Are confidence intervals inference?

Suppose we have a scalar parameter θ, a random variable X with unknown distribution $\mathbb{P}(\cdot)$, and an interval-valued function $x \mapsto C(x)$ such that, no matter the distribution of X, we know that

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \text{where} \quad \mathscr{C} := 1(\theta \in C(X)) \quad (\mathscr{C} \text{ is for "cover"})$$

The interval $C(X)$ is a *valid confidence interval* for θ. This means that if we act as if $\theta \in C(X)$, we will be wrong at most 10% of the time.

Are confidence intervals inference?

Suppose we have a scalar parameter θ, a random variable X with unknown distribution $\mathbb{P}(\cdot)$, and an interval-valued function $x \mapsto C(x)$ such that, no matter the distribution of X, we know that

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \text{where} \quad \mathscr{C} := 1(\theta \in C(X)) \quad (\mathscr{C} \text{ is for "cover"})$$

The interval $C(X)$ is a *valid confidence interval* for θ. This means that if we act as if $\theta \in C(X)$, we will be wrong at most 10% of the time.

When is it reasonable to interpret $\mathscr{C}$ *inferentially*, saying that, when we observe $X = x$, that we subjectively believe that $\theta \in C(x)$ with 90% certainty?

Are confidence intervals inference?

Suppose we have a scalar parameter θ, a random variable X with unknown distribution $\mathbb{P}(\cdot)$, and an interval-valued function $x \mapsto C(x)$ such that, no matter the distribution of X, we know that

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \text{where} \quad \mathscr{C} := 1(\theta \in C(X)) \quad (\mathscr{C} \text{ is for "cover"})$$

The interval $C(X)$ is a *valid confidence interval* for θ. This means that if we act as if $\theta \in C(X)$, we will be wrong at most 10% of the time.

When is it reasonable to interpret $\mathscr{C}$ *inferentially*, saying that, when we observe $X = x$, that we subjectively believe that $\theta \in C(x)$ with 90% certainty?

Not always! Recall, for example, how we can construct silly confidence intervals. Augment the data with a draw $Z \sim \mathrm{Unif}(0, 1)$, and let

$$C(X) = \begin{cases} (-\infty, \infty) & \text{when } Z \leq 0.9 \\ [1337, 1337] & \text{otherwise} \end{cases}.$$

Obviously, no matter what the generating process, $\mathbb{P}(\mathscr{C} = 1) = 0.9$, but it is absurd to assert that we are 90% confident that $\theta = 1337$ because we observed $Z = 0.95$.

Are confidence intervals inference?

Suppose we have a scalar parameter θ, a random variable X with unknown distribution $\mathbb{P}(\cdot)$, and an interval-valued function $x \mapsto C(x)$ such that, no matter the distribution of X, we know that

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \text{where} \quad \mathscr{C} := 1(\theta \in C(X)) \quad (\mathscr{C} \text{ is for "cover"})$$

The interval $C(X)$ is a *valid confidence interval* for θ. This means that if we act as if $\theta \in C(X)$, we will be wrong at most 10% of the time.

When is it reasonable to interpret $\mathscr{C}$ *inferentially*, saying that, when we observe $X = x$, that we subjectively believe that $\theta \in C(x)$ with 90% certainty?

Not always! Recall, for example, how we can construct silly confidence intervals. Augment the data with a draw $Z \sim \mathrm{Unif}(0, 1)$, and let

$$C(X) = \begin{cases} (-\infty, \infty) & \text{when } Z \le 0.9 \\ [1337, 1337] & \text{otherwise} \end{cases}.$$

Obviously, no matter what the generating process, $\mathbb{P}(\mathscr{C} = 1) = 0.9$, but it is absurd to assert that we are 90% confident that $\theta = 1337$ because we observed $Z = 0.95$.

How can we characterize generally and precisely what went wrong?

A pathological confidence interval

Write beliefs as $\mathbb{B}(\cdot)$, to contrast with aleatoric probabiliites $\mathbb{P}()$. So we ask when

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \Rightarrow \quad \mathbb{B}(\mathscr{C} = 1 | X = x) = 0.9$$

A pathological confidence interval

Write beliefs as $\mathbb{B}(\cdot)$, to contrast with aleatoric probabiliites $\mathbb{P}()$. So we ask when

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \Rightarrow \quad \mathbb{B}(\mathscr{C} = 1 | X = x) = 0.9$$

Note that Bayesians define priors and completely specified data generating processes and insist that $\mathbb{B}() = \mathbb{P}()$.

Certainly that suffices. But is it necessary?

A pathological confidence interval

Write beliefs as $\mathbb{B}(\cdot)$, to contrast with aleatoric probabiliites $\mathbb{P}()$. So we ask when

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \Rightarrow \quad \mathbb{B}(\mathscr{C} = 1 | X = x) = 0.9$$

Note that Bayesians define priors and completely specified data generating processes and insist that $\mathbb{B}() = \mathbb{P}()$.

