Some Finitely Additive (Statistical) Decision Theory
or
How Bruno de Finetti might have channeled Abraham Wald

T.Seidenfeld

Based on our T.R.: ***What Finite Additivity Can Add to Decision Theory***

Mark Schervish (CMU)

Teddy Seidenfeld (CMU)

Rafael Stern (Universidade Federal De São Carlos),

and Jay Kadane (CMU)

Organization of this presentation.

1. **Three dominance principles and finitely additive expectations – in increasing strength:**

   Uniform (bounded-away) dominance

   Simple dominance

   ***Admissibility* (aka *Strict Dominance*)**

2. **Finitely additive mixed strategies and Wald's (statistical) *Loss* functions.**

   - **An example involving a discontinuous, strictly proper scoring rule.**

3. **Some results – assuming that *Loss* is bounded below:**

   Existence of a *Minimal, Complete Class of Bayes Decisions*

   Existence of a *Minimax Strategy* and a *Worst-case prior*

   Uniform dominance of never-Bayes decisions for bounded loss – *generalized Rationalizability*

- **But, not all *priors* have Bayes-decisions (!)**

Three dominance principles, in increasing order of logical strength

Fix a partition $\pi = \{\omega_1, \ldots., \omega_n, \ldots\}$, which might be infinite.

An *Act* is a function from π to a set of outcomes *O*.

Assume that outcomes may be compared by preference, at least within each ω.

	ω_1	ω_2	ω_3	…	ω_n	…
Act_1	$o_{1,1}$	$o_{1,2}$	$o_{1,3}$	…	$o_{1,n}$	…
Act_2	$o_{2,1}$	$o_{2,2}$	$o_{2,3}$	…	$o_{2,n}$	…

***Uniform dominance*:**

For each ω_i in π, outcome $o_{2,i}$ is strictly preferred to $o_{1,i}$ by at least $\varepsilon > 0$.

***Simple dominance*:**

For each ω_i in π, outcome $o_{2,i}$ is strictly preferred to outcome $o_{1,i}$.

***Admissibility* (Wald, 1950) – *Strict dominance* (Shimony, 1955):**

For each ω_i — $o_{2,i}$ is weakly preferred to $o_{1,i}$

and for some ω_j — $o_{2,j}$ is strictly preferred to $o_{1,j}$.

Then, by *dominance* applied with partition in π,: Act_2 is *strictly preferred* to Act_1.

de Finetti (1974):

A class $\{X\}$ of real-valued variables defined on a privileged partition of *states*, Ω.

Let P be a (f.a) probability on Ω.

Denote by $\mathcal{E}_P(X)$ the (f.a.) expected value of variable X with respect to Ω.

***Preference* between pairs of variables based on finitely additive expectation:**

- **obeys *Uniform Dominance* in Ω**
- **but fails *Simple Dominance* in Ω.**

***Example$_1$* – Let Ω be countably infinite $\Omega = \{\omega_1, \omega_2, \ldots\}$.**

Consider variables $X(\omega_n) = -1/n$, and the constant $Z(\omega_n) = 0$.

Let P be a (strongly) finitely additive probability $P(\{\omega\}) = 0$.

Then $\mathcal{E}_P(X) = 0 = \mathcal{E}_P(Z)$, so indifference between X and Z.

But Z simply dominates X.

	ω_1	ω_2	ω_3	…	ω_n	…
X	-1	-1/2	-1/3	…	$-1/n$	…
Z	0	0	0	…	0	…

Finitely additive mixed strategies: Making lemonade from lemons.

Example$_2$: **Decision making under certainty: $\Omega = \{\omega\}$.**

Consider the half-open interval of constant rewards, $\mathcal{X} = \{X: 0 \leq X < 1\}$.

Each *pure strategy X* is (uniformly) dominated.

Likewise, each countably additive *mixed strategy* P^σ over $\mathcal{X}$ has expectation < 1.

But let be *P* a f.a. *mixed strategy* over $\mathcal{X}$ where, for each $\varepsilon > 0$, $P[X > 1-\varepsilon] = 1$.

- **Then, $\mathcal{E}_P(X) = 1$.**

In f.a. jargon, P *agglutinates* X at the (missing) value 1.

Elementary Statistical Decision Theory in the fashion of A.Wald.

