Understanding the Epistemic Utility of Imprecise Probabilities in Statistical Inference

Ruobin Gong

Department of Statistics, Rutgers University, USA

EPIMP Inaugural Conference
University of Bristol, UK

Oct 22, 2021

There's more to uncertainty than probabilities

In statistical modeling, imprecise probabilities are useful for capturing structural uncertainty:

- Prior: robust Bayes (Berger, 1990);
- Data: missing and coarse data;
- Model structure (other than the prior and the data):
  - Dempster-Shafer theory of belief functions (Dempster, 2008);
  - Inferential models (Martin & Liu, 2015);
  - Robust frequentist statistics (Huber & Strassen, 1973);
  - Partial identification and incomplete models (e.g. Manski, 2003; Epstein et al., 2016);
  - Statistical disclosure limitation (SDL) and differential privacy (DP).

The (epistemic) utility of IP in statistics

The epistemic utility of a state of full belief K at a world w (Levi, 2004; Konek, 2019):

$$\mathcal{V}(K; w) = a \cdot \mathbb{E}(K; w) + (1 - a) \cdot \mathbb{T}(K; w),$$

where

- $\mathbb{E}(K; w)$ is the truth-value of K at w;
- $\mathbb{T}(K; w)$ is K's degree of informativeness at w; and
- a is a measure of relative priority between “error-avoidance” and “truth-seeking”.

How to formulate $\mathcal{V}(K; w)$ in a typical statistical modeling task?

The (epistemic) utility of IP in statistics

$$\mathcal{V}(K; w) = a \cdot \mathbb{E}(K; w) + (1 - a) \cdot \mathbb{T}(K; w),$$

How to formulate $\mathcal{V}(K; w)$ in a typical statistical modeling task?

Ex. A frequentist confidence interval/calibrated Bayesian posterior interval should:

- Have **coverage** probabilities ideally at, and never below, the nominal level everywhere in the parameter space;
- Be **small** and **sensible** (e.g. connected, bounded, etc).

Ex. (IM; Martin & Liu, 2015) Two principles applicable to a predictive random set $\mathcal{S}$:

- **validity**: $\sup_{\theta \notin A} P_{X|\theta} \left(\text{bel}_X (A; \mathcal{S}) \geq 1 - \alpha \right) \leq \alpha$;
- **efficiency**: $\text{bel}_x (A; \mathcal{S}') \geq \text{bel}_x (A; \mathcal{S})$ for fixed A and every $x \in \mathbb{X}$.

$$\mathcal{V}(K; w) = a \cdot \mathbb{E}(K; w) + (1 - a) \cdot \mathbb{T}(K; w),$$

How to formulate $\mathcal{V}(K; w)$ in a typical statistical modeling task?

Ex. A frequentist confidence interval/calibrated Bayesian posterior interval should:

- Have **coverage** probabilities ideally at, and never below, the nominal level everywhere in the parameter space;
- Be **small** and **sensible** (e.g. connected, bounded, etc).

Ex. (IM; Martin & Liu, 2015) Two principles applicable to a predictive random set $\mathcal{S}$:

- **validity**: $\sup_{\theta \notin A} P_{X|\theta} \left(\mathrm{bel}_X (A; \mathcal{S}) \geq 1 - \alpha\right) \leq \alpha$;
- **efficiency**: $\mathrm{bel}_x (A; \mathcal{S}') \geq \mathrm{bel}_x (A; \mathcal{S})$ for fixed A and every $x \in \mathbb{X}$.

$$\mathcal{V}(K; w) = a \cdot \mathbb{E}(K; w) + (1 - a) \cdot \mathbb{T}(K; w),$$

How to formulate $\mathcal{V}(K; w)$ in a typical statistical modeling task?

Ex. A frequentist confidence interval/calibrated Bayesian posterior interval should:

- Have **coverage** probabilities ideally at, and never below, the nominal level everywhere in the parameter space;
- Be **small** and **sensible** (e.g. connected, bounded, etc).

