



Confidence in Confidence Intervals

Author(s): Janet Bellcourt Pomeranz

Source: *Mathematics Magazine*, Vol. 55, No. 1 (Jan., 1982), pp. 12-18

Published by: Mathematical Association of America

Stable URL: <http://www.jstor.org/stable/2689857>

Accessed: 24/08/2008 19:43

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=maa>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

Confidence in Confidence Intervals

An exoteric view of confidence intervals that involves subjective probability and makes use of computer-simulated sampling.

JANET BELLCOURT POMERANZ

State University of New York Maritime College

Fort Schuyler, Bronx, NY 10465

Picture this. The long rays of late afternoon sunlight are filtering through neatly leveled venetian blinds. Chalk and erasers are in readiness alongside horizontally erased blackboards. Every student in the classroom is flipping through the pages of a textbook on probability and statistics, some casually, a few with great intensity. A bell rings, the professor enters, class is about to begin. The lesson of the day is confidence intervals.

Squinting at yellowed notes and turning as if to address the blackboard, the professor writes on the board as he solemnly recites:

If for a given α
the relation $P(\theta_1 \leq \theta \leq \theta_2) = 1 - \alpha$ holds
we say that the interval $[\theta_1, \theta_2]$ is a
 $100(1 - \alpha)\%$ confidence interval for the parameter θ .

Before the professor has finished transcribing the statement from his notes to the board, the students have swung into action and are busy copying the definition into their spiral notebooks. One student, however, balks. Having attracted attention by clearing his throat and waving frantically, he sputters, "Excuse me, Professor, but what *is* a confidence interval?" The professor peers myopically in the direction of the student. Gesturing toward the blackboard, with great deliberation he repeats the definition verbatim.

Exasperated, the student pleads, "But, Professor, you know I cannot think *abstractly*." Again the professor tries to locate the refractory student, his eyes narrowing. All at once the room has become quiet. With patience, or is it weariness, he rejoins: "On the contrary, you cannot think *except* abstractly."

This story is true, as my own spiral notebook will attest. Why do I recount the anecdote here? Is it because the student's anguish impresses me as genuine, and I wish to commiserate somehow with students everywhere? Perhaps. The real reason, though, is to demonstrate the need for a less abstruse concept of confidence interval. The professor's definition is neat and unambiguous but is hardly enlightening to the average student.

In this paper I take a more tangible approach to confidence intervals, through hypothesis testing. The interpretation of confidence interval which I propose is linked to subjective probability rather than classical probability and is, I believe, simpler than the standard textbook interpretation. As a means of demonstrating in a concrete way the salient features of confidence intervals, I utilize a BASIC program in order to simulate sampling on a computer. Besides the program itself, I include in this paper the results of several demonstrations involving computer-simulated sampling.

From the black-or-white mentality of hypothesis testing to interval estimation

"Do you think that discipline at this college should be stricter?" When I ask this question of

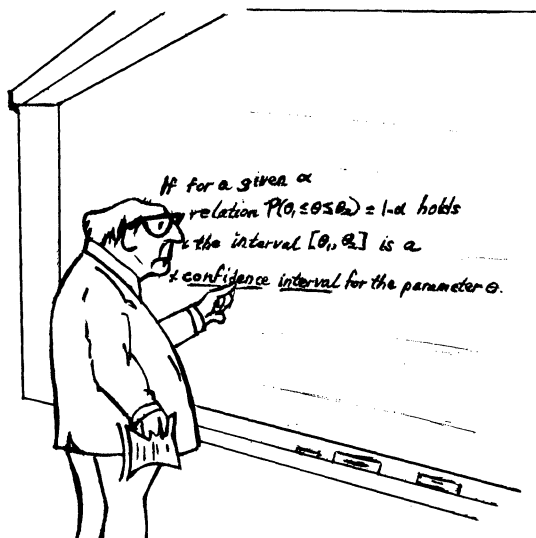


FIGURE 1. "Excuse me, Professor, but what is a confidence interval?"

my statistics class the answer is a resounding, "No." "Do you suppose all students feel this way?" In response, one member of the class ventures the opinion that possibly one in ten students would favor stricter discipline. The stage has now been set for a lesson on hypothesis testing and on the related theme of confidence intervals.

