October 21, 1999

How Incoherent is Fixed-Level Testing?

Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane

It has long been known that the practice of testing all hypotheses at the same level (such as 0.05), regardless of the distribution of the data, is not consistent with Bayesian expected utility maximization. According to deFinetti's "Dutch Book" argument, procedures that are not consistent with expected utility maximization are incoherent and they lead to gambles that are sure to lose no matter what happens. In this paper, we propose a method to measure the degree to which incoherent procedures are sure to lose, so that we can distinguish slightly incoherent procedures from grossly incoherent ones. We present a detailed analysis of testing a simple hypothesis against a simple alternative as a case-study of how the method can work.

KEY WORDS: Bookie; Brier score; Coherence; Escrow; Gambler; Hypothesis testing.

1. INTRODUCTION

Cox (1958) and Lindley (1972) have shown that the practice of testing all hypotheses at the same level, regardless of the distribution of the data, can lead to inadmissibility and incompatibility with Bayesian decision theory. One of the most compelling arguments for Bayesian decision theory and the use of probability to model uncertainty is the "Dutch Book" argument, which says that if you are willing to accept either side of each bet implied by your statements and finite combinations of these together, either

(a) those statements are "coherent," that is they comport with the axioms of probability, or

(b) a gambler betting against you can choose bets so that you are a sure loser.

Excellent introductions to the concepts of coherence and Dutch Book can be found in Shimony (1955), Freedman and Purves (1969), and de Finetti (1974, Section 3.3).

As a practical matter, it is very difficult to structure one's statements of probabilities (i.e. previsions) in such a way that they both reflect one's beliefs and are coherent (see Kadane and Wolfson 1998). Yet the dichotomy above does not allow for discussion of what sets of previsions may be "very" incoherent or just "slightly" incoherent. This paper explores a remedy for this by studying how quickly an incoherent bookie can be forced to lose money. A faster rate of sure financial decline to the bookie, or a faster rate of guaranteed profit to the gambler, is associated with a greater degree of incoherence.

Mark J. Schervish and Joseph B. Kadane are Professor, Department of Statistics and Teddy Seidenfeld is Professor, Departments of Statistics and Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213.
This research was supported by NSF grant DMS-98-01401 and NSF grant DMS-96-26181.

The problem as stated so far requires some normalization. Suppose that a particular combination of gambles yields a sure loss y for the bookie. Then multiplying each gamble by the same constant $k > 0$ will create a combination of gambles that yields sure loss ky. In this paper we explore how to perform the normalization from the bookie's perspective. We introduce normalizations in Section 3. To fix ideas, however, consider that the bookie cannot be assumed to have infinite resources. A wise gambler would want to be sure that the bookie could cover all the bets. One way to do this would be to require the bookie to escrow the maximum amount that the bookie could lose on each gamble separately. Thus we can ask how much the bookie can be forced to lose for sure, given a specified level of escrow that the bookie can offer. This is the subject of Section 4.

2. GAMBLES AND INCOHERENCE

Think of a random variable X as a function from some space T of possibilities to the real numbers. We assume that, for a bounded random variable X, a bookie might announce some values of x such that he/she finds acceptable all gambles whose net payoff to the bookie is $\alpha(X-y)$ for $\alpha > 0$ and $y < x$. Each such x will be called a *lower prevision for* X. In addition, or alternatively, the bookie may announce some values of x such that the gamble $\alpha(X - y)$ is acceptable when $\alpha < 0$ and $y > x$. These x will be called *upper previsions for* X. We allow that the bookie might announce only upper previsions or only lower previsions or both. For example, if X is the indicator I_A of an event A the bookie might announce that he/she finds acceptable all gambles of the form $\alpha(I_A - x)$ for $x < p$ if $\alpha > 0$ but no other gambles, in particular, not for $x = p$. It will turn out not to matter for any of our results whether or not the bookie finds the gamble $\alpha(I_A - p)$ acceptable. In the special case in which x is both an upper prevision and a lower prevision, we call x a *prevision* of X and denote it $P(X)$. Readers interested in a thorough discussion of upper and lower previsions should refer to Walley (1991).

It will be convenient to assume that, whenever both an upper prevision x^+ and a lower prevision x^- have been assessed for the same random variable X, $x^- \leq x^+$, otherwise the bookie is willing to sell X for a certain price and then buy it right back for a higher price. Although such incoherence could be measured, it requires cumbersome bookkeeping that makes general results difficult to understand. See Examples 4 and 5 for examples of what can happen without this assumption. In particular, this assumption implies that there can be at most one prevision of X.

A collection $x_1, \ldots, x_n$ of upper and/or lower previsions for $X_1, \ldots, X_n$ respectively is *incoherent* if there exists $\epsilon > 0$ and a collection of acceptable gambles $\{\alpha_i(X_i - y_i)\}_{i=1}^n$ such that

$$\sup_{t \in T} \sum_{i=1}^{n} \alpha_i(X_i(t) - y_i) < -\epsilon, \tag{1}$$

in which case we say that a *Dutch Book* has been made against the bookie. Of course, we would need $\alpha_i > 0$ and $y_i < x_i$ if x_i is a lower prevision for X_i and we would need $\alpha_i < 0$ and $y_i > x_i$ if x_i is an upper prevision for X_i. When a collection of upper and/or lower previsions is incoherent, we would like to be able to measure how incoherent they are. As we noted earlier, the ϵ in (1) is not a good measure because we could make ϵ twice as big by multiplying all of the α_i in (1) by 2, but the previsions would be the same. Instead, we need to determine some measure of the sizes of the gambles and then consider the left-hand side of (1) relative to the total size of the combination of gambles. This is what we do in Section 3.

3. NORMALIZATIONS

To begin, consider a single acceptable gamble such as $Y = \alpha(X - y)$. There are a number of possible ways to measure the size of Y. For example $\sup_t |Y(t)|$ or $\sup_t -Y(t)$ might be suitable measures. This last one has a nice interpretation. It is the most that the bookie can lose on the one particular gamble. It measures a gamble by its extreme value in the same spirit as Dutch Book measures incoherence in terms of an extreme value (the minimum payoff to the gambler) of a combination of gambles. Alternatively, if we think of the gambler and bookie as adversaries with regard to this one gamble Y, the gambler might want to be sure that the bookie will be able to pay up when the bet is settled. We could imagine that the gambler requests that the bookie place funds in escrow to cover the maximum possible loss. So, for the remainder of the paper, we will call $e(Y) = \sup_t -Y(t)$ the *escrow for gamble* Y. Note that $e(cY) = ce(Y)$ for all $c > 0$. We use the escrow to measure the size of the gamble Y.

Example 1. Let A be an arbitrary event which is neither certain to occur nor certain to fail. Suppose that a lower prevision p is given, and consider the gamble $Y(t) = \alpha(I_A(t) - p)$ with $\alpha > 0$. Then $\sup_t -Y(t) = \alpha p$, and the escrow is $e(Y) = \alpha p$. If an upper prevision q is given and $\alpha < 0$, then $\sup_t -Y(t) = -\alpha(1 - q) = e(Y)$, where $\alpha < 0$.

When we consider more than one gamble simultaneously, we need to measure the size of the entire collection. We assume that the size of (escrow for) a collection of gambles is some function of the escrows for the individual gambles that make up the collection. That is $e(Y_1, \ldots, Y_n) = f_n(e(Y_1), \ldots, e(Y_n))$. In order for a function to be an appropriate measure of size, we have a few requirements. First,

$$f_n(cx_1, \ldots, cx_n) = cf_n(x_1, \ldots, x_n), \quad \text{for all } c > 0, x_1, \ldots, x_n. \tag{2}$$

Equation (2) says that the function f_n must be homogeneous of degree 1 in its arguments so that scaling up all the gambles by the same amount will scale the escrow by that amount as well. Second, since we are not concerned with the order in which gambles are made, we require

$$\begin{aligned} f_n(x_1, \ldots, x_n) &= f_n(y_1, \ldots, y_n), \text{ for all } n, x_1, \ldots, x_n \\ &\quad \text{and all permutations } (y_1, \ldots, y_n) \text{ of } (x_1, \ldots, x_n). \end{aligned} \tag{3}$$

Third, in keeping with the use of escrow to cover bets, we will require that, if a gamble is replaced by one with higher escrow, the total escrow should not go down:

$$f_n(x_1, \ldots, x_n) \text{ is nondecreasing in each of its arguments.} \tag{4}$$

If a gamble requires 0 escrow, we will assume that the total escrow is determined by the other gambles:

$$f_{n+1}(x_1, \ldots, x_n, 0) = f_n(x_1, \ldots, x_n), \text{ for all } x_1, \ldots, x_n \text{ and all } n. \tag{5}$$

Since nobody can lose more than the sum of the maximum possible losses from all of the accepted gambles, we require that

$$f_n(x_1, \ldots, x_n) \leq \sum_{i=1}^{n} x_i, \qquad \text{for all } n \text{ and all } x_1, \ldots, x_n. \tag{6}$$

Small changes in the component gambles should only produce small changes in the escrow, so we require that

$$f_n \text{ is continuous for every } n. \tag{7}$$

Finally, since we have already decided how to measure the size of a single gamble, we require

$$f_1(x) = x. \tag{8}$$

So, if $Y_1, \ldots, Y_n$ is a collection of gambles, we can set $e(Y_1, \ldots, Y_n) = f_n(e(Y_1), \ldots, e(Y_n))$ for some function f_n satisfying (2)–(8) and call $e(Y_1, \ldots, Y_n)$ an *escrow for the collection of gambles*. Every sequence of functions $\{f_n\}_{n=1}^{\infty}$ that satisfy (2)–(8) leads to its own way of defining escrow. Such a sequence is called an *escrow sequence*. Each function in the sequence is an *escrow function*.

