4. Review: the Beta Distribution

Beta Distribution Conjugacy

2/2 points (graded)

In this problem, we will explore the use of the Beta distribution family in Bayesian statistics. This family of distributions is especially suited for the Bayesian framework because it covers a large variety of distribution shapes, only uses two parameters, and simplifies computation as updates may be done by simple addition to the parameters.

For this particular problem, our parameter of interest is p. Our prior distribution is $\text{Beta}\,(\alpha, \beta)$, and conditional on p, we have observations $X_1, X_2, \cdots X_n, \overset{\text{i.i.d}}{\sim} \text{Ber}\,(p)$. As discussed in lecture, the posterior distribution is also a Beta distribution. What are its parameters?

Use **SumXi** for $\sum_{i=1}^{n} X_i$.

$\alpha =$ a + SumXi

$a + SumXi$

✔ **Answer:** a+SumXi+b*0+n*0

$\beta =$ b + n - SumXi

$b + n - SumXi$

✔ **Answer:** b+n-SumXi+a*0

STANDARD NOTATION

Solution:

We use Bayes' theorem, simplified through proportionality notation, for the posterior:

$$\pi\left(p|X_1, \ldots, X_n\right) \propto \pi\left(p\right) L_n\left(X_1, \ldots, X_n|p\right).$$

Throughout this solution, we work in proportionality notation.

- By the definition of the Beta distribution, our prior distribution $\text{Beta}\,(a, b)$ is simply $\pi\left(p\right) \propto p^{a-1}(1-p)^{b-1}$.
- As our observations are i.i.d., we get

$$L_n\left(X_1, \ldots, X_n|p\right) = \prod_{i=1}^{n} L\left(X_i|p\right).$$

Here, $L\left(X_i|p\right) = p^{X_i}\left(1-p\right)^{1-X_i}$ as our observations are from the $\text{Ber}\,(p)$ distribution. Combining like factors, we get that

$$\begin{aligned} L_n\left(X_1, \ldots, X_n|p\right) &= \prod_{i=1}^{n} p^{X_i}\left(1-p\right)^{1-X_i} \\ &= p^{X_1+\ldots+X_n}\left(1-p\right)^{(1-X_1)+\ldots+(1-X_n)} \\ &= p^{\sum_{i=1}^{n} X_i}\left(1-p\right)^{n-\sum_{i=1}^{n} X_i}. \end{aligned}$$

Combining, we get that

$$\begin{aligned} \pi\left(p|X_1, \ldots, X_n\right) &= \left(p^{a-1}(1-p)^{b-1}\right)\left(p^{\sum_{i=1}^{n} X_i}\left(1-p\right)^{n-\sum_{i=1}^{n} X_i}\right) \\ &= p^{(a-1)+\left(\sum_{i=1}^{n} X_i\right)}\left(1-p\right)^{(b-1)+\left(n-\sum_{i=1}^{n} X_i\right)} \end{aligned}$$

We recall that the Beta distribution has the form Beta $(a, b) \propto p^{\alpha-1}(1-p)^{\beta-1}$, so comparing this form to our expression for the posterior gives

$$\alpha = 1 + ((a-1) + (\sum_{i=1}^{n} X_i)) = \boxed{a + \sum_{i=1}^{n} X_i}$$

and

$$\beta = 1 + ((b-1) + (n - \sum_{i=1}^{n} X_i)) = \boxed{b + n - \sum_{i=1}^{n} X_i}.$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Changing the Prior Distribution

1/1 point (graded)

Consider the case where we still have observations $X_1, X_2, \cdots X_n, \overset{\text{i.i.d}}{\sim}$ Ber (p). Suppose we are to change our prior distribution from Beta (a, b) to one of the options listed below. Which of these will result into the posterior distribution being a Beta distribution, regardless of the observations? Assume that the likelihood function is only defined for $0 \leq p \leq 1$.

(Choose all that apply.)

- [x] Unif $(0, 1)$ ✔
- [ ] Unif $(0.3, 0.8)$
- [ ] Ber (0.3)
- [ ] N $(0.5, 0.25)$
- [x] $\pi(p) \propto \sqrt{\frac{1}{p(1-p)}}$ $(0 < p < 1)$ ✔

如果用Ber来当prior，那么概率就只有1或者0的可能

Your observation is right -- whether it's supposed to be included really depends on the application and context. The context here is that we're using the Beta distribution to model distributions over the interval $[0, 1]$, which is why we used this convention here. But definitely don't take this as something fixed, again it really depends on what you're trying to model.

posted a day ago by farrellw00 (Staff)

✔

Solution:

The desired posterior distribution must have the form $\pi(p|X_1, \ldots, X_n) \propto p^{\alpha-1}(1-p)^{\beta-1}$, for some α and β. The likelihood function would also have this form (by the previous problem and since we keep our likelihood function), and we have the relation

$$\pi(p|X_1, \ldots, X_n) = \pi(p) L_n(X_1, \ldots, X_n|p)$$

from Bayes, so we deduce that the prior $\pi(p)$ must also be of the form $p^{\alpha-1}(1-p)^{\beta-1}$ for a (generally) different α and β. Furthermore, the posterior distribution must also have a support of $[0, 1]$.

