Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

A new discrete distribution: properties and applications in medical care

Emilio Gómez Déniz

To cite this article: Emilio Gómez Déniz (2013) A new discrete distribution: properties and applications in medical care, Journal of Applied Statistics, 40:12, 2760-2770, DOI: 10.1080/02664763.2013.827161

To link to this article: https://doi.org/10.1080/02664763.2013.827161

Published online: 09 Aug 2013.

Submit your article to this journal

Article views: 702

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cjas20

Journal of Applied Statistics, 2013
Vol. 40, No. 12, 2760–2770, http://dx.doi.org/10.1080/02664763.2013.827161

Taylor & Francis
Taylor & Francis Group

A new discrete distribution: properties and applications in medical care

Emilio Gómez Déniz*

Department of Quantitative Methods, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain

(*Received 6 March 2013; accepted 17 July 2013*)

This paper proposes a simple and flexible count data regression model which is able to incorporate overdispersion (the variance is greater than the mean) and which can be considered a competitor to the Poisson model. As is well known, this classical model imposes the restriction that the conditional mean of each count variable must equal the conditional variance. Nevertheless, for the common case of well-dispersed counts the Poisson regression may not be appropriate, while the count regression model proposed here is potentially useful. We consider an application to model counts of medical care utilization by the elderly in the USA using a well-known data set from the National Medical Expenditure Survey (1987), where the dependent variable is the number of stays after hospital admission, and where 10 explanatory variables are analysed.

Keywords: confluent hypergeometric function; covariate; health services; uniform distribution; Poisson distribution

1. Introduction

Analysis of count data is required in many areas of economics, social sciences and biometrics. In many cases, the simple Poisson distribution model is not appropriate because it imposes the restriction that the conditional mean of each count variable must equal the conditional variance. As Min and Czado [15] point out, the Poisson distribution is too simple to capture complex structures of count data such as overdispersion. In consequence, various models that are less restrictive than Poisson, and are based on other distributions, have been presented in the statistical literature, including the negative binomial, generalized Poisson and generalized negative binomial (see Cameron and Trivedi [4] and Famoye [12], among others). Other count regression models such as the Poisson-inverse Gaussian have traditionally been used in actuarial contexts with respect to motor-insurance claims [8] and/or to determine premiums in bonus-malus systems [10]. Nevertheless, these regression models have been developed by setting one or more parameters as a specified function of regressors, which in computational terms can be viewed as a disadvantage.

*Email: egomez@dmc.ulpgc.es

This article was originally published with errors. This version has been corrected. Please see Erratum http://dx.doi.org/10.1080/02664763.2013.837642

© 2013 Taylor & Francis

This paper proposes a simple and flexible count data regression model which is able to incorporate the overdispersion that typically occurs in count data sets in economics and the social sciences, and which obtains a larger value in the zero vertex than the Poisson distribution. Like the Poisson distribution, the proposed count distribution depends on only one parameter and its formulation is easy to implement in a survey of fields. In this paper, we consider an application to model counts of medical care utilization by the elderly in the USA using a well-known data set published by the National Medical Expenditure Survey (1987) in which the dependent variable is the number of stays after hospital admission, and analysing 10 explanatory variables.

The distribution we propose is overdispersed and presents simple, closed expressions to estimate the parameters on which it depends, using the moments method. Therefore, it is a candidate for fitting data sets with overdispersion. The examples of real data provided show that the model works very well, and this is confirmed by comparison with the classical Poisson model adjusted by the maximum likelihood method.

Hospitalizations among the elderly population, aged 65 years or more, represent a significant portion of annual expenditure on hospital care. Modelling medical costs is of great interest in health economics and particularly in health insurance [18]. In the USA, for example, utilization, and thus the cost of hospital care among the elderly, is a significant priority for policy-makers because public insurance programmes fund the largest percentage of health care costs for this population, and their number will continue to grow in the future. The situation is very similar in nearly all Western countries; for example, in the UK the proportion of the population aged 65 years or more is currently about 15%, compared with 11% in 1951 and 5% in 1911 [3]. Many more people now survive into their 80s and 90s, and this has led to sharply increased demand for health care. In the coming years, a shrinking working population will be obliged to finance the pensions, health care and other services required by the burgeoning elderly population, and this problem has been exacerbated by the present financial crisis.

The rest of this paper is structured as follows. Section 2 describes the theoretical development of the new count distribution, including some properties and different methods of estimation. An application of the proposed model is examined in Section 3. Finally, some comments are made and conclusions are drawn in Section 4.

