Locally Private Bayesian Inference for Count Models

Aaron Schein UMass Amherst
Zhiwei Steven Wu Microsoft Research
Mingyuan Zhou UT Austin
Hanna Wallach Microsoft Research

1 Introduction

Data from social processes often take the form of discrete observations (e.g., ties in a social network, word tokens in an email) and these observations often contain sensitive information about the people involved. As more aspects of social interaction are digitally recorded, the opportunities for social scientific insights grow; however, so too does the risk of unacceptable privacy violations. As a result, there is a growing need to develop privacy-preserving data analysis methods. Computational social scientists will be more likely to adopt these methods if doing so entails minimal change to their current methodology. Toward that end, under the framework of differential privacy [11], we develop a method for privatizing Bayesian inference for Poisson factorization [25, 6, 33, 14, 16], a broad class of models for discrete data. This class contains some of the most widely used models in social science, including topic models for text corpora [3, 4, 5], genetic population models [18], stochastic block models for social networks [2, 14, 32], and tensor factorization for dyadic data [28, 8, 23, 20, 22]; it further includes deep hierarchical models [19, 34], dynamic models [7, 1, 21], and many others. Our method is general and modular, allowing social scientists to build on (instead of replace) existing derivations and implementations of non-private Poisson factorization. Furthermore, our method satisfies a strong variant of differential privacy—i.e., local privacy—under which the sensitive data is privatized (or noised) before inference. This ensures that no single centralized server need ever store the non-privatized data—a condition that is non-negotiable in many real-world settings. Our method is asymptotically guaranteed to draw samples from the posterior distribution over model parameters conditioned on privatized observations. To derive this method, we rely on a previously unknown relationship between the Skellam [24] and Bessel [30] distributions, as well as a novel reinterpretation of the two-sided geometric noise mechanism [13]; these new results may be of independent interest.

2 Background and problem formulation

Local differential privacy. Differential privacy [11] is a rigorous privacy criterion that guarantees that no single observation in a data set will have a significant influence on the information obtained by analyzing that data set. Here we focus on local differential privacy, which we refer to as local privacy. In the local privacy setting, the observations remain private from even the data analysis algorithm. The algorithm only sees privatized versions of the observations, constructed by adding noise from specific distributions, including the Gaussian and Laplace distributions. The process of adding noise is known as randomized response—a reference to survey-sampling methods originally developed in the social sciences prior to the development of differential privacy [27]. Formally, we say that a randomized response method $\mathcal{R}(\cdot)$ is ϵ-private if for all pairs of observations y and y'

$$P\left(\mathcal{R}(y) \in \mathcal{S}\right) \leq e^{\epsilon} P\left(\mathcal{R}(y') \in \mathcal{S}\right) \tag{1}$$

for all subsets $\mathcal{S}$ in the range of $\mathcal{R}(\cdot)$. If a data analysis algorithm sees only the observations' ϵ-private responses, then we say that the analysis itself satisfies ϵ-local differential privacy.

Private Bayesian inference. In Bayesian statistics, we begin with a probabilistic model $\mathcal{M}$ that relates observable variables Y to latent variables Z via a joint distribution $P_{\mathcal{M}}(Y, Z)$. The goal of inference is then to compute the posterior distribution $P_{\mathcal{M}}(Z \mid Y)$ over the latent variables conditioned on observed values of Y. The posterior is almost always analytically intractable and thus inference involves approximating it. The two most common methods of approximate Bayesian inference are variational inference, wherein we fit the parameters of an approximating distribution $Q(Z \mid Y)$, and

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Markov chain Monte Carlo (MCMC), wherein we approximate the posterior with a finite set of samples $\{Z^{(s)}\}_{s=1}^{S}$ generated via a Markov chain whose stationary distribution is the exact posterior. We can conceptualize each of these methods as an algorithm $\mathcal{A}(Y)$ that returns an approximation to the posterior distribution $P_{\mathcal{M}}(Z \mid Y)$; in general $\mathcal{A}(Y)$ does not satisfy differential privacy. However, if $\mathcal{A}(Y)$ is an MCMC algorithm that returns a single sample from the posterior, it guarantees privacy [10, 26, 12]. Adding noise to posterior samples can also guarantee privacy [31], though this set of noised samples $\{\tilde{Z}^{(s)}\}_{s=1}^{S}$ collectively approximate some distribution $\tilde{P}_{\mathcal{M}}(Z \mid Y)$ that depends on ϵ and is different than the exact posterior (but close, in some sense, and equal when $\epsilon \to 0$). For specific models, we can also noise the transition kernel of the MCMC algorithm to construct a Markov chain whose stationary distribution is again not the exact posterior, but something close that guarantees privacy [12]. An analogous approach can be taken to privatize variational inference, wherein noise is added to the sufficient statistics computed in each iteration [17]. All of these approaches assume the global privacy setting; in contrast, we focus on the local privacy setting for a large class of models for discrete data.

