Journal of Statistical Theory and Applications
Volume 11, Number 3, 2012, pp. 225-235 ISSN 1538-7887

Some Generalizations of Poisson Processes

Seetha Lekshmi, V.*
Dept.of Statistics, Nirmala College
Muvattupuzha, Kerala, India

Catherine Thomas†
Dept. of Statistics, Govt. College
Thrippunithura, Kerala, India

Abstract

In this paper we make an attempt to review count data models developed so far as generalizations of Poisson process. We consider Winkleman's gamma count model and the Weibull count model of Mc Shane et al. The fractional generalization of Poisson process by Mainardi et al. is also considered. A Mittag-Leffler count model is developed and studied in detail. Simulation studies are also conducted.

Key Words and Phrases: Count models, Exponential distribution, Gamma distribution, Mittag-Leffler distribution, Poisson distribution, Weibull distribution

*E-mail: seethalekshmi@gmail.com
†E-mail: catherinebiju@gmail.com

1 Introduction

Count data frequently arise as outcomes of an underlying count process in continuous time. The Poisson count models have widespread popularity among count models. In Poisson count model, the number of arrivals in a given period of time is derived by assuming that the inter arrival time is distributed as exponential. But this model is valid only for the case where data of interest support the restrictive assumption of equi-dispersion. Now researchers are in search of models that allow over dispersion (variance greater than the mean) and under dispersion (mean greater than variance). A number of count models with inter arrival times following Gamma, Weibull etc. are developed by many researchers.The hazard function captures the underlying time dependence of the process. A decreasing hazard function implies that the waiting time is less likely to end the longer it lasts. This situation is referred to as negative duration dependence. An increasing hazard function implies that the waiting time is more likely to end the longer it lasts. This situation is referred as positive duration dependence. No duration dependence corresponds to the case of a constant hazard. The hazard is a constant if and only if the distribution of waiting time is exponential.

Now we consider a general frame work utilized to describe the model that is based upon the equivalence relationship between the inter arrival times and their count model. Let $(\tau_k, k \in N)$ denote the sequence of waiting time between the $(k-1)^{th}$ and k^{th} events. Then the arrival time of the n^{th} event is given by

$$\vartheta_n = \sum_{k=1}^{n} \tau_k, n = 1, 2, ... \tag{1}$$

Let $N(t)$ represent the total number of events in the open interval $(0, t)$. For fixed t, $N(t)$ is a count variable. It follows from the definitions of $N(t)$ and ϑ_n that

$$N(t) < n \Leftrightarrow \vartheta_n \geq t \tag{2}$$

Thus,

$$P[N(t) < n] = P[\vartheta_n \geq t] = 1 - F_n(t) \tag{3}$$

Furthermore,

$$P[N(t) = n] = P[N(t) < n + 1] - P[N(t) < n] = F_n(t) - F_{n+1}(t) \tag{4}$$

Equation (4) provides the fundamental relation between the distribution of arrival times and the distribution of counts. The probability distribution of $N(t)$ can be obtained explicitly for all n from the knowledge of the distribution of ϑ_n.
Let $N(t), t \geq 0$ be a renewal process generated by the distribution function F. Then the process obtained by retaining the points of N(t) with probability p and deleting them with probability (1-p) is called a p-thinned process. Thinning arises in many practical situations where only some of the events are reported. The case of crimes, or the persons affected by particular disease are typical examples of such phenomenon. In such cases we get only a thinned versions of the original process.
The rest of the paper is organized as follows. Gamma count modelis discussed in section 2. The Weibull count model is discussed in section 3. Mittag- Leffler count model is considered in section 4. Fractional generalization of Poisson Process is discussed in section 5. Finally some concluding remarks were given in section 6.

