Imprecise Probabilities in Modern Data Science: Challenges and Opportunities

Ruobin Gong

Department of Statistics, Rutgers University, USA

SIPTA Seminars

June 29, 2022

IP holds great promise for data science

Uncertainty quantification needs more than probabilities.

Data science needs uncertainty quantification.

IP holds great promise for data science

Imprecise probabilities (IP) capture structural uncertainty intrinsic to probabilistic models.

From classic inference problems:

- Robust Bayes: epistemic uncertainty about the prior and the likelihood (Berger, 1990);
- Robust statistics (Huber & Strassen, 1973);
- Fiducial, structural, and functional inference: Dempster-Shafer theory (Dempster, 2008); inferential models (Martin & Liu, 2015);
- Econometrics: partial identification (Manski, 2003);

...to recent examples:

- Satellite conjunction (Balch et al., 2019) ;
- Electoral forecasting (Kreiss et al., 2021);
- Supervised learning, conformal prediction (Cella & Martin, 2021, 2022), E-values?
- Statistical disclosure limitation (SDL) and differential privacy (DP) - more on this later.

Making IP more popular: challenges

IP offers a principled approach to data-driven decision making.

It should be even more popular than it already is!

challenge 1

IP models can defy the intuition derived from precise probability models.

challenge 2

IP models may be difficult to compute.

A statistician's intuition *v.* IP

As an answer to the question "where is θ?", if the statistician reports an interval

$$[\underline{P}, \overline{P}],$$

what does the statistician mean by it?

If the interval represents...

- a frequentist confidence interval,
- a calibrated Bayesian posterior interval, or
- an inferential model (IM) predictive random set,

then the interval

1. is meant to have a **coverage probability** ideally at, and **not below**, the reported nominal level for every value of θ;
2. is usually implied to be **short**, and **sensibly constructed** (e.g. connected, bounded, etc) to the best of the statistician's ability.

These interpretations do not fare well when the interval is derived from a credal set.

A statistician's intuition *v.* IP

Upper and lower probabilities are bounds:

$$\underline{P}(A) = \inf_{P \in \Pi} P(A), \qquad \overline{P}(A) = \sup_{P \in \Pi} P(A).$$

Example. A survey with partial 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

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

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

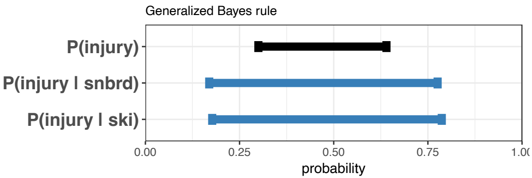
A statistician's intuition *v.* IP

Upper and lower conditional* probabilities are also bounds:

$$\underline{P}_{\mathfrak{B}}(A \mid B) \overset{\text{def}}{=\!=} \inf_{P \in \Pi} \frac{P(A,B)}{P(B)}, \qquad \overline{P}_{\mathfrak{B}}(A \mid B) \overset{\text{def}}{=\!=} \sup_{P \in \Pi} \frac{P(A,B)}{P(B)}.$$

Example. Looking at injury rates for skiers and snowboarders, respectively:

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



Upper and lower probabilities for injury **dilate**, regardless of the conditioning event.

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

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

The most widely used IP updating rule is the generalized $\mathfrak{B}$ayes rule as given before:

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

e.g. as employed in robust Bayesian inference.

A statistician's intuition *v.* IP

How do you feel about dilation?

It's great! | **It's weird!**

✓ Dilation is **not** a violation of Bayesian coherence.

However,

✗ Most statisticians find dilation troubling.

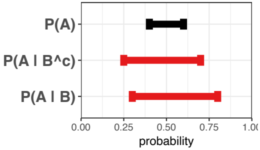
- It creates a loss of *precision* for apparently no good reason;
- Although, a few find it appealing: perhaps a way for Bayesian models to "un-learn"?

Judicious judgment and unsettling updating (Gong & Meng, 2021)

Gen. $\mathfrak{B}$ayes rule

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

Dilation



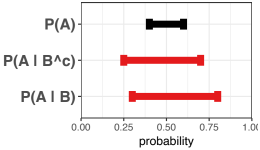
✗ Generalized $\mathfrak{B}$ayes rule may induce dilation.

