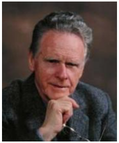
A Gibbs sampler for
a class of random convex polytopes

Ruobin Gong (Rutgers University)

BFF 6.5 Virtual Workshop
Feb 5, 2021



Joint work with Pierre Jacob · Paul Edlefsen · Art Dempster

	Distributional Inference $x \mid \theta \sim P_\theta$	**Functional Inference** $x = G(u, \theta),\ u \sim P_0$
Precise Probability	classic Likelihood and Bayesian inference	Structural inference (Fraser, 1968) Functional models (Dawid & Stone, 1982) Generalized Fiducial inference (Hannig et al., 2016)
Imprecise Probability $(\underline{P}, \overline{P})$	Robust Bayes (Berger, 1994) Robust statistics (Huber & Ronchetti, 2009)	**Dempster-Shafer theory** (Dempster, 1968; Shafer, 1976; Dempster, 2008) Inferential models (Martin & Liu, 2015)

Dempster-Shafer theory

x θ

Dempster-Shafer theory

$$x = G(u, \theta)$$

Dempster-Shafer theory

$$\begin{aligned} x &= G\left(u,\theta\right) \\ &\downarrow \\ \theta &= G^{-1}\left(u,x\right) \end{aligned}$$

The inversion may yield a set-valued mapping: $G^{-1}(\cdot,x) : \Delta \to 2^{\Theta}$.

Dempster's Rule of Combination

$$\begin{aligned} x_1 &= G_1(\theta, u_1) \\ &\vdots \\ x_N &= G_N(\theta, u_N) \end{aligned} \quad \longrightarrow \quad \begin{cases} \theta \in G_1^{-1}(x_1, u_1) \\ \vdots \\ \theta \in G_N^{-1}(x_N, u_N) \end{cases}$$

The **random set** that characterizes post-data inference for θ is

$$\mathcal{F}(\mathbf{u}) = \{\theta \in \Theta : \ \forall n \in [N], \ x_n = G_n(\theta, u_n)\},$$

where $\mathbf{u} \sim \nu_{\mathbf{x}}$, the uniform distr. on the $\mathbf{x}$-dependent feasible subset

$$\mathcal{R}_{\mathbf{x}} = \left\{(u_1, \ldots, u_N) \in [0,1]^N : \exists\theta \in \Theta \ \forall i \in [N] \ x_n = G_n^{-1}(u_n, \theta)\right\}.$$

Challenge: How to efficiently sample $\mathcal{F}(\mathbf{u})$?

Categorical distribution inference

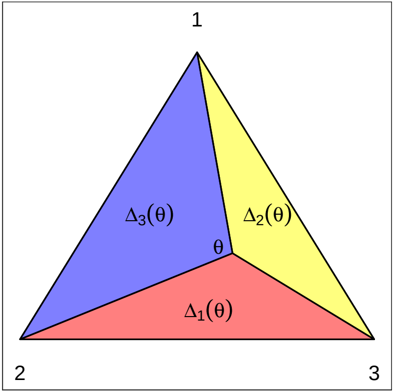
For $n \in [N]$, $x_n \overset{iid}{\sim}$ Categorical(θ) with $\theta = (\theta_k)_{k \in [K]}$. i.e.,

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

Categorical distribution inference

For $n \in [N]$, $x_n \overset{iid}{\sim} \text{Categorical}(\theta)$ with $\theta = (\theta_k)_{k \in [K]}$. i.e.,

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



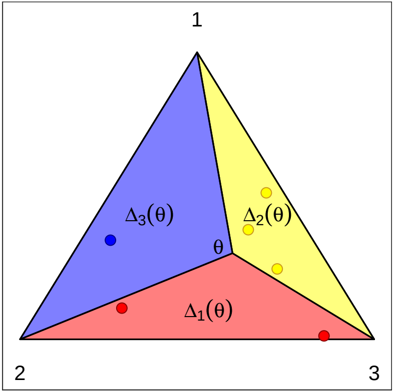
Subsimplex $\Delta_k(\theta)$, for $\theta \in \Delta$:

$$\{z \in \Delta : \forall \ell \in [K] \; z_\ell / z_k \geq \theta_\ell / \theta_k\}.$$