To test hypotheses concerning the proportion p of students in the college population favoring stricter discipline, we take a simple random sample of size n and compute the sample proportion, $\hat{p} = x/n$ where x is the number of students in the sample favoring stricter discipline. Under certain rather general conditions, the sampling distribution of \hat{p} is approximately normal with mean p and standard deviation $\sqrt{p(1-p)/n}$. When are conditions right? Loosely speaking, if n is large enough and p is not too close to either 0 or 1. More precisely, if $np > 5$ when $p \leq 0.5$, and $n(1-p) > 5$ when $p > 0.5$, the normal approximation to \hat{p} is satisfactory [7, p. 114].

We're interested in testing the hypothesis $H_0: p = 0.10$ against the alternative hypothesis $H_A: p \neq 0.10$. The testing procedure calls for our assuming at the outset that H_0 is true, that is, that $p = 0.10$. By choosing $n = 100$, we have $np = 10$ and easily meet the requirement that $np > 5$ when $p \leq 0.5$. Accordingly, the test statistic

$$Z = (\hat{p} - 0.10) / \sqrt{(0.10)(0.90)/100}$$

is approximately standard normal. Having decided to test at the 5% level of significance, we agree to reject H_0 if and only if $|Z| > 1.960$.

Our poll reveals that in a simple random sample of size 100, exactly 15 students favor stricter discipline. A few of the more impetuous members of my class jump to the conclusion that the population proportion p is not 0.10. After all, if p were 0.10, would \hat{p} be as large as 0.15? On calculating the test statistic

$$Z = (0.15 - 0.10) / \sqrt{(0.10)(0.90)/100} = 1.667,$$

though, we are forced to conclude that at the 5% significance level our findings are consistent with the hypothesis that $p = 0.10$. Still, most students are inclined to adjust upward the original estimate of 0.10 for p . Someone notices that the computed value of Z is a little larger than 1.645, the cutoff value for $|Z|$ at the 10% significance level, and I concede that at that level we do reject the hypothesis that $p = 0.10$. At this point some students openly express skepticism of the hypothesis testing process. Is the population proportion 0.10, or isn't it? They simply do not like a

testing procedure which on the basis of the same sample leads either to one conclusion or to its opposite, depending on one's choice of significance level for the test.

Rather than engage in the black-or-white type of thinking that characterizes hypothesis testing, statisticians often turn to confidence intervals. Instead of deciding whether the value of a parameter is equal to some preassumed value, we calculate an interval which has a preassigned probability of containing the unknown parameter. For example, a 90% confidence interval for the population proportion p consists of all values within $1.645\sqrt{\hat{p}(1-\hat{p})/n}$ of the sample proportion \hat{p} . To find a 95% or 99% confidence interval we simply replace 1.645 by 1.960 and 2.576, respectively. Using the sample results of $\hat{p} = 0.15$ and $n = 100$, we compute 90%, 95%, and 99% confidence intervals for p as follows (see FIGURE 2):