Judicious judgment and unsettling updating (Gong & Meng, 2021)

Gen. $\mathfrak{B}$ayes rule

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

Dilation



✗ Generalized $\mathfrak{B}$ayes rule may induce dilation.

✗ Generalized $\mathfrak{B}$ayes rule cannot sharpen vacuous priors.

Judicious judgment and unsettling updating (Gong & Meng, 2021)

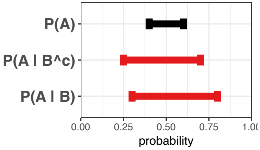
Gen. $\mathfrak{B}$ayes rule

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

$\mathfrak{D}$empster's rule

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

Dilation



- ✗ Generalized $\mathfrak{B}$ayes rule may induce dilation.
- ✗ Generalized $\mathfrak{B}$ayes rule cannot sharpen vacuous priors.
- ✓ $\mathfrak{D}$empster's rule dilates less often than generalized $\mathfrak{B}$ayes rule.

Judicious judgment and unsettling updating (Gong & Meng, 2021)

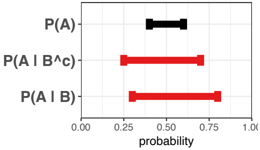
Gen. $\mathfrak{B}$ayes rule

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

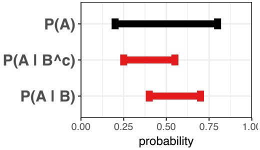
$\mathfrak{D}$empster's rule

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

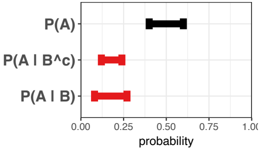
Dilation



Contraction



Sure Loss



- ✗ Generalized $\mathfrak{B}$ayes rule may induce dilation.
- ✗ Generalized $\mathfrak{B}$ayes rule cannot sharpen vacuous priors.
- ✓ $\mathfrak{D}$empster's rule dilates less often than generalized $\mathfrak{B}$ayes rule.
- ✗ Unlike generalized $\mathfrak{B}$ayes rule, $\mathfrak{D}$empster's rule may contract and induce sure loss.

Judicious judgment and unsettling updating (Gong & Meng, 2021)

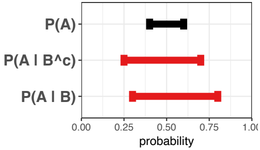
Gen. $\mathfrak{B}$ayes rule

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

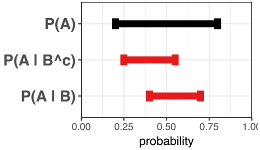
$\mathfrak{D}$empster's rule

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

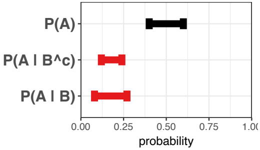
Dilation



Contraction



Sure Loss



From a **statistical** point of view:

- ✗ Generalized $\mathfrak{B}$ayes rule may induce dilation.
- ✗ Generalized $\mathfrak{B}$ayes rule cannot sharpen vacuous priors.
- ✓ $\mathfrak{D}$empster's rule dilates less often than generalized $\mathfrak{B}$ayes rule.
- ✗ Unlike generalized $\mathfrak{B}$ayes rule, $\mathfrak{D}$empster's rule may contract and induce sure loss.

Judicious judgment and unsettling updating (Gong & Meng, 2021)

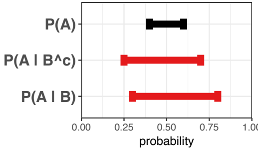
Gen. $\mathfrak{B}$ayes rule

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

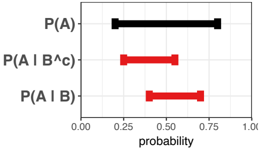
$\mathfrak{D}$empster's rule

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

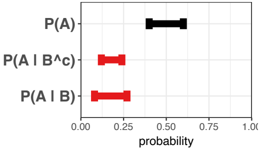
Dilation



Contraction



Sure Loss



From a ~~statistical~~ **coherence** point of view:

- ✓ Generalized $\mathfrak{B}$ayes rule may induce dilation.
- ✓ Generalized $\mathfrak{B}$ayes rule cannot sharpen vacuous priors.
- ✗ $\mathfrak{D}$empster's rule dilates less often than generalized $\mathfrak{B}$ayes rule.
- ✓ Unlike generalized $\mathfrak{B}$ayes rule, $\mathfrak{D}$empster's rule may contract and induce sure loss.

