Modeling Uncertainty with Sets of Probabilities

Ruobin Gong

Rutgers University

Foundations of Probability Seminar

Oct 29, 2018

There's more to uncertainty than probabilities

- Data
- Prior
- Model structure (other than the prior)

There's more to uncertainty than probabilities

- Data: missing & coarse data (Heitjan & Rubin, 1990);
- Prior
- Model structure (other than the prior)

There's more to uncertainty than probabilities

- Data: missing & coarse data (Heitjan & Rubin, 1990);
- Prior: priors neighborhoods in robust Bayes (Berger, 1990);
- Model structure (other than the prior)

There's more to uncertainty than probabilities

- Data: missing & coarse data (Heitjan & Rubin, 1990);
- Prior: priors neighborhoods in robust Bayes (Berger, 1990);
- Model structure (other than the prior)
  - Robust statistics: N-P lemma for hypothesis testing between coherent lower previsions (Huber & Strassen, 1973);
  - Econometrics: partially identified and incomplete models (Kline & Tamer, 2016; Epstein et al., 2016)
  - Under-determined structural equation inference:
    - Fiducial inference (Hannig, 2009);
    - Dempster-Shafer theory (Dempster, 2008);
  - Privacy and data confidentiality.

There's more to uncertainty than probabilities

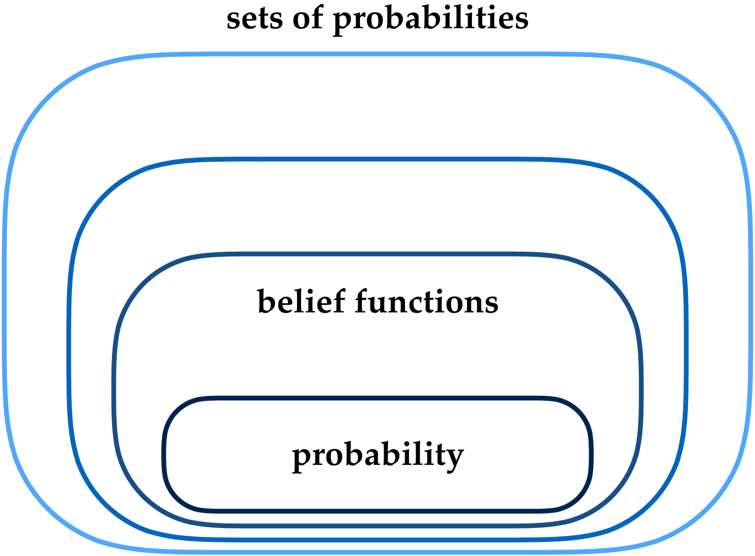
- Data: missing & coarse data (Heitjan & Rubin, 1990);
- Prior: priors neighborhoods in robust Bayes (Berger, 1990);
- Model structure (other than the prior)
  - Robust statistics: N-P lemma for hypothesis testing between coherent lower previsions (Huber & Strassen, 1973);
  - Econometrics: partially identified and incomplete models (Kline & Tamer, 2016; Epstein et al., 2016)
  - Under-determined structural equation inference:
    - Fiducial inference (Hannig, 2009);
    - Dempster-Shafer theory (Dempster, 2008);
  - Privacy and data confidentiality.

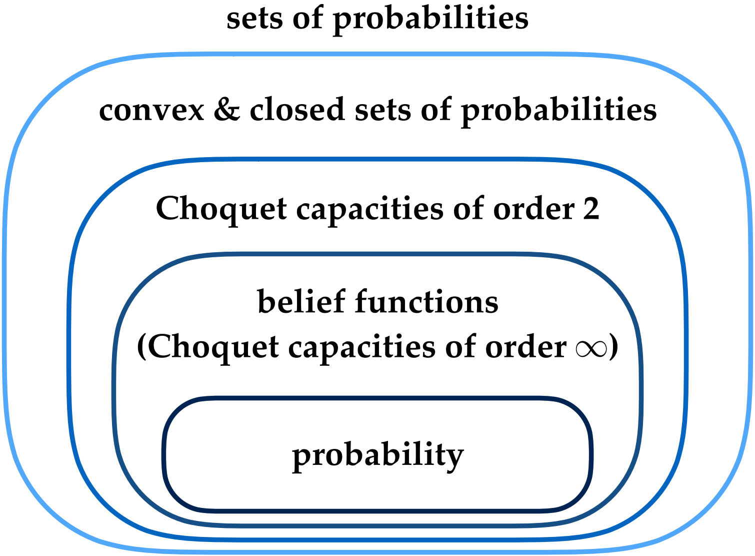
Sets of Probabilities

(Huber & Ronchetti, 2009). Let $\mathcal{M}$ be the set of all probability measures on some measurable space $(\Omega, \mathscr{F}(\Omega))$, and $\mathcal{P} \subset \mathcal{M}$ be a nonempty subset. Define the **lower** and **upper probability** induced by $\mathcal{P}$ as

$$\underline{P}(A) = \inf_{\mathcal{P}} P(A), \quad \overline{P}(A) = \sup_{\mathcal{P}} P(A).$$

$\underline{P}$ and $\overline{P}$ are *conjugate* to each other: $\overline{P}(A) = 1 - \underline{P}(A^c)$.


sets of probabilities
belief functions
probability


sets of probabilities
convex & closed sets of probabilities
Choquet capacities of order 2
belief functions
(Choquet capacities of order ∞)
probability

Agenda

Prior-free multinomial inference with belief function

Sets of probabilities conditioning: rules and paradoxes

Example: survey nonresponse

Q1. Did you injure yourself on the snow last season (**Y**/**N**)?

