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)

- 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
- 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 $(\Omega, \mathcal{F}(\Omega))$, 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 $\infty$)

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:

\[ \mathcal{P}(\Omega) = \{A : A \subseteq \Omega\} \]

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>{Y, N}</th>
<th>{Y, N}</th>
<th>Y</th>
<th>N</th>
<th>{Y, N}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[m(R)]</td>
<td>0.11</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>
What can we say about injury rate, $P(Y)$?

$$P(Y) = P(\{Y, K\}, \{Y, S\})$$

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

The set functions $P$ and $m(R)$ form a belief function, a special kind of lower probability function; $P$ is a plausibility function, a special kind of upper probability function conjugate to $P$: $P(A) = 1 - P(A^c)$. 

<table>
<thead>
<tr>
<th>$Q_1$</th>
<th>Y</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>{Y, N}</th>
<th>{Y, N}</th>
<th>Y</th>
<th>N</th>
<th>{Y, N}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_2$</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td>{K, S}</td>
</tr>
<tr>
<td>$m(R)$</td>
<td>0.11</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>
What can we say about injury rate, $P(Y)$?

<table>
<thead>
<tr>
<th>$Q_1$</th>
<th>Y</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>{Y, N}</th>
<th>{Y, N}</th>
<th>Y</th>
<th>N</th>
<th>{Y, N}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_2$</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td>{K, S}</td>
</tr>
<tr>
<td>$m(R)$</td>
<td>0.11</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>
What can we say about injury rate, $P(Y)$?

$$P(Y) = P(\{(Y, K), (Y, S)\})$$

$$\geq m(\{(Y, K)\}) + m(\{(Y, S)\}) + m(\{(Y, K), (Y, S)\}) = 0.3$$
<table>
<thead>
<tr>
<th>Q₁</th>
<th>Y</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>{Y, N}</th>
<th>{Y, N}</th>
<th>Y</th>
<th>N</th>
<th>{Y, N}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q₂</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td>{K, S}</td>
</tr>
</tbody>
</table>

| m(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**, \( P(Y) \)?

\[
P(Y) \overset{\text{def}}{=} m(\{(Y, K)\}) + m(\{(Y, S)\}) + m(\{(Y, K), (Y, S)\}) = 0.3
\]
<table>
<thead>
<tr>
<th>$Q_1$</th>
<th>Y</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>{Y, N}</th>
<th>{Y, N}</th>
<th>Y</th>
<th>N</th>
<th>{Y, N}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_2$</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td>{K, S}</td>
</tr>
<tr>
<td>$m(R)$</td>
<td>0.11</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>

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

\[
\overline{P}(Y) \overset{\text{def}}{=} m(\{(Y, K)\}) + m(\{(Y, S)\}) + m(\{(Y, K), (Y, S)\}) = 0.3 \\
P(Y) = 1 - P(N)
\]
What can we say about \textbf{injury rate}, \(P(Y)\)?

\[
\overline{P}(Y) \overset{\text{def}}{=} m\left(\{(Y, K)\}\right) + m\left(\{(Y, S)\}\right) + m\left(\{(Y, K), (Y, S)\}\right) = 0.3
\]

\[
\overline{P}(Y) = 1 - P(N) = 0.64
\]
What can we say about **injury rate**, \( P(Y) \)?

\[
P(Y) \overset{\text{def}}{=} m(\{(Y, K)\}) + m(\{(Y, S)\}) + m(\{(Y, K), (Y, S)\}) = 0.3
\]

\[
\overline{P}(Y) = 1 - P(N) = 0.64
\]

The set functions

- \( P : \mathcal{P}(\Omega) \rightarrow [0, 1] \) is a **belief function**, a special kind of **lower probability function**;

- \( \overline{P} \) is a **plausibility function**, a special kind of **upper probability function** conjugate to \( P \): \( \overline{P}(A) = 1 - P(A^c) \).
Prior-free multinomial inference

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

**Multinomial as “partially identified” Poisson.**

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

- \( \lambda \) is aliased with \( T \) and non-identifiable. However:
  - \( q = \lambda / \| \lambda \| \) is free of \( T \) and identifiable.
Prior-free multinomial inference

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

Observed data: \(X = x\).
Prior-free multinomial inference

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

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

\[
X_i = \arg\max_k \left( \sum_{j=0}^{k} E_{ij} \leq \lambda_i \right), \quad i = 1, \ldots, K
\]

where auxiliary variables \(E_{ij} \sim \text{Exp}(1)\) are ancillary to \(\lambda\); \(E_{i0} = 0\).