Ex. (IM; Martin & Liu, 2015) Two principles applicable to a predictive random set $\mathcal{S}$:

- **validity**: $\sup_{\theta \notin A} P_{X|\theta} \left(\text{bel}_X \left(A; \mathcal{S}\right) \geq 1 - \alpha\right) \leq \alpha$;
- **efficiency**: $\text{bel}_x \left(A; \mathcal{S}'\right) \geq \text{bel}_x \left(A; \mathcal{S}\right)$ for fixed A and every $x \in \mathbb{X}$.

The (epistemic) utility of IP in statistics

$$\mathcal{V}(K; w) = a \cdot \mathbb{E}(K; w) + (1 - a) \cdot \mathbb{T}(K; w),$$

How to formulate $\mathcal{V}(K; w)$ in a typical statistical modeling task?

Considerations:

- Can $\mathcal{V}(K; w)$ aid IP model choice? Specifically, when does $\mathcal{V}(K; w)$ attribute superior epistemic utility to <u>imprecise</u> probabilities (as opposed to precise probabilities)?
- The dynamic evaluation of $\mathcal{V}(K; w)$ as the analyst updates K in light of new information;
- When "truth-seeking" may not be desirable: the case of statistical privacy;
- *Computation: what is epistemic utility if without practicality?

Case studies

1. IP for capturing total evidence in hypothesis testing

2. IP updating rules and paradoxes

3. The IP nature of differential privacy

Total evidence in hypothesis testing

Background. Denote

- F: an observable event, and
- $K_{t,M}(F)$: the event that agent learns at time t by method M that F obtains.

The agent's corpus of knowledge at time t consists of

$$F \,\&\, K_{t,M}(F) .$$

GKSS (2021) motivated situations in which a rational agent can identify an observable F that is **sufficient** for $(F, K_{t,M}(F))$, i.e.

$$Cred\,(\cdot \mid F \,\&\, K_{t,M}(F)) = Cred\,(\cdot \mid F)\,,$$

but must resort to IP tools to describe this credence.

In frequentist statistical inference,
recognizing the total evidence (and using IP to capture it) is just as important.

Total evidence in hypothesis testing

(Barnard, 1947) A bag of chrysanthemum seeds: either ♣ or ♣. The statistician wants to study their relative proportion.

Experiment: A lab scientist sows a random sample of seeds and records the flowers' colors.

The level of significance of the observed outcome F is

$$\text{argmax}_{P \in H_0} P\left(R_F\right),$$

i.e. the max prob (under the null hypothesis H_0) of the reference class, R_F, consisting of events deemed as of matching or exceeding significance compared to F.

Scenario 1. The lab scientist reports ♣ $\times$ 9 and ♣ $\times$ 0.

Scenario 1. ♣ × 9 and ♣ × 0. Writing

$$n = (\#♣) + (\#♣),$$
$$r = (\#♣) - (\#♣),$$

the **observational report** is

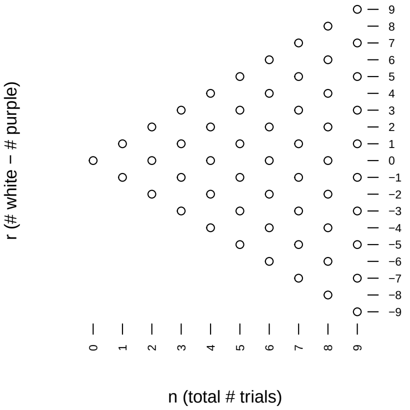
$$F: \quad n = 9, r = 9.$$

The **null hypothesis**:

$$H_0: \quad \left(\frac{n+r}{2}, \frac{n-r}{2}\right) \sim Bin(n, \frac{1}{2}).$$

