Bayesian Statistics VI

Matthieu Vignes

April 2024

Bayesian inference for a binomial proportion

We can now assemble the pieces together and show how to perform Bayesian inference for a continuous parameter, specifically a binomial proportion. It illustrates the mechanics of how we actually calculate the posterior distribution.

We still suppose that we sample 100 elephants from a population, and measure their DNA at a location in their genome ("locus") where there are two types ("alleles"), which it is convenient to label `0` and `1`. In my sample, I observe that 30 of the elephants have the `1` allele and 70 have the `0` allele. What can I say about the frequency (q) of the `1` allele in the population?

Bayesian inference: calculating the posterior

Here we are doing inference for a parameter q that can, in principle, take any value between 0 and 1. That is, we are doing inference for a "continuous" parameter. Bayesian inference for a continuous parameter proceeds in essentially exactly the same way as Bayesian inference for a discrete quantity, except that probability mass functions get replaced by densities.

Specifically remember the form of Bayes Theorem:

$$\text{Posterior} \propto \text{likelihood} \times \text{prior}.$$

We have seen that the likelihood for q is

$$L(q) := P(D|q) \propto q^{30}(1-q)^{70}.$$

You might have heard this likelihood called the "binomial likelihood", because it arises when the data D come from a binomial distribution.

Prior

Recall that the prior distribution is a distribution that is supposed to reflect what we know about q prior to seeing the data. For the purposes of illustration we will assume a uniform prior on q: $q \sim \text{Unif}[0,1]$. That is:

$$p(q) = 1, \ q \in [0,1] \text{ (and 0 otherwise)}.$$

This $\text{Unif}[0,1]$ prior says many things. For example, it says that before seeing the data the idea that $q < 0.5$ is just as plausible as $q > 0.5$. And it says that $q < 0.1$ is just as plausible as $q > 0.9$, or $0.4 < q < 0.5$. And that it is five times less likely than $q < 0.5$ or $q > 0.5$. Etc. If for some reason, these are not equally plausible, then you should use a different prior, at least in principle. However, in practice it is sometimes (but not always) the case that the results of Bayesian inference are robust to the choice of prior distribution, so in such cases, it is common not to worry too much about minor deviations between what you believe and what the prior implies.

For now, we are simply aiming to show how the Bayesian calculations are done under this prior specification.

Posterior calculation

Using Bayes Theorem to combine the prior distribution and the likelihood we obtain:

$$p(q|D) \propto p(D|q)p(q) \propto q^{30}(1-q)^{70} \ (q \in [0,1]).$$

Here, because q is a continuous parameter, this is referred to as the "posterior density" for q.

Now the final "trick" is to notice that this density, $q^30(1-q)^70$ is exactly the density of a Beta distribution. Indeed, specifically it is the density of a Beta$(31, 71)$ distribution. So the posterior distribution for q is Beta$(31, 71)$, and we might write $q|D \sim$ Beta$(31, 71)$.

This kind of "trick" is common in Bayesian inference: you look at the posterior density and "recognise" it as a distribution you know. It turns out that the number of distributions in common use is relatively small, so you only need to learn a few distributions to get sufficiently good at this trick for practical purposes. For example, it is a good start to be able to recognise the following distributions: exponential, binomial, Poisson, Gamma, Beta, Dirichlet, and Normal. If your posterior distribution does not look like one of these, then you may well be in a situation where you need to use computational methods (like Importance Sampling or Markov chain Monte Carlo) to do your computations.

So, in this case we are lucky: the posterior distribution is a nice distribution that we recognise, and this means we can do lots of calculations very easily: R has lots of built-in functions to deal with the Beta distribution, and many analytic properties have been derived (e.g. Wikipedia page on the Beta distribution.) We can use this to summarize and interpret the posterior distribution, as illustrated in this last section.