Making IP more popular: challenges

challenge 1

IP models can defy the intuition derived from precise probability models.

> *Trouble for imprecise probabilities rarely comes in the form of inherent contradictions, but instead is more apt to arise from seeking to preserve consistency at all costs.*
>
> *(Wheeler, 2021, p. 203)*

challenge 2

IP models may be difficult to compute.

Computation *v.* IP (or narrowly, DS theory of belief functions)

Dempster-Shafer theory has found many successful implementations in computer vision, signal processing, and artificial intelligence, where the problems have **discrete** state spaces.

In contrast, statistical applications of the DS theory are limited. Typical inference problems, whether parametric or nonparametric, have **continuous** state spaces.

Computation *v.* IP

Example. Prior-free Multinomial inference.
For $n \in [N]$, $x_n \overset{iid}{\sim}$ Categorical(θ) with $\theta = (\theta_k)_{k \in [K]}$. i.e.,

$$\mathbb{P}(x_n = k) = \theta_k, \quad \forall n, k.$$

We would like to make inference about θ and about future observations.

Computation *v.* IP

Example. Prior-free Multinomial inference.
For $n \in [N]$, $x_n \overset{iid}{\sim}$ Categorical(θ) with $\theta = (\theta_k)_{k \in [K]}$. i.e.,

$$\mathbb{P}(x_n = k) = \theta_k, \quad \forall n, k.$$

We would like to make inference about θ and about future observations.

IP solutions:

- Dempster (1966, 1972)
- Imprecise Dirichlet Model (IDM), Walley (1996):

$$\underline{P}\left(x_{n+1} = j \mid \boldsymbol{x}\right) = \frac{n_j}{n+s}, \qquad \overline{P}\left(x_{n+1} = j \mid \boldsymbol{x}\right) = \frac{n_j + s}{n+s},$$

  for some s fixed representing the number of "hidden" observations.

- Dirichlet-DSM (Lawrence et al., 2009)

Example. Prior-free Multinomial inference.
For $n \in [N]$, $x_n \overset{iid}{\sim}$ Categorical(θ) with $\theta = (\theta_k)_{k \in [K]}$. i.e.,

$$\mathbb{P}(x_n = k) = \theta_k, \quad \forall n, k.$$

We would like to make inference about θ and about future observations.

IP solutions:

- Dempster (1966, 1972)
- Imprecise Dirichlet Model (IDM), Walley (1996):

$$\underline{P}\left(x_{n+1} = j \mid \boldsymbol{x}\right) = \frac{n_j}{n+s}, \qquad \overline{P}\left(x_{n+1} = j \mid \boldsymbol{x}\right) = \frac{n_j + s}{n+s},$$

  for some s fixed representing the number of “hidden” observations.
- Dirichlet-DSM (Lawrence et al., 2009)

Example. Prior-free Multinomial inference.
For $n \in [N]$, $x_n \overset{iid}{\sim}$ Categorical(θ) with $\theta = (\theta_k)_{k\in[K]}$. i.e.,

$$\mathbb{P}(x_n = k) = \theta_k, \quad \forall n, k.$$

We would like to make inference about θ and about future observations.

IP solutions:

- Dempster (1966, 1972)
- Imprecise Dirichlet Model (IDM), Walley (1996):

  $$\underline{P}\,(x_{n+1} = j \mid \boldsymbol{x}) = \frac{n_j}{n+s}, \qquad \overline{P}\,(x_{n+1} = j \mid \boldsymbol{x}) = \frac{n_j + s}{n+s},$$

  for some s fixed representing the number of “hidden” observations.
- Dirichlet-DSM (Lawrence et al., 2009)

Computation *v.* IP

Example. Prior-free Multinomial inference.
For $n \in [N]$, $x_n \stackrel{iid}{\sim}$ Categorical(θ) with $\theta = (\theta_k)_{k\in[K]}$. i.e.,

$$\mathbb{P}(x_n = k) = \theta_k, \quad \forall n, k.$$

We would like to make inference about θ and about future observations.