The **level of significance**:

$$2 \times (1/2)^9 \doteq 0.39\%.$$



Total evidence in hypothesis testing

Scenario 1. ♣ × 9 and ♣ × 0. Writing

$$n = (\#\clubsuit) + (\#\clubsuit),$$
$$r = (\#\clubsuit) - (\#\clubsuit),$$

the **observational report** is

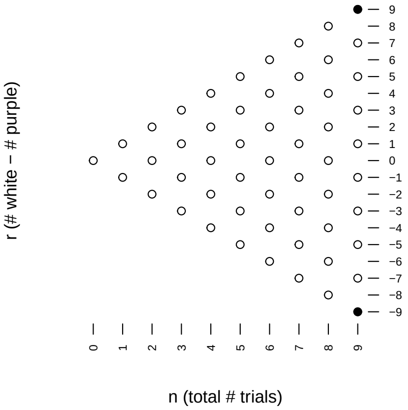
$$F: \quad n = 9, r = 9.$$

The **null hypothesis**:

$$H_0: \quad \left(\frac{n+r}{2}, \frac{n-r}{2}\right) \sim Bin(n, \frac{1}{2}).$$

The **level of significance**:

$$2 \times (1/2)^9 \doteq 0.39\%.$$



Total evidence in hypothesis testing

Scenario 2. The lab scientist reports an additional failed flower: ♣ × 1. The **total evidence** is

$$F' : n = 9, r = 9, N = 10, \text{“premature plant destruction”}.$$

The **new null hypothesis** is

$$H_0' : \left(\frac{n+r}{2}, n_w, \frac{n-r}{2}, n_p\right) \sim$$

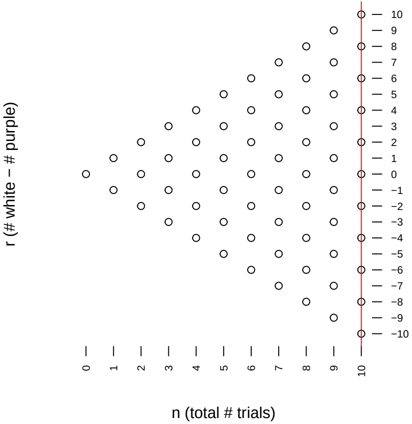
$$\text{Mult}\left(N, \left(\frac{p_w}{2}, \frac{1-p_w}{2}, \frac{p_p}{2}, \frac{1-p_p}{2}\right)\right),$$

for $p_w \in [0,1]$, $p_p \in [0,1]$.

The **new level of significance** is

$$\text{argmax}_{p_w \in [0,1], p_p \in [0,1]} P\left(R_{F'}\right) \doteq 1.07\%,$$

where



$$P\left(R_{F'}\right) = \left[10\left(\frac{p_w + p_p}{2}\right) + \frac{1-p_w}{2}\right]\left(\frac{1-p_w}{2}\right)^9 + \left[10\left(\frac{p_w + p_p}{2}\right) + \frac{1-p_p}{2}\right]\left(\frac{1-p_p}{2}\right)^9.$$

Total evidence in hypothesis testing

Scenario 2. The lab scientist reports an additional failed flower: ♣ × 1. The **total evidence** is

$$F' : n = 9, r = 9, N = 10, \text{“premature plant destruction”}.$$

The **new null hypothesis** is

$$H_0' : \left(\frac{n+r}{2}, n_w, \frac{n-r}{2}, n_p\right) \sim$$

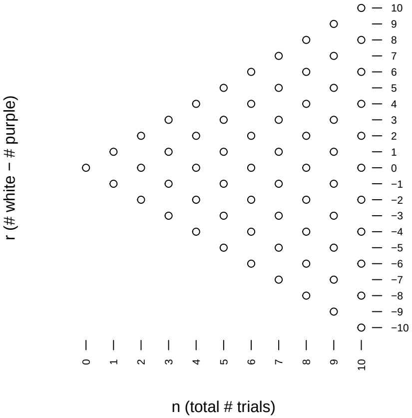
$$\text{Mult}\left(N, \left(\frac{p_w}{2}, \frac{1-p_w}{2}, \frac{p_p}{2}, \frac{1-p_p}{2}\right)\right),$$

for $p_w \in [0,1]$, $p_p \in [0,1]$.