We can find a fairly simple form for all escrow sequences. Combining (8), (4), and (5), we see that $f_n(x_1, \ldots, x_n) \geq \max\{x_1, \ldots, x_n\}$. From (3), we conclude that f_n is a function of the ordered values $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ of $x_1, \ldots, x_n$. That is, $f_n(x_1, \ldots, x_n) = f_n(x_{(1)}, \ldots, x_{(n)})$. Combining these results with (6), we get

$$0 \leq f_n(x_{(1)}, \ldots, x_{(n)}) - x_{(n)} \leq \sum_{i=1}^{n-1} x_{(i)}. \tag{9}$$

Let $\lambda_n(x_{(1)}, \ldots, x_{(n)}) = (f_n(x_{(1)}, \ldots, x_{(n)}) - x_{(n)})/x_{(n)}$ so that

$$f_n(x_{(1)}, \ldots, x_{(n)}) = x_{(n)} \left(1 + \lambda_n(x_{(1)}, \ldots, x_{(n)})\right). \tag{10}$$

In order to satisfy (5), we need $\lambda_n(0, x_{(2)}, \ldots, x_{(n)}) = \lambda_{n-1}(x_{(2)}, \ldots, x_{(n)})$. In order to satisfy (2), λ_n must be invariant under common scale changes for all of its arguments. That is

$$\lambda_n(cx_{(1)}, \ldots, cx_{(n)}) = \lambda_n(x_{(1)}, \ldots, x_{(n)}).$$

Every such function can be written as

$$\lambda_n(x_{(1)}, \ldots, x_{(n)}) = \gamma_n\left(\frac{x_{(1)}}{x_{(n)}}, \ldots, \frac{x_{(n-1)}}{x_{(n)}}\right).$$

In order to satisfy (4), we must have γ_n nondecreasing in each of its arguments. In order to satisfy (9), we must have

$$0 \leq \gamma_n(y_1, \ldots, y_{n-1}) \leq \sum_{i=1}^{n-1} y_i.$$

In summary, every escrow sequence satisfies

$$f_n(x_1, \ldots, x_n) = x_{(n)} \left[1 + \gamma_n\left(\frac{x_{(1)}}{x_{(n)}}, \ldots, \frac{x_{(n-1)}}{x_{(n)}}\right)\right], \tag{11}$$

for some sequence $\gamma_1, \gamma_2, \ldots$ of differentiable functions where $\gamma_1 \equiv 0$ and for $n > 1$ the functions satisfy the following properties:

- $\gamma_n(y_1, \ldots, y_{n-1})$ is defined and continuous for $0 \le y_1 \le y_2 \le \cdots \le y_{n-1} \le 1$,
- γ_n is nondecreasing in each argument,
- $0 \le \gamma_n(y_1, \ldots, y_{n-1}) \le \sum_{i=1}^{n-1} y_i$,
- $\gamma_n(0, y_2, \ldots, y_{n-1}) = \gamma_{n-1}(y_2, \ldots, y_{n-1})$,
- $x[1 + \gamma_n(y_1/x, \ldots, y_{n-1}/x)]$ is nondecreasing in x for all $y_1 \le \cdots \le y_{n-1} \le x$.

(The last condition is equivalent to f_n being nondecreasing in $x_{(n)}$.) It is straightforward to show that every sequence that meets this description satisfies (2)-(8), hence we have characterized escrow sequences.

One simple collection of escrow sequences consists of all sequences in which $\gamma_n(y_1, \ldots, y_{n-1}) = \gamma \sum_{i=1}^{n-1} y_i$ for some common constant $\gamma \in [0, 1]$. In this case, we get a family of escrow functions:

$$f_{\gamma,n}(x_1, \ldots, x_n) = x_{(n)} + \gamma \sum_{i=1}^{n-1} x_{(i)}, \tag{12}$$

for each $0 \le \gamma \le 1$. Another example is $\gamma_n(z_1, \ldots, z_{n-1}) = z_{n-1}$ for $n > 1$. This one makes the total escrow equal to the sum of the two largest individual gamble escrows. Other functions are possible, but we will focus on $f_{\gamma,n}$ for $0 \le \gamma \le 1$. It is easy to see that the two extreme escrow functions correspond to $\gamma = 0$ and $\gamma = 1$:

$$\begin{aligned} f_{0,n}(x_1, \ldots, x_n) &= x_{(n)}, \\ f_{1,n}(x_1, \ldots, x_n) &= \sum_{i=1}^{n} x_i. \end{aligned}$$

We now propose to measure the incoherence of a collection of incoherent previsions based on a normalization by an escrow. For a combination of gambles $Y = \sum_{i=1}^{n} \alpha_i(X_i - y_i)$, define the *guaranteed loss* to be $G(Y) = -\min\{0, \sup_{t \in T} Y(t)\}$. So, Dutch Book can be made if there exists a combination of acceptable gambles whose guaranteed loss is positive. The *rate of guaranteed loss* relative to a particular escrow function f_n is

$$H(Y) = \frac{G(Y)}{f_n(e(Y_1), \ldots, e(Y_n))}, \tag{13}$$

where $Y_i = \alpha_i(X_i - y_i)$. Notice that the rate of guaranteed loss is unchanged if all α_i are multiplied by a common positive number. Also, the rate of guaranteed loss is interesting only when Dutch Book is made, otherwise the numerator is 0. The denominator $f_n(e(Y_1), \ldots, e(Y_n))$ is 0 if and only if $e(Y_i) = 0$ for all i. This will occur if and only if the agent who is required to escrow cannot lose any of the individual gambles, in which case the numerator is 0 as well, and we will then define the rate of guaranteed loss to be 0 (since we cannot guarantee loss). The *extent of incoherence* relative to an escrow and corresponding to a collection of previsions will be the supremum of $H(Y)$ over all combinations Y of acceptable gambles. If the previsions are incoherent then the maximum rate of guaranteed loss is positive, otherwise it is 0.

There is a slightly simpler way to compute the extent of incoherence corresponding to a finite set of previsions than directly from the definition.

Theorem 1. Let $x_1, \ldots, x_n$ be a collection of incoherent upper and/or lower previsions for $X_1, \ldots, X_n$. Define

$$\begin{aligned} g(\alpha_1, \ldots, \alpha_n) &= \sup_{t \in T} \sum_{i=1}^{n} \alpha_i (X_i(t) - x_i), \\ h(\alpha_1, \ldots, \alpha_n) &= f_n(e(\alpha_1[X_1 - x_1]), \ldots, e(\alpha_n[X_n - x_n])). \end{aligned}$$

Then the rate of incoherence is

$$\sup_{\alpha_1, \ldots, \alpha_n} \frac{-g(\alpha_1, \ldots, \alpha_n)}{h(\alpha_1, \ldots, \alpha_n)} \tag{14}$$

or equivalently

$$-\inf_{\alpha_1, \ldots, \alpha_n} g(\alpha_1, \ldots, \alpha_n), \text{subject to } h(\alpha_1, \ldots, \alpha_n) \leq 1. \tag{15}$$

The supremum and infimum are taken over those α_i that have the appropriate signs.

As with all of the more lengthy proofs in this paper, the proof of Theorem 1 is in an appendix, Appendix A in this case. Theorem 1 allows us to ignore the fact that the gamble $\alpha(X - x)$ might not be acceptable when x is a lower or upper prevision for X if we are proving results concerning the rate of incoherence.

Note that if a collection of gambles satisfies the escrow condition $h(\alpha_1, \ldots, \alpha_n) \leq 1$, then every subcollection also satisfies the escrow condition because of (5). Also, note that, since every escrow function f_n is between $f_{0,n}$ and $f_{1,n}$, the maximum and minimum possible rates of incoherence correspond to these two escrows.

4. GENERAL THEOREMS

When we use the bookie's escrow with $e(Y) = \sup_t -Y(t)$ for each individual gamble Y, we call the extent of incoherence the *the maximum rate of guaranteed loss*, since the extent of incoherence is the maximum rate at which the bookie can be forced to lose relative to the particular escrow chosen. We focus on the family of escrows $f_{\gamma,n}$ defined in (12). The corresponding maximum rates of guaranteed loss will be donoted ρ_γ.

For n gambles with escrows $x_1, \ldots, x_n$, the escrow is $x_{(n)} + \gamma \sum_{i=1}^{n-1} x_{(i)}$ and an upper bound on the loss is $\sum_{i=1}^{n} x_{(i)}$, so

$$\rho_\gamma \leq \frac{\sum_{i=1}^{n} x_{(i)}}{x_{(n)} + \gamma \sum_{i=1}^{n-1} x_{(i)}},$$

which is easily seen to be bounded by $n/[(n-1)\gamma + 1]$, an increasing function of n with limit $1/\gamma$ as $n \to \infty$. The upper bound $n/[(n-1)\gamma + 1]$ can actually be achieved by ρ_γ in special cases.

Example 2. Let $X_i(t) = c_i$ for all t be constant random variables with lower previsions $p_i > c_i$ for $i = 1, \ldots, n$. For each acceptable gamble $\alpha_i(X_i - y_i)$ with $c_i < y_i < p_i$ and $\alpha_i > 0$, $\inf_t \alpha(X_i(t) - y_i) = (c_i - y_i)\alpha_i$, since this is the only possible value. The necessary escrow is $(y_i - c_i)\alpha_i$. Let $x_i = (p_i - c_i)\alpha_i$. Then $g(\alpha_1, \ldots, \alpha_n) = \sum_{i=1}^{n} x_i$ for all $\alpha_1, \ldots, \alpha_n > 0$ and $h(\alpha_1, \ldots, \alpha_n) = x_{(n)} + \gamma \sum_{i=1}^{n-1} x_{(i)}$. By choosing the α_i proportional to $1/(p_i - c_i)$, all x_i become the same number and ρ_γ equals $n/[(n-1)\gamma + 1]$. A similar result holds if upper previsions $p_i < c_i$ are given instead of lower previsions for some or all of the constants.

