The Statistical Mechanics of Formal Privacy

Theory and Experiments

by

Mark Fleischer, U.S. Census Bureau, CEDDA

FCSM Conference

October 23, 2024, College Park, MD

The views expressed herein are those of the author and not necessarily those of the U.S. Census Bureau

Acknowledgments

The writing of this article was supported by the U.S. Census Bureau. The author thanks Rolando Rodriguez of the U.S. Census Bureau for his support and encouragement, Alan Weiss of The Mathworks for helpful discussions and Philip Leclerc, Ryan Cumings, Michael Freiman and Brian Finley of the U.S. Census Bureau for helpful comments and suggestions. The views expressed herein do not necessarily reflect the views of the U.S. Census Bureau nor the views of the United States Government.

United States® Census Bureau

Overview:

- The need for Disclosure Avoidance and how it's done
- The perspective offered by Statistical Mechanics in the context of privacy
- The relevant elements of Statistical Mechanics.
  - Simulated Annealing (SA) and Markov Chain Monte Carlo (MCMC) simulations
  - Changing perspective: DP noise injection vs. DP noise evolution.
  - *Unsolving* optimization problems as a noise-evolution method (as opposed to a 'noise-injection' method).
- The Boltzmann Machine Privacy Framework (BMPF)
  - Description of consensus functions for histograms
  - Generating 'candidate' histograms
  - How the BMPF satisfies $(\boldsymbol{\epsilon}(t), \boldsymbol{\delta}(k))$-DP
- Experimental Results
  - MCMC burn-in issues, ways to ameliorate this *i.e.,* stopping criteria.
- Conclusion

United States Census Bureau

Disclosure Avoidance and What it Entails

- Can’t publish actual datasets collected by CB
  - Privacy laws prohibit disclosure. E.g., Title 13, Title 26 and others.
  - Publishing data with many traditional DA methods can still disclose private information (e.g., Governor Weld’s medical records were ‘anonymized’)
  - But good public policy-making requires *some* use of the data.
- How can we publish data yet maintain privacy?
- Differential Privacy (2006), formerly referred to as *epsilon indistinguishability* provides a methodology that guarantees a quantifiable level of privacy via a ‘privacy budget’.
- It infuses data with ‘calibrated noise’ to achieve this quantifiable level of privacy.

United States® Census Bureau

Examples:

- Histograms are a common type of dataset developed at Census
- They reflect counts of entities (people) that are associated with certain mutually exclusive combinations of attributes
- Publishing actual counts can lead to complete privacy loss
- DP modifies these counts in a probabilistic manner such that there is a quantifiable level of privacy yet still maintains usability/utility.

United States® Census Bureau

Definition of Differential Privacy

$$\Pr\{\mathcal{M}(\mathbf{x}')\} \in S\} \leq e^{\epsilon} \Pr\{\mathcal{M}(\mathbf{x}) \in S\} + \delta$$

E.g., adding random noise to pixels in a picture to blur the faces of people in the picture making it hard to identify the people in the picture, yet enabling a fairly accurate counting of the number of people.

Quantifies the fundamental tradeoff between accuracy and privacy.

Lots of way to create 'noisy data':

1. Add random variates to the actual data.
2. Consider the actual data as the 'optimum data' in an optimization problem and produce *sub-optimum* data.

United States® Census Bureau

Simulated Annealing and Markov Chains

Simulated Annealing (SA) circa 1983 is a meta-heuristic that can 'solve' a wide variety of optimization problems.

Hallmarks:

- Based on the Metropolis Algorithm (an accept/reject method), it enables Markov Chain Monte Carlo (MCMC) sampling.

The MAC: Let $\Delta E = E_{\text{cand}} - E_{\text{curr}}$

$$\Pr\{\text{Accept } E_{cand}\} = \begin{cases} e^{-\Delta E/t} & \text{if } \Delta E \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

$$\Pr\{\text{Accept Candidate } j\} = \frac{1}{1 + e^{-\Delta f_{ji}/t}}$$

$$\pi_i p_{ij} = \pi_j p_{ji}$$

$$\pi_i(t) = \frac{e^{-E_i/t}}{\sum_j e^{-E_j/t}}$$

- SA converges in probability under WLLN to the *globally optimal* solution:

