Rates of Incoherence and IP Theory (again!)
A modest proposal to use *Rates of Incoherence* (1999)
as a guide for personal uncertainties about logic and mathematics.

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

Outline

- **Savage's challenge.**

  How to represent uncertainty so that you may learn from

  ***mere thinking and/or computing*?**

  DeFinetti's *Prevision Game*: coherent previsions for math/logical constants.

- **Two existing strategies for replying to Savage's challenge:**
  - **Relax the algebraic closure conditions for measurable spaces**
  - **I.J.Good's *Statistician's Stooge* and the failure of *Total Evidence***

- **A third strategy:**

  Modify the *Prevision Game* to allow for *rates* of incoherence

  - **A simple application with "data" from computations.**

In his (1967) *Difficulties in the theory of personal probability*, Savage writes,

The analysis should be careful not to prove too much; for some departures from theory are inevitable, and some even laudable. For example, a person required to risk money on a remote digit of π would, in order to comply fully with the theory, have to compute that digit, though this would really be wasteful if the cost of computation were more than the prize involved.

For the postulates of the theory imply that you should behave in accordance with the logical implication of all that you know.

Is it possible to improve the theory in this respect, making allowance within it for the cost of thinking, or would that entail paradox, as I am inclined to believe but unable to demonstrate? If the remedy is not in changing the theory but rather in the way in which we are to attempt to use it, clarification is still to be desired.

- **But why is it that *the postulates of the theory imply that you should behave in accordance with the logical implication of all that you know*?**

Begin with a review of deFinetti's *Book* argument for coherent wagering.

The zero-sum (sequential) *Prevision Game* is played between a *Bookie* and a *Gambler*, with a *Moderator* supervising.

Let X be a random variable defined on a space $\Omega = \{\omega_1, \omega_2, \dots\}$ of pairwise-exclusive and mutually-exhaustive possibilities.

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* in state ω is

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

with the opposite payoff to the Gambler.

A simple version of deFinetti'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 $p_i = p(X_i)$ for each r.v. in the set.**

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 in state ω:**

$$\textit{Total payoff to Bookie} = \sum_{i=1}^{n} \alpha_i [X_i(\omega) - p_i].$$

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

Definition: **The *Bookie*'s previsions are *incoherent* if the *Gambler* can choose finitey many 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 **(deFinetti):**

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

***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 – then 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*.**

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

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

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

Corollary 2*: Then a called-off prevision $p(X \parallel E)$ is coherent alongside the (coherent) previsions $p(X)$ for X, and $p(E)$ and E *if and only if

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

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

and is the conditional probability $P(X \mid E)$ if X is an event.

The *Bookie*'s conditional probability $P(\bullet \mid E)$ provides the norm for *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.**

Reflect on Savage's challenge in some detail.

Let X_{π_6} be the variable whose value is the 6th decimal digit of π.

In an ordinary measureable space <Ω, $\mathcal{B}$>,

X_{π_6} is the constant 2,

independent of ω.

In an ordinary measure space, it is certain that the event "X_{π_6} = 2" obtains,

since as a mathematical result, it obtains in each state ω.

Thus, in any ordinary measure space, there is no elbow room for

a non-extreme probability distribuiton about the possible values for X_{π_6}

or for an expectation other than 2 for its value.

- **Any prevision other than P(X_{π_6}) = 2 is incoherent!**

***Strategy #1*: Relax the (algebraic) closure conditions for measurable spaces to accommodate the agent's "boundedly rational" perspective.**

***Instances of Strategy #1*:**

Garber (1983) – Use only sentential operators; no functions, relations, etc.

Gaifman (2004) – Restrict detachment/entailment rules so that logically equivalent expressions are not automatically put into the same equivalence class with respect to a conditional probability P(· | ·).

→ Hacking (1967) – Let it be up to the agent to determine the "space." (?)

De Finetti (1974) – The class of variables that receive a well-defined (coherent) prevision form only a *linear span*, which may be strictly smaller than the space of events formed by the *Boolean closure* of events that receive previsions.

