INTERPRETING STATISTICAL CONFIDENCE

Milo Schield, Augsburg College
Dept. of Business & MIS. 2211 Riverside Drive. Minneapolis, MN 55454

Keywords: Bayesian Inference, Confidence Interval, Probability, Teaching, Epistemology, Philosophy of Science,

Abstract: The traditional interpretation of statistical confidence is mathematically descriptive. As such it provides little guidance for decision making. Three alternate interpretations of confidence are presented: the Bayesian, the pragmatic and one proposed by the author labeled herein as the neoclassical. Each alternate explanation argues that in some way a confidence interval and a numerically equivalent probability should be treated similarly. The neoclassical interpretation argues that one should be indifferent between two bets: betting on a 95% confidence interval to contain the population parameter and betting on a 95% chance that the next ball will be red in drawing from an urn containing 95% red balls. Unlike the Bayesian interpretation, the neoclassical interpretation treats the population parameter as a constant – not as a variable.

The neoclassical interpretation of confidence holds that confidence measures one's strength of belief in a particular claim and thus is psychological and subjective. Yet in some cases, one's confidence *should* be the same in two apparently different situations. This neoclassical interpretation focuses on the 'calculated risk' or objective aspect of confidence and includes aspects from both the Bayesian and traditional viewpoints. Specific recommendations are offered.

1. INTRODUCTION

Teachers and authors spend considerable energy informing students that "being 95% confident that this particular interval includes the population mean" is not the same as saying "there is a 95% chance that this particular interval includes the population mean." Yet many students still act as though being 95% confident had some relation to having a 95% chance.

2. UNCERTAINTY ABOUT CONFIDENCE

Educators are generally uncertain about the operational meaning of confidence. This conclusion is based on the author's experience at the 1997 JSM where thirty attendees were given the following situation.

Situation #1: Suppose we are presented with an urn containing 20 balls: 19 red and 1 blue. And suppose we have randomly sampled values from a normally-distributed population variable whose parameters are known to someone. Suppose we have obtained the particular sample statistics and have generated a particular 95% confidence interval.

Q1. Is betting that 'this particular confidence interval contains the population parameter' operationally equivalent to betting that 'the next ball will be red when randomly drawn from this urn'?
Q2. How would students answer this question (Q1)?
Q3. Do you recall any book that specifically addressed this question (Q1)?

For question Q1, one attendee answered 'Yes', 9 answered 'No' and 10 answered 'Uncertain'. Another 10 did not answer. The respondents indicated that question Q1 was appropriate and deserved an answer. For question, Q2, most attendees indicated 'Yes' along with general laughter (indicating this question was somewhat rhetorical). For question Q3, no one knew of any book in the past 40 years that addressed this particular question.

The same situation was posted on an Australian list (StatEd) and a UK list (Teaching Statistics). A higher percentage of respondents answered "Yes" to the first question, but most of these indicated they were reasoning from a Bayesian perspective. While some cited Bayesian references to this kind of comparison, no one cited a frequentist analysis. When this situation was circulated among professional statisticians in industry most answered 'Yes' to the first question.

Some respondents found this situation difficult to deal with. Some thought we were comparing a confidence interval with a point estimate. Others though we were comparing an inductive activity (confidence interval) with a deductive activity (probability). Still others thought we were comparing an action prior to sampling with an action obtained after sampling. In each case, it seemed as if we were comparing apples and oranges

To better sort out the similarities and differences between a confidence interval and a random draw from an urn, consider the following situation.

Situation #2: Consider the same situation as in #1 but prior to sampling. Is betting that 'the next 95% confidence interval will include the population parameter' operationally equivalent to betting that 'the next random selection from this urn will be a red'?

From a betting perspective, frequentists agree that that these two alternatives are operationally equivalent. Both are prior to sampling and both have a 95% chance of success. From a betting perspective, their differences are irrelevant.

Having better identified the similarity between the confidence interval and the urn, consider four different approaches. The traditional approach is silent or undeclared on the first question (Q1), whereas the Bayesian, the pragmatic, and the author's neoclassical interpretation all agree that the two situations should be equivalent from a betting perspective.

3. TRADITIONAL INTERPRETATION

The traditional interpretation is either silent or undeclared on the relation between confidence and probability in relation to betting. Traditionally, 'confidence' is a technical term – a kind of virtual or hypothetical probability. Statistical confidence is identified by its history. A particular 95% confidence interval is part of a family where 95% of the members contain the population parameter. Thus, prior to sampling one had a 95% chance of selecting a member of this family that contained the population parameter.