$$\lim_{t \to 0^+} \pi_i(t) = \frac{1}{|\mathcal{C}_{\text{opt}}|}$$

United States® Census Bureau

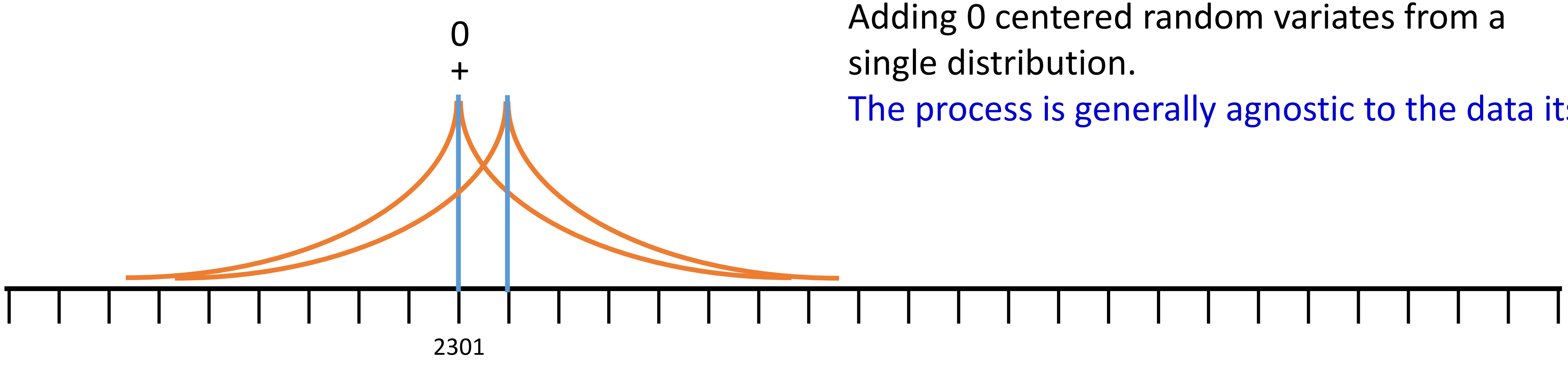
Simulated Annealing

- Used to solve a wide variety of optimization problem by virtue of its simplicity, convergence properties and generalizability.
- Requirements:
  1. A well defined configuration space.
  2. A well defined objective function.
  3. A 'good' candidate generation scheme.
  4. An appropriate 'cooling schedule'. (We'll just need a fixed temperature.)
- MA/SA effectively moves (transitions) from one configuration to another under the influence of these four elements.
- Transition probabilities → Markov Chain

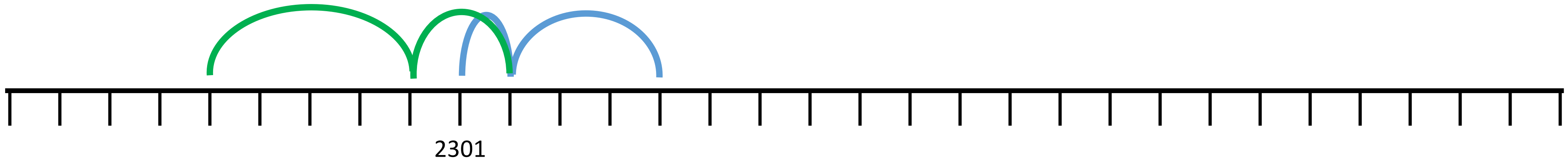
Instead of using SA to find the global optimum (the 'true' configuration), we use it to *move away* from the optimal solutions to find a *sub-optimal* configuration which is equivalent to a 'noisy' configuration by holding the temperature to some positive value.

United States® Census Bureau

Noise-Injection Paradigms



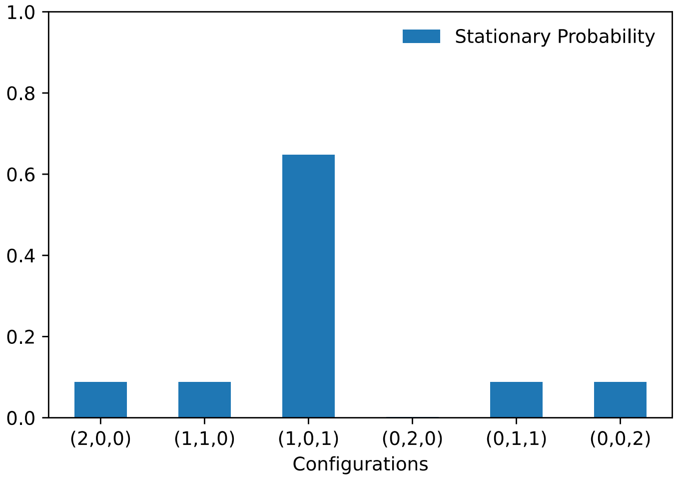
Adding 0 centered random variates from a single distribution.
The process is generally agnostic to the data itself.