Notice that the global upper bound on ρ_γ is 1 for $\gamma = 1$, and this can be achieved in Example 2 even with $n = 1$. The escrow based solely on the maximum, ρ_0, has no global upper bound. Indeed, in Example 2, $\rho_0 = n$ so that ρ_0 increases without bound as there become more incoherent previsions for constants.

As Example 2 illustrates, constant random variables are very special. If $X = c$ is constant, and the bookie assigns an upper and/or lower prevision p on the correct side(s) of c, then the gamble $\alpha(X-p)$ is a constant nonnegative number whenever it is acceptable. On the other hand, if the bookie assigns an upper or lower prevision on the wrong side of c, no matter how close to c, he/she will lose the entire escrow for sure. In fact, if a collection of previsions includes even one incorrect upper or lower prevision for a constant, then $\rho_1 = 1$ is obtained by ignoring all the other previsions and placing the appropriate bet on the constant. This all-or-nothing behavior occurs only with constant random variables, as we illustrate next.

Example 3. Let X take at least two different values, and let $x^m = \inf_t X(t)$ and $x^M = \sup_t X(t)$. Suppose that $p^- > x^M$ is an incoherent lower prevision for X. If $\alpha > 0$, the escrow for gamble $\alpha(X - p^-)$ is $\alpha(p^- - x^m)$. This implies that $\rho_\gamma = (p^- - x^M)/(p^- - x^m) < 1$. Similarly, if $p^+ < x^m$ is an incoherent upper prevision and $\alpha < 0$, the escrow is $\alpha(x^M - p^+)$ and $\rho_\gamma = (x^m - p^+)/(x^M - p^+) < 1$. Note also that $\rho_\gamma < 1$ in these cases no matter how far p^+ and/or p^- are away from the interval $[x^m, x^M]$. Contrast this with the situation in Example 2 in which $\rho_\gamma = 1$ no matter how close the incoherent upper or lower prevision for a constant $X = c$ is to c.

For proving theoretical results, it is convenient to be able to assume that upper and lower previsions have been announced for all random quantities under consideration. Fortunately, it is possible to do this without requiring the bookie to accept gambles that might be considered unfavorable. Suppose that x_+ is an upper prevision for X. Let $x_- = \min\{x_+, \inf_t X(t)\}$. Suppose that we pretend as if x_- is a lower prevision for X. The gambles that this commits the bookie to accept are $\alpha(X - y)$ for $\alpha > 0$ and $y < x_-$. For all t, $\alpha(X(t) - y) > 0$, so the bookie will not have to escrow for such a gamble and he/she will never lose this gamble. Hence, it cannot contribute to the degree of incoherence if we assume that x_- is a lower prevision for X. Similarly, if x_- is an announced lower prevision, then $x_+ = \max\{x_-, \sup_t X(t)\}$ can be taken to be an upper prevision for X without loss of generality.

There are a few other upper or lower previsions that can be inferred from those announced. For example, if x_- is an announced lower prevision for X, let $Y = c - X$ for some constant c and let $y_+ = c - x_-$. Since $\alpha(X(t) - [x_- - \epsilon]) = -\alpha(Y(t) - [y_+ + \epsilon])$ for all t, we have that y_+ is an upper prevision for Y even if the bookie did not realize it. Also, the escrow is the same for both sides of this equation because the values are equal for all t. Similarly, if x_+ is an announced upper prevision for X then $c - x_+$ is a lower prevision for $c - X$. For the case in which $X = I_A$, the indicator of an event A, we can let $c = 1$ and we have that p_- is a lower prevision for A if and only if $1 - p_-$ is an upper prevision for A^C. There is an interesting extension of this to disjoint events for the case of ρ_1.

Result 1. Let $A_1, \ldots, A_m$ be disjoint events with lower previsions $p_1, \ldots, p_m \geq 0$. Let $B = (\cup_{i=1}^m A_i)^C$. Then, the value of ρ_1 is unchanged if we act as if if $q = 1 - \sum_{i=1}^m p_i$ were an upper prevision for B.

Proof. For $\alpha < 0$, write

$$\begin{aligned} \alpha(I_B - [q+\epsilon]) &= \alpha\left(1 - \sum_{i=1}^{m} I_{A_i} - 1 + \sum_{i=1}^{m}\left[p_i - \frac{\epsilon}{m}\right]\right) \\ &= \sum_{i=1}^{m} -\alpha\left(I_{A_i} - \left[p_i - \frac{\epsilon}{m}\right]\right), \end{aligned} \tag{16}$$

which is acceptable since $-\alpha > 0$ and the p_i are lower previsions. The escrow for the far left-hand side of (16) is $-\alpha(1-q-\epsilon) = -\sum_{i=1}^{m} \alpha[p_i - \epsilon/m]$, which is the same as the escrow for the far right-hand side. Hence, even if the bookie does not realize it, he/she is willing to accept all gambles of the form $\alpha(I_B - y)$ for $\alpha > 0$ and $y > q$, hence q is an lower prevision for B.

For the case in which the collection of gambles concerns a partition of the space of possibilities, we have the following result, whose proof is in Appendix B.

Theorem 2. Let $\{A_i\}_{i=1}^{n}$ be a finite partition of the set T of possibilities, with $n > 1$ and all A_i nonempty. Let $q_1, \ldots, q_n$ be upper previsions and let $p_1, \ldots, p_n \geq 0$ be lower previsions for the partition elements. Let $s^+ = \sum_{i=1}^{n} q_i$ and let $s^- = \sum_{i=1}^{n} p_i$. Let $p_{(1)}, \ldots, p_{(n)}$ and $q_{(1)}, \ldots, q_{(n)}$ denote the ordered values of the p_i and q_i respectively.

1. If $s^- > 1$, then
$$\rho_\gamma = \frac{s^- - 1}{p_{(n)} + \gamma \sum_{i=1}^{n-1} p_{(i)}},$$
and all α_j equal to each other achieves the inf in (15). If $0 < p_i < 1$ for all i, then the $\alpha_1, \ldots, \alpha_n$ that achieve the inf in (15) are unique.

2. If $s^+ < 1$, then
$$\rho_\gamma = \frac{1 - s^+}{1 - q_{(1)} + \gamma\left[n - 1 - \sum_{i=2}^{n} q_{(i)}\right]},$$
and α_j all equal to each other achieves the inf in (15). If $0 < q_i < 1$ for all i, then the $\alpha_1, \ldots, \alpha_n$ that achieve the inf in (15) are unique.

Note that ρ_1 is strictly increasing in s^- for $s^- > 1$ and is strictly decreasing in s^+ for $s^+ < 1$. Since lower previsions are never larger than upper previsions, it is easy to see that it is impossible for both $s^- > 1$ and $s^+ < 1$.

The following example illustrates why it is convenient to assume that upper previsions are always at least as large as lower previsions.

Example 4. Let A_1, A_2, A_3 be a partition of T and let $p_1 = 0.5$, $p_2 = 0.4$, and $p_3 = 0.4$ be lower previsions for the three events. Let $q_1 < 0.5$ be an upper prevision for A_1. The most general combination of gambles to consider is

$$\alpha_1(I_{A_1} - 0.5) + \alpha_2(I_{A_2} - 0.5) + \alpha_3(I_{A_3} - 0.4) - \alpha_4(I_{A_1} - q_1),$$

with all $\alpha_i \geq 0$. The sum of the escrows for this combination is

$$0.5\alpha_1 + 0.4\alpha_2 + 0.4\alpha_3 + (1 - q_1)\alpha_4. \tag{17}$$

Subject to (17) equaling 1, the negative of (14) becomes

$$\max\{\alpha_1, \alpha_2 + \alpha_4, \alpha_3 + \alpha_4\} - 1, \tag{18}$$

which we can now try to minimize. It is not difficult to see that we need $\alpha_1 = \alpha_2 + \alpha_4 = \alpha_3 + \alpha_4$ in order to minimize (18). Combining these equations with (17) gives

$$\alpha_2 = \alpha_3 = \frac{1 - (1.5 - q_1)\alpha_1}{q_1 - 0.2}, \quad \alpha_4 = \frac{1.3\alpha_1 - 1}{q_1 - 0.2},$$

and we are trying to minimize α_1 subject to all four $\alpha_i \geq 0$. If $q_1 < 0.2$, we need $\alpha_1 = \alpha_4 = 1/(1.5 - q_1)$ and $\alpha_2 = \alpha_3 = 0$. This corresponds to gambling only using the upper and lower previsions on A_1. If $q_1 > 0.2$, we need $\alpha_1 = \alpha_2 = \alpha_3 = 1/1.3$ and $\alpha_4 = 0$. this corresponds to ignoring the upper prevision on A_1 altogether. For cases with more than three events and more than one upper prevision less than the corresponding lower prevision, we must solve a general linear programming problem.

There is one special case in which we can determine the extent of incoherence even when lower previsions are greater than upper previsions.

Example 5. Assume that there is only one event A of interest and $q < p$ where q is an upper prevision for A and p is a lower prevision. In this case, we can write $\alpha(I_A - q) = -\alpha(I_A^C - (1 - q))$ and treat $1 - q$ as a lower prevision for A^C in the spirit of Result 1. Then we have a partition $\{A, A^C\}$ of T with lower previsions p and $1 - q$ that add up to more than 1. Alternatively, we can treat $1 - p$ as an upper prevision for A^C and then we have a partition of T with upper previsions q and $1 - p$ that add up to less than 1. These are precisely the situations that are covered by Theorem 2. Notice that Theorem 2 gives the same value of ρ_γ regardless of whether we treat p and $1 - q$ as incoherent lower previsions or we treat $1 - p$ and q as incoherent upper previsions. The common value can be expressed as

$$\rho_\gamma = \frac{p - q}{\max\{p, 1 - q\} + \gamma \min\{p, 1 - q\}}.$$

Notice that ρ_γ increases as p increases for fixed q or as q decreases for fixed p. So, the extent of incoherence increases as the incoherent previsions become even more incoherent.