## Summarising and interpreting the posterior

We only show here how to summarize and interpret a posterior distribution that has been computed analytically. The idea when the posterior distribution is not known analytically, but samples can be eextracted from it, is to use enough samples to get empirical summaries. We need to check that the process that generates those samples (spoiler alert: MCMC will be used!) is working properly. For now...

We have a parameter q, whose posterior distribution we have computed to be Beta$(31, 71)$. What does this mean? What statements can we make about q? How do we obtain interval estimates and point estimates for q?

Remember that the posterior distribution represents our uncertainty (or certainty) in q, after combining the information in the data (the likelihood) with what we knew before collecting data (the prior). To get some intuition, we could plot the posterior distribution so we can see what it looks like.

```
q <- seq(0,1,length=100)
plot(q, dbeta(q, 31, 71), main = "Posterior for q", ylab = "density", type = "l")
```

Posterior for q
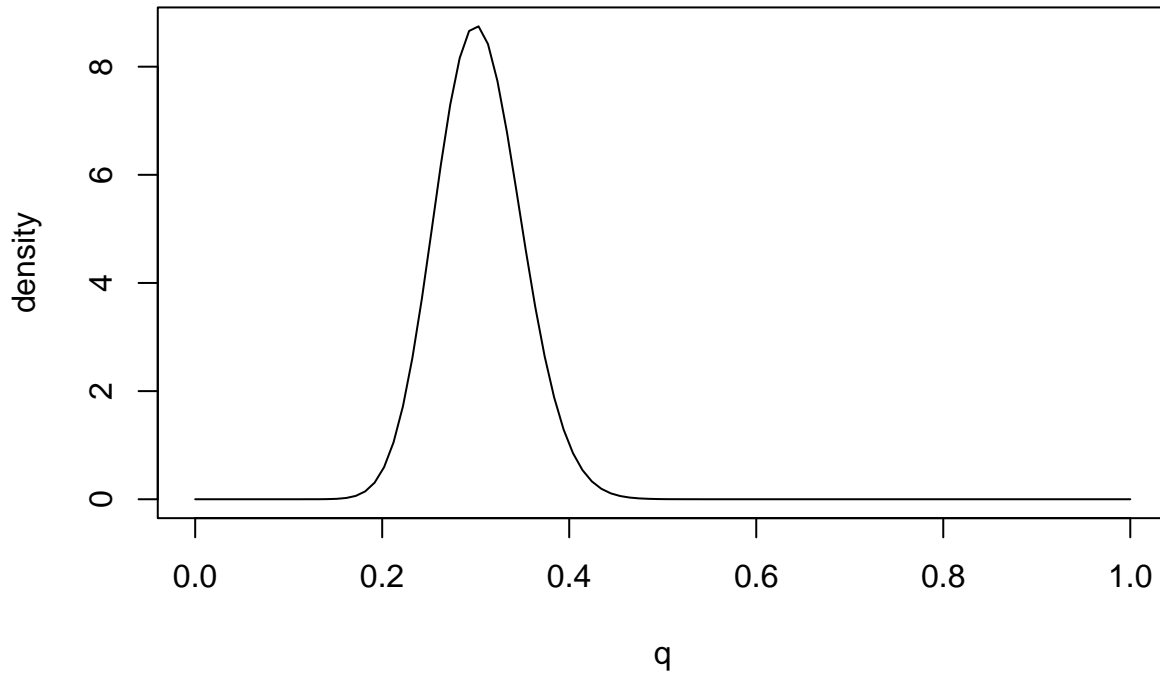
Based on this plot we can visually see that this posterior distribution has the property that q is highly likely to be less than 0.4 (say) because most of the mass of the distribution lies below 0.4. In Bayesian inference we quantify statements like this -that a particular event is "highly likely"- by computing the "posterior probability" of such an event, which is the probability of the event under the posterior distribution.