2. A new competitor count distribution

In this section, we define the new probability function and its properties, following the methodology proposed by Hu *et al.* [14]. They show that if N and X are two random variables denoting the number of particles entering and leaving an attenuator, the probability functions $p(n)$ and $f(x)$ of these two random variables are connected by the binomial decay transformation

$$\Pr(X = x) = \sum_{n=x}^{\infty} \binom{n}{x} p^x (1-p)^{n-x} p(n). \tag{1}$$

It is easy to see that $\Pr(X = x) \geq 0$, $\forall n \in \mathbf{N}$ and since

$$\sum_{x=0}^{\infty} \Pr(X = x) = \sum_{x=0}^{\infty} \sum_{n=x}^{\infty} \binom{n}{x} p^x (1-p)^{n-x} p(n) = \sum_{n=0}^{\infty} p(n) = 1,$$

for $x = 0, 1, \ldots$ and $0 \leq p \leq 1$, expression (1) is a proper probability mass function. Hu *et al.* [14] show that if $p(n)$ is the Poisson distribution with parameter $\lambda > 0$, then $\Pr(X = x)$ is the Poisson distribution with parameter $p\lambda$.

To the best of our knowledge, apart from the paper by Hu *et al.* [14] this kind of methodology to obtain new probability distributions has not been used in the past. We will show that by

interchanging in Equation (1) the binomial distribution and the discrete uniform distribution and maintaining $p(n)$ as the Poisson distribution a simple, single-parameter distribution can be obtained, and that this is competitive with the Poisson distribution.

Assume that X follows a discrete uniform distribution with parameter $n > 0$, i.e. $N \sim \mathcal{U}(n)$, where $x = 0, 1, \ldots, n$ and $p(n)$ is the Poisson distribution with parameter $\lambda > 0$. This is formalized in the following definition.

DEFINITION 1 *We say that a random variable X has a uniform Poisson distribution if it allows the stochastic representation*

$$X \mid N = n \sim \mathcal{U}(n), \tag{2}$$

$$N \sim \mathcal{P}o(\lambda), \tag{3}$$

with n, $\lambda > 0$.

THEOREM 1 *Let* $X \sim \mathcal{UP}(\lambda)$ *be a uniform Poisson distribution as defined in Equations* (2) *and* (3)*, then the probability mass function is given by*

$$\Pr(X = x) = p_x = \frac{\lambda^x \, \mathrm{e}^{-\lambda}}{(x+1)!} \, {}_1F_1(1, x+2, \lambda), \tag{4}$$

with $x = 0, 1, 2, \ldots, \lambda > 0$ *and where* ${}_1F_1(a, b, z)$ *is the confluent hypergeometric function given by*

$${}_1F_1(a, b, z) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} \frac{z^k}{k!},$$

and where $(a)_k = \Gamma(a+k)/\Gamma(a)$ *is the Pochhammer symbol.*

Proof The result follows taking into account that

$$\begin{aligned}
\Pr(X = x) &= \sum_{n=x}^{\infty} \frac{1}{n+1} \frac{\lambda^n \, \mathrm{e}^{-\lambda}}{n!} \\
&= \mathrm{e}^{-\lambda} \sum_{j=0}^{\infty} \frac{1}{x+j+1} \frac{\lambda^{x+j}}{(x+j)!} = \frac{\mathrm{e}^{-\lambda} \lambda^x}{(x+1)!} \sum_{j=0}^{\infty} \frac{j!(x+1)!}{(x+j+1)!} \frac{\lambda^j}{j!} \\
&= \frac{\lambda^x \, \mathrm{e}^{-\lambda}}{(x+1)!} \, {}_1F_1(1, x+2, \lambda).
\end{aligned}$$

■

Using the identity from Kummer's first theorem it is simple to see that the probability mass function (4) can be rewritten as

$$p_x = \frac{\lambda^x}{(x+1)!} \, {}_1F_1(x+1, x+2, -\lambda),$$

and also in terms of the gamma function, $\Gamma(a, z) = \int_z^{\infty} t^{a-1} \, \mathrm{e}^{-t} \, \mathrm{d}t$, as

$$p_x = \frac{1}{\lambda} \frac{\Gamma(x+1) - \Gamma(x+1, \lambda)}{\Gamma(x+1)}, \quad x = 0, 1, \ldots, \lambda > 0,$$

where we now use the identity $\Gamma(a, z) = \Gamma(a) - (z^a/a) {}_1F_1(a, a+1, -z)$, where $\Gamma(z)$ is the gamma function and $\Gamma(a, z)$ is the incomplete gamma function given by $\Gamma(a, z) = \int_z^{\infty} t^{a-1} \, \mathrm{e}^{-t} \, \mathrm{d}t$.

