BU-968-M

Nonoptimality of Randomized Confidence Sets

by

George Casella [1] and Christian Robert [2]

Cornell University and Purdue University

Technical Report #88-9

Department of Statistics
Purdue University

March 1988

[1] George Casella is Associate Professor, Biometrics Unit, Cornell University, Ithaca, NY 14853. This research was supported by National Science Foundation Grant No. DMS 8501973, and was performed while Prof. Casella was Visiting Professor, Department of Statistics, Purdue University, W. Lafayette, IN 47907.

[2] Research partially supported by l'Institut National de la Statistique et des Etudes Economiques (Paris). This research was performed while Prof. Robert was Visiting Assistant Professor at Purdue University.

Nonoptimality of Randomized Confidence Sets

Summary

Randomized confidence sets are used in classical decision theory as a device to obtain optimality results. In particular, in discrete distributions, a uniformly most accurate unbiased (UMAU) confidence set is based on an auxiliary randomizer that is independent of the data. Such procedures are shown to be nonoptimal when evaluated conditionally, using the Buehler-Robinson theory of relevant betting. In particular, for the binomial and Poisson distributions, sets in the sample space are identified on which the UMAU intervals have conditional coverage probability uniformly smaller than the nominal level.

Key Words

Poisson distribution, binomial distribution, relevant betting procedure, α-level Bayes procedure.

AMS Subject Classification (1985)

62C07, 62C99, 62F25, 62F04

1. Introduction

Classical decision theory, as presented in Ferguson (1967) or Lehmann (1986), involves the use of randomized procedures to obtain optimality results. In the case of a convex loss function, such procedures are not required (see, for example, Ferguson, Section 3.4, Theorems 1 and 2). In the cases of both testing and set estimation, common loss functions are not convex and randomized procedures are required.

In general, randomized procedures involve the addition of an extraneous random variable, one that is independent of the data, to achieve a required test size or coverage probability. Thus, if $X_1, \ldots, X_n$ are iid $F(x|\theta)$, where θ is the parameter of interest, a randomized procedure would be based on a statistic (T, V), where T is sufficient for θ, and $V \sim G(v)$, independent of θ. It is usually the case that $F(x|\theta)$ is a discrete probability function and G is an absolutely continuous one. The discreteness of F makes it impossible for a test to achieve any arbitrary α level, but the addition of G alleviates this.

Although such procedures can be classically optimal, consider the following scenario. An experimenter can consult two statisticians and each one gives the correct advice based on a randomized procedure. The experimenter then finds that two opposite conclusions can be legitimately made with the same data. Of course, the randomization is to blame, as any randomized procedure necessarily violates the likelihood principle (Berger and Wolpert 1984). Such behavior is what prompted Casella (1986) to remark that such procedures are to be treated with disdain.

Unfortunately, violation of the likelihood principle (or, more simply, allowing a procedure to take two different actions for the same value of a minimal sufficient statistic) is not a cause for concern for a classical decision theorist. The previously cited theorem in Ferguson (1967) can be interpreted as saying that an auxiliary randomized provides a reduction of the sample space. The performance of any procedure based on the entire sample can be equalled by the performance of a procedure based on only a sufficient statistic plus a randomizer. Since this reduction simplifies the problem, why not use it?

However, deep inside it should be felt that procedures using information that is no more than random noise are nonoptimal. It is the purpose of this paper to demonstrate, using frequency-based criteria, that randomized procedures are nonoptimal. To do so, we must turn to the Buehler-Robinson theory (Buehler 1959, Robinson 1979a, 1979b) of relevant betting, recently discussed by Maatta and Casella (1987) in the context of variance estimation and Casella (1987) in the context of estimation of a multivariate normal mean.

Briefly, if $C(X)$ is a confidence set for a parameter θ, where $X \sim F(x|\theta)$, and $C(X)$ satisfies

$$P_\theta(\theta \in C(X)) \geq 1 - \alpha \qquad \text{for all } \theta \tag{1.1}$$

for some specified value $1-\alpha$, then it would be disturbing if we found a set A in the sample space such that

$$P_\theta(\theta \in C(X)|X \in A) \leq 1 - \alpha - \epsilon \qquad \text{for all } \theta, \tag{1.2}$$

for some $\epsilon > 0$. Such a set is called a ***negatively biased relevant set*** and casts serious doubts on the validity of the assertion (1.1).

These sets were investigated by Buehler, and formalized by Robinson in the following way

Definition 1.1: For the confidence set $C(X)$ with confidence assertion $\beta(X)$, the function $S(X)$ is a ***relevant betting procedure*** if, for some $\epsilon > 0$,

$$E_\theta\{(I(\theta \in C(X)) - \beta(X))S(X)\} \geq \epsilon \, E_\theta|S(X)|,$$

with strict inequality for some θ. If $\epsilon = 0$, then $S(X)$ is a *semirelevant betting procedure*.

A betting procedure is taken to be any bounded function of X. Notice that $S(X) = -I(X \in A)$ reduces the definition to (1.2) and, in general, if $S(X) \leq 0$ for all X and if S is relevant, $S(X)$ is a ***negatively biased*** relevant betting procedure.

The existence of any type of betting procedure allows a nonnegative expected gain for a bettor and thus leads us to question the validity of the probability assertion. However, the existence of a negatively biased procedure is much more serious. Such a procedure says that the unconditional probability assertion is wrong in the worst possible way: the experimenter is overstating confidence uniformly.