$$0.15 \pm 1.645\sqrt{(0.15)(0.85)/100} = [0.09, 0.21],$$

$$0.15 \pm 1.960\sqrt{(0.15)(0.85)/100} = [0.08, 0.22],$$

$$0.15 \pm 2.576\sqrt{(0.15)(0.85)/100} = [0.06, 0.24].$$

There is a close connection between confidence intervals and tests of hypotheses [11, p. 278], [12, p. 116], [15, p. 202], [20, p. 114]. Suppose, for example, that T is a 95% confidence interval for μ , the mean of a normal random variable, with σ^2 known. Then for the same observed sample, we accept the hypothesis $H_0: \mu = \mu_0$ as opposed to the hypothesis $H_A: \mu \neq \mu_0$ at the 5% level of significance if and only if $\mu_0 \in T$. More generally, the acceptance region for a significance level α is exactly the same as the $100(1-\alpha)\%$ confidence interval for μ . Notice in the example given above that $0.10 \in [0.08, 0.22]$; that is, the hypothesized value for p lies in the 95% confidence interval. So we accept $H_0: p = 0.10$ versus $H_A: p \neq 0.10$ at the 5% significance level. But the hypothesized value, 0.10, is also in the 90% confidence interval, and cursory analysis would lead us to accept H_0 at the 10% level too, contrary to our earlier conclusion. Why this discrepancy? The answer lies in the fact that in our example σ^2 is unknown because we don't know the value of p . While it is true that in the testing situation we assume under H_0 that the real value of p is 0.10, in the interval estimation situation the real value of p is not known. In computing a confidence

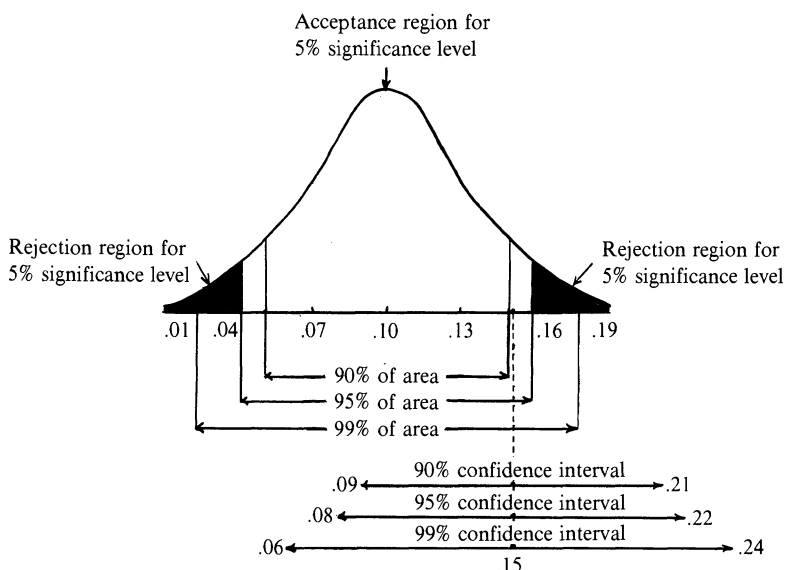


FIGURE 2. Sampling distribution of \hat{p} for samples of size 100 from a population in which $p = 0.10$, and confidence intervals for p based on a sample in which $\hat{p} = 0.15$.

interval for p , we approximate the standard deviation $\sqrt{p(1-p)/n}$ of \hat{p} by $\sqrt{\hat{p}(1-\hat{p})/n}$, and for this reason the equivalence here between confidence interval and the acceptance region of hypothesis testing is only approximate. If we're determined to eliminate the discrepancy, we can do so by utilizing the test statistic

$$Z = (\hat{p} - 0.10) / \sqrt{\hat{p}(1-\hat{p})/n},$$

which is also asymptotically standard normal. Substitution of $\hat{p} = 0.15$ and $n = 100$ in this expression yields $Z = 1.400$, implying that at the 10% level of significance our findings are consistent with the hypothesis that $p = 0.10$. However, the customary practice is to use the test statistic as originally defined, so most statisticians would agree that our original decision to reject H_0 at the 10% significance level happens to be the right one after all.

The subjective side of confidence intervals

Are you a frequentist or a subjectivist? Your answer will influence the way you interpret confidence intervals, for the way we look at confidence intervals is inextricably linked to the way we look at probability itself. There is no unique interpretation of confidence intervals, just as there is no unique interpretation of probability. To say that there is only one viable view of probability is somewhat like insisting that space is Euclidean.