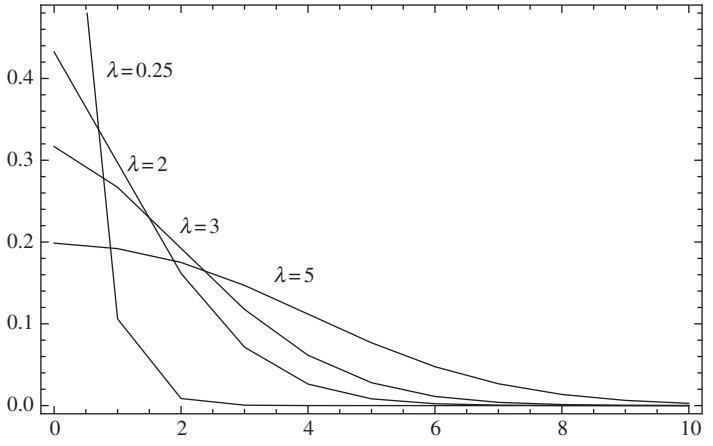

Figure 1. Some appearances (polygons) of the new probability function for different values of the parameter λ.

Taking into account that ${}_1F_1(1,2,\lambda) = (e^{\lambda} - 1)/\lambda$, from Equation (4) we have

$$p_0 = \frac{1 - e^{-\lambda}}{\lambda}. \tag{5}$$

Now, let $\Pr(K = k)$ be the probability function (4) and $\Pr(k \mid \lambda)$ be the probability function of a simple Poisson distribution, then $\Pr(K = 0) \geq \Pr(0 \mid \lambda)$. To see this, consider the function $\Psi(\lambda) = (\lambda + 1)\, e^{-\lambda} - 1$. Since $\Psi(\lambda)$ is a continuous function with $\Psi(0) = 0$, $\Psi(\infty) = -1$ and $\Psi'(\lambda) = -\lambda\, e^{-\lambda} < 0$, we conclude that $\Psi(\lambda) < 0$ and therefore p_0 is larger than $e^{-\lambda}$.

The probabilities of the new distribution can also be computed by using Equation (5) together with

$$p_x = \frac{\lambda}{x+1} \frac{{}_1F_1(1, x+2, \lambda)}{{}_1F_1(1, x+1, \lambda)} p_{x-1}, \quad x = 1, 2, \ldots \tag{6}$$

Using Equation (6) together with the relation $\Gamma(a+1, z) = a\Gamma(a, z) + z^a\, e^{-z}$, it is easy to obtain that

$$\frac{p_x}{p_{x-1}} = 1 - \frac{\Gamma(x+1, \lambda) + x\Gamma(x, \lambda)}{x(\Gamma(x) - \Gamma(x, \lambda))} < 1, \quad x = 1, 2 \ldots,$$

which reveals that the distribution is unimodal with a zero vertex. A study of Figure 1 confirms this feature and that the graph is similar to that of the Poisson distribution when the parameter of the latter is less than 1.

PROPOSITION 1 *The mean and variance of a discrete random variable following the probability function* (4) *are given by*

$$E(X) = \frac{\lambda}{2} \quad \text{and} \tag{7}$$

$$\text{var}(X) = \frac{\lambda(\lambda + 6)}{12}, \tag{8}$$

respectively.

Proof It is known that the mean of the uniform discrete distribution is given by $E(X \mid n) = n/2$. Then, we have the following:

$$E(X) = \sum_{x=0}^{\infty} x p_x = \sum_{x=0}^{\infty} x \sum_{n=x}^{\infty} \frac{1}{n+1} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!}$$
$$= \sum_{n=0}^{\infty} \left(\sum_{x=0}^{n} \frac{x}{n+1} \right) \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} = \sum_{n=0}^{\infty} \frac{n}{2} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} = \frac{\lambda}{2}. \tag{9}$$

The variance can be computed in the same way, using the second-order moment around the origin of the uniform discrete distribution, $E(X^2 \mid n) = n(1+2n)/6$, after straightforward computation. ■