In this paper, we demonstrate that randomized confidence sets are nonoptimal in allowing negatively biased betting. In particular, in Section 2, we exhibit negatively biased semirelevant sets for the Uniformly Most Accurate Unbiased (UMAU) confidence intervals for a binomial or Poisson parameter. In Section 3, we extend a result of Pierce (1973) to show that, in compact parameter spaces, randomized tests allow relevant betting, and hence are nonoptimal.

2. Negatively-Biased Betting Procedures for the Binomial and the Poisson Distributions

We first give some preliminary results concerning the form and behavior of UMPU tests in the binomial and Poisson distributions. Also we note that, although we are principally concerned with the behavior of confidence sets, we can equivalently work with hypothesis tests. If an α level test of H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$ is given by

$$\varphi(X) = \begin{cases} 1 & \text{if } X \in A^c(\theta_0), \\ 0 & \text{if } X \in A(\theta_0), \end{cases}$$

for some acceptance region $A(\theta_0)$, then this defines the $1-\alpha$ confidence set $C_\varphi(X)$, given by

$$\theta \in C_\varphi(X) \Longleftrightarrow X \in A(\theta).$$

It is straightforward to establish that $S'(X)$ is relevant for $C_\varphi(X)$ if and only if

$$E_{\theta_0}\{(\varphi(X) - \alpha)S'(X)\} \geq \epsilon\, E_{\theta_0}|S'(X)| \tag{2.1}$$

for some $\epsilon > 0$. Moreover, $-S'(X)$ is a negative biased betting procedure for C_φ if $S'(X) \geq 0$ satisfies (2.1).

Since inversion of UMPU tests gives UMAU intervals, we equivalently work with either the interval or testing formulation.

2.1 UMPU Tests

Given an observation x of a random variable X whose distribution belongs to a one-parameter family, with parameter θ, it is known (see Lehmann (1986, Chap. 4)) that there does not exist an UMP test for the hypothesis H_1: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$. However, there exists an *UMPU* test, i.e. a test φ_{θ_0} such that

$$E_{\theta_0}(\varphi_{\theta_0}(X)) = \alpha \quad \text{and} \quad E_\theta(\varphi_{\theta_0}(X)) \geq \alpha \quad \text{for every} \quad \theta \neq \theta_0$$

(given a level $0 \leq \alpha \leq 1$). This test is given by

Theorem 2.1. (Lehmann, Chap. 4, Section 2). A UMPU test for the hypothesis H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$, for a given level α is

$$\varphi_{\theta_0}(x) = \begin{cases} 1 & \text{if } x < C_1(\theta_0) \text{ or } x > C_2(\theta_0), \\ \gamma_i(\theta_0) & \text{if } x = C_i(\theta_0), \qquad i = 1, 2, \\ 0 & \text{if } C_1(\theta_0) < \alpha < C_2(\theta_0), \end{cases}$$

where the C_i's and γ_i's $(i = 1, 2)$ are determined by the constraints

$$E_{\theta_0}(\varphi_{\theta_0}(x)) = \alpha \qquad \text{and} \qquad E_{\theta_0}(X\varphi_{\theta_0}(x)) = \alpha\, E_{\theta_0}(X).$$

For the binomial and Poisson distributions, we then have the following consequence, using randomization:

Corollary 2.1.

(i) If $X \sim$ binomial (n, p), a UMPU test for the hypothesis H_0: $p = p_0$ vs. H_1: $p \neq p_0$ satisfies

$$\varphi_{p_0}(x+v) = \begin{cases} 1 & \text{if } x+v < K_1(p_0) \text{ or } x+v > K_2(p_0) \\ 0 & \text{if } x+v \in [K_1(p_0), K_2(p_0)] \end{cases}$$

where v is a realization of $V \sim \mathcal{U}(0,1)$ and $K_1(p_0)$, $K_2(p_0)$ are uniquely determined by

$$E_{p_0}(\varphi_{p_0}(X+V)) = \alpha \qquad \text{and} \qquad E_{p_0}(x\varphi_{p_o}(X+V)) = \alpha\, np_0 \tag{2.2}$$

(ii) If $X \sim$ Poisson (λ), a UMPU test for the hypothesis H_0: $\lambda = \lambda_0$ vs. H_1: $\lambda \neq \lambda_0$ satisfies

$$\varphi_{\lambda_0}(x+v) = \begin{cases} 1 & \text{if } x+v < K_1(\lambda_0) \text{ or } x+v > K_2(\lambda_0) \\ 0 & \text{otherwise} \end{cases}$$

where v is a realization of $V \sim \mathcal{U}(0,1)$ and $K_1(\lambda_0)$, $K_2(\lambda_0)$ are uniquely determined by

$$E_{\lambda_0}(\varphi_{\lambda_0}(X+V)) = \alpha \qquad \text{and} \qquad E_{\lambda_0}(X\varphi_{\lambda_0}(X+V)) = \alpha\, \lambda_0. \tag{2.3}$$

Furthermore, we deduce the following from Lehmann (1986, Lemma 1 – Chap. 5, §5).

Proposition 2.1. When the randomized tests φ are defined as in Corollary 2.1, the functions $K_i (i = 1, 2)$ are strictly increasing.

For the binomial and Poisson distributions, we will now establish that there exist negatively-biased semirelevant betting procedures against the UMPU tests defined in Corollary 2.2.