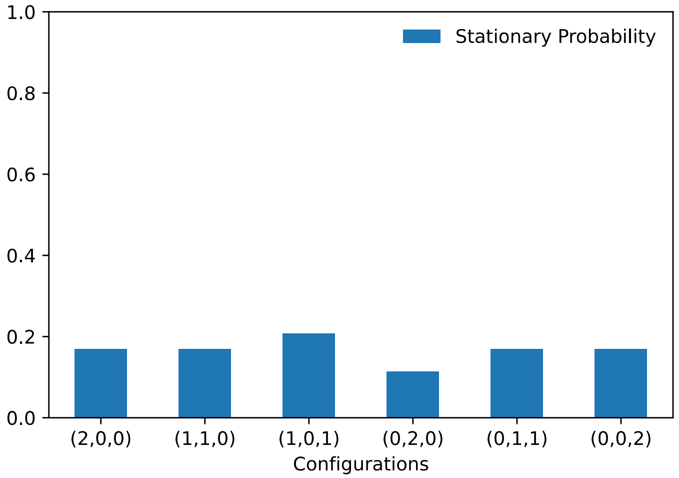
But in a diffusion model, the diffusion model may not be agnostic to the data.

Stationary Distributions

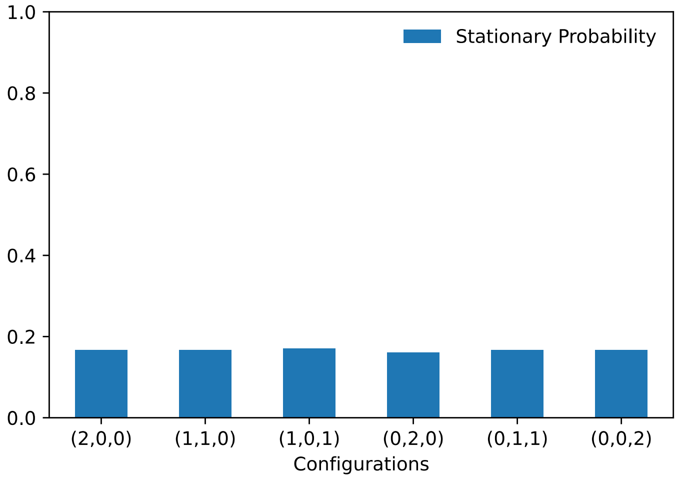
$$\mathcal{C} = \{(2,0,0), (1,1,0), \underline{(1,0,1)}, (0,2,0), (0,1,1), (0,0,2)\}$$



t = 1



t = 10



t = 100

Distributions of the Configuration Space ($\mathcal{C}$) of all histograms with $k = 3$ bins and $N = 2$ individuals

For a fixed N and fixed k, the size of the configuration space = $\binom{N+k-1}{k-1} = \binom{4}{2} = 6$

E.g., if $N = 50$, $k = 5 \rightarrow |\mathcal{C}| = 316251$

United States® Census Bureau

The Main Idea Behind Data Noise Injection Using The Metropolis/Simulated Annealing Approach

- ‘Ground truth’ = ‘observed data’ = ‘sampled data’ = ‘optimal data’
- ‘Disclosed data’ = ‘noise injected data’ = ‘suboptimal data’

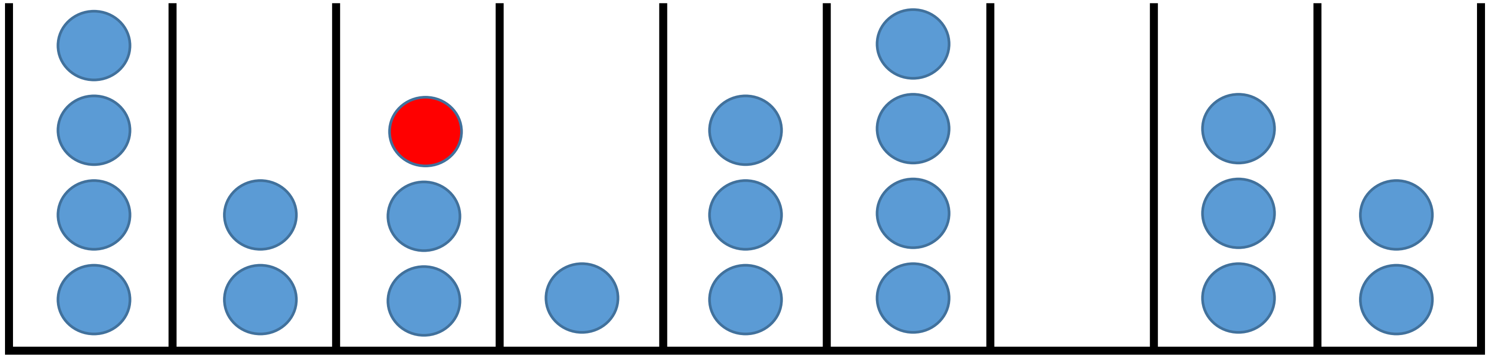
Ground truth/optimal histogram: [1,2,3,20,24] Total count = 50