2 Gamma Count Model

Winkelmann (1995) derived a generalized model by replacing the exponential distribution with a less restrictive non negative distribution, say, gamma distribution. This distribution has non constant hazard function. To derive the model he makes use of the fact that sums of independent gamma distributions are again distributed as gamma. An advantage of this generalized model is that it provides a count data model of substantially higher flexibility than Poisson model at the cost of more additional parameter. It provides an interpretation of over and under dispersion in terms of an underlying sequence of waiting times. This model finds application in the analysis of accident proneness (for instance, airline accidents), labor mobility(the number of changes of employer), the demand for health care services (as measured by the number of doctor consultations in a given time interval) and in economic demography, total fertility (the number of births by a woman).
Let us assume that the waiting times τ_k, are identically and independently distributed(i.i.d.) as gamma. Dropping the index k the density can be written as

$$f(\tau : \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau}; \tau > 0, \alpha, \beta \in R^{+} \tag{5}$$

It has mean $E(\tau) = \frac{\alpha}{\beta}$ and variance $var(\tau) = \frac{\alpha}{\beta^2}$. The hazard function $\lambda(t) = \frac{f(t)}{1-F(t)}$ obeys the equation

$$\frac{1}{\lambda(t)} = \int_0^\infty e^{-\beta u}(1+\frac{u}{\tau})^{\alpha-1}\, du \tag{6}$$

The gamma distribution admits no closed form expression for the tail probabilities and thus no simple formula for the hazard function. From (6) it follows that $\lambda(t)$ is (monotonically) increasing for $\alpha > 1$, decreasing for $\alpha < 1$, and constant (and equal to β) for $\alpha = 1$. Now, consider the arrival time of the n^{th} event

$$\vartheta_n = \tau_1 + \tau_2 + ... + \tau_n, n = 1, 2, ... \tag{7}$$

where τ_i are iid gamma distributed. The reproductive property of the gamma distribution implies that ϑ_n is gamma distributed with density

$$f(\vartheta_n : \alpha, \beta) = \frac{\beta^{n\alpha}}{\Gamma(n\alpha)}\vartheta^{n\alpha-1}e^{-\beta\vartheta} \tag{8}$$

To derive the new count data distribution, we have to evaluate the cumulative distribution function

$$F_n(t) = \frac{1}{\Gamma(n\alpha)}\int_0^{\beta t} u^{n\alpha-1}e^{-u}\, du \tag{9}$$

where the integral is the incomplete gamma function. The right side will be denoted as $G(\alpha n, \beta t)$. Note that $F_0(t) = 1$. The number of event occurrences during the time interval (0,t) has the two parameter distribution function

$$P[N(t) = n] = G(\alpha n, \beta t) - G(\alpha n + \alpha, \beta t) \tag{10}$$

For n=0,1,2,...where $\alpha, \beta \in R^+, G(0, \beta t) = 1$.

Integrating (10) by parts gives

$G(\alpha n, \beta t) = 1 - e^{-\beta t}(1 + \beta t + \frac{(\beta t)^2}{2!} + ... + \frac{(\beta t)^{n\alpha-1}}{(n\alpha-1)!})$

Hence,

$$P[N(t) = n] = G(\alpha n, \beta t) - G(\alpha n + \alpha, \beta t) = e^{-\beta t}\sum_{i=0}^{\alpha-1}\frac{(\beta t)^{\alpha n+i}}{(\alpha n+i)!}, n = 0, 1, 2, ... \tag{11}$$

This can also be written as $P[N(t) = n] = e^{-\beta t}\sum_{r=nk}^{(n+1)k-1}\frac{(\beta t)^r}{r!}$.

For $\alpha = 1$, $f(\tau)$ is the exponential density and (11) simplifies to the Poisson distribution.For non integer values of α, no closed form expression is available for $G(\alpha n, \beta t)$ and thus for $P[N(t) = n]$. Numerical evaluations of the integral can be done based on asymptotic expansions.