Q2. Do you ski or snowboard **(K/S)**?

Example: survey nonresponse

Q1. Did you injure yourself on the snow last season (**Y**/**N**)?

Q2. Do you ski or snowboard **(K/S)**?

Intended sample space:

$$\Omega = \{Y, N\} \times \{K, S\}$$

Actual sample space:

$$\mathscr{P}(\Omega) = \{A : A \subseteq \Omega\}$$

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

What can we say about **injury rate**, $P(Y)$?

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

What can we say about **injury rate**, $P(Y)$?

$$\begin{aligned} P(Y) &= P(\{(Y, K), (Y, S)\}) \\ &\geq m(\{(Y, K)\}) + m(\{(Y, S)\}) + m(\{(Y, K), (Y, S)\}) = 0.3 \end{aligned}$$

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

What can we say about **injury rate**, $P(Y)$?

$$\underline{P}(Y) \stackrel{\text{def}}{=} m(\{(Y,K)\}) + m(\{(Y,S)\}) + m(\{(Y,K),(Y,S)\}) = 0.3$$

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

What can we say about **injury rate**, $P(Y)$?

$$\underline{P}(Y) \stackrel{\text{def}}{=} m(\{(Y,K)\}) + m(\{(Y,S)\}) + m(\{(Y,K),(Y,S)\}) = 0.3$$

$$P(Y) = 1 - P(N)$$

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

What can we say about **injury rate**, $P(Y)$?

$$\underline{P}(Y) \stackrel{\text{def}}{=} m(\{(Y,K)\}) + m(\{(Y,S)\}) + m(\{(Y,K),(Y,S)\}) = 0.3$$

$$\overline{P}(Y) = 1 - \underline{P}(N) = 0.64$$

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

What can we say about **injury rate**, $P(Y)$?

$$\underline{P}(Y) \stackrel{\text{def}}{=} m(\{(Y,K)\}) + m(\{(Y,S)\}) + m(\{(Y,K),(Y,S)\}) = 0.3$$
$$\overline{P}(Y) = 1 - \underline{P}(N) = 0.64$$

The set functions

- $\underline{P} : \mathscr{P}(\Omega) \to [0, 1]$ is a **belief function**, a special kind of **lower probability function**;
- $\overline{P}$ is a **plausibility function**, a special kind of **uppper probability function** conjugate to $\underline{P}$: $\overline{P}(A) = 1 - \underline{P}(A^c)$.

Prior-free multinomial inference

Let $\mathbf{X} = (X_1, \ldots, X_K)$ be a vector of counts in each of K mutually exclusive categories. $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$ is a vector of intensity rates.

Multinomial as "partially identified" Poisson.

Suppose $X_i \sim Pois\,(\lambda_i \cdot T)$, for $i = 1, \ldots, K$:

- $\boldsymbol{\lambda}$ is aliased with T and non-identifiable. However:
- $\mathbf{q} = \boldsymbol{\lambda}/\|\boldsymbol{\lambda}\|$ is free of T and identifiable.

Prior-free multinomial inference

Model state space: $(\mathbf{X}, \boldsymbol{\lambda}) \in \Omega = \mathbb{N}^K \times (\mathbb{R}^+)^K$.

Observed data: $\mathbf{X} = \mathbf{x}$.

Prior-free multinomial inference

Model state space: $(\mathbf{X}, \boldsymbol{\lambda}) \in \Omega = \mathbb{N}^K \times (\mathbb{R}^+)^K$.
Observed data: $\mathbf{X} = \mathbf{x}$.

By Gamma-Poisson count-time duality, posit the sampling model:

$$X_i = \mathrm{argmax}_k \left(\sum_{j=0}^{k} E_{ij} \leq \lambda_i \right), \; i = 1, ..., K$$

where auxiliary variables $E_{ij} \overset{i.i.d.}{\sim} Exp(1)$ are ancillary to $\boldsymbol{\lambda}$; $E_{i0} = 0$.

No prior assumptions on $\boldsymbol{\lambda}$.

Prior-free multinomial inference

Denote $G_i = \sum_{j=0}^{X_i} E_{ij}$ and $\Delta E_i = E_{i(X_i+1)}$. The model implies a collection of subsets of the state space Ω:

$$\{(\mathbf{X}, \boldsymbol{\lambda}) \in \Omega : \mathbf{X} = \mathbf{x}, \boldsymbol{\lambda} \in \otimes_{i=1}^{K}[G_i, G_i + \Delta E_i]\},$$

which is understood as a *random subset generated* by $(\mathbf{G}, \Delta\mathbf{E})$.

