A Mixed Poisson Model with Varying Element Sizes

J.F. Walhin[*•] and J. Paris[†]

∗ *Institut des Sciences Actuarielles, Université Catholique de Louvain*

• *Secura*

† *Institut de Statistique, Université Catholique de Louvain*

Corresponding author :

Jean-François Walhin
Secura
Avenue des Nerviens, 9-31
B-1040 Bruxelles
Belgique
phone : + 32 2 504 82 22
fax : + 32 2 504 82 00
e-mail : jfw@secura-re.com

A Mixed Poisson Model with Varying Element Sizes

Abstract

The data of Bissell (1972a) are counts of the number of flaws in rolls of textile fabric of different lengths. Cox and Snell (1981) remark that "the probability distribution of the number of faults is of interest, especially in its relation to that expected if they occur at random at a fixed rate per unit of length". Given that the simple Poisson model is rejected due to overdispersion, we propose to analyse the data by means of a general family of discrete laws, indexed by a parameter that distinguishes several well-known distributions. Our methodology provides some insight into the quality of the production and a relatively straightforward way to evaluate the probability distribution of the aggregate cost of the defects, avoiding the evaluation of convolutions.

Keywords

Infinitely divisible mixed Poisson process, recursions, convolution, Hofmann process, maximum likelihood.

1 Introduction

When analysing the number of defects in an industrial production process where the events are counted on sample elements of equal magnitude, it is usual to fit the data with a particular distribution and to test the goodness of fit with a standard χ^2 test. Often, it is difficult to reject the proposed choice due to the small number of degrees of freedom. This approach does not give much information on the distribution of defects and hence on the quality of the production.

The problem is different whenever the data are reported on sample elements of different magnitude. Moreover, as indicated in Bissell (1972a) : "notably in industrial sampling, the records may contain only the element sizes and the number of events for each element". The data reported in that paper give the number of faults in rolls of textile fabric with varying length. This constitutes a rare example of data reported on variable element sizes; however, little information is available about the data. They contain only the length of the rolls (in meters) and the number of faults for each roll.

In industrial production, it is hoped that the defects are purely random. In this case, the number of defects is analysed by a homogeneous Poisson process for which the variance is equal to the mean. For Bissell's data, Hinde and Demetrio (1998) observed that the mean and the variance of the number of faults increase with the length of the roll and that overdispersion is present. Therefore, if we let $N(t)$ be the counting process of the number of defects in the interval $(0, t]$, with $N(0) = 0$, the index of overdispersion of the proposed model $\frac{\mathbb{V}ar N(t)}{\mathbb{E}N(t)}$ should be an increasing function of t.

The methodology we propose describes a natural and large family of discrete probability laws indexed by a parameter that is applicable to such variations in element sizes and to overdispersion. Fitting various members of this family reveals that the majority of defects are purely random and that the other defects correspond to a plausible situation, to be expected in practice.

Our model allows an easy way to obtain a prediction of the number of defects in elements of unequal length and an evaluation of the aggregate cost of the defects by recursive formulae, easier to use than classical convolution formulae.

The data of Bissell have been examined on several occasions. Bissell (1972a,b) proposes two formulations of the negative binomial distribution, which is a particular case of our model. Hill and Tsai (1988) used the data as an illustration of the efficiency of maximum quasilikelihood estimation in a negative binomial model reduced to one parameter. Azzalini et al. (1989) used the data as an illustration of nonparametric regression. Hinde and Demetrio (1998) used the data in a comparison of estimation procedures in an overdispersion model. Lindsey (1999) combined a simple loglinear Poisson regression for the counts and a normal probability law for the mixing distribution.

The rest of the paper is set out as follows : Section 2 describes the model. Section 3 proposes a parametric choice within the general framework. Section 4 gives the numerical application. Section 5 discusses the evaluation of aggregate cost and section 6 concludes.

2 The Model

Two formulations have been proposed to explain overdispersion relative to the Poisson probability law. The first is to choose a mixed Poisson probability law for which

$$\Pi(n,t) = \mathbb{P}[N(t) = n] = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} dU(\lambda),$$

where U is the mixing distribution function. Many models have been built by a particular choice of the mixing distribution but this function is not observable from the data and its estimation reveals some problems. We therefore prefer to follow a new avenue. We observe that knowledge of $\Pi(0,t) = \mathbb{P}[N(t) = 0] = \int_0^\infty e^{-\lambda t} dU(\lambda)$, which is the Laplace-Stieltjes transform of U, and thus a completely monotonic function of t, is equivalent to that of U, and that the probabilities $\Pi(n,t) = (-1)^n \frac{t^n}{n!} \Pi^{(n)}(0,t)$ are obtained directly by differentiation of $\Pi(0,t)$. In this formulation, integration is not necessary. Knowledge of $\Pi(0,t)$ identifies not only the mixing function U but also the probability law of $N(t)$, as the probability generating function of $N(t)$ is the function of u, given by $P(u,t) = \Pi(0, t(1-u))$.