2.2 The binomial case

The confidence intervals for p deduced from Corollary 2.1 (i) have been tabulated in Blyth and Hutchinson (1960). The authors have, in particular, established that there exist unique n_i, $\gamma_i(p)$ such that $K_i(p) = n_i + \gamma_i(p)$, with n_i an integer and $\gamma_i(p) \in [0,1)$ $(i = 1,2)$. These numbers are solutions of

$$(2.4) \qquad \begin{cases} \gamma_1 = \dfrac{(n_2-np)\{P_p(n_1 \le X \le n_2-1)-\alpha\}-n_1(1-p)P_p(X=n_1)+n_2(1-p)P_p(X=n_2)}{(n_2-n_1)P_p(X=n_1)} \\ \\ \gamma_2 = \dfrac{(n_1-np)\{P_p(n_1 \le X \le n_2-1)-\alpha\}-n_1(1-p)P_p(X=n_1)+n_2(1-p)P_p(X=n_2)}{(n_2-n_1)P_p(X=n_2)} \end{cases}$$

Before giving the betting procedure against this UMPU test, we need to establish some basic results about the behavior of K_1 and K_2, when p goes towards the extremities, 0 and 1.

Lemma 2.1. For the UMPU binomial test, we have

$$\lim_{p\to 0} K_1(p) = \alpha \quad \text{and} \quad \lim_{p\to 0} K_2(p) = 2 - \alpha$$

and

$$\lim_{p\to 1} K_1(p) = n - 1 + \alpha \quad \text{and} \quad \lim_{p\to 1} K_2(p) = n + 1 - \alpha.$$

Proof. We will only give the proof for $p \to 0$, the result for $p \to 1$ following by symmetry.

As $p \to 0$, $n_1(p) \to 0$ and thus, for small enough p, the equations (2.2) can be written

$$(2.5) \qquad K_1(p)P_p(X=0) + \sum_{n_2(p)+1}^{n} P_p(X=i) + (n_2(p)+1-K_2(p))P_p(X=n_2(p)) = \alpha$$

and

$$(2.6) \qquad n_2(p)(n_2(p)+1-K_2(p))P_p(X=n_2(p)) + \sum_{n_2(p)+1}^{n} i\ P_p(X=i) = \alpha\ np.$$

The fact that $n_1(p)$ goes to 0 as p goes to 0 follows from the first equation: otherwise, the left term would go to 0 while the right one, α, would remain constant, a contradiction.

The limit of (2.5), as p goes to 0, is

$$(2.7) \qquad K_1(0) + (n_2(0)+1-K_2(0)) \lim_{p\to 0} P_p(X=n_2(p)) = \alpha.$$

We will show that $n_2(0) = 1$, from which it follows that $\lim_{p \to 0} K_1(p) = \alpha$.

Consider two cases. First, if $n_2(0) = 0$, we obtain from (2.7) that

$$K_1(0) + 1 - K_2(0) = \alpha.$$

Therefore, since K_2 is continuous, there exists p^* such that $n_2(p) = 0$ for $p \leq p^*$. For these $p \leq p^*$, (2.6) implies

$$\sum_{i=1}^{n} i P_p(x = i) = \alpha \; np,$$

but $\sum_{i=1}^{n} i \; P_p(x = i) = np$. Therefore we have a contradiction.

Now consider $n_2(0) \geq 2$. Equation (2.6) implies

$$n_2(p)(n_2(p) + 1 - K_2(p)) \binom{n}{n_2(p)} p^{n_2(p)-1}(1-p)^{n-n_2(p)} + \sum_{i=n_2(p)+1}^{n} i \binom{n}{i} p^{i-1}(1-p)^{n-i} = \alpha n.$$

Again, this is impossible as the left side goes to 0 as p goes to 0. Thus (2.7) implies that $K_1(0) = \alpha$ and we have from (2.6) that

$$(2 - K_2(p)) \; n(1-p)^{n-1} + \sum_{i=2}^{n} i \; \frac{P_p(x = i)}{p} = \alpha n.$$

The limit of the equality, as p goes to 0, gives $K_2(0) = 2 - \alpha$. □

We are now able to show the main result of this section.

Theorem 2.2. Let s be the indicator function of $[1, n]^c$, i.e.

$$s(t) = \begin{cases} 0 & \text{if } t \in [1, n], \\ 1 & \text{otherwise.} \end{cases}$$

Then

$$E_p[(\varphi_p(X + V) - \alpha)s(X + V)] > 0, \quad \forall \, p \in \,]0, 1[,$$

where φ_p is defined as in Corollary 2.2 (i) and $V \sim \mathcal{U}(0, 1)$.

This result means that the UMPU test φ_p rejects the null hypothesis too often when X is equal to 0 or n. From the discussion at the beginning of the section, it then follows that the procedure that bets against coverage if $X = 0$ is a negatively biased betting procedure against the UMAU interval.

Proof. As φ_p is also an indicator function, note first that we have

$$
\begin{aligned}
(2.8) \qquad & E_p[(\varphi_p(X+V) - \alpha)s(X+V)] \\
& = P_p((X+v) \notin [K_1(p), K_2(p)] \cup [1,n]) - \alpha \, P_p(X+V \in [1,n]).
\end{aligned}
$$

Now consider four exhaustive cases, in which we show that (2.8) is always positive

(i) $1 \le K_1(p) \le K_2(p) \le n$: In this case, (2.8) becomes

$$(1-\alpha)P_p(X+V \notin [1,n]) > 0.$$

(ii) $K_1(p) < 1 \le K_2(p) \le n$: In this case, (2.8) is

$$
\begin{aligned}
E_p[(\varphi_p(X+V) - \alpha)s(X+V)] &= P_p(X+V \notin [K_1(p), n]) - \alpha \, P_p(X+V \notin [1,n]) \\
&= K_1(p)P_p(X=0) + P_p(X=n) - \alpha(P_p(X=0) \\
&\qquad + P_p(X=n)) \\
&= (K_1(p) - \alpha)P_p(X=0) + (1-\alpha)P_p(X=n) > 0,
\end{aligned}
$$

as K_1 is an increasing function and $\lim_{p \to 0} K_1(p) = \alpha$.