Prior-free multinomial inference

Denote $G_i = \sum_{j=0}^{X_i} E_{ij}$ and $\Delta E_i = E_{i(X_i+1)}$. The model implies a collection of subsets of the state space Ω:

$$\{(\mathbf{X}, \boldsymbol{\lambda}) \in \Omega : \mathbf{X} = \mathbf{x}, \boldsymbol{\lambda} \in \otimes_{i=1}^{K} [G_i, G_i + \Delta E_i]\},$$

which is understood as a *random subset generated* by $(\mathbf{G}, \Delta\mathbf{E})$. Down-projected to the $\mathbf{q}$ margin,

$$\mathcal{Q}_{\mathbf{x}} \stackrel{\text{def}}{=\!=} \{\mathbf{q} \in \mathbb{S}_K : \mathbf{X} = \mathbf{x}, \boldsymbol{\lambda} \in \otimes_{i=1}^{K} [G_i, G_i + \Delta E_i], \mathbf{q} = \boldsymbol{\lambda} / \|\boldsymbol{\lambda}\|\}$$

$\mathcal{Q}_{\mathbf{x}}$ is a random convex $(K-1)$-polytope in $\mathbb{S}_K$.

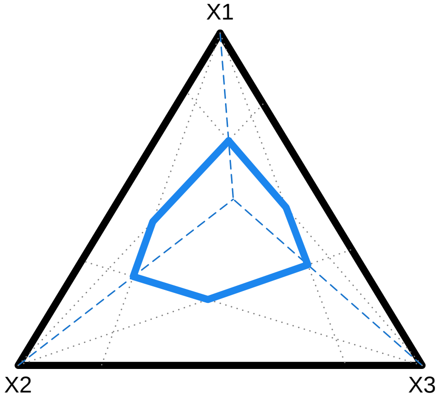


Figure: One realization of $\mathcal{Q}_{\mathbf{x}}$: $\mathbf{x} = (2, 1, 1)$.

$\mathcal{Q}_{\mathbf{x}}$ is a random convex $(K-1)$-polytope in $\mathbb{S}_K$ with two characterizations:

- by vertices (good for Monte Carlo);
- by a system of side inequalities (good for estimation & hypothesis testing).

$\mathcal{Q}_{\mathbf{x}}$ characterized by vertices

Vertices of $\mathcal{Q}_{\mathbf{x}}$ are (dependent) Dirichlet random variables.

Tier 0:	$\frac{1}{\sum G_i}(G_1, ..., G_K)$
Tier 1:	$\frac{1}{\sum G_i + \Delta E_1}(G_1 + \Delta E_1, G_2, ..., G_K)$ $\vdots$
Tier 2:	$\frac{1}{\sum G_i + \Delta E_1 + \Delta E_2}(G_1 + \Delta E_1, G_2 + \Delta E_2, G_3, ..., G_K)$ $\vdots$ $\frac{1}{\sum G_i + \Delta E_{K-1} + \Delta E_K}(G_1, ...G_{K-2}, G_{K-1} + \Delta E_{K-1}, G_K + \Delta E_K)$
$\vdots$	$\vdots$
Tier K-1:	$\frac{1}{\sum G_i + \sum_{i=1}^{K-1} \Delta E_i}(G_1, G_2 + \Delta E_2, ..., G_K + \Delta E_K)$
Tier K:	$\frac{1}{\sum G_i + \sum \Delta E_i}(G_1 + \Delta E_1, ..., G_K + \Delta E_K)$

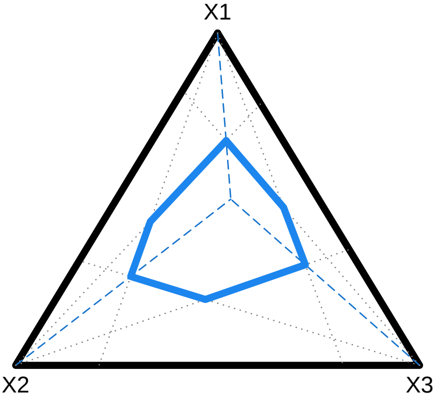


Figure: One realization of $\mathcal{Q}_{\mathbf{x}}$: $\mathbf{x} = (2, 1, 1)$.

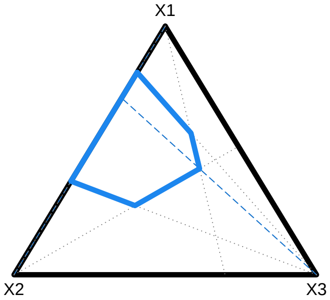
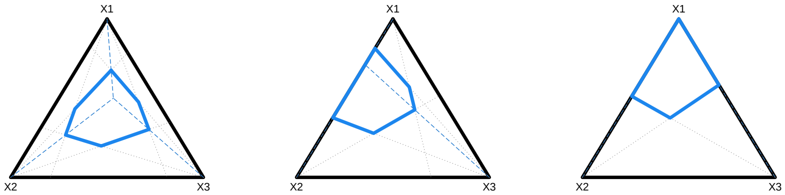


Figure: Realization of $\mathcal{Q}_{\mathbf{x}}$: left, $\mathbf{x} = (3, 1, 0)$; right, $\mathbf{x} = (4, 0, 0)$.