The **new level of significance** is

$$\text{argmax}_{p_w \in [0,1], p_p \in [0,1]} P\left(R_{F'}\right) \doteq 1.07\%,$$

where



$$P\left(R_{F'}\right) = \left[10\left(\frac{p_w+p_p}{2}\right) + \frac{1-p_w}{2}\right]\left(\frac{1-p_w}{2}\right)^9 + \left[10\left(\frac{p_w+p_p}{2}\right) + \frac{1-p_p}{2}\right]\left(\frac{1-p_p}{2}\right)^9.$$

Total evidence in hypothesis testing

Scenario 2. The lab scientist reports an additional failed flower: ♣ × 1. The **total evidence** is

$$F' : n = 9, r = 9, N = 10, \text{“premature plant destruction”}.$$

The **new null hypothesis** is

$$H_0' : \left(\frac{n+r}{2}, n_w, \frac{n-r}{2}, n_p\right) \sim$$

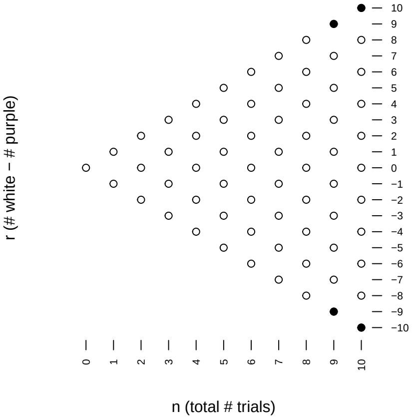
$$\text{Mult}\left(N, \left(\frac{p_w}{2}, \frac{1-p_w}{2}, \frac{p_p}{2}, \frac{1-p_p}{2}\right)\right),$$

for $p_w \in [0,1]$, $p_p \in [0,1]$.

The **new level of significance** is

$$\text{argmax}_{p_w \in [0,1], p_p \in [0,1]} P\left(R_{F'}\right) \doteq 1.07\%,$$

where



$$P\left(R_{F'}\right) = \left[10\left(\frac{p_w+p_p}{2}\right) + \frac{1-p_w}{2}\right]\left(\frac{1-p_w}{2}\right)^9 + \left[10\left(\frac{p_w+p_p}{2}\right) + \frac{1-p_p}{2}\right]\left(\frac{1-p_p}{2}\right)^9.$$

Scenario 3. The lab scientist confessed that he accidentally trod over and killed the young plant. The **total evidence** is now

$$F'' : n = 9, r = 9, N = 10, \text{“premature plant destruction due to a color-agnostic reason”}.$$

The **new null hypothesis** is

$$H_0'' : \left(\frac{n+r}{2}, n_w, \frac{n-r}{2}, n_p\right) \sim$$

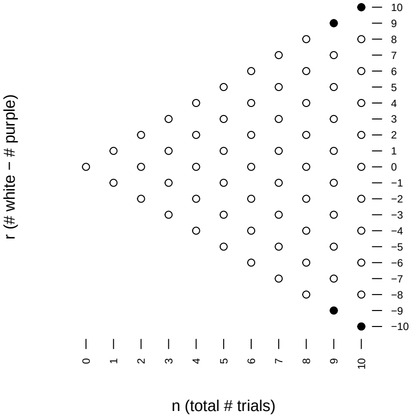
$$\text{Mult}\left(N, \left(\frac{p}{2}, \frac{1-p}{2}, \frac{p}{2}, \frac{1-p}{2}\right)\right),$$

$p \in [0, 1]$. The **new level of significance**

$$\text{argmax}_{p\in[0,1]} P\left(R_{F''}\right) \doteq 0.24\%,$$

where

$$P\left(R_{F''}\right) = 2\left[10p + \frac{1-p}{2}\right]\left(\frac{1-p}{2}\right)^9.$$



Total evidence in hypothesis testing

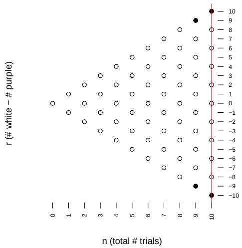
Summary. Three sets of **total evidence of increasing richness**:

1. $F\ : n = 9, r = 9;$
2. $F'\ : n = 9, r = 9, N = 10,$ "premature plant destruction";
3. $F''\ : n = 9, r = 9, N = 10,$ "premature plant destruction due to a color-agnostic reason".

