
In Defense of the Neyman-Pearson Theory of Confidence Intervals

Author(s): Deborah G. Mayo

Source: *Philosophy of Science*, Jun., 1981, Vol. 48, No. 2 (Jun., 1981), pp. 269-280

Published by: The University of Chicago Press on behalf of the Philosophy of Science Association

Stable URL: <https://www.jstor.org/stable/187185>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



The University of Chicago Press and Philosophy of Science Association are collaborating with JSTOR to digitize, preserve and extend access to *Philosophy of Science*

JSTOR

IN DEFENSE OF THE NEYMAN-PEARSON THEORY OF CONFIDENCE INTERVALS*

DEBORAH G. MAYO†

*Department of Philosophy and Religion
Virginia Polytechnic Institute & State University*

In *Philosophical Problems of Statistical Inference*, Seidenfeld argues that the Neyman-Pearson (NP) theory of confidence intervals is inadequate for a theory of inductive inference because, for a given situation, the 'best' NP confidence interval, $[CI_\lambda]$, sometimes yields intervals which are trivial (i.e., tautologous). I argue that (1) Seidenfeld's criticism of trivial intervals is based upon illegitimately interpreting confidence levels as measures of final precision; (2) for the situation which Seidenfeld considers, the 'best' NP confidence interval is not $[CI_\lambda]$ as Seidenfeld suggests, but rather a one-sided interval $[CI_0]$; and since $[CI_0]$ never yields trivial intervals, NP theory escapes Seidenfeld's criticism entirely; (3) Seidenfeld's criterion of non-triviality is inadequate, for it leads him to judge an alternative confidence interval, $[CI_{alt.}]$, superior to $[CI_\lambda]$ although $[CI_{alt.}]$ results in counterintuitive inferences. I conclude that Seidenfeld has not shown that the NP theory of confidence intervals is inadequate for a theory of inductive inference.

1. Introduction. In his recent book, *Philosophical Problems of Statistical Inference* (Chapter 2), Teddy Seidenfeld presents an interesting argument which he takes as discrediting the Neyman Pearson (NP) theory of statistical inference. The statistical methods of NP theory are used widely in the sciences, more than any other theory of inference; and should these methods prove to be inadequate or misconceived, the validity of a good deal of scientific inference will be open to serious question. Thus, when a reputable philosopher of statistics concludes that NP theory is seriously inadequate as a theory of statistical inference, his argument warrants careful scrutiny.

Focusing directly upon confidence intervals (and indirectly upon their corresponding hypotheses tests) Seidenfeld argues that while the 'best' NP confidence interval may be reasonable before observing the data (i.e., on the 'forward look') it may no longer be reasonable once the data is observed (i.e., on the 'backward look'). In the example Seidenfeld con-

*Received August 1980; Revised January 1981.

†I would like to thank Ronald Giere and Teddy Seidenfeld for very useful comments on an earlier draft of this paper. I am very grateful to I. J. Good, Klaus Hinkelmann, and George Shapiro for helpful discussions concerning this paper. I am particularly grateful to Klaus Hinkelmann for our discussions of one-sided confidence intervals.

Philosophy of Science, 48 (1981) pp. 269–280.
Copyright © 1981 by the Philosophy of Science Association.

siders, the confidence interval which is purported to be the 'best' NP interval, $[CI_\lambda]$, is unreasonable on the 'backward look' in the sense that it sometimes produces logically trivial (i.e., tautologous) intervals. Other arguments to the effect that NP theory fails on the 'backward look' are given by Hacking (1965) and Spielman (1973), although these are aimed at hypotheses tests. Seidenfeld claims that there exists a confidence interval $[CI_{alt.}]$, which, while not the 'best' on NP criteria, is actually superior to the 'best' NP interval, $[CI_\lambda]$. The sense in which $[CI_{alt.}]$ is superior to $[CI_\lambda]$ is that it produces trivial intervals less often. The criterion of non-triviality of confidence intervals will be referred to as Seidenfeld's criterion.

