Two Papers on the Comparison of Bayesian and Frequentist Approaches to Statistical Problems of Prediction: Bayesian Tolerance Regions

Author(s): J. Aitchison

Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, 1964, Vol. 26, No. 2 (1964), pp. 161-175

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2984416

Two Papers on the Comparison of Bayesian and Frequentist Approaches to Statistical Problems of Prediction

[Read at a RESEARCH METHODS MEETING of the SOCIETY, February 5th, 1964, Professor D. V. LINDLEY in the Chair]

Bayesian Tolerance Regions

By J. AITCHISON

University of Liverpool

SUMMARY

In the theory of statistical tolerance regions, as usually presented in frequentist terms, there are inherent difficulties of formulation, development and interpretation. The present paper re-examines the basic problem from a Bayesian point of view and suggests that such an approach provides a set of widely applicable, mathematically tractable tools, often more tailored to the requirements of users than the corresponding frequentist tools. For the one-dimensional case, Bayesian intervals are quoted for a number of standard distributions and prior densities, and the customary feature of a Bayesian analysis—that special prior densities give rise to standard frequentist results—is briefly demonstrated. A problem which seems to be of greater practical significance, namely the selection of an optimum tolerance region from a set of possible tolerance regions, is also investigated and the overwhelming advantages of the Bayesian approach are indicated.

1. INTRODUCTION

IN this paper the term *Bayesian* refers to any use or user of prior densities on a parameter space, with the associated application of Bayes's theorem, in the analysis of a statistical problem. The term *frequentist* applies to any analysis or analyst of the "objectivist" school, where the use of prior densities is denied and where there is a tendency to interpret probability solely in terms of relative frequencies in large-scale replication. Typical contributions by frequentists on the problems considered here are Wald and Wolfowitz (1946), Bowker (1947), Wallis (1951) and Fraser (1953).

Bayesian revivalists seem so far to have neglected a promising missionary field in the heart of frequentist territory—the field of statistical tolerance regions. Savage (1954, Section 17.3) briefly dismisses the whole concept of tolerance interval as a slippery one, unamenable to behaviouralistic interpretation. Although Raiffa and Schlaifer (1961, Example 6.1.2) introduce their Bayesian treatment of estimation by a simple illustration, which falls into the general class of tolerance-region problems as defined in Section 3 of this paper, their situation does not belong to the special subclass of these problems which is usually understood to constitute the subject matter of tolerance-region theory. Welch and Peers (1963) obtain mainly asymptotic results on conditions under which frequentist and essentially Bayesian upper (or lower) confidence limits for a real parameter θ are identical. Their approach, however, is staunchly frequentist and they repeatedly assert that their weights are not to be regarded as prior probability densities. Since an upper tolerance limit is merely an

upper confidence limit for a quantile of some fixed order it would be possible, at least for cases where quantiles of fixed order are monotonic functions of θ, to derive from their results asymptotic forms of prior density functions which yield frequentist-type limits. While these forms might provide, for such cases, a means of measuring the agreement or disagreement between frequentist and Bayesian analyses they appear to be of limited practical value in the situations treated here.

It is the purpose of this paper to initiate a mildly evangelical campaign among workers in the field of tolerance regions by demonstrating that the Bayesian approach has much to offer in theoretical and practical terms. Since conversion to a new way of thinking often arises out of self-criticism, one of our early tasks is to undertake a critical re-examination of the probabilistic statement which forms the basis of the frequentist structure of tolerance-region theory. This statement is a complicated one, involving many mathematical difficulties and with a frequentist interpretation often misunderstood—in the author's experience—by users of tolerance regions. Bayesian formulations of the problem are mathematically simpler and appear to be much closer to the spirit in which tolerance regions are used in many practical situations. In standard situations the Bayesian, by persuading himself that his prior densities have a special form, may act as if he were a frequentist, though his probabilistic interpretation will differ from the frequentist. We shall note this aspect in our study of standard univariate distributions where one-dimensional intervals are required. When the problem is one of selecting a region from a set of possible tolerance regions on certain optimality criteria—and this, it is suggested, is the type of problem of practical importance—we shall see that Bayesian theory stands up to the task well, whereas it seems difficult even to formulate the problem sensibly in frequentist form.

In defining the experimental setting of the tolerance-region problem we must be careful about notation, since this is crucial to the development and interpretation of the various probabilistic statements we encounter. The notation adopted here is substantially that of Lindley (1961). We suppose that the uncertainty features of a random experiment $\mathscr{E}$, with outcome space $\mathscr{X}$ and event space $\mathscr{B}$, are described by a probability measure on $\mathscr{B}$, say $P(.\,|\,\theta)$ belonging to a parametric family

$$\{P(.\,|\,\theta);\ \theta \in \Theta\},$$

where Θ is the parameter space. We suppose that each measure $P(.\,|\,\theta)$ admits a density function $p(.\,|\,\theta)$ on $\mathscr{X}$. We do not know which particular parameter value θ_0 is the true one. There is available a set $x^{(n)} = (x_1, \ldots, x_n)$ of outcomes of n (independent) replicates of $\mathscr{E}$; we denote the corresponding density on $\mathscr{X}^n$ by $p^{(n)}(x^{(n)}\,|\,\theta)$ and the associated product probability measure by $P^{(n)}(.\,|\,\theta)$. These are the immediate definitions needed by the frequentist. The Bayesian requires the further concept of a prior probability density $\pi(\theta)$ on Θ; this prior density represents either his prior beliefs, if these are ascertainable or a convenient, sensible form of weighting over the possible parameters. The consequence of observing $x^{(n)}$ is to alter, by way of Bayes's theorem, this prior density into a posterior density $\pi(\theta\,|\,x^{(n)})$ over Θ; the correponding probability measure (over the Borel σ-field defined on Θ) we denote by $\Pi(.\,|\,x^{(n)})$. The Bayesian then decides his course of action solely on this posterior distribution and on the consequences of his possible decisions.