In the next section, we show how Theorem 2 applies to the example of hypothesis testing at a fixed level.

5. TESTING SIMPLE HYPOTHESES AT A FIXED LEVEL

Lindley (1972, p. 14) argues that it is incoherent to test all hypotheses at the same level, such as .05. (See also Seidenfeld, Schervish, and Kadane 1990.) Cox (1958) gave an example of how testing all hypotheses at the same level leads to inadmissibility. In this section, we show how this incoherence and inadmissibility can be measured using the measure of incoherence ρ.

Consider the case of testing a simple hypothesis against a simple alternative. Let f_0 and f_1 be two possible densities for a random quantity X, and let f be the "true" density of X. Suppose that we wish to test the hypothesis $H_0 : f = f_0$ versus the alternative $H_1 : f = f_1$. To write this as a decision problem, let the parameter space and the action space both be $\{0,1\}$ where action $a=0$ corresponds to accepting H_0 and action $a=1$ corresponds to rejecting H_0. Also, parameter i corresponds to $f=f_i$ for $i=0,1$. Let the loss function have the form

$$L(i,a)=\begin{cases} c_i & \text{if } f=f_i \text{ and } a=1-i,\\ 0 & \text{otherwise,}\end{cases} \tag{19}$$

with $c_0, c_1>0$. The Neyman-Pearson lemma says that the most powerful tests of their sizes and the Bayes rules with respect to all priors have the form: For some constant k, choose $a=1$ if $f_1(x)>kf_0(x)$, choose $a=0$ if $f_1(x)<kf_0(x)$, and do whatever you wish (even randomization) if $f_1(x)=kf_0(x)$. Now, suppose that someone chooses a value of k and declares that they prefer the corresponding test to all other tests. One could infer from this choice an "implied prior" over the two possibilities f_0 and f_1. If $\Pr(f=f_0)=p$ and $\Pr(f=f_1)=1-p$, then the Bayes rule is to choose $a=1$ if $pc_0f_0(x)<(1-p)c_1f_1(x)$, which corresponds to $k=pc_0/[(1-p)c_1]$. So $p=c_1k/(c_0+c_1k)$.

Of course, a classical statistician who refuses to use prior and posterior probabilities will not acknowledge the implied prior. However, incoherence will arise if two tests about the same parameter imply different priors. We illustrate this with a version of the example of Cox (1958). Since the only part of the loss function that matters is c_0/c_1, let $c_1=1$. As an example, let f_0 and f_1 be normal distributions with different means θ but the same variance σ^2. Suppose that the hypothesis is $H_0:\theta=0$ versus $H_1:\theta=1$ with $c_0=1$. (The phenomenon we illustrate here applies more generally as shown in Theorem 3.) Suppose that either $\sigma=1$ or $\sigma=0.3$ will be true, but we will not know which until we observe the data. That is, the data consist of the pair (X,σ). Let $\Pr(\sigma=1)=0.5$, so that σ is ancillary. A classical statistician who prefers level 0.05 tests whenever available might think that, after observing σ a conditional level 0.05 test should still be preferred to a test whose conditional level given σ is something else. The most powerful conditional level 0.05 test is to reject $H_0:\theta=0$ if $X>1.645\sigma$. The most powerful marginal level 0.05 test rejects H_0 if $X>0.5+0.9438\sigma^2$ and is the Bayes rule with respect to the prior $\Pr(\theta=0)=0.7199$. The marginal power of the Bayes rule is 0.6227, while the marginal power of the conditional level 0.05 test is 0.6069. Since both tests have the same level, the conditional test is inadmissible.

To see how this inadmissibility translates into incoherence, we interpret the preference of one test δ_1 to another δ_2 as a preference for suffering a loss equal to the risk function of δ_1 to suffering a loss equal to the risk function of δ_2. The risk function of a test δ is

$$R(\theta,\delta)=\begin{cases} c_0 \text{ times size of test } \delta & \text{if } \theta=0,\\ \text{one minus power of test } \delta & \text{if } \theta=1.\end{cases}$$

To say that δ_1 is preferred to δ_2 means that $R(\theta,\delta_2)-R(\theta,\delta_1)$ is an acceptable gamble. In our example, let $\alpha_\delta(\sigma)$ and $\beta_\delta(\sigma)$ denote the size and power of test δ conditional on σ. Also, let $\beta_{cl}(\sigma)$ denote the power of the most powerful level 0.05 test. Then, for each σ, the classical statistician prefers the level 0.05 test to every other test. So, for each σ and all δ that are not the most powerful level 0.05 test, the following gamble is acceptable, even favorable:

$$\begin{aligned} R(\theta,\delta)-R(\theta,\delta_{cl}) &= \begin{cases} (\alpha_\delta(\sigma)-.05)c_0 & \text{if } \theta=0,\\ \beta_{cl}(\sigma)-\beta_\delta(\sigma) & \text{if } \theta=1\end{cases}\\ &= a(I_A-b), \end{aligned} \tag{20}$$

where

$$
\begin{aligned}
A &= \{\theta = 0\}, \\
a &= (\alpha_\delta(\sigma) - .05)c_0 + \beta_\delta(\sigma) - \beta_{cl}(\sigma), \\
b &= \frac{\beta_\delta(\sigma) - \beta_{cl}(\sigma)}{(\alpha_\delta(\sigma) - .05)c_0 + \beta_\delta(\sigma) - \beta_{cl}(\sigma)}.
\end{aligned}
$$

In other words, b is an upper or lower prevision for A depending on whether $a < 0$ or $a > 0$.

We can make use of the construction in (20) to obtain a general result. Theorem 3 has a technical condition (concerning risk sets) that is known to be satisfied for problems of testing simple hypotheses against simple alternatives using fixed sample size and sequential tests. For more detail on risk sets, see Sections 3.2.4 and 4.3.1 of Schervish (1995).

Theorem 3. Let θ be a parameter and let the parameter space Ω consist of two points $\{0, 1\}$. Consider two decision problems D_0 and D_1 both with the same parameter space Ω and with nonnegative loss functions L_0 and L_1. Let the data in problem D_i be denoted X_i. Suppose that the risk sets for the two decision problems are closed from below. Suppose that an agent prefers the admissible decision rule δ_i to all others in problem D_i for $i = 0, 1$. For each decision rule ψ in problem D_i, let $R_i(\theta, \psi)$ denote the risk function. Let $A = \{\theta = 0\}$ and define

$$
\begin{aligned}
a_i(\psi) &= [R_i(0, \psi) - R_i(0, \delta_i)] + [R_i(1, \delta_i) - R_i(1, \psi)], \\
b_i(\psi) &= \frac{R_i(1, \delta_i) - R_i(1, \psi)}{[R_i(0, \psi) - R(0, \delta_i)] + [R_i(1, \delta_i) - R_i(1, \psi)]}.
\end{aligned}
\tag{21}
$$

If ψ is admissible in problem D_i and is not equivalent to δ_i, then

$$
R_i(\theta, \psi) - R_i(\theta, \delta_i) = a_i(\psi)(I_A - b_i(\psi)). \tag{22}
$$

If δ_0 and δ_1 are not Bayes rules with respect to a common prior, then there exist real numbers d_0 and d_1 and decision rules ψ_0 (in problem D_0) and ψ_1 (in problem D_1) such that the two gambles $d_0 a_0(\psi_0)(I_A - b_0(\psi_0))$ and $d_1 a_1(\psi_1)(I_A - b_1(\psi_1))$ are both acceptable, but

$$
d_0 a_0(\psi_0)(I_A - b_0(\psi_0)) + d_1 a_1(\psi_1)(I_A - b_1(\psi_1)) < 0. \tag{23}
$$

Also,

$$
\rho_\gamma = \sup_{p_0 > p_1} \frac{p_0 - p_1}{\max\{p_0, 1 - p_1\} + \gamma \min\{p_0, 1 - p_1\}}, \tag{24}
$$

where the supremum is over all $p_0 > p_1$ such that either δ_i is a Bayes rule with respect to prior p_i for $i = 0, 1$ or δ_i is a Bayes rule with respect to prior p_{1-i} for $i = 0, 1$.

As an example of Theorem 3, return to the test of $H_0 : \theta = 0$ versus $H_1 : \theta = 1$ where $X \sim N(\theta, \sigma^2)$ with σ being one of two known values. Let σ_i be the value of σ for problem D_i for $i = 0, 1$. The implied prior probability of $\{\theta = 0\}$ is $k/(c_0 + k)$ where k comes from the form of the Neyman-Pearson test and c_0 comes from the loss function (19). If a classical statistician decides to test the hypothesis $H_0 : \theta = 0$ at level 0.05

regardless of the value of σ, this will often require choosing two different values of k in the Neyman-Pearson lemma. In fact, $k = \exp(1.645/\sigma - 0.5/\sigma^2)$, and the implied prior is

$$p(\sigma) = \left[1 + c_0 \exp\left(\frac{-1.645}{\sigma} + \frac{0.5}{\sigma^2}\right)\right]^{-1}, \tag{25}$$

not a one-to-one function of σ, but highly dependent on σ. In this problem, the lower boundary of the risk set is a strictly convex differentiable function so that every point on the lower boundary has a unique support line and hence a unique prior such that the corresponding rule is Bayes with respect to that prior. This means that the sup in (24) is unnecessary and

$$\rho_\gamma = \frac{p(\sigma_0) - p(\sigma_1)}{\max\{p(\sigma_0), 1 - p(\sigma_1)\} + \gamma \min\{p(\sigma_0), 1 - p(\sigma_1)\}},$$

if $p(\sigma_0) > p(\sigma_1)$ with a similar formula if $p(\sigma_0) < p(\sigma_1)$. For the case of $\gamma = 1$, this simplifies to

$$\rho_1 = \frac{|p(\sigma_0) - p(\sigma_1)|}{1 + |p(\sigma_0) - p(\sigma_1)|},$$

In this case, the degree of incoherence is a simple monotone function of how far apart the implied priors are for the two level 0.05 tests. Theorem 3 would also apply if one or both of the two decision problems were a sequential decision problem in which the loss equals the cost of observations plus the cost of terminal decision error.