(iii) $1 \le K_1(p) < n < K_2(p)$: This case follows from case (ii) by symmetry, as

$$E_p[(\varphi_p(X+V)-\alpha)s(X+V)] = (1-\alpha)P_p(X=0)+(n+1-\alpha-K_2(p))P_p(X=n) > 0.$$

(iv) $K_1(p) < 1 < n < K_2(p)$: When this case occurs, (2.8) is

$$
\begin{aligned}
E_p((\varphi_p(X+V) - \alpha)s(X+V)) &= (K_1(p) - \alpha)P_p(X=0) + (n+1-\alpha-K_2(p)) \\
&\qquad P_p(X=n) > 0.
\end{aligned}
$$

$\square$

This betting procedure is only semirelevant, because the expectation $E_p[(\varphi(X+V) - \alpha)s(X+V)]$ goes to 0 and p goes to 0 or 1. However, due to Theorem 3.1 of the next section, we know there exists relevant betting procedures. It is easy to construct bets that give increased expected gain over the bet in Theorem 2.2. Consider, for example, betting procedures associated with the functions

$$s_a(t) = \begin{cases} a & \text{if } t \notin [1,n], \\ -1 & \text{if } t \in [1,n] \end{cases} \tag{2.9}$$

which bet against the test when X is equal to 0 or n and for it otherwise. We have then the following result.

Proposition 2.2. For every $a \geq 0$, for every $p \in]0,1[$,

$$E_p[(\varphi_p(X+V) - \alpha)s_a(X+V)] > 0.$$

Furthermore, this expectation is an increasing function of a.

This property is illustrated by the graph in Figure 1, where the expectation $E_p[(\varphi_p(X+V) - \alpha)s(X+V)]$ has been computed for s introduced in Theorem 2.2 and $s_a(a = 1,2,3)$.

Proof. The techniques being the same than in Theorem 2.2, we have skipped the intermediate computations. We have

$E_p((\varphi_p(X+Y) - \alpha)s(X+Y)) =$

$$\begin{cases} (a+1)P_p(X = 0 \text{ or } n)(1-\alpha) \text{ if } 1 \leq K_1(p) \leq K_2(p) \leq n, \\ (a+1)\{(K_1(p) - \alpha)P_p(X=0) + (1-\alpha)P_p(X=n)\} \text{ if } K_1(p) < 1 \leq K_2(p) \leq n, \\ (a+1)\{(1-\alpha)P_p(X=0) + (n+1-\alpha-K_2(p))P_p(X=n)\} \text{ if } 1 \leq K_1(p) \leq n < K_2(p), \\ (a+1)\{(K_1(p)-\alpha)P_p(X=0) + (n+1-\alpha-K_2(p))P_p(X=n)\} \text{ otherwise}, \end{cases}$$

which gives us the desired result. □

Note that these betting procedures are still semirelevant. This problem is due to the fact that $P_p(X = 0)$ goes to 1 as p goes to 0. Therefore, in order to get relevant betting procedures, one should use more complex functions s which could compensate in the extremes.

2.3. The Poisson case.

In the same way as for the binomial distribution, we will exhibit a negatively-biased semirelevant betting procedure for the Poisson distribution. Blyth and Hutchinson (1961) have also tabulated confidence intervals for λ, deduced from Corollary 2.1 (ii); here, too, there exist unique $n_i(\lambda), \gamma_i(\lambda)$ such that $K_i(\lambda) = n_i(\lambda) + \gamma_i(\lambda)$ and the functions K_i are also continuous.

The analog of Lemma 2.1 in this case is

Lemma 2.2. We have

$$\lim_{\lambda \to 0} K_1(\lambda) = \alpha \qquad \text{and} \qquad \lim_{\lambda \to 0} K_2(\lambda) = 2 - \alpha.$$

The second part of Lemma 2.1 is of no interest here. If λ goes to infinity, the Poisson distribution can be approximated by a normal distribution and the reason for randomization disappears. The normal approximation is also valid, for an arbitrary λ, as x goes to infinity (cf. Casella and Robert (1988)). However, we can still propose betting procedures which take action for large values of X (see Proposition 2.3).

Proof. For the Poisson distribution,

$$P_\lambda(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \qquad x = 0, 1, 2, \dots\ ,$$

we have the same property as for the binomial distribution, that $P_\lambda(X = x) = \lambda^x A_\lambda(x)$, with $A_\lambda(x) \neq 0$. This property being the basis of the proof of Lemma 2.1, we can apply the same arguments here. □

We will now establish the analog of Theorem 2.2; as before, the simplest betting procedure is associated with the extreme value. When $X = 0$, it is possible to bet against the test because the conditional level of rejection is more than α.

Theorem 2.3. Let s be the indicator function of $[0, 1]$, i.e.

$$s(t) = \begin{cases} 1 & \text{if } t \in [0, 1), \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E_\lambda[(\varphi_\lambda(X+V)-\alpha)s(X+V)] > 0, \ \ \forall\, \lambda > 0,$$

where $V \sim \mathcal{U}(0,1)$ and φ_λ is defined as in Corollary 2.1 (ii).

Proof. We can write the above expression in the form

$$P_\lambda(X+V \notin (K_1(\lambda), K_2(\lambda) \cup [1,+\infty)) - \alpha\ P_\lambda(X+V \in [0,1]). \tag{2.10}$$

Now consider two cases, in which we show that (2.10) is positive.