Categorical distribution inference

For $n \in [N]$, $x_n \overset{iid}{\sim} \text{Categorical}(\theta)$ with $\theta = (\theta_k)_{k \in [K]}$. i.e.,

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



Subsimplex $\Delta_k(\theta)$, for $\theta \in \Delta$:

$$\{z \in \Delta : \forall \ell \in [K]\ z_\ell / z_k \geq \theta_\ell / \theta_k\}.$$

Sampling mechanism, for $\theta \in \Delta$:

- draw u_n uniform on Δ_n,

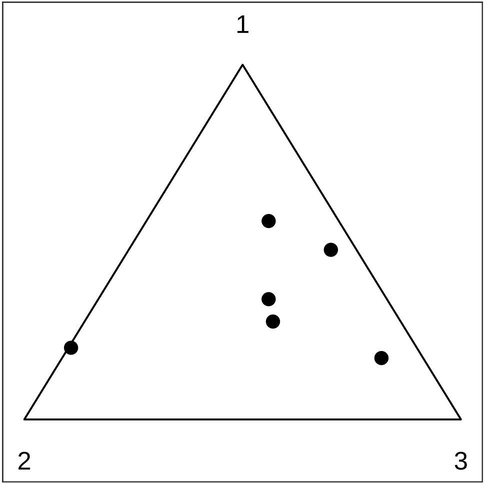
- define x_n such that $u_n \in \Delta_{x_n}(\theta)$,

i.e., $x_n = G(\theta, u_n)$.

Then, $\mathbb{P}(x_n = k) = \text{Vol}(\Delta_k(\theta)) = \theta_k$.

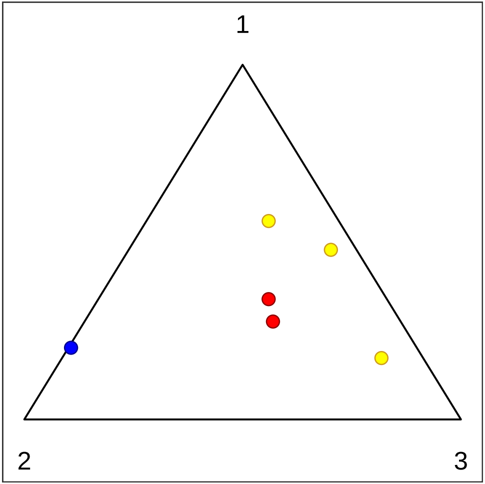
Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

Counts: $(2, 3, 1)$. Let's draw $N = 6$ uniform samples on Δ.



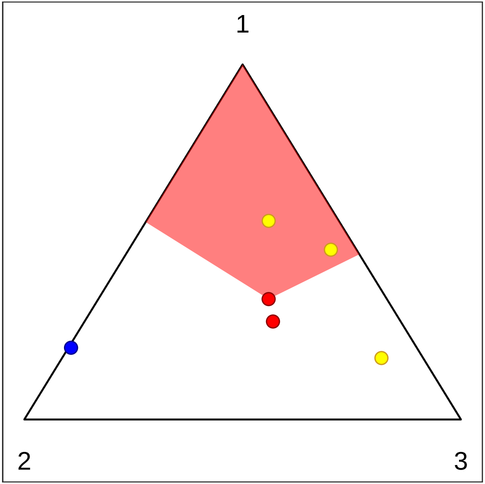
Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

Each u_n is associated to an observed $x_n \in \{1, 2, 3\}$.