Certainly that suffices. But is it necessary?

It is often difficult to plausibly specify everything needed for Bayes. In such cases it can be hard to assert that $\mathbb{B}() = \mathbb{P}()$.

A pathological confidence interval

Write beliefs as $\mathbb{B}(\cdot)$, to contrast with aleatoric probabiliites $\mathbb{P}()$. So we ask when

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \Rightarrow \quad \mathbb{B}(\mathscr{C} = 1 | X = x) = 0.9$$

Note that Bayesians define priors and completely specified data generating processes and insist that $\mathbb{B}() = \mathbb{P}()$.

Certainly that suffices. But is it necessary?

It is often difficult to plausibly specify everything needed for Bayes. In such cases it can be hard to assert that $\mathbb{B}() = \mathbb{P}()$.

We may also want to trade off mathematical or computational effort to achieve $\mathbb{B}() \approx \mathbb{P}()$. Bayes gives no real guidance for doing so.

A pathological confidence interval

Write beliefs as $\mathbb{B}(\cdot)$, to contrast with aleatoric probabiliites $\mathbb{P}()$. So we ask when

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \Rightarrow \quad \mathbb{B}(\mathscr{C} = 1 | X = x) = 0.9$$

Note that Bayesians define priors and completely specified data generating processes and insist that $\mathbb{B}() = \mathbb{P}()$.

Certainly that suffices. But is it necessary?

It is often difficult to plausibly specify everything needed for Bayes. In such cases it can be hard to assert that $\mathbb{B}() = \mathbb{P}()$.

We may also want to trade off mathematical or computational effort to achieve $\mathbb{B}() \approx \mathbb{P}()$. Bayes gives no real guidance for doing so.

I argue that potential answers may be found in *fiducial inference*.

Here, I will follow Ian Hacking's book, *The Logic of Statisical Inference*.

Fiducial inference for confidence intervals

Fiducial inference for confidence intervals requires three key assumptions. The first two are uncontroversial, the third is where things go wrong.

Fiducial inference for confidence intervals

Fiducial inference for confidence intervals requires three key assumptions. The first two are uncontroversial, the third is where things go wrong.

Assumption 1: The logic of support. Formally, any coherent belief function $\mathbb{B}()$ obeys Kolmogorov's axioms in the natural ways. Examples:

- If proposition A and B are mutually incompatible, then $\mathbb{B}(A|B) = 0$.
- If B provides no information about A, then $\mathbb{B}(A|B) = \mathbb{B}(A)$.
- If $B \Rightarrow A$, then $\mathbb{B}(A|B) = 1$. And so on.

The logic of support is needed to even write and manipulate $\mathbb{B}(\cdot)$.

Fiducial inference for confidence intervals

Fiducial inference for confidence intervals requires three key assumptions. The first two are uncontroversial, the third is where things go wrong.

Assumption 1: The logic of support. Formally, any coherent belief function $\mathbb{B}()$ obeys Kolmogorov's axioms in the natural ways. Examples:

- If proposition A and B are mutually incompatible, then $\mathbb{B}(A|B) = 0$.
- If B provides no information about A, then $\mathbb{B}(A|B) = \mathbb{B}(A)$.
- If $B \Rightarrow A$, then $\mathbb{B}(A|B) = 1$. And so on.

The logic of support is needed to even write and manipulate $\mathbb{B}(\cdot)$.

Assumption 2: The frequency principle. If $\mathbb{P}(X)$ is known, then our subjective beliefs correspond with aleatoric probabilities. That is, $\mathbb{B}(X = x) = \mathbb{P}(X = x)$.

Fiducial inference for confidence intervals

Fiducial inference for confidence intervals requires three key assumptions. The first two are uncontroversial, the third is where things go wrong.

Assumption 1: The logic of support. Formally, any coherent belief function $\mathbb{B}()$ obeys Kolmogorov's axioms in the natural ways. Examples:

- If proposition A and B are mutually incompatible, then $\mathbb{B}(A|B) = 0$.
- If B provides no information about A, then $\mathbb{B}(A|B) = \mathbb{B}(A)$.
- If $B \Rightarrow A$, then $\mathbb{B}(A|B) = 1$. And so on.

The logic of support is needed to even write and manipulate $\mathbb{B}(\cdot)$.

Assumption 2: The frequency principle. If $\mathbb{P}(X)$ is known, then our subjective beliefs correspond with aleatoric probabilities. That is, $\mathbb{B}(X = x) = \mathbb{P}(X = x)$.

The third is where things can go wrong for confidence intervals.