$\mathcal{Q}_{\mathbf{x}}$ characterized by a system of inequalities

Each realization of $\mathcal{Q}_{\mathbf{x}}$ constitutes a *density ratio class* (Wasserman, 1992) of probabilities:

$$\mathcal{Q}_{\mathbf{x}} = \left\{ \mathbf{q} \in \mathbb{S}_K : \frac{q_i}{q_j} \geq \frac{G_i}{G_j + \Delta E_j}, \forall (i,j) \in \{1, ..., K\}^2, i \neq j \right\}.$$



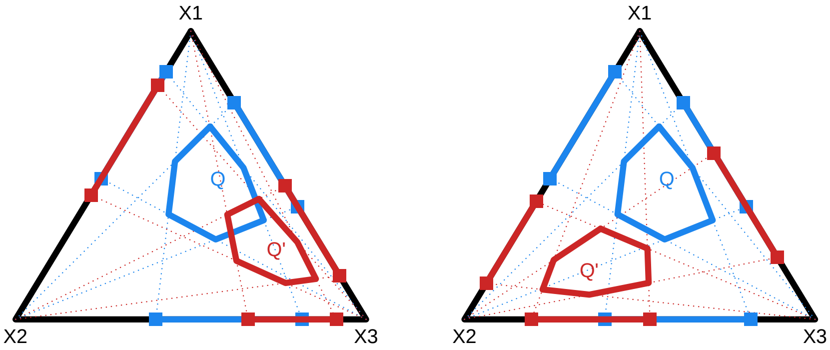
Posterior inference for $\mathbf{q}$

Lemma 3.2.2-2. (polytope intersection)

Let $\mathcal{Q} \subseteq \mathbb{S}_K$. $\mathcal{Q} \cap \mathcal{Q}' \neq \emptyset$ iff $\forall i \neq j$, $i, j \in \{1, ..., K\}$,

$$\mathrm{Proj}_{(i,j)}\left(\mathcal{Q}\right) \cap \mathrm{Proj}_{(i,j)}\left(\mathcal{Q}'\right) \neq \emptyset,$$

where $\mathrm{Proj}_{(i,j)}\left(\mathcal{Q}\right)$ is the projection of $\mathcal{Q}$ onto $\mathbb{S}_2$ spanned by the $(i,j)^{th}$ dimensions.


X1
X2
X3
Q
Q'
X1
X2
X3
Q
Q'

Posterior inference for $\mathbf{q}$: one- and two-sample

Suppose $\mathbf{X}, \mathbf{X}'$ are two independently observed vectors of counts, and $\mathbf{q}, \mathbf{q}' \in \mathbb{S}_K$ their respective proportion vectors.

Theorem 3.2.3. (one-sample)

For the hypothesis $H : \mathbf{q} = \mathbf{q}^*$, the posterior upper probability is

$$\overline{P}(H) = Pr\left(\mathbf{q}^* \in \mathcal{Q}_{\mathbf{x}}\right) = Pr\left(\bigcap_{(i,j)\in\{1,\ldots,K\}^2 : i\neq j} \left(\frac{q_i^*}{q_j^*} \geq \frac{G_i}{G_j + \Delta E_j}\right)\right).$$

Theorem 3.2.4. (two-sample)

For hypothesis $H : \mathbf{q} = \mathbf{q}'$, the posterior upper probability is

$$\overline{P}(H) = Pr\left(\mathcal{Q}_{\mathbf{x}} \cap \mathcal{Q}'_{\mathbf{x}'} \neq \emptyset\right) = Pr\left(\bigcap_{(i,j)\in\{1,\ldots,K\}^2 : i\neq j} \left(\frac{G_i + \Delta E_i}{G_j} \geq \frac{G'_i}{G'_j + \Delta E'_j}\right)\right).$$

Posterior inference for $\mathbf{q}$: test statistic

Let $t : \mathbb{S}_K^2 \to \mathcal{T}$ be a function of multinomial proportion vectors. Define a **one-sample** posterior test statistic

$$\mathcal{R}_{t|\mathbf{x}} = Cl\left(\left\{\mathbf{t}\left(\mathbf{q}, \mathbf{q}_0\right) \in \mathbb{R}^d : \mathbf{q} \in \mathcal{Q}_{\mathbf{x}}\right\}\right),$$

and similarly, a **two-sample** posterior test statistic

$$\mathcal{R}_{t|\mathbf{x},\mathbf{x}'} = Cl\left(\left\{\mathbf{t}\left(\mathbf{q}, \mathbf{q}'\right) \in \mathbb{R}^d : \mathbf{q} \in \mathcal{Q}_{\mathbf{x}}, \mathbf{q}' \in \mathcal{Q}'_{\mathbf{x}'}\right\}\right).$$

$\mathcal{R}_{t|\mathbf{x}}$ and $\mathcal{R}_{t|\mathbf{x},\mathbf{x}'}$ are again random closed subsets of $\mathcal{T}$, whose distributions dictated by that of $\mathcal{Q}_{\mathbf{x}}$ and of $\left(\mathcal{Q}_{\mathbf{x}}, \mathcal{Q}'_{\mathbf{x}'}\right)$ respectively.

