Ordering Rules as Criteria of Inductive Inference

Luc Demortier

April 2, 2004

- Reporting frequentist intervals;
- The meaning of coverage;
- Inductive inference;
- Ordering rules as objective statements about nature;
- Properties of ordering rules;
- Conclusions and recommendations.

Reporting frequentist intervals

Suppose we measure a physics parameter μ and report a frequentist interval $[\mu_1, \mu_2]$ with confidence level γ. If asked to interpret this result, we usually give the following explanation:

> If we repeat the experiment a large number of times, the fraction of those experiments in which the reported interval will actually contain the true value of μ is γ.

Alternatively, we could simply state that:

> The probability that the reported interval contains the true value of μ is γ.

This is correct provided the term "probability" is interpreted in the sense of repeated experiments.

The \$60M question is then: What have we really learned about μ from this?

The meaning of coverage

Two examples will help focus on the issue of interest:

Example 1

Consider the following interval construction procedure: Each time we make a measurement of μ, we also toss an unbiased coin. If the coin falls heads up, we report the whole physical range of μ as interval. If the coin falls tails up, we report an empty interval, such as $[\mu_1, \mu_1]$ for some physical value μ_1. This procedure has 50% coverage, but it doesn't tell us anything useful about μ.

Example 2

Imagine a situation where we estimate the value of μ from two measurements x and y that are independently Gaussian distributed about μ with unit variance. Say we observe $x = 10.00$ and $y = 14.05$. We then calculate two confidence intervals:

1. A 99% C.L. interval using probability density ordering; this yields $[11.32, 12.73]$.
2. A 68% C.L. interval using likelihood ratio ordering; this yields $[11.32, 12.73]$.

Note that we obtained the exact same interval in both cases! Two intervals that are numerically identical have vastly different probabilities for containing the true value of μ, depending on the ordering rule used. Once again, it appears that coverage alone does not teach us anything useful about μ.

Inductive inference

If coverage is useless, how then can we learn something about the true value of μ? How is it for example, that our current knowledge about Planck's constant, $h = 6.62606876(52) \times 10^{-34}$ Js, has so much predictive power? The answer lies in the ordering rule used.

Ordering rules are objective statements that specify what we mean when we say that a given value of μ is "plausible" for the data we observed. They constitute the *inductive* part of statistical inference, telling us how to go from the actual data we observed (the particular) to a statement about a parameter of nature (the general).

In contrast, coverage, accuracy, and bias are ensemble-averaged properties and are therefore purely deductive. They do not allow one to make a meaningful statement about nature from an actual measurement.

Let's look at some examples of ordering rules.

Ordering rules as objective statements about nature

1. A 68% central ordering rule

$[\mu_1, \mu_2]$ is the set of μ values for which the observed data falls between the 16^{th} and 84^{th} percentiles of its distribution.

2. A 68% probability density ordering rule

$[\mu_1, \mu_2]$ is the set of μ values for which the observed data falls within the 68% most probable region of its distribution.

3. A 68% likelihood ratio ordering rule

$[\mu_1, \mu_2]$ is the set of μ values for which the logarithm of the likelihood ratio, evaluated at the observed data, is within half a unit of its maximum.

4. A 68% Feldman-Cousins ordering rule

$[\mu_1, \mu_2]$ is the set of μ values for which the likelihood ratio, evaluated at the observed data, is larger than R_1 to the left of the maximum likelihood estimate $\hat{\mu}$, and larger than R_2 to the right of $\hat{\mu}$. Here, R_1 and R_2 are chosen to achieve 68% coverage.

Properties of ordering rules

- Ordering rules are not probability statements, but "plausibility" statements, about the parameter of interest. In this sense, ordering rules have nothing to do with frequentism.
- Some ordering rules have coverage, but not all; for example, Feldman-Cousins does, but likelihood ratio does not.
- Some ordering rules yield empty intervals, for example the probability density ordering (PDO) rule. Empty intervals do provide useful information however. An empty 68% PDO interval for example, means that for no value of μ is the observed data within the 68% most probable region of its distribution. This could indicate that the observed data is somewhat exceptional, or that the model of the physics process under study is inadequate, or that the ordering rule is inadequate.
- If empty intervals are a problem, it is always legitimate to quote an interval with higher coverage, or several such intervals, or the p value function (see Fraser and Reid).
- Ordering rules should be selected on the basis of properties such as metric independence, etc.

Conclusions and Recommendations

1. Coverage has no useful meaning by itself, but once an ordering rule has been specified, and if that ordering rule allows it, then coverage provides a convenient label that picks out a unique interval among an infinity of intervals that are valid for the observed data.

2. Ordering rules provide objective statements for inductive inference.

3. When reporting a confidence interval, always specify the ordering rule used.

4. There are situations where coverage is more than a convenient label. One example is a repeated testing situation, where coverage is essentially equivalent to the efficiency of a selection procedure.