- **An agent has a set $\mathcal{A}$ of available (*pure strategy*) actions, and there is uncertainty over a set Θ of *parameters* or *states of Nature*. Θ forms a privileged partition.**
- **The agent suffers *loss* $L(\theta; a)$ if she chooses a and θ is the *state of Nature*.**
- **Sometimes the agent is allowed to choose action a using a probability measure (a mixed strategy) δ over $\mathcal{A}$, and (when there are no data) we replace *loss* $L(\theta; \cdot)$ by the *risk* $R(\theta; \delta) = \int_{\mathcal{A}} L(\theta; a)\, \delta(da)$.**

Aside: **As usual, the probability measure $\delta_a(A) = I_A(a)$ for every $A \subseteq \mathcal{A}$ is equivalent to the pure strategy a.**

The agent wants to choose δ to minimize *Risk*: respect *dominance* in Θ.

A.Wald (1950): Respect *Admissibility* for *Risk* in Θ.

Example 3a Brier Score **for two complementary events.**

$\Omega = \{B, B^c\}^2$ where B is also the indicator function I_B for some event B.

$\mathcal{A} = [0, 1]^2$. There are no data.

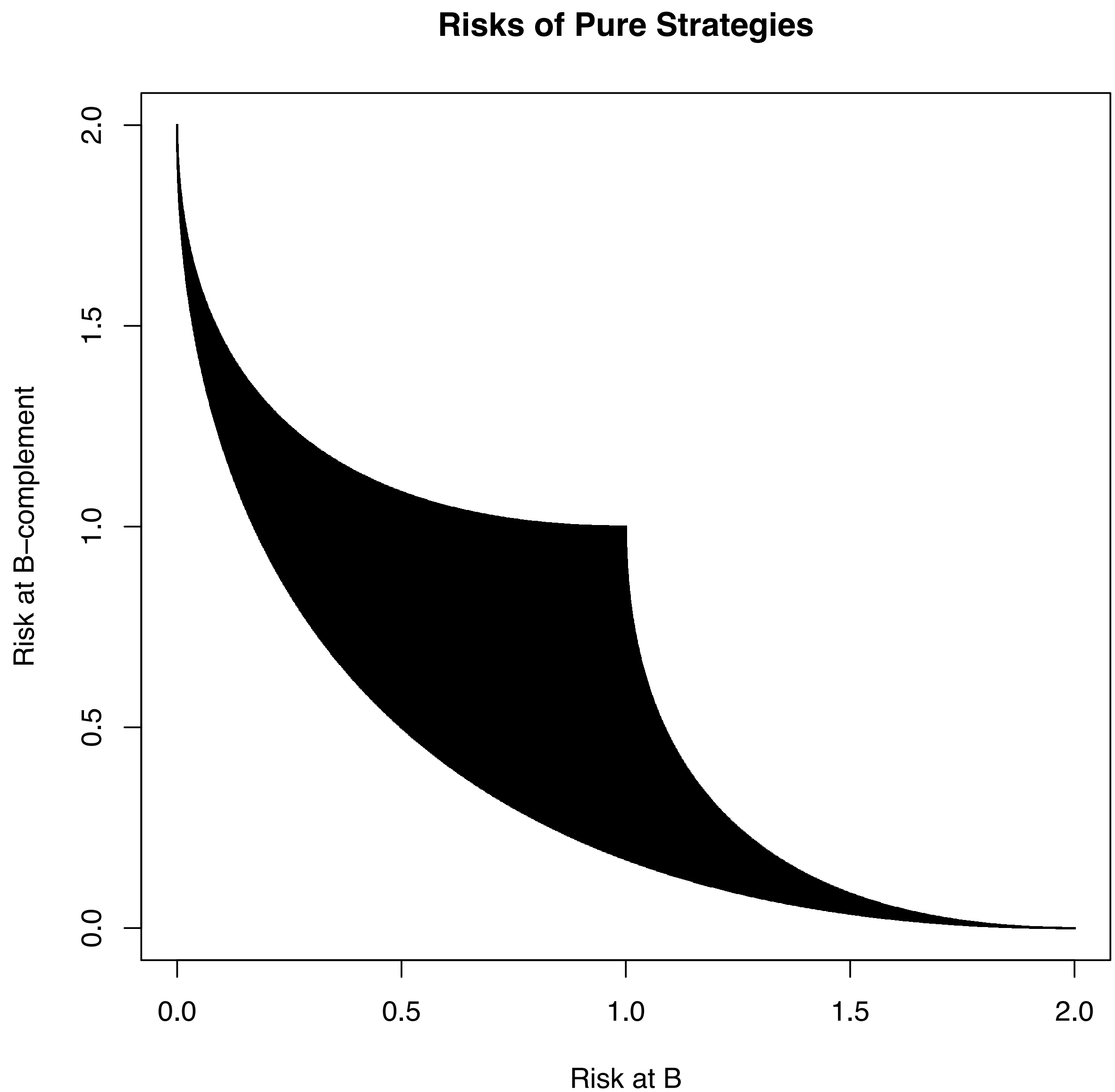
$L(\theta; (a_1, a_2)) = (I_B - a_1)^2 + (I_{Bc} - a_2)^2$

