Bayesian Statistics V

Matthieu Vignes

April 2024

The Likelihood Function

We have seen how one can use the likelihood ratio to compare the support in the data for two fully-specified models. In practice, we often want to compare more than two models - indeed, we often want to compare a continuum of models. This is where the idea of a likelihood function comes from.

Elephant example revisited

In our example here we assumed that the frequencies of different alleles (genetic types) in forest and savanna elephants were given to us. In practice these frequencies themselves would have to be estimated from (possibly historical) data.

For example, suppose we collect data on 100 savanna elephants, and see that 30 of them carry allele 1 at marker 1, while 70 carry the allele 0 at this same marker [again we are treating elephants as haploid to simplify things]. Intuitively we might estimate that the frequency of the 1 allele at that marker is 30/100, or 0.3. But we might think that the data are also consistent with other frequencies near 0.3. For example maybe the data are also consistent with a frequency of 0.29 also. Or 0.28? Or. . .

In this case we have many more than just two models to compare. Indeed, if we allow that the frequency could, in principle lie anywhere in the interval $[0, 1]$ (well, really. . . ?), then we have a continuum of models to compare. Specifically, for each $q \in [0, 1]$ let M_q denote the model that the true frequency of the 1 allele is q. Then, given our observation that 30 of 100 elephants carried allele 1 at marker 1, the likelihood for model M_q is, by the previous definition:

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

And the LR comparing models M_{q1} and M_{q2} is:

$$LR(M_{q1}; M_{q2})) = [q1/q2]^30[(1 - q1)/(1 - q2)]^70.$$

This is an example of what is called a parametric model. A parametric model is collection of models indexed by a parameter vector, often denoted θ, whose values lie in some parameter space, usually denoted Θ. The number of parameters included in the vector θ is called the *dimensionality* of the model or parameter space, and often denoted $\dim(\theta)$.

Here the parameter is q, the parameter space is $[0, 1]$ and the dimensionality is 1.

When computing likelihoods for parametric models, we usually dispense with the model notation and simply use the parameter value to denote the model. So instead of referring to the likelihood for M_q, we just say the "likelihood for q", write it $L(q)$. Similarly for likelihood ratios "for q_1 vs q_2".

We could plot the likelihood function as follows:

```
q <- seq(0, 1, length = 100)
L <- function(q){q^30 * (1-q)^70}
plot(q, L(q), ylab = "L(q)", xlab = "q", type = "l")
```
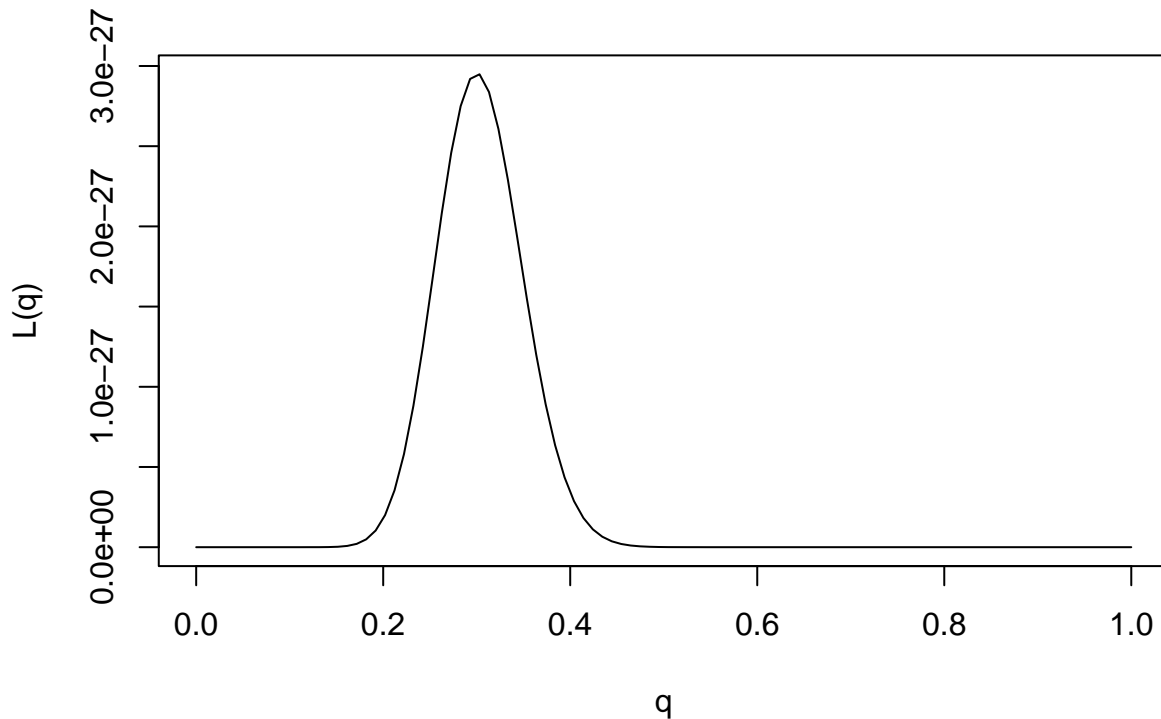