t = 10	*t* = 100
[1,4,5,19,21]	[10,8,5,11,16]
[3,5,6,16,24]	[6,7,10,13,14]
[3,4,4,18,21]	[7,10,11,9,13]

United States® Census Bureau

Notice that the total count in all these histograms = 50

Candidate Generation Mechanism



Histogram

- Negative bin values are impossible.
- Invariants correspond to bins that are ignored.
- Total counts are unchanged.

Can the Metropolis/SA Algorithm Create Markov Chains that are DP?

Define the *configuration space*:

Definition : *Define* $\mathbf{x} \in \mathbb{N}^n$ *and corresponds to a histogram vector of* n *counts.*

Definition : *Given a vector* $\mathbf{x} \in \mathbb{N}^n$ *and where* $\sum_{i=1}^{n} x_i = N$, *define*

$$\mathcal{C}_{\mathbf{x}} = \{\mathbf{x} \in \mathbb{N}^n : \forall \mathbf{x} \in \mathbb{N}^n, \sum_{i=1}^{N} x_i = N\}$$

Can the Metropolis/SA Algorithm Create Markov Chains that are DP?

Define the following objective functions for neighboring configurations **x** and **y:**

$$c_{\mathbf{x}}(\mathbf{z}) = -(\mathbf{z}-\mathbf{x})^{\intercal}\mathbf{W}_{\mathbf{x}}(\mathbf{z}-\mathbf{x}) \textit{ and}$$

Definition : *For all neighboring configurations* $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, *define the global sensitivity*

$$s = \max_{\mathbf{z}\in\mathcal{C}} \max_{\mathbf{x},\mathbf{y}:d(\mathbf{x},\mathbf{y})\leq 2} \|c_{\mathbf{x}}(\mathbf{z}) - c_{\mathbf{y}}(\mathbf{z})\|_2$$

Each dataset/configuration **x** and **y** induces its own Markov Chain: $\pi_{\mathbf{z}|\mathbf{x}}(t) = \dfrac{e^{c_{\mathbf{x}}(\mathbf{z})/t}}{\sum_{\mathbf{z}'} e^{c_{\mathbf{x}}(\mathbf{z}')/t}}$

United States® Census Bureau

The Boltzmann Machine Mechanism

Definition : *Define the* Boltzmann Machine Mechanism $\mathcal{B}_{k,t}(\mathbf{x})$ *as the random configuration of a discrete time, irreducible and aperiodic Markov chain generated by application of the Metropolis algorithm at temperature* t *after* k *iterations given a dataset* $\mathbf{x}$ *where* $\mathbf{x}$ *is the initial configuration of the Markov Chain. This simply corresponds to the* k^{th}*-step transition probability of the Markov Chain. Thus,*

$$\Pr\{\mathcal{B}_{k,t}(\mathbf{x}) = \mathbf{z}\} \equiv \Pr\{X_k(t) = \mathbf{z} | X_0(t) = \mathbf{x}\} \quad \equiv p^{(k)}_{\mathbf{x},\mathbf{z}}$$

Theorem : *Let* $(X_k(t))_{k\geq 0}$ *be an irreducible, aperiodic Markov Chain based on the Metropolis algorithm as in Theorem 4. Then the BMM* $\mathcal{B}_{k,t}(\mathbf{x})$ *satisfies* $(\epsilon(t), \delta(k))-$*DP where* $\delta(k) \to 0$ *as* $k \to \infty$.

