A compound Poisson model with varying element sizes

J.F. Walhin*• and J. Paris*

∗ *Université Catholique de Louvain*

• *Secura Belgian Re*

Abstract

The aim of this note is to extend Bissel's paper (1972) to more general compound Poisson distributions. For the particular example given by Bissel (1972) we are able to distinguish between a pure Poisson and a compound Poisson part for the number of breaks. Therefore we are able to distinguish between the pure hasard breaks and the others.

Keywords

Compound Poisson distribution, recursions, convolution, maximum likelihood, pure hasard.

1. Introduction

In this note, we study the same numerical example than in Bissel (1972) with a more general distribution in order to get a better fit. Moreover we will be able to distinguish between the pure random part of the frequency and the non pure random part of it.

2. The data

The data are from Bissel (1972) and give the number of faults in rolls of textile fabric with varying sizes for the rolls :

Roll no.	Roll length	No. of faults		Roll no.	Roll length	No. of faults
1	551	6		17	543	8
2	651	4		18	842	9
3	832	17		19	905	23
4	375	9		20	542	9
5	715	14		21	522	6
6	868	8		22	122	1
7	271	5		23	657	9
8	630	7		24	170	4
9	491	7		25	738	9
10	372	7		26	371	14
11	345	6		27	735	17
12	441	8		28	749	10
13	895	28		29	495	7
14	458	4		30	716	3
15	642	10		31	952	9
16	492	4		32	417	2

3. The model

If $N(t)$ is the number of faults in a roll of length t (with $N(0) = 0$ naturally), it is a pure birth process. As described in Walhin and Paris (2000) the most general model showing overdispersion relatively to the Poisson which can be interpreted as a mixed and / or a compound Poisson process is obtained by

$$\begin{aligned} \Pi(0,t) &= \mathbb{P}[N(t)=0] = e^{-\delta t-\theta(t)} \\ \Pi(n,t) &= \mathbb{P}[N(t)=n] = (-1)^n \frac{t^n}{n!}\Pi^{(n)}(0,t) \end{aligned}$$

where δ is a non negative constant and $\theta(t)$ a positive function with a completely monotone derivative. So $N(t)$ is the sum of two independent components

$$N(t) = N_1(t) + N^*(t)$$

The first, $N_1(t)$, which is a pure Poisson process with mean δt represents the purely random defaults. The second, $N^*(t)$ which is a mixed and compound Poisson process represents the defaults related to the quality of the production.
The choice
$$\theta'(t) = \frac{p}{(1+ct)^a} \quad p > 0, \quad c > 0, \quad a \geq 0$$
initiated by Hofmann (1955) contains the particular cases $a = 0$ (Poisson), $a = 0.5$ (Poisson Inverse Gaussian), $a = 1$ (Negative Binomial), $a = 2$ (Polya-Aeppli), $a \to \infty, c \to 0, ac \to b$ (Neymann Type A) and so this model encompasses Bissel's (1972) model (Negative Binomial distribution) : $\delta = 0$, $a = 1$.

It is not difficult to show that the density of $N(t)$ can be given by a recursion:

$$\begin{aligned}
f_{N(t)}(0) &= e^{-\theta(t)-\delta t} \\
f_{N(t)}(1) &= \left(\delta t + \frac{pt}{(1+ct)^a}\right) e^{-\theta(t)-\delta t} \\
f_{N(t)}(x) &= \frac{1}{x}\left(\delta t + \frac{pt}{(1+c)^a}\right) f_{N(t)}(x-1) + \\
&\quad \frac{pt}{x(1+ct)^a} \sum_{i=2}^{x} \frac{1}{(i-1)!} \left(\frac{ct}{1+ct}\right)^{i-1} \frac{\Gamma(a+i-1)}{\Gamma(a)} f_{N(t)}(x-i) \quad , x \geq 2
\end{aligned}$$

4. The fit

In this section we apply maximum likelihood in order to find estimates for p, c, a and δ. Standard numerical maximization techniques allow these estimates. However it is necessary to be careful because the loglikelihood is not concave everywhere.

