The Annals of Statistics
1977, Vol. 5, No. 5, 892–898

CONFIDENCE, POSTERIOR PROBABILITY, AND THE BUEHLER EXAMPLE

BY D. A. S. FRASER

University of Toronto

A confidence level is sometimes treated as a probability and accordingly given substantial "confidence." And with certain statistical models a confidence level can also be an objective posterior probability (Fraser and MacKay, 1975). An instance involving such a probability with a binary error was discussed at the Symposium on Foundations of Statistical Inference, University of Waterloo, 1970 (Godambe and Sprott, 1971, page 49). An example by Buehler (ibid., page 337) of a betting strategy and a generalization by Rubin (ibid., page 340) were subsequently presented to support a claim against the objective posterior probability and against too much "confidence." The ordinary Student confidence interval in appropriate contexts is also an objective posterior probability interval. An example by Buehler and Feddersen (1963) has been cited frequently as evidence against the validity of the objective posterior probability and against too much "confidence." This note censures the common procedure of assessment in terms of betting strategies and introduces a modified balanced procedure for such betting assessments. When assessed by this balanced procedure the Buehler and Buehler–Feddersen strategies are faced with large losses and thus do not support claims against the objective posteriors and confidence levels.

1. Introduction. Intermittently, confidence intervals have been the subject of a variety of criticisms. This applies to particular intervals and to the concept itself; see, for example, Welch (1939) concerning the conditional interpretation of an interval for the uniform distribution, and letters to the editor in the Royal Statistical Society, News and Notes, January to October 1976, for pro- and anti-Bayesian comments.

Objective posterior probability intervals are confidence intervals with an additional probability property. These tend to be a subject of criticism from certain statistical extremes: from the Bayesian statisticians for whom statistics is primarily posterior distributions (but only Bayesian posteriors), for example, Lindley in Godambe and Sprott (1971, page 339); or from the classical statisticians for whom posterior distributions are essentially unavailable, for example, Buehler in Godambe and Sprott (1971, page 337).

Objective posterior intervals arise in the context of structural models. A structural model has a class of random variables dependent on the parameter and defined on a fixed probability space (or more generally on a parameter-dependent probability space); other models have a given variable on a space

Received April 1974; revised January 1977.

AMS 1970 *subject classifications*. Primary 60A05; Secondary 62A99.

Key words and phrases. Posterior probability, confidence interval, betting assessment, structural probability, selection, validity of probabilities.

with parameter-dependent probabilities. The range of application for structural models is wide and includes, for example, all the common linear models both with normal and nonnormal distributions. The inference methods available are broadly based and include tests, confidence intervals, likelihood, and marginal likelihood methods, all obtained by standard probability analysis without need for the familiar reduction principles of sufficiency and conditionality.

Objective posterior or structural intervals are obtained as a rather special extension from the basic or core analysis of structural models. These posterior intervals have received more than their share of attention, particularly from the statistical extremes, to the point where for many observers the analysis of structural models is identified with the very special extension involving posterior distributions. This is further complicated by the use by some observers of the term *structural inference* as if there were some special form of inference method with structural models or as if to condemn the general analysis of structural models by association with the sometimes pejorative terms "objective posterior" intervals or "structural" distributions. A recent paper (Fraser, 1976) has examined tests and confidence intervals for the location-scale model and has avoided the special extension to posterior or structural intervals.

An instance of an objective posterior probability with a binary error was discussed at the Symposium on Foundations of Statistical Inference, University of Waterloo, 1970 (Godambe and Sprott, 1971, page 49). An example by Buehler (ibid., page 337) of a betting strategy and a generalization by Rubin (ibid., page 340) were subsequently presented to support a claim against the validity of the objective or structural probability. The Buehler example has been cited more recently (Buehler, 1973) in support of the same claim: "The claim that the experienced gambler would bet in accordance with certain structural probabilities is doubtful in view of an example given by the reviewer (in Godambe and Sprott, 1971, page 340)." In the familiar pattern discussed in the preceding paragraph the reviewer centered his remarks on the example of a structural probability and did not mention the main body of the paper containing various comparisons of inference methods and a theoretical basis for the right Haar prior commonly used by Bayesians. Section 2 shows that in a balanced betting context the Buehler strategy is subject to large losses and thus does not have the properties ascribed to it in support of a claim against the structural probability. For related discussion see Barnard (1971).