- **The only *admissible* actions are $\{(a_1, a_2): a_1 + a_2 = 1\}$,**

  which correspond to the lower boundary of the Risk set – see next slide.

- **Brier Score is a *strictly proper* scoring rule.**

  The Bayesian agent minimizes expected score <u>uniquely</u> by announcing her degrees of belief for (B, B^c): $a_1 = Prob(B)$ and $a_2 = Prob(B^c)$


Risks of Pure Strategies
Risk at B–complement
Risk at B
0.0
0.5
1.0
1.5
2.0

Example 3b: A discontinuous Brier Score.

$\Omega = \{B, B^c\}^2$ where B is also the indicator function I_B for some event B.

$\mathcal{A} = [0, 1]^2$. Again, there are no data.

$$L(\theta; (a_1, a_2)) = (I_B - a_1)^2 + (I_{B^c} - a_2)^2 + (½) \times \begin{cases} (I_{[0,.5]}(a_1) + I_{(.5,1]}(a_2)) & \text{if } \theta = B \\ (I_{(.5,1]}(a_1) + I_{[0,.5]}(a_2)) & \text{if } \theta = B^c \end{cases}$$

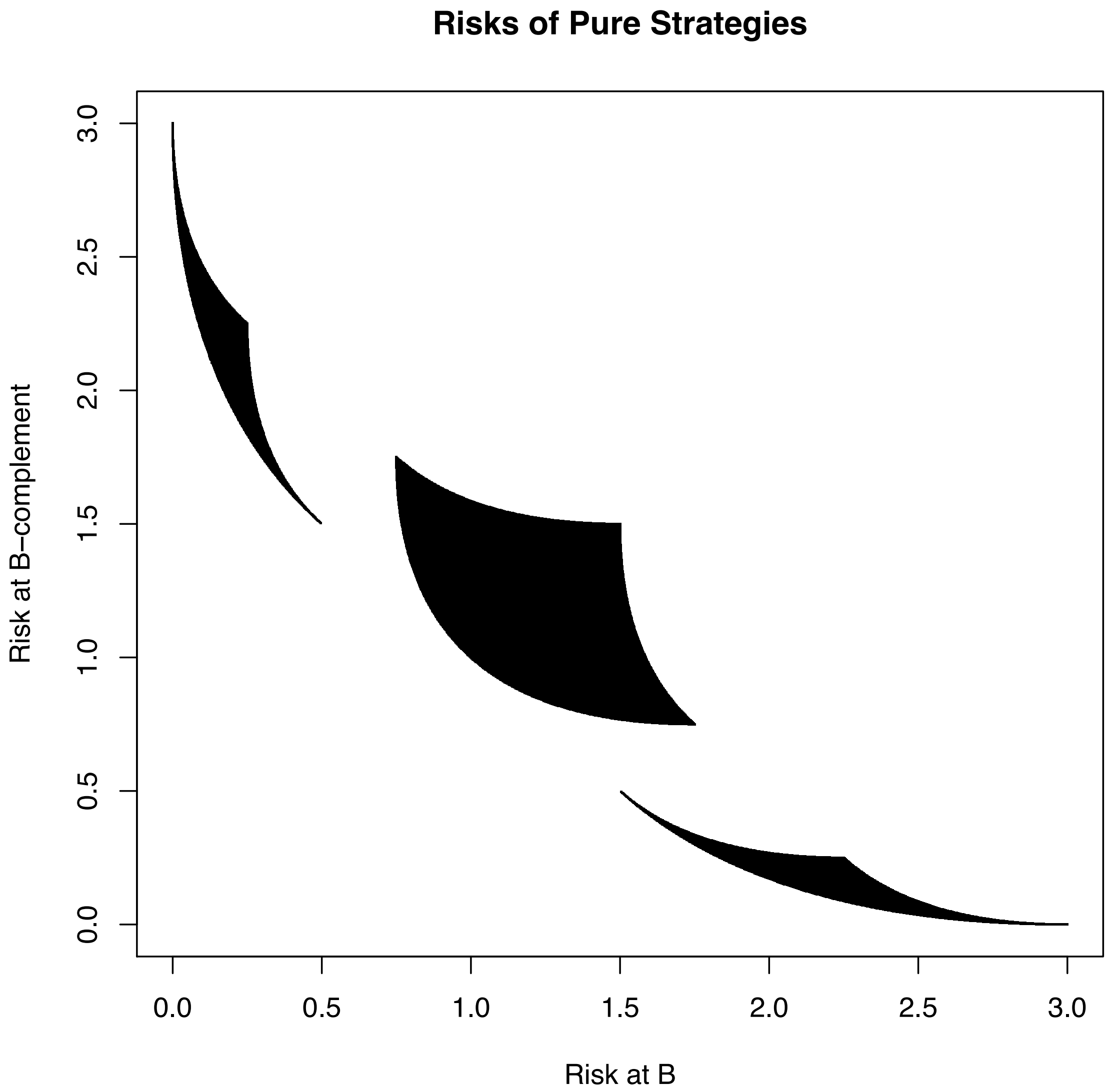
This *Loss* carries an added penalty when the forecast is on the wrong side of ½.