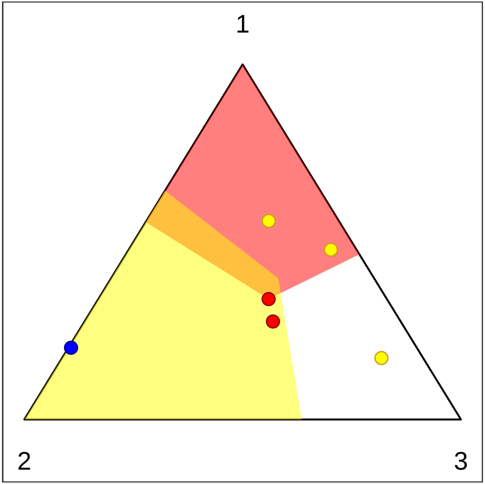
Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

If there exists a feasible θ, it cannot be just anywhere.



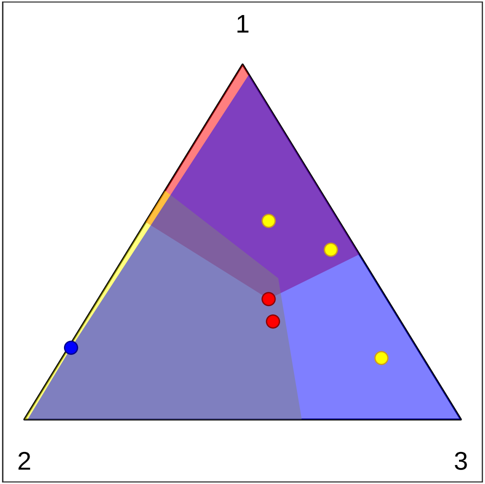
Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

Each category's samples add constraints on θ.



Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

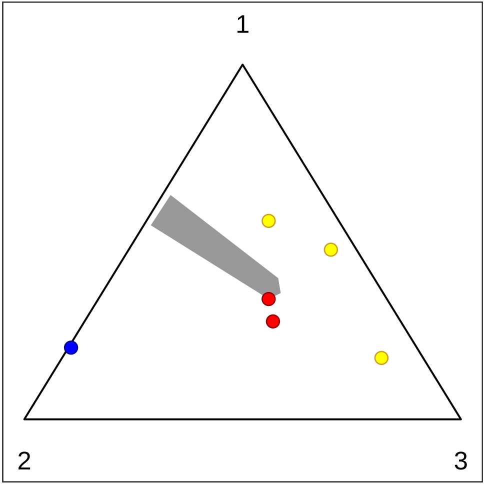
All constraints either define a polytope for θ, or an empty set.



Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

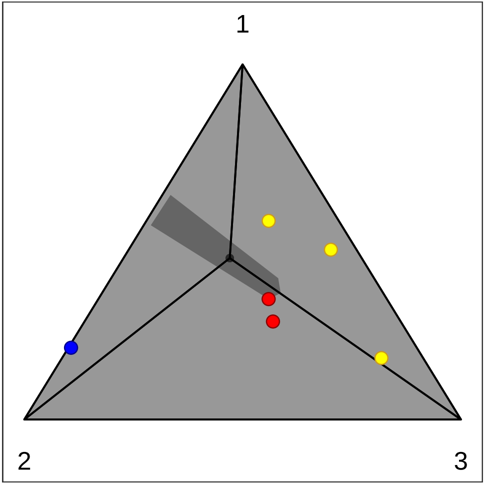
This polytope contains θ values s.t. $\forall n \in [N],\ x_n = G(u_n, \theta)$.

It is one realization of $\mathcal{F}(\mathbf{u})$.



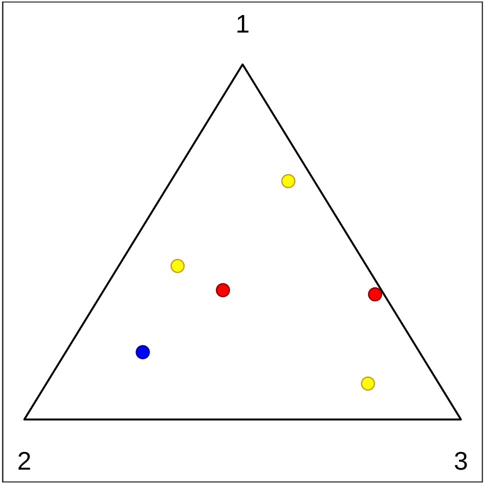
Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

Any θ in this polytope separates the samples appropriately.



Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

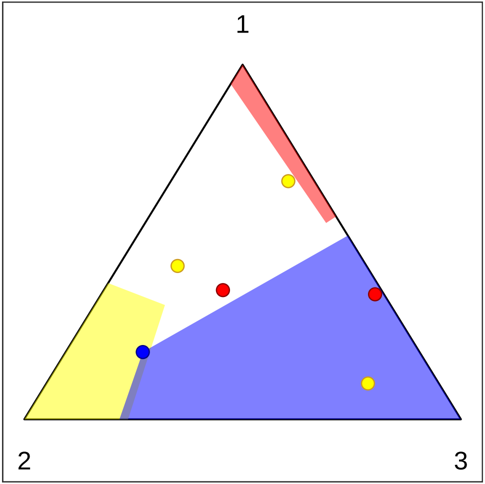
Let's try again with fresh uniform samples on Δ.



Post-data inference: drawing $\mathcal{F}(\mathbf{u})$

Here, there is no $\theta \in \Delta$ such that $\forall n \in [N],\ x_n = G(u_n, \theta)$.

This draw does not constitute a realization of $\mathcal{F}(\mathbf{u})$.

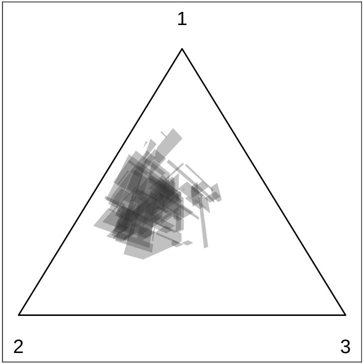


Proposal: Gibbs sampler

The idea is to start from an arbitrary $\mathbf{u} \in \mathcal{R}_{\mathbf{x}}$, and iteratively sample some component (u_n) given others.

To this end, we need a characterization of $\mathcal{R}_{\mathbf{x}}$ in terms of $\mathbf{u}$.

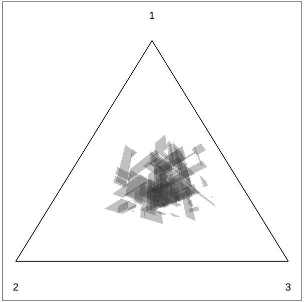
100 polytopes drawn from the proposed sampler for counts $(9, 8, 3)$:



Monte Carlo post-data inference

For a set $B \subset \Delta$ of interest, lower and upper probabilities

$$
\begin{aligned}
\underline{P}(B \mid \mathbf{x}) &= \int \mathbb{1}\left(\mathcal{F}(\mathbf{u}) \subseteq B\right) \nu_{\mathbf{x}}(d\mathbf{u}) \\
&\doteq \frac{1}{T} \sum_{t \in [T]} \mathbb{1}\left(\mathcal{F}\left(\mathbf{u}^{(t)}\right) \subseteq B\right), \\
\overline{P}(B \mid \mathbf{x}) &= \int \mathbb{1}\left(\mathcal{F}(u) \cap B \neq \emptyset\right) \nu_{\mathbf{x}}(d\mathbf{u}) \\
&\doteq \frac{1}{T} \sum_{t \in [T]} \mathbb{1}\left(\mathcal{F}\left(\mathbf{u}^{(t)}\right) \cap B \neq \emptyset\right).
\end{aligned}
$$



Example. Counts: $(7, 5, 8)$, $\hat{\theta}_1 = 0.35$,

$$
\begin{aligned}
\hat{\underline{P}}(0.3 \leq \theta_1 \leq 0.4 \mid \mathbf{x}) &= 0.2, \\
\hat{\overline{P}}(0.3 \leq \theta_1 \leq 0.4 \mid \mathbf{x}) &= 0.62.
\end{aligned}
$$