The ordinary Student confidence interval in appropriate contexts is also an objective posterior probability interval. An example by Buehler and Feddersen (1963) has been cited frequently as evidence against the validity of the objective posterior probability and against too much confidence. Section 3 shows that in a balanced betting context the Buehler–Feddersen strategy is subject to large losses and thus does not have the properties ascribed to it in support of a claim against the structural probability.

Suppose a scientist S has asserted a certain probability. The betting assessment of asserted probabilities envisages a concerned person C who can make bets with S on the basis of asserted probabilities. The common procedure for betting assessment allows the concerned person C the option of selecting the outcomes to bet on, thus betting in general only on a subset of cases. As part of this it is assumed that the scientist S is obligated to accept all bets from the concerned person C.

The common procedure for assessing asserted probabilities thus allows the concerned person C the luxury of selective betting while the defendant has no such privilege. Clearly we should expect such an unbalanced assessment procedure to lead in due course to unbalanced assessments.

Towards a balanced assessment procedure suppose we allow the concerned person C the option of initiating betting either precipitously or on a selected basis. And then to balance the procedure suppose we allow the scientist S in due course to withdraw from betting either precipitously or on a selected basis. In this paper we use this balanced assessment procedure to examine the Buehler and Buehler–Feddersen betting strategies. The larger implications of balanced betting assessments will be discussed elsewhere.

2. An illustration involving binary error. An illustration of confidence and structural probability involving binary error was discussed at the Symposium on Foundations of Statistical Inference, University of Waterloo, 1970. The essentials of the discussion are available in the author's paper "Events, information processing, and the structured model," together with comments and reply as recorded in the symposium publication, Godambe and Sprott (1971). For estimation based on minimum risk see Blackwell (1951).

In the illustration at the symposium, the author referred to an integer designated θ about which no information was available to the audience. A coin acceptable to the audience was assigned the value $+1$ for heads and the value -1 for tails.

The coin was tossed and the resultant upward face was concealed from all members of the audience; let e designate the value for the upward face. The author asked the audience for the value of $P(e = +1)$ and was given an unanimous reply $\frac{1}{2}$ by the audience.

The author then reported the value of $y = \theta + e$ as $y = 3132$. He then asked the audience for the value of $P(e = +1)$, and was given a general reply $\frac{1}{2}$ but some concern and dissent was present.

The author noted that if $e = +1$, θ is 3131, and if $e = -1$, θ is 3133. He then asked the audience for the value of $P(\theta = 3131)$ and was given a reply $\frac{1}{2}$ by many of those present but some substantial dissent arose particularly from those with strong Bayesian commitments.

The only essential change from the first request to the subsequent requests involved the reporting of the value of $\theta + e$; specifically, whether the reporting of the valued $\theta + e$ alters the information concerning the realized but concealed value e. This question of information was examined in a larger context as the topic of the author's paper at the symposium (Godambe and Sprott, 1971, pages 32–41) and no dissent was registered concerning the content of the paper.

The illustration is based on a structural model: a basic probability space ascribing probability $\frac{1}{2}$ to each of the values for the coin; together with an observable variable that is obtained from the probability space by some function in a class $\Phi = \{\theta + \cdot : \theta \text{ an integer}\}$ of functions on the probability space. As mentioned in Section 1 a variety of inference methods are directly available with such a model. The illustration at the Symposium, however, centered on a rather special extension from the common or basic analysis: specifically, the formation of posterior or structural probabilities or intervals.

The posterior distribution for θ has probability $\frac{1}{2}$ at each of 3131 and 3133. A 50% confidence interval is given by the single point $y - 1$; the observed interval is the single point 3131.

Now consider the Buehler betting strategy in its simplest form (Godambe and Sprott, 1971, page 337). The strategy is to bet that $\theta = y - 1$ when $y > 0$ and not to bet otherwise. If $\theta = 0$ or 1, bets are made half the time and the better wins each bet. If $\theta < 0$, no bets are made. If $\theta > 1$, bets are always made and the better wins half of them on the average. This gives an apparent advantage to the bettor and was presented as an example counter to the validity of the structural probabilities.

The Buehler strategy can also be viewed in relation to the confidence interval: bet in favour of the interval if $y > 0$ and do not bet otherwise. Parallel results are obtained.