It is interesting to examine the relationship between ρ_γ and the pair (σ_0, σ_1). For example, suppose that $\sigma_i = 2/\sqrt{n_i}$ for two different values n_0 and n_1. This would correspond to the data consisting of a sample $X_1, \ldots, X_{n_i}$ of independent normal random variables with mean θ and variance 4, where the sample size is n_i in decision problem D_i. Figure 1 is a plot of ρ_γ as a function of (n_0, n_1) for the case $\gamma = 1$ and $c_0 = 19$. Other values of γ produce plots with similar appearances. Of course the values of ρ_γ are higher for $\gamma < 1$. We chose $c_0 = 19$ to correspond to the classical choice of $\alpha = 0.05$. Each curve in Figure 1 corresponds to a fixed value of n_0 and lets n_1 vary from 1 to 150. Notice that each curve touches 0 where $n_1 = n_0$ since there is no incoherence in that case. Some of the curves come close to 0 in another location as well. For example, the $n_0 = 27$ curve comes close to 0 near $n_1 = 2$ and the $n_0 = 2$ curve comes close to 0 near $n_1 = 27$. The reason is that the implied priors corresponding to $\sigma = 2/\sqrt{2}$ and $\sigma = 2/\sqrt{27}$ are nearly the same (0.7137 and 0.7107 respectively), making these two level 0.05 tests nearly coherent. Indeed, the entire curves corresponding to $n_0 = 2$ and $n_0 = 27$ are nearly identical for this same reason. Another interesting feature of Figure 1 is that all of the curves are rising as $n_1 \to \infty$ but not to the same level. As $n_1 \to \infty$, the implied prior on $A = \{\theta = 0\}$ converges to 0. But if n_0 is large also, then the implied prior corresponding to $\sigma = 2/\sqrt{n_0}$ is also close to 0. For example, with $n_0 = 100$, the implied prior is 7.3×10^{-4}. There is not much room for incoherence between 0 and 7.3×10^{-4}, so the curve corresponding to $n_0 = 100$ will not rise very high. On the other hand, with $n_0 = 11$, the implied prior is 0.1691, leaving lots of room for incoherence. In fact, since 0.1691 is the largest possible implied prior in this example, all of the other curves have local maxima near $n_1 = 11$, and the $n_0 = 11$ curve rises higher than all the others as n_1 increases. Since the limiting implied prior is 0 as $n_1 \to \infty$, the height to which the n_0 curve rises as n_1 increases is

$$\frac{|p(2/\sqrt{n_0}) - 0|}{1 + |p(2/\sqrt{n_0}) - 0|} = \left[2 + 19 \exp\left(-0.8224\sqrt{n_0} + 0.125 n_0\right)\right]^{-1}.$$

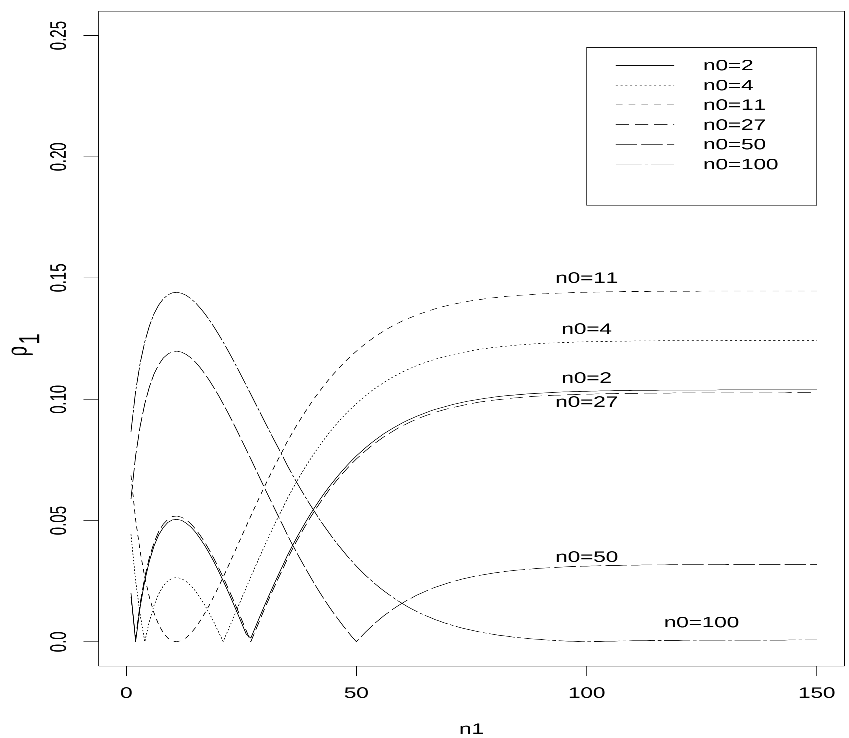

Figure 1. Plot of ρ_1 for level 0.05 testing as a function of n_1 (running from 1 to 150) for various values of n_0 with $c_0 = 19$.

The curve corresponding to $n_0 = 4$ illustrates the original example of Cox (1958), in which the alternative is that θ equals the larger of the two different standard deviations.

Lehmann (1958) offered a rule of thumb for choosing tests based on both their size and their power. One chooses a postive number λ (such as c_0 in the loss function (19)) and then chooses the test so that the probability of type II error equals λ times the probability of type I error. In our case of testing one normal distribution against another one with the same variance σ^2, this procedure will produce the minimax rule with loss (19) if $\lambda = c_0$. When $\lambda = 1$, it is easy to check that Lehmann's suggestion is the Bayes rule with respect to the prior with $\Pr(\theta = 0) = 1/(1 + c_0)$ for all σ. In this special case $\rho_\gamma = 0$ for all γ. However, when $\lambda \neq 1$, each σ leads to a Bayes rule with respect to a different implied prior. Assuming that the test will be to reject H_0 if $X > y$, one must solve the equation

$$\lambda\Phi\left(-\frac{y}{\sigma}\right) = \Phi\left(\frac{y-1}{\sigma}\right). \tag{26}$$

The implied prior, assuming that the loss is still (19), is then

$$p_L(\sigma) = \left[1 + c_0 \exp\left(\frac{0.5 - y}{\sigma^2}\right)\right]^{-1}. \tag{27}$$

When $\lambda = 1$, $y = 1/2$ solves (26). Plugging this into (27) yields $p_L(\sigma) = 1/(1 + c_0)$ for all σ as we noted earlier. Two other limiting cases are of interest. If $\sigma \to \infty$, then y/σ must converge to $\Phi^{-1}(\lambda/[1+\lambda])$ in order for (26) to hold. This would make the type I error probability $1/(1+\lambda)$, and the limit of $p_L(\sigma)$ would be $1/(1+c_0)$. It is not difficult to see that the type I error probability is highest for $\sigma = \infty$, so it must be less than $1/(1+\lambda)$

for all finite σ. If $\sigma \to 0$, then $(y - 1/2)/\sigma^2$ must converge to $\log(\lambda)$ in order for (26) to hold. In this case, $p_L(\sigma)$ converges to $\lambda/(\lambda + c_0)$. For the case of $\lambda = c_0 = 19$, Figure 2 shows the value of ρ_1 with $\sigma_i = 2/\sqrt{n_i}$ for $i = 0, 1$ for various values of n_0 and n_1 in the

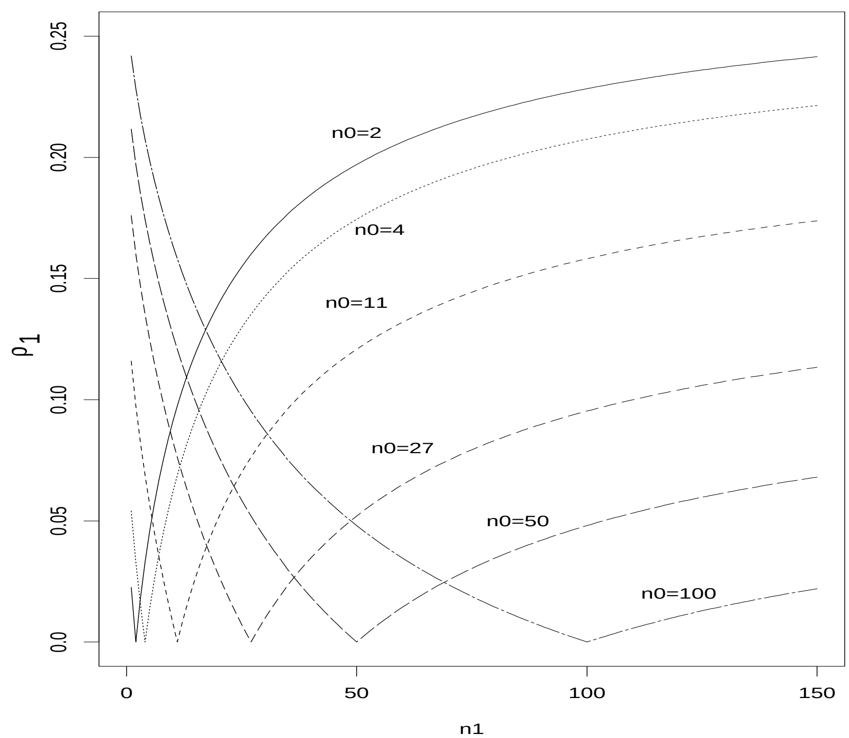

Figure 2. Plot of ρ_1 for minimax rule with $c_0 = 19$ as a function of n_1 (running from 1 to 150) for various values of n_0.

same spirit (and on the same vertical scale) as Figure 1.