Assumption 3: Irrelevance. The precise value of the data $X = x$ is not subjectively informative about whether $\theta \in C(x)$. That is,

$$\mathbb{B}(\theta \in C(x)|X = x) = \mathbb{B}(\theta \in C(x)).$$

Fiducial inference for confidence intervals

Assumption 1: The logic of support.

Assumption 2: The frequency principle.

Assumption 3: Irrelevance.

Confidence intervals are valid inference when

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \Rightarrow \quad \mathbb{B}(\mathscr{C} = 1 | X = x) = 0.9.$$

The above three assumptions are sufficient.

Fiducial inference for confidence intervals

Assumption 1: The logic of support.

Assumption 2: The frequency principle.

Assumption 3: Irrelevance.

Confidence intervals are valid inference when

$$\mathbb{P}(\mathscr{C} = 1) = 0.9 \quad \Rightarrow \quad \mathbb{B}(\mathscr{C} = 1 | X = x) = 0.9.$$

The above three assumptions are sufficient.

Proof:

$$\begin{aligned} \mathbb{B}(\mathscr{C} = 1 | X = x) &= \mathbb{B}(\mathscr{C} = 1) && \text{Irrelevance} \\ &= \mathbb{P}(\mathscr{C} = 1) && \text{The frequency principle} \\ &= \mathbb{P}(\theta \in C(X)) = 0.9. && \text{Construction of } C(\cdot) \end{aligned}$$

The pathological example is caught

Clearly enough, the irrelevance assumption is where things can go wrong. Let's look at our pathological example.

$$C(x) = \begin{cases} (-\infty, \infty) & \text{when } z \leq 0.9 \\ [1337, 1337] & \text{otherwise} \end{cases}.$$

Irrelevance: The precise value of the data $X = x$ is not subjectively informative about whether $\theta \in C(x)$. That is,

$$\mathbb{B}\left(\theta \in C(x) | X = x\right) = \mathbb{B}\left(\theta \in C(x)\right).$$

Our pathological example fails the principle of irrelevance, since knowing $z \geq 0.9$ is very informative about whether $\theta \in C(x)$.

How to use this?

The *invalidity* of a CI can be demonstrated by an ability to predict $1(\theta \in C(x))$ from x.

How to use this?

The *invalidity* of a CI can be demonstrated by an ability to predict $1(\theta \in C(x))$ from x.

Given a candidate confidence interval, constructed using any method, this renders the validity of inference *quantitatively falisifiable*, e.g. through simulation and ML.

How to use this?

The *invalidity* of a CI can be demonstrated by an ability to predict $1(\theta \in C(x))$ from x.

Given a candidate confidence interval, constructed using any method, this renders the validity of inference *quantitatively falisifiable*, e.g. through simulation and ML.

It also admits *degrees of valid inference*, e.g. in the sense that $1(\theta \in C(x))$ may be only slightly predicted by x.

How to use this?

The *invalidity* of a CI can be demonstrated by an ability to predict $1(\theta \in C(x))$ from x.

Given a candidate confidence interval, constructed using any method, this renders the validity of inference *quantitatively falisifiable*, e.g. through simulation and ML.

It also admits *degrees of valid inference*, e.g. in the sense that $1(\theta \in C(x))$ may be only slightly predicted by x.

Computation of $C(x)$ and predictability of $1(\theta \in C(x))$ can in principle be explicitly traded off against one another.

How to use this?

The *invalidity* of a CI can be demonstrated by an ability to predict $1(\theta \in C(x))$ from x.

Given a candidate confidence interval, constructed using any method, this renders the validity of inference *quantitatively falisifiable*, e.g. through simulation and ML.

It also admits *degrees of valid inference*, e.g. in the sense that $1(\theta \in C(x))$ may be only slightly predicted by x.

Computation of $C(x)$ and predictability of $1(\theta \in C(x))$ can in principle be explicitly traded off against one another.

When you think about how, in practice, to evaluate predictability of $1(\theta \in C(x))$, you are forced to grapple explicitly with priors, power, and misspecification, but in *simulation*, not necessarily *modeling*.

How to use this?

The *invalidity* of a CI can be demonstrated by an ability to predict $1(\theta \in C(x))$ from x.

Given a candidate confidence interval, constructed using any method, this renders the validity of inference *quantitatively falisifiable*, e.g. through simulation and ML.

It also admits *degrees of valid inference*, e.g. in the sense that $1(\theta \in C(x))$ may be only slightly predicted by x.

Computation of $C(x)$ and predictability of $1(\theta \in C(x))$ can in principle be explicitly traded off against one another.

When you think about how, in practice, to evaluate predictability of $1(\theta \in C(x))$, you are forced to grapple explicitly with priors, power, and misspecification, but in *simulation*, not necessarily *modeling*.

I think this is very exciting.