It is a simple exercise to show that the mean and variance increase with λ. Remain that for the Poisson distribution the variation of the mean and variance with respect to the parameter is constant and equal to one. Also, it is easy to see that the new distribution in Equation (4) is overdispersed, i.e. the variance is greater than the mean, since the index of dispersion can be written as

$$\frac{\mathrm{var}(X)}{E(X)} = 1 + \frac{\lambda}{6} > 1.$$

Min and Czado [15] point out that the Poisson distribution is too simple to capture complex structures of count data such as overdispersion. Recall that most of the real count data discussed in the literature present overdispersion. For this reason, various models that are less restrictive than Poisson and based on other distributions have been presented in the statistical literature, including the negative binomial, generalized Poisson and generalized negative binomial ([4,12], among others). Nevertheless, with these models additional parameters are included and the estimation procedure of the parameters is more difficult. The new count distribution presented here overcomes these problems and incorporates just one parameter, as in the Poisson distribution.

The next result shows the cumulative distribution function of the new count distribution proposed in this paper.

PROPOSITION 2 *The cumulative distribution function of a random variable following the probability mass function in Equation* (4) *is given by*

$$F(x) = \Pr(X \le x) = 1 - \frac{\lambda^{x+1} \mathrm{e}^{-\lambda}}{(x+2)!} {}_1F_1(2, x+3, \lambda). \tag{10}$$

Proof Using the integral representation of the confluent hypergeometric function given by

$${}_1F_1(a,b,z) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 \mathrm{e}^{zt} t^{a-1} (1-t)^{-a+b-1} \, \mathrm{d}t,$$

we obtain that the cumulative distribution function can be written as

$$F(x) = \sum_{j=0}^{x} \frac{\lambda^j \mathrm{e}^{-\lambda}}{(j+1)!} {}_1F_1(1, j+2, \lambda) = \mathrm{e}^{-\lambda} \int_0^2 \mathrm{e}^{\lambda t} \left(\sum_{j=0}^{x} \frac{[\lambda(1-t)]^x}{j!} \right) \mathrm{d}t$$
$$= \frac{1}{x!} \int_0^1 \Gamma(1+x, \lambda(1-t)) \, \mathrm{d}t.$$

Now, applying integration by parts we have

$$\int_0^1 \Gamma(1+x,\lambda(1-t))\,\mathrm{d}t = t\Gamma(1+x,\lambda(1-t))|_0^1 - \lambda^{x+1}\,\mathrm{e}^{-\lambda}\int_0^1 t\,\mathrm{e}^{\lambda t}(1-t)^x\,\mathrm{d}t$$
$$= \Gamma(x+1) - \lambda^{x+1}\,\mathrm{e}^{-\lambda}\frac{\Gamma(x+1)}{\Gamma(x+3)}\,{}_1F_1(2,x+3,\lambda),$$

from which the result follows directly. ■

The survival function is obtained from Equation (10) and is given by

$$F(x) = 1 - F(x-1) = \frac{\lambda^x\,\mathrm{e}^{-\lambda}}{(x+1)!}\,{}_1F_1(2,x+2,\lambda). \tag{11}$$

Using Equation (4) together with Equation (11), we obtain the failure rate given by

$$r(x) = \frac{\Pr(X=x)}{\bar{F}(x)} = \frac{{}_1F_1(1,x+2,\lambda)}{{}_1F_1(2,x+2,\lambda)}.$$

A plot of this function, which is not presented here, for different values of the parameter λ, reveals that the new probability function has an increasing failure rate.

2.1 *Estimation*

In this section, we propose three methods to estimate the parameter of the new distribution: the zero proportion method, the method of moments and the maximum likelihood method.

The first method is based on the frequency of zeros. This approach usually works well only when the mode of the distribution is zero and the proportion of zeros is relatively high [2]. Because the distribution depends on a single parameter, only one equation is needed to estimate the parameter. Let the proportion of zeros in the sample be given by $\tilde{p}_0$; then by using Equation (5) we need to solve the equation

$$\lambda\tilde{p}_0 + \mathrm{e}^{-\lambda} - 1 = 0,$$

which can be solved numerically.

Assume now that $\underline{x} = (x_1,\ldots,x_t)$ is a random sample of size t from the discrete distribution (4). If $\bar{x}$ is the sample moment, from Equation (7) it is obtained directly that $\hat{\lambda} = 2\bar{x}$ is the moment estimate which always exists, and that it is unique and unbiased, as can be readily shown.

