

Bayesian Statistics with the Zipf-Polylog

github.com/adbd-upc/Bayes-Zipf-Polylog/SEIO-2025/

Víctor Peña, Ariel Duarte-López, Marta Pérez-Casany

Departament d'Estadística i Investigació Operativa, UPC

SEIO 2025 - Lleida



The Zipf-Polylog Family (Valero et al., 2022)

PMF

$$p(y \mid \alpha, \beta) = \frac{\beta^y / y^\alpha}{\text{Li}_\alpha(\beta)}, \quad y \in \{1, 2, \dots\}, \quad \text{Li}_\alpha(\beta) = \sum_{k=1}^{\infty} \frac{\beta^k}{k^\alpha}.$$

Parameter space

$$\Theta = \{0 < \beta < 1, \alpha \in \mathbb{R}\} \cup \{\beta = 1, \alpha > 1\}.$$

Special case $\beta = 1, \alpha > 1$ recovers pure Zipf: $p(y \mid \alpha) \propto 1/y^\alpha$.

Moments

- If $0 < \beta < 1$: $\mathbb{E}[Y^k] < \infty$ for any $\alpha \in \mathbb{R}$.
- If $\beta = 1$, $\mathbb{E}[Y^k] < \infty \Leftrightarrow \alpha > k + 1$

Implemented in R within `library(zipfextR)`

Model Geometry

Fisher information

$$I_{\alpha\alpha} = \text{Var}[\log Y], \quad I_{\alpha\beta} = -\text{Cov}(\log Y, Y)/\beta, \quad I_{\beta\beta} = \text{Var}[Y]/\beta^2$$

$$I(\alpha, \beta) = \begin{bmatrix} I_{\alpha\alpha} & I_{\alpha\beta} \\ I_{\alpha\beta} & I_{\beta\beta} \end{bmatrix}.$$

KL-divergence

Let $\varphi = (\alpha, \beta)$, for a small $\Delta\varphi = (\Delta\alpha, \Delta\beta)$:

$$D_{\text{KL}}(p_\varphi \parallel p_{\varphi+\Delta\varphi}) \approx \frac{1}{2} \left[I_{\alpha\alpha} (\Delta\alpha)^2 + 2 I_{\alpha\beta} \Delta\alpha \Delta\beta + I_{\beta\beta} (\Delta\beta)^2 \right]$$

Also $D_{\text{KL}}(p_\varphi \parallel p_{\varphi+\Delta\varphi}) \approx$ Fisher-Rao metric (replace Δ by differentials).

Numerical Instability

$$D_{\text{KL}}(p_{\varphi} \parallel p_{\varphi+\Delta\varphi}) \approx \frac{1}{2} \left[I_{\alpha\alpha} (\Delta\alpha)^2 + 2 I_{\alpha\beta} \Delta\alpha \Delta\beta + I_{\beta\beta} (\Delta\beta)^2 \right].$$

As $\beta \rightarrow 1^- \dots$

- $I_{\alpha\alpha} = \text{Var}[\log Y] \rightarrow \infty$, if $\alpha \leq 1$,
- $I_{\alpha\beta} = \text{Cov}(\log Y, Y) \rightarrow \infty$, if $\alpha \leq 2$,
- $I_{\beta\beta} = \text{Var}[Y] \rightarrow \infty$, if $\alpha \leq 3$.

Model is “unstable” numerically if $\beta \approx 1$ and $\alpha \geq 3$

Zipf-Polylog as an exponential family

Recall that

$$p(y \mid \alpha, \beta) = \frac{\beta^y / y^\alpha}{\text{Li}_\alpha(\beta)}.$$

If we define

$$\theta_1 = \log \beta, \quad \theta_2 = -\alpha, \quad \psi(\theta_1, \theta_2) = \log \text{Li}_{-\theta_2}(e^{\theta_1}).$$

then

$$\log p(y \mid \alpha, \beta) = \theta_1 y + \theta_2 \log y - \psi(\theta_1, \theta_2).$$

Therefore,

2-parameter exponential family with sufficient statistic $(y, \log y)$

Conjugate Prior & Flat Prior Special Case

From the previous slide:

$$\log p(y \mid \alpha, \beta) = \theta_1 y + \theta_2 \log y - \psi(\theta_1, \theta_2).$$

It's an exponential family, so there exists a conjugate prior:

$$\log \pi(\theta_1, \theta_2) \propto \theta_1 \eta_1 + \theta_2 \eta_2 - \nu \psi(\theta_1, \theta_2).$$

In terms of α and β , and exponentiating back

$$\pi(\alpha, \beta) \propto \beta^{\eta_1} e^{-\eta_2 \alpha} \text{Li}_\alpha(\beta)^{-\nu}.$$

- *Flat prior* (improper) corresponds to $\eta_1 = \eta_2 = \nu = 0$.
- Under the flat prior, the posterior is proper *iff* at least one $y_i > 1$.