Locally private Bayesian inference. We first formalize the general objective of Bayesian inference under local privacy. Given a generative model $\mathcal{M}$ for non-privatized data Y and latent variables Z with joint distribution $P_{\mathcal{M}}(Y, Z)$, we further assume a privatizing noise model $\mathcal{E}$ that generates ϵ-privatized data sets: $\tilde{Y} \sim P_{\mathcal{E}}(\tilde{Y} \mid Y, \epsilon)$. The aim of Bayesian inference is then to form the following posterior:

$$P_{\mathcal{M},\mathcal{E}}(Z \mid \tilde{Y}, \epsilon) = \mathbb{E}_{P_{\mathcal{E}}(Y \mid \tilde{Y}, \epsilon)}\left[P_{\mathcal{M}}(Z \mid Y)\right] = \int P_{\mathcal{M}}(Z \mid Y)\, P_{\mathcal{E}}(Y \mid \tilde{Y}, \epsilon)\, dY. \tag{2}$$

This distribution correctly characterizes our uncertainty about the latent variables Z, conditioned on all of our observations and assumptions—i.e., the privatized data $\tilde{Y}$, the model $\mathcal{M}$, the privatizing noise model $\mathcal{E}$, and the privacy level ϵ. The expansion in equation 2 shows that this posterior implicitly treats the non-privatized data Y as a latent variable and marginalizes over it using the mixing distribution $P_{\mathcal{E}}(Y \mid \tilde{Y}, \epsilon)$ which is itself a posterior that characterizes our uncertainty about Y given $\tilde{Y}$, the noise model $\mathcal{E}$, and the privacy level ϵ. The key observation here is that if we can generate samples from $P_{\mathcal{E}}(Y \mid \tilde{Y}, \epsilon)$, we can use them to approximate the expectation in equation 2, assuming that we already have a method for approximating the non-private posterior $P_{\mathcal{M}}(Z \mid Y)$. In the context of MCMC, alternating between sampling values of the non-privatized data from its complete conditional—i.e., $Y^{(s)} \sim P_{\mathcal{M},\mathcal{E}}(Y \mid Z^{(s-1)}, \tilde{Y}, \epsilon)$—and sampling values of the latent variables—i.e., $Z^{(s)} \sim P_{\mathcal{M}}(Z \mid Y^{(s)})$—constitutes a Markov chain whose stationary distribution contains is $P_{\mathcal{M},\mathcal{E}}(Z, Y \mid \tilde{Y}, \epsilon)$. In scenarios where we already have derivations and implementations for sampling from $P_{\mathcal{M}}(Z \mid Y)$, we need only be able to sample efficiently from $P_{\mathcal{M},\mathcal{E}}(Y \mid Z, \tilde{Y}, \epsilon)$ in order to obtain a locally ϵ-private Bayesian inference algorithm; whether we can do this depends heavily on our assumptions about $\mathcal{M}$ and $\mathcal{E}$.

We note that the objective of Bayesian inference under local privacy, as given in equation 2, is similar to that given by Williams and McSherry [29], who identify their key barrier to inference as being unable to analytically form the marginal likelihood that links the privatized data to the latent variables:

$$P_{\mathcal{M},\mathcal{E}}(\tilde{Y} \mid Z, \epsilon) = \int P_{\mathcal{E}}(\tilde{Y} \mid Y, \epsilon)\, P_{\mathcal{M}}(Y \mid Z)\, dY. \tag{3}$$

In the next section, we show that for a particular class of models for discrete data and a particular discrete noise model, we can analytically form this marginal likelihood and derive an efficient MCMC algorithm that is asymptotically guaranteed to generate samples from the posterior in equation 2.

3 Locally private Poisson factorization

We assume that Y is a count-valued data set. We further assume that each count in this data set is an independent Poisson random variable $y_n \sim \text{Pois}(\mu_n)$ where the count's latent rate parameter μ_n is a function of some globally shared model parameters. This class of models is known as Poisson factorization and, as described in section 1, includes many widely used models in social science. For the purpose of exposition, we will focus on the simple example of Poisson matrix factorization [25, 6], wherein $Y \in \mathbb{Z}_{+}^{D \times V}$ is a count matrix and each element of this matrix y_{dv} is drawn as follows:

$$\mu_{dv} = \boldsymbol{\theta}_d^{\top} \boldsymbol{\phi}_v, \quad y_{dv} \sim \text{Pois}(\mu_{dv}). \tag{4}$$