They impact not only the meaning of the **null hypotheses**:

1. $H_0\ :\quad \left(\frac{n+r}{2}, \frac{n-r}{2}\right) \sim Bin(n, \tfrac{1}{2});$
2. $H_0'\ :\quad \left(\frac{n+r}{2}, n_w, \frac{n-r}{2}, n_p\right) \sim \text{Mult}\left(N, \left(\frac{p_w}{2}, \frac{1-p_w}{2}, \frac{p_p}{2}, \frac{1-p_p}{2}\right)\right);$
3. $H_0''\ :\quad \left(\frac{n+r}{2}, n_w, \frac{n-r}{2}, n_p\right) \sim \text{Mult}\left(N, \left(\frac{p}{2}, \frac{1-p}{2}, \frac{p}{2}, \frac{1-p}{2}\right)\right),$

but also the **reference classes** of events for the hypothesis test.



The **levels of significance**, induced via IP,

$$1)\,0.39\%,\ \ 2)\,1.07\%,\ \text{ and } 3)\,0.24\%,$$

differ mildly in value yet substantially in meaning.

Can an IP epistemic utility theory advise on the choice of statistical models here?

Case studies

1. IP for capturing total evidence in hypothesis testing

2. IP updating rules and paradoxes

3. The IP nature of differential privacy

IP updating rules and paradoxes

Bayes' rule is the central updating rule in precise probability calculations.

For IP, the choice of updating rules is plural – so are the paradoxes:

Rules

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

Paradoxes

- Dilation
- Contraction
- Sure Loss

Example: survey with nonresponse

1. Did you injure yourself on the snow last season (**Y**/**N**)?
2. Do you ski or snowboard **(K/S)**?

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

For example, to learn about **injury rate**:

$$\underline{P}(Y) = 0.11 + 0.10 + 0.09 = 0.3;$$

$$\overline{P}(Y) = 0.11 + 0.10 + 0.09 + 0.08 + 0.06 + 0.20 = 0.64.$$

Example: survey with nonresponse

What is "$P\,(\text{injury} \mid \text{ski})$"?

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:

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

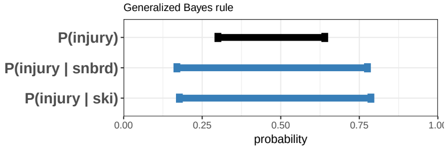
$$\overline{P}_{\mathfrak{B}}\,(\text{injury} \mid \text{ski}) \stackrel{\text{def}}{=\!=} \sup_{P \in \Pi} \frac{P\,(\text{injury}, \text{ski})}{P\,(\text{ski})}.$$

Example: survey with nonresponse

What is "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:



The upper and lower probability intervals for injury **dilate**, regardless of the conditioning event (K or S)!

Dilation (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. $\mathfrak{B}$ayes rule

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

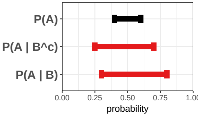
$\mathfrak{D}$empster's rule

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

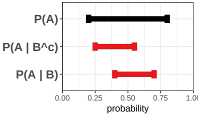
$\mathfrak{G}$eometric rule

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

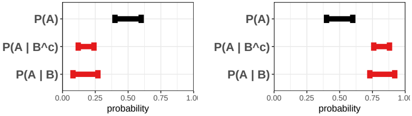
Dilation



Contraction



Sure Loss



Observations (Gong & Meng, 2021):