The curves in Figure 2 are higher for large n_1 than the corresponding curves in Figure 1. This means that, when $c_0 = 19$, Lehmann's procedure with $\lambda = 19$ is more incoherent (as measured by ρ_1) for large values of n_1 than testing at level 0.05. Lehmann (1958) made his suggestion for testing, not to be more coherent than fixed level testing, but rather to avoid a different problem exhibited by fixed level testing. Testing all hypotheses at the same level, regardless of how much data one has, allows the probability of type II error to become much smaller than the probability of type I error as the sample size increases. This amounts to behaving as if the null hypothesis were not very important compared to the alternative. Indeed, the fact that the implied prior goes to zero as the sample size increases reflects this fact. Lehmann's procedure forces the type I and type II errors to decrease together as sample size increases, thereby making sure that both the null and the alternative remain important as the sample size increases. In fact, the implied prior approaches a value strictly between 0 and 1 as sample size increases. What makes Lehmann's procedure less coherent than fixed level testing is the rate at which the implied prior approaches its limit as sample size increases. For Lehmann's procedure, the implied prior differs from its limit by approximately a constant divided by sample size whereas the implied prior for a fixed level test test differs from 0 by approximately $\exp(-cn)$ for some constant c. In this simple testing problem, Lehmann's procedure with $\lambda = 1$ leads to coherent choices of admissible tests for all sample sizes. Lehmann's procedure with

$\lambda = 1$ here corresponds to an implied prior for the null hypothesis of $1/(1+c_0) = 0.05$ when $c_0 = 19$, and an implied prior of $1/2$ when losses are equal ($c_0 = 1$). As we noted, Lehmann's rule gives the minimax risk solution for $\lambda = c_0$. However, as Lindley (1972, p.14) points out, it is not guaranteed that minimax risk solutions from different families of admissible tests correspond to the same Bayes model. In our testing problem, this is what happens with Lehmann's rule when $\lambda \neq 1$, which explains how it suffers a positive degree of incoherence. An alternative procedure to Lehmann's which also lets type I and type II error probabilities decrease as sample size increases, but which is coherent, is to minimize a positive linear combination of those error probabilities.

6. SUMMARY

In this article we introduce a familiy of indices of incoherence of previsions, based on the gambling framework of de Finetti (1974). When a bookie is incoherent, a gambler can choose a collection of gambles acceptable to the bookie that result in a sure loss to the bookie (and a sure gain to the gambler). That is, the gambler can make a *Dutch Book* against the bookie. Our index of incoherence in the bookie's previsions is the maximum guaranteed rate of loss to the bookie that the gambler creates through his/her choice of coefficients, relative to the bookie's escrow. Throughout, we mean by "escrow" an amount needed to cover the bookie's possible losses as developed in Section 3.

In Section 4 we consider how our family of indices applies to combinations of lower and upper previsions made by the bookie on finitely many random variables X_i, $i = 1, \ldots, n$. These one-sided offers from the bookie apply to restrict the gambler (respectively) to positive or negative coefficients. De Finetti's theory, then, is the special case where, for each X_i, the bookie has equal lower and upper previsions.

We focus attention on the case of events over a finite partition. That is, X_i is the indicator function for event A_i. Here, for example, if each pair of the gambler's lower and upper previsions (p_i and q_i, respectively) for A_i is coherent on its own, i.e., $0 \leq p_i \leq q_i \leq 1$, then incoherence arises because the sum s^- (s^+) of the lower (upper) previsions is greater (less) than 1. Theorem 2 reports the gambler's maximin strategy in this game, maximizing the guaranteed minimum rate of loss to the bookie in such a case. Simply, choose equal coefficients with the sign of the coefficient determined by which case of incoherence obtains. It is worth noting that, either as s^- increases above 1 or as s^+ decreases below 1, the maximum rate of loss grows. The more the sum of the bookie's lower or upper previsions depart from their coherent bound, 1, the greater is this index of incoherence.

In Section 5, we apply this idea to identify the degrees of incoherence in two policies for testing simple hypotheses. First, we consider testing at a level that is fixed regardless of the sample size, as in the example of Cox (1958). We show, through a trade of risks, how the gambler can make a "Dutch Book" against a statistician who follows such a testing policy. That is, our index of incoherence coincides with the extent to which the fixed alpha level tests can be dominated by combinations of other tests.

When tests are based on small sample sizes, the degree of incoherence in a fixed-level testing policy is complicated, as illustrated in Figure 1. However, the degree of incoherence between two such tests decreases as the sample sizes for these tests increases. Nonetheless, we do not find this fact sufficient to justify the policy, even with large samples, because the statitstician's near-to-coherent behavior then requires treating one of the hypotheses as practically impossible. That is, the Bayes model that the fixed level testing policy approaches with increasing sample size assigns probability 0 to the null hypothesis. Why bother to collect data if that is your behavioral policy? Obviously, mere coherence of a policy is not sufficient to make it also a reasonable one!

A second testing policy that we examine is due to Lehmann (1958), who proposes admissible tests based on a fixed ratio of the two risks involved, i.e., with a fixed ratio of type I and type II errors denoted by his parameter λ. Except for the case in which that ratio is 1, this too proves to be an incoherent policy for testing two simple hypotheses. Figure 2 shows the plot of the degree of incoherence for Lehmann's rule ($\lambda = 19$) applied to tests with differing sample sizes. Surprisingly, even in a comparison of two tests based on large sample sizes, Lehmann's policy is sometimes more incoherent by our standards than the fixed .05 level policy for the same two sample sizes. Thus, in order to gain the benefits of approximate coherence, it is neither necessary nor sufficient merely to shrink the level of tests with increasing sample sizes, as happens with Lehmann's rule. In tests based on increasing sample sizes Lehmann's policy (λ fixed) is approximately coherent against a Bayes model that assigns equal prior probability to each of the two hypotheses, the implied priors converge to 1/2. Of course, for that prior, the choice of $\lambda = 1$ in Lehmann's rule assures exact coherence at all sample sizes!

Our work on degrees of incoherence, illustrated here with an analysis of testing simple statistical hypotheses, indicates the importance of having finer distinctions than are provided by deFinetti's dichotomy between coherent and incoherent methods. We see the interesting work of Nau (1989, 1992) providing useful algorithms for computing, e.g., instances of our Lemma 1. His approach is more flexible than that described here in that he allows, in addition, *confidence weights* attached to each upper or lower prevision. These permit the bookie to limit the amount risked on a particular wager, at specific odds. Our concept of the extent of incoherence with escrow $\sum_{i=1}^{n} e(Y_i)$ is a special case of *relative incoherence* from Nau (1989, p. 389), when the confidence weights are all equal. We do not think that this level of added flexibility is important for the testing problems discussed here, however.

In conclusion, we believe that approaches like Nau's and those we have developed here and in Schervish, Seidenfeld, and Kadane (1997) permit a more subtle treatment of such longstanding issues as the debate over coherence versus incoherence of some classical statistical practices. That is not the whole problem. Rather, we need to know how far from coherent a particular policy is after we learn that it is incoherent, and learn how it compares with other incoherent methods that have been adopted in practice. We hope to continue this line of investigation in our future work.

APPENDIX A: PROOF OF THEOREM 1

Define

$$\begin{aligned} g_1(\alpha_1, \ldots, \alpha_n, y_1, \ldots, y_n) &= \sup_{t \in T} \sum_{i=1}^{n} \alpha_i (X_i(t) - y_i), \\ h_1(\alpha_1, \ldots, \alpha_n, y_1, \ldots, y_n) &= f_n\left(e(\alpha_1[X_1(t) - y_1]), \ldots, e(\alpha_n[X_n(t) - y_n])\right). \end{aligned}$$

Then $g_1 = -G(\sum_{i=1}^{n} \alpha_i (X_i - y_i))$ and h_1 is the amount the bookie must escrow in order to gamble. Since the previsions are incoherent, there exists at least one combination Y

of gambles with $G(Y) > 0$ and strictly positive escrow, so we do not need to worry about the denominator of (14) being 0 in the vicinity of the supremum. Let ρ_* stand for the right-hand side of (14). Let $\epsilon > 0$ and let

$$\frac{-g(\alpha_{1,0},\ldots,\alpha_{n,0})}{h(\alpha_{1,0},\ldots,\alpha_{n,0})} > \rho_* - \epsilon. \tag{A.1}$$

We will complete the proof by showing that there exist $y_1,\ldots,y_n$ all on the proper sides of the x_is such that

$$\frac{-g_1(\alpha_{1,0},\ldots,\alpha_{n,0},y_1,\ldots,y_n)}{h_1(\alpha_{1,0},\ldots,\alpha_{n,0},y_1,\ldots,y_n)} \geq \rho_* - 2\epsilon. \tag{A.2}$$

Since g_1/h_1 is a continuous function of $(y_1,\ldots,y_n)$ in a neighborhood of $(x_1,\ldots,x_n)$ for fixed αs, there exist $y_1,\ldots,y_n$ such that each y_i is on the proper side of x_i and the ratio in (A.2) is within ϵ of the ratio in (A.1).

APPENDIX B: PROOF OF THEOREM 2

Two lemmas are useful for the proof of Theorem 2.

Lemma 1. Let $\{A_i\}_{i=1}^m$ be a collection of disjoint nonempty events with lower previsions $p_i \geq 0$ for $i = 1,\ldots,m$ such that $s^- = \sum_{i=1}^m p_i \geq 1$ and $m > 1$. Then $\rho = (s^- - 1)/s^-$ and $\alpha_j = 1/s^-$ for all j achieves the inf in (15). If $p_i > 0$ for all i and $s^- > 1$, then the $\alpha_1,\ldots,\alpha_m$ that achieve the inf in (15) are unique.

Proof. If $s^- = 1$, then the lower previsions are coherent and $\rho = 0$. In this case, all α_i equal to each other achieves this value. For the rest of the proof, assume that $s^- > 1$. In this case, we know that there exist $\alpha_1,\ldots,\alpha_m$ (in particular those given in the statement of the lemma) such that $g(\alpha_1,\ldots,\alpha_m) < 0$.