We begin our discussion with an examination in Section 2 of the "simple" version of the tolerance-region problem as usually presented by frequentist writers. The problem is to choose some region $R \in \mathscr{B}$, for which it can be reasonably claimed that a proportion c of outcomes of future replicates of $\mathscr{E}$ will fall in R. If θ_0 were known

the tolerance-region user would select a region $R \in \mathscr{B}$ with the property that $P(R \mid \theta_0) = c$. When R has this property we say that R has *cover* c at θ_0. The selection of a suitable R would then usually pose a relatively simple mathematical problem. It is uncertainty about θ which forces some compromise solution on the user. His experimental data $x^{(n)}$ remove some of this uncertainty about θ. The user thus seeks a region $R(x^{(n)}) \in \mathscr{B}$, which he hopes will provide cover at least c; that is, he hopes that he can say with some "confidence" that $P\{R(x^{(n)}) \mid \theta\} \geqslant c$. He would then be reasonably sure that at least a proportion c of future outcomes of $\mathscr{E}$ fall within $R(x^{(n)})$. It is in the probabilistic formulation and interpretation of the term "reasonably sure" that the frequentist and Bayesian begin to differ.

2. The Frequentist Analysis

The basic statement on which the frequentist builds his theory involves the family of probability measures $\{P^{(n)}(. \mid \theta): \theta \in \Theta\}$ associated with the n-replicate outcome space $\mathscr{X}^n$. Let R be a measurable function or statistic, with domain $\mathscr{X}^n$ and range $\mathscr{B}$ and providing, for each $x^{(n)}$, a subset $R(x^{(n)})$ of $\mathscr{X}$. For given θ, cover c and statistic R, denote by $G(\theta, c, R)$ the set of all $x^{(n)}$ which yield, through R, regions of cover at least c at θ, so that

$$G(\theta, c, R) = [x^{(n)} : P\{R(x^{(n)}) \mid \theta\} \geqslant c]. \tag{1}$$

The frequentist says that we have a satisfactory tolerance statistic R if each probability measure of this set is some pre-assigned value q near unity; that is, if

$$P^{(n)}\{G(\theta, c, R) \mid \theta\} = q \tag{2}$$

for *every* $\theta \in \Theta$.

The popular frequentist view of probability—as the counterpart, in a mathematical model of an experiment, of stable relative frequency in a large number of replicates of the experiment—allows the following interpretation. If we repeat the n-replicate experiment a large number of times, each time obtaining some observation $x^{(n)}$ and each time constructing, through R, a tolerance region $R(x^{(n)})$, then a proportion q of these regions will be satisfactory in providing cover at least c; for this interpretation see, for example, Fraser (1957, p. 116) and Weissberg and Beatty (1960). While this statement is comforting to a statistician making repeated use of R to provide his customers with tolerance regions, its meaning for the individual customer is not at all easy to specify. The author has recently experienced the difficulty of trying to sell the frequentist approach to engineers whose work is crucially concerned with problems of tolerance regions. While there was no difficulty in their understanding of the cover c, the confidence coefficient or *quality* q was a much more elusive concept. This is not surprising, for in many applications there is no reality in the idea of repeated replication of the n-replicate experiment. The frequentist might then argue that the probabilistic statement, while not of direct application to the engineer's needs, is intended to give him comfort through the hope that his particular experiment is one of the lucky ones belonging to this proportion q in a population of hypothetical experiments. This seems to be little removed from admitting two different interpretations for probability—the forced acceptance of a kind of degree-of-belief interpretation for $P^{(n)}$ while retaining the relative frequency interpretation for P. It is this confusion of interpretations, none of which quite fits the real problem, which causes misunderstandings with users.

This confusion can be illustrated by a simple example. Suppose that it is required to specify what fixed daily amount of a perishable commodity it is necessary to supply in order to be "reasonably sure" that the probability of supply meeting demand on any one day is c. We may then decide, on the basis of $x^{(n)}$, daily observations (supposed independent) of demand over a month say, to quote a value $r(x^{(n)})$, where $R(x^{(n)}) = \{-\infty, r(x^{(n)})\}$ is a tolerance interval of cover c and quality q. If, for each of 100 months' observations, we construct a supply system with design values based on the function $r(.)$, then the frequentist interpretation is that, roughly speaking, for a proportion q of these systems will the supply be satisfactory in that the probability of supply meeting demand on any day is at least c. Unfortunately, what is sometimes thought to be the interpretation by the user is that, if he observes one month's demands, quotes a supply value from this and then constructs 100 actual systems based on this one supply value, a proportion q of these systems will function satisfactorily as regards cover. This interpretation is clearly wrong; either all are satisfactory or all are not. The user is, however, so confused by the complexity of the probabilistic statement and the insistence on frequentist interpretations that he falls easily into these misunderstandings.

This is, however, not the only difficulty surrounding the statement (2). The probability $P^{(n)}\{G(\theta, c, R) \mid \theta\}$ depends essentially on θ in the general formulation, and the true value θ_0 of the parameter is unknown. Yet the frequentist determines by his probabilistic statement that this probability is to be q, independent of θ. Two escape routes, other than the pursuit of distribution-free regions which reflects a complete change of attitude, are open to avoid this entanglement with θ. He can alter his statement (2) so that the left-hand side does not depend on θ, though his problem of determining a whole function will remain. He can achieve this by requiring R to satisfy

$$\inf_{\theta \in \Theta} P^{(n)}\{G(\theta, c, R) \mid \theta\} = q, \tag{3}$$

as in Fraser (1957, Definition 5.1). This essentially minimax procedure can place undue emphasis on particular values of θ with possibly awkward effects (Lehmann, 1959, p. 13). Alternatively he could introduce a weighting factor $\pi(\theta)$ associated with each θ and choose R so that

$$\int_{\Theta} d\theta\, \pi(\theta) P^{(n)}\{G(\theta, c, R) \mid \theta\} = q, \tag{4}$$

thus producing "Bayes solutions". In either case it will be observed that he is agreeing to place more emphasis on some values of θ than others and so is moving towards a Bayesian outlook. It is not surprising, therefore, that he tends to avoid this escape route and concentrate on the second.