The benefits with the simple Buehler strategy occur only if $\theta = 0$ or 1. Discussion at the symposium led to a generalization in which benefits are spread thinly to all points on the range for θ. The generalization by Rubin (Godambe and Sprott, 1971, page 340) is to randomly choose a value a in accord with a discrete probability function p having positive probability at each integer and then to bet that $\theta = y - 1$ when $y > a$ and not to bet otherwise. The expected gain is $(p(\theta) + p(\theta - 1))/2$. This gives an apparent positive advantage to the bettor for all θ values and was presented as a more substantial example.

The Buehler strategy in its simple or generalized form has expectation $\geqq 0$ for all θ values and has positive expectation for some θ value.

Now consider the balanced assessment procedure discussed in Section 1. Specifically the concerned person C has the option of initiating betting either precipitously or on a selected basis; this includes the two Buehler strategies. And the scientist S in due course has the option of withdrawing from betting either precipitously or on a selected basis. As a withdrawal tactic suppose that S declines to accept bets when $y > 2k + 1$. We examine first the simpler Buehler strategy.

For the balanced assessment we examine the betting performance with the following sequence of parameter values, $(\theta_1, \theta_2, \cdots)$ where $\theta_i = 2n$ for $2^{n-1} < i \leqq 2^n$.

On the first trial the bet will be completed with probability $\frac{1}{2}$ and a completed bet is won. On the next $2^k - 1$ all bets will be completed and the expected gain for each is 0. On the next 2^k trials there will be on the average 2^{k-1} completed bets and each completed bet is lost. The expected gain per completed bet is

$$\frac{\frac{1}{2}\cdot 1 + (2^k - 1)\cdot 0 + 2^{k-1}\cdot(-1)}{\frac{1}{2} + 2^k - 1 + 2^{k-1}}$$

which is approximately $-\frac{1}{3}$ for moderate-to-large k; the calculation is based on the average number of completed bets and would adjust slightly for deviations from the average. Thus the Buehler strategy can be faced with large losses and the scientist S cannot be viewed as a money machine by the Buehler strategist.

A similar result is available for the generalized Buehler strategy by examining $\theta_i = 2(n + r)$ for $2^{n-1} < i \leqq 2^n$ with r sufficiently large in relation to the randomization function.

3. The Student confidence interval. For a sample $\mathbf{y} = (y_1, \cdots, y_n)'$ from the normal (μ, σ) distribution, the Student $(1 - \alpha)$ confidence interval for μ is $(\bar{y} \pm t_\alpha s_y/n^{\frac{1}{2}})$ where $(-t_\alpha, t_\alpha)$ is a $1 - \alpha$ interval for the Student $(n - 1)$ distribution. For a parallel structural model let $\mathbf{z} = (z_1, \cdots, z_n)'$ be a sample from the standard normal and $\mathbf{y} = (y_1, \cdots, y_n)'$ be some function $\mu\mathbf{1} + \sigma\mathbf{z}$ defined on the standard normal probability space; the confidence interval $(\bar{y} \pm t_\alpha s_y/n^{\frac{1}{2}})$ is also a $1 - \alpha$ structural interval for μ. For a discussion of standard tests and confidence intervals for this location scale model but with parameter dependent distribution shape see Fraser (1976).

Buehler and Feddersen (1963) examine the confidence interval for $n = 2$; a simple 50% interval is $(\min y_i, \max y_i)$. They consider a bettor who selects on the basis of $|\bar{y}|/s_y \leqq$ constant or equivalently on the basis of $|y_2 - y_1| \geqq c|y_1 + y_2|$ with $0 < c < 1$. They show that the probability of covering the parameter μ is at least 51.81% for the selected outcomes.

Now consider the balanced assessment procedure as discussed in Section 1. Specifically consider the concerned person C using the Buehler–Feddersen strategy and consider the scientist with a withdrawal strategy of declining bets when $s_y > k$, a specified constant.

For the balanced assessment we examine the betting performance for a sequence of parameter values $(\theta_1, \theta_2, \cdots)$ where $|\theta_i| > k(1 + 1/c)/2^{\frac{1}{2}}$. The concerned person C tenders a bet if $|y_2 - y_1| \geqq c|y_1 + y_2|$ and the scientist S accepts the bet if $|y_2 - y_1| \leqq (2k)^{\frac{1}{2}}$; the resulting confidence interval $(\min y_i, \max y_i)$ is contained in the interval $(\pm k(1 + 1/c)/2^{\frac{1}{2}})$. Thus the probability of covering the parameter is zero, and expected gain per completed bet is -1. Thus the Buehler–Feddersen strategy can be faced with large losses and it cannot be viewed as an invalidation of the confidence-structural level for the Student interval.