The second formulation explaining overdispersion relative to the Poisson case consists of choosing a compound Poisson probability law that is always infinitely divisible. As the data do not allow us to distinguish between these models, we propose to assume that $N(t)$ is an infinitely divisible mixed Poisson process (see Grandell (1997)) that allows a twofold interpretation. This is simply obtained by restricting $\Pi(0,t)$ to the form $\Pi(0,t) = e^{-\eta(t)}$, where η is a Bernstein function ($\eta \geq 0$, $\eta^{'}$ completely monotonic, $\eta(0) = 0$), see Feller (1971).

The most general form of η is equivalent to $\eta^{'}(t) = \delta + \theta^{'}(t)$ where δ is a positive constant and $\theta^{'}$ is a completely monotonic function (see Berg and Forst (1975)). The probability generating function of the most general case is $\Pi(0, t(1-u)) = e^{-\delta t(1-u)} e^{-\theta(t(1-u))}$ indicating that, in the most general case, $N(t)$ is the sum of two independent components

$$N(t) = N_1(t) + N_2(t),$$

where

- $N_1(t)$ obeys a Poisson probability law with mean δt and describes the purely random defects.

- $N_2(t)$ obeys an infinitely divisible mixed Poisson probability law, which we interpret as being related to the quality of production.

Our model is related to the models described in Gupta et al. (2004) where an additive model and a multiplicative model are introduced. The multiplicative model corresponds to our mixed Poisson model whereas the additive model corresponds to the convolution of a Poisson distribution with another counting distribution. Our model makes sure that the retained distribution is an infinitely divisible mixed Poisson distribution. This happens for the convolution of a Poisson distribution and another distribution such that the convolution is itself a mixed Poisson distribution. In particular, our model shows that the solution is the sum of two (and not more) independent components, one of them being Poisson distributed.

The probability law of $N(t)$ is the convolution of the probability laws of $N_1(t)$ and $N_2(t)$. However, thanks to the extended Panjer's algorithm (see Panjer (1981) and Sundt and Jewell (1981)), we are able to evaluate the probability law of $N(t)$ by recursion. We have

$$\begin{aligned} \Pi(0,t) &= e^{-\delta t - \theta(t)}, \\ \Pi(1,t) &= (\delta t + t\theta'(t))\Pi(0,t), \\ \Pi(n,t) &= (\delta t + t\theta'(t))\frac{\Pi(n-1,t)}{n} + \sum_{i=2}^{n} \frac{(-1)^{i-1}}{n} \frac{t^i}{(i-1)!}\theta^{(i)}(t)\Pi(n-i,t) \quad , \quad n \geq 2. \end{aligned} \tag{1}$$

In our formulation, whenever you know the number of faults in $[0, t]$, you can obtain information for the future since the intensity of the process is $\mathbb{E}[N(t+h) - N(t)|N(t) = k] = h\frac{k+1}{t}\frac{\Pi(k+1,t)}{\Pi(k,t)}$. In particular, when $k = 0$,

$$\mathbb{E}[N(t+h) - N(t)|N(t) = 0] = h\eta'(t). \tag{2}$$

Remark : in order to build our model, it suffices to know the expression of $\Pi(0, t)$ or of $\eta'(t)$, but as the number of observed zero counts is not given, we are unable to estimate those functions directly. We are also unable to use a nonparametric method and to find an estimate for the mixing function U. There are two main reasons : on the one hand, the procedure initiated by Simar (1976) has not been extended to situations of data reported on elements of various sizes and on the other hand, the number of data is such that the result of Mc Kay (1996) applies : "on a sufficiently large class of densities (of U), any estimator must converge more slowly than $n^{-\alpha}$ for any positive α". Therefore it seems natural to choose the parametric option.