1. $\mathfrak{B}$-rule cannot contract nor induce sure loss.
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.
3. Neither $\mathfrak{B}$-rule nor $\mathfrak{G}$-rule can sharpen vacuous priors;
4. Let $\mathcal{B} = \{B, B^c\}$. If $\mathcal{B}$ dilates A under $\mathfrak{G}$-rule, it contracts A under $\mathfrak{D}$-rule. Similarly, if $\mathcal{B}$ dilates A under $\mathfrak{D}$-rule, it contracts A under $\mathfrak{G}$-rule.

In statistical inference:

Can dilation be "error-avoidance"? Can contraction and sure loss be "truth-seeking"?

Gen. $\mathfrak{B}$ayes rule

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

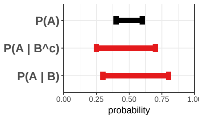
$\mathfrak{D}$empster's rule

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

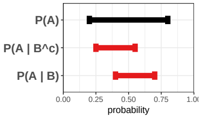
$\mathfrak{G}$eometric rule

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

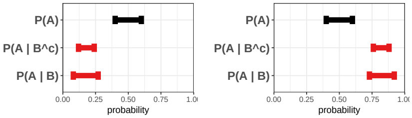
Dilation



Contraction



Sure Loss



Observations (Gong & Meng, 2021):

1. $\mathfrak{B}$-rule cannot contract nor induce sure loss.
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.
3. Neither $\mathfrak{B}$-rule nor $\mathfrak{G}$-rule can sharpen vacuous priors;
4. Let $\mathcal{B} = \{B, B^c\}$. If $\mathcal{B}$ dilates A under $\mathfrak{G}$-rule, it contracts A under $\mathfrak{D}$-rule. Similarly, if $\mathcal{B}$ dilates A under $\mathfrak{D}$-rule, it contracts A under $\mathfrak{G}$-rule.

In statistical inference:

Can dilation be "error-avoidance"? Can contraction and sure loss be "truth-seeking"?

Gen. $\mathfrak{B}$ayes rule

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

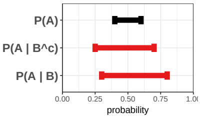
$\mathfrak{D}$empster's rule

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

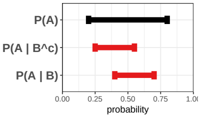
$\mathfrak{G}$eometric rule

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

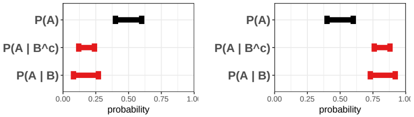
Dilation



Contraction



Sure Loss



Observations (Gong & Meng, 2021):

1. $\mathfrak{B}$-rule cannot contract nor induce sure loss.
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.
3. Neither $\mathfrak{B}$-rule nor $\mathfrak{G}$-rule can sharpen vacuous priors;
4. Let $\mathcal{B} = \{B, B^c\}$. If $\mathcal{B}$ dilates A under $\mathfrak{G}$-rule, it contracts A under $\mathfrak{D}$-rule. Similarly, if $\mathcal{B}$ dilates A under $\mathfrak{D}$-rule, it contracts A under $\mathfrak{G}$-rule.

In statistical inference:

Can dilation be "error-avoidance"? Can contraction and sure loss be "truth-seeking"?

Case studies

1. IP for capturing total evidence in hypothesis testing
2. IP updating rules and paradoxes
3. The IP nature of differential privacy

The IP nature of differential privacy

We wish to learn about the aggregate features of the observed dataset x, while protecting the privacy of the individual respondents. This is accomplished by a *randomized* query function.

