The Similarities Between Frequentist Confidence Intervals and Bayesian Credible Intervals

Ali → Ben

28 Aug 2022

1 Introduction

We're given a likelihood $p(y|x)$ for some observed data y, and we wish to represent our uncertainty in the underlying parameter x. To keep things simple, we'll consider scalar x and y. We'd like to represent this uncertainty with an interval function $I(y)$, represented as a lower function $\underline{x}(y)$ and and upper function $\bar{x}(y)$. Given an observation y, we'd like to form the narrowest such interval and yet still be able to say something like "it's very likely that the x that generated this y is somewhere in the interval $\underline{x}(y)$ to $\bar{x}(y)$."

There are two common ways to form such an interval. A Frequentist might an interval function $I_{\text{freq}}(y)$ guarantees that it covers x say, 95% of the time, if the data y were to be re-drawn infinitely many times. A Bayesian might instead compute the posterior $p(x|y)$, and report a "credible interval" function, $I_{\text{bayes}}(y)$, and guarantee that it captures 95% of the probability mass of the posterior.

The Bayesian credible interval contains the maximum a posterior estimate of the parameter x. As the width of this interval shrinks, it approaches the maximum likelihood estimate from above and from below.

But this is not the case with the Frequentist's confidence interval. As we shrink the width of the Frequentist's confidence interval, it ceases to overlap the maximum likelihood estimate of x. This leaves us in an awkward position of saying that the true value of the parameter is within a certain interval, but the most most likely value of that parameter is outside that interval. We can resolve this issue by defining the interval in a slightly different way.

2 Formal definition of the intervals

An Bayesian α-credible interval $I_{\text{bayes}}(y)$ satisfies $\forall_y \Pr_{x|y}\left[x \in I_{\text{bayes}}(y) \mid y\right] \geq \alpha$. Among all α-credible intervals, we prefer the narrowest one, so we'll cast the above as an optimization problem:

$$\begin{aligned} \min_{\underline{x},\bar{x}} \quad & \|\bar{x} - \underline{x}\|_\infty \\ \text{s.t.} \quad & \min_y \Pr_{x|y}\left[\underline{x}(y) < x < \bar{x}(y)\right] \geq \alpha. \end{aligned} \tag{Bayesian α-credible interval}$$

Here, $\|f\|_\infty \equiv \sup_x |f(x)|$, so the objective $\|\bar{x} - \underline{x}\|_\infty$ favors the narrowest interval for every observation y. This interval is sometimes called the Highest Posterior Density (HPD) interval [1].

The Frequentist's α-confidence interval satisfies $\forall_x \; \Pr_{y|x}\left[x \in I_{\text{freq}}(y)\right] \geq \alpha$. Among all α-confidence intervals, we prefer the narrowest, so we'll define the interval as the solution to an optimization problem:

$$\begin{aligned} \min_{\underline{x},\bar{x}} \quad & \|\bar{x} - \underline{x}\|_\infty \\ \text{s.t.} \quad & \min_x \Pr_{y|x}\left[\underline{x}(y) < x < \bar{x}(y)\right] \geq \alpha. \end{aligned} \tag{Frequentist α-confidence interval}$$

The frequentist objective differs from Neyman's frequentist criterion [2]. Neyman favors the interval that gives the lowest the lowest of acceptance to the wrong parameter. More precisely, if x^* is the true parameter, Neyman favors the interval $I(x)$ that minimizes $\Pr_{y|x^*}\left[\underline{x}(y) \leq x \leq \bar{x}(y)\right]$. for every $x \neq x^*$ (see section III.a of [2]). The problem with this definition is that such an interval need not exist uniformly across all x^*. It is possible for a partcIular interval to be narrowest according to this definition for one particular x^*, but it need not be narrowest for all values of x^*. Said differently, one can only know whether one has found the narrowest interval under this definition only if one knows the true value of the parameter, which defeats the purpose of estimate.