***Two responses to Strategy #1*:**

(*1.1*) (except for Hacking's proposal) Modeling "bounded" rationality:

The weakened closure conditions remain overly restrictive. They do not capture ordinary *uncertainty* about math/logical facts. *YOUR* "bounded" reasoning need not line up with any of these weakened closure rules.

(*1.2*) How does such a "boundly rational" agent make decisions?

None of these proposals includes a decision theory to show how the agent might apply her/his modified personal probabilities.

(and for Hacking): What remains of the old distinction between coherent and incoherent previsions?

***Strategy #2*: Use Good's *Statistician's Stooge* to step around *Total Evidence*.**

- **The *Stooge* can be compelled to censor the data according to the *Statistician*'s directions.**
- **The *Stooge* can learn *X*, but reports only the reduced $g(X) = Y$ to the *Statistician*, according to the *Statistician*'s stipulations about *g*.**

- **Strategic choice of *g* by the *Stooge* allows the *Statistician* to avoid learning too much!**

Continuing Example

Consider a problem in geometric probability that relies on three familiar bits of knowledge from high school geometry.

- **The area of a circle with radius r equals πr^2.**
- **The area of a square is the square of the length of its side.**
- **The Pythagorean Theorem: Given a right triangle, with side lengths a and b and hypotenuse length c, then $a^2 + b^2 = c^2$.**

The *Statistician*'s measure space, $<\Omega, \mathcal{B}, P>$:

Let Ω be the set of points interior to a Circle C with radius r.

Let $\mathcal{B}$ be the algebra of geometric subsets of C generated by ruler-and-compass constructions.

Let P be uniform over points in Ω. A point from Ω is chosen *at random*, with equal probability for congruent subsets of C.

Statistican **knows her probability that the random point falls into region S (∈ $\mathcal{B}$) is the ratio of the** ***area*****(S) to the** ***area*****(C).**

Statistician **is aware that**

$$P(\text{the random point falls into S}) = area(\text{S})/\pi r^2.$$

Let S be a square inscribed inside the Circle C, as in Figure 1.

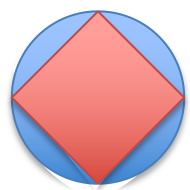


Figure **1**

Then by the Pythagorean theorem and the rule for the area of a square, area(S) = 2*r*² — $\text{area(S)} = 2r^2$**.**

So, ***Statistician*** **is aware that**

$$P(\text{the random point fall into the square S}) = 2/\pi.$$

Connect the *Example* with Savage's challenge, as follows:

> **Suppose *Statistician* is aware that the first five decimals digits in the expansion of π are 3.14159. She cannot identify the 6th decimal digit of π.**

Using the familiar long division algorithm, then *Statistician* is unable to calculate precisely her personal probability, ($2/\pi$), beyond the first 4 digits (0.6366) that the random point is in S.

- **She is unaware of the *<u>value</u>* of her personal probability.**
- **She knows that the 5th digit of her personal probability is either 1 or 2.**

But, e.g, then she is unable to answer whether a bet that the random point is in S at odds of .63662 : .36338 is favorable, fair, or unfavorable.

Application of *Strategy #2* to the *Continuing Example*:

Use a *Statistician's Stooge* to substitute a quantity, θ, for the original uncertain quantity X_{π_6} that the *Stooge* knows (but the *Statistican* does not know) is coextensive with X_{π_6}.

Then *Statistician* may hold non-extreme but coherent probabilities about the substitute variable θ. In this way, familiar numerical methods, including Monte Carlo methods, permit *Statistician* to learn about X_{π_6} by shifting the failure of the *Total Evidence* principle to the *Stooge*.

As an instance of I.J Good's *Statistician's Stooge*, *Stooge*, creates an elementary statistical estimation problem for the quantity 2/π using *iid* repeated draws from the uniform distribution on the circle C.

Stooge **chooses C with center at the origin (0,0) and radius** $r = \sqrt{2}$**.**

Then the inscribed square S has corners with coordinates (±1, ±1).

