Philosophical issues with inferential paradigms

Choose your poison

Subhash R. Lele

Department of Mathematical Sciences
University of Alberta
Canada
Email: slele@ualberta.ca

July 11, 2016

Which experiment do we repeat? (Conditionality principle)

Consider a simple experiment to determine the weight of an item such that it consists of two stages:

Which experiment do we repeat? (Conditionality principle)

Consider a simple experiment to determine the weight of an item such that it consists of two stages:

1. Toss a fair coin and note the outcome whether 'head' or 'tail'.

Which experiment do we repeat? (Conditionality principle)

Consider a simple experiment to determine the weight of an item such that it consists of two stages:

1. Toss a fair coin and note the outcome whether 'head' or 'tail'.
2. If the outcome is 'head', then use a very accurate scale, i.e. $Y \sim N(\mu, 0.01)$

Which experiment do we repeat? (Conditionality principle)

Consider a simple experiment to determine the weight of an item such that it consists of two stages:

1. Toss a fair coin and note the outcome whether 'head' or 'tail'.
2. If the outcome is 'head', then use a very accurate scale, i.e. $Y \sim N(\mu, 0.01)$
3. If the outcome is 'tail', then use a very inaccurate scale, i.e. $Y \sim N(\mu, 10)$

Which experiment do we repeat? (Conditionality principle)

Consider a simple experiment to determine the weight of an item such that it consists of two stages:

1. Toss a fair coin and note the outcome whether 'head' or 'tail'.
2. If the outcome is 'head', then use a very accurate scale, i.e. $Y \sim N(\mu, 0.01)$
3. If the outcome is 'tail', then use a very inaccurate scale, i.e. $Y \sim N(\mu, 10)$

The estimate of μ is the observed data y. What is the uncertainty associated with this number?

Unconditional and Conditional confidence intervals

We actually have two pieces of information: The observation y and *outcome of the coin*: head or tail. In my simulation, I got $w = 1$ and $y = 3.9712$

Conditional confidence interval

- Suppose we know that the outcome of the coin toss was 'head'. Then we know that the scale that we used was very accurate and hence $(y - 1.96 * 0.1, y + 1.96 * 0.1) = (3.78, 4.17)$ is the sensible statement to make.

Unconditional and Conditional confidence intervals

We actually have two pieces of information: The observation y and *outcome of the coin*: head or tail. In my simulation, I got $w = 1$ and $y = 3.9712$

Conditional confidence interval

- Suppose we know that the outcome of the coin toss was 'head'. Then we know that the scale that we used was very accurate and hence $(y - 1.96 * 0.1, y + 1.96 * 0.1) = (3.78, 4.17)$ is the sensible statement to make.
- Suppose we know that the outcome of the coin toss was 'tail'. Then we know that the scale that we used was very inaccurate and hence $y - 1.96 * 3.16, y + 1.96 * 3.16$ is the sensible statement to make.

Unconditional and Conditional confidence intervals

We actually have two pieces of information: The observation y and *outcome of the coin*: head or tail. In my simulation, I got $w = 1$ and $y = 3.9712$

Conditional confidence interval

- Suppose we know that the outcome of the coin toss was 'head'. Then we know that the scale that we used was very accurate and hence $(y - 1.96 * 0.1, y + 1.96 * 0.1) = (3.78, 4.17)$ is the sensible statement to make.
- Suppose we know that the outcome of the coin toss was 'tail'. Then we know that the scale that we used was very inaccurate and hence $y - 1.96 * 3.16, y + 1.96 * 3.16$ is the sensible statement to make.

If we are lucky, we should accept our luck and give a short interval; if we are unlucky, we should give a long interval.

Unconditional confidence interval If we follow the strict definition of confidence interval, we should actually compute these confidence intervals under the repetition of the full experiment: First the coin toss and then the measurement. If we do that, the confidence interval we get is approximately: $(-1.36, 9.25)$.

Unconditional confidence interval If we follow the strict definition of confidence interval, we should actually compute these confidence intervals under the repetition of the full experiment: First the coin toss and then the measurement. If we do that, the confidence interval we get is approximately: $(-1.36, 9.25)$.

- Does this interval make sense to you?