Gibbs sampler: characterizing $\mathcal{R}_{\mathbf{x}}$

(*Sampling model*) $\quad \mathbf{u} \in \Delta_k(\theta) \quad \Leftrightarrow \quad \theta_\ell/\theta_k \leq u_\ell/u_k, \ \forall \ell \in [K].$

Denote

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

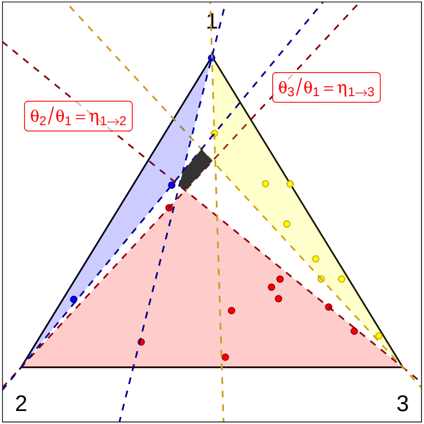
and define $\eta_{k\to k} = 1$ for all $k \in [K]$. We can write

$$\mathcal{R}_{\mathbf{x}} = \left\{ \mathbf{u} \in \Delta^N : \ \exists \theta \in \Delta \quad \forall k, \ell \in [K] \quad \theta_\ell/\theta_k \leq \eta_{k\to\ell} \right\}.$$

Gibbs sampler: characterizing $\mathcal{R}_{\mathbf{x}}$

Counts: $(9, 8, 3)$, $\mathbf{u}$ in $\mathcal{R}_{\mathbf{x}}$.

Values $\eta_{k\to\ell} = \min_{n\in\mathcal{I}_k} u_{n,\ell}/u_{n,k}$ define linear constraints on θ.



Gibbs sampler: characterizing $\mathcal{R}_{\mathbf{x}}$

What are the implications of $\mathbf{u} \in \mathcal{R}_{\mathbf{x}}$?

- There exists $\theta \in \Delta$ such that $\theta_\ell/\theta_k \leq \eta_{k\to\ell}$ for all $k, \ell \in [K]$.

- Then, for all k, ℓ,

$$\frac{\theta_\ell}{\theta_k} \leq \eta_{k\to\ell}, \quad \text{and} \quad \frac{\theta_k}{\theta_\ell} \leq \eta_{\ell\to k}, \quad \text{thus} \quad \eta_{k\to\ell}\eta_{\ell\to k} \geq 1.$$

- If $K \geq 3$: for all k, ℓ, j,

$$\eta_{\ell\to k}^{-1} \leq \frac{\theta_\ell}{\theta_k} = \frac{\theta_\ell}{\theta_j}\frac{\theta_j}{\theta_k} \leq \eta_{j\to\ell}\eta_{k\to j}, \quad \text{thus} \quad \eta_{k\to j}\eta_{j\to\ell}\eta_{\ell\to k} \geq 1.$$

- If $K \geq 4, 5, 6, \ldots$

Main result

if there exists $\theta \in \Delta$ such that $\theta_\ell / \theta_k \leq \eta_{k \to \ell}$ for $k, \ell \in [K]$ then

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

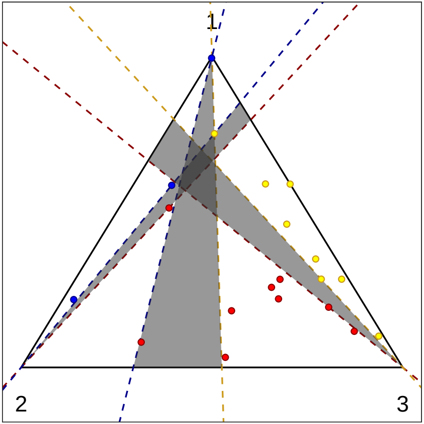
Claim: the reverse implication holds too. This would mean