Let $X_i = (X_{i1}, X_{i2})$ $(i = 1, \ldots, n)$ **be** n **random points drawn by the** ***Stooge*** **using the uniform distribution on C.**

After each draw the Stooge determines whether or not $X_i \in S$**, i.e., whether or not both inequalities obtain:** $-1 \leq X_{ij} \leq +1 \quad (j = 1, 2)$**, which involves examining only the first significant digit of** X_{ij}**.**

Now, the ***Stooge*** **tells** ***Statistican*** **whether the event** Y **occurs on the** i^{th} **trial,** $Y_i = 1$**, if and only if** $X_i \in S$ **for a region S. All the** ***Stooge*** **tells Statistician about the region S is that it belongs to the algebra** $\mathcal{B}$**.**

The Y_i form an *iid* sequence of Bernoulli(θ) variables, with θ = area(S)/2π.

As it happens, θ = 2/π. But this identity is suppressed in the following analysis, with which both *Statistician* and the *Stooge* concur.

Both know that $\sum_{i=1}^{n} Y_i$ is Binomial(n,θ).

Let $\overline{Y}_n = \sum_{i=1}^{n} Y_i/n$ denote the sample average of the Y_i.

$\overline{Y}_n$ is a *sufficient statistic* for θ, i.e., a summary of the *n* draws X_i that preserves all the relevant evidence in a coherent inference about θ based on the data of the n-many *iid* Bernoulli(θ) draws.

The *Stooge* samples with $n = 10^{16}$, obtains $\overline{Y}_n = 0.63661977236$ and carries out ordinary Bayesian reasoning with *Statistician* about the Binomial parameter θ using Statistician's "prior" for θ.

According to what the *Stooge* tells *Statistician*, θ is an uncertain Bernoulli quantity of no special origins.

For convenience, suppose that *Statistician* uses a (uniform) conjugate Beta(1, 1) "prior" distribution for θ, denoted here as P(θ).

So, the *Stooge* reports and given these data, Statistician's "posterior" probability is greater than .999, that $0.63661971 \leq \theta \leq 0.63661990$.

Then, since the *Stooge* knows that $\theta = 2/\pi$, *Stooge* reports that Statistician's probability is at least .999 that the 6^{th} digit of π is 2.

Of course, in order for *Statistician* to reach this conclusion she has to rely on the *Stooge* to suppress the information that S is an inscribed square within C, rather than some arbitrary geometric region within the algebra of ruler-and-compass constructions.

But this information is not fully ignorable, since *Stooge* needs this particular information in order to determine the value of each Y_i.

Response to this use of Good's *Statistician's Stooge*:

(2) How is *Statistician* to formulate precisely what she knows about π and X_{π_6} so as to create the appropriate replacement variable, *e.g.*, *Y* in the Continuing Example, for the *Stooge*?

***Strategy #3*: Modify the *Prevision Game* to allow for *rates* of incoherence**

There are two aspects of deFinetti'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 $\alpha_{X,\,p(X)}\,[\,X(\omega) - p(X)\,]$ to the *Bookie*.**

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

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

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

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

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

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

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

This approach has been explored for more than 50 years!

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 infimum and supremum of probability values from a *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 infimum and supremum of conditional probability values, given *E*, from a *convex* set of (coherent) probabilities, then they are coherent.**

2. **deFinetti'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* $e(Y_1, \ldots, Y_n)$ normalizes the (minimum) guaranteed gains:

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

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

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

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

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

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

   If a 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.

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.

- ***How to reason from an incoherent position.***

  ***Aside*: Here I express results for determinate previsions, rather than working with lower and upper previsions, in order to simplify the analysis of the *Gambler*'s optimal strategy.**

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.

That is, 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 (illustrated for the ***rate of loss*** **– the *Bookie*'s perspective on escrow)**:

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

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**

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

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

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

Corollary 1: ***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.

E_1	E_2		E_3
$X(E_1) = 0$,	$X(E_2) = p(X)$,	and	$X(E_3) = 1$

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

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

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 $p(E_3 \mid \{E_1, E_3\})$

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

***Corollary 2*:**

Let Θ be a finite dimensional parameter space.

