Discrete Zipf-PSS-INAR(1) processes: Properties and Applications

Marta Pérez-Casany, Ariel Duarte-López and Manuel González-Scotto

Universitat Politècnica de Catalunya and Instituto Superior Técnico de Lisboa

ENBIS Conference, Florence Sep. 6-10 2026

Contents

Definition (Zipf):

If X is a r.v. with a Zipf distribution with parameter $\alpha > 1$, then its probability mass function (PMF) is equal to:

$$P(X = x) = \frac{x^{-\alpha}}{\zeta(\alpha)}, \; x \geq 1, \text{ and } \alpha > 1,$$

being $\zeta(\alpha)$ the Riemann zeta function.

It is linear in log-log scale, because

$$\log(P(X = x)) = -\alpha \log(x) - log(\zeta(\alpha)).$$

1. The Zipf and Zipf-PSS distributions

Definition (Stopped-Sum):

A r.v. Y is said to follow a SS distribution iff it there exist a sequence of i.i.d. r.v.'s $(X_i)_i$ such that

$$Y = X_1 + X_2 + X_3 + \cdots + X_N,$$

where N follows a given discrete distribution, **primary distribution**, independent of each one of the X_i, known as **secondary distribution**.

The pfg of a Stopped-Sum distribution is equal to:

$$G(H(z)).$$

This makes them known as *Compound distributions*.

1. The Zipf and Zipf-PSS distributions

Definition (Zipf-PSS):

It is said that a r.v. Y follows a Zipf-PSS distribution with parameters $(\lambda, \alpha) \in \mathbb{R}^+ \times (1, +\infty)$, iff it is a SS distribution with primary distribution a Poisson(λ), and a $Zipf(\alpha)$ as secondary distribution.

if $Y \sim Zipf - PSS(\lambda, \alpha)$, then

$$pgf_Y(z) = \exp\left(\lambda\left(\frac{Li_\alpha(z)}{\zeta(\alpha)} - 1\right)\right), \; z \in (-1, 1).$$

Property: As $\alpha \to +\infty$, $Li_\alpha(z) \to z$ and $\zeta(\alpha) \to 1$, and thus,

$$\text{Zipf-PSS}(\lambda, \alpha) \to \text{Po}(\lambda).$$

Moments

$$E(Y) = \lambda \frac{\zeta(\alpha-1)}{\zeta(\alpha)}, \ \alpha > 2 \quad \text{and} \quad Var(Y) = \lambda \frac{\zeta(\alpha-2)}{\zeta(\alpha)}, \ \alpha > 3.$$

Dispersion Index

$$ID_{Zipf-PSS} = \frac{Var(Y)}{E(Y)} = \frac{\zeta(\alpha-2)}{\zeta(\alpha-1)} > 1 \Longrightarrow \textbf{Overdispersed}$$

Zero-inflation index

$$I_{zero} = 1 + \frac{\log(p_0)}{\mu} = 1 + \frac{ln(e^{-\lambda})}{\lambda \frac{\zeta(\alpha-1)}{\zeta(\alpha)}} = 1 - \frac{\zeta(\alpha)}{\zeta(\alpha-1)} > 0 \Longrightarrow \textbf{Zero-inflated}.$$

2. The INAR(1) process

Definition (INAR(1) process):

An INAR(1) process is a process $(Y_t)_{t\in\mathbb{Z}}$ that follows the recursion:

$$Y_t = \beta \circ Y_{t-1} + \epsilon_t, \quad t \in \mathbb{Z},$$

where

- $(\epsilon)_{t\in\mathbb{Z}}$ is an i.i.d. process in $\mathbb{N}_{\mathbf{0}}$, with **finite mean and variance**, known as **innovations's process**, and
- $\beta \circ Y$ is a **Binomial thinning operator**.

Definition (Zipf-PSS-INAR(1) innovation process):

A Zipf-PSS-INAR(1) innovation process is an INAR(1) process with innovations $\epsilon_t \sim Zipf - PSS(\lambda, \alpha)$ $\forall$ t, and β as a thinning parameter.