Markov Chain Convergence

Theorem : *Let X_k be an irreducible and aperiodic Markov Chain on a finite configuration space $\mathcal{C}$ with stationary distribution $\pi(t)$ at fixed temperature t. Then there exist constants $\alpha \in (0,1)$ and $C > 0$ such that for all state vectors $\mathbf{v}^{[k]}$*

$$\max_{\mathbf{v}^{[0]} \in \mathcal{C}} \|\mathbf{v}^{[k]} - \pi(t)\|_{TV} \leq C\alpha^k \tag{10}$$

⇓

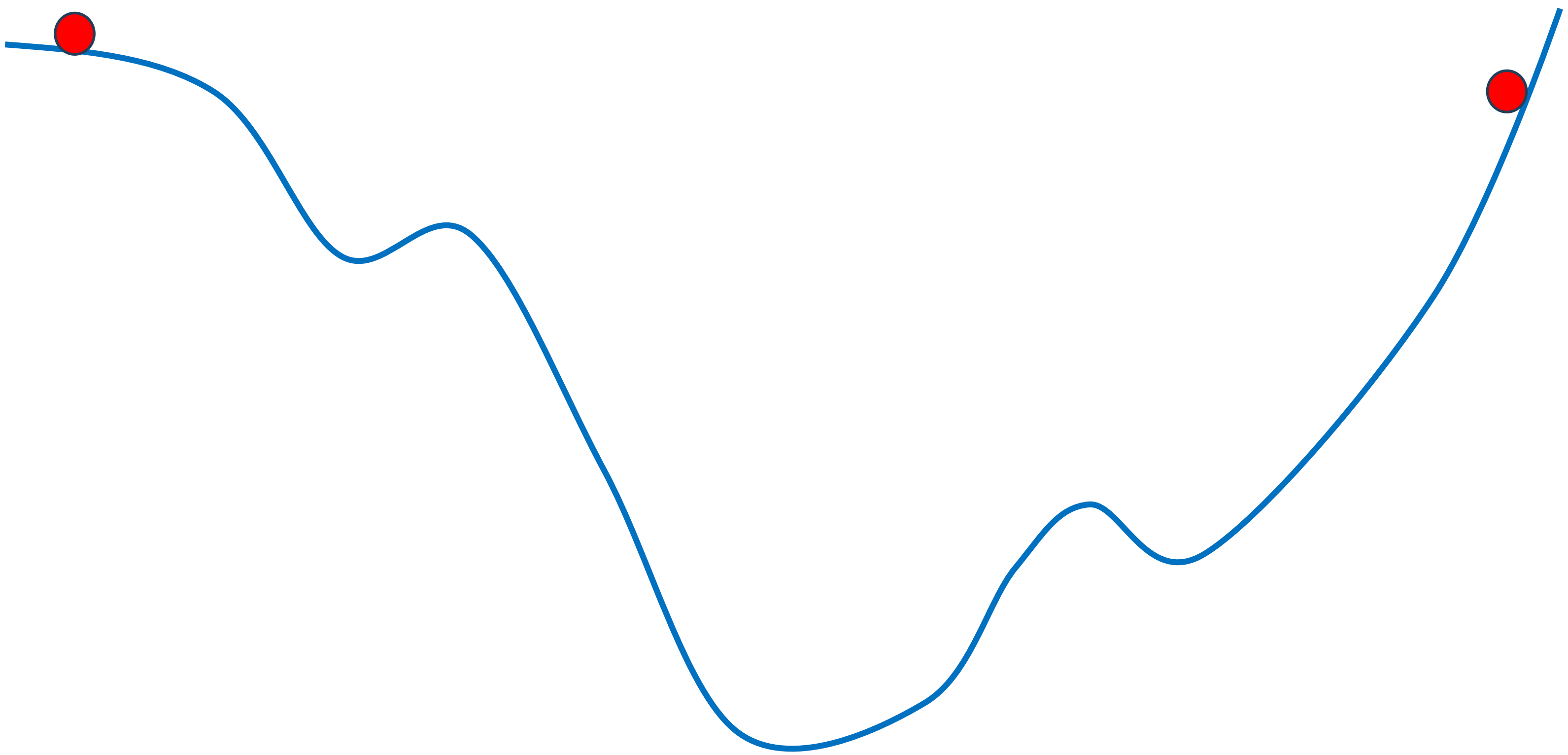
$$\Pr\{\mathcal{B}_{k,t}(\mathbf{x}) \in \mathcal{S}\} \leq e^{\epsilon(t)} \Pr\{\mathcal{B}_{k,t}(\mathbf{y}) \in \mathcal{S}\} + \delta(k)$$

where $\epsilon(t) = \frac{2s}{t}$ and $\delta(k) \rightarrow 0$ as $k \rightarrow \infty$