Let $p(\theta) > 0$ be possibly incoherent non-extreme, *pseudo-prior density function*.

Suppose, that a *pseudo-likelihood density function* $p(X = x \mid \theta)$ has a 0-rate of incoherence, i.e., its conditional probabilities are coherent.

Suppose, also, they are distinct likelihoods for different θ.

Let X_i ($i = 1, \ldots$) form a sequence of conditionally *iid* variables, given θ, according to $p(X = x \mid \theta)$.

Use the *pseudo-Bayes-algorithm* to create a sequence of *pseudo-posterior functions* $p_n(\theta \mid X_{1,\ldots,} X_n)$, $n = 1, \ldots$.

Then, almost surely with respect to the true state, $\theta^* \in \Theta$,

- **the *Neutral* rate of incoherence for the *pseudo-posterior* converges to 0**
- **and that *pseudo-posterior* concentrates on θ^*.**

***Continuing Example (concluded)*:**

Reconsider the version of the Continuing Example, with Good's *Statistician's Stooge*, involving *iid* sampling of bivariate variable X, a point randomly chosen from circle C.

S is a particular inscribed square.

Let $Y_i = 1$, if $X_i \in$ S, and $Y_i = 0$, if $X_i \notin$ S. Let $\theta = 2/\pi = P(Y{=}1 \mid \theta)$.

Supppose *YOU* assign a smooth but incoherent pseudo-prior to θ,

e.g., use a Beta(1, 1) pseudo-prior.

Then, given the sequence, Y_n ($n = 1, \ldots$), by the *Corollary*,

- ***YOUR* pseudo-posteriors, $P_n(\Theta \mid Y_1, \ldots, Y_n)$ converges (uniformly) to $2/\pi$.**
- **With the Neutral Rate, if $X_c(\omega) = c$ is a constant and $P(X_c)$ is a prevision for X_c, then the degree of incoherence for this one prevision is $|c - P(X_c)|$.**
- **Therefore, almost surely, also the Neutral Rate of incoherence in *YOUR* pseudo-posterior converges to 0.**◊ **Example**

Summary

We reviewed three strategies for responding to Savage's challenge.

***Strategy #1*: Change the closure conditions for a measurable space.**

But the modified closure conditions do not align with agent's actual thinking.

***Strategy #2*: Adapt Good's *Statistician's Stooge* and sidestep *Total Evidence*.**

But it is not evident how to capture with a random variable exactly what are the mathematical/logical facts that the *Statistician* overlooks.

Strategy #3: Concede that uncertainty about math/logic is incoherent.

But apply "robust" algorithms (e.g., Bayes' rule) for learning from computations that reduce the agent's 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.

Gaifman, H. (2004) Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements, *Synthese*, 140: 97–119.

Garber, D. (1983) Old Evidence and Logical Omniscience in Bayesian Confirmation Theory, (pp. 99-131) in John Earman, ed., *Testing Scientific Theories*, Minneapolis: Minnesota Univ. Press.

Good, I.J. (1971) Twenty-seven Principles of Rationality" #679 (pp. 15-19) in *Good Thinking*: *The Foundations of Probability and Its Applications*. Minneapolis: Minnesota Univ. Press.

Hacking, I. (1967) Slightly More Realistic Personal Probability. *Phil. Sci.* 34: 311–325.

Savage, L.J. (1967) Difficulties in the Theory of Personal Probability. *Phil. Sci.* 34: 305–310.

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. (2003) Measures of Incoherence. In *Bayesian Statistics 7*. Bernardo, J.M. et al (eds.). Oxford Univ. Press: Oxford.

Seidenfeld,T. , Schervish, M.J., and Kadane, J.B. (2012) What kind of uncertainty is that? *J.Phil.* 109: 516-533.

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

Some basic results about rates of incoherence

***Application-1*: Incoherence for a set of previsions over a (finite) partition.**

Let {E_1, …., 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 and only 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.