In this paper I argue that: (1) the fact that a theory of confidence intervals sometimes produces trivial intervals is not a substantial ground for criticizing that theory, and Seidenfeld's reason for taking it to be a substantial ground for criticism is based upon a common, but illegitimate, interpretation of NP confidence levels; (2) contrary to Seidenfeld's suggestion, the 'best' NP confidence interval is not $[CI_\lambda]$, but rather a one-sided confidence interval, $[CI_0]$; and since $[CI_0]$ always produces intervals which meet Seidenfeld's criterion of non-triviality, $[CI_0]$, and hence, NP theory, is not open to the criticism that Seidenfeld raises; (3) Seidenfeld's criterion is inadequate as a criterion of interval estimates, as it leads him to judge the confidence interval $[CI_{alt.}]$ superior to $[CI_\lambda]$ although $[CI_{alt.}]$ results in counterintuitive inferences. The problem stems from the fact that $[CI_{alt.}]$ is an interval out of which a 'bite' has arbitrarily been taken, causing it to omit parameter values which it ought reasonably to include.

2. $[CI_\lambda]$. Seidenfeld asks us to consider a random variable X having the uniform distribution between 0 and θ (θ positive), with density function $f_X(x)$:¹

$$\begin{aligned} &0, \text{ if } X \leq 0 \\ &1/\theta, \text{ if } 0 < X \leq \theta \\ &0, \text{ if } X > \theta. \end{aligned}$$

Assuming that only a *single* observation is made on X , we are to provide an estimate of the unknown value of θ by giving an interval of values in which θ is presumed to lie. NP theory provides a rule for forming such intervals so that $(1 - \alpha)$ 100% of them will contain the correct value of θ . These are $(1 - \alpha)$ level confidence intervals. For the random variable

¹Throughout, 'X' will be used to refer to the random variable being considered, and 'x' to its value.

with the above uniform density the ‘best’² (two-sided) $(1 - \alpha)$ confidence interval is defined by:

$$X \leq \theta \leq X/\alpha. \quad [CI_\lambda]$$

For example, if X is observed to be 10 and α is set at .05, the 95% $[CI_\lambda]$ is $[10,200]$.

Seidenfeld then asks us to consider a situation in which additional information about θ is known. In addition to knowing that $\theta > 0$, it is now known that there is an upper bound of 15 for θ , i.e., $\theta \leq 15$. Seidenfeld claims that NP theory still recommends the same confidence interval as it did in the case where θ was not known to be truncated, namely $[CI_\lambda]$. Admittedly, $[CI_\lambda]$ is still the ‘best’ two-sided $(1 - \alpha)$ interval, but the appropriateness of a two-sided rather than a one-sided interval for the truncated case is open to question. I shall take up this issue in the next section.

If $[CI_\lambda]$ is still used when θ is truncated at 15, it will sometimes yield intervals which are ‘‘logically trivial’’. For example, if it is observed that $X = 10$, the 95% $[CI_\lambda]$ is no longer $[10,200]$ but rather $[10,15]$. The problem with this interval, according to Seidenfeld, is that it does not tell us anything beyond what we know with probability 1 once the observation is made: namely, that $10 \leq \theta \leq 15$. It is a ‘‘logically trivial’’ or tautologous interval. In fact, any observation: $X \geq 3/4$, leads to a trivial 95% $[CI_\lambda]$. More generally, when it is given that the upper bound for θ is c_u , $[CI_\lambda]$ yields a trivial $(1 - \alpha)$ interval whenever $X \geq \alpha c_u$.

But why are trivial intervals found to be so objectionable to Seidenfeld that he takes them as indicative of a serious deficiency in NP theory? There are two things that Seidenfeld finds objectionable about trivial intervals. For one thing, they do not provide any information beyond what the data provides; and there is no need to have a special procedure for confidence intervals if they tell you no more than you already know. However, this does not mean the theory of NP confidence intervals is inadequate. It may just be that when you request the theory to provide a certain type of interval estimate, and a certain type of observation arises, it can do no more than tell you something which turns out to be