In this scenario, the latent variables are the two factor matrices—i.e., $Z \equiv \{\Theta, \Phi\}$. It is standard to assume independent gamma priors over the elements of both factor matrices; doing so facilitates

efficient Bayesian inference of these matrices via gamma–Poisson conjugacy (when conditioned on Y). As explained in the previous section, we would like to choose a privatizing noise model $\mathcal{E}$ that will enable us we to form the marginal likelihood $P_{\mathcal{M},\mathcal{E}}(\tilde{y}_{dv} \mid \mu_{dv}, \epsilon)$ and sample efficiently from $P_{\mathcal{M},\mathcal{E}}(y_{dv} \mid \tilde{y}_{dv}, \mu_{dv}, \epsilon)$. The standard noise mechanisms in the differential privacy toolbox are real-valued distributions (e.g., Gaussian and Laplace). Unfortunately, this means that they are poor choices in this scenario because a mixture of discrete and real-valued distributions makes it difficult to form $P_{\mathcal{M},\mathcal{E}}(y_{dv} \mid \tilde{y}_{dv}, \mu_{dv}, \epsilon)$ and $P_{\mathcal{M},\mathcal{E}}(\tilde{y}_{dv} \mid \mu_{dv}, \epsilon)$. We therefore take a different approach that allows us to obtain analytic tractability and efficient inference while maintaining strong privacy guarantees. Specifically, we reinterpret an existing, but less standard, privacy mechanism as involving Skellam noise, and derive a general fact relating the Skellam, Bessel, and Poisson distributions.

Two-sided geometric noise. Ghosh et al. [13] showed that adding two-sided geometric noise guarantees ϵ-differential privacy. A two-sided geometric random variable $\tau \sim \text{2Geo}(\epsilon)$ is a (positive or negative) integer $\tau \in \mathbb{Z}$. The PMF for the two-sided geometric distribution is as follows:

$$\text{2Geo}(\tau; \epsilon) = \frac{1-\epsilon}{1+\epsilon}\,\epsilon^{|\tau|} \tag{5}$$

We choose our privatizing noise model $\mathcal{E}$ to be a two-sided geometric distribution. Each non-privatized count y_{dv} is generated by our model $\mathcal{M}$ using equation 4 and then privatized as follows:

$$\tau_{dv} \sim \text{2Geo}(\epsilon), \quad \tilde{y}_{dv}^{(\pm)} := y_{dv} + \tau_{dv}. \tag{6}$$

We use the $(\pm)$ notation to emphasize that $\tilde{y}_{dv}^{(\pm)} \in \mathbb{Z}$ may be positive or negative. Although Ghosh et al. [13] presented the two-sided geometric distribution as the "discretized version of the Laplace distribution," we present a novel reinterpretation of this mechanism that allows us to analytically form the marginal likelihood $P_{\mathcal{M},\mathcal{E}}(\tilde{y}_{dv}^{(\pm)} \mid \mu_{dv}, \epsilon)$. We prove the following lemma in the appendix.

Lemma 1: A two-sided geometric random variable $\tau \sim \text{2Geo}(\epsilon)$ can be generated as follows:

$$\lambda_1 \sim \text{Exp}(\tfrac{\epsilon}{1-\epsilon}), \quad \lambda_2 \sim \text{Exp}(\tfrac{\epsilon}{1-\epsilon}), \quad \tau \sim \text{Skellam}(\lambda_1, \lambda_2), \tag{7}$$

where the Skellam distribution [24] is defined as the marginal distribution over the difference $\tau := g_1 - g_2$ of two independent Poisson random variables $g_1 \sim \text{Pois}(\lambda_1)$ and $g_2 \sim \text{Pois}(\lambda_2)$.

Via equation 7, we can express the generative process for $\tilde{y}_{dv}^{(\pm)}$ in three equivalent ways, each of which provides a unique and necessary insight into interpreting the two-sided geometric noise mechanism.

Process 1	**Process 2**	**Process 3**
—	$\lambda_{dvn} \sim \text{Exp}(\frac{\epsilon}{1-\epsilon})$ for $n \in \{1,2\}$	$\lambda_{dvn} \sim \text{Exp}(\frac{\epsilon}{1-\epsilon})$ for $n \in \{1,2\}$
$\tau_{dv} \sim \text{2Geo}(\epsilon)$	$g_{dvn} \sim \text{Pois}(\lambda_{dvn})$ for $n \in \{1,2\}$	—
$y_{dv} \sim \text{Pois}(\mu_{dv})$	$y_{dv} \sim \text{Pois}(\mu_{dv})$	—
$\tilde{y}_{dv}^{(\pm)} := y_{dv} + \tau_{dv}$	$\tilde{y}_{dv}^{(\pm)} := y_{dv} + g_{dv1} - g_{dv2}$	$\tilde{y}_{dv}^{(\pm)} \sim \text{Skellam}(\lambda_{dv1} + \mu_{dv},\, \lambda_{dv2})$