Application: URV Mail Data

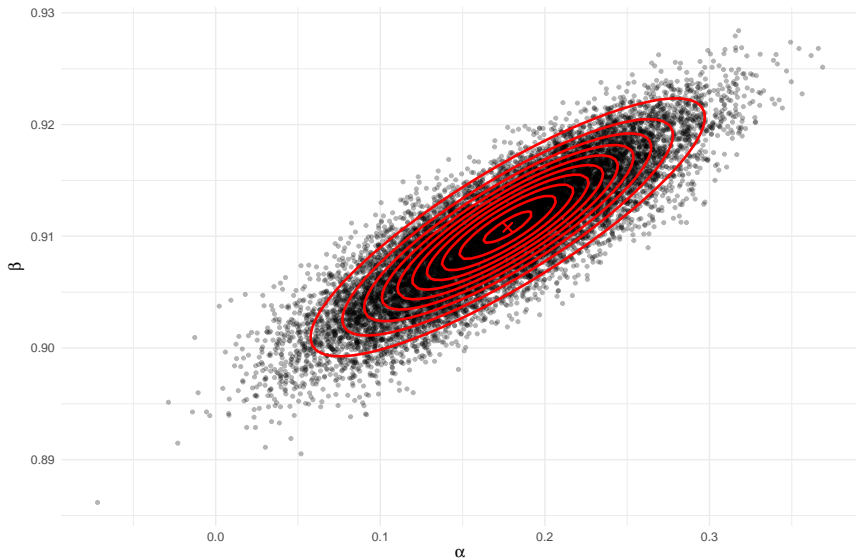
- Network of 1133 mail users at Universitat Rovira i Virgili.
- An edge is created between two nodes if user A sends an email to user B and user B sends an email to user A (5451 edges).
- We model the number of edges that a given node has (how many “contacts” each user has).
- We put a flat prior and run a slice sampler; we compare the results to the Bernstein von Mises approximation

$$\alpha, \beta \mid \text{data} \approx \mathcal{N}_2[(\hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}}), I^{-1}(\hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}})/n]$$

Estimating the Posterior Distribution

Slice Sampler Draws vs BvM Gaussian Contours

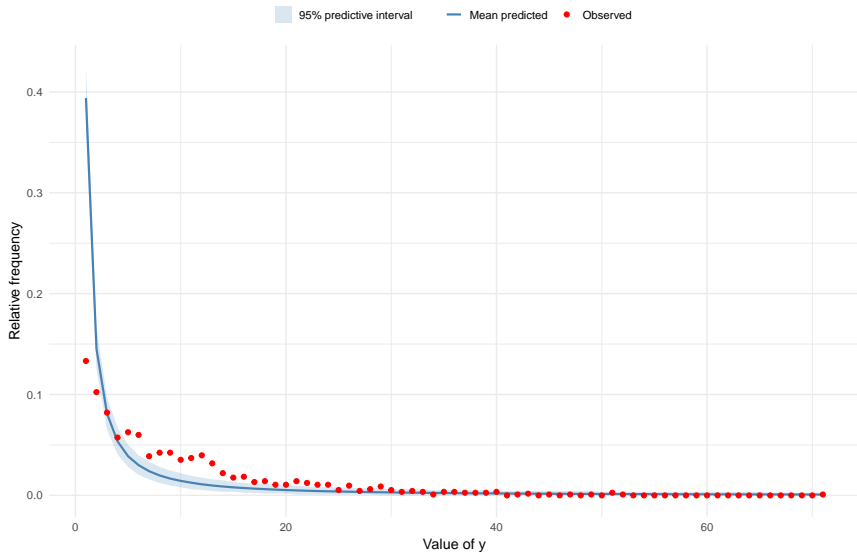
Red cross is the MLE



Zipf

Posterior–Predictive Check: Zipf Relative Frequencies

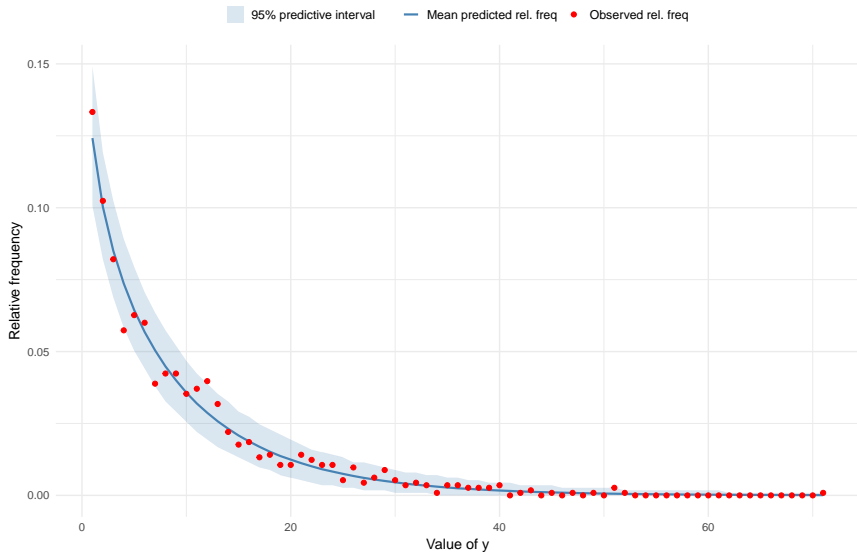
Coverage: 67.6% of observed values within 95% bands



Zipf-Polylog

Posterior-Predictive Check: Relative Frequencies

Coverage: 97.2% of observed values within 95% bands



Conclusions & Future Work

- Zipf-Polylog is useful for modeling degree sequences; offers additional (needed) flexibility with respect to Zipf.
- Bayesian-frequentist reconciliation through Bernstein von Mises
- Working on a Bayesian test for Zipf vs interior of parameter space of Zipf-Polylog.
- It's an exponential family, so we could build GLMs based on the Zipf-Polylog.
- Interesting geometry we could study through Fisher-Rao metric, scalar curvature (here, it's just a determinant), etc. Could also compute geodesics and other quantities.

Thanks!

ADBD GitHub repo:

<https://github.com/adbd-upc>

Funding: Project PID2023-148158OB-I00 awarded by MCIU and AEI



MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



PI: Pedro Delicado (UPC)

References I

Valero, J., M. Pérez-Casany, and A. Duarte-López (2022). The Zipf-Polylog distribution: Modeling human interactions through social networks. *Physica A: Statistical Mechanics and its Applications* 603, 127680.