²A ‘best’ confidence interval is understood as one which is uniformly most accurate (UMA) for appropriate values of the parameter of interest. A *two-sided* $(1 - \alpha)$ confidence interval is UMA if, regardless of the true value of the parameter, it is less probable for it to contain incorrect values of the parameter than it is for any other two-sided $(1 - \alpha)$ confidence interval to do so. That is, $[CI^*]$ is the UMA two-sided confidence interval at level $(1 - \alpha)$ if for any other two-sided $(1 - \alpha)$ confidence interval $[CI]$, $P(\theta' \in [CI^*] | \theta) \leq P(\theta' \in [CI] | \theta)$ for all θ and θ' where θ' is an incorrect value of the parameter. Neyman refers to the UMA confidence interval as the ‘shortest’, but this is misleading since this criterion does not directly refer to the length of the interval; the UMA confidence interval is not the same as the uniformly narrowest.

obviously true, given additional background information. Yet, it seems that what disturbs Seidenfeld is not just that NP theory sometimes yields an interval which is obviously true, but that it presents the interval as a $(1 - \alpha)100\%$ interval and not as a 100% interval. The problem with this, according to Seidenfeld, is that trivial intervals are “true not merely at the $(1 - \alpha)$ level, but at the 100% level.” (p.56) But finding this problematic seems based on a misinterpretation of NP confidence levels—one which is not uncommon.

Seidenfeld claims (using a formulation of Hacking) that:

it seems reasonable to say, *before knowing the data*, that the $[CI_\lambda]$ interval will cover the true value of the unknown parameter with probability $(1 - \alpha)$; however, *having seen the value x* , it may be unreasonable to maintain the probability statement or use it to express a degree of confidence in the interval generated by the $[CI_\lambda]$ rule. (pp. 56–57)

It must be stressed, however, that having seen the value x , NP theory *never* permits one to conclude that the specific confidence interval formed covers the true value of θ with either $(1 - \alpha)100\%$ probability or $(1 - \alpha)100\%$ degree of confidence. Seidenfeld’s remark seems rooted in a (not uncommon) desire for NP confidence intervals to provide something which they cannot legitimately provide; namely, a measure of the degree of probability, belief, or support that an unknown parameter value lies in a specific interval. Following Savage (1962), the probability that a parameter lies in a specific interval may be referred to as a measure of *final precision*. While a measure of final precision may seem desirable, and while confidence levels are often (wrongly) interpreted as providing such a measure, no such interpretation is warranted. Admittedly, such a misinterpretation is encouraged by the word “confidence”.

Whenever $X \geq 3/4$, the 95% $[CI_\lambda]$ (for θ truncated at 15) yields an interval in which θ lies with probability 1. That is, the final precision of a trivial interval is 1. One is tempted to suppose, along with Seidenfeld, that in such cases $[CI_\lambda]$ is really a 100% and not a 95% confidence interval. But this is to identify incorrectly confidence levels with measures of final precision. Even a trivial 95% $[CI_\lambda]$ is still a 95% confidence interval. The reason is that it arose from a rule which is known to generate confidence intervals 95% of which will cover the true value of θ , regardless of what θ is. Hence, *before the data* is observed it is known that there is a probability of .95 that the interval that $[CI_\lambda]$ will generate will contain the true value of θ . This is a measure of *initial precision*. Once the data is known, this probability measure cannot legitimately be attached to the specific interval formed (i.e., it cannot be used as a measure of final precision). In the context of confidence intervals NP theory views

θ not as a random variable, but as a fixed quantity; and hence, the probability that θ is in a specific interval is always either 0 or 1! And finding the final precision to be either of these extreme values is not at all inconsistent with the initial precision being some non-extreme value between 0 and 1.