3 Weibull Count Model

Winkelmann (1995) points out that the Weibull distribution can be preferred in duration analysis for its closed form hazard function. But he does not pursue such a model. Mc Shane et.al (2008) derived count models based on Weibull inter arrival times. They developed a Weibull inter arrival process which nests the exponential as a well known special case. This model via shape parameter being less than, equal to, or greater than one can capture over dispersed, equi dispersed and under dispersed data respectively. The basic Weibull count model was derived by assuming that the inter arrival times are independent and identically distributed as Weibull with probability density function(pdf) $f(t) = act^{c-1}e^{-at^c}(c, a \in R^+)$, and the corresponding cumulative density function (cdf) $F(t) = 1 - e^{-at^c}$, which simplifies to the exponential model when $c = 1$.

Using Taylor series expansion, the cdf and pdf of the Weibull distribution are obtained as,

$$F(t) = \sum_{j=1}^{\infty} \left[\frac{(-1)^{j+1}(at^c)^j}{\Gamma(j+1)} \right] \tag{12}$$

$$f(t) = \sum_{j=1}^{\infty} \left[\frac{(-1)^{j+1}cja^j t^{cj-1}}{\Gamma(j+1)} \right] \tag{13}$$

Using (4), we obtain the following recursive relationship for the Weibull count model.

$$\begin{aligned} P_n(t) &= \int_0^t F_{n-1}(t-s)f(s)ds - \int_0^t F_n(t-s)f(s)ds \\ &= \int_0^t P_{n-1}(t-s)f(s)ds \end{aligned} \tag{14}$$

Before proceeding to develop the general solution to the problem we note that $F_0(t) = 1$ for all t and $F_1(t) = F(t)$. In general the Weibull count model probabilities are given by

$$P(N(t) = n) = P_n(t) = \sum_{j=n}^{\infty} \left[\frac{(-1)^{j+n}(at^c)^j \alpha_j^n}{\Gamma(cj+1)} \right]; n = 0, 1, 2, ...$$

where $\alpha_j^0 = \frac{\Gamma(cj+1)}{\Gamma(j+1)}; j = 0, 1, 2, ...$ and

$$\alpha_j^{n+1} = \sum_{m=n}^{j-1} \left[\alpha_m^n \frac{\Gamma(cj - cm + 1)}{\Gamma(j - m + 1)} \right]; n = 0, 1, 2, ..for\ j = n + 1, n + 2, ...$$

It may be noted that when c=1, t=1, this model reduces to the Poisson count model. This model can handle both over dispersed and under dispersed data.

4 Mittag-Leffler Count Model

Pillai (1990) introduced Mittag Leffler distribution. Jayakumar and Pillai (1993) used Mittag-Leffler distributions to model autoregressive time series. The basic Mittag-Leffler count model was derived by assuming that the inter arrival times are independent and identically distributed as Mittag-Leffler distribution with probability density function

$$f(x) = \sum_{k=1}^{\infty} \left[\frac{(-1)^{k-1} k\alpha (ax)^{k\alpha - 1}}{\Gamma(1 + k\alpha)} \right], x > 0, 0 < \alpha \leq 1, a > 0$$

and the corresponding cumulative density function (cdf)

$$F(x) = \sum_{k=1}^{\infty} \left[\frac{(-1)^{k-1} (ax)^{k\alpha - 1}}{\Gamma(1 + k\alpha)} \right], x > 0, 0 < \alpha \leq 1, a > 0$$

which simplified to the exponential model when $\alpha = 1$.

Equation (4) provided the fundamental relation between the distribution of arrival times and the distribution of counts. Let $P_n(t) = P[N(t) = n] = F_n(t) - F_{n+1}(t)$.To obtain equation (4) we use the recursive relationship of the form