Three ways to generate $\tilde{y}_{dv}^{(\pm)}$. The first way is useful for showing that our MCMC algorithm guarantees privacy, since two-sided geometric noise is an existing privacy mechanism. The second way represents the two-sided geometric noise in terms of a pair of Poisson random variables with exponentially distributed rates; in so doing, it reveals the auxiliary variables that facilitate inference. The third way marginalizes out all three Poisson random variables (including y_{dv}), so that $\tilde{y}_{dv}^{(\pm)}$ is directly drawn from Skellam distribution, which also happens to be the desired marginal likelihood $P_{\mathcal{M},\mathcal{E}}(\tilde{y}_{dv} \mid \mu_{dv}, \epsilon)$ under two-sided geometric noise. To derive the second and third ways, we use lemma 1, the definition of the Skellam distribution, and the additive property of two or more Poisson random variables.

4 MCMC algorithm

In this section, we derive an efficient way to draw samples from $P_{\mathcal{M},\mathcal{E}}(y_{dv} \mid \tilde{y}_{dv}, \mu_{dv}, \epsilon)$. As explained in section 2, this is all we need to obtain a locally private MCMC algorithm for drawing samples of the latent variables—i.e., the factor matrices—given the privatized data, provided that we already have a

method for drawing samples of the latent variables given the non-private data. To derive our MCMC algorithm, we rely on a previously unknown relationship between three discrete distributions—the Poisson, Skellam, and Bessel distributions. We prove the following lemma in the appendix.

Lemma 2: Consider two Poisson random variables $y_1 \sim \text{Pois}(\lambda_1)$ and $y_2 \sim \text{Pois}(\lambda_2)$. Their minimum $m := \min\{y_1, y_2\}$ and their difference $\delta := y_1 - y_2$ are deterministic functions of y_1 and y_2. Then, if not conditioned on y_1 and y_2, the random variables m and δ can be marginally generated as follows:

$$\delta \sim \text{Skellam}(\lambda_1, \lambda_2), \quad m \sim \text{Bessel}\left(|\delta|, 2\sqrt{2\lambda_1\lambda_2}\right). \tag{8}$$

Yuan and Kalbfleisch [30] give details of the Bessel distribution, which can be sampled efficiently [9].[1]

Thus, to generate two independent Poisson random variables, we can first generate their difference δ and then their minimum m. Because $\delta = y_1 - y_2$, if δ is positive, then y_2 must be the minimum $y_2 = m$ and thus $y_1 = \delta - m$. In practice, this means that if we only get to observe the difference of two Poisson-distributed counts, we can "recover" the counts by drawing a Bessel random variable.

Gibbs sampling. The input to this algorithm is the privatized data set $\tilde{Y}^{(\pm)}$. Assuming that $\tilde{y}_{dv}^{(\pm)} \sim \text{Skellam}(\lambda_{dv1} + \mu_{dv}, \lambda_{dv2})$, as described in the previous section, we can represent $\tilde{y}_{dv}^{(\pm)}$ explicitly as the difference between two latent non-negative counts: $\tilde{y}_{dv}^{(\pm)} = \tilde{y}_{dv}^{(+)} - g_{dv2}$. We further define the minimum of these non-negative counts to be $m_{dv} = \min\{\tilde{y}_{dv}^{(+)}, g_{dv2}\}$. Given randomly initialized factor matrices, we can sample a value of m_{dv} from its conditional posterior, which is a Bessel distribution:

$$\left(m_{dv} \mid -\right) \sim \text{Bessel}\left(|\tilde{y}_{dv}^{(\pm)}|,\, 2\sqrt{(\lambda_{dv1}+\mu_{dv})\lambda_{dv2}}\right). \tag{9}$$

Using this value, we can then compute $\tilde{y}_{dv}^{(+)}$ and g_{dv2} (which are determined by m_{dv} and $\tilde{y}_{dv}^{(\pm)}$):

$$\tilde{y}_{dv}^{(+)} := m_{dv}, \quad g_{dv2} := \tilde{y}_{dv}^{(+)} - \tilde{y}_{dv}^{(\pm)} \quad \text{if } \tilde{y}_{dv}^{(\pm)} \leq 0 \tag{10}$$

$$g_{dv2} := m_{dv}, \quad \tilde{y}_{dv}^{(+)} := g_{dv2} + \tilde{y}_{dv}^{(\pm)} \quad \text{otherwise.} \tag{11}$$