The second approach is to attempt to choose the function R in such a way that $P^{(n)}\{G(\theta, c, R) \mid \theta\}$ is indeed independent of θ. The frequentist thus directs his attention to the search for such a pivotal statistic. For example, in the well-known case where $p(x \mid \theta)$ is a $N(\mu, \sigma^2)$ density, he uses $R(x^{(n)}) = (\bar{x} + k_1 s, \bar{x} + k_2 s)$, where $\bar{x}$ and s have their usual meanings and k_1 and k_2 are constants, and so achieves a disengagement from $\theta = (\mu, \sigma)$. While we would not seek to deny the reasonableness of this choice—based presumably on the fact that $(\bar{x}, s)$ is minimally sufficient for (μ, σ) and on an attempt to reflect, in the linear expressions $\bar{x} + k_1 s$ and $\bar{x} + k_2 s$, the linearity of percentiles in μ and σ—no case seems to be made out by the frequentist that there is just one such pivotal function of the minimal statistic or, alternatively, that the one chosen

is in some sense the best of such pivotal functions. Moreover, in cases where no minimal sufficient statistic exists, it is difficult to see on what principle, other than that of expediency, the frequentist could choose his pivotal statistic.

We shall see later, in Section 5, that the frequentist's difficulties in the three aspects of formulation, interpretation and mathematical development grow when he is faced with a situation, where there are certainly many possible R which satisfy his requirements and he has to incorporate some additional optimality criterion in his choice of R.

We can express the frequentist approach in terms which allow an easier comparison with the Bayesian formulations. For we can assign a utility $U(R, \theta)$ to the choice of a tolerance region R when the parameter value is θ in the following way:

$$U(R, \theta) = \begin{cases} 1 & \text{if} \quad P(R \mid \theta) \geqslant c, \\ 0 & \text{if} \quad P(R \mid \theta) < c. \end{cases} \tag{5}$$

This is simply an expression of the negative of a loss function in frequentist decision theory. The frequentist then wishes to choose his tolerance-region statistic $R(.)$ in such a way that the expectation, with respect to the $P^{(n)}(. \mid \theta)$ measure, of this utility is q. This is so since $U\{R(.), \theta\}$ is merely the indicator function of the set $G(\theta, c, R)$ and so

$$\int_{\mathscr{X}^n} dx^{(n)}\, U\{R(x^{(n)}), \theta\} p^{(n)}(x^{(n)} \mid \theta) = P^{(n)}\{G(\theta, c, R)\}.$$

3. The Bayesian Formulations

A number of statisticians—in particular, Lindley (1961), Raiffa and Schlaifer (1961) and Savage (1962)—have vigorously advocated a rethinking of statistical theory and practice in Bayesian terms. I have no wish here to enter into the general controversy over the relative merits of frequentist and Bayesian statistical analyses. My intention is rather to point out that the Bayesian approach must be regarded as a serious alternative to, or at least a useful complement of, the frequentist one in tolerance-region problems.

The Bayesian bases his action on two components—first, his posterior measure $\Pi(. \mid x^{(n)})$ and secondly, his *utility function* U, a real-valued function defined on $\mathscr{B} \times \Theta$ with typical value $U(R, \theta)$, the utility of choosing tolerance region R when the parameter value is θ. His action is to choose R so as to maximize his expected utility, the expectation being taken with respect to the $\Pi(. \mid x^{(n)})$ measure; that is, he maximizes

$$W(R, x^{(n)}) = \int_{\Theta} d\theta\, U(R, \theta)\, \pi(\theta \mid x^{(n)}) \tag{6}$$

with respect to R.

Stated in these general terms, a tolerance-region problem can be described as a decision problem in which *the decision space is the whole or part of the event space* $\mathscr{B}$, the decision function or tolerance-region statistic being a measurable function with domain $\mathscr{X}^n$ and range $\mathscr{B}$. Underlying the selection of a tolerance region $R \in \mathscr{B}$ there is usually the suggestion that R will contain a high proportion of the outcomes of future replicates of $\mathscr{E}$ or most of the "important" outcomes of future replicates. The choice of region should thus depend on our assessment of the advantage or disadvantage associated with R in relation to each possible outcome x. This dependence can

in fact be given full expression in our choice of utility function U. For suppose that, corresponding to each x, it is possible to assign quite a specific advantage or *value* $V(R, x)$ attaching to R. Then we would wish to choose R so as to maximize the expectation of $V(R, .)$ with respect to the marginal distribution of x, assessed in the knowledge that we have observed the outcome $x^{(n)}$ in the n-replicate experiment. This marginal distribution has density

$$g(x \mid x^{(n)}) = \int_{\Theta} d\theta\, p(x \mid \theta)\, \pi(\theta \mid x^{(n)}) \quad (x \in \mathscr{X}). \tag{7}$$

Thus we require to maximize

$$\int_{\mathscr{X}} dx\, V(R, x)\, g(x \mid x^{(n)}) = \int_{\Theta} d\theta\, \pi(\theta \mid x^{(n)}) \int_{\mathscr{X}} dx\, V(R, x)\, p(x \mid \theta) \tag{8}$$

with respect to R, and so we arrive at the maximization of (6), taking

$$U(R, \theta) = \int_{\mathscr{X}} dx\, V(R, x)\, p(x \mid \theta) \tag{9}$$

as the utility derived from the value function V. The utility specification therefore leads to as wide a class of problems as the value specification.

