

The Similarities Between Frequentist Confidence Intervals and Bayesian Credible Intervals

Ali \rightarrow Ben

August 27, 2022

1 Introduction

We're a likelihood $p(y|x)$, have observed some data y , and wish to represent our uncertainty in its underlying parameter x . We'd like to represent this uncertainty with an interval function $I(y)$, represented as a lower function $\underline{x}(y)$ and an upper function $\bar{x}(y)$. A Frequentist might inspect the likelihood $p(y|x)$ and report a confidence interval function $I_{\text{freq}}(y)$ and guarantee that it covers x 95% of the time if the data y are 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. To simplify matters, we'll consider scalar x and y .

In this story, both Frequentists and Bayesians seek an interval function $I(y)$. But to the Frequentist, the Bayesian makes an untenable assumption: when x is a universal constant, not a random variable, its posterior is a nonsensical object.

At the root of this dispute is what each of these camps mean by the word "probability". But I argue that the Frequentist's confidence interval has an undesirable behavior: In the limit where we seek very low probability intervals, it does not bracket the maximum likelihood estimate. The Bayesian credible interval, on the other hand, does converge to the maximum a posterior estimate.

2 The Bayesian's Credible Interval

An α -credible interval satisfies

$$\forall_y \Pr_{x|y} [x \in I_{\text{bayes}}(y) \mid y] = \alpha.$$

Since y is random, $I_{\text{bayes}}(y)$ is random, so the left hand side of this inequality is also stochastic because it is conditioned on the random variable y .

The above condition admits many solutions. Any interval that satisfies the above condition at a particular y can be shifted slightly and re-scaled slightly to obtain another interval that satisfies the condition.

Among all α -credible intervals, we prefer those that are short in some sense. To do that, we'll cast the above as an optimization problem:

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

$$\text{s.t. } \min_y \Pr_{x|y} [\underline{x}(y) < x < \bar{x}(y)] \geq \alpha. \tag{2}$$

Here, $\|f\|_{\infty} \equiv \sup_x |f(x)|$.

When $\alpha \rightarrow 0$, these intervals converge to the peak of the posterior. In the next section, I'll show that the penalizing the width of the interval this way causes the Frequentist interval to not converge to the maximum likelihood estimate. In preparation for that exposition, let's first see why the Bayesian interval converges to the MAP estimate.

At opt, the constraint is tight everywhere: $\forall_y \Pr_{x|y} [\underline{x}(y) < x < \bar{x}(y)] = \alpha$, and when $\alpha \rightarrow 0$, $\|\bar{x} - \underline{x}\|_{\infty} \rightarrow 0$, meaning that $\forall_y \bar{x}(y) - \underline{x}(y) \rightarrow 0$. A linear approximation of the constraint around opt

gives for all y ,

$$\Pr_{x|y} [\underline{x}(y) < x < \bar{x}(y)] = \int_{-\infty}^{\infty} p(x|y) \mathbf{1} [\underline{x}(y) < x] \mathbf{1} [x < \bar{x}(y)] dx \quad (3)$$

$$= \int_{\underline{x}(y)}^{\bar{x}(y)} p(x|y) dx \quad (4)$$

$$\approx p_{x|y}(\underline{x}(y)|y) \cdot (\bar{x}(y) - \underline{x}(y)). \quad (5)$$

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

$$\min_{\underline{x}, \bar{x}} \|\bar{x} - \underline{x}\|_{\infty} \quad (6)$$

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

which can be solved pointwise at each y :

$$\min_{\underline{x}(y) \geq \bar{x}(y)} \bar{x}(y) - \underline{x}(y) \quad (8)$$

$$\text{s.t. } p_{x|y}(\underline{x}(y)|y) \cdot (\bar{x}(y) - \underline{x}(y)) = \alpha. \quad (9)$$

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

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

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 $p(x|y)$.

3 The Frequentist's Confidence Interval

An α confidence interval must satisfy

$$\forall_x \Pr_{y|x} [x \in I_{\text{freq}}(y)] = \alpha. \quad (10)$$

While $I_{\text{freq}}(y)$ is random (its location and width depend on y), the left hand side of the inequality is deterministic. The only source of randomness, y , appears inside the \Pr .