If $\{A_i\}_{i=1}^m$ do not form a partition, let $n = m+1$, $A_n = (\cup_{i=1}^m A_i)^C$, $\alpha_n = 0$, and $p_n = 0$. If $\{A_i\}_{i=1}^m$ do form a partition, then let $n = m$. In either case, the combination of gambles in this problem is

$$\sum_{i=1}^n \alpha_i(I_{A_i} - p_i) = \sum_{i=1}^n \alpha_i I_{A_i} - \sum_{i=1}^n \alpha_i p_i. \tag{A.3}$$

Let $c = \sum_{i=1}^n \alpha_i p_i$. Then $g(\alpha_1,\ldots,\alpha_n) = \max\{\alpha_1,\ldots,\alpha_n\} - c$. Clearly having all $\alpha_i = 0$ does not achieve the inf in (15), so assume that at least one $\alpha_i > 0$. In such cases, $h(\alpha_1,\ldots,\alpha_n) = \sum_{i=1}^m \alpha_i p_i = c > 0$. (Since $g < 0$, we cannot have $c = 0$.) Clearly, if $c < 1$, we can make g smaller (strictly smaller if all $p_i > 0$) by scaling up the α_i to make $c = 1$. So assume that $c = 1$. Now, $g = \max\{\alpha_1,\ldots,\alpha_m\} - 1$. We minimize g by making the largest α_i as small as possible subject to $\sum_{i=1}^m \alpha_i p_i = 1$.

If the α_i are not all equal, then it is easy to see that we can lower the largest ones by raising the smallest ones while maintaining the constraint $\sum_{i=1}^m \alpha_i p_i = 1$. If $0 < p_i$ for all $i = 1,\ldots,m$, then this maneuver strictly lowers g. This implies that g is minimized by choosing all of the α_i equal to the same value, which must, by the constraint, be the value $1/\sum_{i=1}^m p_i$. Plugging this value for all α_i into the formula for g yields the value of ρ stated in the lemma. Uniqueness in the case $0 < p_i$ for all $i = 1,\ldots,m$ follows from the series of strict decreases in g that occurred in the above discussion.

The next lemma handles the case in which the sum of upper previsions is too small.

Lemma 2. Let $\{A_i\}_{i=1}^n$ be a collection of $n > 1$ disjoint nonempty events with upper previsions $0 \le q_i$ and $s^+ = \sum_{i=1}^n q_i < 1$. Let

$$g'(\alpha_1, \ldots, \alpha_n) = \sup_{t \in \cup_{i=1}^n A_i} \sum_{i=1}^n \alpha_i (I_{A_i}(t) - q_i).$$

Then

- $\alpha_j = -1/(n - s^+)$ for all j will minimize g' subject to $h(\alpha_1, \ldots, \alpha_n) \le 1$. Call the minimum value ρ'.
- If $\{A_i\}_{i=1}^n$ is a partition, then $\rho' = (1 - s^+)/(n - s^+)$ and $\alpha_j = -1/(n - s^+)$ for all j achieves the inf in (15).

If all $q_i > 0$, the $\alpha_1, \ldots, \alpha_n$ that minimize g' are unique.

Proof. We have $g'(\alpha_1, \ldots, \alpha_n) = \max\{\alpha_1, \ldots, \alpha_n\} - c$, where $c = \sum_{i=1}^n \alpha_i q_i$, and $g' = g$ in the partition case. Let $\alpha_j = \max\{\alpha_1, \ldots, \alpha_n\}$, so that $g'(\alpha_1, \ldots, \alpha_n) = \alpha_j - c$. Then

$$g'(\alpha_j, \ldots, \alpha_j) = \alpha_j - c + \sum_{i \ne j} (\alpha_i - \alpha_j) q_i \le g'(\alpha_1, \ldots, \alpha_n), \tag{A.4}$$

since $\alpha_i \le \alpha_j$ for all $i \ne j$ and $q_i \ge 0$. Also,

$$h(\alpha_j, \ldots, \alpha_j) = h(\alpha_1, \ldots, \alpha_n) + \sum_{i \ne j} (\alpha_i - \alpha_j)(1 - q_i) \le h(\alpha_1, \ldots, \alpha_n),$$

since $\alpha_i \le \alpha_j$ for all $i \ne j$ and $1 - q_i \ge 0$. It follows that g' is minimized (subject to $h \le 1$) by setting all α_i equal to each other and rescaling them to make $h = 1$. Uniqueness in the case $0 < q_i$ for all $i = 1, \ldots, m$ follows from the fact that the inequality in (A.4) is strict if all $q_i > 0$ and the α_i are not all equal.

Proof of Theorem 2. For part 1, suppose that we contemplate a combination of gambles

$$\sum_{i=1}^n \alpha_i (I_{A_i} - p_i) + \sum_{i=1}^n \beta_i (I_{A_i} - q_i), \tag{A.5}$$

with all $\alpha_i \ge 0$ and all $\beta_i \le 0$. We will show that, in order to maximize ρ, we must have all $\beta_i = 0$. Together with Lemma 1, this will imply that the conclusion from part 1 holds. Although not acceptable, consider the following alternate combination of gambles

$$\sum_{i=1}^n \alpha_i (I_{A_i} - p_i) + \sum_{i=1}^n \beta_i (I_{A_i} - p_i). \tag{A.6}$$

The gamble (A.6) equals (A.5) minus $\sum_{i=1}^n \beta_i (p_i - q_i)$ in all circumstance. This value is nonnegative, and it also equals the additional escrow required for (A.6) over that required for (A.5). It follows that the maximum value for ρ based on gambles of the form (A.6) will be at least as close to 1 as (hence at least as large as) the value based on (A.5). We will prove that all $\beta_i = 0$ is necessary to maximize ρ based on (A.6), hence it is also necessary to maximize ρ based on (A.5) because the two gambles are the same when all $\beta_i = 0$. Rewrite (A.6) as

$$\sum_{i=1} (\alpha_i + \beta_i)(I_{A_i} - p_i). \tag{A.7}$$

This gamble has the same payoff as (A.6) but its escrow is no larger because $(\alpha_i+\beta_i)p_i \le \alpha_i p_i$ whenever $\alpha_i+\beta_i > 0$ and $-(\alpha_i+\beta_i)(1-p_i) \le -\beta_i(1-p_i)$ whenever $\alpha_i+\beta_i < 0$. It follows that the maximum value for ρ based on gambles of the form (A.7) will be at least as large as the value based on (A.6). We prove below that all $\alpha_i+\beta_i \ge 0$ is necessary to maximize ρ based on (A.7), hence all $\beta_i = 0$ is also necessary to maximize ρ based on (A.6) because the two gambles are the same when all $\beta_i = 0$.

For convenience, define $\gamma_i = \alpha_i + \beta_i$. Suppose first that all $\gamma_i < 0$. Let $c = \sum_{i=1}^{n} \gamma_i p_i$. Then $g(\gamma_1,\ldots,\gamma_n) = \max\{\gamma_1,\ldots,\gamma_n\} - c$. Since $\sum_{i=1}^{n} p_i > 1$, $c < \max\{\gamma_1,\ldots,\gamma_n\}$. It follows that $g(\gamma_1,\ldots,\gamma_n) > 0$ which cannot provide a maximum for ρ. Hence at least one $\gamma_i \ge 0$. Next, suppose that for some j $\gamma_j < 0$, $g(\gamma_1,\ldots,\gamma_n) < 0$, and $h(\gamma_1,\ldots,\gamma_n) \le 1$. Since at least one $\gamma_i \ge 0$, we know that $\gamma_j < \max\{\gamma_1,\ldots,\gamma_n\}$. We show next that replacing γ_j by 0 will not raise the value of g and it will not violate the constraint $h \le 1$. Without loss of generality, let $\gamma_1 < 0$. Then $c' = \sum_{i=2}^{n} \gamma_i p_i \ge c$ and

$$g(0,\gamma_2,\ldots,\gamma_n) = \max\{0,\gamma_2,\ldots,\gamma_n\} - c' = g(\gamma_1,\ldots,\gamma_n) - c' + c.$$

Since $c' \ge c$, we have not raised g by replacing γ_1 by 0. Also, if $p_1 > 0$, g strictly decreases. Clearly $h \le 1$ still holds because setting $\gamma_1 = 0$ is equivalent to deleting the first gamble. It follows that we cannot have a single $\gamma_i < 0$ if we wish to maximize ρ.

For part 2, we can argue as above that the maximum ρ achievable by a gamble of the form (A.5) is no larger than that available by a gamble of the form

$$\sum_{i=1}(\alpha_i+\beta_i)(I_{A_i} - q_i), \tag{A.8}$$

and that the two maxima are the same if maximizing (A.8) requires all $\alpha_i+\beta_i \le 0$. For convenience, define $\gamma_i = \alpha_i+\beta_i$. Let $c = \sum_{i=1}^{n} \gamma_i q_i$. Then $g(\gamma_1,\ldots,\gamma_n) = \max\{\gamma_1,\ldots,\gamma_n\} - c$. Suppose first that all $\gamma_j > 0$. Since $\sum_{i=1}^{n} q_i < 1$, we have $c < \max\{\gamma_1,\ldots,\gamma_n\}$ and $g > 0$, which cannot provide a maximum for ρ. Hence at least one $\gamma_i \le 0$. Next, suppose that some $\gamma_j > 0$, $g'(\gamma_1,\ldots,\gamma_n) < 0$, and $h(\gamma_1,\ldots,\gamma_n) \le 1$. We show now that replacing all of the positive γ_j by 0 will not raise the value of g and it will not violate the constraint $h \le 1$. Without loss of generality, let $\gamma_{k+1},\ldots,\gamma_n > 0$ and $\gamma_1,\ldots,\gamma_k \le 0$. Let $c' = \sum_{i=1}^{k} \gamma_i q_i = c - \sum_{i=k+1}^{n} \gamma_i q_i$ so that

$$g(\gamma_1,\ldots,\gamma_k,0,\ldots,0) \le -c' = \sum_{i=k+1}^{n} \gamma_i q_i - c < g(\gamma_1,\ldots,\gamma_n),$$

because $\sum_{i=k+1}^{n} \gamma_i q_i < \max\{\gamma_{k+1},\ldots,\gamma_n\}$. So, we have lowered g by replacing the positive γ_js by 0. Since replacing $\gamma_{k+1},\ldots,\gamma_n$ by 0 is equivalent to deleting a few gambles from the collection, $h \le 1$ still holds. It follows that we cannot have a single $\gamma_i > 0$ if we wish to maximize ρ. Lemma 2 completes the proof of part 2.