- **The only *admissible* actions are $\{(a_1, a_2): a_1 + a_2 = 1\}$.**
- **This discontinuous Brier Score is a *strictly proper* scoring rule.**

The Bayesian agent (uniquely) minimizes expected score by announcing her degrees of belief for (B, B^c): $a_1 = Prob(B)$ and $a_2 = Prob(B^c)$

but ...

$L(\theta; (a_1, a_2))$ is a point in a two-dimensional set $[0, 3]^2$.



Recall: The *admissible* options are on the lower boundary.

The shaded risk set has the properties that for pairs (*p*, 1-*p*):

Top **From (0, 3) down to but not including (.5, 1.5) are the points on the <u>lower boundary</u>, which correspond to $0 \leq p < .5$.**

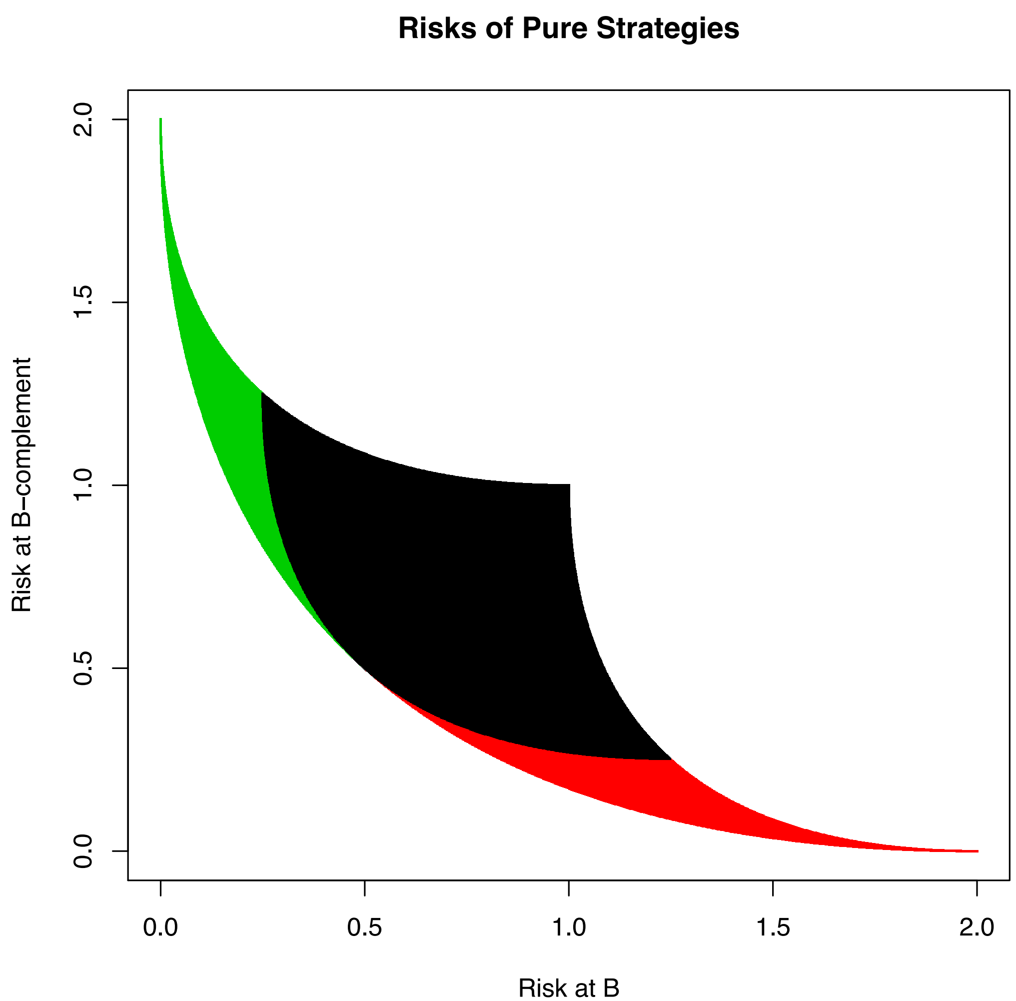
***Middl*e** **In the middle section, only the point (1, 1) is on the <u>lower boundary</u>, corresponding to p = .5.**

