March 13, 2002

Measures of incoherence: How not to gamble if you must

MARK J. SCHERVISH, TEDDY SEIDENFELD and JOSEPH B. KADANE
Carnegie Mellon University, USA

SUMMARY

The degree of incoherence, when previsions are not made in accordance with a probability measure, is measured by the rate at which an incoherent bookie can be made a sure loser. We consider each bet from three points of view: that of the gambler, that of the bookie, and a neutral viewpoint. From each viewpoint, we define an normalization for each bet, and the sure loss for incoherent previsions is divided by the normalization to determine the rate of incoherence. Several different definitions of normalization are considered in order to determine plausible ranges for the degree of incoherence. We give examples of the measurement of incoherence of of some classical statistical procedures.

Keywords: BOOKIE; COHERENCE; ESCROW; GAMBLER.

1. INTRODUCTION

de Finetti (1974) describes the criterion of coherence for probabilities assigned to events and previsions assigned to general random variables. The idea is that, if one were to use these previsions as fair prices to pay for gambles on the random quantities, then the previsions are incoherent if and only if there exists a finite combination of the gambles that is guaranteed to lose at least some positive amount. To put this in mathematically precise language, let $X_1, \ldots, X_n$ be bounded random variables, that is bounded functions from a set of states of nature S to the reals. For $i = 1, \ldots, n$, let p_i be the prevision of X_i, so that $\alpha_i(X_i - p_i)$ is a fair gamble for all sufficiently small $|\alpha_i|$ values. The previsions are *incoherent* if there exist values $\alpha_1, \ldots, \alpha_n$ such that

$$\sup_s \alpha_i(X_i(s) - p_i) < 0. \tag{1.1}$$

The previsions are *coherent* if they are not incoherent. If (1.1) holds, we say that *Dutch book* has been made against the person who offered the previsions.

de Finetti (1974) merely partitions all prevision assignments into two classes: coherent and incoherent. It seems reasonable, however, to expect that some collections of previsions are more incoherent than others. It is our goal in this paper to set up a framework in which one can attempt to measure how incoherent is a collection of incoherent previsions.

We begin by considering a slightly more general situation than that described above, both because it strengthens the results and because the most interesting examples are of the more general form. Think of previsions being assigned to random variables by a bookie who is then going to take bets from one or more gamblers. Suppose that the bookie chooses the prevision p for a bounded random variable X. The gain to the bookie when the state of nature is s and the bookie accepts the gamble $\alpha(X-p)$ from a gambler is $\alpha(X(s)-p)$. It is common in some gambling situations (horse racing is a common example) for bookies to offer only one-sided previsions. That is, the bookie offers a prevision p for X, but accepts only gambles of the form $\alpha(X-p)$ for $\alpha > 0$. Such a prevision p will be called a *lower prevision*. Similarly, if the bookie accepts only gambles of the form $\alpha(X-p)$ for $\alpha < 0$, we shall call p an *upper prevision* for X. It should be easy to see that, if a bookie is willing to accept the gamble $\alpha(X-p)$ with $\alpha > 0$, then the bookie should also be willing to accept the gamble $\alpha(X-x)$ for every $x < p$, since the payment to the bookie from this second gamble is greater than that of the first by $p-x$ no matter what state of nature occurs. Similarly, if p is an upper prevision, the bookie should be willing to accept the gamble $\alpha(X-x)$ for all $x > p$ when $\alpha < 0$. In the future when we refer to a coefficient of the *appropriate sign*, we shall mean a positive α for a gamble based on a lower prevision and a negative α for a gamble based on an upper prevision.

Example 1. A simple example of incoherence occurs when a lower prevision is greater than an upper prevision. Suppose that $p > q$ where p is a lower prevision for X and q is an upper prevision for X. Then the bookie is willing to accept the following two gambles, $\alpha(X-p)$ and $-\alpha(X-q)$ for $\alpha > 0$. The sum of these two gambles is $\alpha(q-p) < 0$ no matter what state of nature occurs. There is a sense in which, the larger $p-q$ is, the more the bookie stands to lose from such gambles. Of course, the bookie loses more the larger α is, regardless of how large $p-q$ is. However, α is more a measure of how much the gambler wishes to bet than it is a measure of the bookie's incoherence.

We would like to try to measure the incoherence of the bookie's previsions by looking at how much the bookie can be forced to lose relative to how much the gambler needs to bet in order to force the loss. As we noted in Example 1, if the gambler bets twice as much (double α) the bookie loses twice as much, even though the previsions have not changed. We need to be able to normalize the combinations of bets that a bookie is willing to accept in order to extract the degree of incoherence from the size of the loss that the bookie incurs.

In summary, incoherence arises when a bookie offers previsions such that there exists a combination of gambles for which the bookie is guaranteed to lose at least some positive amount. A gambler can increase the guaranteed loss by increasing the sizes of his/her bets with the bookie. We shall try to measure the incoherence of incoherent previsions by looking at how large the guaranteed loss can be made relative to some normalization for the collection of gambles that the gambler chooses.