Posterior Hellinger statistic

Hellinger distance between two discrete probability distributions:

$$h\left(\mathbf{q},\mathbf{q}'\right)=\frac{1}{\sqrt{2}}\cdot\sum_{i=1}^{K}\left(q_i^{1/2}-q_i^{1/2}\right)^2.$$

The two-sample posterior Hellinger distance $\mathcal{R}_{h|\mathbf{x},\mathbf{x}'}$ is the closed interval

$$[h_{\min},h_{\max}]\stackrel{\text{def}}{=\!=}\left[\min_{\mathbf{q}\in\mathcal{Q}_{\mathbf{x}},\mathbf{q}'\in Q'_{\mathbf{x}'}}h\left(\mathbf{q},\mathbf{q}'\right),\max_{\mathbf{q}\in\mathcal{Q}_{\mathbf{x}},\mathbf{q}'\in Q'_{\mathbf{x}'}}h\left(\mathbf{q},\mathbf{q}'\right)\right].$$

Posterior Hellinger statistic: under H_0

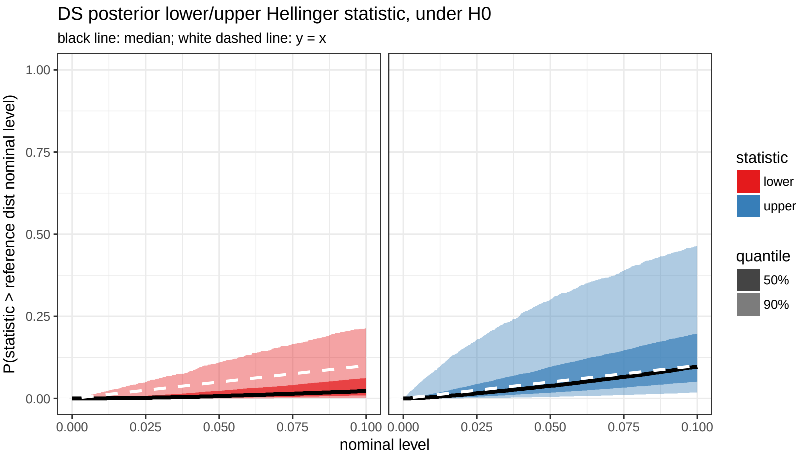


Figure: $K = 5$, $N = 100$, $N' = 108$, $\mathbf{q} = (0.12, 0.20, 0.24, 0.32, 0.12)$. Under H_0, $\mathbf{q}' = \mathbf{q}$. Under H_1, $\mathbf{q}' \propto (0.06, 0.06, 0.12, 0.256, 0.048)$.

Posterior Hellinger statistic: under H_1

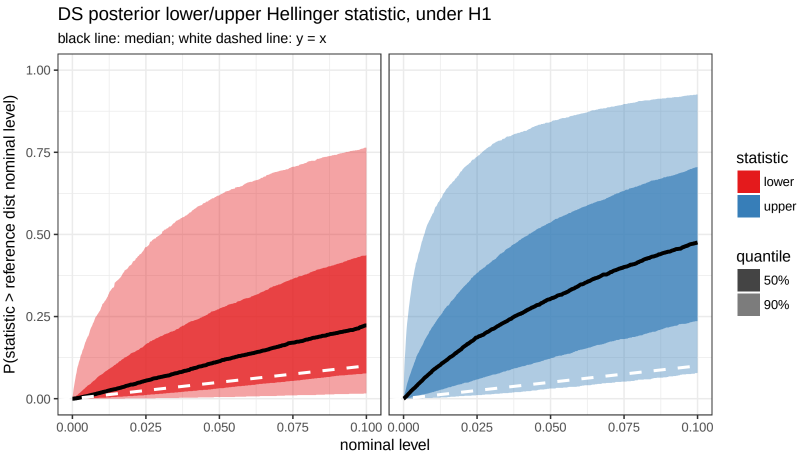
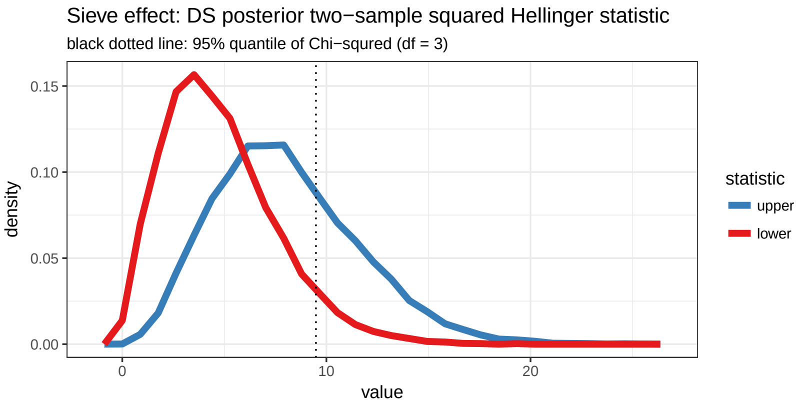


Figure: $K = 5$, $N = 100$, $N' = 108$, $\mathbf{q} = (0.12, 0.20, 0.24, 0.32, 0.12)$. Under H_0, $\mathbf{q}' = \mathbf{q}$. Under H_1, $\mathbf{q}' \propto (0.06, 0.06, 0.12, 0.256, 0.048)$.