The problem treated by Raiffa and Schlaifer (1961) in their Example 6.1.2 has a V specification. If the daily demand of a perishable commodity falls outside the chosen tolerance interval $(-\infty, r)$ the supplier loses the potential sale of $x-r$ units, say at a loss $a(x-r)$; if demand falls inside the interval then $r-x$ units are lost through deterioration at a loss $b(r-x)$, say. Thus they take

$$V(R, x) = \begin{cases} -a(x-r) & (x \geqslant r), \\ -b(r-x) & (x < r). \end{cases} \tag{10}$$

While a detailed specification of V and U may be possible in many problems there are cases where $V(R, x)$ depends only on whether x falls inside or outside R and not on the actual value of x. For example, the main elements of concern to an engineer, who chooses the design of an electrical supply system so that only those demands x which lie in R are met, may be the capital cost of R and the possibility of the system failing through demand exceeding supply, with the consequent costly repair to the system and his prestige as designer; see, for instance, the example of Section 5. Another problem of demand-and-supply type is the choice of the height of the lower deck of a double-decker bus. The supply value is the height r provided and a typical demand is the height x of a standing passenger. All passengers with $x < r$ are accommodated comfortably and it is difficult to apportion different degrees of discomfort for passengers with $x \geqslant r$. Again, if a decision is to be taken, not by a single Bayesian but by a group of Bayesians each with his own utility and prior density functions, it may be necessary for the group, as a compromise measure, to adopt a V specification of the type just discussed, namely

$$V(R, x) = \begin{cases} B(R) & (x \notin R), \\ C(R) & (x \in R). \end{cases} \tag{11}$$

This leads, by (9), to the equivalent U specification

$$U(R, \theta) = A(R)P(R|\theta) + B(R), \tag{12}$$

where $A(R) = C(R) - B(R)$.

The important characteristic of the utility function defined by (12) is that it depends on θ only through the cover $P(R|\theta)$ provided by R at θ. Any utility function displaying this characteristic will be said to lead to a *restricted* tolerance region problem. It should be observed that while (12) gives the most general form of restricted U arising from a V specification there are other forms of restricted U, for example, that given by (5). The restricted form of the problem is the one traditionally associated with the branch of statistics known as tolerance-region theory.

The mathematical problem associated with either the general or restricted formulation above is, relative to the frequentist approach, the simple task of maximizing a function. The technique of maximization will depend largely on the form of U or V and, since the choice of this in any particular application requires specialized knowledge, we shall not pursue this essentially computational aspect further. Later, in Section 5, we shall study some interesting aspects of the restricted problem and its frequentist counterpart. In the remainder of this Section we return to a comparison with the frequentist treatment of Section 2.

The restricted Bayesian approach is not practicable with the utility given by (5) since clearly the maximum utility of 1 occurs when $R = \mathscr{X}$. If we therefore compromise by deciding to accept an expected utility of q near 1, so that we choose R to satisfy

$$W(R, x^{(n)}) = q, \tag{13}$$

then we obtain the Bayesian formulation closest to the frequentist. A useful alternative view of this formulation is provided in the following way. For given cover c and region R, denote by $H(c, R)$ the subset of Θ consisting of parameter values at which R gives cover at least c; that is,

$$H(c, R) = \{\theta : P(R|\theta) \geqslant c\}. \tag{14}$$

Then, since $U(R, .)$ is the indicator function of the set $H(c, R)$, relation (13) states that R is to be chosen so that

$$Q(R) = \Pi\{H(c, R)|x^{(n)}\} = q. \tag{15}$$

Our Bayesian rationale may then be expressed as follows. If we knew the true parameter value θ_0 we would have no difficulty in deciding on an appropriate R with cover c at θ_0. Since we do not know θ_0 we must make our choice taking account of our opinion or knowledge at the time of choice. After we have experimented and observed $x^{(n)}$ this is provided by the measure $\Pi(.|x^{(n)})$ associated with Θ. By choosing R so that $H(c, R)$ has $\Pi(.|x^{(n)})$ measure q we are thus, as Bayesians, saying that we are strongly of the opinion that the unknown parameter value is such that R provides at it cover at least c; the strength of the opinion is measured by the quality $q = Q(R)$. This seems a reasonable resolution of the problem and is, we believe, the kind of reasoning which takes place in the minds of many users of tolerance regions. We shall see in Section 4 that the formulation based on (15), with a choice of prior density which in a sense describes a display of ignorance about the parameter, leads to exactly the intervals quoted by the frequentist school.

The customary advantage of the Bayesian approach can now be seen. It substitutes for the choice of a whole *function* R defined on $\mathscr{X}^n$ in the face of entanglement with the parameter θ, the choice of a *region* R for the particular $x^{(n)}$ observed, free from any complication with unknown parameter values. The price that has to be paid for this great increase in mathematical tractability is the admission into the argument of prior densities. Now such an admission may be anathema to the frequentist, but in tolerance-region problems it is how I believe many decision-makers act. An engineer, faced with the design of a supply system as in Section 2, will, if he cannot experiment, behave as a rather pessimistic Bayesian and arrive at his design value by some weighing up of the likely values of θ. Indeed, this is the kind of information which is sometimes laid down for him in his guide books. In recent conversation with some engineers the view was put that it was widely recognized that the prior weights deducible from these books were often ridiculously pessimistic and based on the flimsiest of evidence, but that engineers acted by the book because this was their safeguard in law. Perhaps one method of breaking away from this situation is for the progressive engineer to use this quoted information as a prior density function, to carry out any possible experiments of observing demand and to base his design on a Bayesian tolerance region. In the unlikely event of his being called to account for his choice of design it would be up to the Bayesian to appear as expert witness for the defence. It is indeed difficult to see how else it would be sensible to combine the guide-book information and the experimental evidence other than through a Bayesian analysis.