$$P_n(t) = \int_0^t F_{n-1}(t - s) f(s) ds - \int_0^t F_n(t - s) f(s) ds$$

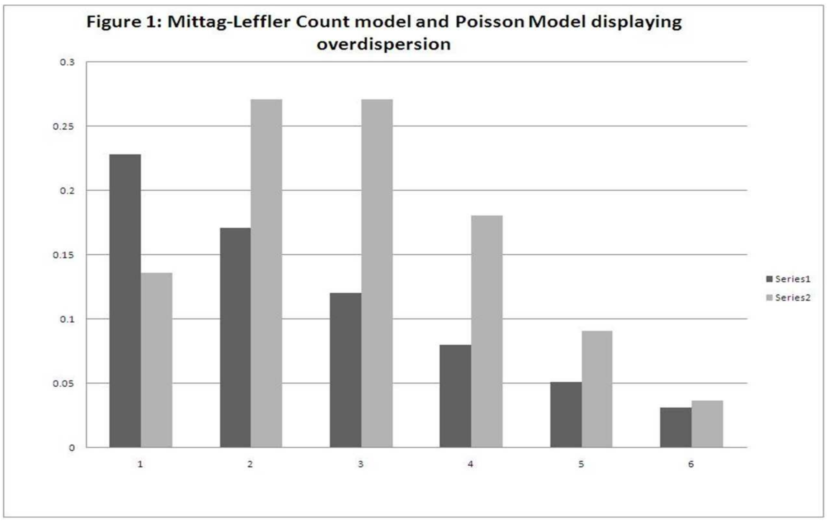
$$= \int_0^t P_{n-1}(t - s) f(s) ds$$

It may be noted that $F_0(t) = 1$ and $F_1(t) = F(t)$ for every t. Therefore we have
$P_0(t) = F_0(t) - F_1(t) = 1 - F(t)$.

Following the steps in equation(14) we obtain

$$P_n(t) = \sum_{j=n}^{\infty} \left[\frac{(-1)^{(j-n)} \binom{j}{n} (at)^{j\alpha}}{\Gamma(1 + j\alpha)} \right], n = 0, 1, 2, ...$$

Now we have the following results.



(i) When $\alpha = 1$, Mittag-Leffler count model reduces to Poisson count model. Hence

$$P_n(t) = \sum_{j=n}^{\infty} \left[\frac{(-1)^{(j-n)} \binom{j}{n} (at)^{j\alpha}}{\Gamma(1 + j\alpha)} \right] = \frac{e^{-at}(at)^n}{n!}$$

(ii) The mean and variance of the Mittag-Leffler count model exists. $Mean = \frac{(at)^\alpha}{\Gamma(1+\alpha)}$ and $Variance = \frac{(at)^\alpha}{\Gamma(1+\alpha)} + \frac{2(at)^{2\alpha}}{\Gamma(1+2\alpha)} - \left[\frac{(at)^\alpha}{\Gamma(1+\alpha)}\right]^2$.

(iii) The hazard function of the Mittag-Leffler distribution is a decreasing function of time. Therefore the distribution displays negative duration dependence. Negative duration dependence causes over dispersion. This model can handle over dispersed as well as equi dispersed data when $0 < \alpha < 1$ and $\alpha = 1$ respectively.

(iv) If $P_n(t), t > 0$ is a Mittag-Leffler count process, then the p-thinned process is also a Mittag-Leffler process for each $p \in (0, 1)$. Details on p-thinned processes are available in Anil(2001).

(v) It can be seen that this distribution coincides with alpha Poisson distribution given by Anil(2001). It is worthwhile to say that different properties of the Mittag-Leffler distribution, generalized Mittag-Leffler distribution and geometric Mittag-Leffler distribution are discussed in Jose et al. (2010), Seetha Lekskmi and Jose (2002,2004).

As a demonstration of these findings Figure 1 displays the probability histograms for the Mittag-Leffler and Poisson count models. Both the Mittag-Leffler and Poisson models are intentionally chosen to have identical means but their variances are different. In figure 1, we have the probability histograms for an over dispersed Mittag Leffler distribution parameters $\alpha = 0.5$ and $a = 3.14$.