(i) $K_1(\lambda) \geq 1$: Expression (2.10) is then

$$P_\lambda(X+V \in [0,1]) - \alpha \ \ P_\lambda(X+V \in [0,1]) > 0$$

(ii) $K_1(\lambda) < 1$: As $K_2(\lambda) > 1$ for every λ (see Lemma 2.2), expression 2.10 is

$$\begin{aligned} &P_\lambda(X+V < K_1(\lambda)) - \alpha\ P_\lambda(X+V \in [0,1]) \\ &= K_1(\lambda)P_\lambda(X=0) - \alpha\ P_\lambda(X=0) \\ &= (K_1(\lambda)-\alpha)P_\lambda(X=0) > 0, \end{aligned}$$

because of Lemma 2.2. □

Once again, the extremum of the possible values is a valuable candidate to build up the betting set; the UMPU test φ_λ rejects more than α when $X = 0$. In fact, it seems possible to go beyond 1, i.e. to take s as the indicator function of the interval $[0, 1+\epsilon]$ if ϵ is small enough ($K_1(\lambda) - \alpha(1+\epsilon\lambda)$ must remain positive as λ goes to 0). But, when $n = 10$, $\alpha = 0.05$, numerical study shows that $\epsilon = 0.01$ is too large.

As for the binomial distribution, we have also considered betting procedures associated with the functions

$$s_a(t) = \begin{cases} a & \text{if } t \in [0,1], \\ -1 & \text{otherwise.} \end{cases} \tag{2.11}$$

The analog of Proposition 2.2 is then

Proposition 2.3. For every $a \geq 0$, for any $\lambda > 0$,

$$E_\lambda[(\varphi_\lambda(X+V)-\alpha)s_a(X+V)] > 0.$$

Furthermore, this expectation is an increasing function of a.

This result is illustrated by the graph in Figure 2 where we represent the expectations $E_\lambda[(\varphi_\lambda(X+V)-\alpha)s(X+V)]$ for s defined as in Theorem 2.3 and s_a of (2.11) for a equal to 1 and 2.

Proof. As previously, it is easy to show that

$$E_\lambda[(\varphi_\lambda(X+V)-\alpha)s_a(X+V)] = \begin{cases} (a+1)(1-\alpha)P_\lambda(X=0) & \text{if } 1 \leq K_1(\lambda), \\ (a+1)(K_1(\lambda)-\alpha)P_\lambda(X=0) & \text{if } K_1(\lambda) < 1, \end{cases}$$

which is sufficient to establish the result. □

Note that all these betting procedures are semirelevant. As the parameter space is not compact, we are not sure whether there exist relevant betting procedures for the Poisson distribution. The case of a compact parameter space will be discussed in the next section.

3. Necessary and sufficient condition of existence of relevant betting procedures.

For point estimation, it is well known that admissible estimators are (proper) Bayes estimators or limits of such estimators in many cases (see, e.g., Brown (1986)). For confidence intervals, when the parameter space is finite, Pierce (1973) establishes that a confidence region admits a relevant betting procedure if and only if it is not essentially δ-level Bayes (see definition below). We generalize this result to the case where the parameter space is compact and where the measure induced by the observed random variable is a continuous function of the parameters. It is, in particular, the case when the distribution is discrete with compact parameter space (binomial, geometric, hypergeometric, logarithmic distributions, ...).

Let us first recall the notion of δ-level Bayes confidence procedure. Let X be a random variable on $\mathcal{X}$ with density f_θ w.r.t. a measure λ, where $\theta \in \Omega$. If $C(x)$ is a confidence region, we denote $C(x,\theta)$ the associated indicator function, i.e.

$$C(x,\theta) = \begin{cases} 1 & \text{if } \theta \in C(x), \\ 0 & \text{otherwise.} \end{cases}$$

For $0 < \delta < 1$, the procedure C is *δ-level Bayes* for a prior distribution π on Ω if

$$E^{\pi(\theta|x)}[C(x,\theta)] = \delta \qquad \text{for every } x \in \mathcal{X}, \tag{3.1}$$

where $\pi(\theta|x)$ is the posterior distribution of θ; C will be said *essentially δ-level Bayes* if (3.1) holds for almost (λ) all $x \in \mathcal{X}$.

The proof of the main result relies upon a separation theorem in topological spaces (see, e.g., Parthasarthy (1967, Theorem 1.6)).

Lemma 3.1. If $\mathcal{X}$ is a metric space, A and B two disjoint non-empty closed convex sets of $\mathcal{X}$, there exist $\gamma \in \mathbf{R}$ and $\Lambda \in \mathcal{X}^*$, the topological dual of $\mathcal{X}$, such that

$$\Lambda x < \gamma < \Lambda y, \qquad \text{for every } x \in A, \;\; y \in B.$$

From this proposition, we have the following theorem

Theorem 3.1. Let X be a random variable on $\mathcal{X}$ with distribution P_θ such that $\theta \in \Omega$, compact, and the measure induced by P_θ is a continuous function of θ. Then, for any confidence region $C(X)$, $C(X)$ admits a relevant betting procedure if and only if C is not essentially δ-level Bayes for any $\delta \in]0,1[$.

Proof. Let f_θ be the density of X with respect to a measure λ. First suppose that C is essentially δ-level Bayes for the prior distribution π, i.e. that

$$\int_\Omega (\delta - C(x,\theta)) f_\theta(x) d\pi(\theta) = 0, \qquad \text{for almost } (\lambda) \text{ all } x.$$

Therefore, for every $s \in L_\infty(\mathcal{X}, \mathcal{B}, \lambda)$,

$$\begin{aligned} &\int_\Omega \left[\int_\mathcal{X} s(x)(\delta - C(x,\theta)) f_\theta(x) d\lambda(x) \right] d\pi(\theta) \\ &= \int_\mathcal{X} s(x) \int_\Omega (\delta - C(x,\theta)) f_\theta(x) d\pi(\theta) d\lambda(x) \\ &= 0, \end{aligned}$$

by application of Fubini's theorem (as Ω is compact). This implies that there exists no relevant betting procedure against C.

