Far East Journal of Theoretical Statistics
Volume 44, Number 1, 2013, Pages 83-92
Published Online: August 2013
Available online at http://pphmj.com/journals/fjts.htm
Published by Pushpa Publishing House, Allahabad, INDIA

COUNT MODEL WITH GAMMA AND ERLANG INTERARRIVAL TIME DISTRIBUTIONS

V. Seetha Lekshmi and Catherine Thomas

Department of Statistics
Nirmala College
Muvattupuzha
Kerala, India
e-mail: seethalekshmi@gmail.com

Department of Statistics
Government College
Tripunithura
Kerala, India
e-mail: catherinebiju@gmail.com

Abstract

In this paper, a new count model with gamma interarrival time distribution is introduced and its properties are studied. The Erlang model is developed as a special case and studied in detail. Various characteristics like mean function, variance function, probability generating function, hazard rate function, etc. are derived. The model is applied to a real data set on interarrival times of customers arrived in a bank counter at Muvattupuzha, Kerala, India.

2010 Mathematics Subject Classification: 60K05.

Keywords and phrases: count model, gamma count model, Erlang count model.

Communicated by K. K. Azad

Received March 9, 2013

1. Introduction

Poisson model is the basic regression model for count data (i.e., the number of events in a given interval of time). For the Poisson model the conditional variance equals the conditional mean. In most real situations conditional variance exceeds the conditional mean (over-dispersion) or conditional variance is less than conditional mean (under-dispersion). In these cases, Poisson model is inefficient and leads to biased inference (Winkelmann [5]).

Note that Poisson process corresponds to a sequence of independent and identically distributed exponential waiting times (interarrival times) (see Cox [2]). A generalized model can be derived by replacing the exponential distribution by another nonnegative distribution. Weibull and gamma nest the exponential distribution and allow for non-constant (monotone) hazard function (duration dependence). Weibull distribution is mostly preferred in duration analysis for its closed form hazard function. But gamma distribution is preferred here for its reproductive property (since sum of independent gamma random variables is again gamma distributed).

Recently many researchers have attempted to develop new count models as alternatives to the Poisson count process. McShane et al. [4] developed a Weibull count model whereas Jose and Abraham [3] developed a Mittag-Leffler Count model. Seetha Lekshmi and Catherine [6] also developed some generalizations of Poisson count process.

In this paper, we develop a new approach for constructing count models to obtain a generalized Poisson model by replacing the exponential distribution by gamma distribution which is a generalization of exponential distribution. This model has the following merits: First, our count model is based upon an assumed gamma interarrival process which nests the exponential as a special case. Second, this model can handle over-dispersed, under-dispersed as well as equi-dispersed data. Third, the gamma interarrival time model is richer than the exponential model because it allows non-constant hazard rates (duration dependence). Fourth, we can implement the model entirely in standard computing software.

The rest of the paper is organized as follows. In Section 2, gamma count model is discussed. Erlang count model is introduced and discussed in Section 3. Application to a real data set on customer arrivals in a bank counter is considered in Section 4.

2. Gamma Count Model

Let us describe a general framework utilized to describe the model that is based upon the relationship between the interarrival times and their count model equivalent. Let τ_k be the waiting time between $(k-1)^{\text{th}}$ and k^{th} events. The arrival time of the n^{th} event is given by

$$\nu_n = \sum_{k=1}^{n} \tau_k, \quad n = 1, 2, 3, \ldots.$$

Let us assume that the waiting times are independent and identically distributed gamma variables. The density can be written as

$$f(t, \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta t} t^{\alpha-1}; \quad \alpha, \beta \in R^{+}. \tag{1}$$

For $t > 0$, $E(t) = \dfrac{\alpha}{\beta}$ and $Var(t) = \dfrac{\alpha}{\beta^2}$, the hazard function is given by $\lambda(\tau) = \dfrac{f(t)}{\overline{F}(t)}$, where $\overline{F}(t) = 1 - F(t)$ and $F(t)$ is the distribution function. By reproductive property of the gamma distribution,

$$f(v_n; \alpha, \beta) = \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} e^{-\beta v} v^{n\alpha-1}. \tag{2}$$

As a new approach the cumulative distribution is obtained using the relationship between incomplete gamma integral and Poisson probabilities, so that,

$$\begin{aligned} F_n(t) &= G(\alpha n, \beta t) \\ &= \frac{1}{\Gamma(n\alpha)} \int_0^{\beta t} u^{n\alpha-1} e^{-u} du \end{aligned}$$

$$= \sum_{u=n\alpha}^{\infty} \frac{e^{-\beta t}(\beta t)^u}{u!},$$

$$F_{n+1}(t) = \sum_{u=(n+1)\alpha}^{\infty} \frac{e^{-\beta t}(\beta t)^u}{u!}.$$

Hence we can obtain

$$P_n(t) = P[N(t) = n]$$

$$= F_n(t) - F_{n+1}(t)$$

$$= \sum_{u=n\alpha}^{(n+1)\alpha-1} \frac{e^{-\beta t}(\beta t)^u}{u!}, \quad n = 0, 1, 2, \ldots. \tag{3}$$

Table 1 and Table 2 give values of gamma count model probabilities for different values of α and β at $t = 1, 2, 3$.