Because $\tilde{y}_{dv}^{(+)}$ is defined to be the sum of two independent Poisson random variables—i.e., $\tilde{y}_{dv}^{(+)} = y_{dv} + g_{dv1}$—we can then sample y_{dv} from its conditional posterior, which is a binomial distribution:

$$\left(y_{dv} \mid -\right) \sim \text{Binom}\left(y_{dv}^{(+)}, \tfrac{\mu_{dv}}{\mu_{dv}+\lambda_{dv1}}\right) \tag{12}$$

Equations 9 through 12 constitute a method for sampling a value of y_{dv} from $P_{\mathcal{M},\mathcal{E}}(y_{dv} \mid \tilde{y}_{dv}, \mu_{dv}, \epsilon)$. Given this value, we can then draw samples of θ_{dk} and ϕ_{kv} from their conditional posteriors, which are the same as in non-private Poisson factorization. Finally, we can also sample λ_{dv1} and λ_{dv2}:

$$\left(\lambda_{dvn} \mid -\right) \sim \Gamma\left(1 + g_{dvn}, \tfrac{\epsilon}{1-\epsilon} + 1\right) \text{ for } n \in \{1, 2\} \tag{13}$$

Equation 13 follows from gamma–Poisson conjugacy and the fact that the exponential prior over λ_{dvn} can be expressed as a gamma prior with shape parameter equal to one—i.e., $\lambda_{dvn} \sim \Gamma(1, \frac{\epsilon}{1-\epsilon})$.

Equations 9–13, along with the conditional posteriors for θ_{dk} and ϕ_{kv}, define an MCMC algorithm that is asymptotically guaranteed to generate samples from $P_{\mathcal{M},\mathcal{E}}(\Theta, \Phi \mid \tilde{Y}^{(\pm)}, \epsilon)$ as desired.

5 Results and future directions

Preliminary experimental results (in the appendix) demonstrate the utility of our method—in particular, its ability to construct an accurate posterior over the non-privatized data. To our knowledge, our method is the only locally private Bayesian inference algorithm for Poisson factorization. We note that our approach can be also used to derive a globally private Bayesian inference algorithm, which would allow us to compare our method with existing work [17]; however we leave this contribution for the future. Finally, we can also use our approach to privatize variational inference.

[1] We have released our implementation of Bessel sampling. It is the only open-source version of which we are aware: `https://github.com/aschein/fatwalrus/blob/master/fatwalrus/bessel.pyx`

References

[1] Ayan Acharya, Joydeep Ghosh, and Mingyuan Zhou. Nonparametric Bayesian factor analysis for dynamic count matrices. arXiv:1512.08996, 2015.

[2] Brian Ball, Brian Karrer, and Mark E. J. Newman. Efficient and principled method for detecting communities in networks. *Physical Review E*, 84(3):036103, 2011.

[3] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[4] Wray Buntine and Aleks Jakulin. Applying discrete PCA in data analysis. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, pages 59–66, 2004.

[5] John Canny. GaP: a factor model for discrete data. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 122–129, 2004.

[6] Ali Taylan Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009.

[7] Laurent Charlin, Rajesh Ranganath, James McInerney, and David M. Blei. Dynamic Poisson factorization. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 155–162, 2015.

[8] Eric C. Chi and Tamara G. Kolda. On tensors, sparsity, and nonnegative factorizations. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1272–1299, 2012.

[9] Luc Devroye. Simulating Bessel random variables. *Statistics & Probability Letters*, 57(3): 249–257, 2002.

[10] Christos Dimitrakakis, Blaine Nelson, Aikaterini Mitrokotsa, and Benjamin I. P. Rubinstein. Robust and private Bayesian inference. In *International Conference on Algorithmic Learning Theory*, pages 291–305, 2014.

[11] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Theory of Cryptography Conference*, volume 3876, pages 265–284, 2006.

[12] James Foulds, Joseph Geumlek, Max Welling, and Kamalika Chaudhuri. On the theory and practice of privacy-preserving Bayesian data analysis. 2016.

[13] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.

[14] Prem K. Gopalan and David M. Blei. Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences*, 110(36):14534–14539, 2013.

[15] Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz. *Univariate discrete distributions*. 2005.

[16] John Paisley, David M. Blei, and Michael I. Jordan. Bayesian nonnegative matrix factorization with stochastic variational inference. In Edoardo M. Airoldi, David M. Blei, Elena A. Erosheva, and Stephen E. Fienberg, editors, *Handbook of Mixed Membership Models and Their Applications*, pages 203–222. 2014.

[17] Mijung Park, James Foulds, Kamalika Chaudhuri, and Max Welling. Private topic modeling. *arXiv:1609.04120*, 2016.