Unconditional confidence interval If we follow the strict definition of confidence interval, we should actually compute these confidence intervals under the repetition of the full experiment: First the coin toss and then the measurement. If we do that, the confidence interval we get is approximately: $(-1.36, 9.25)$.

- Does this interval make sense to you?
- To me, the conditional intervals make sense but if we follow the frequentist approach as per the definition, the unconditional confidence interval is the correct one (in fact, it is **optimal!**)

- This phenomenon is not limited to artificial examples.

- This phenomenon is not limited to artificial examples.
- In linear regression, the variance of the estimator of the regression coefficient depends on the configuration of the covariates. If the covariates are spread out, we get a good estimate of the slope; otherwise we get bad estimates.

- This phenomenon is not limited to artificial examples.
- In linear regression, the variance of the estimator of the regression coefficient depends on the configuration of the covariates. If the covariates are spread out, we get a good estimate of the slope; otherwise we get bad estimates.
- The standard calculations reflect this but are they 'kosher' according to the strict definition of the confidence interval?

- This phenomenon is not limited to artificial examples.
- In linear regression, the variance of the estimator of the regression coefficient depends on the configuration of the covariates. If the covariates are spread out, we get a good estimate of the slope; otherwise we get bad estimates.
- The standard calculations reflect this but are they 'kosher' according to the strict definition of the confidence interval?
- They are 'kosher' only if we condition on the covariate values. That is, we only repeat the experiment under the same covariate values.

- This phenomenon is not limited to artificial examples.
- In linear regression, the variance of the estimator of the regression coefficient depends on the configuration of the covariates. If the covariates are spread out, we get a good estimate of the slope; otherwise we get bad estimates.
- The standard calculations reflect this but are they 'kosher' according to the strict definition of the confidence interval?
- They are 'kosher' only if we condition on the covariate values. That is, we only repeat the experiment under the same covariate values.
- If the covariate values themselves are random (for example, in observational studies), we should be computing the unconditional confidence intervals, integrating over the variation in the covariate values too.

- This phenomenon is not limited to artificial examples.
- In linear regression, the variance of the estimator of the regression coefficient depends on the configuration of the covariates. If the covariates are spread out, we get a good estimate of the slope; otherwise we get bad estimates.
- The standard calculations reflect this but are they 'kosher' according to the strict definition of the confidence interval?
- They are 'kosher' only if we condition on the covariate values. That is, we only repeat the experiment under the same covariate values.
- If the covariate values themselves are random (for example, in observational studies), we should be computing the unconditional confidence intervals, integrating over the variation in the covariate values too.
- Scientifically the unconditional confidence interval does not make sense and hence we do not use it.

- The value of the coin toss gives us no information about the parameter of interest but tells us about the uncertainty.

- The value of the coin toss gives us no information about the parameter of interest but tells us about the uncertainty.
- The values of the covariates give us no information about the slope but tell us about the uncertainty associated with the parameter estimate.

- The value of the coin toss gives us no information about the parameter of interest but tells us about the uncertainty.
- The values of the covariates give us no information about the slope but tell us about the uncertainty associated with the parameter estimate.
- Such quantities are called 'ancillary statistics'.

- The value of the coin toss gives us no information about the parameter of interest but tells us about the uncertainty.
- The values of the covariates give us no information about the slope but tell us about the uncertainty associated with the parameter estimate.
- Such quantities are called 'ancillary statistics'.
- Conditionality principle says that we should use inference conditional on such ancillary statistics or 'relevant subsets' of the sample space defined by them.

- The value of the coin toss gives us no information about the parameter of interest but tells us about the uncertainty.
- The values of the covariates give us no information about the slope but tell us about the uncertainty associated with the parameter estimate.
- Such quantities are called 'ancillary statistics'.
- Conditionality principle says that we should use inference conditional on such ancillary statistics or 'relevant subsets' of the sample space defined by them.
- These are non-unique. That is a BIG problem with the frequentist approach.

The researcher has to decide which experiment is supposed to be repeated.