3 The Choice of the Model

In our formulation, it suffices to choose the function θ'. We take $\theta'(t) = \frac{p}{(1+ct)^a}$, a choice proposed by Hofmann (1955) and studied by Walhin and Paris (2002). The parameter a distinguishes among the probability laws for $N_2(t)$: $a = 0$ (Poisson), $a = 0.5$ (Poisson Inverse Gaussian, PIG for short), $a = 1$ (Negative Binomial, NB for short), $a = 2$ (Polya-Aeppli), $a \to \infty, c \to 0, ac \to b$ (Neyman Type A, NTA for short). Clearly δ and p are not identifiable when $a = 0$.

The evaluation of the probability function of $N(t)$ is given by formula (1) :

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

In the most general model we always have $\lim_{t\to\infty}\mathbb{P}[N(t)=0] = 0$ if $\delta > 0$. However if $\delta = 0$, we can have $\lim_{t\to\infty}\mathbb{P}[N(t)=0] > 0$ if $a > 1$, indicating that there is a positive chance that infinite rolls do not produce any defect due to the quality of the production, which seems impossible in practice. So if we assume $\delta = 0$, we have to restrict the choice of the probability law to $0 \leq a \leq 1$, which explains the popular use of the Negative Binomial ($a = 1$) and Poisson Inverse Gaussian ($a = 0.5$) distributions. Bissell's (1972a,b) model (NB) is obtained by $\delta = 0$ and $a = 1$.

Mean and variance can be obtained by different methods. The easiest method is the use of the factorial cumulant generating function of $N(t)$ (see Kendall and Stuart (1962) p75) which reduces to $-\eta(-tu)$ and gives by differentiation

$$\begin{aligned}
\mathbb{E}N(t) &= (\delta+p)t \\
\mathbb{V}arN(t) &= (\delta+p)t + pact^2.
\end{aligned}$$

In the general model, the index of overdispersion depends on the four parameters of the model :

$$\frac{\mathbb{V}arN(t)}{\mathbb{E}N(t)} = 1 + \frac{pact}{p+\delta} = 1 + \frac{act}{1+\frac{\delta}{p}}.$$

For a good production, $\delta + p$ has to be low but a high standard of quality of production requires that the fraction $\frac{\delta}{p}$ be large. In that case, most defaults are purely random.

Note that if $\delta = 0$ and a particular value of a is chosen, then the index of overdispersion is reduced to depend only on c.

4 The Fit

Let us assume that there are n observations $(t_i, N(t_i))$.

The parametric approach has the advantage that the parameters may be estimated by maximum likelihood. Standard numerical maximization techniques (all the calculations have been done using Microsoft Excel, and in particular the optimization module of Excel) allow these estimates.

To compare the possible models, we will use the following goodness of fit criteria :

- l : the maximized loglikelihood.
- $AIC = l - r$: the Akaike Information Criterion (see Akaike (1973)).
- $BIC = l - \frac{1}{2}r\ln(n)$: the Bayesian Information Criterion (see Schwartz (1978)).

where r is the number of estimated parameters. Table 1 summarizes the numerical results.

	Model	δ	p	c	a	ac	l	AIC	BIC
1	Poisson	0	0.0151		0		-93.91	-94.91	-95.64
2	Hofmann	0	0.0150	0.0086	0.2402		-87.53	-90.53	-92.72
3	PIG	0	0.0150	0.0037	0.5		-87.54	-89.54	-91.00
4	NB	0	0.0153	0.0024	1		-88.01	-90.01	-91.47
5	NTA	0	0.0151			0.0015	-88.02	-90.02	-91.48
6	PIG + Poisson	0.0066	0.0083	0.0076	0.5		-87.35	-90.35	-92.54
7	NB + Poisson	0.0097	0.0055	0.0077	1		-87.35	-90.35	-92.54
8	NTA + Poisson	0.0117	0.0032			0.0099	-86.93	-89.93	-92.12

Table 1: Summary details of fitting various members of the class of models

The first model estimates the parameter of the Poisson probability law. In model 2, we assume $\delta = 0$ and we estimate the three parameters of the general probability law. In models $3, 4, 5$ we assume $\delta = 0$ and we restrict the choice of the law by assuming respectively $a = 0.5$, $a = 1$, $a \to \infty$. As Lindsey (1999) observed, "the change in AIC for the effect of length is much less when overdispersion is taken into account". In models $6, 7$ and 8 we leave the parameter δ free and restrict respectively $a = 0.5$, $a = 1$ and ac being constant.