How is confidence in the 95% confidence interval, $[0.08, 0.22]$, to be construed? Most authors [3, p. 106], [5, p. 157], [10, p. 245], [14, p. 164], [15, p. 176], [16, p. 303], [17, p. 204], [18, p. 275], [19, p. 298] advocate the strict interpretation that our confidence is in the *process* of taking random samples and obtaining intervals, and not in any specific interval, such as $[0.08, 0.22]$ in the present case. These authors affirm that

$$P(\hat{p} - 1.960\sqrt{\hat{p}(1-\hat{p})/n} \leq p \leq \hat{p} + 1.960\sqrt{\hat{p}(1-\hat{p})/n}) = 0.95,$$

but they deny that

$$P(0.08 \leq p \leq 0.22) = 0.95.$$

In other words, once the endpoints of the interval have been evaluated, the probability statement no longer holds. Why this subtle distinction? Because the unknown proportion p is either *in* the interval $[0.08, 0.22]$ or *not* in that interval; there's really nothing random about it. The randomness lies in the process of generating the 95% confidence intervals. Inasmuch as 95% of the time (in the long run) this process yields intervals which contain p , it is to the process itself that we ascribe a probability of 0.95.

This strict interpretation of confidence interval springs from a **frequentist** interpretation of probability, in which the aspect of probability which is stressed is the tendency displayed by some chance devices to produce stable relative frequencies. This view identifies probability with the limit of a relative frequency: to say that p is the probability that an A is a B is simply to say that p is the *limit of the relative frequency of B's among A's* (as the number of observed A 's is increased without bound) [9, p. 4]. For example, when flipping a coin a great many times, a frequentist would interpret $P(\text{heads}) = 1/2$ to mean that in the limit, half of the flips would be heads. Are you nodding in agreement with all this? If so, then you can call yourself a frequentist.

But the end result of interval estimation is a specific interval, such as the interval $[0.08, 0.22]$ we've been considering. Exactly what meaning does *it* have? Not much, in the school of the frequentists, as we have seen. There is, though, another view of confidence intervals, one which is tied to a different view of probability. The theory of **subjective probability** has been created to enable one to talk about probabilities when the frequency viewpoint does not apply [1, p. 61]. This is often the case, since probability when used as a guide in life refers, not to the frequency with which a random variable will take a certain value, but to the likelihood that a certain *constant* will have a certain value [8, p. 25].

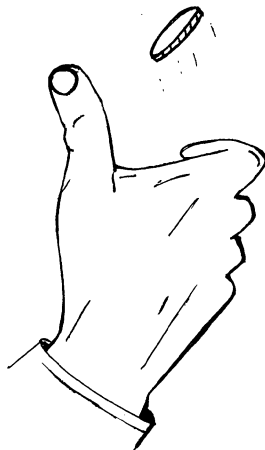


FIGURE 3. $P(\text{heads}) = \frac{1}{2}$: what this means to you depends on whether you are a frequentist or a subjectivist.

The basic thesis of the subjective theory of probability is that probability statements are statements concerning actual degrees of belief [8, p. 30]. In the example of flipping a coin, a subjectivist would interpret $P(\text{heads}) = .5$ to reflect the betting odds for heads, i.e., he would say there were even odds for heads vs. tails. Thus in the subjective theory of probability, we define the probability of an event as a number between 0 and 1 that reflects our personal assessment of the chance that the event will occur. Notice that this value is not uniquely determined, inasmuch as it depends on the inclination of the person whose degree of belief that probability represents. This nonuniqueness clearly differentiates subjective probabilities from classical probabilities, which derive either from the full understanding of the mechanics of a process such as dice shooting or from extensive empirical observation of relative frequencies [21, p. 214]. Aside from lacking the precision of classical probabilities, subjective probabilities are problematical in other ways.

In practice, how do we determine the subjective probability of an event? The easiest way is to compare events, determining relative likelihoods. For example, to find $P(E)$, the probability of event E , compare E with E^c , the complement of E . If we feel that E is about twice as likely to happen as E^c , then we fix $P(E)$ at $2/3$ [1, p. 62]. Put another way, our degree of belief in a given statement may be indicated by the odds at which we're willing to bet on its truth [8, p. 5]. If this approach appeals to you, consider yourself a subjectivist. Be aware, however, of the problems that are peculiar to subjective probabilities. In addition to the lack of precision caused by the varying degrees of belief different persons have in the same statement, subjective probabilities are prone to a certain fuzziness even when just one person is involved. This fuzziness is suggested by our own inconsistency when we're forced to compare events several times, especially if we don't realize we're making the same comparison or can't remember what our previous choice was [13, p. 373]. We can't let these probabilities become too fuzzy, however, because even subjective probabilities are bound by certain laws. For example, we have to guard against this type of irrational assignment of probabilities: $P(A) = 1/3$, $P(B) = 1/3$, $P(\text{either } A \text{ or } B \text{ occurring}) = 3/4$ [1, p. 62].