Table 1

				$\alpha = 1$					
	$t = 1$			$t = 2$			$t = 3$		
β	$P_1(t)$	$P_2(t)$	$P_3(t)$	$P_1(t)$	$P_2(t)$	$P_3(t)$	$P_1(t)$	$P_2(t)$	$P_3(t)$
0.5	0.3033	0.0758	0.0126	0.3679	0.1839	0.0613	0.3347	0.2510	0.1255
1	0.3678	0.1839	0.0613	0.2707	0.2707	0.1805	0.1493	0.2240	0.2240
1.5	0.3347	0.2510	0.1255	0.1493	0.2240	0.2240	0.0499	0.1125	0.1687
2	0.2707	0.2707	0.18044	0.0733	0.1465	0.1954	0.0149	0.0446	0.0892
2.5	0.2052	0.2565	0.2138	0.0337	0.0842	0.1404	0.0041	0.0157	0.0389
3	0.1494	0.2240	0.2240	0.0149	0.0446	0.0892	0.0011	0.0050	0.0149
3.5	0.1057	0.1850	0.2158	0.0064	0.0223	0.0521	0.0003	0.0015	0.0053
4	0.0733	0.1465	0.1954	0.0027	0.0107	0.0286	0.0001	0.0004	0.0017

Table 2

					$\alpha = 2$				
	$t = 1$			$t = 2$			$t = 3$		
β	$P_1(t)$	$P_2(t)$	$P_3(t)$	$P_1(t)$	$P_2(t)$	$P_3(t)$	$P_1(t)$	$P_2(t)$	$P_3(t)$
0.5	0.0884	0.0017	0.00001	0.2453	0.0184	0.0006	0.3765	0.0471	0.0043
1	0.2453	0.0184	0.0006	0.4511	0.0902	0.0155	0.4481	0.1344	0.0720
1.5	0.3765	0.0612	0.0043	0.4481	0.2689	0.0720	0.2812	0.1898	0.2105
2	0.4511	0.1263	0.0155	0.3418	0.1954	0.1637	0.1338	0.18739	0.2983
2.5	0.4703	0.2004	0.0378	0.2246	0.3509	0.2507	0.0544	0.0729	0.2832
3	0.4481	0.2689	0.0720	0.1339	0.1339	0.2983	0.0199	0.0675	0.2082
3.5	0.4007	0.3209	0.1156	0.0745	0.2189	0.2980	0.0063	0.0139	0.1281
4	0.3419	0.3517	0.1637	0.0393	0.0572	0.2617	0.0022	0.1380	0.0692

3. Erlang Count Model

In this section, we consider the gamma count model given by (3) for integer values of α and this model can be referred to as Erlang count model. A detailed study of Erlang count model is carried out using Laplace transform techniques. Mean function, variance function, probability generating function and hazard rate function are considered. The Erlang count model probabilities are obtained as

$$P_n(t) = \sum_{u=n\alpha}^{(n+1)\alpha-1} \frac{e^{-\beta t}(\beta t)^u}{u!}, \quad n = 0, 1, 2, \ldots.$$

This means that it is a count model with interarrival time distributed as Erlang with parameters β and α, where α is an integer. The Erlang distribution with parameters β and α is given by

$$f(x; \beta, \alpha) = \frac{\beta^\alpha}{\Gamma\alpha} e^{-\beta x} x^{\alpha-1}; \ \alpha = 1, 2, \ldots, \quad \beta > 0$$

$$= 0 \qquad \text{otherwise.}$$

3.1. Mean function

Expected number of counts for the Erlang model is

$$M(t) = E(N(t)) = \sum_{n=1}^{\infty} F_n(t). \tag{4}$$

Taking Laplace transform on both sides, we get

$$M^*(s) = \frac{f^*(s)}{s(1 - f^*(s))}. \tag{5}$$

We have $f^*(s) = \left(\frac{\beta}{s+\beta}\right)^{\alpha}$.

Hence, $M^*(s) = \frac{\beta^{\alpha}}{s[(s+\beta)^{\alpha} - \beta^{\alpha}]}$.