No prior assumptions on \(\lambda\).
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 \( \Omega \):

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

which is understood as a *random subset generated* by \((G, \Delta E)\).
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 $\Omega$:

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

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

$$Q_x \overset{\text{def}}{=} \{q \in S_K : X = x, \lambda \in \otimes_{i=1}^{K} [G_i, G_i + \Delta E_i], q = \lambda/\|\lambda\|\}$$
\(Q_x\) is a random convex \((K - 1)\)-polytope in \(\mathbb{S}_K\).

**Figure:** One realization of \(Q_x\): \(\mathbf{x} = (2, 1, 1)\).
$Q_x$ is a random convex $(K - 1)$-polytope in $S_K$ with two characterizations:

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

Vertices of \( Q_x \) are (dependent) Dirichlet random variables.

Tier 0:
\[
\frac{1}{\sum G_i} \ (G_1, \ldots, G_K)
\]

Tier 1:
\[
\frac{1}{\sum G_i + \Delta E_1} \ (G_1 + \Delta E_1, G_2, \ldots, G_K)
\]
\[
\vdots
\]

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

Tier K-1:
\[
\frac{1}{\sum G_i + \sum_{i=1}^{K-1} \Delta E_i} \ (G_1, G_2 + \Delta E_2, \ldots, G_K + \Delta E_K)
\]

Tier K:
\[
\frac{1}{\sum G_i + \sum \Delta E_i} \ (G_1 + \Delta E_1, \ldots, G_K + \Delta E_K)
\]
Figure: One realization of $Q_x$: $x = (2, 1, 1)$. 
Figure: Realization of $Q_x$: left, $x = (3, 1, 0)$; right, $x = (4, 0, 0)$. 
$Q_x$ characterized by a system of inequalities

Each realization of $Q_x$ constitutes a *density ratio class* (Wasserman, 1992) of probabilities:

$$Q_x = \left\{ q \in S_K : \frac{q_i}{q_j} \geq \frac{G_i}{G_j + \Delta E_j}, \forall (i, j) \in \{1, ..., K\}^2, i \neq j \right\}.$$
Lemma 3.2.2-2. (polytope intersection)

Let $Q \subseteq S_K$. $Q \cap Q' \neq \emptyset$ iff $\forall i \neq j, i, j \in \{1, \ldots, K\}$,

$$\text{Proj}_{(i,j)}(Q) \cap \text{Proj}_{(i,j)}(Q') \neq \emptyset,$$

where $\text{Proj}_{(i,j)}(Q)$ is the projection of $Q$ onto $S_2$ spanned by the $(i, j)^{th}$ dimensions.
Posterior inference for $q$: one- and two-sample

Suppose $X, X'$ are two independently observed vectors of counts, and $q, q' \in S_K$ their respective proportion vectors.

**Theorem 3.2.3. (one-sample)**

For the hypothesis $H : q = q^*$, the posterior upper probability is

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

**Theorem 3.2.4. (two-sample)**

For hypothesis $H : q = q'$, the posterior upper probability is

$$
\overline{P}(H) = Pr (Q_x \cap Q'_{x'} \neq \emptyset) = Pr \left( \bigcap_{(i,j) \in \{1,\ldots,K\}^2 : i \neq j} \left( \frac{G_i + \Delta E_i}{G_j} \geq \frac{G'_i}{G'_j + \Delta E'_j} \right) \right).
$$
Let $t : \mathbb{S}^2_k \to \mathcal{T}$ be a function of multinomial proportion vectors. Define a **one-sample** posterior test statistic

$$R_{t|x} = \text{Cl} \left( \left\{ t(q, q_0) \in \mathbb{R}^d : q \in Q_x \right\} \right),$$

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

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

$R_{t|x}$ and $R_{t|x, x'}$ are again random closed subsets of $\mathcal{T}$, whose distributions dictated by that of $Q_x$ and of $(Q_x, Q'_{x'})$ respectively.
Posterior Hellinger statistic

Hellinger distance between two discrete probability distributions:

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

The two-sample posterior Hellinger distance \( R_{h|x,x'} \) is the closed interval

\[ [h_{\min}, h_{\max}] \overset{\text{def}}{=} \left[ \min_{q \in Q_x, q' \in Q'_x} h(q, q'), \max_{q \in Q_x, q' \in Q'_x} h(q, q') \right]. \]
Posterior Hellinger statistic: under $H_0$

Figure: $K = 5$, $N = 100$, $N' = 108$, $q = (0.12, 0.20, 0.24, 0.32, 0.12)$. Under $H_0$, $q' = q$. Under $H_1$, $q' \propto (0.06, 0.06, 0.12, 0.256, 0.048)$. 
Posterior Hellinger statistic: under $H_1$