Possible real situation

Assume that Y_t measures the number of people at time t in a given check-in lane.

Assuming $\epsilon_t \sim Zipf - PSS(\lambda, \alpha)$ means that:

- i) it exists a random number of families, $N \sim Po(\lambda)$, that arrive at the check in of an airport from $t-1$ to t.
- ii) Each family contributes with a X_i number of members at the queue that comes from a Zipf distribution.

3. Zipf-PSS-INAR(1) innovation process

Properties:

- The arrivals distribution does not depend on β.
- It is a queuing process of the type Zipf-PSS / Geom / ∞.
- It tends to the Poisson-INAR(1) process when α increases.
- Index of dispersion of the marginal distribution

$$ID_Y = \frac{\sigma_Y^2}{\mu_Y} = \frac{1}{1+\beta}\left[\frac{\zeta(\alpha-2)}{\zeta(\alpha-1)} + \beta\right] > 1 \Longrightarrow \textbf{Overdispersion}$$

- Probability at zero of the marginal distribution

$$p_0 = \exp\left(\lambda \sum_{j=1}^{+\infty}\left(\frac{Li_\alpha(1-\beta^j)}{\zeta(\alpha)} - 1\right)\right) > 0 \Longrightarrow \textbf{zero-inflated}$$

4. Zipf-PSS-INAR(1) marginal process

Definition (Zipf-PSS-INAR(1) marginal process):

A Zipf-PSS-INAR(1) marginal process is the INAR(1) process with the $Zipf - PSS(\lambda, \alpha)$ as stationary marginal distribution and β as a thinning parameter.

Possible real situation:

Assume that $Y_t \sim Zipf - PSS(\lambda, \alpha)$ measures the number of individuals in a supermarket queue at time t. This means that:

i) The number of individuals in the queue doesn't depend on the survival parameter.

ii) The new arrivals to the queue depends on β.

4. Zipf-PSS-INAR(1) marginal process

Properties:

- The marginal distribution is **overdispersed** and **zero-inflated**.
- The innovation's distribution is also a PSS, and is **overdispersed** and **zero-inflated**.
- As larger (smaller) is the probability β, fewer (more) terms will appear in the sum of the innovation's distribution.
  - As β tends to one, it tends to the degenerate distribution at zero.
  - As β tends to zero, it tends to the pgf of the Zipf-PSS distribution.
- It is a discrete queuing model of the type G/Geom/ $+\infty$.

Example 1:

Daily number of downloads of the TEX editor from June 2006 to February 2007.

obs	Mean	Variance	ID	corr	p_0
266	2.37	7.26	3.06	0.26	0.28

Table: Tex Downloads: Length, mean, variance, ID and probability at zero.

$$1 + \frac{log(p_0)}{\overline{x}} = 1 + \frac{log(0.28)}{2.37} = 0.4628 > 0,$$

The data are overdispersed and zero-inflated.

5. Fitting count time series

Model	**Method**	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{r}$	$\hat{p}$	SRSS	AIC_C
PINAR(1)	ML	0.18	-	1.95	-	-	464.80	-
NBINAR(1)	ML	0.14	-	-	1.00	0.33	201.70	-
Zipf-PSS-INAR(1) Innovat.	CML	0.15	2.39	1.16	-	-	-	1092.728
		(0.04)	(0.12)	(0.11)				
Zipf-PSS-INAR(1) Innovat.	**CLS**	0.26	3.21	1.38	-	-	**155.23**	-
Zipf-PSS-INAR(1) Marg.	**CML**	0.16	2.47	1.48	-	-	-	**1090.48**
		(0.01)	(0.01)	(0.10)				
Zipf-PSS-INAR(1) Marg.	CLS	0.26	3.21	1.38	-	-	155.80	-

Table: Tex Downloads: Estimation method and parameter estimations for the Poisson, NB, Zipf-PSS innovations and Zipf-PSS marginal processes, jointly with their SRSS.