It can easily be verified that for $\alpha = 1$, $M^*(s) = \frac{\beta}{s^2}$.

Mean function, $M(t) = \beta t$.

For $\alpha = 2$, we have $M^*(s) = \frac{\beta^2}{s(s^2 + 2\beta s)}$. Hence, $M(t) = \frac{1}{2}\beta t - \frac{1}{4}$ $+ \frac{1}{4} e^{-2at}$.

3.2. Variance function

The variance function can be obtained from the relation

$$V(t) = E[N(t)]^2 - [E[N(t)]]^2. \tag{6}$$

The Laplace transform of $E[(t)]^2$ is given by

$$M_1^*(s) = \frac{\beta^{\alpha}[(s+\beta)^{\alpha} + \beta^{\alpha}]}{s[(s+\beta)^{\alpha} - \beta^{\alpha}]^2}.$$

When $\alpha = 1$, $V(t) = \beta t$.

When $\alpha = 2$, $V(t) = \frac{1}{4}\beta t + \frac{1}{16} - \frac{1}{2}\beta t e^{-2\beta t} - \frac{1}{16} e^{-4\beta t}$.

3.3. Probability generating function

The probability generating function of the model is given by

$$P(t, z) = \sum_{n=0}^{\infty} z^n P[N(t) = n].$$

The corresponding Laplace transform is

$$P^*(s, z) = \frac{(s+\beta)^\alpha - \beta^\alpha}{s[(s+\beta)^\alpha - z\beta^\alpha]}.$$

When $\alpha = 1$, $P(t, z) = e^{-\beta t(1-z)}$.

When $\alpha = 2$, $P(t, z) = \dfrac{e^{-\beta t(1+\sqrt{z})}}{2\sqrt{z}}[1 + e^{2\beta t\sqrt{z}}(1+\sqrt{z}) + \sqrt{z}]$.

3.4. Hazard rate function

The hazard rate function is given by

$$h(t) = \frac{f(t)}{\overline{F}(t)},$$

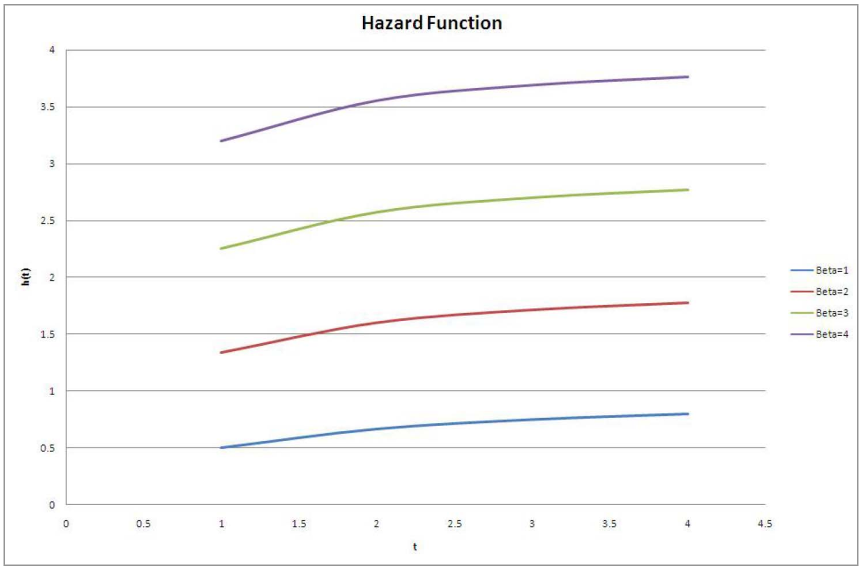
where $\overline{F}(t) = P_0(t)$. Hence,

$$h(t) = \frac{\dfrac{\beta^\alpha e^{-\beta t} t^{\alpha-1}}{(\alpha-1)!}}{\sum_{u=0}^{\alpha-1} \dfrac{e^{-\beta t}(\beta t)^u}{u!}}.$$

When $\alpha = 1$, $h(t) = \beta$, $\alpha = 2$, $h(t) = \dfrac{\beta^2 t}{1+\beta t}$ and $\alpha = 3$,

$$h(t) = \frac{\beta^3 t^2}{2\left(1 + \frac{\beta t}{1!} + \frac{(\beta t)^2}{2!}\right)}.$$

The following figure displays the hazard rate function of Erlang count model for $\alpha = 2, \ \beta = 1, 2, 3, 4.$ Clearly, it is increasing function of time t.