Although the moment estimator always exists and it is easy to obtain, it often provides unfeasible estimates and therefore maximum likelihood estimates are more convenient. The maximum likelihood estimates of the model are obtained by maximizing the following log-likelihood function:

$$\ell(\lambda) = t(\bar{x}\log\lambda - \lambda) + \sum_{i=1}^{t}\log\,{}_1F_1(1,x_i+2,\lambda) - \sum_{i=1}^{t}\log(x_i+1)!, \tag{12}$$

from which, by differentiating with respect to λ and equating to 0, the following likelihood equation is obtained

$$t\left(\frac{\bar{x}}{\lambda} - 1\right) + \sum_{i=1}^{t}\frac{1}{x_i+2}\frac{{}_1F_1(2,x_i+3,\lambda)}{{}_1F_1(1,x_i+2,\lambda)}. \tag{13}$$

Again, Equation (13) cannot be solved explicitly. It must be solved by the numerical method or by directly maximizing the log-likelihood function in Equation (12). Because the global maximum

of the log-likelihood surface is not guaranteed, different initial values of the parametric space can be considered as a seed point; then, by using the `FindMaximum` function of Mathematica software package v.8.0 [17] methods such as Newton, PrincipalAxis and QuasiNewton will produce the same result. The standard errors of the parameter estimates were obtained, approximately, from the Hessian matrix by the conventional approach. Thus, we find that the second derivative of the log-likelihood function is given by

$$\frac{\mathrm{d}^2\ell(\lambda)}{\mathrm{d}\lambda^2} = -\frac{t\bar{x}}{\lambda^2} + \sum_{i=1}^{t} \frac{2}{(x_i+2)(x_i+3)} \frac{{}_1F_1(3,x_i+4,\lambda)}{{}_1F_1(1,x_i+2,\lambda)} - \sum_{i=1}^{t} \frac{1}{(x_i+2)^2} \left[\frac{{}_1F_1(2,x_i+3,\lambda)}{{}_1F_1(1,x_i+2,\lambda)}\right]^2. \tag{14}$$

It is verified, by using integral representation of the confluent hypergeometric function, that

$$E\left(\frac{{}_1F_1(3,X+4,\lambda)}{(X+2)(X+3){}_1F_1(1,X+2,\lambda)}\right) = \frac{1}{6},$$

where the expectation is taken with respect to Equation (4).

The analytic expression for the last expectation in Equation (14) is not feasible. For large t, for computational purposes, this is evaluated by ignoring the expectation operator, thus

$$E\left(-\frac{\mathrm{d}^2\ell(\lambda)}{\mathrm{d}\lambda^2}\right)\Big|_{\lambda=\hat{\lambda}} \approx \frac{1}{2\hat{\lambda}} - \frac{t}{3} + \sum_{i=1}^{t} \frac{1}{(x_i+2)^2} \left[\frac{{}_1F_1(2,x_i+3,\hat{\lambda})}{{}_1F_1(1,x_i+2,\hat{\lambda})}\right]^2,$$

where $\hat{\lambda}$ is the maximum likelihood estimate of λ.

In practice, the line of argument followed is that the dependent variable is a count variable, and that including covariates would be an appropriate method of analysis. See, for example Duan *et al.* [11], Christensen *et al.* [7], Cameron *et al.* [5], Cartwright *et al.* [6] and Deb and Trivedi [9], among others, with respect to health services.

Following this approach, a new reparameterization of the distribution in Equation (4) can be obtained from Equation (7) by assuming $\lambda = 2\theta$ from which the mean is now the parameter θ. This is done because the practitioner usually wishes the model to include covariates and this is a suitable way of doing so. The most common specification for the mean parameter θ is exponential, ensuring the non-negativity of θ. That is,

$$\log\theta_i = \sum_{s=1}^{q} x_{is}\beta_s, \quad i = 1,\ldots,t,$$

obtaining the conventional log-linear model such that $E(X) = \exp\{\beta^\top \mathbf{x}\}$, where $\mathbf{x}$ is the vector of covariates and $\beta = (\beta_1,\ldots,\beta_q)^\top$ is an unknown vector of regression coefficients.

The log-likelihood of the new model with covariates is proportional to

$$\ell(\beta_1,\ldots,\beta_q) \propto \sum_{i=1}^{t} y_i \sum_{s=1}^{q} x_{is}\beta_s - 2\sum_{i=1}^{t}\theta_i + \sum_{i=1}^{t} \log({}_1F_1(1,y_i+2,\theta_i)).$$

The normal equations are

$$\frac{\partial\ell}{\partial\beta_j} = -\sum_{i=1}^{t} y_i x_{ij} - 2\sum_{i=1}^{t} x_{ij}\theta_i + 2\sum_{i=1}^{t} \frac{x_{ij}\theta_i {}_1F_1(2,y_i+3,2\theta_i)}{(y_i+2){}_1F_1(1,y_i+2,2\theta_i)} = 0,$$

for $j = 1,2,\ldots,q$.