Suppose now that C is not essentially δ-level Bayes. Therefore, for every prior distribution π on Ω, there exists a set $N_\pi \subset \mathcal{X}$ with $\lambda(N_\pi) > 0$ and

$$\int_\Omega (\delta - C(x,\theta)) f_\theta(x) d\pi(\theta) \neq 0 \qquad \text{for } x \in N_\pi.$$

If $\mathcal{A}$ is the collection of Borel sets associated with the topology induced by the metric on Ω and $\mathcal{M}(\Omega)$ is the space of probability measures on $(\Omega, \mathcal{A})$, $\mathcal{M}(\Omega)$ is a compact space for the weak topology (see Parthasarathy (1967, Theorem 6.4)), as Ω is compact. Let us define the function ψ, from $\mathcal{M}(\Omega)$ into $L_1(\mathcal{X}, \mathcal{B}, \lambda)$ by

$$\psi(\pi)(x) = \int_\Omega (\delta - C(x,\theta)) f_\theta(x) d\pi(\theta)$$

for every $x \in \mathcal{X}$ and every $\pi \in \mathcal{M}(\Omega)$. If C is not essentially Bayes, $\psi(\pi)$ is not identically null, for every $\pi \in \mathcal{M}(\Omega)$.

We will now establish that *the closure of $\psi(\mathcal{M}(\Omega))$ does not contain the null function.* By application of Lemma 3.1, we know then that there exists $s \in (L_1(\mathcal{X}, \mathcal{B}, \lambda))^* = L_\infty(\mathcal{X}, \mathcal{B}, \lambda)$ such that

$$\int_\mathcal{X} s(x)\psi(\pi)(x) d\lambda(x) = \int_\mathcal{X} s(x) \int_\Omega (\delta - C(x,\theta)) f_\theta(x) d\pi(\theta) d\lambda(x) > 0$$

for every $\pi \in \mathcal{M}(\Omega)$. This implies, by Fubini's theorem and by considering the Dirac priors on Ω, that, for every $\theta \in \Omega$,

$$\int_{\mathcal{X}} s(x)(\delta - C(x,\theta)) f_\theta(x) d\lambda(x) > 0,$$

i.e. that s is a relevant betting procedure for C.

We now establish the intermediary result. Suppose there exists a sequence $(\pi_n)_n$ in $\mathcal{M}(\Omega)$ such that $(\psi(\pi_n))_n$ converges in $L_1(\mathcal{X}, \mathcal{B}, \lambda)$ to 0, the null function. Then, as $\mathcal{M}(\Omega)$ is compact, there exists $\pi \in \mathcal{M}(\Omega)$, and a subsequence $(\pi_{n_k})_k$ converging, in the weak sense, towards π. Let A_1 be the subset of $\mathcal{X}$ defined by

$$\{x \in \mathcal{X}; \int_\Omega (\delta - C(x,\theta) f_\theta(x) d\pi(\theta) \geq 0\}.$$

For every $\epsilon > 0$, there exists k_0 such that, for $k > k_0$,

$$||\psi(\pi_{n_k})||_1 = \int_{\mathcal{X}} |\psi(\pi_{n_k})(x)| d\lambda(x) < \epsilon.$$

This implies, for $k > k_0$,

$$|\int_{A_1} \psi(\pi_{n_k})(x) d\lambda(x)| < \int_{A_1} |\psi(\pi_{n_k})(x)| d\lambda(x) < \epsilon,$$

or, by Fubini's theorem,

$$|\int_\Omega \left[\int_{A_1} (\delta - C(x,\theta)) f_\theta(x) d\lambda(x)\right] d\pi_{n_k}(\theta)| < \epsilon.$$

We have assumed that the function

$$\begin{aligned} m(\theta) &= \int_{A_1} (\delta - C(x,\theta)) f_\theta(x) d\lambda(x) \\ &= \delta P_\theta(A_1) - P_\theta(A_1 \cap C_\theta) \end{aligned}$$

is a continuous function of θ (where $C(x,\theta)$ is the indicator function of C_θ, i.e. $x \in C_\theta \iff \theta \in C(x)$).

Therefore, by definition of the weak topology, the sequence $\left(\int_\Omega m(\theta) d\pi_{n_k}(\theta)\right)_k$ is converging towards $\int_\Omega m(\theta) d\pi(\theta)$. There exists then k_1 such that, for $k > k_1$,

$$|\int_\Omega m(\theta) d\pi_{n_k}(\theta) - \int_\Omega m(\theta) d\pi(\theta)| < \epsilon.$$

Thus, for $k > \max(k_0, k_1)$,

$$|\int_\Omega m(\theta)d\pi(\theta)| < 2\epsilon$$

or

$$|\int_{A_1}\int_\Omega (\delta - C(x,\theta))f_\theta(x)d\pi(\theta)d\lambda(x)| < 2\epsilon,$$

by Fubini's theorem. Due to the definition of A_1, this implies

$$\int_\Omega (\delta - C(x,\theta))f_\theta(x)d\pi(\theta) = 0, \text{ for every } x \in A_1.$$

The same technique, applied to $A_2 = A_1^c$, leads to

$$\int_\Omega (\delta - C(x,\theta))f_\theta(x)d\pi(\theta) = 0 \quad \text{for every } x \in \mathcal{X}.$$

We have then got a contradiction; 0 does not belong to $\overline{\psi(\mathcal{M}(\Omega))}$. □

A randomized procedure, since it violates the likelihood principle, cannot be essentially δ level Bayes and thus, if the parameter space is compact, a relevant betting procedure exists.

