ldots, X_n | \lambda)$ from the previous parts, which is

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

Now, we apply Bayes' formula to get that

$$\pi(\lambda | X_1, \ldots, X_n) \propto \pi(\lambda)\, L(X_1, \ldots, X_n | \lambda)$$

$$\propto (\lambda^{a-1} e^{-\frac{\lambda}{b}})\,(\lambda^{n} e^{-(\sum X_i)\lambda})$$
$$\propto e^{-(\sum X_i + \frac{1}{b})\lambda} \lambda^{a+n-1}.$$

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 1 of 3 attempts

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

Solution:

As we could observe from the above scenario, it is when we have a **more general** prior distribution that it is more likely for the posterior to contain it. This is especially true if the specific distribution is contained in the general distribution.

On the other hand, having a general likelihood function distribution, such as the Gamma function, will introduce more terms in the pdf that will increase the number of variables in the posterior as well. This makes the posterior more general, so it's less likely to be contained in the posterior distribution and hence less likely to be a conjugate prior. So, having a **more specific** likelihood function distribution will make it more likely to have a conjugate prior.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 5 Bayesian statistics:Homework 9: Bayesian Statistics / 1. Practice with Priors