- The distribution Unif $(0, 1)$ is a suitable prior because it by definition has support on $[0, 1]$ and can be written as $p^0(1-p)^0$. In fact, it is equivalent to Beta $(1, 1)$.
- The distribution Unif $(0.3, 0.8)$ is not a suitable prior because its support is $[0.3, 0.8]$, which is not the whole interval $[0, 1]$ that's required for a Beta distribution. Even though on this prior's support, it is proportional to Beta $(1, 1)$, using this as the prior will yield a posterior with 0 weight on $[0, 0.3]$ and $[0.8, 1]$, which could not be a Beta distribution.

- The distribution $\mathsf{Ber}\,(0.3)$ is not a suitable prior because its support is the binary set $\{0,1\}$, which is not the whole interval $[0,1]$ that's required for a Beta distribution. As a side note, one could see that the posterior distribution would actually turn out to be proportional to a Bernoulli distribution, so we still have a conjugate prior in this case.

- The distribution $\mathsf{N}\,(0.5, 0.25)$ is not a suitable prior because it has an e exponential and thus can't be expressed in the form $p^{\alpha-1}(1-p)^{\beta-1}$. Note that in this case, even though the support is $\mathbb{R}$, this alone is not a sufficient argument to rule this out as our likelihood function being defined only on $[0,1]$ would constrain the posterior to $[0,1]$.

- The distribution $\pi\,(p) \propto \sqrt{\frac{1}{p(1-p)}}$ $(0 < p < 1)$ is a suitable prior because it has support $(0,1)$ and can be written as $p^{-\frac{1}{2}}\,(1-p)^{-\frac{1}{2}}$, so it is in fact $\mathsf{Beta}\,\left(\frac{1}{2}, \frac{1}{2}\right)$.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

Changing the Likelihood Function

1/1 point (graded)

Now, consider the case where our prior distribution is still $\mathsf{Beta}\,(a,b)$. Suppose we are to change our conditional likelihood model from $\mathsf{Ber}\,(p)$ to one of the options listed below, still maintaining the i.i.d. assumption for the observations. Which of these will necessarily result into the posterior distribution being a Beta distribution, regardless of the observations? Assume that we only define the likelihood over $0 \le p \le 1$. (Choose all that apply.)

- [x] $\mathsf{Geom}\,(p)$ ✔
- [ ] $\mathsf{Poiss}\,\left(\frac{1}{p}\right)$
- [ ] $\mathsf{Unif}\,([0,p])$
- [ ] $\mathsf{N}\,(0,p^2)$

✔

Solution:

By a similar reasoning as in the previous part, we need the likelihood function $L_n\,(X_1,\ldots,X_n|p)$ to be expressible, after simplifying through proportionality notation, in the form $p^{\alpha-1}(1-p)^{\beta-1}$. From the i.i.d. assumption, we could write $L_n\,(X_1,\ldots,X_n|p) = L\,(X_1|p)\ldots L\,(X_n|p)$. As the functions are the same, we would need the general form for the likelihood function $L\,(X_i|p)$ to also be of this form.

- $\mathsf{Geom}\,(p)$ will make $L\,(X_i|p) \propto p^1\,(1-p)^{X_i-1}$, so it is of this desired form.

- $\mathsf{Poiss}\,\left(\frac{1}{p}\right)$ will make $L\,(X_i|p) \propto \left(\frac{1}{p}\right)^x e^{-\frac{1}{p}}$, which is not of the desired form due to the extra $e^{-\frac{1}{p}}$ component.

- $\mathsf{Unif}\,([0,p])$ does not work because the overall likelihood $L_n\,(X_1,\ldots,X_n|p)$ would be 0 for all $p \le \min X_i$, making it undefined at some points in $[0,1]$ and hence would not be of the desired form.

- $\mathsf{N}\,(0,p^2)$ will make $L\,(X_i|p) \propto \frac{1}{p}e^{-\frac{x^2}{2p^2}}$, which is not of the desired form due to the extra $e^{-\frac{x^2}{2p^2}}$ component.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 5 Bayesian statistics:Lecture 18: Jeffrey's Prior and Bayesian Confidence Interval / 4. Review: the Beta Distribution