Other related problems

- Multiple comparisons
  - Two anthropological surveys (same individuals in the survey): One collects information only about the character of interest (say, Height) and the other collects information on all the different characters (say, Height, Weight, Age ...)
  - *Should the confidence interval for Height be different for two surveys?*
- Nuisance parameters
  - *How do we conduct inference about one of the parameters in the presence of other parameters?*

- The main problem with frequentist measure of uncertainty is that it is a **pre-data** measure whereas we would like to have a **post-data** measure of uncertainty.

- The main problem with frequentist measure of uncertainty is that it is a **pre-data** measure whereas we would like to have a **post-data** measure of uncertainty.
- This is similar to saying that the GPS unit that I use to locate an animal is highly accurate. It tells us nothing about the accuracy of the individual measurement that we have. In principle, it could be quite far apart (indeed with small probability).

- The main problem with frequentist measure of uncertainty is that it is a **pre-data** measure whereas we would like to have a **post-data** measure of uncertainty.
- This is similar to saying that the GPS unit that I use to locate an animal is highly accurate. It tells us nothing about the accuracy of the individual measurement that we have. In principle, it could be quite far apart (indeed with small probability).
- Is it possible to give post-data uncertainty of the inferential statement?

Issues with the Bayesian approach

- Specification of the prior distribution is the tricky part, especially when we multi-parameter situation with possible correlations between parameters.

Issues with the Bayesian approach

- Specification of the prior distribution is the tricky part, especially when we multi-parameter situation with possible correlations between parameters.
- Eliciting priors from the experts is extremely difficult.

Issues with the Bayesian approach

- Specification of the prior distribution is the tricky part, especially when we multi-parameter situation with possible correlations between parameters.
- Eliciting priors from the experts is extremely difficult.
- We do not even know which parameters are estimable and which ones are not estimable.

Issues with the Bayesian approach

- Specification of the prior distribution is the tricky part, especially when we multi-parameter situation with possible correlations between parameters.
- Eliciting priors from the experts is extremely difficult.
- We do not even know which parameters are estimable and which ones are not estimable.
- Prior distribution on non-identifiable parameters is impossible to elicit.

Issues with the Bayesian approach

- Specification of the prior distribution is the tricky part, especially when we multi-parameter situation with possible correlations between parameters.
- Eliciting priors from the experts is extremely difficult.
- We do not even know which parameters are estimable and which ones are not estimable.
- Prior distribution on non-identifiable parameters is impossible to elicit.
- Complete class theorem says that there **exists** a prior that will lead to the best decision. But it neither tells us how to construct such a prior nor does it indicate how far our chosen prior is from such an ideal prior distribution. This is not very useful in practice.

Do we need frequentist probability to evaluate Bayesians?

Do we need frequentist probability to evaluate Bayesians?

- What does credible interval mean? How do we validate its correctness? The credible interval is to be interpreted as: "I believe the parameter belongs to ...". I have seldom seen a Bayesian paper saying 'I believe ...' when interpreting the credible intervals or Bayesian hypothesis testing.

Do we need frequentist probability to evaluate Bayesians?

- What does credible interval mean? How do we validate its correctness? The credible interval is to be interpreted as: "I believe the parameter belongs to ...". I have seldom seen a Bayesian paper saying 'I believe ...' when interpreting the credible intervals or Bayesian hypothesis testing.
- What does prediction interval mean? How do we validate its correctness? Are we thinking of replicating the experiment and checking how many times the interval contains the actual outcome?

Do we need frequentist probability to evaluate Bayesians?

- What does credible interval mean? How do we validate its correctness? The credible interval is to be interpreted as: "I believe the parameter belongs to ...". I have seldom seen a Bayesian paper saying 'I believe ...' when interpreting the credible intervals or Bayesian hypothesis testing.
- What does prediction interval mean? How do we validate its correctness? Are we thinking of replicating the experiment and checking how many times the interval contains the actual outcome?
- When we say we are willing to bet in a certain way, do we implicitly mean that we believe that we will win a certain proportion of times? Is this a back door entry to a frequentist thinking?

Stanford Encyclopedia of Philosophy

> *Ascertainability This criterion requires that there be some method by which, in principle at least, we can ascertain values of probabilities. It merely expresses the fact that a concept of probability will be useless if it is impossible in principle to find out what the probabilities are...*

Is the Bayesian probability ascertainable without some aid from frequentist notion?