The second partial derivatives are given by

$$\frac{\partial^2 \ell}{\partial \beta_j^2} = -\sum_{i=1}^{t} x_{ij}^2 \theta_i + 2\sum_{i=1}^{t} \frac{x_{ij}^2 \theta_i G_1(y_i, \theta_i)}{(y_i+2)(y_i+3)} - 4\sum_{i=1}^{t} \frac{x_{ij}^2 \theta_i^2}{(y_i+2)^2} G_2(y_i, \theta_i),$$

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{t} x_{ij} x_{il} \theta_i + 2\sum_{i=1}^{t} \frac{x_{ij} x_{il} \theta_i G_1(y_i, \theta_i)}{(y_i+2)(y_i+3)} - 4\sum_{i=1}^{t} \frac{x_{ij} x_{il} \theta_i^2}{(y_i+2)^2} G_2(y_i, \theta_i),$$

for $j = 1, 2, \ldots, q$ and $j \neq l$, where

$$G_1(y_i, \theta_i) = \frac{(y_i+3)\,{}_1F_1(2, y_i+3, 2\theta_i) + 4\theta_i\,{}_1F_1(3, y_i+4, 2\theta_i)}{{}_1F_1(1, y_i+2, 2\theta_i)} \quad \text{and}$$

$$G_2(y_i, \theta_i) = \left(\frac{{}_1F_1(2, y_i+3, 2\theta_i)}{{}_1F_1(1, y_i+2, 2\theta_i)}\right)^2.$$

3. Illustrative example

In this section, we examine an application of the proposed method to analyse the number of stays after hospital admission in the USA among the elderly population, aged 65 years or more. These hospital stays account for a significant portion of the annual public sector expenditure on hospital care, because public insurance programmes in the USA fund the largest percentage of health care costs for this population, and it is estimated that the size of this population will continue to grow in the future.

The data analysed consist of 4406 individuals covered by Medicare, the USA public insurance programme, and were obtained from the Journal of Applied Econometrics 1997 Data Archive at http://www.econ.queensu.ca/jae/1997-v12.3/deb-trivedi/. They were originally used by Deb and Trivedi [9] in their analysis of various measures of health care utilization using a sample of 4406 single-person households in 1987. For convenience, we model the number of stays after hospital admission (HOSP) as the dependent variable, although another count variable could be taken as the dependent variable if another study were required, such as the number of medical visits. Fundamentally, the convenience of this approach is based on the fact that by testing all the count variables appearing in the data, the variable HOSP simultaneously presents a higher number of zero values and a larger index of dispersion. The number of stays after hospital admission has two interesting features: the presence of overdispersion and the existence of a very high proportion of non-users (80.36%). The mean and the standard deviation of this variable are 0.30 and 0.75, respectively. The explanatory variables are shown in Table 1. Deb and Trivedi [9] give details about the definition of these variables and the summary statistics.

Medicaid is the USA health programme for individuals and families with low incomes and resources; it is jointly funded by the federal government and by the states, and is managed by the latter.

Since the dependent variable is overdispersed (the sample mean is 0.296 and the sample variance is 0.557), the Poisson model seems to be inadequate for estimating these count data. Figure 2 shows the HOSP distribution of the study obtained from the Poisson distribution, on the one hand, and from the proposed distribution, on the other. There is a clear spike of extra zeros representing the non-hospitalization of the elderly population, and the new distribution clearly produces a better fit to the data. Thus, the log-likelihood values obtained using the new distribution and the Poisson distribution without covariates are -3193.28 and -3304.51, respectively.

Table 1. Explanatory variables description.