5 Fractional Generalization of Poisson Process

Mainardi et al. (2004,2005) provide a generalization of pure Poisson process via fractional calculus. For a renewal process we have, $\vartheta_n = \sum_{k=1}^{n} \tau_k, n = 1, 2, ...$ Let $\phi(t)$ denote the density function and $\Phi(t)$ denote the cumulative distribution function.Then we have,

$$\phi(t) = \tfrac{d}{dt}\Phi(t) \text{ or } \Phi(t) = P[T \leq t] = \int_0^t \phi(t^{'})dt^{'}$$

In other words, if a non negative random variable represents the life time of the technical system, then $\Phi(t)$ is referred to as failure probability and $\Psi(t) = P[T \geq t] = 1 - \Phi(t)$ refers to the survival probability. Let $N(t)$ denote the counting function

$$N(t) = \begin{cases} 0 & for \quad 0 \leq t < t_1 \\ max\,\{k|\tau_k \leq t, k = 0, 1, 2, , ...\} & for \quad t \geq \tau_1 \end{cases}$$

represents the number of events occurred before t. In particular $\Psi(t) = P[N(t) = 0]$. Set $F_1(t) = \Phi(t); f_1(t) = \phi(t)$ and $F_k(t) = P[\tau_k \leq t]$ and $f_k(t) = \frac{d}{dt}F_k(t), k \geq 1$, where $F_k(t)$ represents the sum of the first k waiting times is less than or equal to t, and f_k(t)be its density. Then $P[\tau_k \leq t, \tau_{k+1} > t] = \displaystyle\int_0^t f_k(t^{'})\Psi(t - t^{'})dt^{'}$. It can be easily recognized that $f_k(t)$ turns out to be the k- fold convolution of $\Phi(t)$ with itself.

If the inter arrival time distribution is exponential, then we have,

$$P[N(t) = k] = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, t \geq 0, k = 0, 1, 2, ... \tag{15}$$

The probability distribution related to the sum of k i.i.d. exponential random variables is known as the Erlang's distribution of order k. The corresponding density is given by

$$f_k(t) = \frac{e^{-\lambda t}(\lambda)^k(t)^{k-1}}{\Gamma(k)}, t \geq 0, k = 0, 1, 2, ... \tag{16}$$

The Erlang's distribution function of order k turns out to be

$$F_k(t) = \int_0^t f_k(t^{'})dt^{'} = 1 - \sum_{n=0}^{k-1} \left[\frac{e^{-\lambda t}(\lambda t)^n}{n!}\right] = \sum_{n=k}^{\infty} \left[\frac{e^{-\lambda t}(\lambda t)^n}{n!}\right], t \geq 0.$$

They also establish that the survival probability for the Poisson renewal process obeys the ordinary differential equation

$$\frac{d}{dt}\Psi(t) = -\lambda\Psi(t), t \geq 0; \Psi(0^+) = 1 \tag{17}$$

A fractional generalization of the Poisson renewal process is simply obtained by generalizing the differential equation (17) by replacing the first derivative with integro-differential operator tD_*^α that is interpreted as the fractional derivative of order α in Caputo's sense. The Caputo derivative of order $\alpha \in (0, 1]$ of a well-behaved function $f(t)$ in R^+ is

$$tD_*^\alpha f(t) = \begin{cases} \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{f^{(1)}(\tau)}{(t-\tau)^\alpha d\tau} & 0 < \alpha < 1 \\ \frac{d}{dt} f(t) & \alpha = 1 \end{cases}$$

It can alternatively be written in the form

$$tD_*^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t \frac{f(\tau)}{(t-\tau)^\alpha} d\tau - \frac{t^{-\alpha}}{\Gamma(1-\alpha)} f(0^+)$$

$$= \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t \frac{f(\tau) - f(0^+)}{(t-\tau)^\alpha} d\tau, 0 < \alpha < 1$$

Writing the equation(17), for $\lambda = 1$, we get

$$tD_*^\alpha \Psi(t) = -\Psi(t), t \geq 0, \qquad 0 < \alpha < 1, \qquad \Psi(0^+) = 1 \tag{18}$$

The solution of equation (18) is given by

$$\Psi(t) = E_\alpha(-t^\alpha), t \geq 0, 0 < \alpha < 1$$

Hence we have, $\phi(t) = -\frac{d}{dt}\Psi(t) = -\frac{d}{dt}E_\alpha(-t^\alpha), t \geq 0, 0 < \beta < 1.$