Seidenfeld's objection (pp. 56–57) says in effect: "Having seen the value x , it may be unreasonable to use $(1 - \alpha)$ as a measure of final precision." To this the NP theorist could reply that he never intended for a confidence level to be interpreted as a measure of final precision; and that he never attempted to supply such a measure, believing, as he does, that such measures are illegitimate. It is not the fault of NP theory that by misinterpreting confidence levels an invalid measure of final precision results. However, what really seems to be at the heart of Seidenfeld's criticism is the supposition that unless NP confidence intervals can be interpreted as providing a measure of final precision, they are inadequate for making inferences about estimates. In other words, in order for NP intervals to be useful for inference one is forced to interpret them illegitimately. This supposition is common among critics of NP theory: a theory is deemed inadequate for inference unless it can use data to assign a probability to a claim about the value of θ , i.e., unless it can provide a measure of final precision. On the basis of his criticism, Seidenfeld concludes that "the NP theory cannot serve as an adequate replacement for an inductive logic, especially for inverse inference." (p. 37) But his argument does not warrant this conclusion. All that can be concluded is that NP confidence levels do not provide measures of final precision, and thus NP confidence intervals cannot provide an inverse inference (i.e., an inference from sample data to a population parameter) of the form: the probability that θ is in the interval $[a, b]$ equals p , where a, b are numbers. But NP theory never claimed to provide techniques for an inverse inference of this form. The whole point of the NP program was to build a theory in which inverse inference involves not final but rather initial precision. (In the case of confidence intervals, the initial precision is the probability (before observing the data) that the confidence interval which the data will generate will contain the correct parameter value.) And Seidenfeld has not shown that the form of inference that the NP program provides is inadequate for inductive logic, unless such a logic is seen as requiring a measure of final precision, which would be to beg the question against NP theory. In my own estimation, the NP solution to the problem of inverse inference can provide an adequate inductive logic, and NP confidence intervals can be interpreted in a way which is both legitimate and useful for making inferences. But much remains to be done in setting out the logic of confidence interval estimation before this claim can be supported—a task which requires a separate paper.

3. $[CI_0]$. Even if the occasional production of trivial intervals is to be taken as a deficiency of a confidence interval theory, I would argue that Seidenfeld's example has still not shown that NP theory has such a deficiency. My argument is that for the case Seidenfeld considers (i.e., θ truncated) the 'best' NP interval is not the two-sided interval $[CI_\lambda]$ as he suggests, but rather a one-sided interval $[CI_0]$; and $[CI_0]$ produces only non-trivial intervals. To see how $[CI_0]$ is generated, it is helpful to see how one and two-sided tests may be used to generate one and two-sided confidence intervals, respectively.

A NP test is a rule which designates which observations will be taken to reject the null or test hypothesis in favor of an alternative hypothesis. These observations make up the *critical region*, and it is specified so that there is a given probability, α , the *size* of the test, that the observation will fall in it if the null hypothesis is true. NP confidence intervals are intimately related to NP tests in that a confidence interval consists of those null hypotheses (concerning the value of θ) which the observed data does not reject. That is, if C_T is the critical region of a test of $h_0: \theta = \theta_0$ with size α , then the $(1 - \alpha)$ confidence interval generated by x is $[CI_T] = \{\theta_0: x \notin C_T\}$. This relationship is of interest because 'best' (uniformly most powerful) tests can be used to generate 'best' (uniformly most accurate) confidence intervals. Seidenfeld generates the two-sided interval $[CI_\lambda]$ from the two-sided test of $h_0: \theta = \theta_0$ vs. $h_1: \theta \neq \theta_0$ with critical region $C_\lambda = \{x: (x \leq 0) \text{ or } (x > \theta_0) \text{ or } (x/\theta_0 \leq \alpha)\}$. In the case where θ is truncated from above, however, it seems that a one-sided test would generate a more appropriate confidence interval; namely, one which is one-sided.

In a two-sided interval one specifies both a lower and an upper confidence bound for the parameter θ ,³ while in a one-sided interval only either an upper or a lower bound is specified. A lower bound may be of interest when θ is the breaking strength of a metal; an upper bound when θ is the toxicity of a drug, or the probability of some undesirable event. A good discussion of one-sided confidence intervals occurs in Lehmann (1959). I suggest that in the situation where θ is truncated from above, a one-sided (lower) confidence interval is called for. The one-sided test which can be used to generate such a one-sided (lower) interval is the test of $h_0: \theta \leq \theta_0$ vs. $h_1: \theta > \theta_0$. The uniformly most powerful test of size α for these hypotheses, still assuming only one observation is to be made on X , has critical region $C_0 = \{x: x \geq \theta_0 (1 - \alpha)\}$ and the corresponding uniformly most accurate confidence interval is $[CI_0] = \{\theta: \theta \geq X/(1 - \alpha)\}$. Since we are assuming that θ is known to have an upper

³Consideration of confidence intervals for more than one parameter is omitted here as it is not relevant for the purposes of this paper.