4. Bayesian Tolerance Intervals

4.1. *A General Result*

We now investigate the case where $\mathscr{X}$ is all or part of the real line. Our main purpose is to derive, for some standard distributions with associated reasonably rich families of prior densities, Bayesian tolerance intervals as defined by (15); and further, to show their relationship to the corresponding frequentist intervals. An excellent account of criteria for the choice of suitable "conjugate" prior densities, together with the main properties of these families in relation to the experimental density, is given by Raiffa and Schlaifer (1961, Chapter 3) and there is no need to reproduce their advocacy here. We catalogue our results by way of the experimental density.

Before we consider these special densities, however, we can derive a simple result for determining an upper tolerance limit $r = r(x^{(n)})$ for the case where θ is a real parameter. Let us define by $d_c(\theta)$ the c probability point of the $p(x|\theta)$ distribution, so that

$$\int_{-\infty}^{d_c(\theta)} dx\, p(x|\theta) = c \tag{16}$$

and let $\delta_c(x^{(n)})$ be the c probability point of the $\pi(\theta|x^{(n)})$ distribution. We then have that

$$\Pi\{\theta : \delta_{1-q}(x^{(n)}) \leqslant \theta | x^{(n)}\} = q,$$

so that, if $d_c(\theta)$ decreases as θ increases,

$$\Pi[\theta : d_c\{\delta_{1-q}(x^{(n)})\} \geqslant d_c(\theta) | x^{(n)}] = q.$$

Thus

$$r(x^{(n)}) = d_c\{\delta_{1-q}(x^{(n)})\} \tag{17}$$

gives an upper tolerance limit of cover c and quality q. The corresponding limit when $d_c(\theta)$ increases as θ increases is

$$r(x^{(n)}) = d_c\{\delta_q(x^{(n)})\}. \tag{18}$$

4.2. *Upper Tolerance Limits for the Gamma Distribution*

Here

$$p(x|\theta) = \theta^k x^{k-1} e^{-\theta x}/\Gamma(k) \quad (x>0), \tag{19}$$

so that

$$d_c(\theta) = 2\chi^2(2k;\, c)/\theta, \tag{20}$$

where $\chi^2(2k;\, c)$ is the c probability point of a $\chi^2(2k)$ distribution. For this case $d_c(\theta)$ decreases as θ increases and so the limit (17) applies. We use here a gamma prior density

$$\pi(\theta) = b^a\, \theta^{a-1} e^{-b\theta}/\Gamma(a) \quad (\theta>0), \tag{21}$$

which results in a gamma posterior density

$$\pi(\theta|x^{(n)}) = (b+z)^{a+nk}\, \theta^{a+nk-1} e^{-(b+z)\theta}/\Gamma(a+nk) \quad (\theta>0),$$

where $z = x_1+\ldots+x_n$. Hence

$$\delta_{1-q}(x^{(n)}) = 2\chi^2(2a+2nk;\, 1-q)/(b+z)$$

and so an upper tolerance limit is

$$r(x^{(n)}) = d_c\{\delta_{1-q}(x^{(n)})\} = (b+z)\,\chi^2(2k;\, c)/\chi^2(2a+2nk;\, 1-q). \tag{22}$$

Lower tolerance limits may be similarly obtained. The corresponding frequentist limit, based on the sufficiency of z for θ, can be shown to be

$$z\chi^2(2k, c)/\chi^2(2nk;\, 1-q). \tag{23}$$

Note that this corresponds to the case $a = b = 0$ of the Bayesian result. This has the appearance of being very reasonable, for $E(\theta) = a/b$ and $\mathrm{var}(\theta) = a/b^2$ and, letting $a\to 0, b\to 0$ so that $a/b\to\mu$, we have a limiting density with finite mean and infinite variance, which signifies considerable ignorance about θ. For the dangers of such interpretations of vagueness about parameters, however, see Raiffa and Schlaifer (1961, Section 3.3.4).

4.3. *Upper Tolerance Limits for the Normal Distribution*

For this case $\theta = (\mu, \sigma)$, $\Theta = \{\theta:\, \sigma>0\}$ and $p(x|\theta)$ is the $N(\mu, \sigma^2)$ density. We define $\bar{x}$ and s in the usual way by $n\bar{x} = \Sigma x_i$ and $(n-1)s^2 = \Sigma(x_i-\bar{x})^2$. The conjugate prior density is of normal-gamma type with

$$\pi(\mu, \sigma) \propto (1/\sigma)\exp\{-\tfrac{1}{2}b(\mu-a)^2/\sigma^2\}(1/\sigma)^w \exp(-\tfrac{1}{2}wv/\sigma^2); \tag{24}$$

we then say that $\pi(\mu, \sigma)$ is a $N\Gamma(a, b, v, w)$ density. It follows very simply that $\pi(\mu, \sigma|x^{(n)})$ is a $N\Gamma(A, B, V, W)$ density, where

$$B = b+n, \quad A = (ba+n\bar{x})/B, \tag{25}$$

$$W = \begin{cases} w+n & (b>0), \\ w+n-1 & (b=0), \end{cases} \tag{26}$$

$$V = \{wv+ba^2+(n-1)s^2+n\bar{x}^2-BA^2\}/W. \tag{27}$$

Since $d_c(\theta) = \mu + \nu_c \sigma$, where ν_c is the c probability point of the $N(0, 1)$ distribution, we require by (15) to choose r so that

$$\Pi\{(\mu, \sigma) : \mu + \nu_c \sigma \leqslant r \,|\, x^{(n)}\} = q. \tag{28}$$