Sieve analysis: vaccine equi-efficacy



A **precise** probability model updates according to, e.g., Bayes rule.

$$P(A) \quad \Longrightarrow \quad P(A \mid B)$$

What if P is **imprecise**?

Updating sets of probabilities

Rules

1. Generalized $\mathfrak{B}$ayes rule
2. $\mathfrak{D}$empster's rule
3. $\mathfrak{G}$eometric rule

Paradoxes

- Dilation
- Contraction
- Sure Loss

Survey nonresponse: what's $P(\text{injury} \mid \text{ski})$?

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

One option can be:

$$\underline{P}_{\mathfrak{B}}(\text{injury} \mid \text{ski}) \stackrel{\text{def}}{=\!=} \inf_{P \in \Pi} \frac{P(\text{injury}, \text{ski})}{P(\text{ski})}$$

$$\overline{P}_{\mathfrak{B}}(\text{injury} \mid \text{ski}) \stackrel{\text{def}}{=\!=} \sup_{P \in \Pi} \frac{P(\text{injury}, \text{ski})}{P(\text{ski})}$$

Survey nonresponse: what's P(injury | ski)?

Q_1	Y	Y	N	N	$\{Y, N\}$	$\{Y, N\}$	Y	N	$\{Y, N\}$
Q_2	K	S	K	S	K	S	$\{K, S\}$	$\{K, S\}$	$\{K, S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

Generalized $\mathfrak{B}$ayes rule:

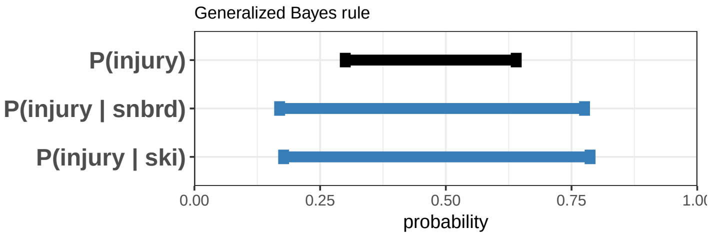
$$\underline{P}_{\mathfrak{B}}(\text{injury} \mid \text{ski}) \stackrel{\text{def}}{=\!=} \inf_{P\in\Pi} \frac{P(\text{injury}, \text{ski})}{P(\text{ski})}$$

$$\overline{P}_{\mathfrak{B}}(\text{injury} \mid \text{ski}) \stackrel{\text{def}}{=\!=} \sup_{P\in\Pi} \frac{P(\text{injury}, \text{ski})}{P(\text{ski})}$$

Survey nonresponse: what's $P\,(\text{injury} \mid \text{ski})$?

Q_1	Y	Y	N	N	$\{Y,N\}$	$\{Y,N\}$	Y	N	$\{Y,N\}$
Q_2	K	S	K	S	K	S	$\{K,S\}$	$\{K,S\}$	$\{K,S\}$
$m(\mathbf{R})$	0.11	0.10	0.13	0.13	0.08	0.06	0.09	0.10	0.20

Generalized $\mathfrak{B}$ayes rule:



"Posterior" interval grows wider regardless of sport type!

Definition: Dilation (cf. Seidenfeld & Wasserman (1993))

Let $A \in \mathscr{B}(\Omega)$, $\mathcal{B}$ a Borel-measurable partition of Ω, Π be a closed and convex set of probability measures on Ω, $\underline{P}$ its lower probability function, and $\underline{P}_\bullet$ the conditional lower probability function supplied by the updating rule "$\bullet$".

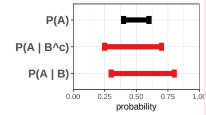
Say that $\mathcal{B}$ **strictly dilates** A under the $\bullet$-rule if

$$\sup_{B \in \mathcal{B}} \underline{P}_\bullet(A \mid B) < \underline{P}(A) \leq \overline{P}(A) < \inf_{B \in \mathcal{B}} \overline{P}_\bullet(A \mid B).$$

Gen. 𝔅ayes rule

$\overline{P}_{\mathfrak{B}}(A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)}$

Dilation



Gen. $\mathfrak{B}$ayes rule

$$\overline{P}_{\mathfrak{B}}(A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)}$$

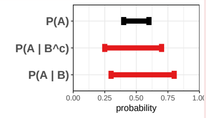
$\mathfrak{D}$empster's rule

$$\overline{P}_{\mathfrak{D}}(A \mid B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)}$$

$\mathfrak{G}$eometric rule

$$\underline{P}_{\mathfrak{G}}(A \mid B) = \frac{\inf_{P \in \Pi} P(A \cap B)}{\inf_{P \in \Pi} P(B)}$$

Dilation



𝔊en. 𝔅ayes rule

$$\overline{P}_{\mathfrak{B}}(A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)}$$

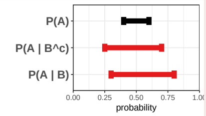
𝔇empster's rule

$$\overline{P}_{\mathfrak{D}}(A \mid B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)}$$

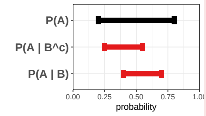
𝔊eometric rule

$$\underline{P}_{\mathfrak{G}}(A \mid B) = \frac{\inf_{P \in \Pi} P(A \cap B)}{\inf_{P \in \Pi} P(B)}$$