bound c_u , the resulting $(1 - \alpha)$ confidence interval is defined by:

$$X/(1 - \alpha) \leq \theta \leq c_u. \tag{CI_0}$$

What makes $[CI_0]$ of significant interest for the case in which θ is truncated, is that it produces non-trivial intervals for all possible observations (provided $\alpha > 0$), and hence it satisfies Seidenfeld's criterion for confidence intervals.⁴ For example, when $X = 10$, $\alpha = .05$, $c_u = 15$, $[CI_0]$ is $[10.53, 15]$ which is a non-trivial interval. $[CI_0]$, I claim, is appropriate for the case where θ is truncated because in such a case one already knows what the upper bound for θ is, hence, one is only interested in setting a lower bound for θ ; and $[CI_0]$ is the 'best', unbiased, one-sided (lower) confidence interval for θ when X has the uniform distribution.⁵ The NP theorist should no more be expected to apply the confidence interval appropriate when θ is not truncated (i.e., $[CI_\lambda]$) to a situation in which θ is truncated, than he should be expected to apply a two-sided test to a situation in which a one-sided test is appropriate. And we have seen that $[CI_\lambda]$ and $[CI_0]$ correspond to a two-sided and a one-sided test,

⁴As Seidenfeld has recently pointed out to me in private conversation, when $x > (1 - \alpha) c_u$, $[CI_0]$ contains points which are impossible for θ to have, assuming it is truncated at c_u . Hence whenever $x \geq (1 - \alpha) c_u$, $[CI_0]$ collapses to the limiting case of the interval; namely, $\theta = c_u$. Still these cases are not trivial in Seidenfeld's sense since it is not certain that $\theta = c_u$ unless $x = c_u$. Nor does it seem correct to consider the case where $x = c_u$ trivial. For in that case there simply is no more information to report other than that $\theta = c_u$; any other estimate of θ would be less informative.

One might be troubled by the fact that, strictly speaking, when $x \geq (1 - \alpha) c_u$, $[CI_0]$ is not an interval but rather a point. One way of getting around this is to let $[CI_0]$ be set equal to a non-trivial interval of the form $\left[\frac{(b-1)x + c_u}{b}, c_u \right]$ some $b \geq 0$, whenever $x \geq (1 - \alpha) c_u$. (For example, letting $b = 2$ results in $[CI_0]$ extending from the midpoint of x and c_u , to c_u .) This is, in effect, to make $[CI_0]$ a *conditional interval*, conditional as it is on whether or not $x \geq (1 - \alpha) c_u$. Letting $[CI_0]$ be conditional in this way only serves to improve its confidence level which would now be greater than $1 - \alpha$. Such a confidence interval is termed a "conservative $1 - \alpha$ confidence interval". However it is not possible to determine how much greater than $1 - \alpha$ it is without having a prior distribution for θ .

⁵The criterion for a *best one-sided confidence interval* is set out in Neyman (1937). It is required that the interval be uniformly most accurate (UMA) only for the appropriate parameter points. An interval $[CI^*]$ is the UMA one-sided lower confidence interval at level $(1 - \alpha)$ if for any other one-sided lower $(1 - \alpha)$ confidence interval $[CI]$, $P(\theta \in [CI^*] | \theta) \leq P(\theta \in [CI] | \theta)$ for any $\theta > \theta'$. $[CI_0]$ may be shown to be the UMA one-sided lower $(1 - \alpha)$ confidence interval.

$[CI_0]$ may also be shown to be an *unbiased* one-sided lower confidence interval. A one-sided lower confidence interval $[CI]$ is an *unbiased one-sided lower $(1 - \alpha)$ confidence interval* if $P(\theta \in [CI] | \theta) \leq 1 - \alpha$ for all $\theta > \theta'$. (See Lehmann, pp. 176-177 or Kendall, p. 118.) That $[CI_0]$ is an unbiased one-sided lower $(1 - \alpha)$ confidence interval may be demonstrated as follows: Assuming that $\theta \leq c_u$, $P(\theta \in [CI_0] | \theta > \theta') = P(X/(1 - \alpha) \leq \theta' < \theta \leq c_u | \theta > \theta') = P(X \leq \theta'(1 - \alpha) | \theta > \theta') = \theta'(1 - \alpha)/\theta < 1 - \alpha$ since $\theta > \theta'$.