Definition (ϵ-Differential Privacy; Dwork et al., 2006)

An $\mathscr{F}$-measurable randomized function T is ϵ-differentially private if for all pairs of datasets x, x' such that $d_H(x, x') = 1$ and all $S \in \mathscr{F}$,

$$Pr\left(T\left(x'\right) \in S\right) \leq \exp(\epsilon) \cdot Pr(T(x) \in S).$$

Ex (Laplace mechanism). Let $t : \mathcal{X} \to \mathbb{R}^d$ be a deterministic query of interest. Define the random query

$$T(x) = t(x) + E,$$

where $E \sim \text{Lap}_d(GS_t/\epsilon)$ a product Laplace variable, and GS_t the global sensitivity of t.

The privacy-accuracy tradeoff

less accuracy $\Leftrightarrow$ larger noise $\Leftrightarrow$ smaller privacy loss budget ϵ $\Leftrightarrow$ **more privacy**

The IP nature of differential privacy

We wish to learn about the aggregate features of the observed dataset x, while protecting the privacy of the individual respondents. This is accomplished by a *randomized* query function.

Definition (ϵ-Differential Privacy; Dwork et al., 2006)

An $\mathscr{F}$-measurable randomized function T is ϵ-differentially private if for all pairs of datasets x, x' such that $d_H(x, x') = 1$ and all $S \in \mathscr{F}$,

$$Pr\left(T\left(x'\right) \in S\right) \leq \exp(\epsilon) \cdot Pr(T(x) \in S).$$

Ex (Laplace mechanism). Let $t : \mathcal{X} \to \mathbb{R}^d$ be a deterministic query of interest. Define the random query

$$T(x) = t(x) + E,$$

where $E \sim \text{Lap}_d(GS_t/\epsilon)$ a product Laplace variable, and GS_t the global sensitivity of t.

The privacy-accuracy tradeoff

less accuracy $\Leftrightarrow$ larger noise $\Leftrightarrow$ smaller privacy loss budget ϵ $\Leftrightarrow$ **more privacy**

The IP nature of differential privacy

Definition (Interval of Measures; DeRobertis & Hartigan, 1981)

Let Ω be the set of all σ-finite measures on $(\mathcal{T}, \mathscr{F})$, and $L, U \in \Omega$ be a pair satisfying $L \leq U$, that is, $L(S) \leq U(S)$ for all $S \in \mathscr{F}$. Then, the convex set of measures

$$\mathcal{I}(L, U) = \{P \in \Omega : L \leq P \leq U\}$$

is called an *interval of measures*. L and U are called the *lower* and *upper measures*, respectively.

Note.

- IoM can be used to describe robust neighborhoods of *sampling distributions*. Lavine (1991)'s recursive algorithm computes various upper and lower posterior quantities;
- When L and U have densities with respect to some σ-finite dominating measure ν, the IoM defines a **density ratio class** probability neighborhood, which is invariant with respect to Bayesian updating (Wasserman, 1992) and is immune to dilation.

Definition (Interval of Measures; DeRobertis & Hartigan, 1981)

Let Ω be the set of all σ-finite measures on $(\mathcal{T}, \mathscr{F})$, and $L, U \in \Omega$ be a pair satisfying $L \leq U$, that is, $L(S) \leq U(S)$ for all $S \in \mathscr{F}$. Then, the convex set of measures

$$\mathcal{I}(L, U) = \{P \in \Omega : L \leq P \leq U\}$$

is called an *interval of measures*. L and U are called the *lower* and *upper measures*, respectively.

Note.

- IoM can be used to describe robust neighborhoods of *sampling distributions*. Lavine (1991)'s recursive algorithm computes various upper and lower posterior quantities;
- When L and U have densities with respect to some σ-finite dominating measure ν, the IoM defines a **density ratio class** probability neighborhood, which is invariant with respect to Bayesian updating (Wasserman, 1992) and is immune to dilation.