Bottom **From (but not including) (1.5, .5) to (3, 0) are the points on the lower boundary, which correspond to $.5 < p \leq 1$.**

So, points in the middle section (other than (1,1)) are *inadmissible* though some are not dominated by (1,1).

But those are dominated too, but only by other *inadmissible* options.

The discontinuous (strictly proper) Brier Score carves up the continuous Brier Score.



3. Some decision-theory results in the fashion of Wald (1950)

***Definitions*:**

Call a subclass $C \subseteq \mathcal{A}$ of available decisions *Complete*

if for each decision $\delta \notin C$ there is $\delta_0 \in C$ where δ_0 *dominates* δ

in the sense of *admissibility*.

Call a subclass $C \subseteq \mathcal{A}$ of available decisions *Minimally Complete*

if C is complete and no proper subset of C is Complete.

- **If there exists a *Minimally Complete* class it consists of the *admissible* decisions.**

- **In *Example* 3_b (discontinuous Brier), there is no *Minimally Complete* class.**

  And using countably additive mixed strategies does not help this way.

BUT, augment the decision space by allowing (merely) f.a. mixed strategies.

Then, these f.a. mixed strategies fill in the missing lower boundary for *Risk*.

For example, consider f.a. mixed strategies δ_1 and δ_2 with the features that

$\forall\varepsilon > 0,\qquad \delta_1\{ a_1: .5 - \varepsilon < a_1 < .5 \} = 1$

and $\qquad \delta_2\{ a_1: .5 < a_1 < .5 + \varepsilon \} = 1,$

and $\qquad$ where $a_2 = 1 - a_1$.

Then $\qquad R(\theta; \delta_1) = (.5, 1.5)$ **and** $R(\theta; \delta_2) = (1.5, .5)$

***Aside*: As $R(\theta; (.5,.5)) = (1,1)$, the 3 risk points $R(\theta; \delta_1)$, $R(\theta; (.5,.5))$ and $R(\theta; \delta_2)$ are colinear.**

Some results – assuming that *Loss* is bounded below:

Strategies are (f.a.) *mixed strategies*.

Strategy δ is *admissible* if there is no strategy δ' such that

$$(\forall\theta) \quad R(\theta;\delta) \geq R(\theta;\delta') \quad \text{and} \quad (\exists\theta) \quad R(\theta;\delta) > R(\theta;\delta').$$

Strategy δ^0 is *Bayes* with respect to a f.a. "prior" probability λ on Θ if

$$\int_\Theta R(\theta;\delta^0)\,\lambda(d\theta) = \mathit{infimum}_\delta \int_\Theta R(\theta;\delta)\,\lambda(d\theta).$$

Strategy δ^* is *minimax* provided that

$$\mathit{supremum}_\theta\, R(\theta;\delta^*) = \mathit{infimum}_\delta\, \mathit{supremum}_\theta\, R(\theta;\delta)$$

Denote the *Bayes-risk* for δ *wrt* "prior" λ by $r(\lambda,\delta) = \int_\Theta R(\theta;\delta)\,\lambda(d\theta)$.

A "prior" λ^* on Θ is *least favorable* provided that

$$\mathit{infimum}_\delta\, r(\lambda^*,\delta) = \mathit{supremum}_\lambda\, \mathit{infimum}_\delta\, r(\lambda;\delta).$$

- **Assume that the Loss function is bounded below, that decision rules are mixed strategies, and that “prior” probabilities are finitely additive.**

***Theorem*$_1$: The decision rules whose risks form the lower boundary of the risk function constitute a minimal complete class of admissible rules.**

Each admissible rule is a Bayes rule.