IP solutions:

- Dempster (1966, 1972)
- Imprecise Dirichlet Model (IDM), Walley (1996): - analytical

$$\underline{P}\left(x_{n+1} = j \mid \boldsymbol{x}\right) = \frac{n_j}{n+s}, \qquad \overline{P}\left(x_{n+1} = j \mid \boldsymbol{x}\right) = \frac{n_j + s}{n+s},$$

  for some s fixed representing the number of "hidden" observations.
- Dirichlet-DSM (Lawrence et al., 2009) - analytical

Example. Prior-free Multinomial inference.
For $n \in [N]$, $x_n \overset{iid}{\sim}$ Categorical(θ) with $\theta = (\theta_k)_{k \in [K]}$. i.e.,

$$\mathbb{P}(x_n = k) = \theta_k, \quad \forall n, k.$$

We would like to make inference about θ and about future observations.

IP solutions:

- Dempster (1966, 1972) - a Gibbs sampler is devised 50 years later (Jacob et al., 2021)
- Imprecise Dirichlet Model (IDM), Walley (1996): - analytical

$$\underline{P}\,(x_{n+1} = j \mid \boldsymbol{x}) = \frac{n_j}{n+s}, \qquad \overline{P}\,(x_{n+1} = j \mid \boldsymbol{x}) = \frac{n_j + s}{n+s},$$

  for some s fixed representing the number of "hidden" observations.
- Dirichlet-DSM (Lawrence et al., 2009) - analytical

A Gibbs sampler for a class of random convex polytopes

Pierre E. Jacob[1], Ruobin Gong[2], Paul T. Edlefsen[3], Arthur P. Dempster[1]
[1]Harvard University, USA [2]Rutgers University, USA [2]Fred Hutchinson Cancer Research Center, USA

Summary

We present a Gibbs sampler for the Dempster-Shafer (DS) statistical inference for categorical distributions. The algorithm targets a class of random convex polytopes [2] corresponding to the *structure of the second kind* model of [1]. Central to the sampler is an equivalence between the iterative **vertex configuration constraints** and the **non-negativity of cycles** in a fully connected directed graph. The sampler demonstrates good convergence for reasonable category and sample sizes.

Notation

The observations are $\mathbf{x} = (x_n)_{n\in[N]}$ with $x_n \in [K]$ (categories), with which we infer the parameters $\theta = (\theta_1, \ldots, \theta_K) \in \Delta$ (**the K-simplex**) of a categorical distribution:

$$\mathbb{P}(x_n = k) = \theta_k.$$

A **polytope** is a set of points $z \in \mathbb{R}^K$ satisfying linear inequalities of the form $Mz \le c$. For a given $\mathbf{x} \in [K]^N$, $\mathcal{I}_k$ is the set of indices $\{n \in [N] \colon x_n = k\}$. The counts are $N_k = |\mathcal{I}_k|$ and $\sum_{k\in[K]} N_k = N$. A **uniform** variable Z over set $\mathcal{S}$ is written $Z \sim \mathcal{S}$.

Sampling Mechanism and Feasible Set

Let $(\Delta_k(\theta))_{k\in[K]}$ be the **subsimplices** that partition Δ by θ, with $\mathrm{Vol}(\Delta_k(\theta)) = \theta_k$. Define

$$x_n = \sum_{k\in[K]} k \mathbb{1}(u_n \in \Delta_k(\theta)),$$

where $u_n \sim \Delta$. This sampling mechanism is invariant to permutation of category labels. It is also equivalent to the *Gumbel-max trick*.

Lemma 5.2 of [1]. For $k \in [K]$, $\theta \in \Delta$ and $u_n \in \Delta$, $u_n \in \Delta_k(\theta)$ if and only if $u_{n,\ell}/u_{n,k} \ge \theta_\ell/\theta_k$ for all $\ell \in [K]$.

The set of all possible realizations of $\mathbf{u}$ which could have produced the data $\mathbf{x}$ for some θ is

$$\mathcal{R}_{\mathbf{x}} = \{(u_1, \ldots, u_N) \in \Delta^N \colon \exists \theta \in \Delta \;\; \forall n \in [N] \;\; u_n \in \Delta_{x_n}(\theta)\},$$

and given a realization of $\mathbf{u} \in \mathcal{R}_{\mathbf{x}}$, by definition the non-empty **feasible set** for θ is

$$\mathcal{F}(\mathbf{u}) = \{\theta \in \Delta \colon \forall n \in [N] \;\; u_n \in \Delta_{x_n}(\theta)\},$$

so that $\mathcal{R}_{\mathbf{x}} = \{\mathbf{u} \colon \mathcal{F}(\mathbf{u}) \ne \emptyset\}$.