As $t \rightarrow \infty$, $\epsilon(t) \rightarrow 0$ As $t \rightarrow 0$, $\epsilon(t) \rightarrow \infty$

United States® Census Bureau

Convergence of Two or More Markov Chains



Convergence of Two or More Markov Chains

Theorem:
Given p irreducible, aperiodic Markov Chains $X_m^{[k]}, m = 1, 2, \ldots, p$ and a metric space $d(\cdot, \cdot)$ where

$$M_m^{[k]} = d\left(f(X_m^{[k]}), f(\mathbf{x}_{\text{OPT}})\right)$$

where $\forall\, m,\; M_m^{[k]} \xrightarrow{\text{Prob}} 0$ as $k \to \infty$ and some function

$$Y^{[k]} = g(X_1^{[k]}, X_2^{[k]}, \ldots, X_p^{[k]})$$

such that

$$Y^{[k]} \to 0 \text{ as } k \to \infty$$

and

$$Y^{[k]} = 0 \Leftrightarrow f(X_1) = f(X_2) = \ldots = f(X_p)$$

then for any chain m and k sufficiently high

$$\Pr\{M_m^{[k]} = 0 | Y^{[k]} = 0\} > \Pr\{M_m^{[k]} = 0\}.$$

United States® Census Bureau

Convergence of Two or More Markov Chains

Corollary: Let p be the number of independent Markov Chains and let $Y^{[k,p]}$ be the metric among the p chains as defined above at time index k. Then for any chain m with $p \geq 2$ and k sufficiently large, then

$$\Pr\{M_m^{[k]} = 0 | Y^{[k,p+1]} = 0\} > \Pr\{M_m^{[k]} = 0 | Y^{[k,p]} = 0\} > \Pr\{M_m^{[k]} = 0\}.$$

United States® Census Bureau

Stopping Criteria:

How can we apply the foregoing theorems when we are not converging to the 'true' data?

We can define a random variable that converges to 0 based on the ergodic theorem:

$$\frac{1}{n}\sum_{k=0}^{n-1} c_{\mathbf{x}}(X_1^{[k]}) \xrightarrow{a.s.} \sum_{\mathbf{z}\in\mathcal{C}} \pi_{\mathbf{z}|\mathbf{x}}(t) c_{\mathbf{x}}(\mathbf{z}) \text{ as } n \to \infty$$

Convergence to the expected objective function value.

$$\lim_{k\to\infty} \left| \frac{1}{n}\sum_{n=0}^{n-1} c_{\mathbf{x}}(X_1^{[k]}) - \sum_{\mathbf{z}\in\mathcal{C}} \pi_{\mathbf{z}|\mathbf{x}}(t) c_{\mathbf{x}}(\mathbf{z}) \right| = 0$$

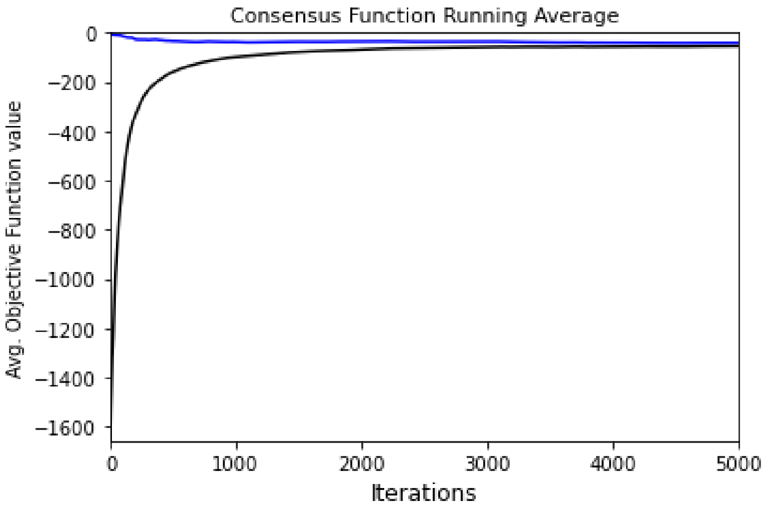
$$M_1^{[k]} \equiv \frac{1}{n}\sum_{n=0}^{n-1} c_{\mathbf{x}}(X_1^{[k]}) - \sum_{\mathbf{z}\in\mathcal{C}} \pi_{\mathbf{z}|\mathbf{x}}(t) c_{\mathbf{x}}(\mathbf{z}) \to 0$$