We come now to the main point of this section of the paper. Certainly the subjective rather than frequency viewpoint of probability applies when we speak of the probability that there will be another frost this spring, or that the stock market will fall, or that the U.S. will retain the America's Cup. So too, does the subjective concept apply to the probability that the proportion of smokers in the U.S. is between 0.3 and 0.4, or the probability that the proportion of students in the college favoring stricter discipline is between 0.08 and 0.22. In keeping with this view we adopt a looser interpretation of confidence intervals. In short, we place our confidence in the specific interval as well as in the general process of generating the confidence intervals, by our assertion

that

$$P(0.09 \leq p \leq 0.21) = 0.90,$$

$$P(0.08 \leq p \leq 0.22) = 0.95,$$

$$P(0.06 \leq p \leq 0.24) = 0.99.$$

Some authors [2, p. 254], [4, p. 262], [11, p. 237] espouse this loose interpretation of confidence intervals, but they do so with an apology for what they deem a lack of precision or even an abuse of language. To my way of thinking, this interpretation requires no apology. My proposal is that having noted with those of the frequentist persuasion that the process of generating 95% confidence intervals works 95% of the time, let us simply set 0.95 equal to the probability that a specific 95% confidence interval contains the population proportion p .

Demonstrations on confidence intervals through computer-simulated sampling

As suggested by the opening anecdote of this paper, students sometimes find these concepts hard to grasp. One way of reducing the abstractions to more concrete terms is to illustrate the meaning of confidence intervals through computer-simulated sampling.

Using this technique, you can demonstrate for yourself that the process of generating 95% confidence intervals yields intervals which, 95% of the time, contain the population proportion p , provided the assumptions underlying the process are satisfied (and if they are not satisfied, that the process fails). The BASIC program listed below will enable you to do just that. This program produces a large number m of simple random samples of size n from a population in which the relevant proportion p is $1/L$ and for each sample computes a $100(1 - \alpha)\%$ confidence interval for p . Then the program computes the percentage of the m intervals that actually contain p . Notice that in the second line, the program directs you to "ENTER L, N, M, A, AND Z." Choose your population proportion p to be of the form $1/2, 1/3, 1/4, \dots$, and then let L be the integer $1/p$. As indicated above, "N" is the sample size and "M" is the number of samples to be generated. For 90%, 95%, or 99% confidence intervals, let "A" equal 0.10, 0.05, 0.01, respectively, and let "Z" equal 1.645, 1.960, 2.576, respectively. Other combinations of "A" and "Z" are permissible provided, of course, that "A" equals the probability that the standard normal variable is greater than "Z."

```
100 REM- CONFIDENCE INTERVALS FOR P
110 INPUT "ENTER L, N, M, A, AND Z"; L,N,M,A,Z
120 RANDOMIZE: P = 1/L: B = 100*(1 - A):T = 0
130 DIM P(M), S(M): FOR J = 1 TO M: X = 0
140 FOR I = 1 TO N: R = INT(RND*L)
150 X = X - (R = 0): NEXT I
160 P(J) = X/N: S(J) = SQR(P(J)*(1 - P(J))/N)
170 T = T - ((ABS(P(J) - P) < (Z*S(J))): NEXT J
180 C = 100*T/M
190 LPRINT "FOR L, N, M EQUAL TO ";L,N,M
200 LPRINT "% OF ";B;"% CONFIDENCE INTERVALS CONTAINING P: ";C
210 END
```