Once sampling occurs, the population parameter is either included in a particular 95% confidence interval or else it is not. The probability that a particular 95% confidence interval contains the population parameter is either zero or one (assuming that the concept of probability is meaningful when applied to a fact). Thus, for any particular confidence interval, the level of confidence never equals the probability that that particular interval contains the parameter of interest. After sampling, confidence is not the same as probability.

The traditional presentation involves three claims: the first says what confidence is not, the second says what it's about and the third says how it's numerically determined.

1. What it is not: After sampling, 'confidence' is not the probability that the population parameter (which is a constant) is in the confidence interval based on the sample statistic (which is also a constant).

2. What it's about: 'Confidence' is confidence in a process. "The statement that 'we are 95% **confident** that the unknown μ lies between 452 and 480' is shorthand for saying, 'We arrived at these numbers by a method that gives correct results 95% of the time.' " (Moore and McCabe, p. 111)

3. How it's determined: 'Confidence' is numerically determined by an objective probability: the probability that one could select a confidence interval that contained the population parameter prior to sampling.

Classical frequentists are ambivalent about whether statistical confidence has any relation to psychological confidence. Some indicate 'No' ('confidence' is an unfortunate choice of name) while others indicate 'Yes' (the higher the confidence the more one is justified in believing the claim is true.) Each answer leads to further questions. If the relationship is denied, then is 'confidence' misleading ('prior-probability' would have been better)? Does statistical confidence have any psychological import or any relevance to choice or action? If the relationship is asserted, then what is the nature, extent and justification of this relationship? If statistical confidence is objective and psychological confidence is subjective, then how are they related?

Even when frequentists try to be very specific, students find ambiguity in their statements about probability and confidence. Algebraically, $P(|\bar{x} - \mu| < 2\ SE)$ means $P(|\tilde{\bar{x}} - \mu_o| < 2\ SE)$ rather than $P(|\bar{x}_o - \tilde{m}| < 2\ SE)$. Here the tilde (~) indicates a random variable and the subscript ($_o$) indicates a constant. Verbally, when we say 'this interval' we can mean either 'this particular interval' (after sampling) or 'this kind of interval' (prior to sampling). But without these auxiliary clues, students are unsure of what is being said.

4. BAYESIAN INTERPRETATION

The Bayesian interpretation defines 'probability' as one's strength of belief in the truth of a statement whose truth-value is unknown. Bayesians can speak of the *probability* the dinosaurs were destroyed by the effects of a meteor whereas classical frequentists would not. In estimating a quantitative parameter, a subjective prior probability distribution is informed by sampled data using Bayes rule to yield a posterior distribution. In so doing, "the Bayesian approach regards θ [the population parameter] as random *in the sense that we have certain beliefs about its value,* and think of the interval as fixed once the datum is available." (Lee, p. 51).

Some Bayesians consider ‘confidence’ as being a classical dead-end and find it to be a sign of poor thinking. At best, ‘confidence’ is a way that classical frequentists try to achieve Bayesian results without agreeing to the Bayesian premises. Other Bayesians might agree that “the uncritical use of confidence interval estimates may imply unreasonable assumptions about the investigator’s prior knowledge concerning the parameter being estimated.” (Hamburg, 1987, p. 699)

Classical frequentists make several objections. They do not share the Bayesian premises about the nature of probability and thus consider the Bayesian argument as being valid but unsound. They feel that Bayesian argument may allow epistemic uncertainty about the truth of a claim (the measurement of the speed of light is a random variable) to justify asserting metaphysical uncertainty (the speed of light is a random variable). Finally, they worry that including any subjectivity undermines the objectivity one wants in science.

5. PRAGMATIC INTERPRETATION

Pragmatists treat probability and confidence as indistinguishable but without adequate justification. Given a particular confidence interval after sampling, pragmatists say that confidence equals probability. Specifically, ‘being 95% confident that this confidence interval contains the population parameter’ is the same as saying ‘a 95% chance that this confidence interval contains the population parameter.’ The following are Pragmatist arguments along with traditional responses.

Necessity. If the idea of a confidence interval is to be of any value, then it must provide a basis for action as does probability, so we treat a particular 95% confidence interval *as if there were* a ‘95% chance that it contains the population parameter’. Traditional response: “Necessity” may explain why one acts; it does not automatically justify one’s action.

Symmetry. If $\bar{x}$ is within 2 standard errors of μ 95% of the time, then the converse must be true. Traditional response: Symmetry works so long as one ignores the distinction between a random variable ($\tilde{\bar{x}}$) and a constant (μ_o). This symmetry doesn’t justify treating a constant (μ_o) as a random variable ($\tilde{m}$).