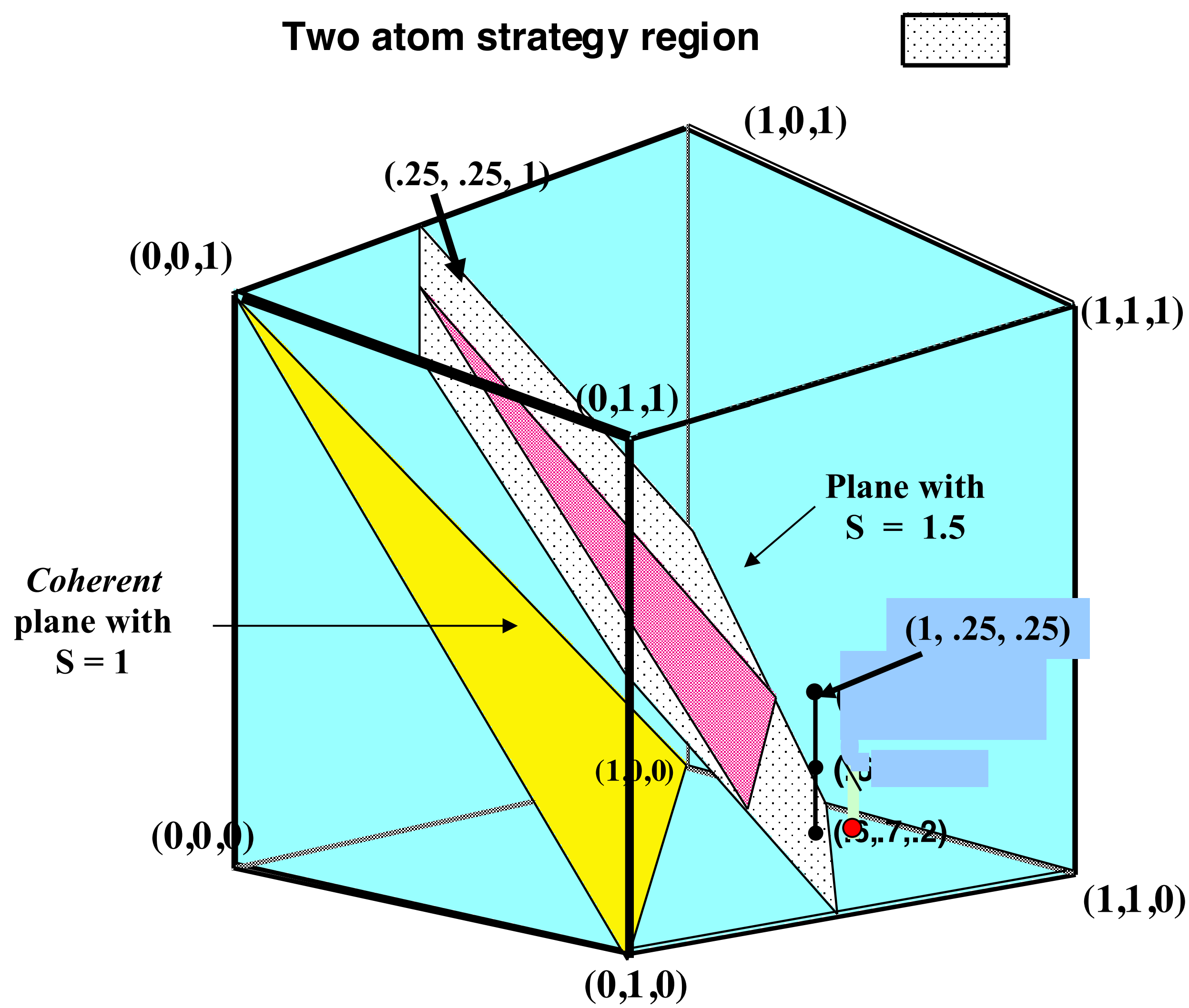
(2) If $\sum_{i=1}^{n} p_*(E_i) = q > 1$, then the *Gambler* maximizes the *rate of gain* by choosing the stakes according to the following rule:

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

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

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

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

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

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 α = .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