3 What happens to the intervals as $\alpha \to 0$?

As the intervals get narrowest, these intervals converge to different estimates from below and from above. Not surprisingly, I_{bayes} converges to the maximum a posteriori estimate of x. But I_{freq} converges to an interval that does not encompass the maximum likelihood estimate.

In preparation for that exposition, let's first see why the Bayesian interval converges to the MAP estimate as $\alpha \to 0$. At opt, the constraint is tight everywhere: $\forall_y \Pr_{x|y}[\underline{x}(y) < x < \bar{x}(y)] = \alpha$, and when $\alpha \to 0$, $\|\bar{x} - \underline{x}\|_\infty \to 0$, meaning that $\forall_y \bar{x}(y) - \underline{x}(y) \to 0$. A Taylor expansion of the constraint around opt gives for all y,

$$\Pr_{x|y}[\underline{x}(y) < x < \bar{x}(y)] = \int_{-\infty}^{\infty} p(x|y) \mathbf{1}[\underline{x}(y) < x] \mathbf{1}[x < \bar{x}(y)] \, dx \tag{1}$$

$$= \int_{\underline{x}(y)}^{\bar{x}(y)} p(x|y) \, dx \tag{2}$$

$$\approx p_{x|y}(\underline{x}(y)|y) \cdot (\bar{x}(y) - \underline{x}(y)). \tag{3}$$

So when $\alpha \to 0$, the optimization problem amounts to

$$\min_{\underline{x},\bar{x}} \quad \|\bar{x} - \underline{x}\|_\infty \tag{4}$$

$$\text{s.t. } \forall_y \; p_{x|y}(\underline{x}(y)|y) \cdot (\bar{x}(y) - \underline{x}(y)) = \alpha. \tag{5}$$

This can be solved pointwise at each y:

$$\min_{\underline{x}(y) \le \bar{x}(y)} \quad \bar{x}(y) - \underline{x}(y) \tag{6}$$

$$\text{s.t. } p_{x|y}(\underline{x}(y)|y) \cdot (\bar{x}(y) - \underline{x}(y)) = \alpha. \tag{7}$$

The constraint implies $\bar{x}(y) - \underline{x}(y) = \alpha / p_{x|y}(\underline{x}(y)|y)$. Plugging this back into the objective gives the optimization problem

$$\min_{\underline{x}(y) \le \bar{x}(y)} \quad \alpha / p_{x|y}(\underline{x}(y)|y),$$

or more simply, for all y,

$$\max_{\underline{x}(y)} p_{x|y}(\underline{x}(y)|y).$$

meaning that $\underline{x}(y)$ is the peak of the posterior $p(x|y)$, as expected.

Now let's apply the same program to the frequentist interval. As $\alpha \to 0$, the constraint becomes, up to a first order approximation,

$$\Pr_{y|x}[\underline{x}(y) < x < \bar{x}(y)] = \int_{-\infty}^{\infty} p(y|x) \mathbf{1}[\underline{x}(y) < x] \mathbf{1}[x < \bar{x}(y)] \, dy \tag{8}$$

$$= \int_{-\infty}^{\infty} p(y|x) \mathbf{1}\left[y < \underline{x}^{-1}(x)\right] \mathbf{1}\left[\underline{x}^{-1}(x) < y\right] \, dy \tag{9}$$

$$= \int_{\bar{x}^{-1}(x)}^{\underline{x}^{-1}(x)} p(y|x) \, dy \tag{10}$$

$$\approx p_{y|x}\left(\bar{x}^{-1}(x)|x\right) \cdot \left(\underline{x}^{-1}(x) - \bar{x}^{-1}(x)\right). \tag{11}$$

Plugging this back into the constraint gives

$$\min_{\underline{x},\bar{x}} \quad \|\bar{x} - \underline{x}\|_\infty \tag{12}$$

$$\text{s.t. } \forall_x \; p_{y|x}\left(\bar{x}^{-1}(x)|x\right) \cdot \left(\underline{x}^{-1}(x) - \bar{x}^{-1}(x)\right) = \alpha. \tag{13}$$