Laxity. Although frequentists say the probability of a fixed outcome is either zero or one, some are lax in talking about the ‘probability’ of a prize being behind a door (even though the prize is already in place as in the Monty Hall three-door problem). Although Bayesians often say that probability is subjective, some grant the ‘probability’ of flipping a fair coin heads is 50% (assuming no other knowledge). Given this laxity by others, pragmatists see no reasons to ‘split hairs’ distinguishing confidence from probability. Traditional response: Laxity by others does not justify laxity by oneself.

Identity. If 95% of these intervals contain the parameter, there must be a 95% chance that any particular interval contains the parameter. Traditional response: Identity prior to sampling does not guarantee identity between confidence and probability after sampling.

Utility. Students do not need to understand this subtle distinction between probability and confidence. It confuses rather than enlightens. Traditional response: Utility motivates but does not always justify.

Neo-classical response: Even if Pragmatists are wrong in saying “confidence and probability are indistinguishable”, it may be that confidence is practically the same as probability. But the nature and extent of this similarity needs clarification and an adequate justification.

6. NEOCLASSICAL INTERPRETATION

The neoclassical interpretation contains elements of the traditional, pragmatic and Bayesian approaches. It agrees with the pragmatists and the Bayesians that confidence should be related to probability. It agrees with the Bayesians that confidence is psychological – a measure of one’s strength of belief in the truth of a claim. It agrees with the traditionalists that fixed parameters should not be treated as random variables and that statistical confidence should be objective.

The neoclassical approach differs from the traditional, pragmatic and Bayesian approaches. Whereas the traditional approach is silent on betting on a single confidence interval, the neoclassical approach asserts that one should bet *as though* the chance of winning with a 95% confidence interval equals the chance of winning with a 95% probability of drawing a red ball from an urn. It differs from the pragmatic by offering a systematic justification. It disagrees with the Bayesians by not using a prior probability, by not using Bayes Rule and by not speaking of the probability of a constant being in a particular interval.

Probability by itself does not lead to action. Rather probability justifies confidence and confidence justifies action. Thus we use probability to calibrate confidence. This approach involves three steps as shown in Table I. Table II illustrates the same process in drawing a ball from an urn. The goal is to show that confidence after the fact is operationally equivalent to a numerically-equivalent probability before the fact.

Table I. Claims about confidence intervals from a normal population where Std. Error = $\sigma/\sqrt{n}$.

Confidence Intervals	*Context of uncertainty (what)*		
Description of uncertainty	**Sample has not been drawn.** μ is unknown $\tilde{\tilde{x}}$ is variable and unknown	**Sample has been drawn. but statistics are unknown** μ is unknown $\bar{x}_?$ is constant but unknown	**Sample has been drawn. and statistics are known** μ is unknown $\bar{x}_o$ is constant and known
Objective Frequentist or a-priori	**Classical/traditional** $P[(\mu_o - 2\ SE) \le \tilde{\tilde{x}} \le (\mu_o + 2\ SE)] = .954$	$P[(\mu_o - 2SE) \le \bar{x}_? \le (\mu_o + 2SE)] = 0$ or 1	$P[(\mu_o - 2SE) \le \bar{x}_o \le (\mu_o + 2SE)] = 0$ or 1
Confidence Strength of belief that $\lvert \bar{x} - \mu_o \rvert \le$ 2 SE	ß ① **95% confidence 95% confident** $\lvert \tilde{\tilde{x}} - \mu_o \rvert \le$ 2 Std.Errors	Þ **95% confidence 95% confident** $\lvert \bar{x}_? - \mu_o \rvert \le$ 2 Std.Error	Þ **"Neoclassical"** **95% confidence 95% confident** $\lvert \bar{x}_o - \mu_o \rvert \le$ 2 Std.Error

Table II. Claims about drawing a red ball from an urn containing 19 red balls and one blue ball.

Color of ball	*Context of uncertainty (what)*		
Description of uncertainty	**Ball has not been drawn**	**Ball has been drawn but color is unknown**	**Ball has been drawn. Color is known (not red)**
Objective Frequentist or a-priori	**Classical/traditional** P(red ball) =95%	P(ball is red) = 0 or 1	P(ball is red)= 0
Confidence Strength of belief that ball is red	ß ① **95% confidence 95% confident**	Þ **95% confidence 95% confident**	Þ **0% confidence 0% confident**

The first row indicates the status of probability from a classical perspective. The neo-classical approach preserves this row but moves to the second row by identifying the intimate relationship between confidence and probability both taken prior to sampling.