Dilation



Contraction



Gen. 𝔅ayes rule

$$\overline{P}_{\mathfrak{B}}(A \mid B) = \sup_{P\in\Pi} \frac{P(A\cap B)}{P(B)}$$

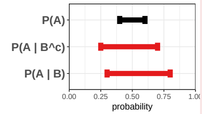
𝔇empster's rule

$$\overline{P}_{\mathfrak{D}}(A \mid B) = \frac{\sup_{P\in\Pi} P(A\cap B)}{\sup_{P\in\Pi} P(B)}$$

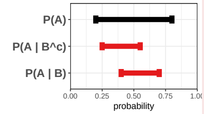
𝔊eometric rule

$$\underline{P}_{\mathfrak{G}}(A \mid B) = \frac{\inf_{P\in\Pi} P(A\cap B)}{\inf_{P\in\Pi} P(B)}$$

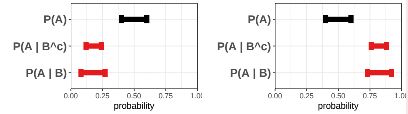
Dilation



Contraction



Sure Loss



Gen. $\mathfrak{B}$ayes rule

$$\overline{P}_{\mathfrak{B}}(A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)}$$

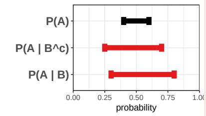
$\mathfrak{D}$empster's rule

$$\overline{P}_{\mathfrak{D}}(A \mid B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)}$$

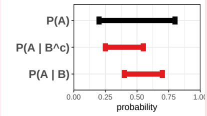
$\mathfrak{G}$eometric rule

$$\underline{P}_{\mathfrak{G}}(A \mid B) = \frac{\inf_{P \in \Pi} P(A \cap B)}{\inf_{P \in \Pi} P(B)}$$

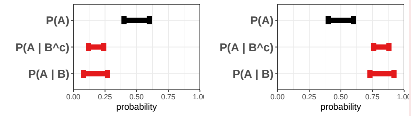
Dilation



Contraction



Sure Loss



Thm 1. $\mathfrak{B}$-rule cannot contract nor induce sure loss.

Gen. $\mathfrak{B}$ayes rule

$$\overline{P}_{\mathfrak{B}}\,(A \mid B) = \sup_{P\in\Pi} \frac{P(A\cap B)}{P(B)}$$

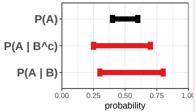
$\mathfrak{D}$empster's rule

$$\overline{P}_{\mathfrak{D}}\,(A \mid B) = \frac{\sup_{P\in\Pi} P(A\cap B)}{\sup_{P\in\Pi} P(B)}$$

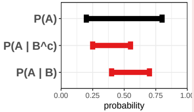
$\mathfrak{G}$eometric rule

$$\underline{P}_{\mathfrak{G}}\,(A \mid B) = \frac{\inf_{P\in\Pi} P(A\cap B)}{\inf_{P\in\Pi} P(B)}$$

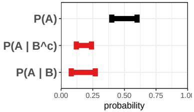
Dilation

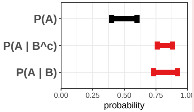


Contraction



Sure Loss





Thm 1. $\mathfrak{B}$-rule cannot contract nor induce sure loss.

Thm 2. Conditioning using $\mathfrak{B}$-rule results in a superset of posterior probabilities than $\mathfrak{D}$- and $\mathfrak{G}$-rules.

- Thus, if either $\mathfrak{D}$- or $\mathfrak{G}$-rule dilates, $\mathfrak{B}$-rule dilates.

Gen. 𝔅ayes rule

$$\overline{P}_{\mathfrak{B}}(A \mid B) = \sup_{P\in\Pi} \frac{P(A\cap B)}{P(B)}$$

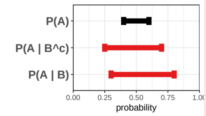
𝔇empster’s rule

$$\overline{P}_{\mathfrak{D}}(A \mid B) = \frac{\sup_{P\in\Pi} P(A\cap B)}{\sup_{P\in\Pi} P(B)}$$

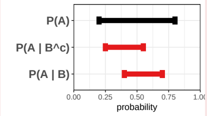
𝔊eometric rule

$$\underline{P}_{\mathfrak{G}}(A \mid B) = \frac{\inf_{P\in\Pi} P(A\cap B)}{\inf_{P\in\Pi} P(B)}$$

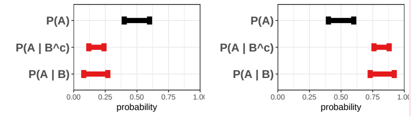
Dilation



Contraction



Sure Loss



Thm 3. Neither 𝔅-rule nor 𝔊-rule can sharpen vacuous priors.

