A Nonparametric Approach for Estimating Aggregate Loss

Gabriela González
York University

May 25, 2013

joint work with Hanna Jankowski

The Problem Setup:

$$S = \sum_{i=1}^{N} X_i$$

$X_i := \Omega_X \to (0, \infty)$ *Severity distribution*
$N := \Omega_N \to 0, 1, 2, ...$ $\perp X_i \forall i$ $in(1, ..., N)$
X_i *IID*

Proposition

- Assume a Zero Modified Discrete Log-Concave Distribution for the number of claims
- While gaining robustness, the proposed model will preserve efficiency
- Compare our approach with Panjer's Recursion method

The Problem Setup:

$$S = \sum_{i=1}^{N} X_i$$

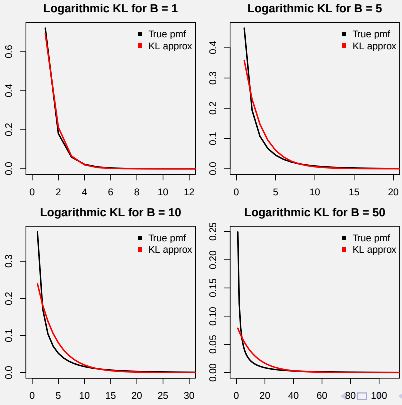
$X_i := \Omega_X \rightarrow (0, \infty)$ *Severity distribution*
$N := \Omega_N \rightarrow 0, 1, 2, ... \quad \perp X_i \forall i \quad in(1, ..., N)$
$X_i \quad IID$

Proposition

- Assume a Zero Modified Discrete Log-Concave Distribution for the number of claims
- While gaining robustness, the proposed model will preserve efficiency
- Compare our approach with Panjer's Recursion method

Popular Distributions for the number of claims N:

- Poisson(λ) $\rightarrow$ belong to LC class
- geometric(p)
- negative binomial(r,p)
- binomial(n,p)
- logarithmic(B) $\rightarrow$ not member of LC class



Panjer's Method:

Definition

$$k \frac{p_N(k)}{p_N(k-1)} = ak + b, \quad \text{for} \qquad k = 1, 2, 3, \ldots$$

Model Advantages

- The distribution of the Aggregate Loss is recursively computed
- Simple first approach for model identification

For a sample size of 10,000 with an $std = 0.01581139$

True Distribution	Unidentified	Poisson	Geometric	Binomial	Logarithmic	Negative Binomial
Negative Binomial	0.003	0.002	0.842	0.000	0.000	0.153
Binomial	0.032	0.689	0.000	0.279	0.000	0.000
Geometric	0.392	0.001	0.607	0.000	0.000	0.000
Poisson	0.119	0.845	0.001	0.022	0.000	0.013

The proposed model for N:

$$p_N(k) \quad = \quad \alpha\delta_0^k + (1-\alpha)\rho(k)$$

The proposed model for N:

$$p_N(k) = \alpha\delta_0^k + (1-\alpha)\rho(k)$$

where...

$$\delta_0^k = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\rho(k) = e^{\varphi(k)} \quad \text{for} \qquad k = 1, 2, 3, \ldots$$

$\alpha \in (0, 1)$

The proposed model for N:

$$p_N(k) \quad = \quad \alpha\delta_0^k + (1-\alpha)\rho(k)$$

where. . .

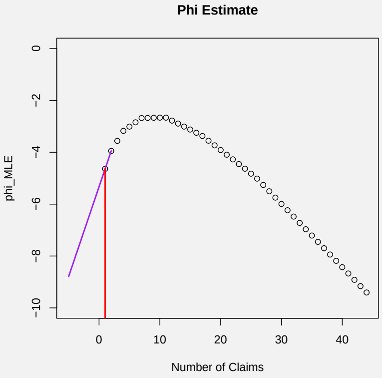
$$\delta_0^k \quad = \quad \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\rho(k) \quad = \quad e^{\varphi(k)} \quad \text{for} \qquad k = 1, 2, 3, \ldots$$

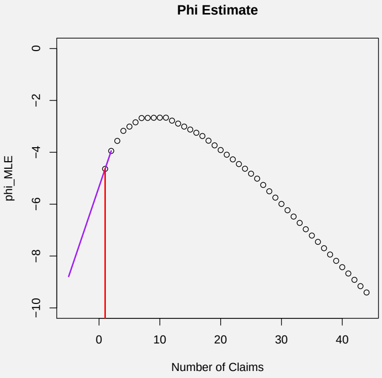
$\alpha \in (0, 1)$

Identifiable?