EXCLHLTH	A dummy variable which takes the value 1 if self-perceived health is excellent
POORHLTH	A dummy variable which takes the value 1 if self-perceived health is poor
NUMCHRON	A count variable giving the number of chronic diseases and condition (cancer, heart attack, etc.)
AGE	Age divided by 10
MALE	A dummy variable which takes the value 1 if the patient is male
MARRIED	A dummy variable for marital status
FAMINC	Family income in $10,000
EMPLOYED	A dummy variable which takes the value 1 if the patient is employed
PRIVINS	A dummy variable which takes the value 1 if the patient is covered by private health insurance
MEDICAID	A dummy variable which takes the value 1 if the patient is covered by Medicaid

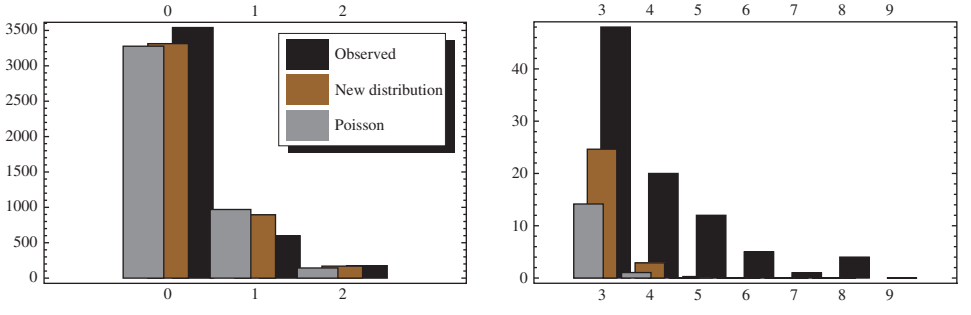

Figure 2. Distribution of the 4406 data observed for days of hospitalization, according to the Poisson method and the new distribution proposed here.

Table 2. Estimates from the new count distribution and the Poisson distribution (in parentheses).

Variable	Parameter	Estimate	S.E.	t-Wald	Pr $>\|t\|$
	β_1	−3.530 (−3.376)	0.37 (0.34)	−9.46 (−9.78)	0.00 (0.00)
EXCLHLTH	β_2	−0.725 (−0.726)	0.18 (0.17)	−4.04 (−4.13)	0.00 (0.00)
POORHLTH	β_3	0.627 (0.618)	0.07 (0.06)	8.44 (9.12)	0.00 (0.00)
NUMCHRON	β_4	0.274 (0.263)	0.02 (0.02)	13.42 (14.34)	0.00 (0.00)
AGE	β_5	0.197 (0.178)	0.04 (0.04)	4.20 (4.09)	0.00 (0.00)
MALE	β_6	0.154 (0.131)	0.06 (0.06)	2.29 (2.11)	0.02 (0.03)
MARRIED	β_7	−0.043 (−0.039)	0.07 (0.06)	−0.62 (−0.59)	0.53 (0.55)
FAMINC	β_8	0.005 (0.007)	0.01 (0.01)	0.50 (0.76)	0.61 (0.45)
EMPLOYED	β_9	0.023 (0.022)	0.11 (0.10)	0.20 (0.20)	0.83 (0.84)
PRIVINS	β_{10}	0.200 (0.197)	0.08 (0.07)	2.38 (2.53)	0.02 (0.01)
MEDICAID	β_{11}	0.227 (0.236)	0.11 (0.09)	2.08 (2.36)	0.03 (0.02)
$\ell_{\max} = -2951.33\,(-3042.83)$					
AIC = 5924.66 (6107.66)					
Vuong test = 0.725					

Table 2 presents the maximum likelihood estimates of the Poisson method and the new distribution regression model including an intercept β_1 (the regression estimate when all variables in the model are evaluated at zero). All of these, except MARRIED, FAMINC and EMPLOYED, are significant at 5%. We used the value of the maximum of the log-likelihood function ($\ell_{\max}$) and of the Akaike information criterion (AIC) [1] to compare the estimated models. A model with a minimum AIC value is to be preferred. The table shows that the new distribution surpasses the

performance of the Poisson model, both in terms of the maximized value of the log-likelihood function and of the AIC.

As in the AIC, Vuong [16] and Denuit *et al.* [10] set the information criterion in a testing framework, where the null hypothesis is that the two competing models are equally close to the true model. Under the null hypothesis that the models are indistinguishable, the test statistic is asymptotically distributed standard normal. The Vuong statistic for the new model against the Poisson regression model is equal to 0.725, providing some evidence of superiority of the new model in front of the Poisson model.

Apart from the log-likelihood and the AIC values, there are also some differences in the results obtained by the two models. For example, there are significative differences between the values of the estimated regressors. The new model predicts a higher use of the health service when self-perceived health is poor, when the number of chronic diseases is greater, with greater patient age, when the patient is male, when the patient is employed and finally, when the patient has private health insurance.