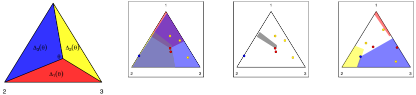


Figure 1. Left: partition of Δ into $(\Delta_k(\theta))_{k\in[K]}$ with $K = 3$. Each point $u_n \in \Delta$ defines, for a fixed $x_n \in [K]$, a set of $\theta \in \Delta$ such that $u_n \in \Delta_{x_n}(\theta)$. With $(x_1, x_2, x_3) = (1, 3, 2)$, there may (mid left & mid right) – or may not (right) – exist a $\theta \in \Delta$ such that $u_n \in \Delta_{x_n}(\theta)$ for $n = 1, 2, 3$.

References

[1] Dempster, A. P. (1966), 'New methods for reasoning towards posterior distributions based on sample data', *The Annals of Mathematical Statistics*, 37(2), 355-374.
[2] Dempster, A. P. (1972), 'A class of random convex polytopes', *The Annals of Mathematical Statistics* pp. 260-272.
[3] Jacob, P. E., Gong, R., Edlefsen, P. T., and Dempster, A. P. (2021). A Gibbs sampler for a class of random convex polytopes (with discussion). *Journal of the American Statistical Association*, doi:10.1080/01621459.2021.1881523

The Gibbs Sampler

Our main contribution is the Markov chain Monte Carlo (MCMC) algorithm targeting the uniform distribution on $\mathcal{R}_{\mathbf{x}}$:

$$\nu_{\mathbf{x}}(u_1, \ldots, u_N) = \mathrm{Vol}\left(\mathcal{R}_{\mathbf{x}}\right)^{-1} \mathbb{1}\left((u_1, \ldots, u_N) \in \mathcal{R}_{\mathbf{x}}\right).$$

The sets $\mathcal{F}(\mathbf{u})$ obtained when $\mathbf{u} \sim \nu_{\mathbf{x}}$ constitute the class of random convex polytopes studied in [2], and the result of *Dempster's rule of combination* on the N observations.

Define $\forall k \in [K]$ and $\forall \ell \in [K]$,

$$\eta_{k\to\ell}(\mathbf{u}) = \min_{n\in\mathcal{I}_k} \frac{u_{n,\ell}}{u_{n,k}},$$

which depend on the observations through $(\mathcal{I}_k)$. We have that $\theta \in \mathcal{F}(\mathbf{u})$ is *equivalent* to $\theta_\ell/\theta_k \le \eta_{k\to\ell}(\mathbf{u})$ for $\ell, k \in [K]$.

Proposition 3.1. There exists $\theta \in \Delta$ satisfying $\theta_\ell/\theta_k \le \eta_{k\to\ell}$ for all $k, \ell \in [K]$ if and only if the values $(\eta_{k\to\ell})$ satisfy

$$\forall L \in [K] \quad \forall j_1, \ldots, j_L \in [K], \quad \eta_{j_1\to j_2}\eta_{j_2\to j_3} \ldots \eta_{j_L\to j_1} \ge 1.$$

It suffices to restrict the above inequalities to distinct indices $j_1, \ldots, j_L$.

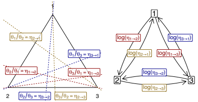


Figure 2. Two views on the constraints. Left: the values $\eta_{k\to\ell}$ define linear constraints $\theta_\ell/\theta_k = \eta_{k\to\ell}$. Right: the log values are weights on the edges of a complete directed graph.