Identifiability

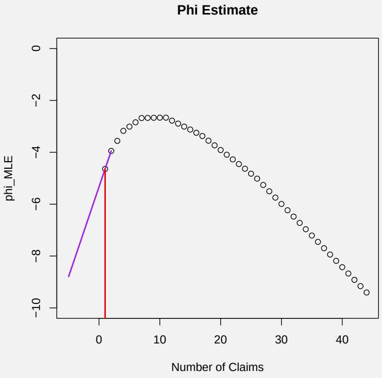


Identifiability



Restriction

Identifiability



Restriction

$$\rho(k) = \begin{cases} e^{\varphi(k)} & \text{for} \quad k = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

MLE Estimates:

$$\widehat{p_N}(k) \quad = \quad \widehat{\alpha}\delta_0^k + (1-\widehat{\alpha})\widehat{\rho}(k)$$

$$\widehat{\alpha} \quad = \quad \frac{\mathbb{I}(N_i(0))}{n} \quad \text{for} \qquad i = 1, \ldots, n$$

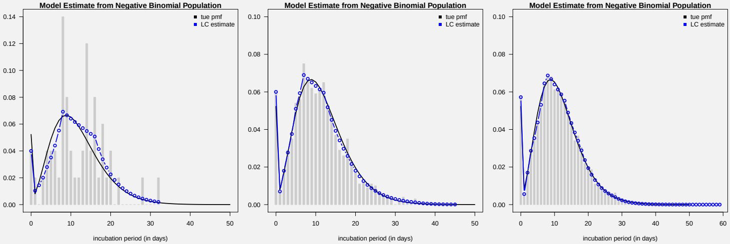
$$\widehat{\rho} \quad = \quad \begin{cases} e^{\widehat{\varphi(k)_{MLE}}} & \text{for} \qquad k = 1, 2, 3, \ldots \\ 0 & \qquad \text{otherwise} \end{cases}$$

MLE Estimates:

$$\widehat{p_N}(k) = \widehat{\alpha}\delta_0^k + (1-\widehat{\alpha})\widehat{\rho}(k)$$

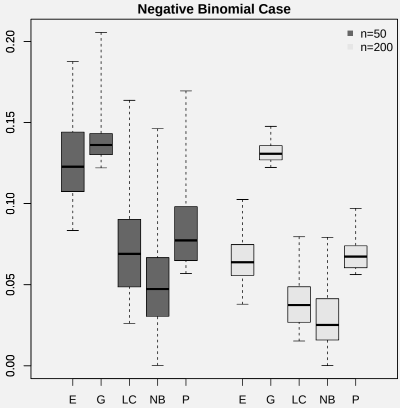
$$\widehat{\alpha} = \frac{\mathbb{I}(N_i(0))}{n} \quad \text{for} \quad i = 1, \ldots, n$$

$$\widehat{\rho} = \begin{cases} e^{\widehat{\varphi(k)_{MLE}}} & \text{for} \quad k = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

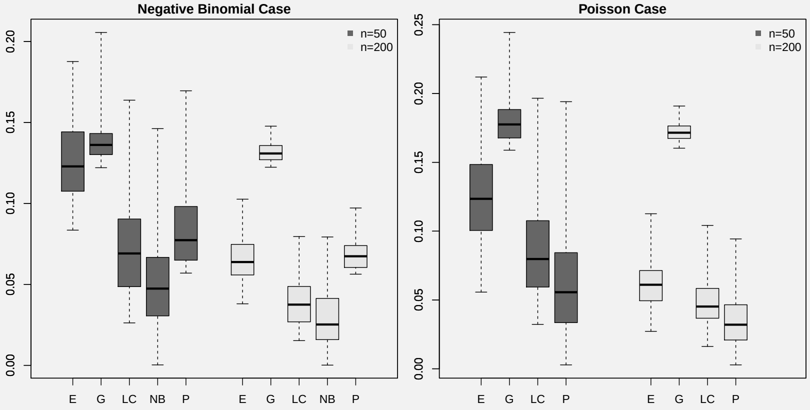


l_2 Distances from the true PMF:

l_2 Distances from the true PMF:



l_2 Distances from the true PMF:



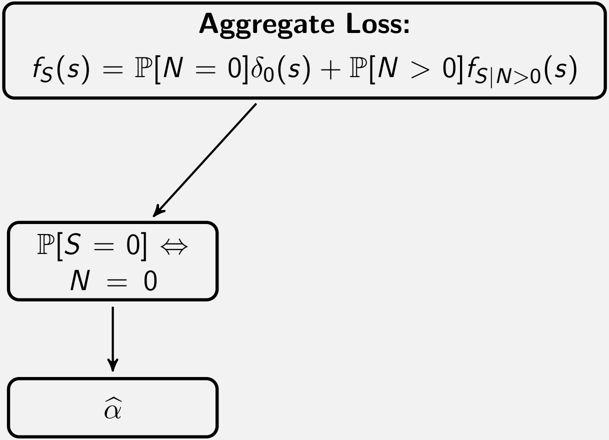
The Estimate of the Aggregate Loss:

The Estimate of the Aggregate Loss:

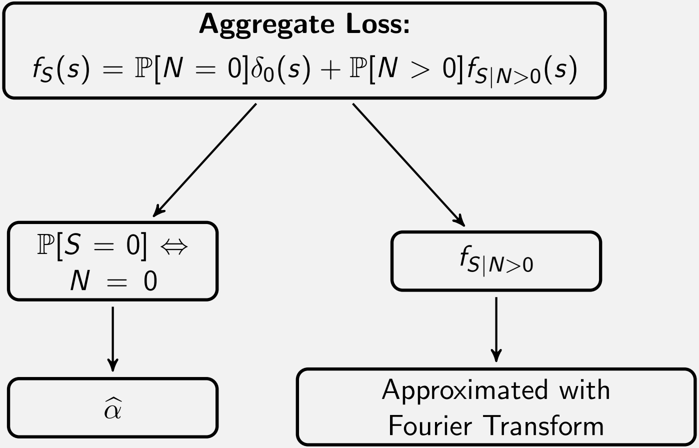
Aggregate Loss:

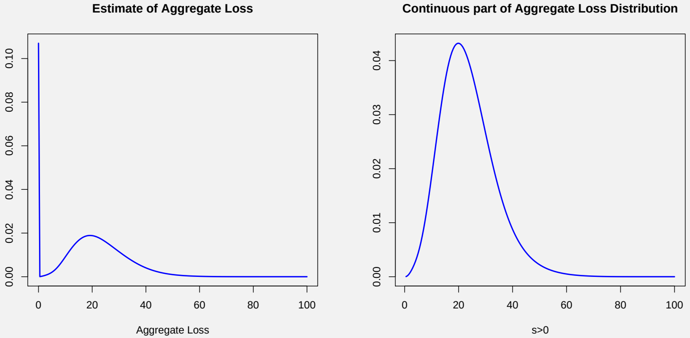
$$f_S(s) = \mathbb{P}[N = 0]\delta_0(s) + \mathbb{P}[N > 0]f_{S|N>0}(s)$$

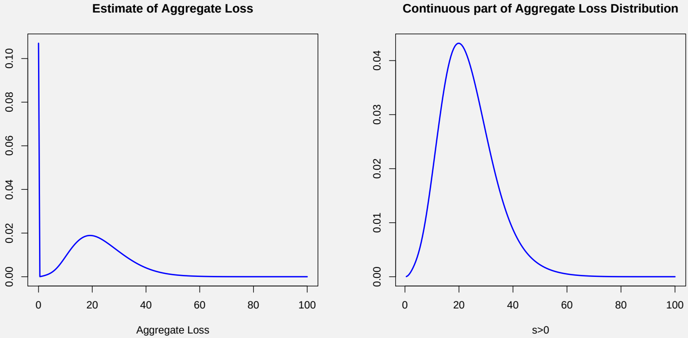
The Estimate of the Aggregate Loss:



The Estimate of the Aggregate Loss:




Estimate of Aggregate Loss
0.00
0.02
0.04
0.06
0.08
0.10
0
20
40
60
80
100
Aggregate Loss
Continuous part of Aggregate Loss Distribution
0.00
0.01
0.02
0.03
0.04
0
20
40
60
80
100
s>0



Future work

- Prove consistency of the proposed model
- Generalize the assumptions on the claim severity
- Provide a goodness of fit statistic under a discrete log-concave distribution

THANK YOU

$$f_{S|N>0}(s) = \int_{-\infty}^{\infty} \phi_S(-2\pi i t) e^{2\pi i t s} dt$$