5. Fitting count time series

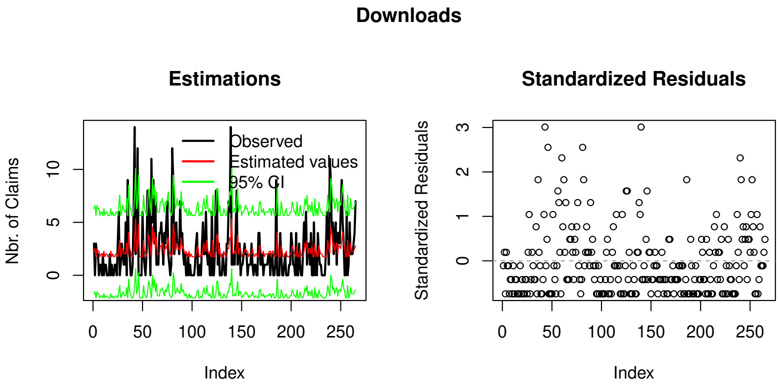

Figure: Tex Downloads: On the left hand side, observed and fitted values jointly with confidence intervals for the Zipf-PSS marginal process, and the CML parameter estimations. On the right hand side one has the plot of the raw residuals.

Example 2:

Number of claims performed by male claimants, between ages of 35 to 54 years reported to Richmond, BC workers compensation Board. The data can be downloaded from: `https://rdrr.io/github/projecttsinteger/tsintegerpackage/man/claims.html`

obs	Mean	Variance	ID	corr	p_0
120	6.13	11.7	1.91	0.56	0.00

Table: Claims 4: Length, mean, variance, ID and probability at zero.

The data are underdispersed and zero-deflated.

5. Fitting count time series

Model	**Meth.**	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{r}$	$\hat{p}$	SRSS	AIC$_C$
PINAR(1)	ML	0.47	-	5.19	-	-	40.50	-
NBINAR(1)	ML	0.47	-	-	536.00	0.99	40.02	-
Zipf-PSS-INAR(1) Innovat.	CML	0.52 (0.08)	4.47 (1.39)	4.44 (1.06)	-	-	48.07	579.89
Zipf-PSS-INAR(1) Marg.	**CML**	0.52 (0.01)	4.86 (0.01)	9.35 (0.53)	-	-	**26.95**	578.77

Table: Claims 3: Estimation method and parameter estimations for the Poisson, NB, Zipf-PSS innovations and marginal processes, jointly with their SRSS and the AIC$_C$. In brackets the standard errors of the parameter estimations of the Zipf-PSS processes.

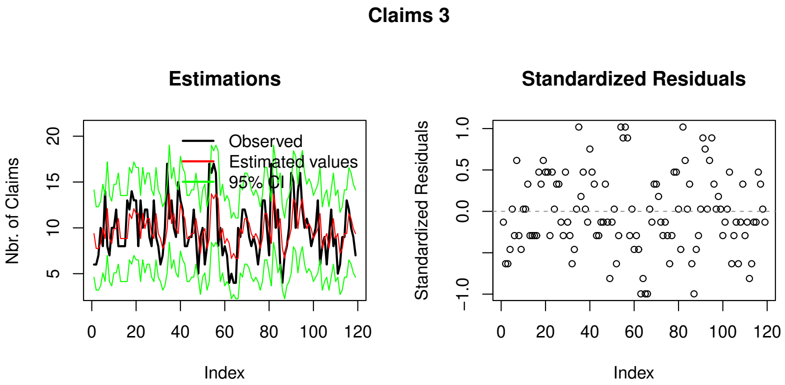

Figure: Claims 3: On the left hand side, observed and fitted values jointly with confidence intervals for the Zipf-PSS marginal process, and the CML parameter estimations. On the right hand side one has the plot of the standarized residuals.

Thank you very much for your attention!!!!