If we introduce new parameters (ξ, η) by

$$\xi = B^{\frac{1}{2}}(\mu - A)/\sigma, \quad \eta = V/\sigma, \tag{29}$$

then the inequality in (22) becomes

$$(\xi + \nu_c B^{\frac{1}{2}})/\eta < B^{\frac{1}{2}}(r - A)/V. \tag{30}$$

Now the posterior distribution of (ξ, η) is such that ξ and η are independent with ξ distributed as $N(0, 1)$ and η as $\{\chi^2(W)/W\}^{\frac{1}{2}}$, and so $(\xi + \nu_c B^{\frac{1}{2}})/\eta$ has a non-central t-distribution with W degrees of freedom and non-centrality parameter $\nu_c B^{\frac{1}{2}}$. If $t(W, \nu_c B^{\frac{1}{2}}; q)$ denotes the q probability point of such a non-central t-distribution then it follows from (30) that

$$B^{\frac{1}{2}}(r - A)/V = t(W, \nu_c B^{\frac{1}{2}}; q)$$

and so

$$r = A + VB^{-\frac{1}{2}} t(W, \nu_c B^{\frac{1}{2}}; q). \tag{31}$$

When $b = w = 0$ we have by (25), (26) and (27) that

$$r = \bar{x} + sn^{-\frac{1}{2}} t(n - 1, \nu_c n^{\frac{1}{2}}; q), \tag{32}$$

which is the familiar frequentist limit. The improper prior density associated with $b = w = 0$ is of Jeffreys type and is usually interpreted as a display of ignorance about the parameters μ and σ; see Jeffreys (1961, Section 3.4.1).

4.4. *Upper Tolerance Limits for Poisson and Binomial Distributions*

For discrete distributions over non-negative integers we define $d_c(\theta)$ as the smallest integer satisfying the inequality

$$\sum_{x=0}^{d_c(\theta)} p(x \,|\, \theta) \geqslant c, \tag{33}$$

and the Bayesian upper tolerance limit $r(x^{(n)})$ is defined by

$$\Pi\{\theta : d_c(\theta) \leqslant r \,|\, x^{(n)}\} = q. \tag{34}$$

For the Poisson experimental density

$$p(x \,|\, \theta) = e^{-\theta} \theta^x / x! \quad (x = 0, 1, 2, \ldots) \tag{35}$$

with gamma prior density (21) we obtain a gamma posterior density with

$$\delta_q(x^{(n)}) = 2\chi^2(2a + 2z; q)/(b + n). \tag{36}$$

Since $d_c(\theta)$, which is easily obtainable from Poisson tables, increases with θ, the limit $r(x^{(n)})$ is readily obtained by (18). Again, as in Section 4.2, the case $a = b = 0$ produces a frequentist tolerance limit.

A similar type of analysis can be applied to the binomial experimental density with a beta prior density. For difficulties in the interpretation of the case which leads to frequentist results see Raiffa and Schlaifer (1961, pp. 63–65).

4.5. *Finite Tolerance Intervals for the Normal Distribution*

Finally in this Section we consider the problem of obtaining a Bayesian tolerance interval (r_1, r_2) for the normal experimental density with prior density (24). By (15) we have to choose a pair (r_1, r_2) to satisfy the probabilistic relation

$$\Pi[(\mu, \sigma) : \Phi\{(r_2-\mu)/\sigma\} - \Phi\{(r_1-\mu)/\sigma\} \geqslant c \,|\, x^{(n)}] = q, \tag{37}$$

where Φ denotes the distribution function of a $N(0, 1)$ distribution, and the transformation (29) provides the equivalent inequality

$$\Phi\{B^{-\frac{1}{2}}\xi + (r_2-A)\,V^{-1}\eta\} - \Phi\{B^{-\frac{1}{2}}\xi + (r_1-A)\,V^{-1}\eta\} \geqslant c.$$

If $k_1 = (r_1-A)\,V^{-1}, k_2 = (r_2-A)\,V^{-1}$ satisfy the new probabilistic statement then

$$r_1 = A + k_1 V, \quad r_2 = A + k_2 V. \tag{38}$$

It can be shown that the length of the interval $(k_2-k_1)\,V$ is minimized when $k_2 = -k_1 = k$ say, so that $(A-kV, A+kV)$ would be a satisfactory tolerance interval. Because of the distributional properties of (ξ, η) the value of k is exactly that given in frequentist tables; see, for example, Owen (1962, Section 5.4), where $B^{-\frac{1}{2}}\xi$ corresponds to the estimator of μ based on B observations and η the independent estimator of σ with W degrees of freedom. Again the case $a = b = 0$ yields the usual frequentist interval $(\bar{x}-ks, \bar{x}+ks)$ based on the outcome $x^{(n)}$ of n replicates.

5. Optimum Tolerance Regions

On the assumption that the frequentist has chosen his pivotal statistic or the Bayesian his prior density, the tolerance-interval problem, based on (2) or (15) and developed in Section 4, usually leads to a unique sensible solution. This is not so, however, when $\mathcal{X}$ is higher-dimensional. There may then be many statistics R satisfying (2) and many regions satisfying (15) so that formulations based on the frequentist utility (5) are incomplete. A problem of this type is the so-called "diversity problem" of engineers, illustrated by the following simple example.