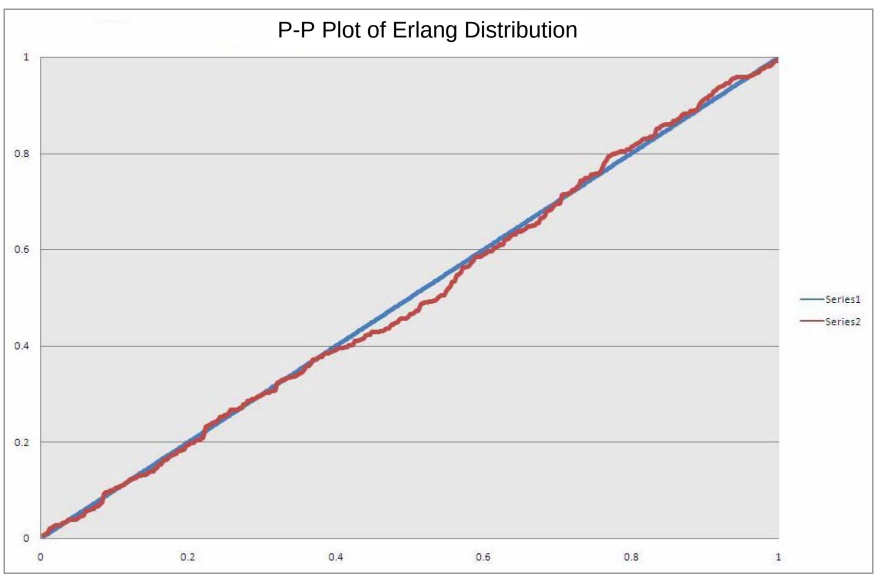
If the hazard function is a decreasing function of time, then the distribution displays negative duration dependence. If the hazard function is an increasing function of time, then the distribution displays positive duration dependence. In both cases the conditional probability of a current occurrence depends on the time since the last occurrence rather than on the number of previous events.

There is a link between duration dependence and dispersion. Without making the assumptions on the exact distribution of τ, a limiting result can be obtained. Denote the mean and the variance of the waiting time distribution by $E(\tau) = \mu$ and $Var(\tau) = \sigma^2$ and the coefficient of variation by $\nu = \frac{\sigma}{\mu}$. Let $\lambda(t)$ be the hazard function. The distribution displays negative duration dependence for $\frac{d\lambda(t)}{dt} < 0$ and positive duration dependence for $\frac{d\lambda(t)}{dt} > 0$. Assume that the hazard function is monotonic. Then by Barlow and Proschan [1], $\frac{d\lambda(t)}{dt}(< 0 = 0 > 0) \Rightarrow \nu(> 1 = 1 < 1)$.

4. Application to a Real Data Set

In this section, we apply the model to a data on the interarrival time of customers arrived on a given day in 2011 at a bank counter in Muvattupuzha, Kerala, India. All interarrival times are expressed in minutes and total number of customers arrived in the bank on a randomly selected day is 350. Here the conditional mean is greater than the conditional standard deviation (mean = 7.649 and S.D = 5.404). Thus this data set is under-dispersed and hence we can apply the Erlang count model. Using maximum likelihood method of estimation, the values of the parameters are given by $\alpha = 2$ and $\beta = 3.814.$ To test whether there is significant difference between the observed interarrival time distribution and Erlang distribution we use the Kolmogrov Smirnov test. Let H_0 : the data follows Erlang distribution. Here the calculated value of the Kolmogrov Smirnov test statistic is 0.04077 and critical value at 5 percent level of significance is 0.077 showing that the Erlang assumption for interarrival times is valid.

The following figure gives the P-P plot of the Erlang distribution.



The P-P plot is very close to the straight line joining (0, 0) and (1, 1). This shows that the Erlang distribution is a good fit to the data. Thus based on Kolmogrov Smirnov statistic and P-P plot we conclude that the null hypothesis of Erlang distribution as a good fit to the data is acceptable. From the graph we can conclude that it is a good fit.

References

[1] R. E. Barlow and F. Proschan, Mathematical Theory of Reliability, John Wiley, New York, 1965.

[2] D. R. Cox, Renewal Theory, John Wiley, New York, 1972.

[3] K. K. Jose and B. Abraham, A count model based on Mittag-Leffler interarrival times, Statistica 71 (2011), 501-514.

[4] Blake McShane, M. Adrian, E. T. Bradlow and P. S. Fader, Count models based on Weibull interarrival times, J. Bus. Econom. Statist. 26(3) (2008), 369-378.

[5] R. Winkelmann, Duration dependence and dispersion in count-data models, J. Bus. Econom. Statist. 13 (1995), 467-474.

[6] V. Seetha Lekshmi and Catherine Thomas, Some generalizations of Poisson processes, J. Stat. Theory Appl. 11 (2012), 225-235.