2. NORMALIZATION

As we noted earlier, we need normalizations to measure the sizes of gambles. There are two important aspects of normalization that we have chosen to separate. The first is the normalization of a single gamble, and the second is the normalization of sums of gambles. Treating a sum of gambles as a single gamble obscures the fact that the coefficients can be chosen individually and that the previsions are given individually. For a single gamble, there are several possible ways to measure the size. Let X be a random quantity with prevision p, and let α be a constant with the appropriate sign. The normalization for $\alpha(X-p)$ will be allowed to depend on all three of α, X, and p. (When we say that the normalization depends on X, we mean that it depends on the function X from states of nature to the real numbers, not on the unknown value $X(s)$. For example, the normalization can depend on the maximum and/or minimum value of X, etc.) Some examples of normalizations include the following three:

1. *The bookie's escrow:* $\max\{0, -\inf_s \alpha[X(s)-p]$.
2. *The gambler's escrow:* $\max\{0, \sup_s \alpha[X(s)-p]$.
3. *The neutral normalization:* $|\alpha|$.

The two escrows have interpretations in terms of betting. If the bookie requires each gambler to show that they have sufficient funds to pay off any bets they might lose, then the gambler's escrow, being the most the gambler can lose, will cover the bet. Similarly, the bookie's escrow will cover the the bookie's largest possible loss. Notice that all three of the normalizations listed above have the following two properties:

- The normalization is always nonnegative.
- If α is changed to $c\alpha$ for $c>0$, the normalization gets multiplied by c as well.

These are the important properties that we require of a normalization. Although there are many normalizations that satisfy these two conditions, we shall only consider the three above in this paper.

When we combine several gambles, we need to be able to normalize the combination. We have chosen to require the normalization of a combination of gambles to depend solely on the normalizations of the individual gambles that go into the combination. The reason for this is that, once we start to combine gains and losses, we are starting to do some of the work of measuring incoherence. For example, consider $X_1 = I_A$, the indicator of an event with prevision p and $X_2 = I_{A^C}$, the indicator of the complement with prevision q. If we combine $(I_A - p) + (I_{A^C} - q)$, we get $1-p-q$, which already tells us whether or not the previsions are coherent. Our assumption is designed to help us separate the normalization from the measurment of incoherence.

Schervish, Seidenfeld & Kadane (2002b) give a list of criteria that a normalization should satisfy, including the assumption just mentioned. We summarize these here. Let $f_n(x_1, \ldots, x_n)$ denote the normalization for a combination of n gambles whose individual normalizations are $x_1, \ldots, x_n$.

1. f_n is homogeneous of degree 1, that is, $f_n(cx_1, \ldots, cx_n) = cf_n(x_1, \ldots, x_n)$ for all $c > 0$.
2. f_n is invariant under permutations of its arguments.
3. f_n is nondecreasing in each argument.
4. $f_{n+1}(x_1, \ldots, x_n, 0) = f_n(x_1, \ldots, x_n)$.
5. The normalization is no larger than the sum of the individual normalizations.
6. f_n is continous.
7. $f_1(x) = x$.

The first condition implies that scaling up a collection of gambles by the same amount will not change the rate of incoherence. The second condition expresses the fact that we do not care in what order the gambles are written. The third condition expresses the idea that bigger gambles should require larger normalization. The fourth condition says that if a gamble does not require any normalization, then the other gambles should determine the normalization for the combination. The fifth condition is like the triangle inequality for norms, saying that the whole is no greater than the sum of the parts. It makes particular sense if the normalization is thought of as an escrow. The sixth condition says that small changes in individual normalizations should produce small changes in the overall normalization. The seventh condition merely expresses the fact that the individual normalizations are just that.

Based on these criteria, Schervish, Seidenfeld & Kadane (2002b) characterize the collection of functions f_n that can be used for normalization. In particular, they find that the largest and smallest normalizations arise from the following two functions:

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

In this paper, we shall consider only normalizations based on these two functions, which we call the *max* and *sum* normalizations.

In summary, a normalization is defined by two choices. One is a choice of normalization for individual gambles, and the other is a choice of normalization function for combining individual normalizations. We have given specific examples of three individual normalizations and two combination functions. By pairing these arbitrarily, we produce six different examples of how to normalize combinations of gambles. We shall refer to these pairings with names such as neutral/sum, gambler's escrow/max, etc. We shall assume throughout this paper that the same individual normalization is chosen for all individual gambles when they are being combined.

3. RATE OF INCOHERENCE

We are now in position to say how we will measure incoherence. Consider a finite collection of gambles $Y_i = \alpha_i(X_i - p_i)$ for $i = 1, \ldots, n$. The *guaranteed loss* from the combination $Y = \sum_{i=1}^n Y_i$ is

$$G(Y) = -\min\left\{0, \sup_s Y(s)\right\}.$$

Dutch book has been made with this combination if and only if $G(Y) > 0$. It is clear that if each Y_i is changed to cY_i for $c > 0$, then Y changes to cY and $G(cY) = cG(Y)$. We shall measure the rate of incoherence of Y by first dividing $G(Y)$ by a normalization. If $e(Y_i)$ is our chosen normalization for individual gamble Y_i, then our normalization for Y is $f_n(e(Y_1), \ldots, e(Y_n))$, for some function f_n satisfying the conditions mentioned earlier. The *rate of guaranteed loss* for Y is then

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

For a fixed collection of random variables $X_1, \ldots, X_n$ and previsions $p_1, \ldots, p_n$, the *rate of incoherence* is the maximal value of $H(Y)$ over all choices of $\alpha_1, \ldots, \alpha_n$ that have the appropriate signs.

Example 2. Consider an event A and its complement A^C. Suppose that a bookie offers lower previsions of 0.6 for both of these events. These are clearly incoherent, and there are many combinations of gambles that make Dutch book. All such combinations must be of the form $Y = \alpha_1(I_A - 0.6) + \alpha_2(I_{A^C} - 0.6)$ for $\alpha_1, \alpha_2 > 0$. In this case, Y only takes two values, $0.4\alpha_1 - 0.6\alpha_2$ and $0.4\alpha_2 - 0.6\alpha_1$. Hence,

$$G(Y) = \begin{cases} 0 & \text{if } \alpha_1 \geq 1.5\alpha_2 \text{ or } \alpha_2 \geq 1.5\alpha_1, \\ 0.6\min\{\alpha_1, \alpha_2\} - 0.4\max\{\alpha_1, \alpha_2\} & \text{otherwise.} \end{cases}$$

That is, Dutch book is made if and only if both α_1/α_2 and α_2/α are less than 1.5. For the neutral/sum normalization, the normalization equals $\alpha_1 + \alpha_2$. So, when Dutch book can be made,

$$\begin{aligned} H(Y) &= \frac{0.6\min\{\alpha_1, \alpha_2\} - 0.4\max\{\alpha_1, \alpha_2\}}{\alpha_1 + \alpha_2} \\ &= r - 0.4, \end{aligned}$$

where $r = \min\{\alpha_1, \alpha_2\}/(\alpha_1 + \alpha_2)$ can be any number between 0.4 and 0.5. Clearly, we maximize $H(Y)$ by choosing $r = 0.5$. This makes the rate of incoherence 0.1 in this example. We can achieve this rate by choosing $\alpha_1 = \alpha_2$.

If, instead of neutral/sum, we had used neutral/max, then the normalization would have been $\max\{\alpha_1, \alpha_2\}$ and then $H(Y) = 0.6\frac{r}{1-r} - 0.4$, where r is the same as before. This is also maximized with $r = 0.5$, and the rate of incoherence is then 0.2. This illustrates a fact that is quite general. Since $f_{0,n}$ and $f_{1,n}$ are respectively the smallest and largest normalizations, they lead respectively to the largest and smallest rates of incoherence when combined with a common individual normalization. In the case of events, the neutral normalization is the sum of the gambler's escrow and the bookie's escrow whenever the prevision is between 0 and 1. Hence, rates of incoherence based on the two escrows will be larger than those based on the neutral normalization.

4. CHOOSING BETWEEN NORMALIZATIONS

We have introduced at least six ways to normalize sums of gambles when computing rates of incoherence. Each has its advantages and disadvantages. There are, however, some properties that we have been able to determine for some of them. The two escrows have interpretations in terms of amounts needed to cover the bets. We have not yet found any operational interpretation for the neutral normalization. The escrows have the property that a prevision of p for X and a prevision of cp for cX (with $c > 0$) will have equivalent effects on any rate of incoherence. This is not true of the neutral normalization. For example, suppose that we wish to include $\alpha(X - p)$ in a combination of gambles, where p is a coherent lower prevision for X and $\alpha > 0$. The three individual normalizations for $(\alpha/c)(cX - cp)$ are $\alpha(\sup X - p)$ (gambler's escrow), $-\alpha(\inf X - p)$ (bookie's escrow) and α/c (neutral). Notice that the first two are the same as they would be for the gamble $\alpha(X - p)$, whereas the third is different. One consequence of this is the following. If it turns out that $\alpha(X - p)$ is included in a combination of gambles Y that is incoherent, and if $c > 1$, then $H(Y)$ based on the neutral normalization will be larger if one uses $(\alpha/c)(cX - cp)$ instead, because this version requires smaller normalization and produces the same payoffs as $\alpha(X - p)$. We don't find this to be a troubling feature of the neutral normalization because incoherent agents are free to choose previsions other than cp for cX even after they have chosen p as the prevision of X. From this point of view, X and cX are different random variables, and there is no compelling reason that their effects on the rate of incoherence need be related.

Schervish, Seidenfeld & Kadane (2002b) have established some additional properties of the various normalizations. They have established conditions under which the rate of incoherence is continuous as a function of the random variables and their previsions. Continuity of the rate of incoherence means (loosely) if $\sup_s |X_1(s) - X_2(s)| < \epsilon$ and $|p_1 - p_2| < \epsilon$, then substituting X_2 with prevision p_2 for X_1 with prevision p_1 should make a small change to the rate of incoherence. Schervish, Seidenfeld & Kadane (2002b) show that the neutral/sum normalization gives a continuous rate of incoherence and all neutral normalizations give continuous rates of incoherence if we have only finitely many gambles from with to choose. Normalizations based on the escrows give continuous rates of incoherence only under the assumption that each individual prevision is coherent by itself (and not just barely so). That is, if p is a lower prevision for X, then $p < \sup_s X(s)$ and if p is an upper prevision for X, then $\inf_s X(s) < p$. Equality in either of these would lead to 0 normalization for an individual gamble, and rates of incoherence are not continuous when 0 normalizations occur.

Example 3. Consider a constant random variable $X = c$. Suppose that an incoherent bookie specifies a lower prevision $p > c$. The gambler's escrow will be 0, since the gambler cannot lose a bet of the form $\alpha(X - p)$ with $\alpha > 0$. The rate of incoherence is ∞ in this case. The bookie's escrow is $\alpha(p - c)$ and so is $G(Y)$, hence the rate of incoherence is 1 no matter how far p is from c. If $p = c$, the rate of incoherence drops to 0 for both escrows. For the neutral normalization, the normalization is α and $G(Y) = \alpha(p - c)$, so the rate of incoherence is $p - c$, which increases as p increases, as intuition might suggest.

Another property considered by Schervish, Seidenfeld & Kadane (2002b) is dominance. Suppose that two different bookies offer previsions for the same random variables. Suppose that for a specific set of coefficients for the gambles that makes a Dutch book, the second bookie's losses are always larger than the first's. Then we say that the first bookie dominates the second with respect to those coefficients. Intutively, one might expect the function H computed for the first bookie to be smaller than the function H computed for the second bookie. This will be true whenever the normalization is neutral. For the bookie's escrow, we need to assume that none of the individual gambles be a sure winner for the bookie. For the gambler's escrow, we need to assume that none of the individual gambles be a sure winner for the gambler.

On balance, it might appear that the neutral normalizations have better mathematical properties than the others, but we shall continue to work with all of them for the remainder of this paper.

5. CLASSICAL INFERENCE

Inference techniques used by non-Bayesians have sometimes been criticized on the grounds that they are incoherent. If they are indeed incoherent, then we should be able to make Dutch book and measure the rate of incoherence. For example, suppose that one chooses to test all null hypotheses at level 0.05 regardless of how the data arose or how many observations are available. It is well-known (see Cox, 1958 and Lindley, 1972 for examples) that there are cases in which such behavior runs afoul of admissibility if not coherence. Two obvious stumbling blocks stand in the way of applying the concept of coherence to classical inferences. First, classical inferences do not provide previsions. In particular, they do not provide probabilities or expected values for unknown quantities. They are often based upon probabilities and expectations for random variables that used to be unknown, but have since been observed. The unknown quantities never become the subject of a probabilistic calculation.

Secondly, classical statisticians are not prepared to gamble based on their inferences.

This last point suggests a fruitful avenue to pursue. If nothing is at stake, who cares what inference is made? So, suppose that there is a decision problem with a loss function. Classical statisticians are willing to talk about decision theory. Indeed, they have developed a theory of risk functions, admissibility, minimaxity, etc. When statisticians choose one decision rule δ_0 over another δ_1, they express a preference for suffering the risk function $R(\theta, \delta_0)$ to $R(\theta, \delta_1)$. We choose to interpret such a preference by saying that $R(\theta, \delta_1) - R(\theta, \delta_0)$ is a favorable gamble. If inferences are incoherent we should be able to combine such favorable gambles to make Dutch book.

5.1 Simple Hypothesis Testing

Consider first the case of testing a simple null hypothesis against a simple alternative hypothesis. (Schervish, Seidenfeld & Kadane, 2002a consider this situation in more detail.) Suppose that the loss function is has the simple 0–1 form. That is, the loss is 0 if the correct hypothesis is chosen, and the loss is 1 of the incorrect hypothesis is chosen.

Example 4. Suppose that X has a normal distribution with mean θ and variance σ^2, where σ^2 is known, but we want to test the null hypothesis $H_0 : \theta = 0$ versus the alternative $H_1 : \theta = 1$. Suppose that our classical statistician wants to use the most powerful level 0.05 test no matter what σ^2 is. For example, if $\sigma^2 = 1$, the most powerful level 0.05 test is to reject H_0 if $X > 1.645$. In general, we can write the risk function of the most powerful level α_0 test δ_{α_0} as

$$\begin{aligned} R(\theta, \delta_{\alpha_0}) &= \begin{cases} \alpha_0 & \text{if } \theta = 0, \\ \Phi(\Phi^{-1}(1-\alpha_0) - 1/\sigma) & \text{if } \theta = 1, \end{cases} \\ &= [\alpha_0 - \Phi(\Phi^{-1}(1-\alpha_0) - 1/\sigma)] I_{\{0\}}(\theta) + \Phi(\Phi^{-1}(1-\alpha_0) - 1/\sigma). \end{aligned}$$

So, if the statistician prefers the level 0.05 test to the level 0.1 test when $\sigma = 1$, the difference in risk functions is

$$R(\theta, \delta_{0.1}) - R(\theta, \delta_{0.05}) = 0.1796[I_{\{0\}}(\theta) - 0.7217],$$

indicating that 0.7217 is a lower prevision for the event $\{\theta = 0\}$. (Technically, all we can really say is that (0.7217×0.1796) is a lower prevision for $0.1796 I_{\{0\}}$.) Suppose also that the statistician prefers the level 0.05 test to the level 0.01 test when $\sigma = 2$. Then the difference of the risk functions is

$$R(\theta, \delta_{0.01}) - R(\theta, \delta_{0.05}) = -0.1322[I_{\{0\}}(\theta) - 0.6975],$$

making it appear as if 0.6975 is an upper prevision for $\{\theta = 0\}$. An upper prevision that is lower than a lower prevision ought to be incoherent. If we combine the first gamble with $\alpha_1 = 1/0.1796$ and the second gamble with

$\alpha_2 = 1/0.1322$, we get the combination with constant value -0.0242. Hence, we have made Dutch book. For the two gambles being discussed, the rate of incoherence using bookie's escrow/sum normalization is

$$\frac{0.0242}{0.7217 + (1 - .6975)} = 0.02363.$$

In Example 4, we could have chosen different risk functions to trade with the classical statistician. For example, when $\sigma = 1$, we could trade the risk function of the level 0.06 test instead of the risk function of the level 0.1 test. Similarly, when $\sigma = 2$, we could trade the risk function of the level 0.04 test instead of the level 0.01 test. Each of these alternative trades leads to two different gambles that provide different Dutch books, and different rates of incoherence. We might ask which trades, if any, lead to the largest rate of incoherence. Schervish, Seidenfeld & Kadane (2002a) presents a theorem that says essentially the following. Use bookie's escrow/sum normalization. For each σ, the level most powerful α_0 test is a Bayes rule with respect to a unique prior $p(\sigma) = \Pr(\theta = 0)$. Let σ_0 and σ_1 be two different possible variances. Let p_i $(i = 0, 1)$ be the two priors $p_i = p(\sigma_i)$ for $i = 0, 1$. Assume that $p_0 > p_1$. For each $i = 0, 1$, consider all possible trades of the risk function for some test δ for the risk function of the most powerful level α_0 test. The largest possible rate of incoherence from combining two such risk function trades is $(p_0 - p_1)/(1 - p_1 + p_0)$. A precise statement and proof of this result can be found in Schervish, Seidenfeld & Kadane (1997).

Example 5. We can return to Example 4 and ask what is the largest rate of incoherence. In that example, each of the two level 0.05 tests is the Bayes rule with respect to a unique prior. When $\sigma = 1$, the level 0.05 test is the Bayes rule with respect to the prior $p_0 = 0.7586$, and when $\sigma = 2$, the level 0.05 test is the Bayes rule with respect to the prior $p_1 = 0.6676$. The theorem quoted above says that the maximum rate of incoherence for bookie's escrow/sum normalization is $(0.7586 - 0.6676)/(1 + 0.7586 - 0.6676) = 0.08341$.

5.2 Minimax Estimation

Hypothesis testing at a fixed level is not the only incoherent classical inference. Minimax estimation also suffers from the same malady. Consider a binomial random variable X with parameters n and θ. The minimax estimator of θ is

$$\delta_n(X) = \frac{\sqrt{n}/2 + X}{\sqrt{n} + n}.$$

For each sample size n, δ_n is the Bayes rule with respect to some (actually several) prior distribution, but the prior must change as n changes. If indeed it is incoherent to use the minimax rule for different sample sizes, we should be able to make Dutch book by trading risk functions.

Example 6. Suppose that we offer to trade the risk function of a Bayes rule with respect to a different beta distribution prior for the risk function of the minimax rule. The risk function of the the Bayes rule $\beta_{n,\gamma}$ with respect to a beta distribution with parameters γ and γ is

$$R(\theta, \beta_{n,\gamma}) = \frac{n\theta(1-\theta) + \gamma^2(1-2\theta)^2}{(n+2\gamma)^2}.$$

If $n = 1$, the minimax rule is $\delta_1(X) = (0.5 + X)/2$ and the minimax risk is $1/16$. Suppose that we offer to trade the risk function of the Bayes rule with respect to the beta distribution prior with parameters 1.3 and 1.3, $\beta_{1,1.3}(X) = (1.3 + X)/3.6$. If $n = 4$, the minimax rule is $\delta_4(X) = (1 + X)/5$ with minimax risk $1/36$. In this decision problem, we might offer to trade the risk function of the rule $\beta_{4,0.8}$. If we make these two trades, we can then combine the two favorable gambles as follows:

$$\alpha_1[R(\theta, \beta_{4,0.8}) - R(\theta, \delta_4)] + \alpha_2[R(\theta, \beta_{1,1.3}) - R(\theta, \delta_1)]. \tag{5.1}$$

As a function of θ, (5.1) is a quadratic that is symmetric around $\theta = 1/2$. The maximum value can be minimized by choosing α_1 and α_2 so that the function is constant. Using neutral/sum normalization, the rate of incoherence in this case is 0.00032, and it is achieved with $\alpha_1 = 9.679$ and $\alpha_2 = 1$.

As in Example 5, we can replace the two trades in Example 6 by different trades that might achieve higher rates of incoherence. In general, to make

$$\alpha_1[R(\theta, \beta_{n_1,\gamma_1}) - R(\theta, \delta_{n_1})] + \alpha_2[R(\theta, \beta_{n_2,\gamma_2}) - R(\theta, \delta_{n_2})]$$

constant, we need one of γ_i to be less than $\sqrt{n_i}/2$ and the other to be larger. Then, we can choose

$$\alpha_{3-i} = \left| \frac{4\gamma_i^2 - n_i}{(n_i + 2\gamma_i)^2} \right|,$$

for $i = 1, 2$. The constant value of the combination of gambles is then

$$\alpha_1 \left(\frac{\gamma_1^2}{(n_1 + 2\gamma_1)^2} - \frac{1}{4 + 8\sqrt{n_1} + 4n_1} \right) + \alpha_2 \left(\frac{\gamma_2^2}{(n_2 + 2\gamma_2)^2} - \frac{1}{4 + 8\sqrt{n_2} + 4n_2} \right). \tag{5.2}$$

Take the negative of this, normalize it, and then maximize the ratio by choice of γ_1 and γ_2. In Example 6, with neutral/sum normalization, the choices that provide the maximum are $\gamma_1 = \gamma_2 = 0.653$, and the rate of incoherence for these choices is 0.0011.

We have not yet solved the general problem of finding the combination of risk function trades that leads to the largest rate of incoherence. To do that, we would have to consider the risk functions of all possible Bayes rules. Since every Bayes rule is the posterior mean of θ, we can write each Bayes rule as

$$\delta(x) = \frac{\int_{[0,1]} \theta^{x+1}(1-\theta)^{n-x} d\mu(\theta)}{\int_{[0,1]} \theta^{x}(1-\theta)^{n-x} d\mu(\theta)}, \tag{5.3}$$

where μ is an arbitrary prior on $[0, 1]$. Notice that (5.3) depends only on the first $n+1$ moments of the prior distribution μ. Feller (1972, Sec. VII.3) shows that, for every possible set of first $n+1$ moments, there is a prior concentrated on the points $k/(n+1)$, for $k = 0, \ldots, n+1$ that has those moments. Hence, the collection of all Bayes rules can be parameterized by the finite-dimensional collection of all prior distributions concentrated on $n+2$ equally spaced points. Although the resulting maximization problem is not likely to have a closed-form solution, a numerical algorithm should be possible to solve it.

5.3 Testing a Sharp Null Hypothesis

Suppose that X has a normal distribution with unknown mean θ and known variance σ^2, and we wish to test the null hypothesis $H_0 : \theta = 0$ versus the alternative $H_1 : \theta \neq 0$. Several authors have highlighted sharp differences between Bayesian methods and testing such hypotheses at fixed levels. (See Berger & Sellke, 1987 and Schervish, 1996 for two examples.) We can ask whether such testing strategies are incoherent. For example, suppose again that the loss function is of the following type

$$L(\theta, a) = \begin{cases} c & \text{if } \theta = 0 \text{ and } a = 1, \\ 1 & \text{if } \theta \neq 0 \text{ and } a = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

Then, the risk function of the uniformly most powerful unbiased (UMPU) level α_0 test δ_{α_0} has the following property for every $\alpha_0 < 1$:

$$\lim_{|\theta| \to \infty} R(\theta, \delta_{\alpha_0}) = 0.$$

It follows that, for every test δ,

$$\limsup_{|\theta| \to \infty} R(\theta, \delta) - R(\theta, \delta_{\alpha_0}) \geq 0.$$

Hence, if we combine finitely many such risk function trades, the combination will also have nonnegative lim sup as $|\theta| \to \infty$. It follows that we cannot make Dutch book by trading risk functions.

One could argue that, in practical problems, $|\theta|$ is bounded. Since UMPU tests remain UMPU when extreme portions of the parameter space are removed (as long as an interval around the null remains), we could try to make Dutch book with a bounded parameter space. Even this is not possible. The power function of every test is continuous, hence, for every test δ,

$$\lim_{\theta \to 0} R(\theta, \delta) - R(\theta, \delta_{\alpha_0}) = -\frac{1}{c}[R(0, \delta) - R(0, \delta_{\alpha_0})].$$

It follows that, for every finite combination of risk function trades, the value at 0 will have magnitude c times as big as but the opposite sign of the limit of the values as $\theta \to 0$. Hence, these two values cannot both be negative, and we still cannot make Dutch book. The arguments given here apply regardless of how one chooses the level α_0 as a function of σ^2. One could use the same level for all values of σ^2 or one could arbitrarily choose different values of α_0 as σ^2 changes.

If it is coherent to choose the level of the test as an arbitrary function of σ^2, one would expect that there must be at least one prior distribution such that these choices would be Bayes rules (or formal Bayes rules) with respect to each such prior. Indeed, we have been able to identify those prior distributions that have this property when the level is chosen to be the same value α_0 for all values of σ^2.

Theorem 1. *Let μ be a prior such that, for every test δ the trade of the risk function of δ for the risk function of the UMPU level α_0 test has nonnegative expected value for all σ. Then μ is finitely additive and has the following properties. Let $p_0 = \mu(\{0\})$, $q = \inf_{a>0} \mu(\{\theta : 0 < |\theta| < a\})$ and $r = \inf_{b>0} \mu(\{\theta : |\theta| > b\})$. Then $q = cp_0$ and $p_0 + q + r = 1$.*

Proof. Suppose that we trade the risk function of the UMPU level α_1 test for the risk function of the UMPU level α_0 test. The trade has the value $c(\alpha_1 - \alpha_0)$ for $\theta = 0$ and for $\theta \neq 0$, the value is

$$\begin{aligned} &\Phi\left(\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)-\frac{\theta}{\sigma}\right)-\Phi\left(-\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)-\frac{\theta}{\sigma}\right) \\ &\qquad -\Phi\left(\Phi^{-1}\left(1-\frac{\alpha_1}{2}\right)-\frac{\theta}{\sigma}\right)+\Phi\left(-\Phi^{-1}\left(1-\frac{\alpha_1}{2}\right)-\frac{\theta}{\sigma}\right). \end{aligned} \tag{5.5}$$

For each $\theta \neq 0$, the value in (5.5) goes to 0 as $\sigma \to 0$ and it goes to $\alpha_0 - \alpha_1$ as $\sigma \to \infty$. From these facts, it follows that the limit as $\sigma \to 0$ of the expected value of (5.5) is $(\alpha_1 - \alpha_0)(cp_0 - q)$. Unless $q = cp_0$, this can be made negative by appropriate choice of α_1. Hence, $q = cp_0$ is necessary in order for the expected value to be nonnegative.

Next, notice that every prior as described in the theorem has the property that the expected value of (5.5) is 0. Finally, suppose that $p_0 + q + r < 1$. Then there exists a bounded interval $[a, b]$ such that $\mu([a, b]) > 0$. Let $\alpha_1 > \alpha_0$. Then (5.5) is negative for all $\theta \neq 0$ and the expected value is strictly negative. Hence $1 = p_0 + q + r$. $\square$

Although the prior in Theorem 1 gives nonnegative expected value to every risk function trade of the form (5.5), it also gives expected value 0 to every risk function trade between any two UMPU test regardless of what their levels are. That is, all UMPU tests are equally good.

5.4 The Significance Level Depends on the Loss Function

When asked "How should I choose the size of a test?" the classical statistician sometimes responds by saying that the size should depend on the costs of type I and/or type II errors. For example, the more serious is a type I error, the smaller the size should be. Consider loss functions of the form (5.4). The larger c gets, the smaller should be the significance level (size) of the preferred test. Theorem 1 suggests that even this interpretation will not stand up to Dutch book, if the parameter space is bounded. Notice that Theorem 1, when applied to decision problems with different loss functions (different values of c) says that the only prior distributions that support letting the size of the test depend only on c are the ones for which $\Pr(|\theta| > b) = 1$. If the parameter space is bounded, no prior will support such preferences.

Example 7. Suppose that we have two different decision problems both with loss functions of the form (5.4) but with two different values of c, c_0 and c_1. Call these problem 0 and problem 1 respectively. Suppose that the parameter space is bounded, that is, we know that $|\theta| < b$. Suppose that our classical statistician prefers the UMPU level α_i test in problem i for $i = 0, 1$, where α_0 is not necessarily the same as α_1. We can make Dutch book against such preferences. Without loss of generality, assume that $c_1 > c_0$. Suppose that we trade the risk function of a UMPU level β_i test for the risk function of the UMPU level α_i test in problem i. Then the values of these trades at $\theta = 0$ are $c_i(\beta_i - \alpha_i)$ for $i = 0, 1$. The values of such trades for $\theta \neq 0$ will depend on σ, and they will have a form similar to (5.5). The rate of incoherence depends on many factors, including the bound on the parameter space. Here, we shall merely illustrate how Dutch book can be made. Suppose that $\alpha_0 = 0.1$ and $\alpha_1 = 0.05$ while $c_0 = 1$ and $c_1 = 2$. In problem 0, our classical statistician prefers the level 0.1 test to all others regardless of σ, and in problem 1, he/she prefers the level 0.05 test to all others regardless of σ. Consider problem 0 with $\sigma = 1.234$ and $\beta_0 = 0.11$ and problem 1 with $\sigma = 1$ and $\beta_1 = 0.045$. Suppose that the parameter space is bounded at 4, that is $|\theta| \leq 4$. If we combine the two trades, described in this example with coefficients γ and $1 - \gamma$, the value at $\theta = 0$ is $0.01\gamma - 0.01(1 - \gamma) = 0.02\gamma - 0.01$. This will be negative for all $\gamma \in [0, .5)$. With $\gamma = 0.49223$, the value at $\theta = 0$ is -1.55×10^{-4}, and the max for $0 < |\theta| < 4$ is also -1.55×10^{-4}. The value $\gamma = 0.49223$ achieves the rate of incoherence for these two trades for all parameters spaces of the form $\{\theta : |\theta| \leq b\}$ for $b \in [2, 5.4]$. For b outside of this range, a different rate of incoherence occurs.