Gen. 𝔅ayes rule

$$\overline{P}_{\mathfrak{B}}(A \mid B) = \sup_{P\in\Pi} \frac{P(A\cap B)}{P(B)}$$

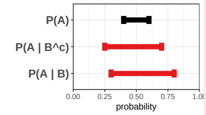
𝔇empster's rule

$$\overline{P}_{\mathfrak{D}}(A \mid B) = \frac{\sup_{P\in\Pi} P(A\cap B)}{\sup_{P\in\Pi} P(B)}$$

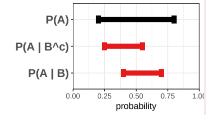
𝔊eometric rule

$$\underline{P}_{\mathfrak{G}}(A \mid B) = \frac{\inf_{P\in\Pi} P(A\cap B)}{\inf_{P\in\Pi} P(B)}$$

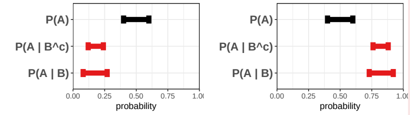
Dilation



Contraction



Sure Loss



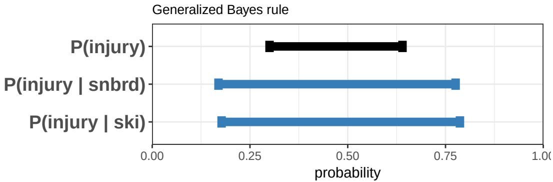
Thm 3. Neither 𝔅-rule nor 𝔊-rule can sharpen vacuous priors.

Thm 4. Counteractions of 𝔇-rule and 𝔊-rule. Let $\mathcal{B} = \{B, B^c\}$:

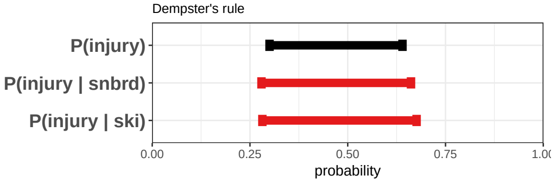
- If $\mathcal{B}$ dilates A under 𝔊-rule, it contracts A under 𝔇-rule;
- If $\mathcal{B}$ dilates A under 𝔇-rule, it contracts A under 𝔊-rule.

Survey nonresponse: conditional injury rate

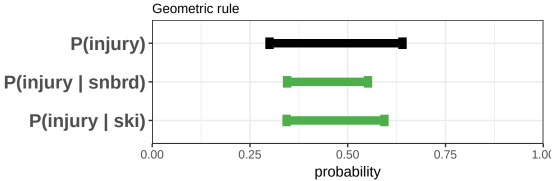
$\mathfrak{B}$-rule:



$\mathfrak{D}$-rule:

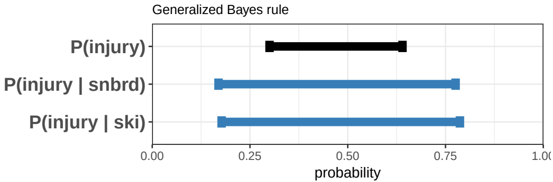


$\mathfrak{G}$-rule:

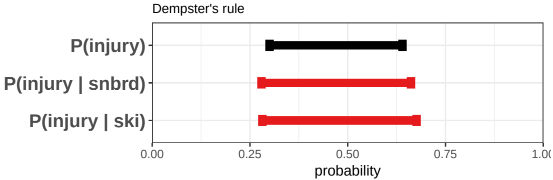


Survey nonresponse: conditional injury rate

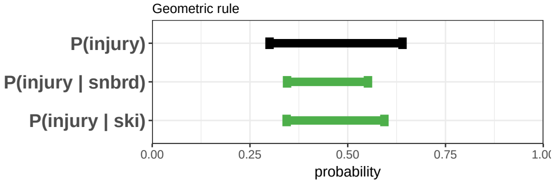
$\mathfrak{B}$-rule: sport type **dilates** injury rate



$\mathfrak{D}$-rule: sport type **dilates** injury rate



$\mathfrak{G}$-rule: sport type **contracts** injury rate




© Tom Roeleveld (CC BY-NC 2.0)
There's more to uncertainty than probabilities
http://www.ISIPTA2019.ugent.be
3 - 6 July
Ghent, Belgium
The 20-year anniversary edition of the
world's main forum on imprecise probabilities

Gong, R. (2018) *Low-resolution Statistical Modeling with Belief Functions* (doctoral dissertation). Harvard University, Cambridge, MA.

Berger, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, *25*(3), 303–328.

Dempster, A. P. (2008). The Dempster-Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, *48*(2), 365–377.

Epstein, L. G., Kaido, H., & Seo, K. (2016). Robust confidence regions for incomplete models. *Econometrica*, *84*(5), 1799–1838.

Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica*, *19*(2), 491–544.

Heitjan, D. F., & Rubin, D. B. (1990). Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association*, *85*(410), 304–314.

Huber, P. J., & Ronchetti, E. M. (2009). *Robust statistics*. Wiley.

Huber, P. J., & Strassen, V. (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *The Annals of Statistics*, *1*(2), 251–263.

Kline, B., & Tamer, E. (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics*, *7*(2), 329–366.

Seidenfeld, T., & Wasserman, L. (1993). Dilation for sets of probabilities. *The Annals of Statistics*, *21*(3), 1139–1154.

Wasserman, L. (1992). Invariance properties of density ratio priors. *The Annals of Statistics*, *20*(4), 2177–2182.