APPENDIX C: PROOF OF THEOREM 3

If ψ is admissible, then neither ψ nor δ_i can dominate the other and the two expressions in square brackets in the formula for $a_i(\psi)$ have the same sign. Also, $0 \le b_i(\psi) \le 1$. If $a_i(\psi) > 0$, then $b_i(\psi)$ is a lower prevision for A. If $a_i(\psi) < 0$, then $b_i(\psi)$ is an upper prevision for A.

Since the δ_i and ψ_i are all going to be admissible, we cannot have (23) unless $a_i(\psi_i) > 0$ and $a_{1-i}(\psi_{1-i}) < 0$ for either $i = 0$ or $i = 1$. Next, we show that we can find ψ_0 and ψ_1 such that for one i (either 0 or 1) $R_i(0, \psi_i) > R_i(0, \delta_i)$ and $R_{1-i}(0, \psi_{1-i}) < R_{1-i}(0, \delta_{1-i})$ and hence $a_i(\psi_i) > 0$ and $a_{1-i}(\psi_{1-i}) < 0$. We argue indirectly. If this were not possible, then $R_i(0, \delta_i)$ would either have to be the largest possible value of $R_i(0, \cdot)$ for both $i = 0, 1$ or they would both have to be the smallest possible value for both $i = 0, 1$, otherwise we could get below one and above the other. If both δ_0 and δ_1 are admissible and have the smallest possible risk values at $\theta = 0$ then they are both Bayes rules with respect to the prior that puts probability 1 on $\theta = 0$. Since they are not both Bayes with respect to the same prior, they cannot both have the smallest possible risk at $\theta = 0$. If they both had the highest possible risk at $\theta = 0$ then they would both have the smallest possible risk at $\theta = 1$ (since they are both admissible). In this case, they would both be Bayes rules with respect to the prior that put all probability on $\theta = 1$.

Next, we show that there exist ψ_0 and ψ_1 such that (23) occurs. For $0 \le q \le 1$, construct decision problem D_q with the same parameter space but with data (Y, Z_q) where Z_q is independent of (X_0, X_1) and has Bernoulli distribution with probability q. If $Z_q = 1$, $Y = X_1$ and we are presented with problem D_1. If $Z_q = 0$, $Y = X_0$ and we are presented with problem D_0. (In this way, D_0 and D_1 have the same meaning as their original definitions.) A decision rule in problem D_q will consist of a pair (ψ_0, ψ_1) where ψ_i is a decision rule in problem D_i and the pair is interpreted as meaning that we use rule ψ_1 if $Z_q = 1$ and we use ψ_0 if $Z_q = 0$. The risk function for a rule (ψ_0, ψ_1) in problem D_q is $qR_1(\theta, \psi_1) + (1-q)R_0(\theta, \psi_0)$. The Bayes risk in problem D_q of the rule (ψ_0, ψ_1) with respect to a prior $p_0 = \Pr(\theta = 0)$ is

$$\begin{aligned} & p_0[qR_1(0,\psi_1) + (1-q)R_0(0,\psi_0)] + (1-p_0)[qR_1(1,\psi_1) + (1-q)R_0(1,\psi_0) \\ &= q[p_0R_1(0,\psi_1) + (1-p_0)R_1(1,\psi_1)] + (1-q)[p_0R_0(0,\psi_0) + (1-p_0)R_0(1,\psi_0)] \end{aligned} \tag{A.9}$$

From A.9, we see that (ψ_0, ψ_1) is a Bayes rule in problem D_q with $0 < q < 1$ if and only if ψ_i is a Bayes rule in problem D_i for $i = 0, 1$. The assumption that δ_0 and δ_1 are not Bayes rules with respect to a common prior means that (δ_0, δ_1) is not a Bayes rule in problem D_q if $0 < q < 1$. Define $R_q(\theta, \cdot)$ to be the risk function in problem D_q. Then

$$R_q(\theta, (\psi_0, \psi_1)) - R_q(\theta, (\delta_0, \delta_1)) = (1-q)a_0(\psi_0)(I_A - b_0(\psi_0)) + qa_1(\psi_1)(I_A - b_1(\psi_1)). \tag{A.10}$$

Then, for every $0 < q < 1$ there exists a rule (ψ_0, ψ_1) that dominates (δ_0, δ_1) in problem D_q, making the left-hand side of (A.10) negative for both $\theta = 0$ and $\theta = 1$ and hence satisfying (23).

For the remainder of the proof, we will assume that ψ_0 and ψ_1 have been chosen so that (23) holds and so that $b_0(\psi_0)$ is a lower prevision and $b_1(\psi_1)$ is an upper prevision. (That is, $a_0(\psi_0) > 0$, $a_1(\psi_1) < 0$, and $b_0(\psi_0) > b_1(\psi_1)$.) For each such pair (ψ_0, ψ_1), the discussion in Example 5 establishes that

$$\rho_\gamma = \frac{b_0(\psi_0) - b_1(\psi_1)}{\max\{b_0(\psi_0), 1 - b_1(\psi_1)\} + \gamma \min\{b_0(\psi_0), 1 - b_1(\psi_1)\}},$$

which increases as either $b_0(\psi_0)$ increases or as $b_1(\psi_1)$ decreases. In order to make ρ_γ as large as possible, we need to make $b_0(\psi_0)$ as large as possible while making $b_1(\psi_1)$ as small as possible by choice of ψ_0 and ψ_1. Since $b_i(\psi_i)$ depends only on ψ_i, we can try to maximize $b_0(\psi_0)$ by choice of ψ_0 and try to minimize $b_1(\psi_1)$ by choice of ψ_1. These can be done separately. Looking at the formula for b_i in (21), we see that $b_i(\psi_i) = (1 + v_i(\psi_i))^{-1}$, where

$$v_i(\psi_i) = \frac{R_i(0, \psi_i) - R_i(0, \delta_i)}{R_i(1, \delta_i) - R_i(1, \psi_i)}$$

is the negative of the slope of the line connecting the two points $(R_i(1, \psi_i), R_i(0, \psi_i))$ and $(R_i(1, \delta_i), R_i(0, \delta_i))$ on the lower boundary of the risk set for problem D_i. Each point on the lower boundary of the risk set (a convex set) is the risk function of an admissible rule and a Bayes rule with respect to some prior. For each point on the lower boundary of the risk set, the slope of every support line at that point is minus the odds in favor of $\{\theta = 0\}$ for a prior such that the corresponding rule is Bayes with respect to the prior. The lower boundary of the risk set is a convex and continuous curve. Since $R_0(1, \psi_0) < R_0(1, \delta_0)$, we increase $b_0(\psi_0)$ (decrease $v_0(\psi_0)$) by moving $(R_0(1, \psi_0), R_0(0, \psi_0))$ as close to $(R_0(1, \delta_0), R_0(0, \delta_0))$ as possible. The limit of $v_0(\psi_0)$ is minus the largest absolute value of all slopes of support lines at $(R_0(1, \delta_0), R_0(0, \delta_0))$. This slope is minus the odds ratio $p_0/(1 - p_0)$ for the prior $p_0 = \Pr(\theta = 0)$ with the largest value of p_0 such that δ_0 is Bayes with respect to the prior. By similar reasoning, the largest value of $v_1(\psi_1)$ (giving the smallest value of $b_1(\psi_1)$) is minus the smallest absolute value of odds ratios $p_1/(1-p_1)$ for priors $p_1 = \Pr(\theta = 0)$ such that δ_1 is Bayes with respect to the prior. So ρ_γ has the value stated in (24).

REFERENCES

Cox, D. R. (1958), "Some problems connected with statisitical inference," *The Annals of Mathematical Statistics*, 29, 357–363.

de Finetti, B. (1974), *Theory of Probability*, volume 1, New York: John Wiley.

Freedman, D. A. and Purves, R. A. (1969), "Bayes' method for bookies," *The Annals of Mathematical Statistics*, 40, 1177–1186.

Kadane, J. B. and Wolfson, L. (1998), "Experiences in elicitation," (with comments), *Journal of the Royal Statistical Society*, Ser. D (The Statistician), 47, 3–19.

Lehmann, E. (1958), "Significance level and power," *The Annals of Mathematical Statistics*, 29, 1167–1176.

Lindley, D. (1972), *Bayesian Statistics, A Review*, Philadelphia: SIAM.

Nau, R. F. (1989), "Decision analysis with indeterminate or incoherent probabilities," *Ann. Oper. Res.*, 19, 375–403.

——— (1992), "Indeterminate probabilities on finite sets," *The Annals of Statistics*, 20, 1737–1767.

Schervish, M. J. (1995), *Theory of Statistics*, New York: Springer-Verlag.

Schervish, M. J., Seidenfeld, T., and Kadane, J. B. (1997), "Two measures of incoherence: How not to gamble if you must," Technical Report 660, Carnegie Mellon University, Statistics.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (1990), "Decisions without ordering," in *Acting and Reflecting*, ed. W. Sieg, Dordrecht: Kluwer Academic Publishers, pp. 143–170.

Shimony, A. (1955), "Coherence and the axioms of probability," *Journal of Symbolic Logic*, 20, 1–28.

Walley, P. (1991), *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall.