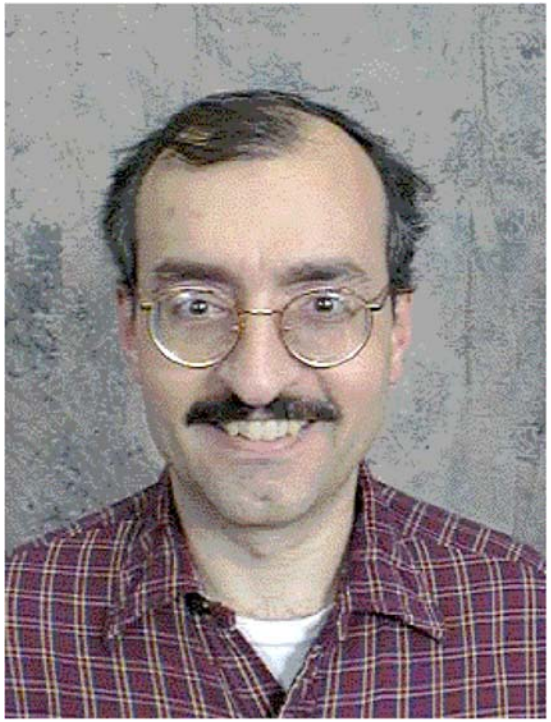
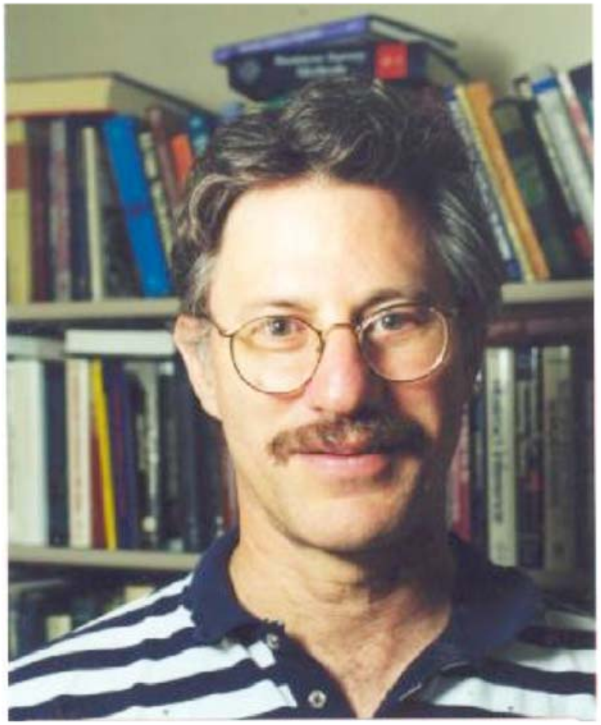
Degrees of Incoherence: a framework for Bayes/non-Bayes compromises

Or, How I learned to Reduce my Incoherence

Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane
Carnegie Mellon University



Mark Schervish



Jay Kadane

Outline

- **De Finetti's *coherence* game, adapted for 1-sided wagers**

- **Modifying the coherence game to allow for *rates* of incoherence**

    - **A theory of *escrow* for normalizing sure-gains from a Book**

    - **Different escrows, and their purposes.**

- **Two Applications**
    - **How incoherent are Non-Bayes Statistical procedures?**

      Setting the level of a statistical test as function of sample size.

    - **How to make decisions from an incoherent position?**

      You don't have to be Coherent to use Bayes' rule!

Begin with a sketch of de Finetti's *Book* argument for coherent wagering.

A Zero-Sum (sequential) game is played between a *Bookie* and a *Gambler*, with a *Moderator* supervising.

Let X be a random variable defined on a space Ω of possibilities, a space that is well defined for all three players by the Moderator.

The *Bookie's* prevision $p(X)$ on the r.v. X has the operational content that,

when the *Gambler* fixes a real-valued quantity $\alpha_{X,\, p(X)}$

then the resulting payoff to the *Bookie* is

$$\alpha_{X,\, p(X)} [\, X - p(X) \,],$$

with the opposite payoff to the Gambler.

A simple version of de Finetti's *Book* game proceeds as follows:

1. **The *Moderator* identifies a (possibly infinite) set of random variables $\{X_i\}$.**
2. **The *Bookie* announces a prevision, a *fair price* $p_i = p(X_i)$ for buying and selling each r.v. in the set $\{X_i\}$.**
3. **The *Gambler* then chooses (*finitely many*) non-zero terms $\alpha_i = \alpha_{X_i, p(X_i)}$.**
4. **The *Moderator* settles up and awards Bookie (Gambler) the respective SUM of his/her payoffs:** ***Total payoff to Bookie*** $= \sum_{i=1}^{n} \alpha_i [X_i - p_i]$.

Total payoff to Gambler $= -\sum_{i=1}^{n} \alpha_i [X_i - p_i]$.

***Definition*:**

The *Bookie'*s previsions are *incoherent* if the *Gambler* can choose terms α_i that assures her/him a (*uniformly*) positive payoff, regardless which state in Ω obtains – so then the *Bookie* loses for sure.

A set of previsions is *coherent*, if not incoherent.

***Theorem* (de Finetti):**

A set of previsions is coherent *if and only if*

each prevision *p(X)* is the expectation for *X* under a common (finitely additive) probability *P*.

That is, $$p(X) = E_{P(\bullet)}[X] = \int_\Omega X \, dP(\bullet)$$

***Two Corollaries*:**

***Corollary 1*: When the random variables are *indicator functions* for events $\{E_i\}$, so that the gambles are simple bets – with the α's then the stakes in a winner-take-all scheme**

The previsions p_i are coherent

if and only if

Each prevision is the probability $p_i = P(E_i)$, for some (f.a.) probability P.

On conditional probability:

Definition: A *called-off prevision* $p(X \| E)$ for X,

made by the *Bookie* on the condition that event E obtains,

has a payoff scheme to the Bookie*:* $\alpha_{X\|E}\, E[\, X - p(X \| E)\,]$.

Corollary 2: Then a called-off prevision $p(X \| E)$ is coherent alongside the

(coherent) previsions $p(X)$ for X, and $p(E)$ and E *if and only if*

$p(X \| E)$ is the *conditional expectation* under P for X, given E.

That is, $p(X \| E) = E_{P(\cdot\,|E)}[X] = \int_\Omega X\, dP(\bullet|E)$ and is $P(X \mid E)$ if X is an event.

- In this sense, the *Bookie*'s conditional probability distribution $P(\bullet|E)$ is the norm for her/his *static called-off* bets.
- Coherence of *called-off* previsions is not to be confused with the norm for a *dynamic learning rule*, e.g., when the Bookie learns that *E* obtains.

There are two aspects of de Finetti's coherence criterion that we relax.

1. Previsions may be *one-sided*, to reflect a difference between *buy* and *sell* prices for the *Bookie*, which depends upon whether the *Gambler* chooses a *positive* or *negative* α-term in the payoff $\alpha_{X,\, p(X)}\, [\, X - p(X)\,]$ **to the *Bookie*.**

For positive values of α, allow the *Bookie* to fix a <u>maximum</u> *buy*-price.

- **Betting on event E, this gives the *Bookie*'s lower probability $p_*(E)$,**

$$\alpha^{+}\, [\, E - p_*(E)\,].$$

For negative values of α, allow the *Bookie* to fix a <u>minimum</u> *sell*-price.

- **Betting against event E, this gives the *Bookie*'s upper probability $p^*(E)$,**

$$\alpha^{-}\, [\, E - p^*(E)\,].$$

At odds between the lower and upper probabilities, the *Bookie* rather not wager!

This approach has been explored for more than 50 years!

***(See* http://www.sipta.org/ the *Society for Imprecise Probabilities, Theories and Practices*)**

For example, when dealing with upper and lower probabilities:

Theorem **[C.A.B. Smith, 1961]**

- **If the *Bookie*'s one-sided betting odds $p_*(\bullet)$ and $p^*(\bullet)$ correspond, respectively, to the minimum and maximum of probability values from a *closed, convex* set of (coherent) probabilities, then the Bookie's wagers are coherent: then the *Gambler* can make no *Book* against the *Bookie*.**

- **Likewise, if the *Bookie*'s one-sided *called-off* odds $p_*(\bullet \,||E)$ and $p^*(\bullet \,||\mathrm{E})$ correspond to the minimum and maximum of conditional probability values, given *E*, from a *closed, convex* set of (coherent) probabilities, then they are coherent.**

2. De Finetti's coherence criterion is dichotomous.

- **A set of (one-sided) previsions is *coherent* – then no *Book* is possible,**

  or it is not, and then the previsions form an *incoherent* set.

BUT, are all incoherent sets of previsions equally *bad*, equally *irrational*?

- **Rounding a coherent probability distribution to 10 decimal places and rounding the same distribution to 2 decimal places may both produce "incoherent" betting odds. Are these two equally defective?**

- **Some Classical statistical practices are non-Bayesian – they have no Bayes models. Are all non-Bayesian statistical practices equally irrational?**

ESCROWS for Sets of Gambles when a Book is possible

In order to normalize the *guaranteed gains* that the *Gambler* can achieve by making *Book* against the *Bookie*, we introduce an ESCROW function.

Let $Y_i = \alpha_i(X_i - p_i)$ be a wager that is *acceptable* to the *Bookie*.

Let $G(Y_1, \ldots, Y_n)$ be the (minimum) *guaranteed gains* to the *Gambler* from a *Book* formed with gambles acceptable to the (incoherent) *Bookie*.

An *escrow function* $e(Y_1, \ldots, Y_n)$ is introduced to normalize the (minimum) guaranteed gains, as follows:

Where H is the intended *measure* or *rate* of incoherence,

$$H(Y_1, \dots, Y_n) = \frac{G(Y_1, \dots, Y_n)}{e(Y_1, \dots, Y_n)}$$

Here are 7 conditions that we impose on an Escrow function,

$$e(Y_1, \dots, Y_n) = f_n(Y_1, \dots, Y_n) .$$

1. **For one (simple) gamble, Y, the player's escrow**

$$e(Y) = f(Y) = Z$$

is her/his *maximum possible loss* from an outcome of Y.

2. $$e(Y_1, \dots, Y_n) = f_n(e(Y_1), \dots, e(Y_n)) = f_n(Z_1, \dots, Z_n).$$

The escrow of a set of gambles is a function of the individual escrows.

3. $$f_n(cZ_1, \ldots, cZ_n) = c\, f_n(Z_1, \ldots, Z_n) \text{ for } c > 0.$$

Scale invariance of escrows.

4. $$f_n(Z_1, \ldots, Z_n) = f_n(Z_{\pi(1)}, \ldots, Z_{\pi(n)})$$

Invariance for any permutation $\pi(\bullet)$.

5. $f_n(Z_1, \ldots, Z_n)$ **is non-decreasing and continuous in each of its arguments.**

6. $$f_n(Z_1, \ldots, Z_n, 0) = f_n(Z_1, \ldots, Z_n)$$

When a particular gamble carries no escrow, the total escrow is determined by the other gambles.

7. $$f_n(Z_1, \ldots, Z_n) \leq \Sigma_i Z_i$$

The total escrow is bounded above by the sum of the individual escrows.

Then:

- **As a lower bound,** $f_n(Z_1, \ldots, Z_n) \geq Max\{Z_i\}$

- **Thus, with** $e(Y_1, \ldots, Y_n) = Max\{Z_i\}$,

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

is the largest possible (least charitable) measure.

- **Thus when** $e(Y_1, \ldots, Y_n) = \Sigma_i Z_i$, **then** H **is the smallest (most charitable) measure of incoherence.**

Here we work with the most charitable measure of incoherence:

The total escrow for a set of gambles is the *sum* of the individual escrows.

When the escrow reflects the (incoherent) Bookie's *exposure* in the set of gambles, we call the measure *H* the *Bookie*'s guaranteed rate of loss.

When the escrow reflects the *Gambler*'s exposure, we call the measure *H* the *Gambler*'s *guaranteed rate of gain*.

Also, we have a third perspective, *neutral* between the *Bookie*'s and *Gambler*'s exposures, which we use for singly incoherent previsions, as might obtain with failures of mathematical or logical omniscience.

The third (*neutral*) perspective uses an escrow: $e(Y) = |\alpha|$.

In the case of simple bets, this escrow is the magnitude of the stake.

The *neutral* escrow results in a measure of coherence *H* that is *continuous* in both the random variables and previsions, unlike the case with the measures of guaranteed rates of loss or gain, above.

Some basic results in this theory

Let $\{E_1, \ldots., E_n\}$ form a partition, and let $0 \leq p_*(E_i) \leq p^*(E_i) \leq 1$ be the Bookie's lower and upper probabilities for these events.

So, we assume that no prevision is incoherent alone.

$$\text{Let } \sum_{i=1}^{n} p_*(E_i) = q \quad \text{and} \quad \sum_{i=1}^{n} p^*(E_i) = r, \text{ and}$$

So, the *Bookie* is incoherent if either $q > 1$ or $r < 1$.

Theorem **(for *rate of loss* – the *Bookie*'s escrow):**

(1) If $\sum_{i=1}^{n} p_*(E_i) > 1$, then the *Gambler* maximizes the guaranteed rate of loss by choosing the stakes (α's) equal and positive. $H = [q - 1] / q$

(2) If $\sum_{i=1}^{n} p^*(E_i) < 1$, then the *Gambler* maximizes the guaranteed rate of loss by choosing the stakes (α's) equal and negative. $H = [1 - r] / [n - r]$

(3) If the $p_*(E_i)$, $p^*(E_i) \neq 0$, then these *maximin* solutions are unique.

What about efficient Bookmaking from the perspective of the *Gambler*'s escrow, the *guaranteed rate of gain*?

***Example*: If the *Bookie*'s incoherent lower odds are (.6, .7, .2) on {E_1, E_2, E_3}, then we note the following, by the previous *Theorem*:**

Equal stakes ($\alpha_1 = \alpha_2 = \alpha_3 > 0$) maximizes the *rate of loss*, with $H = 1/3$.

Then, since the *Gambler*'s escrows has the same total in this case as the *Bookie* under this strategy, equal stakes by *Gambler* produces a *rate of gain* of 1/3.

- **However, the *Gambler* can improve on this rate, upping it to 3/7,**

by setting $\alpha_1 = \alpha_2 > 0$ and setting $\alpha_3 = 0$.

This situation is generalized as follows.

Reorder the atoms so that the *Bookie*'s odds are not decreasing:

$p_j \geq p_i$ whenever $j \geq i$. Again, assume that $0 \leq p_j \leq 1$.

Theorem (for *rate of gain*– the *Gambler*'s escrow):

(1) If $\sum_{i=1}^{n} p^*(E_i) = r < 1$, then the *Gambler* maximizes the *rate of gain* by choosing the stakes equal and negative.

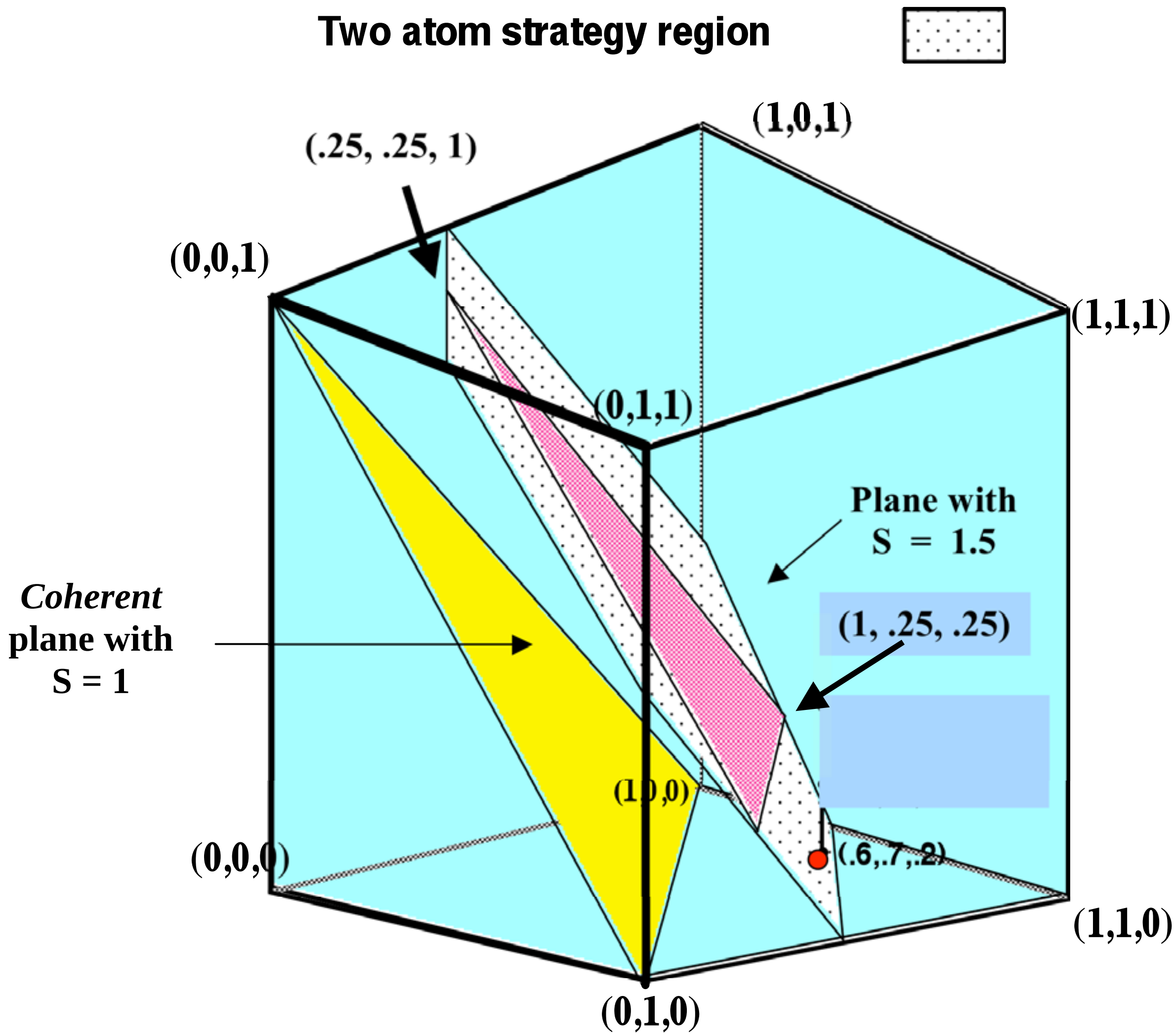
(2) If $\sum_{i=1}^{n} p_*(E_i) = q > 1$, then the *Gambler* maximizes the *rate of gain* by choosing the stakes according to the following rule:

Let k^* be the first k such that $\sum_{i=n-k+1}^{n} p_{*i} \geq 1 + (k-1)p_{n-k}$

with $k^* = n$ if this equality always fails.

Then the *Gambler* sets the α_i all equal and positive for $i \geq n-k^*+1$,

and sets $\alpha_i = 0$ for all $i < n - k^*$.



For the ***rate of gain***, when the ***Bookie***'s incoherent previsions lie in the dotted region the ***Gambler*** uses only 2 previsions, but uses all 3 in the pink region.

***Application-1*: Statistical Hypothesis Testing at a Fixed (.05) level (See Cox, 1958)**

Null hypothesis $\mathbf{H_0}$**:** $X \sim \mathrm{N}[0, \sigma^2]$ ***vs.*** ***Alternative hypothesis*** $\mathbf{H_1}$**:** $X \sim \mathrm{N}[1, \sigma^2]$

Testing a *simple* null *vs* a *simple* alternative, so that the *N-P* Lemma applies.

For each value of the variance, as might result from using different sample sizes, by the N-P Lemma there is a family of *Most Powerful* (best) Tests.

Let us examine the familiar convention to give preference to tests of level $\alpha = .05$**.**

α is the chance of a type-1 error. β is the chance of a type-2 error.

Table of the best β-values for seven α-values and six σ-values.

σ =	.250	.333	.400	.500	1.000	1.333
α	*best* β-values					
.010	.047	.250	.431	.628	.908	.942
.030	.017	**.131**	.268	.452	.811	.871
.040	.012	.106	.227	.401	.773	.841
.050	.009	**.088**	.196	.361	.740	**.814**
.060	.007	.074	.172	.328	.710	.789
.070	.006	.064	.153	.300	.683	**.766**
.100	.003	.043	.111	.236	.611	.702

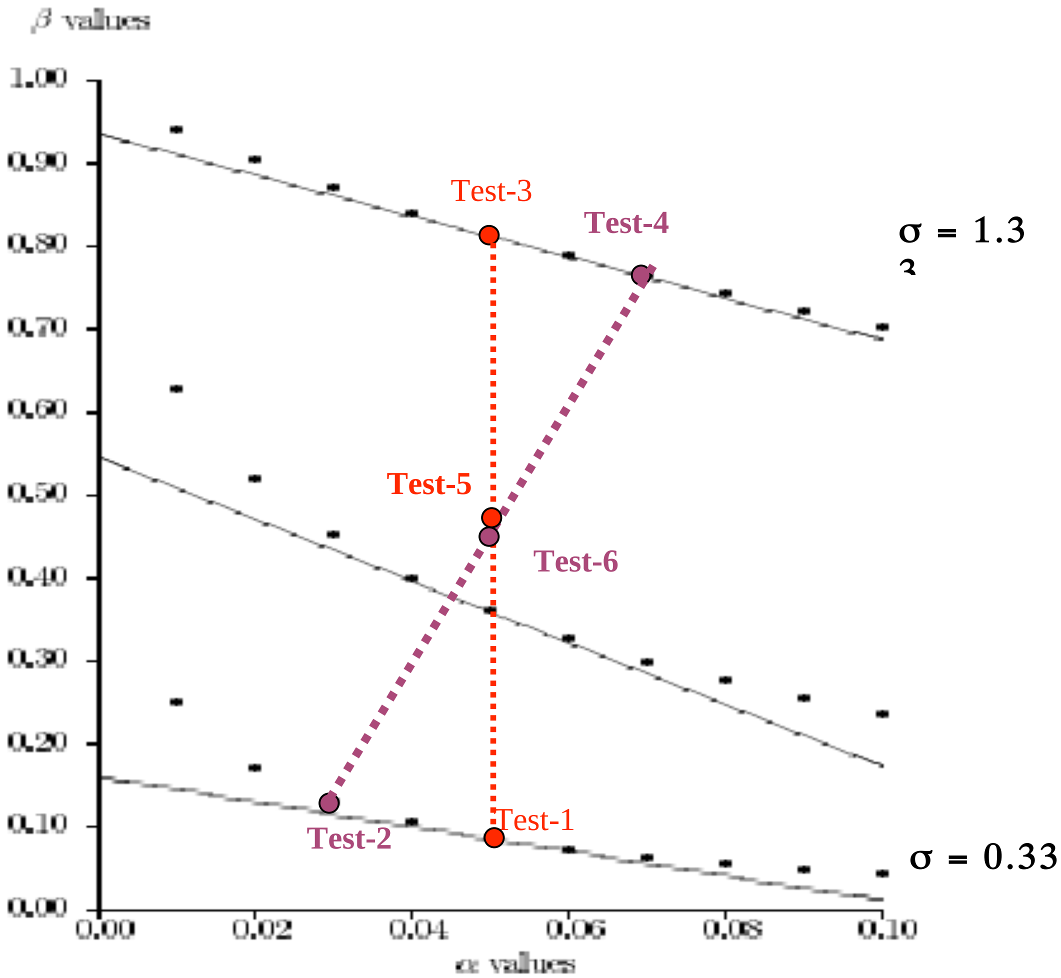
With the convention to choose the best test of level α = .05, the following results:

With σ = 1.333, Test_1: (α = .05; β = .814) is chosen over Test_2: (α = .07; β = .766).

With σ = 0.333 Test_3: (α = .05; β = .088) is chosen over Test_4: (α = .03; β = .131).

But the mixed Test_5 = .5 $\text{Test}_1 \oplus$ .5 Test_3 has (α = .05; β = .451).

Whereas mixed Test_6 = .5 $\text{Test}_2 \oplus$ .5 Test_4 has (α = .05; β = .449), which is better!


β values
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00
0.02
0.04
0.06
0.08
0.10
α values
Test-3
Test-4
Test-5
Test-6
Test-1
Test-2
σ = 1.33
σ = 0.33

To apply our measures of incoherence, we have to get the Statistician to wager.

A *Classical* (non-Bayesian) Statistician will not admit to (non-trivial) odds on the rival hypotheses in this problem, but will compare tests by their RISK, so see if one (weakly) dominates another. In which case the dominated test is *inadmissible*.

The *RISK* (loss) function R of a statistical test T of H_0 vs H_1.

$$R(\theta, T \mid \sigma) = \begin{cases} \alpha(\sigma) & \text{if } \theta = 0 \text{ (the level of the test)} \\ \beta(\sigma) & \text{if } \theta = 1 \text{ (the chance of a type-2 error)} \end{cases}$$

A Classical Statistician who follows the *convention* prefers admissible tests at the .05 level over other tests.

This Statistician may be willing to trade away (to payout) the risk of the preferred test in order to receive (to be paid) the risk of another test, with a different level.

Trading RISKS between tests this way is represented by:

$$R(\theta, T_{\alpha(\sigma)} \mid \sigma) \text{ -- } R(\theta, T_{.05} \mid \sigma) = \begin{cases} \alpha(\sigma) - .05, & \text{if } \theta = 0 \text{ (the null)} \\ \beta_{T_{\alpha(\sigma)}}(\sigma) - \beta_{T_{.05}}(\sigma), & \text{if } \theta = 1 \text{ (alternative)} \end{cases}$$

which *is* of the form of a de Finetti *prevision*:

$$= a(E - b)$$

where $E = H_0$, **i.e. the null hypothesis** $\theta = 0$

$$a = [\alpha(\sigma) - .05 + \beta_{T_{\alpha(\sigma)}}(\sigma) - \beta_{T_{.05}}(\sigma)]$$

and $b = [\beta_{T_{.05}}(\sigma) - \beta_{T_{\alpha(\sigma)}}(\sigma)] \;/\; [\alpha(\sigma) - .05 + \beta_{T_{.05}}(\sigma) - \beta_{T_{\alpha(\sigma)}}(\sigma)]$

Here is a chart of the *rate of loss* to the Classical Statistician who trades .05-level tests based on two samples of sizes (n_0, n_1). Each curve is identified by the size of the first sample, n_0.

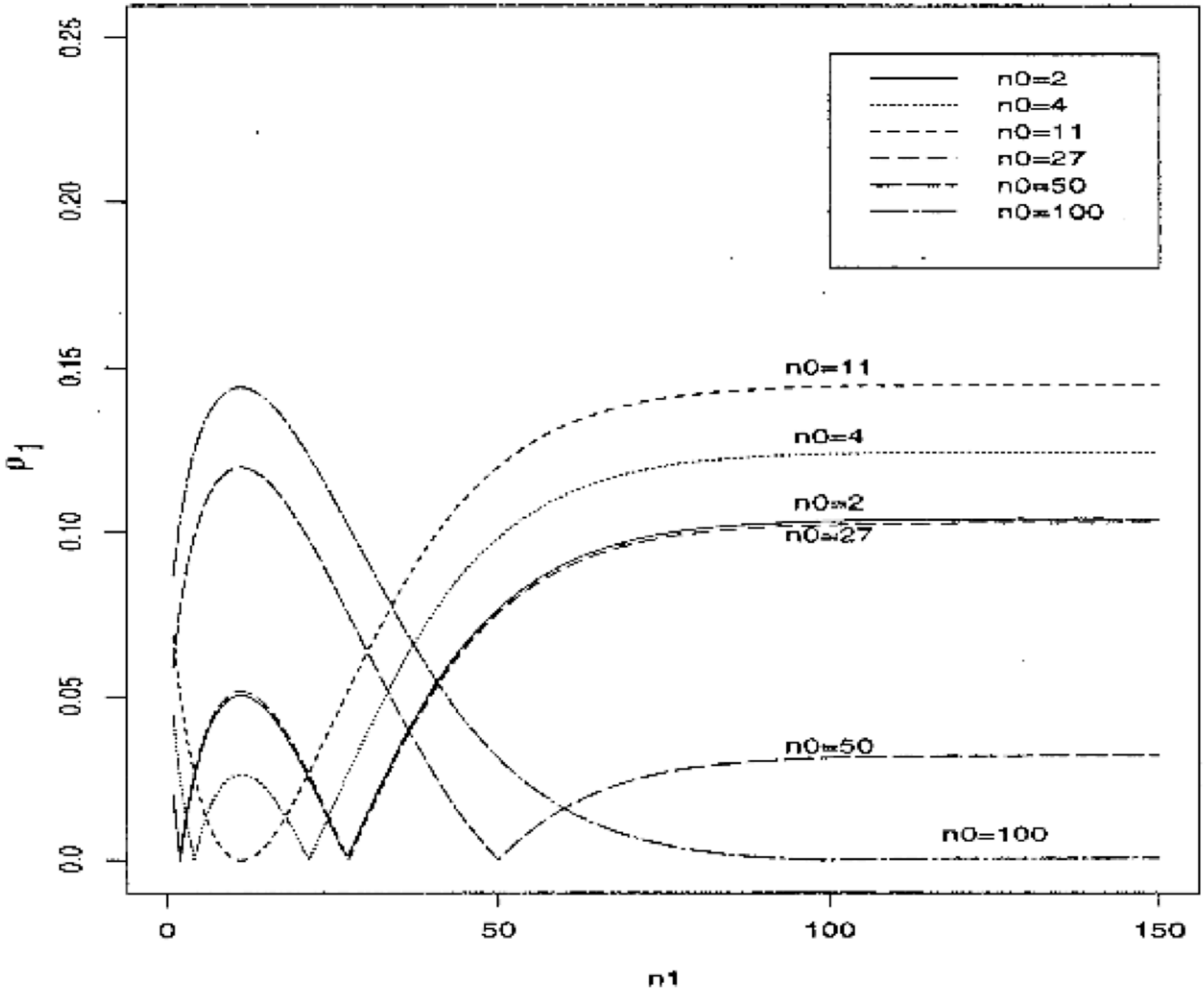


Figure 1. Plot of ρ_1 for level 0.05 testing as a function of n_1 (running from 1 to 150) for various values of n_0 with $c_0 = 19$.

Application-2: **How to wager from an incoherent position.**

Aside: **We restrict ourselves to previsions, rather than using lower and upper previsions, in order to simplify the analysis of the *Gambler*'s optimal strategy.**

As before, let $\{E_1, \ldots., E_n\}$ form a partition, and let $0 \leq p(E_i) \leq 1$ be the Bookie's previsions for these *n*-many events.

Again, we assume that no one of these previsions is incoherent, by itself.

Let $\sum_{i=1}^{n} p(E_i) = q$. It might be that $q \neq 1$, so that the *Bookie*'s previsions are incoherent.

- **Now, the Moderator introduces a new random variable *X*, measurable with respect to this partition, i.e., $X = \sum_i x_i E_i$, and calls upon the *Bookie* to give a prevision for *X*, $p(X)$.**

- **What can the Bookie do with the value of $p(X)$ to avoid increasing her/his measure of incoherence?**

For notational ease, order the events so that $x_1 \leq x_2 \leq \ldots \leq x_n$.