Example. An electrical supply system of the type illustrated in Fig. 1 is to be designed. "Operations" of the system are independent and the loads "demanded" at

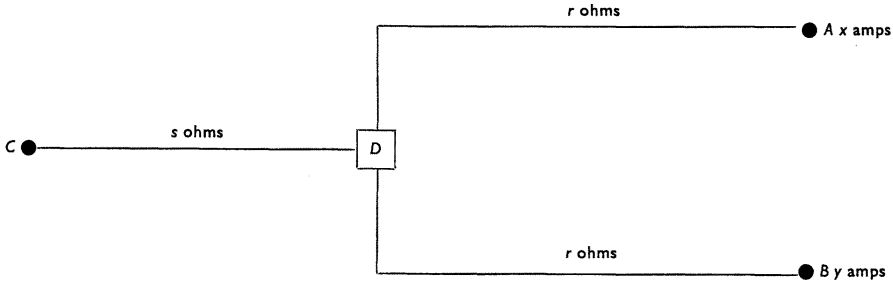

FIG. 1. Electrical supply system.

terminals A and B—of x and y amps, say—are independently and identically distributed, the variation from operation to operation being described by a measure $P(.\,|\,\theta)$ associated with the (x, y)-plane. The system succeeds at an operation if the drops in voltage between C and A and between C and B are both less than v volts. A design consists of choosing the sizes of the cables AD, BD and CD, the sizes being

most conveniently expressed in terms of the associated resistances, say r, r and s ohms. Then clearly the cover at θ provided by the *design* (r, s) is given by $P(R|\theta)$, where

$$R = \{(x, y) : rx + s(x+y) < v, ry + s(x+y) < v\}.$$

If there is available information on the potential demands on the system—possibly the observed demands on an already existing system which has been so overdesigned that failure is virtually impossible—then we might attempt to regard the choice of design (r, s) as the choice of a tolerance region R to satisfy (15), or, in the frequentist case, a tolerance-region statistic R to satisfy (2). But here there are clearly many statistics and many regions satisfying (2) and (15), respectively, and it is obvious that, to make the problem sensible practically, we must take account of the cost $K(R)$ associated with the construction and/or operation of the design (r, s). Regarded as a restricted problem, the utility must take account of both $P(R|\theta)$ and $K(R)$.

Even in the one-dimensional case it is difficult to feel satisfied with the frequentist formulation characterized by (2) or its Bayesian counterpart, for the utility $U(R, \theta)$ in a restricted problem will usually take account of two competitive factors. The more extensive the set R the greater is the cover provided by it at θ and so the greater will be the utility. On the other hand, there must usually be disadvantages attached to increasing R, such as increased cost or length of interval; if not, why, in any practical problem, do we not quote $\mathscr{X}$ as a region which certainly provides cover 1? In the remainder of this Section we term this disadvantage factor a cost $K(R)$ and find it convenient to speak in terms of the operations of systems such as the supply system of the above example.

If a cost $K(R)$—such as running cost, possibly with some allowance for capital cost or depreciation—can be assigned to each operation of the system, then it may be appropriate to consider the following restricted V specification.

$$V(R, x) = \begin{cases} \lambda_1 - K(R) & (x \notin R), \\ \lambda_2 - K(R) & (x \in R). \end{cases} \tag{39}$$

Here λ_1 is the (negative) monetary value associated with failure of an operation and λ_2 the monetary value associated with success of an operation. The corresponding U specification is, by (12),

$$U(R, \theta) = \lambda P(R|\theta) - K(R),$$

where $\lambda = \lambda_2 - \lambda_1$. The Bayesian formulation in terms of the maximization of $W(R, x^{(n)})$, as given by (6), is then straightforward. For instance, simple calculations show that, for the experimental and prior densities of Section 4.2, the region $R = (0, r)$ must be chosen so as to maximize

$$\lambda I_{r/(b+z+r)}(k, a+nk) - K(R), \tag{40}$$

where I denotes the incomplete beta-function (Pearson, 1934). This presents no computational difficulties. The frequentist, on the other hand, is faced with $-U(R, \theta)$ as his loss function in the general decision problem and hence with the formidable task of finding a *function* R which maximizes

$$\int_{\mathscr{X}^n} dx^{(n)} p(x^{(n)}|\theta) [\lambda P\{R(x^{(n)})|\theta\} - K\{R(x^{(n)})\}] \tag{41}$$

for every θ. It is quite possible that no such function exists and he would be forced to adopt one of the other escape routes from the entanglement with θ indicated in Section 2. It seems to be a fair generalization that the more sophisticated the specification of the utility or loss function the more the Bayesian enjoys the situation, whereas the more the frequentist encounters difficulties.

When the cost associated with the system is essentially the capital cost $K(R)$ of constructing the design or region R and there is pressure to provide cover c, then a more suitable form of utility function may be given by

$$U(R, \theta) = \begin{cases} \lambda_1 - K(R) & \text{if} \quad P(R \mid \theta) < c, \\ \lambda_2 - K(R) & \text{if} \quad P(R \mid \theta) \geqslant c. \end{cases} \tag{42}$$

We then have, by (6), that

$$W(R, x^{(n)}) = \lambda \Pi\{H(c, R) \mid x^{(n)}\} - K(R) = \lambda Q(R) - K(R), \tag{43}$$

where $\lambda = \lambda_2 - \lambda_1$ and an additive constant has been omitted. The difficulty with such a formulation is that it may be practically impossible to assess λ. In these circumstances, if the pressure to provide cover c is strong, it is interesting to study the following resolution. From the set of regions satisfying the guarantee of cover given by (15) choose one which minimizes the cost.

