

Bayesian inference for a binomial proportion

Matthew Stephens

University of Chicago

January 8, 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 [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 you should simply ignore this comment.]

Suppose 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 say) 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:

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

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

Likelihood

Here the likelihood for q is

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

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 *prior* to seeing the data. For the purposes of 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 $U[0, 1]$ 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 $0.4 < q < 0.5$. 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:

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

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 $q|D \sim \text{Beta}(31, 71)$.

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 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 recognize, 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](#).) We can use this to summarize and interpret the posterior distribution, as illustrated [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.
Example.
- In cases where you do not recognize the posterior distribution, you may need to use computational methods to compute quantities of interest.