A new underdispersed count model with applications

Amanda Fernández-Fontelo[1], *Pere Puig*[2]

[1] amanda@mat.uab.cat, Departament de matemàtiques, UAB, Spain.
[2] ppuig@mat.uab.cat, Departament de matemàtiques, UAB, Spain.

Abstract

Underdispersion is a less common phenomenon which appears in count data. In the literature, there exist many models which deal with the problem of overdispersion, but not with underdispersion. COM-Poisson model is one of the few which is able to deal with underdispersion. In this work we propose a new count model generated from a Birth-death process which will be applied to analyse several examples. One of them will be the number of chromosomal aberrations in irradiated cells that is useful in dosimetry.

Keywords: Birth and death process; COM-Poisson distribution; Underdispersion

1. The M/M/1 model.

The M/M/1 is the most elementary process in queuing theory whose state space is $\{0, 1, 2, \cdots\}$. It is characteristic because the arrivals occur according to a Poisson process with rate λ (interarrival times are exponentially distributed with mean equal to $\frac{1}{\lambda}$), service times are exponentially distributed with mean $\frac{1}{\mu}$ and the customers are served following the discipline first-come, first-served. That is, when the service is complete the customer leaves the queue.
The stationary distribution of the process is the limiting distribution when time t tends to infinity. M/M/1 process only has stationary distribution when, on average, arrivals happen slower than service completions. In other words, when $\rho = \frac{\lambda}{\mu}$, which is defined as the proportion of time the server is working, is less than 1. The M/M/1 process can be understood as a continuous time Markov process, and it is also a Birth-death process.
In this talk we consider an M/M/1 model with state dependent service rate, that is, $\mu_n = \mu f(n)$, where n is the number of customers in the queue.

An important point in a M/M/1 queue is the calculation of $p_n(t)$ $n = 0, 1, 2, \cdots$, that is the probability that at time t there are n customers in the system. To obtain these probabilities, it is necessary to use the memoryless property of the exponential distributions of the times in the system and the assumption that $\nabla t \to 0$. Hence,

$$p_0\left(t + \nabla t\right) = \left(1 - \lambda \nabla t\right) p_0(t) + \mu_n \nabla t p_1(t) + o\left(\nabla t\right), \tag{1}$$

$$p_n\left(t + \nabla t\right) = \lambda \nabla t p_{n-1}(t) + \left(1 - \left(\lambda + \mu_n\right)\right) \nabla t\right) p_n(t) + \mu_n \nabla t p_{n+1}(t) + o\left(\nabla t\right). \tag{2}$$

By letting $\nabla t \to 0$ and assuming a steady state ($p_n^{'}(t) = 0$ for all n), we obtain the following set of difference equations for the probabilities p_n,

$$\mu_n p_1 = \lambda p_0 \tag{3}$$

$$(\lambda + \mu_n)p_n = \lambda p_{n-1} + \mu_{n+1} p_{n+1}, \tag{4}$$

Therefore, the probabilities can be calculated recursively from (4). Even more, in some situations a closed expression of these probabilities can also be found. The value of p_0 can be obtained from the condition that all the probabilities have to sum 1.

The idea of customers in a queue can be used to model some biological events, as we will see in section 4.

2. The COM$-$Poisson distribution.

The origin of the COM-Poisson distribution is the paper of Richard et all. (1962), where the authors proposed a power function for μ_n taking into account that $f(n) = n^c$. Specifically, they considered the following model for the service rate in a queuing system,

$$\mu_n = n^c \mu, \tag{5}$$

where n is the number of units in the system and c a constant which indicates the degree to which the service rate of the system is affected by the system state.
Applying the equations (3) and (4), it was deduced a general expression for the probabilities,

$$p_n = \frac{\rho}{(n!)^c} p_0 \qquad n \geq 0, \tag{6}$$

and p_0 can be obtained using that the sum of the probabilities is 1.

The distribution of (6) is known as the Conway-Maxwell-Poisson distribution, also called COM-Poisson. One important property of this distribution is that the parameter c models the overdispersion and underdispersion of the data.

3. A new underdispersed count model

The idea to incorporate an accurate behaviour of μ_n which is able to change the speed of the service according to the size of the queue, allow us to propose an exponential function.

$$\mu_n = \mu e^{\beta n}, \tag{7}$$

that is, $f(n) = e^{\beta n}$. Hence, n is also the number of customers in the queue, and β is the constant which indicates the degree to which the service rate of the system is affected by the system state.

Applying the method described above, that is the equations (3) and (4) and replacing μ_n by (7), the following solution is obtained,

$$p_n = \frac{\rho^n}{e^{\beta n\left(\frac{n+1}{2}\right)}} p_0 \qquad n \geq 0. \tag{8}$$

As the sum of the probabilities is 1, p_0 can be expressed as $p_0 = \frac{1}{\sum\limits_{n=0}^{\infty} \frac{\rho^n}{e^{\beta n\left(\frac{n+1}{2}\right)}}}$.

