Two-sided geometric distribution

\[ 2\text{Geom}(\tau; \alpha) = \frac{1 - \alpha}{1 + \alpha} \]

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

Geometric distribution

\[ \text{Geom}(g; p) = (1 - p)^{g - 1} p \]

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

Discrete analog to the exponential distribution

Poisson factorization

Two-sided geometric Poisson factorization

\[ y_{dv} \sim \text{Pois}(\mu_{dv}); \quad \mu_{dv} = \theta_d^T \phi_v \]

Locally private Bayesian inference

The usual story: non-private Bayesian inference:

Define a model relating data \( Y \) to latent variables \( Z \): \( P_M(Y; Z) \)

Objective: Posterior of latent variables given data:

\[ P_M(Z \mid Y) \]

New objective: Posterior over latent variables given privatized data

\[ P_{M,\epsilon}(Z \mid \tilde{Y}, \epsilon) = \mathbb{E}_{P_M(Z \mid Y)} \left[ P_M(Z \mid Y) \right] = \int dY P_M(Z \mid Y) P_{\epsilon}(Y \mid \tilde{Y}, \epsilon) \]

If you can sample from \( P_M(Z \mid Y) \) and from \( P_{\epsilon}(Y \mid \tilde{Y}, \epsilon) \)

you can approximate this posterior.

Bessel distribution

\[ \text{Bes}(g; v, a) = \frac{\left( \frac{g}{v} \right)^{2v + w}}{y^{1/2} \Gamma(2v + 1) \Gamma(a)} \]

"Underdispersed Poisson" (variance always less than mean)

MCMC Inference

Gibbs sampling. The input to this algorithm is the privatized data set \( Y^{\epsilon/2} \). Assuming that \( y_{dv}^{(+)} \sim \text{Skell}(\\lambda_{dv1}, 2) \), \( y_{dv}^{(-)} \sim \text{Skell}(\\lambda_{dv2}, 2) \) as described in the previous section, we can express \( y_{dv}^{(+)} \) explicitly as the difference between two latent non-negative counts \( y_{dv}^{(+)} = \tilde{y}_{dv}^{(+)} - \tilde{y}_{dv}^{(-)} \). We further define the minimum of these non-negative counts to be \( m_{dv} = \min(\tilde{y}_{dv}^{(+)}, \tilde{y}_{dv}^{(-)}) \). Given randomly initialized factor matrices, we can sample a value of \( m_{dv} \) from its conditional posterior, which is a Bessel distribution:

\[ m_{dv} \sim \text{Bessel}(v, 2\sqrt{\\lambda_{dv1} + \lambda_{dv2}}) \]

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 iterations 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).