That Theorem 3.1 can be extended to a general parameter space Ω is not clear. Pierce (1973) proposed an alternative definition of a Bayes set to be used in a general case, but the class of confidence procedures satisfying the definition is actually very large. The technical problems in general parameter spaces become quite impressive, making a result like Theorem 3.1 extremely difficult to establish. It is probably the case, however, that a confidence procedure must be a limit of Bayes sets to be free to relevant betting.

References

Berger, J. and Wolpert, R. (1984). *The Likelihood Principle.* Institute of Mathematical Statistics Monograph Series, Hayward, California.

Blyth, C. and Hutchinson, D. (1960). Table of Neyman-Shortest unbiased confidence intervals for the binomial parameter. *Biometrika* **47(3)**, 381–391.

Blyth, C. and Hutchinson, D. (1961). Table of Neyman-shortest unbiased confidence intervals for the Poisson parameter. *Biometrika* **48 (1–2)**, 191–194.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families.* Institute of Mathematical Statistics Monograph Series, Hayward, California.

Buehler, R. (1959). Some validity criteria for statistical inference. *Ann. Math. Stat.* **30**, 845–863.

Casella, G. (1986). Refining binomial confidence intervals. *Can. J. Stat.* **14(2)**, 113–129.

Casella, G. (1987). Conditionally acceptable recentred set estimators. *Ann. Stat.* **15(4)**, 1363–1371.

Casella, G. and Robert, C. (1988). Refining Poisson Confidence Intervals, Technical Report #88-7, Purdue University, West Lafayette, IN.

Ferguson, T. (1967). *Mathematical Statistics: a Decision-Theoretic approach.* Academic Press, NY.

Lehmann, E. (1986). *Testing Statistical Hypothesis.* Wiley, NY (2nd edition).

Maatta, J. and Casella, G. (1987). Conditional properties of interval estimators of the normal variance. *Ann. Stat.* **15(4)**, 1372–1388.

Parthasarathy, T. (1967). *Probability measures on metric spaces.* Academic Press, NY.

Pierce, D. (1973). On some difficulties in a frequency theory of inference. *Ann. Stat.* **1(2)**, 241–250.

Robinson, G. (1979a). Conditional properties of statistical procedures. *Ann. Stat.* **7(4)**, 742–755.

Robinson, G. (1979b). Conditional properties of statistical procedures for location and scale parameters. *Ann. Stat.* **7(4)**, 756–771.