Proposition 3.2. Let $\mathbf{u} \in \mathcal{R}_{\mathbf{x}}$, and define $\eta_{k\to\ell}$ as before. Let $k \in [K]$. Define for $\ell \in [K]$,

$$\theta_\ell = \frac{\exp(-\min(\ell \to k))}{\sum_{\ell'\in[K]} \exp(-\min(\ell' \to k))}$$

where $\min(\ell \to k)$ is the **minimum value** over **all paths** from ℓ to k, in a fully connected directed graph with **weight** $\log \eta_{j\to\ell}$ on edge (j, ℓ). Then the conditional distribution of $\mathbf{u}_{\mathcal{I}_k}$ under $\nu_{\mathbf{x}}$,

$$\nu_{\mathbf{x}}(d\mathbf{u}_{\mathcal{I}_k} | \mathbf{u}_{[N]\setminus\mathcal{I}_k}),$$

is indeed the uniform distribution on $\Delta_k(\theta)^{N_k}$.

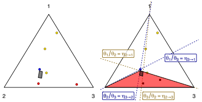


Figure 3. Given $\mathbf{u} \in \mathcal{R}_{\mathbf{x}}$ (left), the sampler drops components $\mathbf{u}_{\mathcal{I}_k}$ for some $k \in [K]$ (red dots) and draws new $\mathbf{u}_{\mathcal{I}_k}$ (red squares in right) from the above conditional distribution, with support being the shaded triangle.

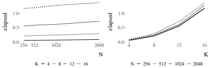


Figure 4. Elapsed time (s) for 100 iterations, as a function of N for different K (left) and as a function of K for different N (right).

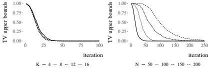


Figure 5. Upper bounds on total variation distance between $\mathbf{u}^{(t)}$ and $\nu_{\mathbf{x}}$ against t, varying K with 10 counts per category (left), and varying N with $N/(K = 5)$ counts per category (right).

Making IP more popular: challenges and opportunities

challenge 1

IP models can defy the intuition derived from precise probability models.

challenge 2

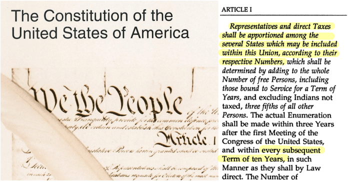
IP models may be difficult to compute.

opportunity

Statistical disclosure limitation (SDL) and differential privacy (DP).

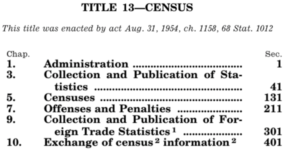
Modern data curators seek to meet two goals at once:

1. To **disclose** key statistics/use cases of the database, in accordance with its legal, policy, and/or ethical mandates.
2. To protect the **privacy** of individuals with trust-worthy guarantees.


The Constitution of the United States of America

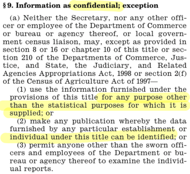
ARTICLE I

Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, *three fifths of all other Persons.* The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct. The Number of


TITLE 13—CENSUS

This title was enacted by act Aug. 31, 1954, ch. 1158, 68 Stat. 1012

Chap.		Sec.
1.	**Administration**	1
3.	**Collection and Publication of Statistics**	41
5.	**Censuses**	131
7.	**Offenses and Penalties**	211
9.	**Collection and Publication of Foreign Trade Statistics** [1]	301
10.	**Exchange of census [2] information** [2]	401


§ 9. Information as confidential; exception

(a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison, may, except as provided in section 8 or 16 or chapter 10 of this title or section 210 of the Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations Act, 1998 or section 2(f) of the Census of Agriculture Act of 1997—

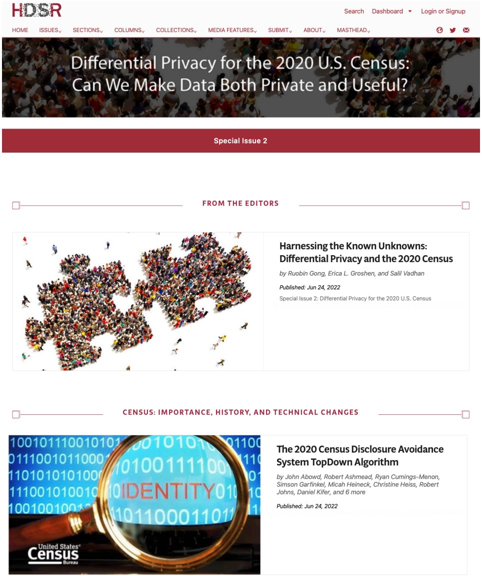
(1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or

(2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or

(3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports.