As before, we assume that $x_1 \leq p(X) \leq x_n$, so that by itself $p(X)$ is coherent.

Define $\mu = \sum_i x_i p_i$

You may think of μ as the *pseudo-expectation* for X with respect to the *Bookie*'s incoherent *distribution* $P(\bullet)$ for the x_i.

Theorem for the ***rate of loss* – using the *Bookie*'s perspective on escrow**:

The *Bookie* can avoid increasing the *rate of loss* with a previsions for *X*, as follows:

- **If $q < 1$, choose $p(X)$ to satisfy**

$$\mu + \frac{1-q}{n-1}\sum_{i=1}^{n-1} x_i \;\leq\; p(X) \;\leq\; \mu + \frac{1-q}{n-1}\sum_{i=2}^{n} x_i$$

- **If $q > 1$, choose $p(X)$ to satisfy**

$$\mathit{max}\{\, x_1,\ \mu - (q-1)x_n \,\} \;\leq\; p(X) \;\leq\; \mathit{min}\{\, x_n,\ \mu - (q-1)x_1 \,\}$$

- **If $q = 1$, choose $p(X)$ to satisfy the Bayes solution**

$$\mu = p(X).$$

Theorem **for the *rate of gain* – using the *Gambler'*s escrow:**

The *Bookie* can avoid increasing the *rate of gain* by setting a prevision for *X* as:

Choose $p(X)$ to satisfy

$$\mu + (1-q)x_1 \leq p(X) \leq \mu + (1-q)x_n$$

***Corollary*: *You don't have to be coherent to like Bayes' rule*!**

Consider a ternary partition $\{E_1, E_2, E_3\}$ with previsions $\{p_1, p_2, p_3\}$.

Let X be the *r.v.* for the called-off wager on E_3 *vs* E_1, called-off if E_2 obtains.

$$\underline{E}_1 \qquad\qquad \underline{E}_2 \qquad\qquad \underline{E}_3$$

$$X(E_1) = 0, \quad X(E_2) = p(X), \quad \text{and} \quad X(E_3) = 1$$

Thus, $\alpha(X - p(X))$ has the respective payoffs:

$$-\alpha p(X) \qquad\qquad 0 \qquad\qquad \alpha(1 - p(X))$$

Then, e.g., with $q < 1$, the *Bookie* wants to satisfy the inequalities:

$$p_2 p(X) + p_3 \leq p(X) \leq p_2 p(X) + p_3 + (1-q)$$

If the *Bookie* uses a pseudo-Bayes value, the inequality is *automatic*, as follows:

$p(X) = p(E_3 \,||\, \{E_1, E_3\}) = p_3/(p_1+p_3)$ = "as if" calculating $p(E_3 \,|\, \{E_1, E_3\})$.

Hence, betting like a coherent Bayesian makes sense even if you are incoherent!

Summary

- **De Finetti's dichotomous theory of 2-sided (*fair*) previsions may be relaxed to permit measures of incoherence for 1-sided (*lower* and *upper*) previsions.**

- **There is more than one measure of incoherence, reflecting different perspectives: *escrow* functions, used for normalizing sure-losses from a *Book*.**

- **These measures of incoherence may be applied to modulate longstanding debates over Classical *vs.* Bayesian statistical methods.**

- **It is feasible to reason from an incoherent position, to determine what new previsions will not increase the already existing rate of incoherence.**

Selected References

Cox, D.R. (1958) Some problems connected with statistical inference. *Ann. Math. Stat.* 29, 357-363.

deFinetti, B. (1974) *Theory of Probability*, Vol. 1. Wiley: New York.

Schervish, M.J., Seidenfeld,T., and Kadane, J.B. (1997) Two Measures of Incoherence. Tech. Report #660, Department of Statistics, Carnegie Mellon University, Pgh. PA 15213. (Postscript available @ http://www.stat.cmu.edu/.)

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

Schervish, M.J., Seidenfeld, T., and Kadane, J.B. (2002) A rate of incoherence applied to fixed-level testing. *Phil. Sci.* 69 #3 S248-S264.

Schervish, M.J., Seidenfeld, T., and Kadane, J.B. (2002) Measuring Incoherence *Sankhya* 64 A, 561-587.

Schervish, M.J., Seidenfeld, T., and Kadane, J.B. (2003) Measures of Incoherence. In *Bayesian Statistics 7*. Bernardo, J.M. et al (eds.). Oxford Univ. Press: Oxford.

Smith, C.A.B. (1961) Consistency in Statistical Inference and Decision *J. Royal Stat. Soc.* B, 23 1-25.