7. NEOCLASSICAL JUSTIFICATION

Step 1: From probability to confidence prior to sampling. This step uses the Principal Principle (Howson and Urbach, 1993, Page 240). "The principle states that if the objective, physical probability of a random event (in the sense of its limiting relative frequency in an infinite sequence of trials) were known to be r, and if no other relevant information were available, then the appropriate subjective degree of belief that the event will occur on any particular trial would also be r." [Underscore added] Notice that the Principal Principle is normative. The words 'appropriate' and 'would' entail the normative 'should'. In this situation, this principle says that prior to sampling, one *should* be 95% confident that the population parameter will be included in the next random 95% confidence interval. For a fair coin, one should be 50% confident the next flip will be heads.

Confidence in the process is time-independent; it applies prior to taking a sample as well as subsequent to sampling. And one should be willing to bet on that.

How does this principle accommodate the fact that confidence is basically personal and subjective? How can subjective confidence be determined objectively? Since we can't measure confidence directly, we can only identify a particular gamble and let each individual determine their personal confidence in relation to that objective uncertainty. In that sense, we calibrate confidence objectively.

Step 2: From confidence prior to sampling to confidence after the sample is obtained yet while the sample statistic and associated confidence interval are not yet known. In this step, one's level of confidence *should* remain unchanged since one has no new knowledge about the sample itself. Of course the probability this particular interval includes the population parameter is either zero or one (but we don't know which).

Step 3: After sampling, from confidence without particular knowledge to confidence with particular knowledge. In this step, we now discover the sample statistics and the particular values of the associated confidence interval. If we knew the population parameter, we could now discern whether the probability is zero or one in this particular case. But the population parameter is still unknown and knowing the sample mean gives us no new knowledge about the relation of this particular sample-based confidence interval and the population parameter.

So long as one has no outside information by which to relate this particular sample to the population, one's level of confidence about the distribution of distances between sample statistic and population parameter *should* remain unchanged.

8. NEOCLASSICAL SUMMARY

The neoclassical approach holds that from a betting perspective one should treat a 95% confidence interval the same way one would treat a 95% chance in drawing from an urn since one should have 95% confidence in each of them.

By locating the uncertainty in one's strength of belief rather than in the next sample or in the value of the population parameter, the neoclassical interpretation treats both the confidence interval and the value of the population parameter as being constants.

In this neoclassical approach, probability retains its classical frequentist interpretation. The source of uncertainty is located within one's lack of knowledge rather than within a physical property of nature.

To better understand this neoclassical relation between probability and confidence, consider drawing a ball from an urn containing 20 balls: 19 red and 1 blue. The associated probabilities of the ball being red are shown in Table II for different circumstances. Note that in this case, unlike the confidence interval, seeing the color of the ball forces the confidence to equal the probability. Step 1 and 2 in flipping a fair coin are similar to steps 1 and 2 in drawing a colored ball from an urn. In both cases, *confidence after the random event is the same as probability before the random event – provided one is ignorant of the outcome.*

Notice that the confidence in step 1 of the urn (95%) equals the confidence in step 3 of the confidence interval (95%). This is the justification needed to support treating a 95% confidence interval as one would treat a 95% chance of success in a bet. While students may go too far by saying there is a 95% chance that a given interval contains the parameter, they are still close to the mark in saying one should treat a 95% confidence interval as one would treat a 95% chance of success in a bet.

9. CONCERNS

The following concerns have been raised by those who have heard this presentation.

Even if this conclusion is true in some ideal world, it seldomly holds in the real world. It is much harder to obtain a random sample from a continuous population than to obtain a random sample from an urn. So in practice, this prescription may not be justified. Response: this applies to all statistical inference.

This use of 'prescriptive' is easily misunderstood. Response: Normally 'prescribe' specifies an action; here it specifies a level of 'confidence'. This statistical prescription does not say that one should bet twice as much with 60% confidence as with 30% confidence. It 'prescribes' only in a weak sense of equivalent actions. Regardless of how one would bet on the next draw from an urn containing 95% successes, one should bet the same way on a particular 95% confidence interval.

If confidence and probability are similar for betting, students may conclude they are identical. Response: If students jump from similarity to identity, then one must decide which is more important: understanding the similarity or understanding the difference (the lack of identity). The neoclassical approach argues similarity is more important than difference from a decision-making perspective.