Alternative models to the Poisson and the new distribution proposed here require additional parameters. This is the case of the negative binomial distribution. We fitted the data with this distribution and obviously an improvement in the likelihood and the AIC values was obtained (-2856.42 and 5734.85, respectively). Nevertheless, the variables PRIVINS and MEDICAID are non-significant under this latter model, and because of the large number of variables considered the computation time required to obtain the maximum likelihood estimates of the parameters is much greater than in the Poisson distribution and the count distribution proposed here.

4. Final comments

The model proposed can be generalized to truncated and censored jointly dependent count regression models. Censored samples may be obtained when high counts are not observed, or may be imposed by survey design. Given the survival function of the proposed distribution, a censored model can be obtained straightforwardly. According to Greene [13], truncation applies when sample data are drawn from a subset of a larger population, and is a characteristic of the distribution from which the subset of observations in a sample is drawn. On the other hand, count data sometimes contain a large number of zeros (as is the case for the response variable studied here). When there are too many zeros, a zero-inflated distribution or a hurdle model can be used. This can be readily implemented for the proposed distribution, and will be the object of future research.

Acknowledgements

The research was partially funded by ECO200914152 (MICINN, Spain). The author is indebted to the referees for helpful comments which, without doubt, helped to improve an earlier version of the paper.

References

[1] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, in *Proceedings of the 2nd International Symposium on Information Theory*, N. Petrov and F. Csadki, eds., Akademiai Kiado, Budapest, 1973, pp. 267–281.

[2] T. Alanko and J.C. Duffy, *Compound binomial distribution for modelling consumption data*, Statistician 45(3) (1996), pp. 269–286.

[3] H.P. Bartlett and D.R. Phillips, *Ageing in the United Kingdom: A review of demographic trends, recent policy developments and care provisions*, Korean J. Popul. Dev. 24(2) (1995), pp. 181–195.

[4] A.C. Cameron and P.K. Trivedi, *Regression Analysis of Count Data*, Cambridge University Press, New York, 1998.

[5] C. Cameron, P.K. Trivedi, F. Milne, and J. Piggot, *A microeconometric model of the demand for health care and health insurance in Australia*, Rev. Econ. Stud. 55 (1988), pp. 85–106.

[6] W. Cartwright, S.T. Hu, and L.-F. Huang, *Impact of varying Medigap insurance coverage on the use of medical services of the elderly*, Appl. Econ. 2(4) (1992), pp. 529–539.
[7] S. Christensen, S. Long, and J. Rodgers, *Acute health care costs for the aged Medicare population: Overview and policy options*, Millbank Q. 65 (1987), pp. 397–425.
[8] C. Dean, J.F. Lawless, and G.E. Willmot, *A mixed Poisson-inverse-Gaussian regression model*, Can. J. Stat. 17(2) (1989), pp. 171–181.
[9] P. Deb and P.K. Trivedi, *Demand for medical care by the elderly: A finite mixture approach*. J. Appl. Econ. 12(3) (1997), pp. 313–336.
[10] M. Denuit, X. Maréchal, S. Pitrebois, and J-F. Walhin, *Actuarial Modelling of Claims Counts. Risk Classification, Credibility and Bonus-Malus Systems*, John Wiley & Sons Ltd, Chichester, UK, 2007.
[11] N. Duan, W. Manning, C. Morris, and J. Newhouse, *A comparison of alternative models for the demand for medical care*, J. Bus. Econ. Stat. 1 (1983), pp. 115–126.
[12] F. Famoye, *Generalized binomial regression model*, Biom. J. 37(5) (1995), pp. 581–594.
[13] W. Greene, *Econometric Analysis*, Vol. 3, Prentice Hall, Upper Saddle River, NJ, 2003.
[14] Y. Hu, X. Peng, T. Li, and H. Guo, *On the Poisson approximation to photon distribution for faint lasers*, Phys. Lett. 367 (2007), pp. 173–176.
[15] A. Min and C. Czado, *Testing for zero-modification in count regression models*, Stat. Sin. 20 (2010), pp. 323–341.
[16] Q.H. Vuong, *Likelihood ratio tests for model selection and non-nested hypotheses*, Econometrica 57(2) (1989), pp. 307–333.
[17] S. Wolfram, *The Mathematica Book*, Wolfram Media, Inc., Champaign, IL, 2003.
[18] X. Zhao and X. Zhou, *Estimation of medical costs by copula models with dynamic change of health status*, Insur. Math. Econ. 51 (2012), pp. 480–491.