The IP nature of differential privacy

Definition (ϵ-Differential Privacy; Dwork et al., 2006)

An $\mathscr{F}$-measurable randomized function T is ϵ-differentially private if for all pairs of datasets x, x' such that $d_H(x, x') = 1$ and all $S \in \mathscr{F}$,

$$Pr\left(T\left(x'\right) \in S\right) \leq e^{\epsilon} \cdot Pr\left(T\left(x\right) \in S\right).$$

Proposition (ϵ-DP as Interval of Measures)

Let T be a random variable defined on $(\mathcal{T}, \mathscr{F})$, and P, Q be probability measures associated with $T(x)$ and $T(x')$ respectively. T is ϵ-differentially private iff for all $x, x' \in \mathcal{X}$ such that $d_H(x, x') = 1$,

$$P \in \mathcal{I}\left(L_\epsilon, U_\epsilon\right), \qquad \text{where } L_\epsilon = e^{-\epsilon} Q, \quad U_\epsilon = e^{\epsilon} Q.$$

Moreover, if P and Q have densities p and q with respect to a suitable measure (such as Lebesgue or counting), then for all $t \in \mathcal{T}$,

$$e^{-\epsilon} q(t) \leq p(t) \leq e^{\epsilon} q(t).$$

The IP nature of differential privacy

less accuracy $\Leftrightarrow$ larger noise $\Leftrightarrow$ smaller privacy loss budget ϵ $\Leftrightarrow$ **more privacy**

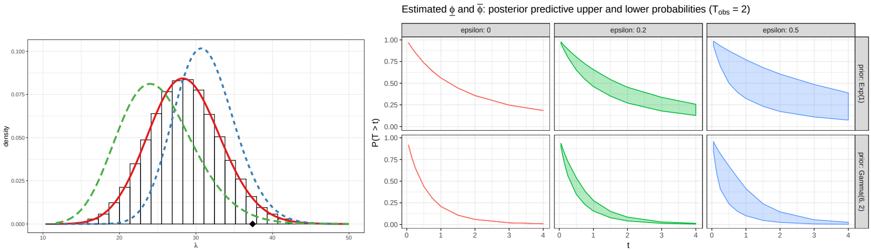


Figure: Left: (Gong, 2019) a privatized query ($\epsilon < \infty$) is statistically less informative than a non-privatized one ($\epsilon = \infty$); Right: Smaller ϵ induces narrower posterior predictive IoM over neighboring datasets, delivering more privacy.

Can an IP epistemic utility theory reflect moral desiderata such as privacy requirements?

Summary

Considerations for an IP epistemic utility theory in statistical applications:

- An aid to IP model choice, and an advocate for imprecision;
- Assessment of model behavior in dynamic learning;
- Moral desiderata (e.g. privacy);
- *Computation.

♣ Thank you! ♣

Bibliography

Barnard, G. A. (1947). The meaning of a significance level. *Biometrika*, *34*(1/2), 179–182.

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.

DeRobertis, L., & Hartigan, J. (1981). Bayesian inference using intervals of measures. *The Annals of Statistics*, *9*(2), 235–244.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265–284).

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

GKSS. (2021). Total evidence and learning with imprecise probabilities. In *International symposium on imprecise probability: Theories and applications* (pp. 161–168).

Gong, R. (2019). Exact inference with approximate computation for differentially private data via perturbations. *arXiv:1909.12237*.

Gong, R., & Meng, X.-L. (2021). Judicious judgment meets unsettling updating: dilation, sure loss, and Simpson's paradox (with discussion). *Statistical Science*, *36*(2), 169–190.

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

Konek, J. (2019). Ip scoring rules: foundations and applications. In *International symposium on imprecise probabilities: Theories and applications* (pp. 256–264).

Lavine, M. (1991). Sensitivity in Bayesian statistics: the prior and the likelihood. *Journal of the American Statistical Association*, *86*(414), 396–399.

Levi, I. (2004). *Mild contraction: Evaluating loss of information due to loss of belief.* Oxford University Press.

Manski, C. F. (2003). *Partial identification of probability distributions.* Springer Science & Business Media.

Martin, R., & Liu, C. (2015). *Inferential Models: reasoning with uncertainty.* CRC Press, Boca Raton, FL.

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.