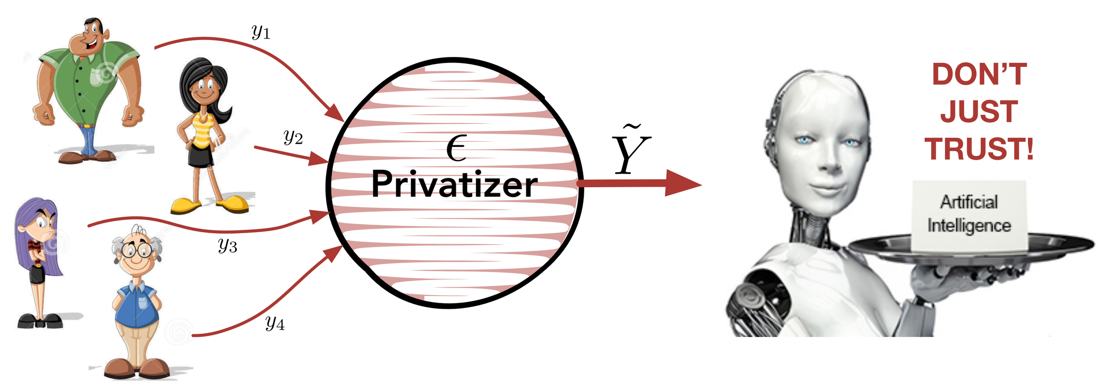
# Locally Private Bayesian Inference for Count Models

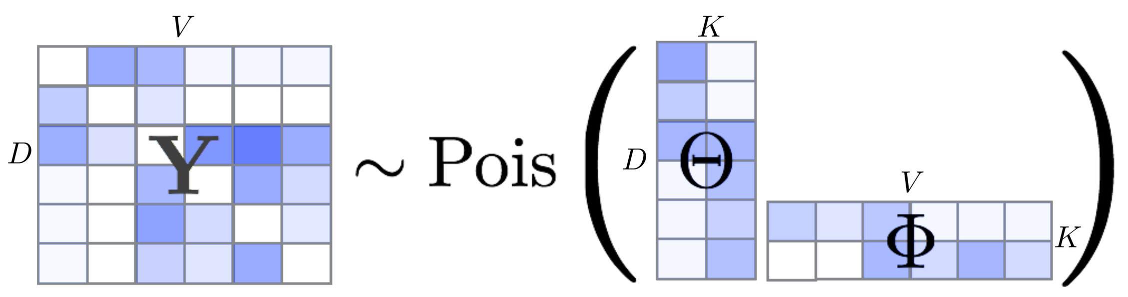
Aaron Schein — UMass Amherst
Zhiwei Steven Wu — Microsoft Research
Mingyuan Zhou — Univ. of Texas Austin
Hanna Wallach — Microsoft Research

## Local differential privacy

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




## Poisson factorization




$$y_{dv} \sim \mathrm{Pois}(\mu_{dv}) \qquad \mu_{dv} = \boldsymbol{\theta}_d^T \boldsymbol{\phi}_v$$

## Locally private Bayesian inference

**The usual story: non-private Bayesian inference:**

Define a model relating data $Y$ to latent variables $Z$:

$$P_{\mathcal{M}}(Y, Z)$$

Objective: Posterior of latent variables given data:

$$P_{\mathcal{M}}(Z \mid Y)$$

**Add local privacy:**

Define a privatizing noise model:

$$P_{\mathcal{E}}(\tilde{Y} \mid Y, \epsilon)$$

**New objective:** Posterior over latent variables given *privatized* data

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

If you can sample from $P_{\mathcal{M}}(Z \mid Y)$ and from $P_{\mathcal{E}}(Y \mid \tilde{Y}, \epsilon)$ you can approximate this posterior.