Since $[CI_0]$ corresponds to the test of $h_0: \theta \leq \theta'$ vs. $h_1: \theta > \theta'$, the statement of unbiasedness is asserting that $[CI_0]$ is less likely to contain false parameter values than true ones.

respectively. To use $[CI_\lambda]$ when given the additional information that θ is truncated is simply to ignore this additional information; by not ignoring it, $[CI_0]$ becomes a more relevant interval. If I am right, and in the case where θ is truncated NP theory should recommend $[CI_0]$ rather than $[CI_\lambda]$, then NP theory escapes Seidenfeld's criticism entirely.

I conclude that even if a theory of confidence intervals is to be deemed inadequate on the grounds that it sometimes produces trivial intervals (grounds which I argue are weak), Seidenfeld's example fails to provide such grounds, though it was intended to do so. The reason is that whether or not θ is truncated, NP theory recommends intervals which are always non-trivial: if θ is not truncated it advocates $[CI_\lambda]$ which is always non-trivial and if θ is truncated it recommends $[CI_0]$ which also is always non-trivial. Conceivably there is a different example in which NP theory recommends an interval, which, for some observations is trivial. All I am claiming is that Seidenfeld has not presented such an example.

4. $[CI_{alt.}]$. We can go further; for even if another example arises in which NP theory does recommend a confidence interval which sometimes yields trivial intervals, or even if one insists that $[CI_\lambda]$ rather than $[CI_0]$ is to be recommended in the example we have been considering, it does not follow that an alternative confidence interval which less often yields trivial intervals is superior, even on the 'backward look'. Seidenfeld provides such a confidence interval as an alternative to $[CI_\lambda]$. This alternative $(1 - \alpha)$ interval, $[CI_{alt.}]$ is defined as:

$$X \leq \theta \leq X/(1 - [.1 \cdot \alpha]) \text{ and } X/(1 - [1.1 \cdot \alpha]) \leq \theta. \quad [CI_{alt.}]$$

For $\alpha = .05$ $[CI_{alt.}]$ becomes:

$$X \leq \theta \leq X/.995 \text{ and } X/.945 \leq \theta.$$

$[CI_{alt.}]$ arises from the test of size α ($\alpha \leq .9$) with critical region $C_{alt.}$ given by:

$$C_{alt.} = \{x: (1 - [1.1 \cdot \alpha]) \theta_0 \leq x \leq (1 - [.1 \cdot \alpha]) \theta_0\} \\ \cup \{x: x < 0 \text{ or } x > \theta_0\} \text{ where } h_0: \theta = \theta_0.$$

According to Seidenfeld, considering $[CI_{alt.}]$ further discredits NP criteria because while $[CI_{alt.}]$ is not 'best' on NP criteria, it is actually superior, at least when θ is truncated, to the 'best' NP interval; namely, $[CI_\lambda]$. But the only sense in which $[CI_{alt.}]$ is superior to $[CI_\lambda]$ is that it produces trivial intervals for less observations on X . More specifically, when it is known that $\theta \leq c_u$, $[CI_{alt.}]$ produces trivial intervals for $X \geq (1 - [.1 \cdot \alpha])c_u$, while $[CI_\lambda]$ produces trivial intervals for $X \geq \alpha c_u$. As long as $\alpha \leq .9$, more observations lead $[CI_\lambda]$ to yield trivial intervals than

lead $[CI_{alt.}]$ to do so. For example, if $\alpha = .05$, $c_u = 15$, $[CI_\lambda]$ is trivial whenever $X \geq .75$, while the only time $[CI_{alt.}]$ is trivial is when $X \geq 14.925$. On Seidenfeld's criterion, then, $[CI_{alt.}]$ is superior to $[CI_\lambda]$. On these grounds Seidenfeld finalizes his argument against NP theory. He states:

In conclusion, it is claimed that the NP theory fails on the 'after trial' analysis since the 'best' interval, based on the 'best' test can be logically trivial (for almost all of the sample space); whereas a NP *inferior* test and *inferior* interval has superior 'after-trial' consequences. (p. 58)