Here $E_\beta(z)$ represents the Mittag-Leffler function with parameter β defined in the complex plane by the power series given by $E_\alpha(z) = \sum_{n=0}^{\infty}\left[\frac{z^n}{\Gamma(1+\alpha n)}\right], \alpha > 0, z \in C$. Hence as a generalization of Poisson process , the Mittag-Leffler process is obtained. If the inter arrival distributions are Mittag-Leffler distributions, then the generalization of equation (15) is given by

$$P[N(t) = k] = \frac{t^{k\alpha}}{k!} E_\alpha(-t^\alpha), t \geq 0, k = 0, 1, 2, ...$$

This is referred to as the generalized Poisson distribution. The generalization of equation (16) is given by

$$f_k(t) = \alpha \frac{t^{k\alpha-1}}{(k-1)!} E_\alpha^{(k)}(-t^\alpha)$$

where $E_\alpha^{(k)} = \frac{d^k}{dz^k} E_\alpha(z)$.

This is referred to as generalized Erlang's distribution. Further more, it can be shown that the Mittag-Leffler distribution is the limiting distribution under thinning of a generic renewal process with waiting time density of power law character.

6 Conclusions

In this paper we make an attempt to review count data models developed so far as generalizations of Poisson process. The most popular model among these is the Poisson count model. But this model is valid only for the case where the data of interest supports the restrictive assumption of equi-dispersion. But the other models allow both over and under dispersion. Also we consider the fractional generalization of the Poisson process by viewing the count model as a renewal process. Further works on generalized Mittag-Leffler count models are progressing.

Acknowledgements

The authors acknowledge the financial support from the University Grants Commission, New Delhi for carrying out this research. The authors also express their gratitude to the reviewers and editor for the valuable suggestions for improving the presentation of results in the paper.

References

[**1**] Anil,V.(2001). A generalized Poisson distribution and its applications. Journal of the Kerala Statistical Association,Vol.12, Dec.2001, 11-22.

[**2**] Jayakumar, K. and Pillai,R.N.(1993).The first order autoregressive Mittag-Leffler process. Journal of Applied Probability, 30, 462-466.

[**3**] Jose,K.K., Uma,P., Seetha Lekshmi, V. and Haubold, H.J.(2010).Generalized Mittag-Leffler Distributions and Processes for Applications in Astrophysics and Time Series Modeling. Proceedings of the Third UN/ESA/NASA Workshop on the International Helio Physical Year, 2007, Springer-Verlag, Berlin.

[**4**] McShane,B., Moshe Adrian, Eric T. Bradlow and Peter S. Fader(2008).Count models based on Weibull inter arrival times.Journal of Business and Economics Statistics,Vol.25, No.3, 369-377 .

[**5**] Mainardi,F., Gorenflo,R. and Scalas,E.(2004).A fractional generalization of the Poisson Processes. Vietnam Journal of Mathematics, Vol.32,53-64.

[**6**] Mainardi,F., Gorenflo,R.and Vivoli,A.(2005). Renewal processes of Mittag-Leffler and wright type. Fractional Calculus and Applied Analysis, An International Journal for Theory and Applications, Vol.8,No.1,7-38.

[**7**] Pillai, R.N.(1990).On Mittag-Leffler functions and its related distributions. Annals of Institute of Statistical Mathematics,42,157-161.

[**8**] Seetha Lekshmi, V. and Jose, K.K.(2002). Geometric Mittag-Leffler Tailed autoregressive processes. Far East Journal of Theoretical Statistics, 6(2) 147-153.

[**9**] Seetha Lekshmi, V. and Jose, K.K.(2004). Geometric Mittag-Leffler distributions and processes. Journal of Applied Statistical Science, 13,4,335-342.

[**10**] Winkelmann,R.(1995). Duration Dependence and Dispersion in Count-Data Models. Journal of Business and Economic Statistics, October, Vol. 13, No. 4, 467-474.