Four equivalent generative processes

### 1

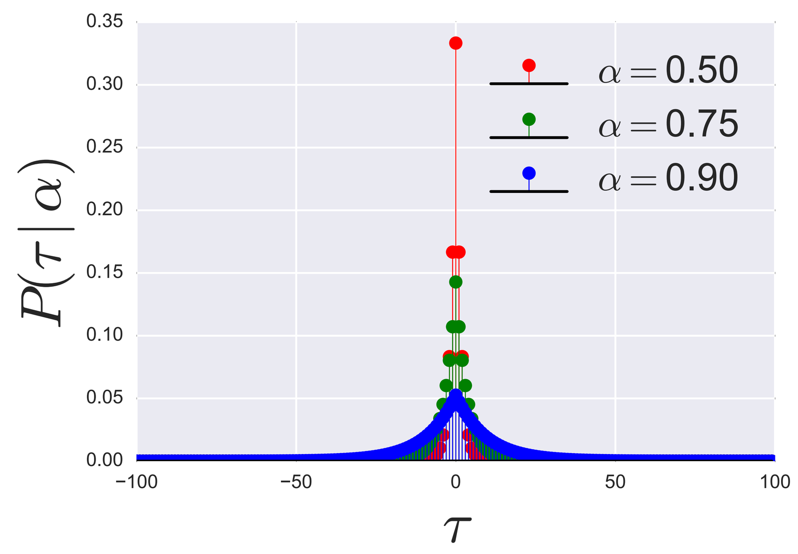
**Two-sided geometric** distribution

$$2\mathrm{Geom}(\tau; \alpha) = \frac{1-\alpha}{1+\alpha} \alpha^{|\tau|}$$

Introduced by Ghosh et al. (2012) as a privatizing mechanism with $\epsilon = -\ln \alpha$

Discrete analog to the Laplace distribution

Marginal distribution for the difference of two iid geometric random variables




$$\tau_{dv} \sim 2\mathrm{Geo}(\alpha)$$

$$y_{dv} \sim \mathrm{Pois}(\mu_{dv})$$

$$\tilde{y}_{dv}^{(\pm)} := y_{dv} + \tau_{dv}$$

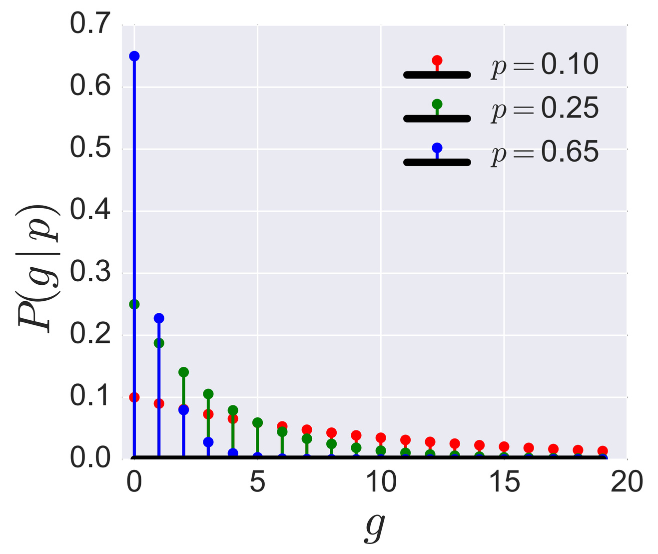
### 2

**Geometric** distribution

$$\mathrm{Geom}(g; p) = (1-p)^g\, p$$

Special case of the negative binomial distribution (for shape parameter equal to 1)

Discrete analog to the exponential distribution




$$g_{dv1} \sim \mathrm{Geom}(1-\alpha) \qquad g_{dv2} \sim \mathrm{Geom}(1-\alpha)$$

$$y_{dv} \sim \mathrm{Pois}(\mu_{dv})$$

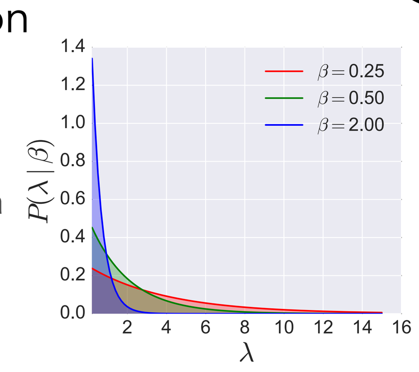
$$\tilde{y}_{dv}^{(+)} := y_{dv} + g_{dv1} \qquad \tilde{y}_{dv}^{(\pm)} := \tilde{y}_{dv}^{(+)} - g_{dv2}$$

### 3

**Exponential** distribution

$$\mathrm{Exp}(\lambda; \beta) = \beta e^{-\beta\lambda}$$

Special case of the gamma distribution (for shape parameter equal to 1)