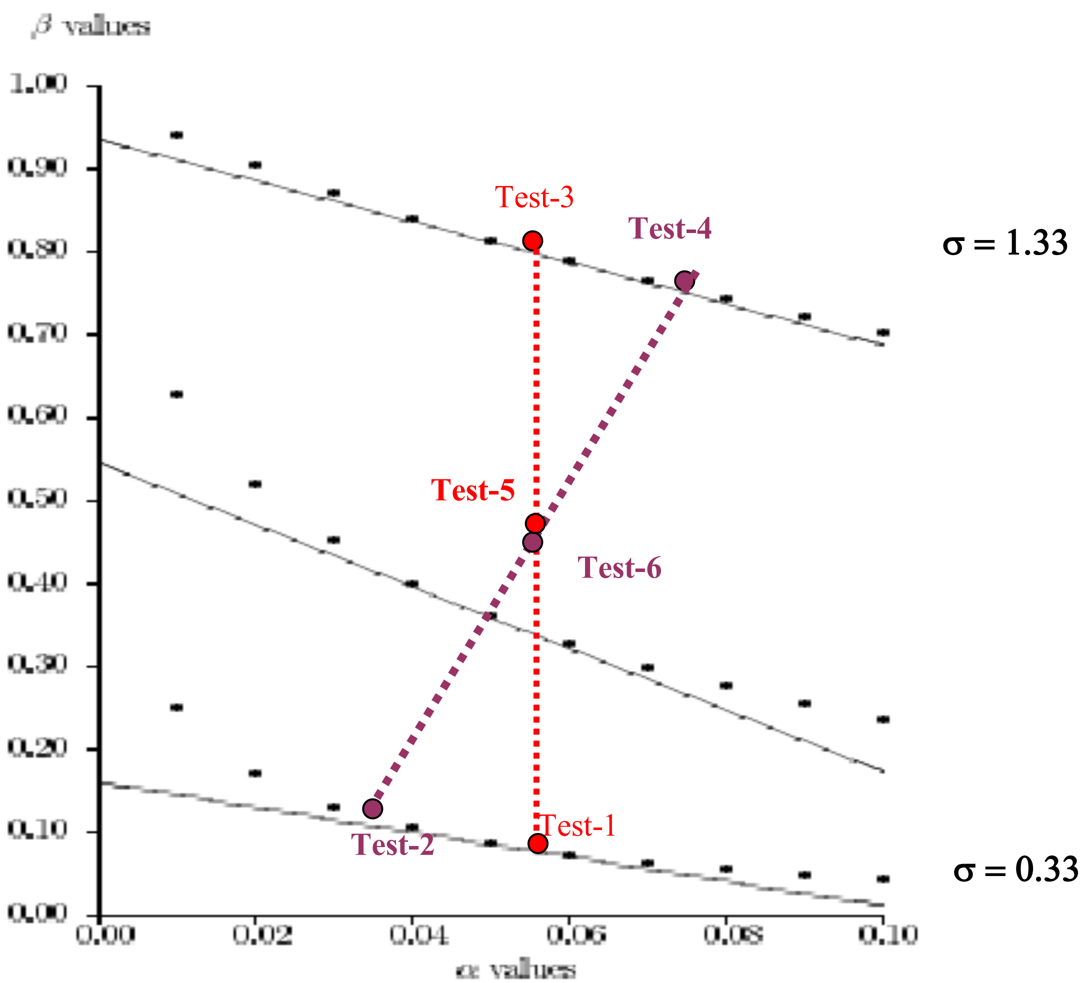
With the convention to choose the best test of level α = .05, the following results:

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

With $\sigma = 0.333$ Test$_3$: ($\alpha = .05$; $\beta = .088$) is chosen over Test$_4$: ($\alpha = .03$; $\beta = .131$).

But the mixed Test$_5$ = .5 Test$_1$ $\oplus$ .5 Test$_3$ has ($\alpha = .05$; $\beta = .451$).