The value of θ that maximises the likelihood function is referred to as the "maximum likelihood estimate" (MLE), and usually denoted $\hat{\theta}$:

$$\hat{\theta} := \text{argmax}_\theta(L(\theta)).$$

Provided the data are sufficiently informative, and the number of parameters is not too large, MLEs tend to be sensible. In this case we can see that the maximum likelihood estimate is $q = 0.3$, which also corresponds to our intuition.

Note that from the likelihood function we can easily compute the likelihood ratio for any pair of parameter values! And just as with comparing two models, it is not the likelihoods that matter, but the likelihood ratios. That is you can divide the likelihood function by any constant without affecting the likelihood ratios. One way to emphasize this is to standardize the likelihood function so that its maximum is at 1, by dividing $L(\theta)/L(\hat{\theta})$.

```
q <- seq(0, 1, length = 100)
L <- function(q){q^30 * (1-q)^70}
plot(q, L(q)/L(0.3), ylab = "L(q)/L(qhat)", xlab = "q", type = "l")
```
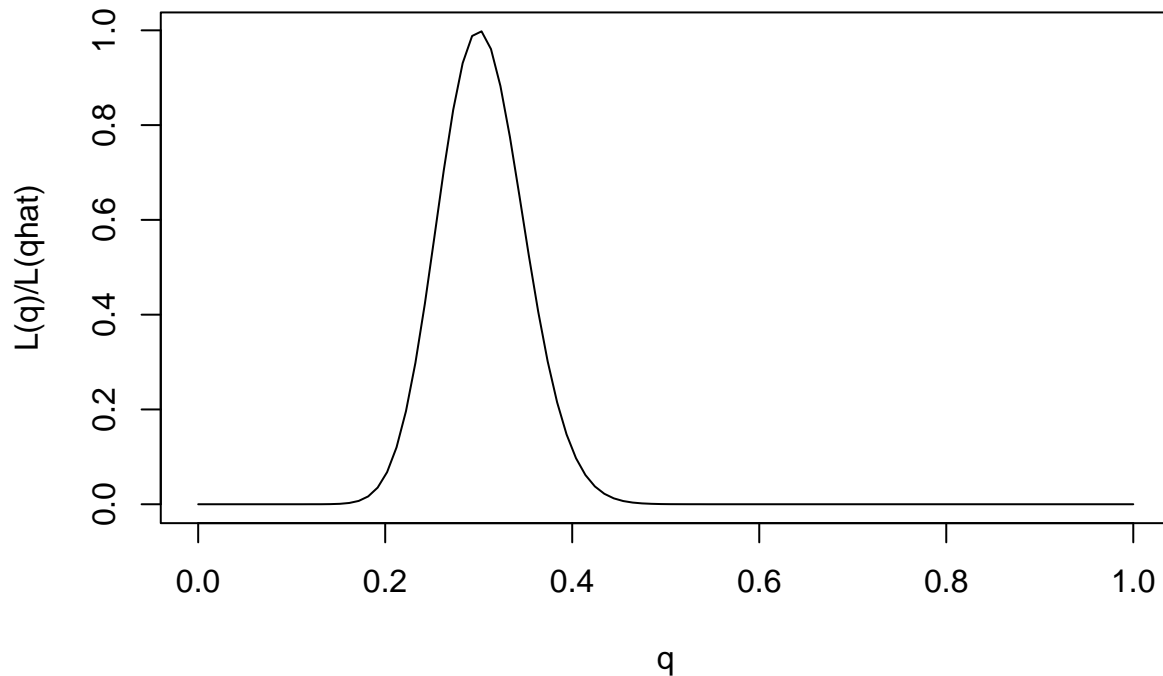
Note that for some values of q the likelihood ratio compared with $q = 0.3$ is very close to 0. These values of q are so much less consistent with the data that they are effectively excluded by the data. Just looking at the picture we might say that the values of q less than 0.15 or bigger than 0.5 are pretty much excluded by the data. We will see later how Bayesian analysis methods can be used to make this kind of argument more formal.