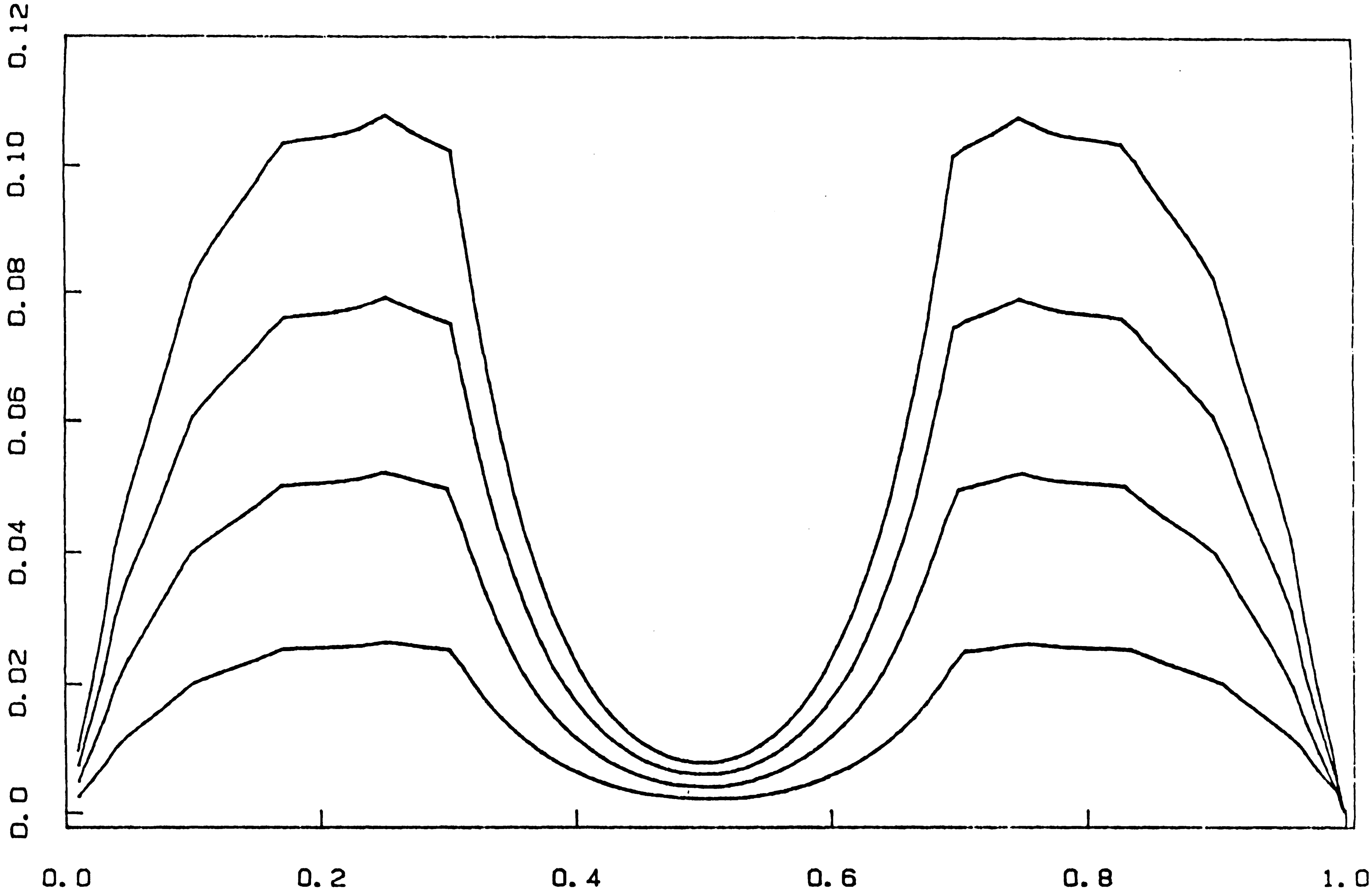

Figure 1: Expected Gain $E_p[(\varphi_p(X+V)-\alpha)s(X+V)]$ for s in Theorem 2.2 and (2.9) with $a = 1,2,3$. These curves are increasing in a.

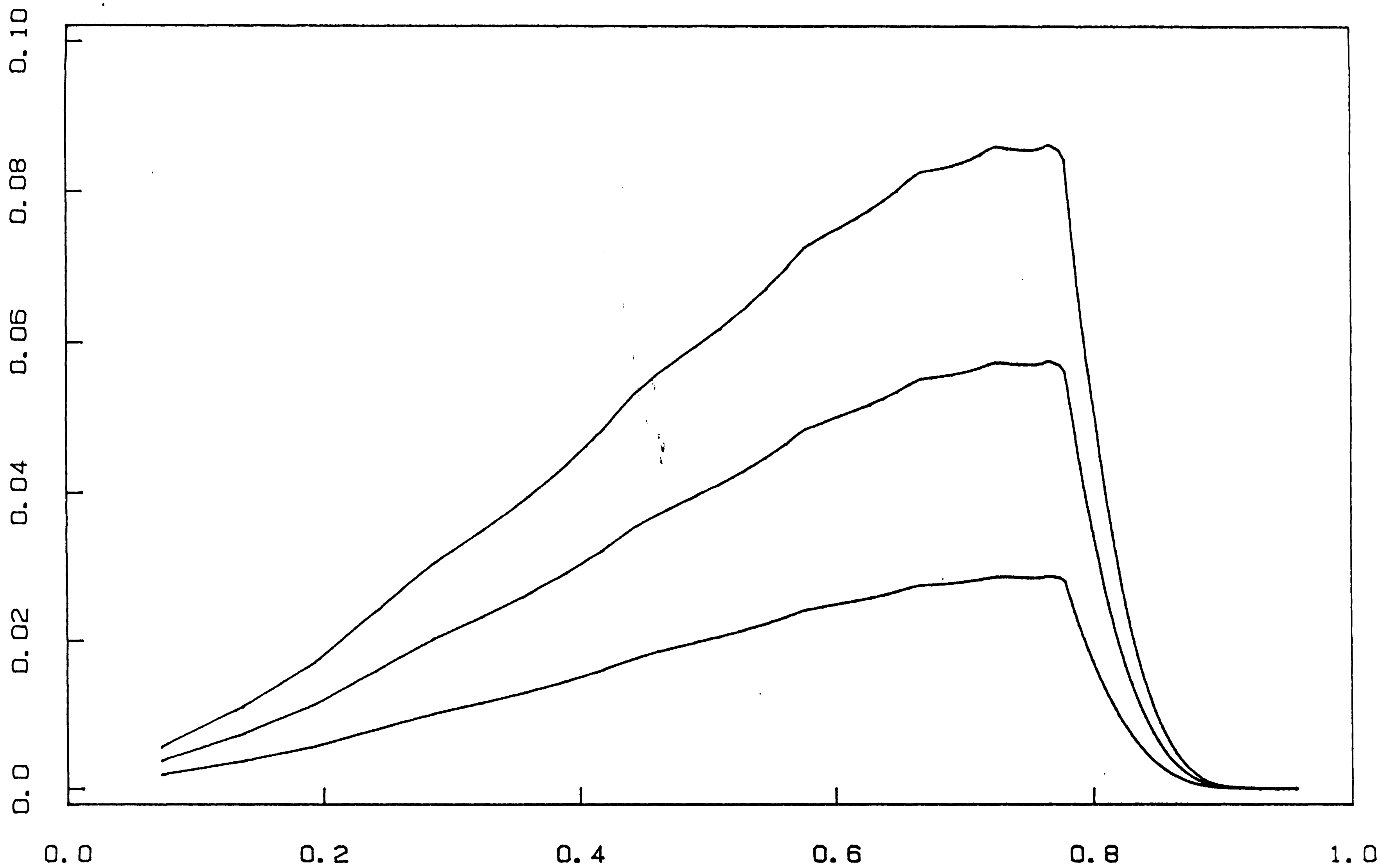

Figure 2: Expected Gain $E_\lambda[(\varphi_\lambda(X+V)-\alpha)s(X)]$ for s in Theorem 2.3 and (2.11). Lowest curve is for s in Theorem 2.3, and the other two are for s in (2.11) with $a = 1, 2$. These curves are increasing in a. The horizontal axis is $\lambda/1-\lambda$.