With our model, equation (2) becomes

$$\begin{aligned}
\mathbb{E}[N(t+h) - N(t)|N(t) = 0] &= h\eta'(t) \\
&= h[\delta + pe^{-act}] \quad \text{in the NTA+Poisson case} \\
&= h[\delta + \frac{p}{(1+ct)^a}] \quad \text{else.}
\end{aligned}$$

Due to the fact that $e^{-act} \leq \frac{1}{(1+ct)^a}$, we see from (2) that the data indicate a high quality of production. Indeed a good production process is such that a roll of a certain length without defect implies the smallest frequency of defects in a longer roll.

From the point of view of the penalized goodness of fit criteria, we observe that PIG and NB may be acceptable choices within our general family. However, the absence of a purely random component does not seem tenable for physical reasons, even if both distributions have the property $\lim_{t\to\infty} \mathbb{P}[N_2(t) = 0] = 0$. This illustrates what was announced in the introduction on the validation of a statistical model. We therefore retain the NTA + Poisson model even if it is a little overfitting.

5 Aggregate Cost

In this section we provide some additional information regarding the use of our model. We assume that the distributions of both the number of defects and the cost of defects are known a priori. It is interesting to obtain the distribution of the aggregated cost, e.g. for insurance buying purposes. We will see that this is obtainable at a relatively low cost, thanks to some recursion formulae.

We assume that the cost for repairing a fault is given by a random variable X that is discrete and distributed along equidistant mass points. We also assume that the realizations of X are independent and independent of $N(t)$. We are interested in the probability distribution of the total cost $S(t) = X_1 + \cdots + X_{N(t)}$, where $N(t)$ is the number of flaws in a roll of length t. This kind of model is well known in actuarial sciences.

The usual way to obtain the probability distribution of $S(t)$ is to use the convolution formula:

$$f_{S(t)}(x) = \sum_{n=0}^{\infty} \Pi(n,t) f_X^{*n}(x),$$

where $f_X(x) = \mathbb{P}[X = x]$ is the probability function of X,
$f_{S(t)}(x) = \mathbb{P}[S(t) = x]$ is the probability function of $S(t)$,
$f_X^{*n}(x) = \mathbb{P}[X_1 + \cdots + X_n = x]$ is the probability function of the n-fold convolution of X with the convention that $f_X^{*0}(0) = 1$.

But with Hofmann's distribution, a double recursion may be performed in order to find the distribution of $S(t)$ (see Walhin (2000) for details) avoiding the evaluation of convolutions. As the probability law of $N(t)$ is infinitely divisible, we can write $N(t) = \xi_1 + \cdots + \xi_{L(t)}$ where the ξ_i are integer valued random variables having probability law depending on t and $L(t)$ is a Poisson random variable.

We then define an intermediate random variable $V(t)$ as $V(t) = X_1 + \cdots + X_\xi$. Its distribution is given by

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

where $p_1 = \frac{t\theta'(t)+\delta t}{\theta(t)+\delta t}$ and $p_2 = -\frac{t^2}{2}\frac{\theta''(t)}{\theta(t)+\delta t}$.

Then the distribution of $S(t)$ is obtained from the distribution of $V(t)$ by :

$$\begin{aligned}
f_{S(t)}(0) &= e^{-\theta(t)(1-f_X(0))}, \\
f_{S(t)}(x) &= \frac{\theta(t)+\delta t}{x}\sum_{i=1}^{x} i f_{V(t)}(i) f_{S(t)}(x-i) \quad , x \geq 1.
\end{aligned}$$

For our particular model NTA+Poisson, the recursive formulae reduce to

$$\begin{aligned}
f_{V(t)}(0) &= 1 - \frac{p(1-e^{-bt(1-f_X(0))}) + \delta bt(1-f_X(0))}{p(1-e^{-bt}) + \delta bt} \\
f_{V(t)}(x) &= \frac{bt}{x}\sum_{i=1}^{x} i f_X(i) f_{V(t)}(x-i) + \alpha_1 f_X(x) + \alpha_2 f_X^{*2}(x) \quad , \quad x \geq 1 \\
\alpha_1 &= bt\frac{pe^{-bt}+\delta}{p(1-e^{-bt}) + \delta bt} \\
\alpha_2 &= -\frac{b^2t^2\delta}{2(p(1-e^{-bt}) + bt\delta)} \\
f_{S(t)}(0) &= e^{-(\frac{p}{b}(1-e^{-bt})+\delta t)(1-f_X(0))} \\
f_{S(t)}(x) &= \frac{p(1-e^{-bt})+\delta bt}{bx}\sum_{i=1}^{x} i f_{V(t)}(i) f_{S(t)}(x-i) \quad , \quad x \geq 1.
\end{aligned}$$