To simplify this, we'll rely on a first-order Taylor approximation that relates the vertical width of the interval band to its horizontal width. Using $\frac{\bar{x}(y) - \underline{x}(y)}{\underline{x}^{-1}(x) - \bar{x}^{-1}(x)} \approx \dot{\bar{x}}(y)$, we can write the problem as

$$\min_{\underline{x}^{-1},\bar{x}^{-1}} \quad \|\left(\underline{x}^{-1} - \bar{x}^{-1}\right) \cdot \dot{\underline{x}}^{-1}\|_\infty \tag{14}$$

$$\text{s.t. } \forall_x \; p_{y|x}\left(\bar{x}^{-1}(x)|x\right) \cdot \left(\underline{x}^{-1}(x) - \bar{x}^{-1}(x)\right) = \alpha, \tag{15}$$

The presence of $\dot{\underline{x}}^{-1}$ in the objective means the probelm can't be solved pointwise: the best value of $\underline{x}$ and $\bar{x}$ at a particular y depends on its value at nearby y's through $\dot{\underline{x}}$. For the interval to coincide with the peak of $p(y|x)$, the derviative must be a constant, or equivalently, $\underline{x}^{-1}$ must be a linear function of y. This happens, for example, when x is the location parameter of a distribution.

4 A Frequentist Confidence Interval That Converges to the Maximum Likelihood Estimate

Here is a Frequentist confidence interval that does converge to the maximum likelihood estimate of x as $\alpha \to 0$. Instead of penalizing $\|\bar{x} - \underline{x}\|_\infty$, we'll penalize the width of the *inverse* of the confidence interval:

$$\min_{\underline{x},\bar{x}} \quad \|\underline{x}^{-1} - \bar{x}^{-1}\|_\infty \tag{16}$$

$$\text{s.t.} \quad \min_x \Pr_{y|x} [\underline{x}(y) < x < \bar{x}(y)] \geq \alpha. \tag{17}$$

As we take $\alpha \to 0$, $\bar{x}^{-1}(x) \to \underline{x}^{-1}(x)$ pointwise, and equation (12) becomes

$$\min_{\underline{x},\bar{x}} \quad \|\underline{x}^{-1} - \bar{x}^{-1}\|_\infty \tag{18}$$

$$\text{s.t.} \ \forall_x \ p_{y|x} \left(\bar{x}^{-1}(x)|x\right) \cdot \left(\underline{x}^{-1}(x) - \bar{x}^{-1}(x)\right) = \alpha. \tag{19}$$

This problem can be solved pointwise for each x:

$$\min_{\underline{x}^{-1}(x) > \bar{x}^{-1}(x)} \quad \underline{x}^{-1}(x) - \bar{x}^{-1}(x) \tag{20}$$

$$\text{s.t.} \ p_{y|x} \left(\bar{x}^{-1}(x)|x\right) \cdot \left(\underline{x}^{-1}(x) - \bar{x}^{-1}(x)\right) = \alpha. \tag{21}$$

After substituting the constraint into the objective, this gives

$$\max_{\bar{x}^{-1}(x)} \quad p_{y|x} \left(\bar{x}^{-1}(x)|x\right) \tag{22}$$

In other words, for each x, $\bar{x}^{-1}(x)$ the most probable observation under x, or equivalently, for each observation y, $\bar{x}(y)$ is the x under which y is most probable. That's exactly the definition of the maximum likelihood estimate of x under an observation y.

5 Conclusion

I've shown that the obvious ways to penalize the width of a confidence interval gives confidence intervals that don't converge to the MLE. Neyman's 1933 paper on the subject offers a variety of other ways to canonicalize intervals, but some of them don't admit solutions, and as far as I can see, none of them converge to the MLE. For the confidence to converge to the MLE requires an uncomfortable penalty on its width.

References

[1] Edwin T Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[2] Jerzy Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 236(767):333–380, 1937.