***Theorem*$_2$: There exists a minimax decision rule and a corresponding least-favorable prior.**

Each minimax rule is Bayes *wrt* each least-favorable prior.

***Theorem*$_3$ (*Rationalizability for infinite games*):**

Assume that the loss function is bounded above and below.

Suppose that δ_0 is a decision rule that is *not* Bayes for any prior.

Then there is decision rule δ_1 and $\varepsilon > 0$ such that

$$(\forall\theta) \quad R(\theta; \delta_0) > R(\theta; \delta_1) + \varepsilon.$$

That is, then there is a rival δ_1 that *uniformly dominates* δ_0 in Risk.

3. But not all *priors* have Bayes-decisions (!)

One of the challenges associated with (merely) f.a. expectations is that the order of integration matters – *Fubini's Theorem* has restricted validity.

So, even though the *Risk* function has a closed (lower) boundary composed of Bayes decisions, it does *not* follow that for an arbitrary "prior" λ on Θ there is a Bayes decision δ^0 *wrt* λ, where

$$\int_\Theta R(\theta; \delta^0)\, \lambda(d\theta) \;=\; \mathit{infimum}_\delta \int_\Theta R(\theta; \delta)\, \lambda(d\theta).$$

Example$_4$:

Parameter space, $\Theta = (0, 1)$

$\mathcal{A}$ is the set of all non-empty open subintervals of (0, 1).

That is, $\mathcal{A} = \{ (x, y): 0 \leq x < y \leq 1\}$.

Denote by *Len*$[(x, y)] = y - x$, the length of interval *a*.

The *Loss* function reflects a goal of *anti-estimation* for θ:

$$L(\theta; a) = I_a(\theta)/Len[a] + (1-Len[a])/10$$

Consider a (strongly) f.a. "prior" $\lambda^\#$ on Θ where, for each $y > 0$,

$$\lambda^\#\{\theta: 0 < \theta < y] = 1.$$

In f.a. jargon, $\lambda^\#$ *agglutinates* its mass at the (missing) $\theta = 0$.

The *Bayes risk* with respect to $\lambda^{\#}$ satisfies, for each $n = 1, 2, \ldots$,

$$r(\lambda^{\#}; (1/n, 1)\,) = \int_{\Theta} R(\theta; (1/n, 1)\, \lambda^{\#}(d\theta)$$

$$= \int_{\Theta} [\mathrm{I}_a(\theta)/Len[a] + (1\text{-}Len[a])/10]\, \lambda^{\#}(d\theta)$$

$$= \int_{\Theta} [\mathrm{I}_{(1/\mathrm{n},\, 1)}(\theta) \,/\, (\mathrm{n}\text{-}1)/\mathrm{n}] \;+\; (1/n)/10]\, \lambda^{\#}(d\theta)$$

$$= 0 + 1/10n = 1/10n.$$

Hence, $$\mathit{infimum}_{\delta}\; r(\lambda^{\#}; \delta) = 0.$$

- **But, there is no decision rule $\delta^{\#}$ with *Bayes risk* $r(\lambda^{\#}; \delta^{\#}) = 0$.**

To see this, note that, by indirect reasoning:

If $r(\lambda^{\#}; \delta) = 0$, then – from the 2nd term in the *Loss* function – $\mathcal{E}_{\lambda\#}[Len(\delta)] = 1$.

But then, because of the order of integration, with the "prior" $\lambda^{\#}(d\theta)$ integration on the outside – from the 1st term in the *Risk* function –

$$\int_{\Theta} (\mathrm{I}_{\delta}(\theta)/Len[\delta])\, \lambda^{\#}(d\theta) \;>\; 0.$$

SUMMARY

We have reviewed how the use of some (merely) f.a. mixed strategies convert

the failure of simple dominance – a *lemon*,

into the closure of the lower-boundary for a (bounded-below) statistical *Loss function*, understood in the fashion of A. Wald – *lemonade*!

It follows that there exist:

- **a *Minimal Complete Class* of *Admissible* decisions, each of which is Bayes with respect to some (f.a.) "prior";**
- **a Minimax rule and Worst-case "prior" for which the Minimax is Bayes;**

and

- **a generalized *Rationalizability result* where each never-Bayes decision is uniformly dominated by some alternative (mixed strategy) decision.**

BUT – not every "prior" has its Bayes rule.