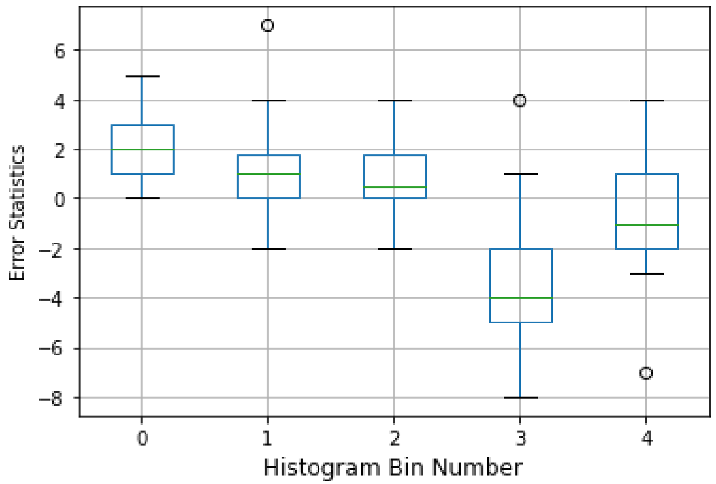
$$M_2^{[k]} \equiv \frac{1}{n}\sum_{n=0}^{n-1} c_{\mathbf{x}}(X_2^{[k]}) - \sum_{\mathbf{z}\in\mathcal{C}} \pi_{\mathbf{z}|\mathbf{x}}(t) c_{\mathbf{x}}(\mathbf{z}) \to 0$$

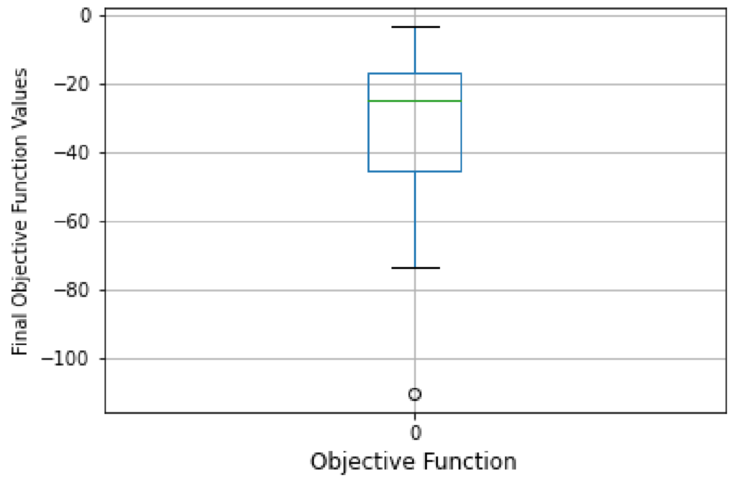
$$Y^{[k]} = M_1^{[k]} - M_2^{[k]} \xrightarrow{\text{prob}} 0$$