Whereas mixed Test$_6$ = .5 Test$_2$ $\oplus$ .5 Test$_4$ has ($\alpha = .05$; $\beta = .449$), 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-2
Test-1
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:

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

which *is* of the form of a deFinetti *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

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

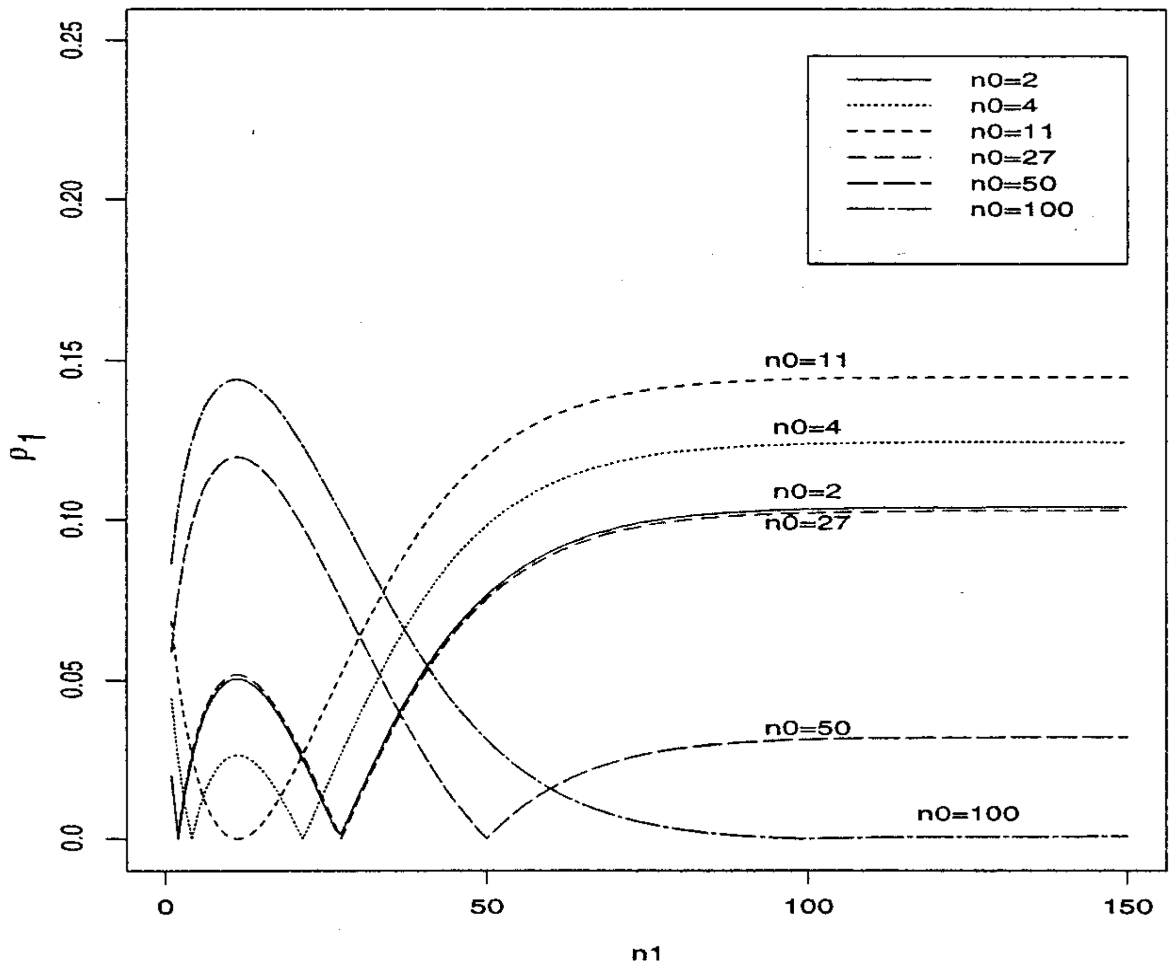


Figure 1. Plot of ρ_1 for level 0.05 testing as a function of n_1 (running from 1 to 150) for various values of n_0 with $c_0 = 19$.

Here is a chart of the resulting *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.