[18] Jonathan K. Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

[19] Rajesh Ranganath, Linpeng Tang, Laurent Charlin, and David Blei. Deep exponential families. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pages 762–771, 2015.

[20] Aaron Schein, John Paisley, David M. Blei, and Hanna Wallach. Bayesian Poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1045–1054, 2015.

[21] Aaron Schein, Hanna Wallach, and Mingyuan Zhou. Poisson–gamma dynamical systems. In *Advances in Neural Information Processing Systems 29*, pages 5005–5013, 2016.

[22] Aaron Schein, Mingyuan Zhou, David M. Blei, and Hanna Wallach. Bayesian Poisson Tucker decomposition for learning the structure of international relations. In *Proceedings of the 33rd International Conference on Machine Learning*, 2016.

[23] Mikkel N. Schmidt and Morten Morup. Nonparametric Bayesian modeling of complex networks: an introduction. *IEEE Signal Processing Magazine*, 30(3):110–128, 2013.

[24] John G. Skellam. The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society, Series A (General)*, 109:296, 1946.

[25] Michalis K. Titsias. The infinite gamma–Poisson feature model. In *Advances in Neural Information Processing Systems 21*, pages 1513–1520, 2008.

[26] Yu-Xiang Wang, Stephen Fienberg, and Alex Smola. Privacy for free: posterior sampling and stochastic gradient Monte Carlo. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 2493–2502, 2015.

[27] Stanley L. Warner. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

[28] Max Welling and Markus Weber. Positive tensor factorization. *Pattern Recognition Letters*, 22 (12):1255–1261, 2001.

[29] Oliver Williams and Frank McSherry. Probabilistic inference and differential privacy. In *Advances in Neural Information Processing Systems 23*, pages 2451–2459, 2010.

[30] Lin Yuan and John D. Kalbfleisch. On the Bessel distribution and related problems. *Annals of the Institute of Statistical Mathematics*, 52(3):438–447, 2000.

[31] Zuhe Zhang, Benjamin I. P. Rubinstein, and Christos Dimitrakakis. On the differential privacy of Bayesian inference. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, pages 2365–2371, 2016.

[32] Mingyuan Zhou. Infinite edge partition models for overlapping community detection and link prediction. In *Proceedings of the 18th Conference on Artificial Intelligence and Statistics*, pages 1135–1143, 2015.

[33] Mingyuan Zhou and Lawrence Carin. Augment-and-conquer negative binomial processes. In *Advances in Neural Information Processing Systems 25*, pages 2546–2554, 2012.

[34] Mingyuan Zhou, Yulai Cong, and Bo Chen. The Poisson gamma belief network. In *Advances in Neural Information Processing Systems 28*, pages 3043–3051, 2015.

A Formal privacy guarantee

We present a slight generalization of the two-sided geometric noise mechanism [13] and show that it can serve as a randomized response method for ensuring local privacy. We assume that each observation is a count-valued vector of length V whose l_1 norm is at most Δ. The two-sided geometric noise mechanism takes an observation and adds independently generated two-sided geometric noise to each element, as in equation 6. We show that this mechanism is a randomized response method—i.e., for all pairs of observations $\boldsymbol{y}$ and $\boldsymbol{y'}$, the distributions of their responses are almost identical.

Theorem 1: Fix any $\epsilon > 0$. For all pairs of observations $\boldsymbol{y}$ and $\boldsymbol{y}'$ such that $\|\boldsymbol{y}\|_1, \|\boldsymbol{y}'\|_1 \leq \Delta$, then

$$P\left(\mathcal{R}(\boldsymbol{y}) \in \mathcal{S}\right) \leq e^{\Delta\epsilon} P\left(\mathcal{R}(\boldsymbol{y}') \in \mathcal{S}\right) \tag{14}$$

for all subsets $\mathcal{S} \subseteq \mathbb{Z}_+^V$, where $\mathcal{R}(\cdot)$ is the two-sided geometric noise mechanism, parameterized by ϵ.

In the context of Poisson matrix factorization, $Y \in \mathbb{Z}_+^{D \times V}$ is a count matrix and each row of this matrix $\boldsymbol{y}_d \in \mathbb{Z}_+^V$ is an observation. Here Δ is domain-specific—for example, if Y is a matrix of document–word counts, then Δ might be the maximum document length. The MCMC algorithm described in section 4 only sees the privatized versions of the observations, constructed via the two-sided geometric mechanism. Therefore, it satisfies $\Delta\epsilon$-local differential privacy, as desired.