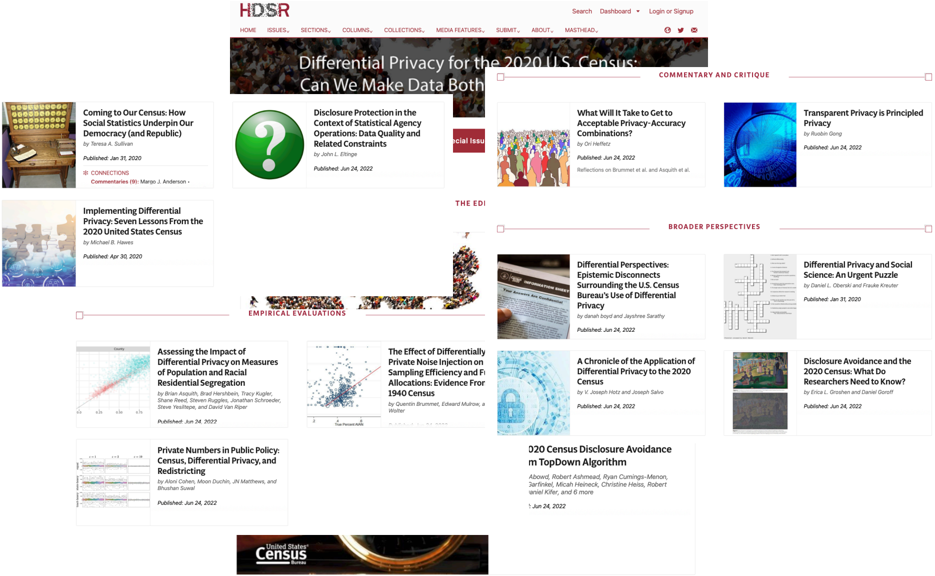
For example, the U.S. Census Bureau bears the constitutional mandate to enumerate the population every 10 years for apportionment. It is also bound by Title 13 of U.S. Code to protect respondent confidentiality.

The U.S. Census Bureau adopts differential privacy



Harvard Data Science Review (https://hdsr.mitpress.mit.edu)

The U.S. Census Bureau adopts differential privacy



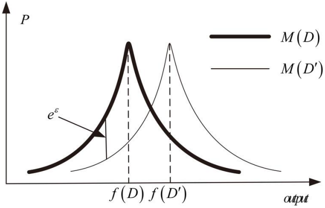
Harvard Data Science Review (https://hdsr.mitpress.mit.edu)

The mechanism of differential privacy

We wish to learn about aggregate features of a database, while protecting the privacy of the individual respondents.

A randomized function $T(\boldsymbol{x}; \boldsymbol{r})$ is said to be ϵ-differentially private (Dwork et al., 2006) if for all *neighboring* databases $(\boldsymbol{x}, \boldsymbol{x}')$ and all measurable S,

$$\frac{\Pr(T(\boldsymbol{x}; \boldsymbol{r}) \in S \mid \boldsymbol{x})}{\Pr(T(\boldsymbol{x}'; \boldsymbol{r}) \in S \mid \boldsymbol{x}')} \leq \exp(\epsilon).$$



https://www.ons.gov.uk/peoplepopulationandcommunity

That is, differentially private mechanisms conceal the confidential database x, by infusing crafted noise r into the data product T for release:

$$\boldsymbol{x} \longrightarrow T(\boldsymbol{x}; \boldsymbol{r})$$

The privacy-utility tradeoff

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

Differential privacy: how can IP help

- Supply a **rigorous** vocabulary for data and inferential imprecision under privacy constraints;

  e.g. Imprecise Dirichlet Model (IDM) again: lower and upper probabilities

  $$\underline{P}\left(x_{n+1} = j \mid \boldsymbol{x}\right) = \frac{n_j}{n+1}, \qquad \overline{P}\left(x_{n+1} = j \mid \boldsymbol{x}\right) = \frac{n_j + 1}{n+1},$$

  accord to posterior expectations based on two hypothetical databases that differ by $s = 1$;

  e.g. Interval of Measures

Differential privacy: how can IP help

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.