I have already argued in Section 3 that the 'best' NP interval (for the case where θ is truncated) is $[CI_0]$ which is never logically trivial. But even if it is assumed that NP theory recommends $[CI_\lambda]$ for this case, it seems that Seidenfeld's argument rests upon other faulty assumptions. His argument assumes that if one can find an interval other than the 'best' NP interval, and if there is any respect in which this alternative interval is preferable to the 'best' NP interval once the data is in, then it can be concluded that NP theory "fails on the 'after trial' analysis". This is false in general because the alternative interval found may be worse than the 'best' NP interval in any number of other respects, rendering it inadequate for estimating the parameter in question. No statistical inference theory pretends to provide inferences which will be seen to be best in all respects once the data is in. In suggesting 'best' inferences NP theory attempts to strike a balance among several competing criteria for inferences (e.g., size, power, bias.) (Often there is no one 'best' NP inference, but rather a number of possible inferences, each satisfying different criteria.) Thus, finding an alternative interval preferable in some respect to the 'best' NP interval does not justify Seidenfeld's conclusion that "NP theory cannot serve as an adequate replacement for an inductive logic. . . ." For, $[CI_{alt.}]$ is only superior to $[CI_\lambda]$ with respect to Seidenfeld's criterion; it may fail miserably on other criteria rendering it unsuitable for estimating θ . One such criterion will be pointed out below.

Moreover, why should it even be supposed that satisfying Seidenfeld's criterion is a desirable thing for an interval estimate to do? Nowhere does Seidenfeld provide reasons for thinking that an interval that is certain to contain the true value of a parameter is to be avoided. By calling such intervals "trivial" he certainly makes them sound undesirable, but one could equally well (and perhaps more accurately) call them "certain" intervals. In Section 2 I suggested one reason that one might object to trivial intervals, but this was seen to be based on mistakenly supposing that a confidence level is to be a measure of final precision. Surely if an inference theory led to trivial intervals for all problems of interval esti-

mation, that theory would be uninformative. But for certain problems of interval estimation it may very well be that an interval which is trivial for most observations is the most adequate interval, all things considered. As such, Seidenfeld has not provided grounds for his claim that consideration of $[CI_{alt.}]$ serves to discredit NP theory. In fact, it seems that consideration of $[CI_{alt.}]$ can be taken to show the opposite of what Seidenfeld intended it to show. Rather than discredit NP criteria, it can be seen as discrediting the criterion under which $[CI_{alt.}]$ is deemed superior to $[CI_{\lambda}]$ —namely, Seidenfeld's criterion. The reason is that $[CI_{alt.}]$ may be seen to lead to counterintuitive inferences.

The superiority of $[CI_{alt.}]$ to $[CI_{\lambda}]$ if it exists, must show up in the cases where $[CI_{\lambda}]$ yields trivial intervals (i.e., θ is truncated); for Seidenfeld acknowledges the inferiority of $[CI_{alt.}]$ when θ is not truncated, since then it yields only infinite intervals. In addition, its alleged superiority is to be found on the 'backward look' for Seidenfeld is not claiming it is superior on the 'forward look'.⁶ So we must consider a case where the data is in. Suppose, for example, that θ is truncated at 15, $\alpha = .05$, and $X = 10$. Then $[CI_{alt.}] = 10 \leq \theta \leq 10.05$ and $10.58 \leq \theta \leq 15$, a disconnected interval; and $[CI_{\lambda}] = 10 \leq \theta \leq 15$, a trivial interval. In order to avoid triviality $[CI_{alt.}]$ arbitrarily takes a 'bite' out of the trivial interval given by $[CI_{\lambda}]$, and thereby omits parameter values which it ought reasonably to include. There is nothing in the inference situation which provides a basis for thinking that θ is not in the open interval (10.05, 10.58); indeed the data provides positive reasons for judging these to be plausible values for θ . Hence, it appears that even on the 'backward look' $[CI_{\lambda}]$ is a more sensible interval estimate to infer than is $[CI_{alt.}]$. It is preferable, in this example, to report an interval within which it is known that the parameter must lie, than to report one which wrongly suggests that the parameter value does not lie in a region somewhere in the middle of the interval. Arbitrarily taking a 'bite' out of the interval also leads to counterintuitive results in the tests which correspond to $[CI_{alt.}]$. For example, the observation $X = 10$ would be taken to reject $\theta_0 = 10.06$ but to accept both $\theta_0 = 10.04$ and $\theta_0 = 15$ —a result which is counterintuitive. Even if $C_{alt.}$ is understood as the critical region of a one-sided rather than a two-sided test, something as which it is more plausibly seen (although Seidenfeld does not claim it to be), it still is counterintuitive to accept