B Proof of lemma 1

Lemma 1: A two-sided geometric random variable $\tau \sim \text{2Geo}(\epsilon)$ can be generated as follows:

$$\lambda_1 \sim \text{Exp}(\tfrac{\epsilon}{1-\epsilon}), \quad \lambda_2 \sim \text{Exp}(\tfrac{\epsilon}{1-\epsilon}), \quad \tau \sim \text{Skellam}(\lambda_1, \lambda_2), \tag{15}$$

where the Skellam distribution [24] is defined as the marginal distribution over the difference $\tau := g_1 - g_2$ of two independent Poisson random variables $g_1 \sim \text{Pois}(\lambda_1)$ and $g_2 \sim \text{Pois}(\lambda_2)$.

Proof: A two-sided geometric random variable $\tau \sim \text{2Geo}(\epsilon)$ can be generated by taking the difference of two independent and identically distributed geometric random variables as follows:[2]

$$g_1 \sim \text{Geom}(1-\epsilon), \quad g_2 \sim \text{Geom}(1-\epsilon), \quad \tau := g_1 - g_2. \tag{16}$$

The geometric distribution is a special case of the negative binomial distribution, with shape parameter equal to one [15]; the negative binomial distribution can be augmented as a mixture of Poisson distributions with a gamma mixing distribution. We can therefore re-express equation 16 as follows:

$$\lambda_1 \sim \text{Gam}(1, \tfrac{\epsilon}{1-\epsilon}), \quad \lambda_2 \sim \text{Gam}(1-\epsilon), \quad g_1 \sim \text{Pois}(\lambda_1), \quad g_2 \sim \text{Pois}(\lambda_2), \quad \tau := g_1 - g_2. \tag{17}$$

Finally, a gamma distribution with shape equal to one is an exponential distribution, while the difference of two independent Poisson random variables is marginally a Skellam random variable [24].

C Proof of lemma 2

Lemma 2: Consider two Poisson random variables $y_1 \sim \text{Pois}(\lambda_1)$ and $y_2 \sim \text{Pois}(\lambda_2)$. Their minimum $m := \min\{y_1, y_2\}$ and their difference $\delta := y_1 - y_2$ are deterministic functions of y_1 and y_2. Then, if not conditioned on y_1 and y_2, the random variables m and δ can be marginally generated as follows:

$$\delta \sim \text{Skellam}(\lambda_1, \lambda_2), \quad m \sim \text{Bessel}\left(|\delta|, 2\sqrt{2\lambda_1\lambda_2}\right). \tag{18}$$

Proof:

$$\begin{aligned} P(y_1, y_2) &= \text{Pois}(y_1; \lambda_1)\,\text{Pois}(y_2; \lambda_2) \\ &= \frac{\lambda_1^{y_1}}{y_1!}\, e^{-\lambda_1}\, \frac{\lambda_2^{y_2}}{y_2!}\, e^{-\lambda_2} \\ &= \frac{(\sqrt{\lambda_1\lambda_2})^{y_1+y_2}}{y_1!\, y_2!}\, e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{(y_1-y_2)\,/\,2}. \end{aligned} \tag{19}$$

[2] See `https://www.youtube.com/watch?v=V1EyqL1cqTE`.

If $y_1 \geq y_2$, then

$$P(y_1, y_2) = \frac{(\sqrt{\lambda_1\lambda_2})^{y_1+y_2}}{I_{y_1-y_2}(2\sqrt{\lambda_1\lambda_2})\, y_1!\, y_2!}\, e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{(y_1-y_2)\,/\,2} I_{y_1-y_2}(2\sqrt{\lambda_1\lambda_2})$$
$$= \text{Bessel}(y_2;\, y_1 - y_2, 2\sqrt{\lambda_1\lambda_2})\, \text{Skellam}(y_1 - y_2; \lambda_1, \lambda_2); \tag{20}$$

otherwise

$$P(y_1, y_2) = \frac{(\sqrt{\lambda_1\lambda_2})^{y_1+y_2}}{I_{y_2-y_1}(2\sqrt{\lambda_1\lambda_2})\, y_1!\, y_2!}\, e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_2}{\lambda_1}\right)^{(y_2-y_1)\,/\,2} I_{y_2-y_1}(2\sqrt{\lambda_1\lambda_2})$$
$$= \text{Bessel}(y_1;\, y_2 - y_1, 2\sqrt{\lambda_1\lambda_2})\, \text{Skellam}(y_2 - y_1; \lambda_2, \lambda_1)$$
$$= \text{Bessel}(y_1; -(y_1 - y_2), 2\sqrt{\lambda_1\lambda_2})\, \text{Skellam}(y_1 - y_2; \lambda_1, \lambda_2). \tag{21}$$