Type	δ	p	c	a	ac	l
Poisson	0	0.0151		0		-93.91
Negative Binomial	0	0.0153	0.0024	1		-88.01
Poisson Inverse Gaussian	0	0.0150	0.0037	0.5		-87.54
Hofmann	0	0.0150	0.0086	0.2402		-87.53
Neymann Type A	0	0.0151			0.0015	-88.02
Negative Binomial + Poisson	0.0097	0.0055	0.0077	1		-87.35
Poisson Inverse Gaussian + Poisson	0.0066	0.0083	0.0076	0.5		-87.35
Neymann Type A + Poisson	0.0117	0.0032			0.0099	-86.93

Our model has the following good properties

$$\begin{aligned}
(1) \qquad \mathbb{E}[N(t+1) - N(t) | N(t) = 0] &= \theta'(t) + \delta \\
(2) \qquad \lim_{t\to\infty} \mathbb{P}(N^*(t) = 0) &> 0 \quad \text{if } a > 1
\end{aligned}$$

Clearly we expect that a roll without default implies the smallest frequency of defaults in a longer roll. So we would like to have the smallest value for (1). It is interesting to note, that

with our model constrained to fixed ac, $\theta'(t)$ is minimized for the Neymann Type A choice (this is easily seen by a Taylor development).
Furthermore we expect that the quality of the production implies that it is possible to have infinite rolls without defaults. So (2) implies that $N^*(t)$ should have the property $a > 1$. Clearly we can not have infinite rolls without defaults because of the purely random effects. This is the reason why a Poisson component is introduced in our model. Finally, it is clear that from an industrial point of view, we expect that most of the defaults should occur randomly. The best fit, given by the Neymann Type A + Poisson shows very much this desirable property : 78.5% of the frequency comes from the Poisson component.

The preceding considerations combined with the maximum likelihood fit tell us that the model to be retained is Neymann Type A + Poisson.

5. Aggregate cost

It might be interesting to know the total cost for repairing the defaults. Assume that the cost for repairing a fault is given by a random variable X that is arithmetic. We are interested in

$$S = X_1 + \cdots + X_{N(t)}$$

where $N(t)$ is the number of faults in a roll of lenght t.
This kind of model is well known in actuarial sciences.
With the Hofmann's distribution, it is interesting to note that the evaluation of the distribution S does not need the evaluation of convolutions :

$$f_S(x) = \sum_{n=0}^{\infty} \mathbb{P}(N(t) = n) f_X^{*n}(x)$$

Instead a double recursion may be performed to find the distribution of S (see Walhin and Paris (2000) for details) :

$$\begin{aligned}
f_V(0) &= 1 - \frac{\theta(t - tf_X(0)) + \delta(t - tf_X(0))}{\theta(t) + \delta t} \\
f_V(x) &= \frac{1}{1 - \frac{ct}{1+ct} f_X(0)} \left[\frac{ct}{1+ct} \sum_{i=1}^{x} (1 + (a-2)\frac{i}{x}) f_X(i) f_V(x-i) \right. \\
&\qquad \left. + p_1 f_X(x) + (p_2 - \frac{ct}{1+ct}\frac{a}{2} p_1) f_X^{*2} \right] \quad , x \geq 1 \\
f_S(0) &= e^{-\theta(t)(1 - f_X(0))} \\
f_S(x) &= \frac{\theta(t) + \delta t}{x} \sum_{i=1}^{x} i f_V(i) f_S(x-i) \quad , x \geq 1
\end{aligned}$$

where

$$
\begin{aligned}
p_1 &= \frac{t\theta'(t) + \delta t}{\theta(t) + \delta t} \\
p_2 &= \frac{ct}{1+ct}\left(1 + \frac{a-2}{n}\right) p_1
\end{aligned}
$$

REFERENCES

Bissel, A., A negative binomial model with varying element sizes, Biometrika, 59 (1972) 435-441

Hofmann, M.,Uber zusammengesetzte Poisson-Prozesse und ihre Anwendungen in der Unfall versicherung, Bull. of the Swiss Actuaries, 55 (1955) 499-575

Kestemont, R.M. and Paris, J., Sur l'ajustement du nombre de sinistres, Bull. of the Swiss Actuaries, 85 (1985) 157-164

Walhin, J.F. and Paris, J., A large family of discrete and overdispersed probability laws, forthcoming

Jean-François Walhin*•
José Paris*
* Institut de Statistique
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve
Belgique e-mail : walhin@stat.ucl.ac.be
• Secura Belgian Re
Rue Montoyer, 12
B-1000 Bruxelles
Belgique