$$\begin{aligned} \mathcal{R}_{\mathbf{x}} &= \{\mathbf{u} : \exists \theta \quad \forall k, \ell \in [K] \quad \theta_\ell / \theta_k \leq \eta_{k \to \ell}\} \\ &= \{\mathbf{u} : \forall L \in [K] \quad \forall j_1, \ldots, j_L \in [K] \quad \eta_{j_1 \to j_2} \eta_{j_2 \to j_3} \ldots \eta_{j_L \to j_1} \geq 1\}. \end{aligned}$$

i.e. $\mathcal{R}_{\mathbf{x}}$ is represented by relations among components (u_n).
From here we can work out conditional distributions under $\nu_{\mathbf{x}}$,
leading to a Gibbs sampler.

Gibbs sampler: characterizing $\mathcal{R}_\mathbf{x}$

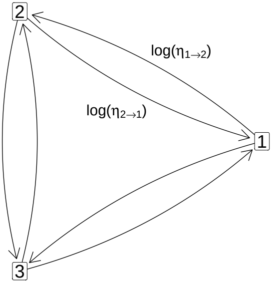
Counts: $(9, 8, 3)$. $\mathbf{u}$ is in $\mathcal{R}_\mathbf{x}$. Inequalities $\eta_{k\to\ell}\eta_{\ell\to k} \geq 1$ are shaded. Inequalities $\eta_{k\to j}\eta_{j\to\ell}\eta_{\ell\to k} \geq 1$ reflected by the common intersection.



An interesting connection to graphs

Consider a fully connected graph with K vertices,

and with weight $\log \eta_{k \to \ell}$ on edge (k, ℓ).



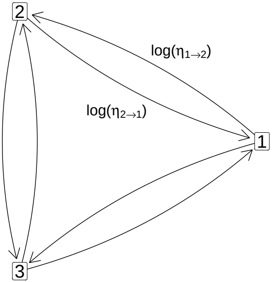
value of a path $=$ sum of the weights of edges in the path

cycle $=$ path from a vertex to itself

An interesting connection to graphs

Consider a fully connected graph with K vertices,

and with weight $\log \eta_{k \to \ell}$ on edge (k, ℓ).



value of a path $=$ sum of the weights of edges in the path

negative cycle $=$ path from a vertex to itself with negative value

An interesting connection to graphs

$$\mathcal{R}_{\mathbf{x}} = \{\mathbf{u} : \forall L \in [K] \quad \forall j_1, \ldots, j_L \in [K] \quad \eta_{j_1 \to j_2} \eta_{j_2 \to j_3} \ldots \eta_{j_L \to j_1} \geq 1\}.$$

In this fully connected graph,

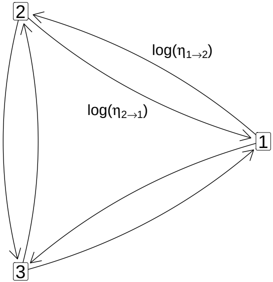
the ordered indices $j_1 \to j_2 \to \cdots \to j_L \to j_1$ form a **cycle**.

Thus, $\forall L \in [K] \quad \forall j_1, \ldots, j_L \in [K]$

$$\eta_{j_1 \to j_2} \ldots \eta_{j_L \to j_1} \geq 1$$

$$\Leftrightarrow \log(\eta_{j_1 \to j_2}) + \ldots + \log(\eta_{j_L \to j_1}) \geq 0$$

$\Leftrightarrow$ there are **no negative cycles** in the graph.