Among all α -confidence intervals, we prefer those that are short in some sense. As before, it's tempting to cast the optimization problem

$$\min_{\underline{x}, \bar{x}} \|\bar{x} - \underline{x}\|_{\infty} \quad (11)$$

$$\text{s.t. } \min_x \Pr_{y|x} [\underline{x}(y) < x < \bar{x}(y)] \geq \alpha. \quad (12)$$

As we take $\alpha \rightarrow 0$, $\underline{x}(y) \rightarrow \bar{x}(y)$ pointwise, but the interval does not converge to the maximum likelihood x . Indeed,

$$\Pr_{y|x} [\underline{x}(y) < x < \bar{x}(y)] = \int_{-\infty}^{\infty} p(y|x) \mathbf{1} [\underline{x}(y) < x] \mathbf{1} [x < \bar{x}(y)] dy \quad (13)$$

$$= \int_{-\infty}^{\infty} p(y|x) \mathbf{1} [y < \underline{x}^{-1}(x)] \mathbf{1} [\underline{x}^{-1}(x) < y] dy \quad (14)$$

$$= \int_{\bar{x}^{-1}(x)}^{\underline{x}^{-1}(x)} p(y|x) dy \quad (15)$$

$$\approx p_{y|x}(\bar{x}^{-1}(x)|x) \cdot (\underline{x}^{-1}(x) - \bar{x}^{-1}(x)). \quad (16)$$

Replacing the constraint with this approximation near opt gives, as $\alpha \rightarrow 0$,

$$\min_{\underline{x}, \bar{x}} \quad \|\bar{x} - \underline{x}\|_\infty \quad (17)$$

$$\text{s.t. } \forall_x p_{y|x} (\bar{x}^{-1}(x)|x) \cdot (\underline{x}^{-1}(x) - \bar{x}^{-1}(x)) = \alpha. \quad (18)$$

To simplify this, we'll use $\frac{\bar{x}(y) - \underline{x}(y)}{\underline{x}^{-1}(x) - \bar{x}^{-1}(x)} \approx \dot{\bar{x}}(y)$ when $x = \bar{x}(y)$ to write the problem as

$$\min_{\underline{x}^{-1}, \bar{x}^{-1}} \quad \|(\underline{x}^{-1} - \bar{x}^{-1}) \cdot \dot{\bar{x}}^{-1}\|_\infty \quad (19)$$

$$\text{s.t. } \forall_x p_{y|x} (\bar{x}^{-1}(x)|x) \cdot (\underline{x}^{-1}(x) - \bar{x}^{-1}(x)) = \alpha, \quad (20)$$

The presence of $\dot{\bar{x}}^{-1}$ in the objective means we can't solve this problem pointwise. In particular, our choice of ℓ_∞ norm to penalize the width of the interval implies that the interval does not converge to the maximum likelihood estimate.

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 $\alpha \rightarrow 0$. Instead of penalizing $\|\bar{x} - \underline{x}\|_\infty$, we'll penalize the width of the inverse of the confidence interval:

$$\min_{\underline{x}, \bar{x}} \quad \|\underline{x}^{-1} - \bar{x}^{-1}\|_\infty \quad (21)$$

$$\text{s.t. } \min_x \Pr_{y|x} [\underline{x}(y) < x < \bar{x}(y)] \geq \alpha. \quad (22)$$

As we take $\alpha \rightarrow 0$, $\bar{x}^{-1}(x) \rightarrow \underline{x}^{-1}(x)$ pointwise, and equation (17) becomes

$$\min_{\underline{x}, \bar{x}} \quad \|\underline{x}^{-1} - \bar{x}^{-1}\|_\infty \quad (23)$$

$$\text{s.t. } \forall_x p_{y|x} (\bar{x}^{-1}(x)|x) \cdot (\underline{x}^{-1}(x) - \bar{x}^{-1}(x)) = \alpha. \quad (24)$$

We can solve this problem pointwise for each x :

$$\min_{\underline{x}^{-1}(x) > \bar{x}^{-1}(x)} \quad \underline{x}^{-1}(x) - \bar{x}^{-1}(x) \quad (25)$$

$$\text{s.t. } p_{y|x} (\bar{x}^{-1}(x)|x) \cdot (\underline{x}^{-1}(x) - \bar{x}^{-1}(x)) = \alpha. \quad (26)$$

After substituting the constraint into the objective, this gives

$$\max_{\bar{x}^{-1}(x)} \quad p_{y|x} (\bar{x}^{-1}(x)|x) \quad (27)$$

In other words, for each x , $y = \bar{x}^{-1}(x)$ the most probable observation under y , 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.