

Bayesian inference for a binomial proportion

Matthew Stephens

University of Chicago

January 12, 2026

See [here](#) for a PDF version of this vignette.

Overview

This vignette illustrates how to perform Bayesian inference for a continuous parameter, specifically a binomial proportion. Specifically, it illustrates the mechanics of how we actually calculate the posterior distribution.

You should be familiar with the concepts of the [likelihood function](#) and [Bayesian inference for discrete random variables](#). You should also be familiar with the binomial distribution and the [Beta distribution](#).

Motivation

Technical note: To simplify this problem, I have assumed that elephants are haploid, which they are not. If you do not know what this means then you should simply ignore this comment.

Suppose we sample 100 elephants from a population, and measure their DNA at one location (“locus”) in their genome, where there are two types (“alleles”). We label these alleles as “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 the same way as Bayesian inference for a [discrete quantity](#) except that probability mass functions get replaced by densities.

Remember Bayes Theorem:

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

To apply this, we need to have both the prior distribution and the likelihood.

Likelihood

Here, the likelihood for q is

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

where D here denotes the data. This expression comes from the fact that the data consist of 30 “1” alleles (each of which occur with probability q) and 70 “0” alleles (each of which occur with probability $1 - q$), and we assume that the samples are independent. (You might have heard this likelihood called the “binomial likelihood”, because it arises when the data come from a binomial distribution.)

Prior

Recall that the prior distribution is a distribution that is supposed to reflect what we know about q before (“prior to”) seeing the data. For illustration, we will assume a uniform prior on q ,

$$q \sim U[0, 1].$$

That is,

$$p(q) = 1, \quad q \in [0, 1].$$

This prior says many things. For example, it says that, before seeing the data, a q less than 0.5 is just as plausible as a q greater than 0.5. It also says that a q less than 0.1 is just as plausible as a q greater than 0.9 or a q between 0.5 and 0.5. If for some reason these are not equally plausible, then you should use a different prior. 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, and in such cases it is common not to worry too much about minor discrepancies 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.

Posterior calculation

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

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

Because q is a continuous parameter, this is called 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](#). Specifically, it is the density of a Beta(31, 71) distribution. So the posterior distribution for q is Beta(31, 71), which we write as $q | D \sim \text{Beta}(31, 71)$.

This kind of trick is common in Bayesian inference: you look at the posterior density and “recognize” 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 recognize 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 be in a situation where you need to use computational methods like [Importance Sampling][importance_saampling] or [Markov chain Monte Carlo](#) to do your computations.

In this case, we were lucky: the posterior distribution is a distribution that we recognize, and this means we can do many calculations very easily. R has many built-in functions for calculations with the Beta distribution, and many analytic properties have been derived (e.g., [see the Wikipedia](#)

[page](#).) And we can use this result to summarize and interpret the posterior distribution, as we illustrate [here](#).

Summary

- To compute the posterior density of a continuous parameter, up to a normalizing 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 recognize. If so, you can often use known properties of that distribution to compute quantities of interest. See [here](#) for an example.
- In cases where you do not recognize the posterior distribution, you may need to use computational methods like importance sampling or Markov chain Monte Carlo to compute quantities of interest.