The log-likelihood

Just as it can often be convenient to work with the log-likelihood ratio, it can be convenient to work with the log-likelihood function, usually denoted $l(\theta)$ (lower case L). As with likelihood ratios, unless otherwise specified, we use log base e. Here is the log-likelihood function:

```r
q <- seq(0, 1, length = 100)
l <- function(q){30*log(q) + 70*log(1-q)}
plot(q, l(q) - l(0.3), ylab = "l(q)-l(qhat)", xlab = "q", type = "l")
```
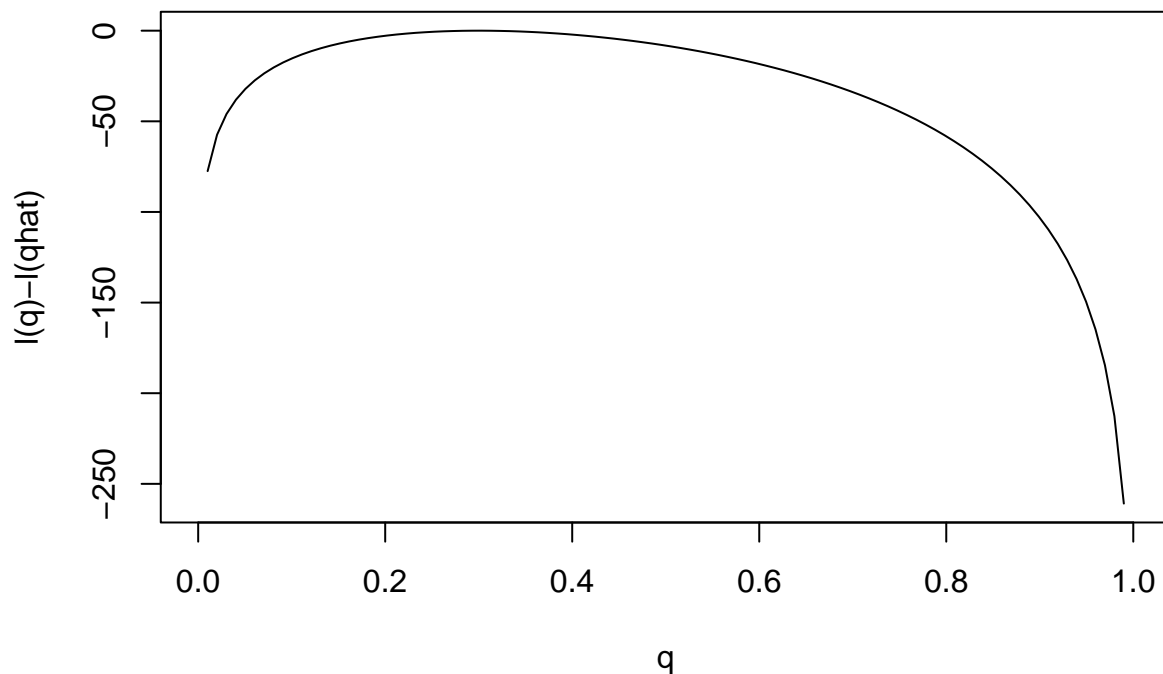
Changes in the log-likelihood function are referred to as "log-likelihood units". For example the difference in the support for $q = 0.3$ and $q = 0.35$ is $l(0.3)\text{-}l(0.35) =$ 0.5630377 log-likelihood units. Again, remember that it is differences in l that matter, not the actual values.