Note that, if $\beta = 0$ then $p_0 = 1 - \rho$. However, for other values of β, the exact expression of p_0 is difficult to obtain, and has to be numerically approximated.

4. Example of application

Data is composed of the number of aberrations in cells irradiated with different doses of radiation. The description and the experimental procedure can be found in Pujol et al. (2014). They considered the samples of cells irradiated with doses 1, 3, 5, 7, 10, 15, 20 and 25. The means, variances and dispersion indexes for each sample are summarized in Table 2 in Pujol et al. (2014).

We have detected that the dose is related with parameter β. Therefore, we define a suitable function for this parameter as a function of the applied dose. After several explorations, we chose the link-function $\beta = \frac{\theta_1}{d+\theta_2} + \theta_3 d$, where d is the dose. It was selected taking into account several criteria: minimum AIC, the significance of the parameters, and the behaviour of the empirical and theoretical means and variances.

The selected model was estimated by maximum likelihood. The following estimates were obtained:

$$\hat{p}_n = \frac{(3.930)^n}{e^{\left(\frac{2.202}{d-0.369}+0.001d\right)n\left(\frac{n+1}{2}\right)}} \hat{p}_0 \qquad n \geq 0, \tag{9}$$

Parameter	**Estimate**	**Lower limit (95%)**	**Upper limit (95%)**
ρ	3.930	3.387	4.474
θ_1	2.202	1.986	2.419
θ_2	-0.369	-0.424	-0.316
θ_3	0.001	0.0008	0.0015

Table 1: MLE of the selected model.

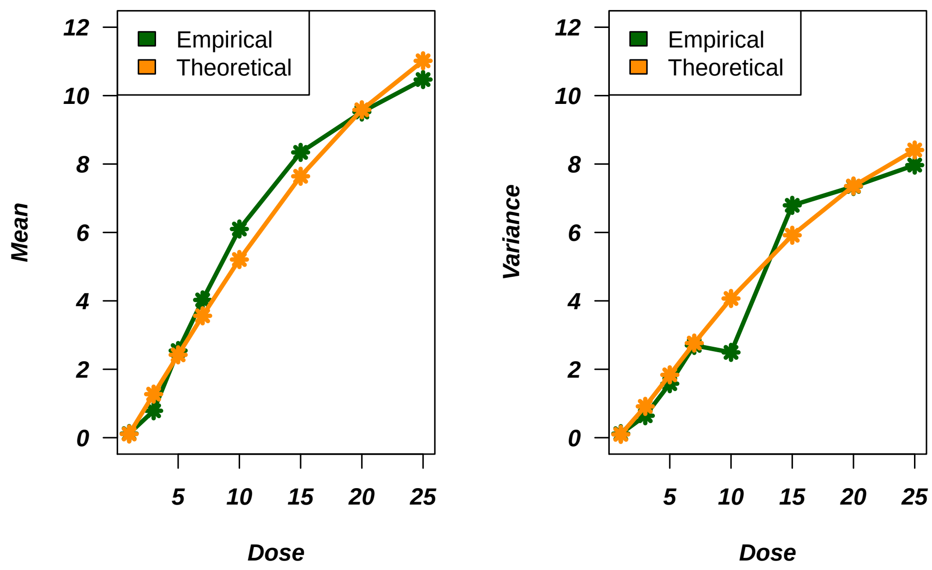

Figure 1: Empirical and theoretical means and variances of the observed and estimated data.

Taking into account the interpretation of a queuing model with state dependent service rate, the relation between this model and the example exposed can be understood in the following way: when the cell is irradiated, the particles arrive to the cell and damage it. These arrivals follow a Poisson process. At the time that a particle arrive and produce damage, the repair mechanism of the cell (service) begins to repair the damage. While this mechanism is working, other damages arrive to the cell. If the mechanism is able to repair the aberrations, these aberrations leave the queue.

5. Acknowledgments

This work was funded by the grant MTM2012-31118 from the Ministry of Economy and Competitiveness of Spain.

6. Bibliography

[1] Richard, W., Conway, W., and Maxwell, W.L. (1962). *A Queueing Model with State Dependent Service Rate.* The Journal of Industrial Engineering **12**, 132−136.

[2] Pujol, M., Barquinero, J.F., Puig, P., Puig, R., Caballín, M.R. and Barrios, L. (2014). *A New Model of Biodosimetry to Integrate Low and High Doses.* PloS ONE **9(12)**: e114137. doi:10.1371/journal.pone.0114137.