6. GENERAL RESULTS

Schervish, Seidenfeld & Kadane (1997, 2002a) prove several general results about incoherent previsions for elements of a partition. To summarize these, let $A_1, \ldots, A_n$ be a partition. That is events $A_1, \ldots, A_n$ must occur and no two of them can occur simultaneously. If upper previsions for these events add up to $q < 1$, then the rate of incoherence is $(1-q)/(n-q)$, $(1-q)/q$ and $(1-q)/n$ for bookie's escrow/sum, gambler's escrow/sum and neutral/sum normalizations respectively. If lower previsions for these events add to $p > 1$, the rate of incoherence for bookie's escrow/sum normalization is $(p-1)/p$. For the other two normalizations based on sum, the rate of incoherence depends on whether any partial sums of the lower previsions are strictly more than 1.

Example 8. Let A_1, A_2, A_3 form a partition. Suppose that the lower previsions of 0.1, 0.8 and 0.7 are given for these events respectively. Suppose that we use neutral/sum normalization. The rate of guaranteed loss for the combination of gambles $(I_{A_2} - 0.8) + (I_{A_3} - 0.7)$ is 0.25. If we use all three gambles, the rate is 0.2. Hence, the maximal rate is achieved by combining fewer than three gambles.

Additional results concern a partition together with a simple random variable X measurable with respect to that partition. That is, $X = \sum_{i=1}^{n} x_i I_{A_i}$. Let p_i be both an upper prevision and a lower prevision for A_i for $i = 1, \ldots, n$. Let p_X be both an upper prevision and a lower prevision for X. Let $\mu = \sum_{i=1}^{n} x_i p_i$ and $\delta = p_X - \mu$. Schervish, Seidenfeld & Kadane (1997) address the question of how much, if any, the added prevision for X increases the rate of incoherence already computed for the previsions of the partition elements. Not surprisingly, the degree to which the added prevision increases the rate of incoherence depends on both δ and how incoherent the previsions are for the events in the partition. Indeed, there are cases in which the previsions of the events in the partition are sufficiently incoherent that the additional prevision might not increase the rate of incoherence at all if δ is small enough. To summarize these results, for bookie's escrow/sum normalization, let $q = \sum_{i=1}^{n} p_i$, and assume that $x_1 < x_2 < \cdots < x_n$. Then the values of p_X that do not increase the rate of incoherence above what it was for the previsions of the partition elements are as follows:

$$
\begin{aligned}
\mu + \frac{1-q}{n-1} \sum_{i=1}^{n-1} x_i \leq p_X \leq \mu + \frac{1-q}{n-1} \sum_{i=2}^{n} x_i \qquad & \text{if } q < 1, \\
\max\{x_1, \mu - (q-1)x_n\} \leq p_X \leq \min\{x_n, \mu - (q-1)x_1\} \qquad & \text{if } q > 1, \\
p_X = \mu \qquad & \text{if } q = 1.
\end{aligned}
$$

For the gambler's escrow/sum normalization, when $q < 1$, the corresponding set of p_X is $\mu + (1-q)x_1 \leq p_X \leq \mu + (1-q)x_n$. This last range has an interesting interpretation. The p_X values that do not increase the rate of incoherence are those formed by computing the "expected value" of X using the incoherent previsions and then placing the remaining probability $1-q$ anywhere else between the smallest and largest possible values of X. Schervish, Seidenfeld & Kadane (2000) describe how the incoherent previsions that add up to $q < 1$ are related to lower probabilities from ϵ-contamination models. In particular, the lower and upper expectation of the simple random variable X discussed above turn out to be $\mu + (1-q)x_1$ and $\mu + (1-q)x_n$.

7. DISCUSSION

In this article we introduce several 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. Each of our indices 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 a normalization. We introduced some properties that we want a normalization to have, and we identified largest and smallest normalizations amongst special classes of normalizations.

We then illustrated the methods for measuring incoherence with several examples of incoherent inferences. These included testing simple hypotheses at fixed levels, minimax estimation, and choosing the size of a test based solely on the loss function.

ACKNOWLEGMENTS

This research was supported by NSF grant DMS-9801401.

REFERENCES

Berger, J. O. & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of P values and evidence (with Discussion). *J. Am. Statist. Assoc.* **82**, 112–122.

Cox, D. R. (1958). Some problems connected with statisitical inference. *Ann. Math. Statist.* **29**, 357–363.

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

Feller, W. (1972). *An Introduction to Probability Theory and Its Applications*, 2 ed. vol. 2, New York: Wiley.

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

Schervish, M. J. (1996). *P*-values: What they are and what they are not. *Amer. Statist.* **50**, to appear.

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

Schervish, M. J., Seidenfeld, T. & Kadane, J. B. (2000). How sets of coherent probabilities may serve as models for degrees of incoherence. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **8**, 347–355.

Schervish, M. J., Seidenfeld, T. & Kadane, J. B. (2002a). How incoherent is fixed-level testing? In *Proceedings of the 2000 Meeting of the Philosophy of Science Association*, Ed. J. Barett, pp. to appear. Chicago: Philosophy of Science Association. Also available as T.R.704, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, 15213, USA.

Schervish, M. J., Seidenfeld, T. & Kadane, J. B. (2002b). Measuring incoherence. *Sankhya* **??**, to appear.