**Figure:** $K = 5$, $N = 100$, $N' = 108$, $q = (0.12, 0.20, 0.24, 0.32, 0.12)$. Under $H_0$, $q' = q$. Under $H_1$, $q' \propto (0.06, 0.06, 0.12, 0.256, 0.048)$. 
Sieve analysis: vaccine equi-efficacy

Sieve effect: DS posterior two-sample squared Hellinger statistic

black dotted line: 95% quantile of Chi-squared (df = 3)
A precise probability model updates according to, e.g., Bayes rule.

\[ P(A) \implies P(A \mid B) \]

What if \( P \) is imprecise?
### Updating sets of probabilities

**Rules**

1. Generalized Bayes rule
2. Dempster’s rule
3. Geometric rule

**Paradoxes**

- Dilation
- Contraction
- Sure Loss
Survey nonresponse: what’s $P(\text{injury} \mid \text{ski})$?

<table>
<thead>
<tr>
<th>$Q_1$</th>
<th>Y</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>{Y, N}</th>
<th>{Y, N}</th>
<th>Y</th>
<th>N</th>
<th>{Y, N}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_2$</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td></td>
</tr>
<tr>
<td>$m(R)$</td>
<td>0.11</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>

One option can be:

$$P_{\#3} (\text{injury} \mid \text{ski}) \overset{\text{def}}{=} \inf_{P \in \Pi} \frac{P(\text{injury, ski})}{P(\text{ski})}$$

$$\overline{P}_{\#3} (\text{injury} \mid \text{ski}) \overset{\text{def}}{=} \sup_{P \in \Pi} \frac{P(\text{injury, ski})}{P(\text{ski})}$$
Survey nonresponse: what’s $P(\text{injury} \mid \text{ski})$?

<table>
<thead>
<tr>
<th>$Q_1$</th>
<th>Y</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>${Y, N}$</th>
<th>${Y, N}$</th>
<th>Y</th>
<th>N</th>
<th>${Y, N}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_2$</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>${K, S}$</td>
<td>${K, S}$</td>
<td>K</td>
<td>S</td>
<td>${K, S}$</td>
</tr>
<tr>
<td>$m(\mathbf{R})$</td>
<td>0.11</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Generalized Bayes rule:

$$P_{QB}(\text{injury} \mid \text{ski}) \overset{\text{def}}{=} \inf_{P \in \Pi} \frac{P(\text{injury}, \text{ski})}{P(\text{ski})}$$

$$\overline{P}_{QB}(\text{injury} \mid \text{ski}) \overset{\text{def}}{=} \sup_{P \in \Pi} \frac{P(\text{injury}, \text{ski})}{P(\text{ski})}$$
Survey nonresponse: what’s \( P(\text{injury} | \text{ski}) \)?

<table>
<thead>
<tr>
<th>( Q_1 )</th>
<th>Y</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>{Y, N}</th>
<th>{Y, N}</th>
<th>Y</th>
<th>N</th>
<th>{Y, N}</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q_2 )</td>
<td>K</td>
<td>S</td>
<td>K</td>
<td>S</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td>{K, S}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m(R) )</td>
<td>0.11</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
<td>0.08</td>
<td>0.06</td>
<td>0.09</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Generalized Bayes rule:

\[
P(\text{injury})
\]

\[
P(\text{injury} | \text{snbrd})
\]

\[
P(\text{injury} | \text{ski})
\]

“Posterior” interval grows wider regardless of sport type!
**Definition: Dilation** (cf. Seidenfeld & Wasserman (1993))

Let $A \in \mathcal{B}(\Omega)$, $\mathcal{B}$ a Borel-measurable partition of $\Omega$, $\Pi$ be a closed and convex set of probability measures on $\Omega$, $P$ its lower probability function, and $P_\bullet$ the conditional lower probability function supplied by the updating rule “$\bullet$”.

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

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

\[ \overline{P_B} (A | B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)} \]

Dilation

\[
\begin{array}{c|cc|c|ccc|cc}
 & P(A) & P(A | B^c) & P(A | B) \\
\hline
P(A) & 0.00 & 0.25 & 0.50 & 0.75 & 1.00 \\
\end{array}
\]

Thm 1. B-rule cannot contract nor induce sure loss.