	L	N	M	A	% CONFIDENCE INTERVALS CONTAINING P
(1)	4	40	100	0.05	94
(2)	20	40	100	0.05	88
(3)	100	40	100	0.05	38
(4)	10	10	100	0.05	63
(5)	10	40	100	0.05	90
(6)	10	80	100	0.05	89
(7)	4	40	100	0.10	88
(8)	4	40	100	0.05	95
(9)	4	40	100	0.01	99

TABLE 1

The usefulness of this program lies in its ability to generate and analyze a large number of samples within a relatively short period of time. You can vary the parameters at will and see the effect immediately. TABLE 1 displays data obtained on a number of runs of this program. The number M of samples was 100 throughout, each of the other parameters was varied in turn. In rows (1) through (3) $P = 1/L$ was varied, in rows (4) through (6) n was varied, and in rows (7) through (9), α was varied.

Notice in TABLE 1 that only in two cases did the number of confidence intervals containing p fall far below the number expected. This happened in row (3), where the population proportion p is a small 0.01, and in row (4), where the sample size n is only 10. In both instances, np is less than the required 5: $np = 0.4$ for row (3) and $np = 1$ for row (4). Thus these experimental results are in agreement with the theory.

Other applications of this program will occur to the creative teacher. And to struggling students as well.

References

- [1] J. O. Berger, *Statistical Decision Theory*, Springer-Verlag, New York, 1980.
- [2] J. D. Braverman, *Fundamentals of Business Statistics*, Academic, New York, 1978.
- [3] R. E. Chandler, *The statistical concepts of confidence and significance*, Statistical Issues, edited by R. E. Kirk, Brooks/Cole, Monterey, CA, 1972.
- [4] B. Gelbaum and J. G. March, *Mathematics for the Social and Behavioral Sciences*, Saunders, Philadelphia, PA, 1969.
- [5] D. G. Haack, *Statistical Literacy: A Guide to Interpretation*, Duxbury, N. Scituate, MA, 1979.
- [6] I. Hacking, *The Emergence of Probability*, Cambridge, New York, 1975.
- [7] P. G. Hoel and R. J. Jessen, *Basic Statistics for Business and Economics*, Wiley, New York, 1971.
- [8] H. E. Kyburg, Jr., *Probability and the Logic of Rational Belief*, Wesleyan Univ. Press, Middletown, CT, 1961.
- [9] H. E. Kyburg, Jr., and H. E. Smokler, Editors, *Studies in Subjective Probability*, Wiley, New York, 1964.
- [10] L. Lapin, *Statistics for Modern Business Decisions*, Harcourt Brace Jovanovich, New York, 1978.
- [11] H. J. Larson, *Introduction to Probability Theory and Statistical Inference*, Wiley, New York, 1968.
- [12] S. G. Levy, *Inferential Statistics in the Behavioral Sciences*, Holt, Rinehart, and Winston, New York, 1968.
- [13] R. D. Luce and H. Raiffa, *Games and Decisions*, Wiley, New York, 1957.
- [14] E. Lukacs, *Probability and Mathematical Statistics*, Academic, New York, 1972.
- [15] H. J. Malik and K. Mullen, *Applied Statistics for Business and Economics*, Addison-Wesley, Reading, MA, 1975.
- [16] P. L. Meyer, *Introductory Probability and Statistical Applications*, Addison-Wesley, Reading, MA, 1970.
- [17] R. L. Mills, *Statistics for Applied Economics and Business*, McGraw-Hill, New York, 1977.
- [18] D. S. Moore, *Statistics: Concepts and Controversies*, Freeman, San Francisco, CA, 1979.
- [19] F. Mosteller, R. E. K. Rourke, and G. B. Thomas, *Probability with Statistical Applications*, 2nd ed., Addison-Wesley, Reading, MA, 1973.
- [20] M. G. Natrella, *The relation between confidence intervals and tests of significance*, Statistical Issues, edited by R. E. Kirk, Brooks/Cole, Monterey, CA, 1972.
- [21] H. Schwarz, *The use of subjective probability methods in estimating demand*, Statistics: A Guide to the Unknown, edited by J. M. Tanur, Holden-Day, San Francisco, CA, 1972.