Assuming a discrete uniform distribution for the cost of repairs between 0 and 4, i.e. $f_X(i) = 0.2, \quad i = 0, 1, 2, 3, 4$, we are able to compute the distribution of the aggregate costs at different times. Table 2 gives the distribution for $t = 200$ as a numerical illustration.

x	$f_{V(200)}(x)$	$f_{S(200)}(x)$
0	0.187	0.119
1	0.189	0.059
2	0.190	0.074
3	0.192	0.092
4	0.195	0.113
5	0.010	0.079
6	0.009	0.078
7	0.008	0.073
8	0.006	0.064
9	0.004	0.049
10	0.003	0.042
11	0.002	0.035
$\vdots$	$\vdots$	$\vdots$

Table 2: Distribution of the aggregate cost at $t = 200$

6 Conclusion

We have shown in this paper that the Hofmann process, convolved with a Poisson process, may be of interest in order to analyse industrial problems where defects are counted on varying element sizes. The model has interesting properties that can be given a physical interpretation in the case of data like that of Bissell (1972a). Moreover, it is tractable due to the recursions giving the probability function. It is also easy to obtain the law of the aggregate cost of defects thanks to a two-stage recursion. As the model encompasses many models encountered in the literature and remains tractable, it is an appropriate means for analysing such industrial processes.

Acknowledgments

We would like to thank the editor and two anonymous referees for their useful comments which have led to a better presentation of the paper.

References

1. Akaike, H. (1973), Information Theory and an Extension of the Maximum Likelihood Principle, pages 267-281. In Petrov, B.N. and Csaki, F. (editors) (1973), Second International Symposium on Information Theory), Akademiai Kiado, Budapest.

2. Azzalini, A., Bowman, A.W. and Härdle, W. (1989), On the Use of Nonparametric Regression for Model Checking, Biometrika, 76, 1-11.

3. Berg, C. and Forst, G. (1975), Potential Theory on Locally Compact Abelian Groups, Springer-Verlag, Berlin.

4. Bissell, A.F. (1972a), A Negative Binomial Model with Varying Element Sizes, Biometrika, 59, 435-441.

5. Bissell, A.F. (1972b), Another Negative Binomial Model with Varying Element Sizes, Biometrika, 59, 691-693.

6. Cox, D.R. and Snell, E.J. (1981), Applied Statistics, Principles and Examples, Chapman and Hall, London.

7. Feller, W. (1971) An Introduction to Probability Theory and its Applications Vol II (third edition), Wiley, New York.

8. Grandell, J. (1997), Mixed Poisson Processes, Chapman and Hall.

9. Gupta, P.L., Gupta, R.C. and Ong, S.H. (2004), Modelling Count Data by Random Effect Poisson Model, Sankhya, 66, 548-565.

10. Hill, J.R. and Tsai, C.L. (1988), Calculating the Efficiency of Maximum Quasilikelihood Estimation, Applied Statistics, 37, 219-230.

11. Hinde, J. and Demétrio, C. (1998), Overdispersion : Models and Estimation, Computational Statistics and Data Analysis, 27, 151-170.

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

13. Kendall, M.G. and Stuart, A. (1962), The Advanced Theory of Statistics, vol1. Distribution Theory (second edition),London, Charles Griffin.

14. Lindsey, J.K. (1999), Models for Repeated Measurements (second edition), Oxford University Press.

15. Mc Kay, I. (1996), A Note on Density Estimation for Poisson Mixtures, Statistics and Probability Letters, 27, 255-258.

16. Panjer, H.H. (1981), Recursive Evaluation of a Family of Compound Distributions, Astin Bulletin, 12, 22-26.

17. Schwartz, G., (1978), Estimating the dimension of a model. The Annals of Statistics, 6, 461-464.

18. Simar, L. (1976), Maximum Likelihood Estimation of a Compound Poisson Process, Annals of Statistics, 4, 1200-1209.

19. Sundt, B. and Jewell, W.S. (1981), Further Results on Recursive Evaluation of Compound Distributions, Astin Bulletin, 12, 27-39.

20. Walhin, J.F. (2000), Recursions for Actuaries with Applications in the Field of Reinsurance and Bonus-Malus Systems, PhD thesis, Institut de Statistique, Université Catholique de Louvain.

21. Walhin, J.F. and Paris, J. (2002), A General Family of Overdispersed Probability Laws, Belgian Actuarial Bulletin, 2, 92-96.