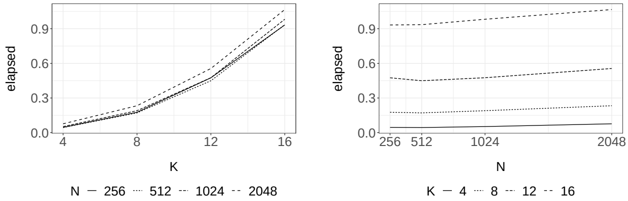
Gibbs sampler

- Initialize by obtaining $\mathbf{u}^{(0)} \in \mathcal{R}_{\mathbf{x}}$.
- At each time $t \geq 1$, for each category $k \in [K]$,
  1. compute $\theta^\star$ such that, for $n \in \mathcal{I}_k$,
     u_n given other components is uniform on $\Delta_k(\theta^\star)$.
  2. Draw $u_n^{(t)} \sim \Delta_k(\theta^\star)$ for $n \in \mathcal{I}_k$.
  3. Update $\eta_{k\to\ell}^{(t)} \leftarrow \min_{n\in\mathcal{I}_k} u_{n,\ell}^{(t)}/u_{n,k}^{(t)}$ for $\ell \in [K]$.

In step 1, $\theta^\star$ is obtained by computing the shortest path in a graph with weights $\eta_{k\to\ell}^{(t)}$ on edge (k,ℓ); e.g. Berkelaar et al. (2004); Csardi & Nepusz (2006).

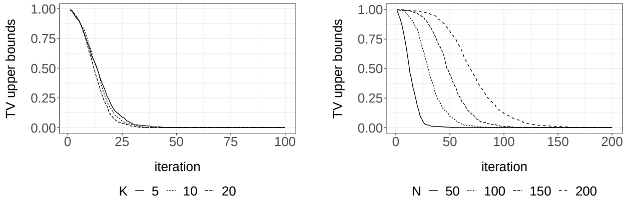
Performance

Cost in seconds for 100 full sweeps.



Let $\nu^{(t)}$ by the distribution of $\mathbf{u}^{(t)}$ after t iterations.

$$\text{TV}(\nu^{(t)}, \nu_{\mathbf{x}}) = \sup_A |\nu^{(t)}(A) - \nu_{\mathbf{x}}(A)|.$$



Summary

We proposed a Gibbs sampler for random sets encapsulating post-data Dempster-Shafer inference for Categorical distributions.

- A workable representation of feasible set $\mathcal{R}_{\mathbf{x}}$;
- Equivalence in graph theory for efficient computation.

The Gibbs sampler allows for straightforward

- Addition of categories: $K \to K + 1$;
- Addition of observations: $N \to N + 1$.

Extensions of the Categorical distribution include models for hierarchical counts, hidden Markov models, etc.

Jacob, Gong, Edlefsen & Dempster. *A Gibbs sampler for a class of random convex polytopes* (to appear in JASA with discussion). On ArXiv and Researchers.one.

R package available at `https://github.com/pierrejacob/dempsterpolytope`.

Bibliography

Berger, J. O. (1994). An overview of robust Bayesian analysis. *Test*, *3*(1), 5–124.

Berkelaar, M., Eikland, K., & Notebaert, P. (2004). lpsolve: Open source (mixed-integer) linear programming system. *Eindhoven U. of Technology*, *63*.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695. Retrieved from `http://igraph.org`

Dawid, P. A., & Stone, M. (1982). The functional-model basis of fiducial inference. *The Annals of Statistics*, *10*(4), 1054–1067.

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, *30*(2), 205–247.

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.

Fraser, D. A. (1968). *Structural inference.* John Wiley & Sons, New York, NY.

Hannig, J., Iyer, H., Lai, R. C., & Lee, T. C. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, *111*(515), 1346–1361.

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

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

Shafer, G. (1976). *A mathematical theory of evidence.* Princeton University Press, Princeton, NJ.

In case you ask...

Proof of the $\mathcal{R}_{\mathbf{x}}$ characterization

Proof of claim: "inequalities" $\Rightarrow$ "$\exists \theta: \ \theta_\ell / \theta_k \leq \eta_{k \to \ell} \ \forall k, \ell$".

$\min(k \to \ell) :=$ minimum value of path from k to ℓ in the graph.

It is finite $\forall k, \ell$ because no negative cycles in the graph.