⁶One of the 'forward looking' criteria in which Seidenfeld notes $[CI_{alt.}]$ is deficient is unbiasedness. However, if $[CI_{alt.}]$ is construed as a certain one-sided lower confidence interval rather than as a two-sided interval, then it is biased against just those parameter values against which such a one-sided lower interval would be expected to be biased. (See the definition of an unbiased one-sided lower confidence interval in note #5.) If it is construed this way then bias is not really a problem for $[CI_{alt.}]$. It is for this reason that bias is not a problem for the one-sided interval $[CI_0]$.

values of θ_0 both above and below a value of θ_0 which is rejected. Moreover, these examples are typical of the sort of counterintuitive results to which $[CI_{alt.}]$ gives rise.

Since, even on the 'backward look' it appears that $[CI_{alt.}]$ provides less sensible inferences than does $[CI_\lambda]$, Seidenfeld's criterion, under which $[CI_{alt.}]$ is superior to $[CI_\lambda]$, must be rejected as an inadequate or irrelevant criterion for interval estimates.

5. Conclusion. My remarks in defense of NP theory against Seidenfeld's criticism may be summarized as follows: (1) Firstly, I have argued that the occasional production of confidence intervals which for some observations are trivial is not adequate grounds for deeming a theory of confidence intervals inadequate. The possession of additional information limiting the range of a parameter may give rise to trivial intervals, but there is nothing false or misleading about them. What leads Seidenfeld to deem them misleading, and hence grounds for objecting to NP theory, is the supposition that such intervals should be considered as 100% and not as $(1 - \alpha)100\%$ intervals; but this is to interpret illegitimately confidence levels as measures of final precision. Moreover, Seidenfeld has not shown that NP confidence intervals cannot be useful for inductive inference without being illegitimately interpreted. (2) Secondly, I have argued that in the case where the parameter of the uniform distribution is truncated NP theory would recommend, not $[CI_\lambda]$ as Seidenfeld claims, but the one-sided interval $[CI_0]$; and, unlike $[CI_\lambda]$, $[CI_0]$ never yields trivial intervals, and hence is not open to Seidenfeld's criticism. (3) Lastly, I claim that Seidenfeld's criterion of non-triviality is shown to be inadequate for judging statistical inferences when one considers the alternative confidence interval $[CI_{alt.}]$; for on Seidenfeld's criterion $[CI_{alt.}]$ is superior to $[CI_\lambda]$ and yet $[CI_{alt.}]$ yields less sensible inferences than does $[CI_\lambda]$. $[CI_{alt.}]$ avoids triviality only at the price of arbitrarily taking a 'bite' out of the trivial interval, thereby omitting parameter values which ought reasonably to be included.

In conclusion, Seidenfeld has not shown that NP theory is inadequate as a theory for inductive inference as he had claimed to.

REFERENCES

- Hacking, I. (1965), *Logic of Statistical Inference*. Cambridge: Cambridge University Press.
- Kendall, M. G. (1961), *The Advanced Theory of Statistics*, Vol. II. London: Charles Griffin.
- Lehmann, E. L. (1959), *Testing Statistical Hypotheses*. New York: Wiley.
- Mood, A. M., Graybill, F. A., and Boes, D. C. (1950), *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
- Neyman, J. (1937), "Outline of a Theory of Statistical Estimation Based on the Classical

- Theory of Probability", *Philosophical Transactions of the Royal Society of London*, Ser. A, No. 767, 236: 333–380.
- Savage, L. J. (1962), *The Foundations of Statistical Inference*. London: Methuen.
- Seidenfeld, T. (1979), *Philosophical Problems of Statistical Inference*. Dordrecht: Reidel.
- Spielman, S. (1973), "A Refutation of the Neyman-Pearson Theory of Testing", *British Journal for the Philosophy of Science* 24: 201–222.