Notice that the scale of the y axis in this plot was set to span 10 log likelihood units. Setting the scale in this way makes sure the plot focuses on the parts of the parameter space that have more than minuscule support from the data (in this case, LR no smaller than $1/\exp(10)$). Without this the plot can be much harder to read. For example, here is the plot using the default scale selected by R:

```
plot(q, l(q)-l(0.3), ylab = "l(q) - l(qhat)", xlab = "q", type = "l")
```
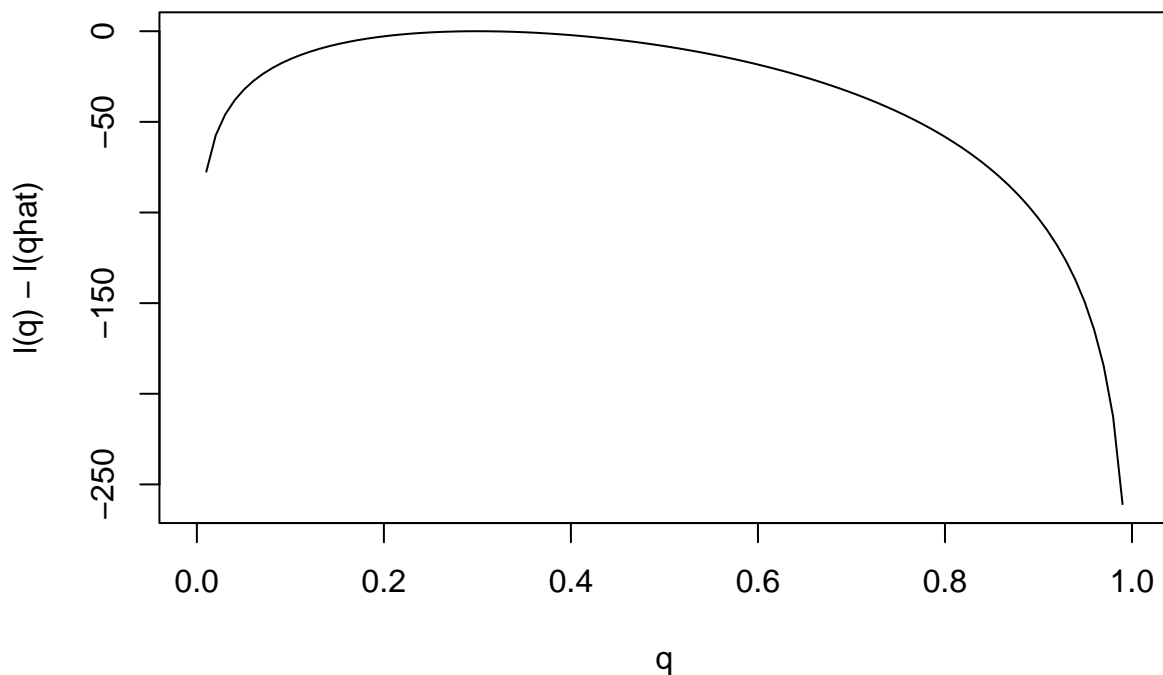

Notice how different this plot looks to the eye even though it is exactly the same curve being plotted (just

different y axis scale). It is worth thinking about what scale you use when plotting log-likelihoods (and, of course, figures in general!).