Define θ via $\theta_k = \exp(\min(K \to k)) / \sum_{j \in [K]} \exp(\min(K \to j))$.

Then $\theta \in \Delta$. Also, for all k, ℓ

$$\min(K \to \ell) \leq \min(K \to k) + \log(\eta_{k \to \ell})$$

therefore $\theta_\ell / \theta_k \leq \eta_{k \to \ell}$.

Conditional distributions

We can obtain conditional distributions of u_n for $n \in \mathcal{I}_k$ given $(u_n)_{n \notin \mathcal{I}_k}$ with respect to $\nu_{\mathbf{x}}$:

u_n given $(u_n)_{n \notin \mathcal{I}_k}$ are i.i.d. uniform in $\Delta_k(\theta^\star)$,

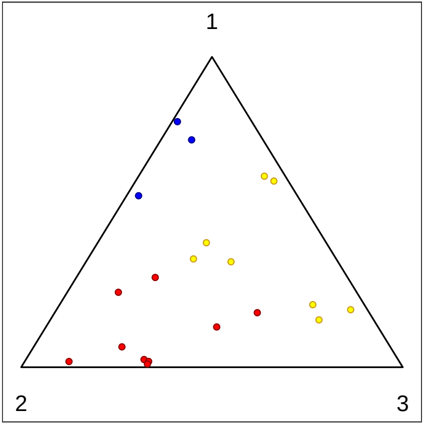
where $\theta^\star_\ell \propto \exp(-\min(\ell \to k))$ for all ℓ,

with $\min(\ell \to k) :=$ minimum value of path from ℓ to k.

Note: $\min(\ell \to k)$ can be computed in polynomial time.

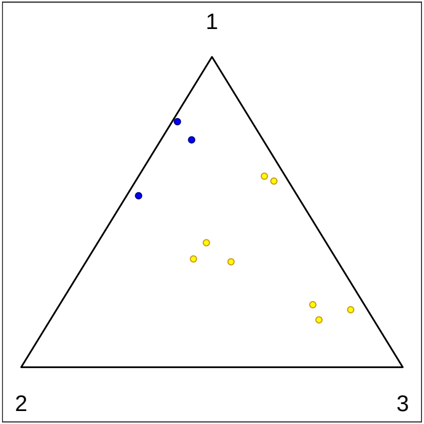
Conditional distributions

Counts: $(9, 8, 3)$. What is the conditional distribution of $(u_n)_{n \in \mathcal{I}_k}$ given $(u_n)_{n \notin \mathcal{I}_k}$ under $\nu_{\mathbf{x}}$?



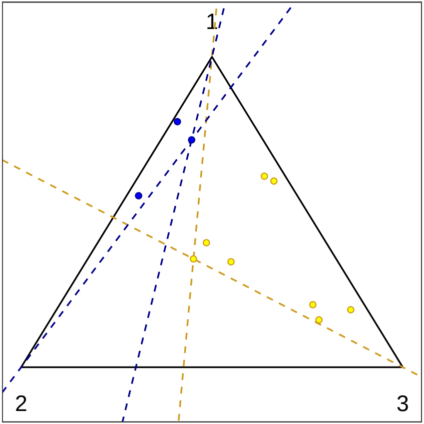
Conditional distributions

Counts: $(9, 8, 3)$. What is the conditional distribution of $(u_n)_{n\in\mathcal{I}_k}$ given $(u_n)_{n\notin\mathcal{I}_k}$ under $\nu_{\mathbf{x}}$?



Conditional distributions

Counts: $(9, 8, 3)$. What is the conditional distribution of $(u_n)_{n \in \mathcal{I}_k}$ given $(u_n)_{n \notin \mathcal{I}_k}$ under $\nu_{\mathbf{x}}$?



Conditional distributions

Counts: $(9, 8, 3)$. What is the conditional distribution of $(u_n)_{n \in \mathcal{I}_k}$ given $(u_n)_{n \notin \mathcal{I}_k}$ under $\nu_{\mathbf{x}}$?