If

$$m := \min\{y_1, y_2\}, \quad \delta := y_1 - y_2, \tag{22}$$

then

$$y_2 = m, \quad y_1 = m + \delta \quad \text{if } \delta \geq 0 \tag{23}$$
$$y_1 = m, \quad y_2 = m - \delta \quad \text{otherwise} \tag{24}$$

and

$$\begin{vmatrix} \frac{\partial y_1}{\partial m} & \frac{\partial y_1}{\partial \delta} \\ \frac{\partial y_2}{\partial m} & \frac{\partial y_2}{\partial \delta} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix}^{\delta \geq 0} \begin{vmatrix} 1 & 0 \\ 1 & -1 \end{vmatrix}^{\delta < 0} = 1, \tag{25}$$

so

$$P(m, \delta) = P(y_1, y_2) \begin{vmatrix} \frac{\partial y_1}{\partial m} & \frac{\partial y_1}{\partial \delta} \\ \frac{\partial y_2}{\partial m} & \frac{\partial y_2}{\partial \delta} \end{vmatrix}$$
$$= \text{Bessel}(m; |\delta|, 2\sqrt{\lambda_1\lambda_2})\, \text{Skellam}(\delta; \lambda_1, \lambda_2). \tag{26}$$

D Preliminary experiments

We implemented our method for privatizing Bayesian inference for Poisson matrix factorization, as described in sections 3 and 4, and ran preliminary experiments using synthetic data. We generated synthetic data for $\epsilon \in \{0.1, 0.25, 0.5, 0.0\}$ according to the following generative process:

$$\theta_{dk} \sim \Gamma(0.25, 4), \quad \phi_{kv} \sim \Gamma(0.25, 4) \tag{27}$$
$$\mu_{dv} = \boldsymbol{\theta}_d^\top \boldsymbol{\phi}_v, \quad y_{dv} \sim \text{Pois}(\mu_{dv}) \tag{28}$$
$$\tau_{dv} \sim 2\text{Geo}(\epsilon), \quad \tilde{y}_{dv}^{(\pm)} := y_{dv} + \tau_{dv}. \tag{29}$$

Equation 28 is identical to equation 4, while equation 29 is identical to equation 6.

Given the privatized data $\tilde{Y}$, we ran our MCMC algorithm and then checked how well the posterior samples of the non-privatized data Y matched the ground truth. For most elements of Y, we found that the histogram of posterior samples was tightly concentrated around the true value for reasonable privacy levels (i.e., $\epsilon > 0.1$). We show a representative sample for $\epsilon = 0.5$ in the left side of figure 1.

We also ran our MCMC algorithm conditioning on the true factor matrix Φ in addition to $\tilde{Y}$ to alleviate the "label-switching" problem. This enabled us to directly compare posterior samples of Θ to the ground truth. Because this procedure does not guarantee privacy, we used it only as a diagnostic: if the algorithm is effective at sampling Y, then conditioning on Φ should allow it to closely recover the true Θ. From the right side of figure 1, we can see that even for $\epsilon = 0.1$, the algorithm eventually recovers the ground truth, suggesting that it is indeed effective at sampling Y.

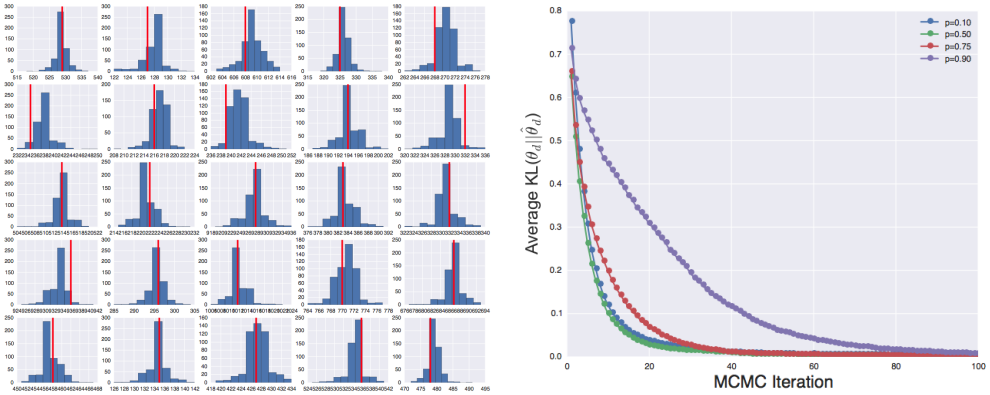

Figure 1: *Left:* Histograms of posterior samples of Y for 25 randomly selected elements for $\epsilon = 0.5$. In each plot, the red line indicates the true value of y_{dv}, which is known to us, but not the MCMC algorithm. *Right:* Recovery of Θ—i.e., the average KL divergence between posterior samples and the ground truth—conditioned on the true Φ for $p \in \{0.1, 0.25, 0.5, 0.9\}$, where $p = 1 - \epsilon$.