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

▶ Thus, if either D- or G-rule dilates, B-rule dilates.
Gen. Bayes rule

\[ \overline{P}_B (A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)} \]

Dempster’s rule

\[ \overline{P}_D (A \mid B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)} \]

Geometric rule

\[ \overline{P}_G (A \mid B) = \frac{\inf_{P \in \Pi} P(A \cap B)}{\inf_{P \in \Pi} P(B)} \]

Dilation

\[
\begin{array}{c}
P(A) \quad P(A \mid B^c) \quad P(A \mid B) \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c|c}
\text{probability} & 0.00 & 0.25 & 0.50 & 0.75 & 1.00 \\
\hline
P(A) & - & - & - & - & - \\
P(A \mid B^c) & - & - & - & - & - \\
P(A \mid B) & - & - & - & - & - \\
\end{array}
\]

Thm 1. B-rule cannot contract nor induce sure loss.

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

▶ Thus, if either D- or G-rule dilates, B-rule dilates.
General Bayes rule

\[ \overline{P}_B (A | B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)} \]

Dempster’s rule

\[ \overline{P}_D (A | B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)} \]

Geometric rule

\[ P_G (A | B) = \frac{\inf_{P \in \Pi} P(A \cap B)}{\inf_{P \in \Pi} P(B)} \]

**Dilation**

- \( P(A) \)
- \( P(A | B^c) \)
- \( P(A | B) \)

**Contraction**

- \( P(A) \)
- \( P(A | B^c) \)
- \( P(A | B) \)
\[ \overline{P}_\mathbb{B} (A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)} \]
\[ \overline{P}_\mathbb{D} (A \mid B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)} \]
\[ P_\mathbb{G} (A \mid B) = \frac{\inf_{P \in \Pi} P(A \cap B)}{\inf_{P \in \Pi} P(B)} \]
Thm 1. \( \mathcal{B} \)-rule cannot contract nor induce sure loss.
Gen. Bayes rule
\[ P_B(A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)} \]

Dempster’s rule
\[ \overline{P}_D(A \mid B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)} \]

Geometric rule
\[ P_G(A \mid B) = \frac{\inf_{P \in \Pi} P(A \cap B)}{\inf_{P \in \Pi} P(B)} \]

Thm 1. B-rule cannot contract nor induce sure loss.

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

▶ Thus, if either D- or G-rule dilates, B-rule dilates.
Gen. Bayes rule
\[ \overline{P_B}(A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)} \]

Dempster’s rule
\[ \overline{P_D}(A \mid B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)} \]

Geometric rule
\[ \overline{P_G}(A \mid B) = \inf_{P \in \Pi} \frac{P(A \cap B)}{P(B)} \]

Dilation

Contraction

Sure Loss

Thm 3. Neither $B$-rule nor $G$-rule can sharpen vacuous priors.
**Gen. Bayes rule**

\[ P_{\Sigma} (A \mid B) = \sup_{P \in \Pi} \frac{P(A \cap B)}{P(B)} \]

**Dempster’s rule**

\[ P_{\Omega} (A \mid B) = \frac{\sup_{P \in \Pi} P(A \cap B)}{\sup_{P \in \Pi} P(B)} \]

**Geometric rule**

\[ P_{\Theta} (A \mid B) = \frac{\inf_{P \in \Pi} P(A \cap B)}{\inf_{P \in \Pi} P(B)} \]

**Dilation**

<table>
<thead>
<tr>
<th>P(A)</th>
<th>P(A \mid B)</th>
<th>P(A \mid B^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>0.50</td>
<td>0.75</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Contraction**

<table>
<thead>
<tr>
<th>P(A)</th>
<th>P(A \mid B)</th>
<th>P(A \mid B^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>0.75</td>
<td>0.50</td>
<td>0.25</td>
</tr>
</tbody>
</table>

**Sure Loss**

<table>
<thead>
<tr>
<th>P(A)</th>
<th>P(A \mid B)</th>
<th>P(A \mid B^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.75</td>
<td>0.50</td>
</tr>
<tr>
<td>0.00</td>
<td>0.25</td>
<td>0.50</td>
</tr>
</tbody>
</table>

**Thm 3.** Neither \( B \)-rule nor \( G \)-rule can sharpen vacuous priors.

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

- If \( B \) dilates \( A \) under \( G \)-rule, it contracts \( A \) under \( D \)-rule;
- If \( B \) dilates \( A \) under \( D \)-rule, it contracts \( A \) under \( G \)-rule.
Survey nonresponse: conditional injury rate

\textbf{B}-rule:

\textbf{D}-rule:

\textbf{G}-rule:
Survey nonresponse: conditional injury rate

\( \mathbb{B} \)-rule: sport type dilates injury rate

\( \mathcal{D} \)-rule: sport type dilates injury rate

\( \mathcal{G} \)-rule: sport type contracts injury rate
There’s more to uncertainty than probabilities

http://www.ISIPTA 2019.ugent.be

3 - 6 July
Ghent, Belgium

The 20-year anniversary edition of the world’s main forum on imprecise probabilities