We don't need this elaborate proof since authors already say the same thing: "confidence is confidence in the process – not in the outcome." Response: But authors don't say that one should bet on the process after sampling in the same way one would bet prior to sampling. If students are to understand this statement, we must make these implications explicit.

10. OBJECTION

Howson and Urbach (1993, p. 240) assert that the use of the Principal Principle is unjustified whenever one is dealing with specific numerical values of a statistic (after the sample is obtained). They argue as follows:

"For example, the physical probability of getting a number of heads greater than 5 in 20 throws of a fair coin is 0.86…. According to the Principal Principle … 0.86 is also the confidence you should place in any particular trial of 20 throws of a fair coin producing a number of heads greater than 5. Suppose a trial is made and 2 heads are found in a series of 20 throws with a coin that is known to be fair. To infer that we should now be 86% confident that 2 is greater than 5 would be absurd and a misapplication of the Principal Principle…. Mistaking this (applying the Principal Principle to fixed numbers) appears to be the fallacy implicit in the subjective-confidence interpretation of confidence intervals…. The principle … does not license the substitution of numbers for the (variables), so the desired inference from experimentally measured intervals to subjective confidences is blocked."

While this particular example supports their conclusion, the argument may lack generality. The equality of probability and confidence fails in this case because the outcome gives new knowledge that 'trumps' prior uncertainty and thus changes one's confidence. *But not all outcomes give one this kind of knowledge.* Thus, it seems permissible to use the Principal Principle in the context Howson and Urbach accept (step 1 as shown above) and then use other arguments to go from confidence before sampling to confidence after sampling (steps 2 and 3 as shown above).

11. RECOMMENDATIONS

1. *Say what 'confidence' is rather than what it is not.* Identify confidence as a psychological concept that measures one's strength of belief in the truth of a claim. Acknowledge that usually it cannot be assessed or compared much less prescribed or calibrated.

2. *Identify the role of confidence in action.* Acknowledge that confidence is the basis for action and probability is a basis for having confidence.

3. *Use confidence prior to sampling rather than only using it after sampling.* This upholds the role of confidence in going from probability to action and sets the stage for using it after a sample has been taken.

4. *Assert that the similarities between confidence and probability are typically more important than their differences.* This is certainly true in decision making. It may not be true in doing philosophy of science.

5. *If probability and confidence are used to identify a difference in reality, then use them consistently.* After flipping a fair coin, the chance of heads is no longer 50% in classical terms. Prior to seeing the result say, "We are 50% confident that this coin has been flipped heads up." In this case, the outcome is a fact of reality just as in the case of a confidence interval.

6. *Avoid over-stressing that probability equals zero or one after taking a sample.* While true, this emphasis on the difference between confidence and probability easily undermines student's awareness of their similarity from a betting perspective.

7. *Use 'confident' to emphasize that statistical confidence has a psychological import.* Given a particular 95% confidence interval, we are 95% confident that this interval includes the population parameter.

8. *Assert that from a decision-making perspective, one's confidence in a confidence interval is operationally equivalent to that in an equivalent probability.* Having confidence in a random process means we can and should bet on the process regardless of whether or not we have already sampled. Note that we 'bet on the process' in taking calculated risks whenever we play a game of cards. After the cards are dealt, the uncertainty about the distribution of cards is purely mental.

9. *Use wording that emphasizes the location of the uncertainty as being in one's mind rather than in the value of the parameter*. In betting on a 95% confidence interval, we might say, "we are willing to take a *5% risk* that the interval does not actual cover the true value." (Utts, Page 338.) We might say, "we have 19 to 1 odds of having drawn a sample whose confidence interval includes the population parameter".

REFERENCES

Hamburg, Morris (1987). *Statistical Analysis for Decision Making*, 3rd Ed., Harcourt, Brace Jovanovich,.

Howson, Colin and Peter Urbach, (1993). *Scientific Reasoning* 2nd Ed. Open Court Publishing.

Kelley, David (1994). *The Art of Reasoning*. 2nd Ed.

Lee, Peter (1989) *Bayesian Statistics, An Introduction.* Edward Arnold

Moore, David & George McCabe (1993). *Introduction to the Practice of Statistics*, 2nd ed., Freeman.

Utts, Jessica (1991), *Seeing Through Statistics*, Duxbury Press.

Acknowledgments: Dan Brick, University of St. Thomas, prompted consideration of this topic by asserting that students need to relate confidence to action. Gerald Kaminski, Thomas V.V. Burnham, Dr. Julie Naylor and Dr. Robert Raymond made helpful suggestions. Dr. Schield can be reached at schield@augsburg.edu.