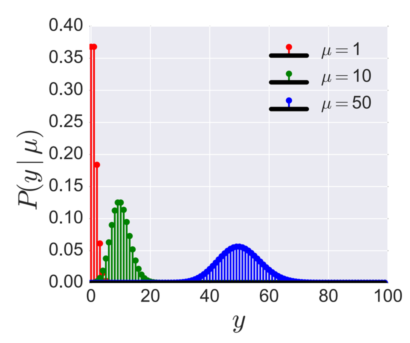

$$\lambda_{dv1} \sim \mathrm{Exp}\left(\tfrac{\alpha}{1-\alpha}\right) \qquad \lambda_{dv2} \sim \mathrm{Exp}\left(\tfrac{\alpha}{1-\alpha}\right)$$

**Poisson** distribution

$$\mathrm{Pois}(y; \mu) = \frac{\mu^y}{y!} e^{-\mu}$$

Mean equal to variance

Closed under addition




$$g_{dv1} \sim \mathrm{Pois}(\lambda_{dv1}) \qquad g_{dv2} \sim \mathrm{Pois}(\lambda_{dv2})$$

$$y_{dv} \sim \mathrm{Pois}(\mu_{dv})$$

$$\tilde{y}_{dv}^{(+)} := y_{dv} + g_{dv1} \qquad \tilde{y}_{dv}^{(\pm)} := \tilde{y}_{dv}^{(+)} - g_{dv2}$$

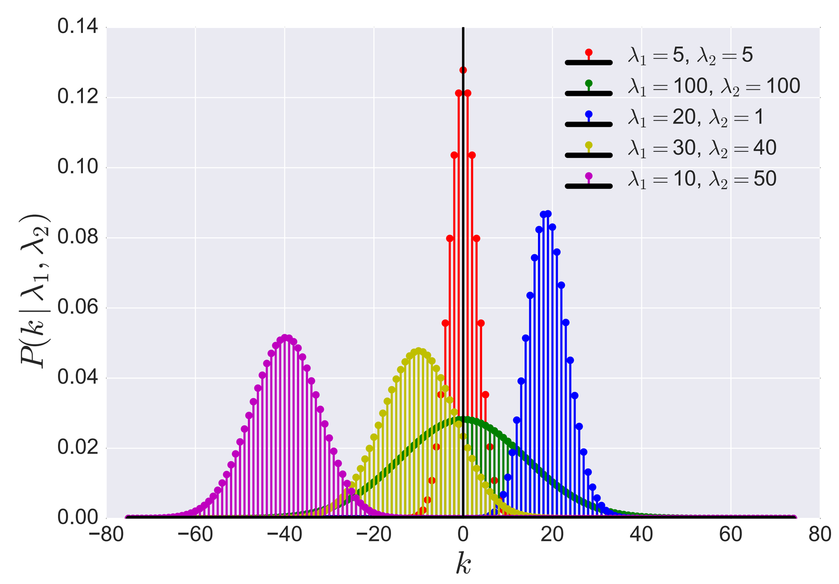
### 4

$$\lambda_{dv1} \sim \mathrm{Exp}\left(\tfrac{\alpha}{1-\alpha}\right) \qquad \lambda_{dv2} \sim \mathrm{Exp}\left(\tfrac{\alpha}{1-\alpha}\right)$$

**Skellam** distribution

$$\mathrm{Skell}(k; \lambda_1, \lambda_2) = \frac{\left(\frac{\lambda_1}{\lambda_2}\right)^{k/2} I_{|k|}(2\sqrt{\lambda_1 \lambda_2})}{e^{(\lambda_1 + \lambda_2)}}$$

Marginal distribution for the difference of two iindependent Poisson random variables




$$\tilde{y}_{dv}^{(\pm)} \sim \mathrm{Skell}(\lambda_{dv1} + \mu_{dv},\ \lambda_{dv2})$$

## MCMC Inference

*Lemma 2:* Consider two Poisson random variables $y_1 \sim \mathrm{Pois}(\lambda_1)$ and $y_2 \sim \mathrm{Pois}(\lambda_2)$. 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 $\delta$ can be marginally generated as follows:

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

**Gibbs sampling.** The input to this algorithm is the privatized data set $\tilde{Y}^{(\pm)}$. Assuming that $\tilde{y}_{dv}^{(\pm)} \sim \mathrm{Skellam}(\lambda_{dv1} + \mu_{dv}, \lambda_{dv2})$, 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:

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

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

$$y_{dv}^{(+)} := m_{dv}, \quad g_{dv2} := y_{dv}^{(+)} - y_{dv}^{(\pm)} \quad \text{if } y_{dv}^{(\pm)} \le 0 \tag{10}$$

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

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

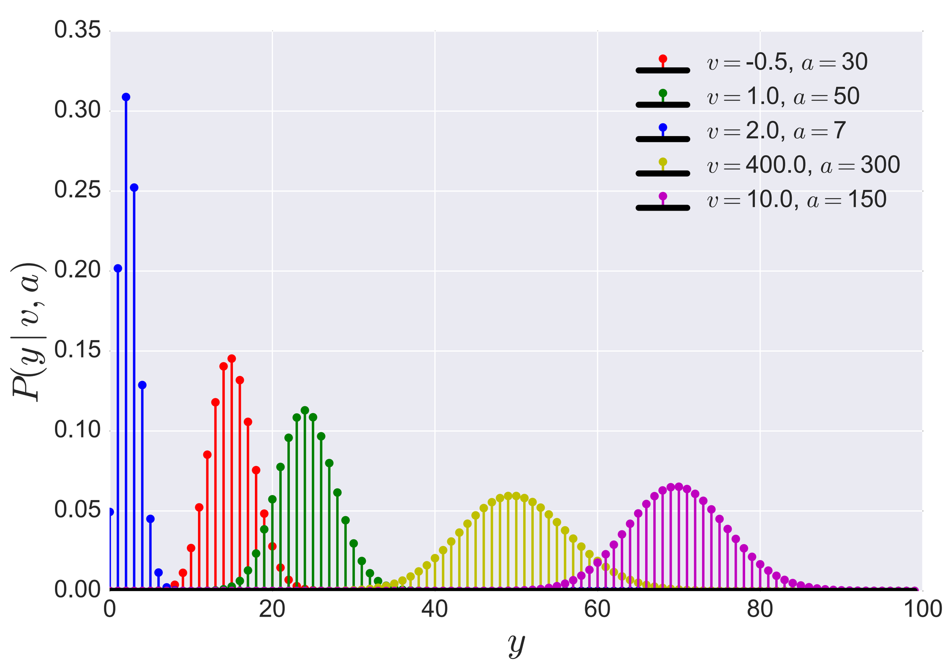
$$(y_{dv} \mid -) \sim \mathrm{Binom}\left(y_{dv}^{(+)}, \tfrac{\mu_{dv}}{\mu_{dv} + \lambda_{dv1}}\right) \tag{12}$$

Equations 9 through 12 constitute a method for sampling a value of $y_{dv}$ from $P_{\mathcal{M},\mathcal{E}}(y_{dv} \mid \tilde{y}_{dv}, \mu_{dv}, \epsilon)$.

**Bessel** distribution

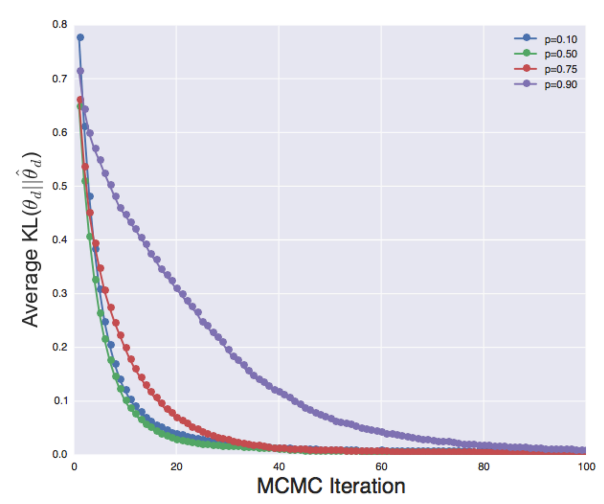
$$\mathrm{Bes}(y; v, a) = \frac{\left(\frac{a}{2}\right)^{2y+v}}{y!\,\Gamma(y+v+1) I_v(a)}$$

"Underdispersned Poisson"
(variance always less than mean)




## Preliminary results

**Experiment 1:** We generate synthetic data from the above model. Then, clamping one factor matrix to its true value, we perform posterior inference over all other variables .We calculate the average KL divergence over Gibbs iteration of the columns of the other factor matrix from the true value.




**Experiment 2:** We generate synthetic data from the above model. We run posterior inference over all parameters. For 25 random entries in the data matrix, we plot the histogram of posterior samples of the true underlying count (shown in red).