For example, in this case we can compute the (posterior) probability that $q < 0.4$ or $P(q < 0.4|D)$. Since we know the posterior distribution is a $\text{Beta}(31, 71)$ distribution, this probability is easy to compute using the `pbeta` function in R:

```r
pbeta(0.4,31,71)
```

```
## [1] 0.9792202
```

So we would say that "The posterior probability that $q < 0.4$ is 0.98".

Interval estimates

We can extend this idea to assess the certainty (or "confidence"... careful, slipery slope :)) that q lies in any interval. For example, from the plot, it looks like q will very likely lie in the interval $[0.2, 0.4]$ because most of the posterior distribution mass lies between these two numbers. To quantify how likely we compute the (posterior) probability that q lies in the interval $[0.2, 0.4]$, $P(q \in [0.2, 0.4]|D)$. Again, this can be computed using the `pbeta` function in R:

```r
pbeta(0.4,31,71) - pbeta(0.2,31,71)
```

```
## [1] 0.9721229
```

Thus, based on our prior and the data, we would be highly confident (probability approximately 97%) that q lies between 0.2 and 0.4. That is, $[0.2, 0.4]$ is a 97% Bayesian Confidence Interval for q. Bayesian Confidence Intervals are often referred to as "Credible Intervals" (and also often abbreviated to CI) or compatilibility interval (McElreath 2021).

In practice, it is more common to compute Bayesian Confidence Intervals the other way around: specify the

level of confidence we want to achieve and find an interval that achieves that level of confidence. This can be done by computing the *quantiles* of the posterior distribution. For example, the 0.05 and 0.95 quantiles of the posterior would define a 90% Bayesian Confidence Interval.

In our example, these quantiles of the Beta distribution can be computed using the `qbeta` function in R:

```r
qbeta(0.05,31,71)
```

```
## [1] 0.2315858
```

```r
qbeta(0.95,31,71)
```

```
## [1] 0.38065
```

So $[0.23, 0.38]$ is a 90% Bayesian Confidence Interval for q. It is 90% because there is a 5% chance of it being above 0.23 and 5% of it being above 0.38. Notice that other quantiles can produce a 90% Bayesian CI. We sometimes look for the *highest posterior density interval* (HPDI), which are often uniquely defined.

Point estimates

In some cases we might be happy to give our "best guess" for q, rather than worrying about our uncertainty. That is, we might be interested in giving a *point estimate* for q. Essentially this boils down to summarising the posterior distribution by a single number.

When q is a continuous-valued variable, as here, the most common Bayesian point estimate is the mean (or expectation) of the posterior distribution, which is called the *posterior mean*. The mean of the Beta$(31, 71)$ distribution is $31/(31 + 71) = 0.3$. So we would say "The posterior mean for q is 0.3."

An alternative to the mean is the median. The median of the Beta(31,71) distribution can be found using qbeta:

```r
qbeta(0.5, 31, 71)
```

```
## [1] 0.3026356
```

So we would say "The posterior median for q is 0.3".

The mode of the posterior (*posterior mode*) is another possible summary, although this perhaps makes more sense in settings where q is a discrete variable rather than a continuous variable as here.

Exercises

1. Suppose you are interested in a parameter μ and obtain a posterior distribution for μ to be normal with mean 0.2 and standard deviation 0.4. Find

   a. a 90% Credible Interval for μ.
   b. a 95% Credible Interval for μ.
   c. a point estimate for μ.

2. Explore the Beta distribution. See the wikipedia page above. Plot the density (is it intuitive to you?) for different values of the parameters a and b. Equal, non-equal values, above or below 1. What is the spacial case $a = b = 1$? Try to get sketches before plotting them using `dbeta`.

Summary

- To compute the posterior density of a continuous parameter, up to a normalising constant, you multiply the likelihood by the prior density.

- In simple cases you may find that the result is the density of a distribution you recognise. If so, you can often use known properties of that distribution to compute quantities of interest.

- In cases where you do not recognize the posterior distribution, you may need to use computational methods to compute quantities of interest.