4.3 The Three Goals of Inference

Zongyi Liu

2023-06-09

4.3 The Three Goals of Inference

4.3.1 Estimation of Parameter Values

Estimation of parameter values means determining the extent to which we believe in each possible parameter value.

The posterior distribution can be narrow, with most of the probability piled heavily over a small range of θ. In this case, we fairly certain about the possible values of θ.

On the other hand, the posterior probability distribution could be wide, spread over a larger range of θ.

4.3.2 Prediction of Data Values

Using our current beliefs, we may want to predict the probability of future data values. To avoid notational conflicts later, I'll denote a data value as y. The predicted probability of data value y is determined by averaging the predicted data probabilities across all possible parameter values, weighted by the belief in the parameter values:

$$p(y) = \int d\theta p(y|\theta)p(\theta)$$

As an example, consider the prior beliefs in the top panel of Figure 4.1. For those beliefs, the predicted probability of getting a head is

$$\begin{aligned}
p(y=H) &= \sum_{\theta} p(y=H|\theta)p(\theta) \\
&= p(y=H|\theta=0.25)p(\theta=0.25) \\
&\quad + p(y=H|\theta=0.50)p(\theta=0.50) \\
&\quad + p(y=H|\theta=0.75)p(\theta=0.75) \\
&= 0.25 \times 0.25 + 0.50 \times 0.50 + 0.75 \times 0.25 \\
&= 0.5
\end{aligned}$$

4.3.3 Model Comparison

One of the nice features of Bayesian model comparison is that there is an automatic accounting for model complexity when assessing the degree to which we should believe in the model. This might be best explained with an example.

The complex model has many more available values for θ, and so it has much more opportunity to fit arbitrary data sets. For example, if a sequence of coin flips has 37% heads, the simple model does not have a θ value very close to that outcome, but the complex model does. On the other hand, for θ values that are in both the simple and complex models, the prior probability on those values in the simple model is much higher than in the complex model.

Therefore, if the actual data we observe happens to be well accommodated by a θ value in the simple model, we will believe in the simple model more than the complex model, because the prior on that θ value in the simple model is so high.

The complex model can be the winner if the data are not adequately fit by the simple model. For example, consider a case in which the observed data have just 1 head and 11 tails. None of the θ values in the simple model is very close to this outcome. But the complex model does have some θ values near the observed proportion, even though there is not a strong belief in those values.

4.3.4 Why Bayesian Inference can be Difficult

There are many reason that the Bayesian Inference might be hard, and the first is that we might need to compute a difficult integral. The traditional way is to use likelihood functions with "conjugate" prior functions. A prior function that is conjugate to the likelihood function simply makes the posterior function come out with the same functional form as the prior.

Another potential difficulty of Bayesian inference is determining a reasonable prior. What distribution of beliefs should we start with, over all possible parameter values or over competing models? This question may seem daunting, but in practice it is typically addressed in straightforward manner.

4.3.5 Bayesian Reasoning in Everyday Life

4.3.5.1 Holmesian Deduction

One example of implementing the Bayesian reasoning in everyday life has been immortalized in the words of Sherlock Holmes to his friend Dr. Watson: "How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?"

This can be reinterpreted as "How often have I said to you that when $p(D|\theta_i) = 0$ for all $i \neq j$, then, no matter how small the prior $p(\theta_j) > 0$ is, the posterior $p(\theta_j|D)$ must equal one."

4.3.5.2 Judicial Exoneration

The reverse of Holmes' logic is also commonplace. For example, when an object d'art is found fallen from its shelf, our prior beliefs may indict the house cat, but when the visiting toddler is seen dancing next to the shelf, then the cat is exonerated. This downgrading of a hypothesis is sometimes called "explaining away" of a possibility by verifying a different one. This sort of exoneration also follows from Bayesian belief updating: When $p(D\|\theta\_j)$ is higher, then, even if $p(D|\theta_i)$ is unchanged for all $i \neq j$, $p(\theta_i|D)$ is lower.