4. Conclusions. The binary illustration involved errors that were ± 1 with equal probability. The simple Buehler strategy is to bet that θ is towards zero if $y > 0$. A generalized strategy is to bet inwards based on a randomly located center. This strategy is successful against a *distribution* for θ; for in such cases the true θ is *inward* from the extremes of the real line.

The scheme of considering single values for θ or a distribution for θ seems on the surface to be fully general. In this paper, however, we have examined diverging sequences of parameter values; such a sequence of course does not represent a distribution although in a loose sense it is a distribution about ∞.

A balanced assessment involving the Buehler strategy in this larger context shows that the strategy can be a large money loser. The strategy thus does not have the uniform winning properties ascribed to it and thus does not reflect against the confidence-structural probability interval.

The second illustration examined the common Student confidence interval. The Buehler–Feddersen strategy is to select instances having a *relatively* large standard deviation.

A balanced assessment involving the Buehler-strategy and the scientist accepting bets based on small values of the standard deviation shows that the strategy can be a money loser. The strategy does not have the uniform winning properties ascribed to it and thus does not reflect against the Student interval.

5. Addendum.[1] I agree that it is the fairness of the comparison between confidence-objective probabilities and ordinary probabilities that should concern us primarily.

In the discussion, "behaving like ordinary probability" turns into "assessing in the context of ordinary probability"—a sequence of repetitions on a physical system with a single θ input or a *distribution* for θ input. But this does *not* cover possibilities such as diverging sequences of θ input values. For assessment in the simpler context all that one needs is subsequences; the balance has nothing to add.

The betting context is used to establish a norm between two players for assessing probabilities. If this context is designed to protect one player against certain eventualities (the diverging sequence), then the protected player has an advantage. Thus the proposal for fair comparisons and balanced assessments.

REFERENCES

BARNARD, G. A. (1971). Discussion of the paper by Godambe and Thompson. *J. Roy. Statist. Soc. Ser. B* **33** 376–378.

BLACKWELL, DAVID (1951). On the translation problem for discrete variables. *Ann. Math. Statist.* **22** 393–399.

BUEHLER, R. J. and FEDDERSEN, A. P. (1963). Note on a conditional property of Student's t. *Ann. Math. Statist.* **34** 1098–1100.

BUEHLER, R. J. (1973). *Math. Reviews* 6514, **46** 1120.

FRASER, D. A. S. and MACKAY, JOCK (1975). Parameter factorization and inference based on significance, likelihood, and objective posterior. *Ann. Statist.* **3** 559–572.

FRASER, D. A. S. (1976). Necessary analysis and adaptive inference. *J. Amer. Statist. Assoc.* **71** 99–113.

[1] Refers to discussion immediately following.

GODAMBE, V. P. and SPROTT, D. A. (1971). *Foundations of Statistical Inference.* Holt, Rinehart and Winston, Toronto.

WELCH, B. L. (1939). On confidence limits and sufficiency with special reference to parameters of location. *Ann. Math. Statist.* **10** 58-69.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
TORONTO, ONTARIO M5S 1A1
CANADA

DISCUSSION

ROBERT J. BUEHLER

University of Minesota

The point at issue is the "validity" of objective posterior probabilities. Perhaps our disagreements are largely semantic since "validity" is defined neither in my 1970/1971 example nor in the present paper. I intended validity to mean "behaving like ordinary probability," where ordinary probability means direct frequency probability exhibiting randomness in the von Mises sense.

It is not fairness of the game between S and C that should concern us primarily, but rather fairness of the comparison between objective posterior probabilities and ordinary probabilities. For the latter S survives any strategy of C (essentially a von Mises axiom); for the former he does not. Thus objective posterior probabilities behave differently from ordinary probabilities.

It is not surprising that in the "balanced assessment procedure" the scientist S is seen to win. Since this would not have happened with ordinary probabilities, the example in no way refutes the claim of nonvalidity, but in fact strengthens it.

I readily concede that the experienced gambler would bet under the balanced assessment procedure. I continue to doubt that he would bet if the opponent alone had the option of accepting or declining. Conceivably he would do so if operating on a finitely additive prior assigning zero probability to every θ value, but then any posterior assignment would serve as well as the objective posterior (Buehler, 1976).

REFERENCE

BUEHLER, ROBERT J. (1976). Coherent preferences. *Ann. Statist.* **4** 1051-1064.