- **Model diagnostics** is inherently a frequentist concept. In the Bayesian literature, distance of the data from the predictive distribution is used to study model adequacy. This can tell us something is wrong with the model but what it cannot tell us is whether it is the prior or the likelihood that is wrong.

- **Model diagnostics** is inherently a frequentist concept. In the Bayesian literature, distance of the data from the predictive distribution is used to study model adequacy. This can tell us something is wrong with the model but what it cannot tell us is whether it is the prior or the likelihood that is wrong.
- **Well calibrated Bayesian?** Some Bayesian philosophers want to have a procedure that is Bayesian in principle but also valid in the frequentist sense. They try to choose a prior that satisfies this condition. The question of which experiment should I repeat rears is ugly head again.

Non informative or objective Bayesian approach

Can we choose a prior that will *'let the data speak'*?

- There is no proper definition of a non-informative prior. Flat priors or priors with large variance are not non-informative.

Non informative or objective Bayesian approach

Can we choose a prior that will *'let the data speak'*?

- There is no proper definition of a non-informative prior. Flat priors or priors with large variance are not non-informative.
- In practice, only independent component priors are chosen although clearly the parameters are bound to be related. This is just for the sake of convenience and does not reflect actual expert opinion.

Non informative or objective Bayesian approach

Can we choose a prior that will *'let the data speak'*?

- There is no proper definition of a non-informative prior. Flat priors or priors with large variance are not non-informative.
- In practice, only independent component priors are chosen although clearly the parameters are bound to be related. This is just for the sake of convenience and does not reflect actual expert opinion.
- In the past, priors were chosen for mathematical convenience. Now they are chosen for computational convenience and convergence of the MCMC. They seldom reflect expert opinion.

Non informative or objective Bayesian approach

Can we choose a prior that will *'let the data speak'*?

- There is no proper definition of a non-informative prior. Flat priors or priors with large variance are not non-informative.
- In practice, only independent component priors are chosen although clearly the parameters are bound to be related. This is just for the sake of convenience and does not reflect actual expert opinion.
- In the past, priors were chosen for mathematical convenience. Now they are chosen for computational convenience and convergence of the MCMC. They seldom reflect expert opinion.
- The results from the non-informative priors are NOT similar to the frequentist results. This is a complete myth and a number of examples (and, theoretical explanations) exist showing the falsity of this statement.

- There are frequentist or likelihood based methods, e.g. meta-analysis, that can be used to combine past data or related data or expert opinion with the data at hand. One does not have to follow the Bayesian approach to achieve this goal. This is another myth spread by the Bayesian practitioners.

- There are frequentist or likelihood based methods, e.g. meta-analysis, that can be used to combine past data or related data or expert opinion with the data at hand. One does not have to follow the Bayesian approach to achieve this goal. This is another myth spread by the Bayesian practitioners.
- The non-informative or objective Bayesian analysis is nothing but BINO (Bayesian In Name Only).

- Does it even make sense to talk about uncertainty about events/experiments that are not replicable?
- When we say equivalent experiments, do we mean similar but not necessarily identical? Is it ever possible to repeat an experiment *identically*? Can we even toss a coin exactly the same way? What do we mean by repeating an experiment?
- Example for combining past data with future data using the likelihoods vs prior-posterior paradigm (which one is better?)
- Example of combining 'related' data but not the exact same experiment using hierarchical modelling and likelihood (similar to expert opinion)
- Example of checking if the prior is 'good' by generating data from the predictive distribution and using that to see if it would have added any information.
- Why should Bayesians be worried about sensitivity to the priors? Are they bolstering weak data with priors?

SUMMARY

- If you want to be a frequentist, you have to decide on which experiment you are (hypothetically) repeating.
- If you are a proper Bayesian, you have to choose a prior. You are answering the question: What is my belief having observed these data? Is this what we want in science?
- An objective Bayesian is neither a Bayesian nor a frequentist. It suffers from the shortcomings of both approaches. Priors do not reflect expert opinion; posteriors have no frequentist meaning.
- Sophie's choice: Bayesian myths or frequentist follies?

(An impossibly difficult choice, especially when forced onto someone. The choice is between two unbearable options, and it is essentially a no-win situation)