The frequentist, trying to formulate his corresponding problem, has first the formidable task of discovering the set $\mathscr{R}$ of all statistics satisfying (2). Even if he can find $\mathscr{R}$, how does he then proceed to choose the statistic which gives minimum cost? There is unlikely to be a statistic $R^* \in \mathscr{R}$ which provides uniformly minimum-cost regions in the sense that

$$K\{R^*(x^{(n)})\} = \min_{R \in \mathscr{R}} K\{R(x^{(n)})\}, \tag{44}$$

for all $x^{(n)} \in \mathscr{X}^n$. He may then decide to take as a suitable tolerance-region statistic that R which satisfies (44) for the particular $x^{(n)}$ he observes, but this suffers from the defect, unpopular with frequentists, that the statistic used depends on the particular $x^{(n)}$ which turns up rather than on some properties of measures over the outcome space. Clearly there are great difficulties involved even in the formulation of the frequentist problem.

The Bayesian, freed from the necessity of determining whole functions in the presence of unknown θ, is able to formulate the problem much more simply. We can, in fact, pose two different problems and exploit a useful duality property of them.

Fixed-quality q problem. We wish the selected region R to satisfy (15) so that the quality $Q(R)$ of the region is fixed at q. If $\mathscr{S}(q)$ denotes the set of all such regions, that is, if

$$\mathscr{S}(q) = \{R : Q(R) = q\}$$

we then wish to minimize $K(R)$ over $\mathscr{S}(q)$.

Fixed-cost k problem. We suppose that the cost of the selected region is to be fixed at k so that we are interested only in regions in the set

$$\mathscr{T}(k) = \{R : K(R) = k\}.$$

We then wish to choose a region $R^*(k)$ which maximizes $Q(R)$ over this set; that is,

$$Q\{R^*(k)\} = \max_{R \in \mathscr{T}(k)} Q(R) = L(k), \tag{45}$$

say. In what follows we make the assumption, reasonable in many applications, that the maximum quality $L(k)$ is a continuous and strictly increasing function of k.

Solutions $R^*(k)$ of fixed-cost problems lead easily to the solution of the fixed-quality q problem. If, in fact, we determine k^* such that $L(k^*) = q$ then $R^*(k^*)$ is an optimum tolerance region of fixed quality q; the minimum cost is k^*. For, suppose that R_1 with cost $k_1 < k^*$ and not $R^*(k^*)$ is optimum. Then we have

$$q = Q(R_1) \leqslant \max_{R \in \mathscr{T}(k_1)} Q(R) = L(k_1) < L(k^*) = q,$$

a contradiction, so that $R^*(k^*)$ is in fact an optimum fixed-quality q region.

Solutions of fixed-quality problems could, by similar reasoning, provide a solution of the fixed-cost problem. Since the set $\mathscr{T}(k)$ is usually easily defined and in no way depends on θ the fixed-cost problem is likely to be a straightforward maximization problem. From its solutions for different k the function $L(k)$ can be explored to obtain k^* and hence $R^*(k^*)$.

It is not our purpose here to explore further the computational aspects of this problem, which, for complicated systems, are considerable. Our objective has been the limited one of showing that the Bayesian approach, in contrast to the frequentist one, to the search for optimum tolerance regions leads to well-formulated problems.

6. Some General Remarks

It is my hope that this paper and that by Mr Thatcher which follows are provocative enough to lead to a discussion which will sweep tolerance-region theory out of the doldrums in which it seems to be at present becalmed. While I have no doubt that frequentists and Bayesians will try to create such a storm as to drive the craft into port, their own or their opponents'—I am not sure which—I hope that the direction of motion may be influenced by the views of users of tolerance regions. A very useful contribution to the discussion could be the opinion of users as to which type of problem—the general or the restricted type—they have met, what utility or loss functions they regard as useful and how applicable they, as users, judge frequentist and Bayesian analyses.

Acknowledgements

I am grateful to members of the Hospital Engineering Research Unit of the University of Glasgow, who, by their patient attention to my explanations of frequentist tolerance-region theory and by their reluctance to accept its concepts and interpretation, first drew my attention to the need for an alternative approach. I should also like to thank both referees for their helpful, constructive criticism.

References

Bowker, A. H. (1947), "Tolerance limits for normal distributions", in *Selected Techniques of Statistical Analysis*, edited by Eisenhart, C., Hastay, M. W. and Wallis, W. A., Chapter 2. New York: McGraw-Hill

Fraser, D. A. S. (1953), "Non-parametric tolerance regions", *Ann. math. Statist.*, **24**, 44–55.

—— (1957), *Nonparametric Methods in Statistics*. New York: Wiley.

Jeffreys, H. (1961), *Theory of Probability*. Oxford University Press.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*. New York: Wiley.

LINDLEY, D. V. (1961), "The use of prior probability distributions in statistical inference and decision", *Proc. 4th Berkeley Symp.*, **1**, 453–469.
OWEN, D. B. (1962), *Handbook of Statistical Tables*. London: Pergamon.
PEARSON, K. (1934), *Tables of the Incomplete Beta-Function*. Cambridge University Press.
RAIFFA, H. and SCHLAIFER, R. (1961), *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University.
SAVAGE, L. J. (1954), *The Foundations of Statistics*. New York: Wiley.
—— (1962), *The Foundations of Statistical Inference*. London: Methuen.
WALD, A. and WOLFOWITZ, J. (1946), "Tolerance limits for a normal distribution", *Ann. math. Statist.*, **17**, 208–215.
WALLIS, W. A. (1951), "Tolerance intervals for linear regressions", *Proc. 2nd Berkeley Symp.*, **1**, 43–51.
WEISSBERG, A. and BEATTY, G. H. (1960), "Tables of tolerance-limit factors for normal distributions", *Technometrics*, **2**, 483–500.
WELCH, B. L. and PEERS, H. W. (1963), "On formulae for confidence points based on integrals of weighted likelihoods", *J. R. statist. Soc.* B, **25**, 318–329.