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

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

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

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

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

KL-divergence

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

$$D_{\mathrm{KL}}(p_\varphi \,\|\, p_{\varphi+\Delta\varphi}) \approx \frac{1}{2}\Big[I_{\alpha\alpha}\,(\Delta\alpha)^2 + 2\,I_{\alpha\beta}\,\Delta\alpha\,\Delta\beta + I_{\beta\beta}\,(\Delta\beta)^2\Big]$$

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

Numerical Instability

$$D_{\mathrm{KL}}(p_\varphi \,\|\, p_{\varphi+\Delta\varphi}) \approx \frac{1}{2}\Big[I_{\alpha\alpha}\,(\Delta\alpha)^2 + 2\,I_{\alpha\beta}\,\Delta\alpha\,\Delta\beta + I_{\beta\beta}\,(\Delta\beta)^2\Big].$$

As $\beta \to 1^-$...

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

$\boxed{\text{Model is “unstable” numerically if } \beta \approx 1}$

Zipf-Polylog as an exponential family

Recall that

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

If we define

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

then

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

Therefore,

$$\boxed{\text{2-parameter exponential family with sufficient statistic } (y, \log y)}$$

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

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

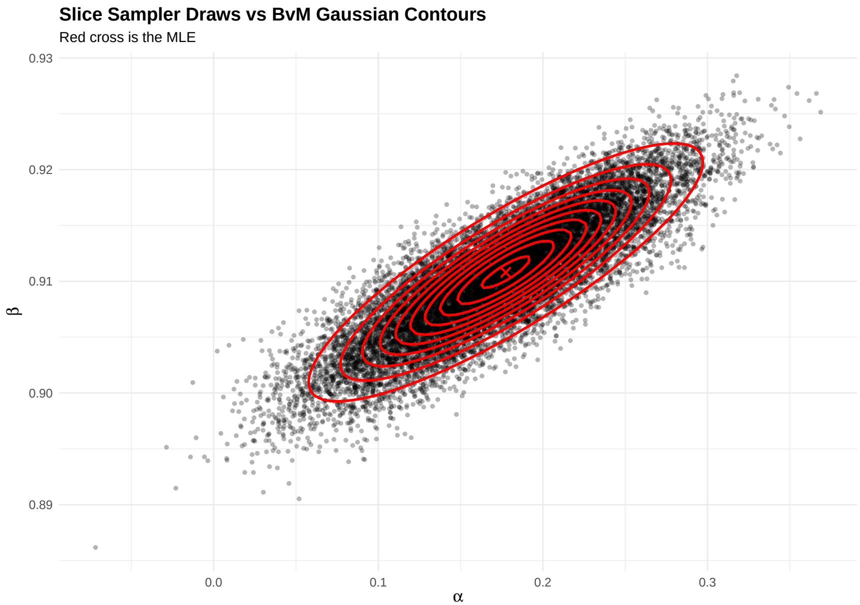
- *Flat prior* (improper) corresponds to $\eta_1 = \eta_2 = \nu = 0$.
- Under the flat prior, the posterior is proper *iff* at least one $y_i > 1$.

Application: URV Mail Data

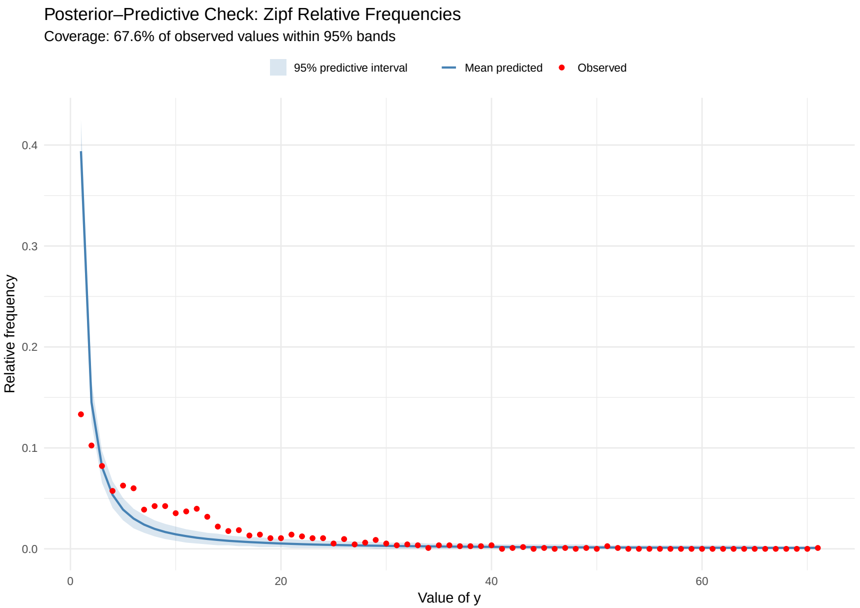
- Network of 1133 mail users at Universitat Rovira i Virgili.
- An edge is created between two nodes if user A sends an email to user B and user B sends an email to user A (5451 edges).
- We model the number of edges that a given node has (how many "contacts" each user has).
- We put a flat prior and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

$$\alpha, \beta \mid \texttt{data} \approx N_2[(\widehat{\alpha}_{\mathrm{ML}}, \widehat{\beta}_{\mathrm{ML}}), I^{-1}(\widehat{\alpha}_{\mathrm{ML}}, \widehat{\beta}_{\mathrm{ML}})/n]$$

Estimating the Posterior Distribution

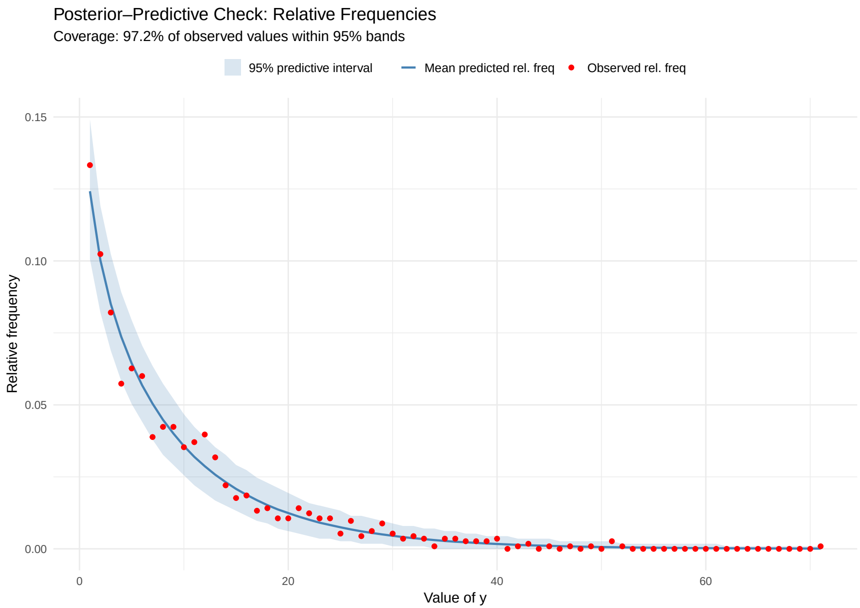


Zipf



Posterior–Predictive Check: Zipf Relative Frequencies
Coverage: 67.6% of observed values within 95% bands
95% predictive interval
Mean predicted
Observed
Relative frequency
0.4
0.3
0.2
0.1
0.0
0
20
40
60
Value of y

Zipf-Polylog



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

AGENCIA
ESTATAL DE
INVESTIGACIÓN

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.