Differential privacy: how can IP help

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 $S \in \mathscr{F}$,

$$\frac{\Pr(T(x) \in S)}{\Pr(T(x') \in S)} \leq \exp(\epsilon).$$

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 $d_H(x, x') = 1$,

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

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

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

Differential privacy: how can IP help

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

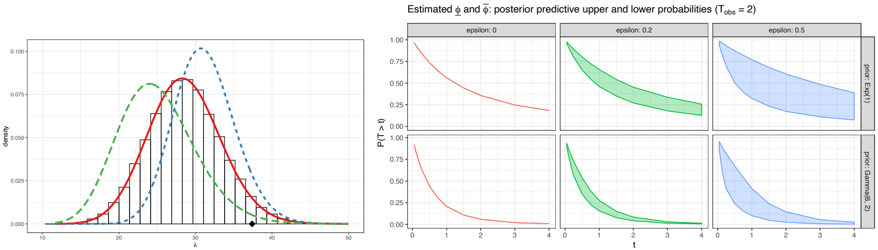


Figure: Left: (Gong, 2019) a privatized query ($\epsilon < \infty$) is statistically less informative than a non-privatized one ($\epsilon = \infty$); Right: Smaller ϵ induces narrower posterior predictive IoM over neighboring datasets, delivering more privacy.

Differential privacy: how can IP help

- Supply a **rigorous** vocabulary for data and inferential imprecision under privacy constraints;
  - e.g. Imprecise Dirichlet Model (IDM) again: lower and upper probabilities

    $$\underline{P}\left(x_{n+1}=j \mid \boldsymbol{x}\right)=\frac{n_j}{n+1}, \qquad \overline{P}\left(x_{n+1}=j \mid \boldsymbol{x}\right)=\frac{n_j+1}{n+1},$$

    accord to posterior expectations based on two hypothetical databases that differ by $s=1$;
  - e.g. Interval of Measures
- Help draw principled inferential conclusions from privatized data, particularly when the privacy mechanism is **not transparent** (cf. Gong, 2022);
  - e.g. This may happen if the privatization scheme is complex and/or algorithmically defined;
- Characterize **non-** and **partial-identification** incurred by special cases of the privacy mechanism and privacy loss budgeting policy.
  - e.g. Komarova & Nekipelov (2020)

Summary

IP holds great promise in data science. It should be more popular than it already is.

- Reconcile with existing intuition derived from precise probability models;
- Develop accessible computation for moderately sized applications;
- Explore (and exploit!) the role of IP in statistical privacy and differential privacy.

Thank you

ruobin.gong@rutgers.edu · Research supported by the National Science Foundation

Bibliography

Balch, M. S., Martin, R., & Ferson, S. (2019). Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society A*, *475*(2227), 20180565.

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

Cella, L., & Martin, R. (2021). Valid inferential models for prediction in supervised learning problems. In *International symposium on imprecise probability: Theories and applications* (pp. 72–82).

Cella, L., & Martin, R. (2022). Validity, consonant plausibility measures, and conformal prediction. *International Journal of Approximate Reasoning*, *141*, 110–130.

Dempster, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *The Annals of Mathematical Statistics*, *37*(2), 355–374.

Dempster, A. P. (1972). A class of random convex polytopes. *The Annals of Mathematical Statistics*, *43*(1), 260–272.

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).

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

Gong, R. (2022). Transparent privacy is principled privacy. *Harvard Data Science Review (Special Issue 2)*.

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.

Jacob, P. E., Gong, R., Edlefsen, P. T., & Dempster, A. P. (2021). A Gibbs sampler for a class of random convex polytopes (with discussion). *Journal of the American Statistical Association*, *116*(535), 1181-1192.

Komarova, T., & Nekipelov, D. (2020). Identification and formal privacy guarantees. *arXiv preprint arXiv:2006.14732*.

Kreiss, D., Schollmeyer, G., & Augustin, T. (2021). Towards improving electoral forecasting by including undecided voters and interval-valued prior knowledge. In *International symposium on imprecise probability: Theories and applications* (pp. 201–209).

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

Lawrence, E. C., Vander Wiel, S., Liu, C., & Zhang, J. (2009). *A new method for multinomial inference using Dempster-Shafer theory* (Tech. Rep.). Los Alamos National Laboratory.

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.

Walley, P. (1996). Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society: Series B (Methodological)*, 3–57.

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

Wheeler, G. (2021). Comment: Moving beyond sets of probabilities. *Statistical Science*, *36*(2), 201–204.