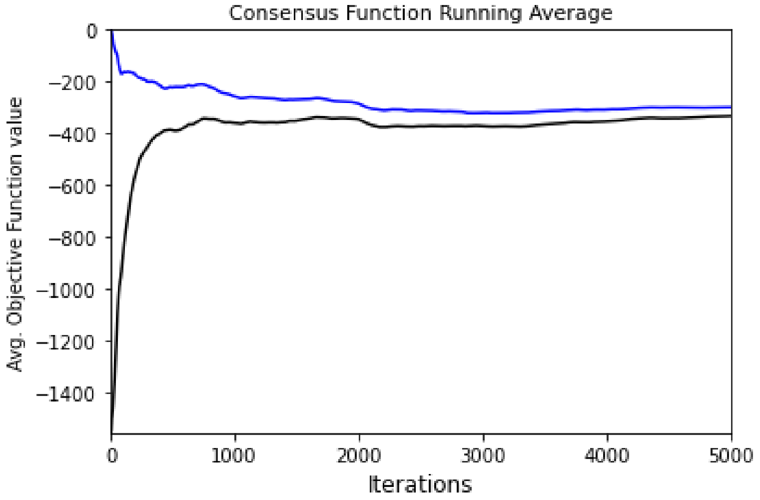
United States® Census Bureau

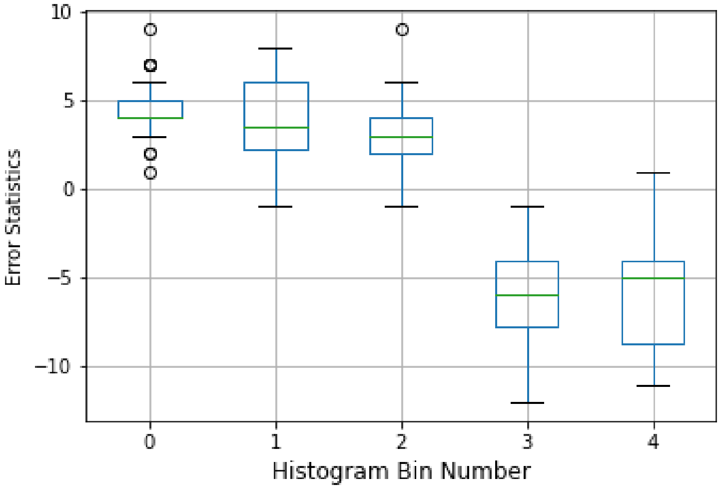
Experiments: *t = 10, 50; Run Length = 5000; Replications = 30*

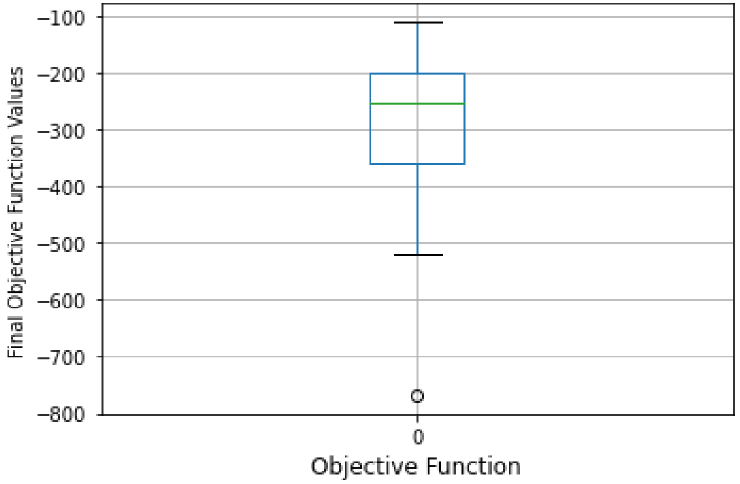












United States® Census Bureau

Histogram: (1, 2, 3, 20, 24)

Conclusion

- Metropolis/Simulated Annealing Algorithm can serve as a basis for noise injection.
- Boltzmann Machine Mechanism can provide ‘noise-evolution’.
- Constructing candidate generation mechanisms can obviate the need for post-processing.
- Resulting Markov Chains satisfies $(\epsilon(t), \delta(k)) -$ DP.
- Flexibility in implementation: data object types and objective functions.
- Possibility of *tuning* sensitivity by modifying the weight matrices.
- Bears some striking similarities to the Exponential Mechanism.

United States® Census Bureau

References:

- Flamm, D.: Ludwig Boltzmann and His Influence on Science. Studies in History and Philosophy of Science Part A **14** (1983) 255–278
- Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In Halevi, S., Rabin, T., eds.: Theory of Cryptography, Berlin, Heidelberg, Springer Berlin Heidelberg (2006) 265–284
- Hinton, G., Seijnowski, T.: Optimal perceptual inference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, D.C (1983) 448–453
- Manco, G., Pirrò, G.: Differential privacy and neural networks: A preliminary analysis. In Guidotti, R., Monreale, A., Pedreschi, D., Abiteboul, S., eds.: Personal Analytics and Privacy. An Individual and Collective Perspective, Cham, Springer International Publishing (2017) 23–35
- Wang, Y.X., Fienberg, S.E., Smola, A.: Privacy for free: Posterior sampling and stochastic gradient monte carlo (2015)
- Yildirim, S.: On the use of penalty mcmc for differential privacy. arXiv: Computation (2016)
- Yildirim, S., Ermis, B.: Exact mcmc with differentially private moves: Revisiting the penalty algorithm in a data privacy framework. Statistics and Computing **29** (2019)
- Le, T.T., Simmons, W.K., Misaki, M., Bodurka, J., White, B.C., Savitz, J., McKinney, B.A.: Differential privacy-based evaporative cooling feature selection and classification with relief-F and random forests. Bioinformatics **33** (2017) 2906–2913
- Li, B., Chen, C., Liu, H., Carin, L.: On connecting stochastic gradient mcmc and differential
- Lalley, S.P.: Convergence rates of markov chains (1997)
- Fleischer, M.A.: Cybernetic optimization by simulated annealing: Accelerating convergence by parallel processing and probabilistic feedback control. Journal of Heuristics **1** (1996)

United States® Census Bureau