

Extension of Exponential Count Model and Its Application to Emissions of Beta Particles from a Nuclear Reaction

Manoj Kumar^{1*}, Sanjay Kumar Singh² and Umesh Singh²

¹Central University of Rajasthan, Kishangarh-305817

²Department of Statistics and DST-CIMS, Banaras Hindu University, Varanasi-221005

Email: manustats@gmail.com

Abstract In this paper we propose a new generalized counting process with extension of exponential inter-arrival time distribution. This new model is a generalization of the exponential distribution. The computational intractability is overcome by deriving the extension of exponential count model using polynomial expansion. The hazard function of this new model is an increasing then decreasing function of time, so that the distribution displays positive then negative duration dependence. The model is applied to a real data set on inter arrival times between emissions of beta particles from a nuclear reaction. This count model can be simulated by Markov Chain Monte-Carlo (MCMC) methods, using Metropolis-Hastings algorithm. Our model has many nice features and its comparable with other existing competitive models. It has computational simplicity and there exist of moments that can be used for under-dispersed data.

Keywords: Count Model, Moments, polynomial expansion.

1 Introduction

The Poisson model for count data is used in cases where the number of arrivals in a given time period assumes exponentially distributed inter arrival times. This is truly valid only in the case where the data of interest support the restrictive assumption of equi-dispersion (i.e., the conditional variance equals the conditional mean). But typically Poisson model is inefficient, if the data of interest present over-dispersion (the conditional variance exceeds the conditional mean) and under-dispersion (the conditional mean exceeds the conditional variance) or if the waiting time between the events are independent but not exponential (which would lead to the Poisson distribution for counts). Moreover, the waiting time may follow some other distribution with non constant hazard function. If the hazard function is a decreasing function of time, the distribution displays negative duration dependence. If the hazard function is an increasing function of time, the distribution displays positive duration dependence. In each cases, the conditional probability of occurrence depends on the time since the last occurrence rather than on the number of previous events. The probability of occurrence of an event at small increment of time, $(t + \Delta t)$ is influenced by the occurrence of at least one event up to time $(t + \Delta t)$.

For more details about Poisson process with a sequence of independently and identically exponentially distributed waiting times see Cox [3]. To derive a generalized model we replace the exponential distribution with a less restrictive non negative distributions i.e., extension of exponential (EE) distribution. Possible candidates are the Weibull (see McShane et al.,[6]), the gamma (including generalized gamma), (see Winkelmann [11]), and the log normal distributions.

In this paper we develop a new generalized count model, EE distribution in place of the exponential one. Then we have formulated count model. Before we develop the EE count model, we first set the stage by laying out the main properties that the EE count model developed here embodies. The model generalizes the most commonly used extant models such as the Poisson thus, when a simple structure is sufficient, most of the researcher will see it through the estimated model parameters. Furthermore, standard inferential procedures have been used to estimate the parameters like maximum likelihood, which can be used to compare different specifications for model suitability. The main motivation behind this model has considered specially for underdispersed data, which are likely to be seen in practice. Its feasibility to work with simulation-based method is computationally justified. On the other hand this model allows also for the incorporation of person-level heterogeneity reflecting the fact that individuals interarrival

rates may vary quite substantially across the population. Since, first our count model is based upon an assumed EE interarrival process, which generalizes the exponential one, and second, we demonstrate that the EE count model, via the shape parameter α being less than 1 is considered due to mathematical simplicity for underdispersed data. The EE interarrival time story is richer than the exponential story, because it allows for nonconstant hazard rates (duration dependence), see Figure1. Applying polynomial expansion to deriving our considered model. More details about polynomial expansion for different models have been discussed by many authors such as Bradlow, Hardie, and Fader[2], Everson and Bradlow[4], and Miller, Bradlow, and Dayaratna [7] etc.

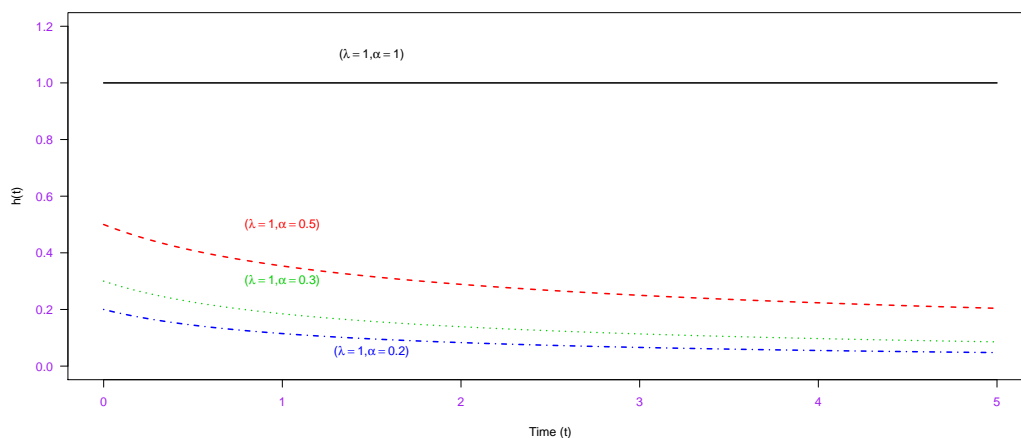


Figure 1. Hazard rate of EE Count model.

The reminder of this article is as follows. In the next section, we present a description about EE count model via polynomial expansions. Section 3 contains the properties of the EE count model. In section 4 we present the results of a simulation study/ numerical illustration pertaining to the EE count model. Applicability of the EE count model to real data is explained in section 5. Finally conclusions are provided.

2 Extension of Exponential Count Model

EE distribution is an alternative to the gamma, Weibull and the exponentiated exponential distribution. This only allows for decreasing or constant hazard when their respective probability densities are monotonically decreasing (see Figure2). In such a situation EE distribution is more applicable as compare to the gamma, Weibull and the Exponentiated exponential distribution. Since we can be used as an alternative to the gamma and Weibull, are these distributions have inability to models data having mode fixed at zero. There are many physical processes which have their mode fixed at zero by definition. The model is applicable in infant mortality situation. On the other hand, for example, consider daily rainfall and daily snowfall, for most locations in the world. However, if one tries to fit the gamma and Weibull distributions to the data, the resulting fitted probability densities may not always have their modes at zero, (see Nadaraja [8]). On the other hand, the Weibull and exponentiated exponential distributions have not allowed for an increasing hazard function when their respective probability density functions(pdf's) are monotonically decreasing. While the EE distribution is more applicable than the existing life time distributions.

The cumulative density function of an EE distribution is given as

$$F(t) = 1 - e^{1-(1+\lambda t)^\alpha} = 1 - e^{-(-1+(1+\lambda t)^\alpha)}; \quad \alpha > 0, \lambda > 0, t > 0. \quad (1)$$

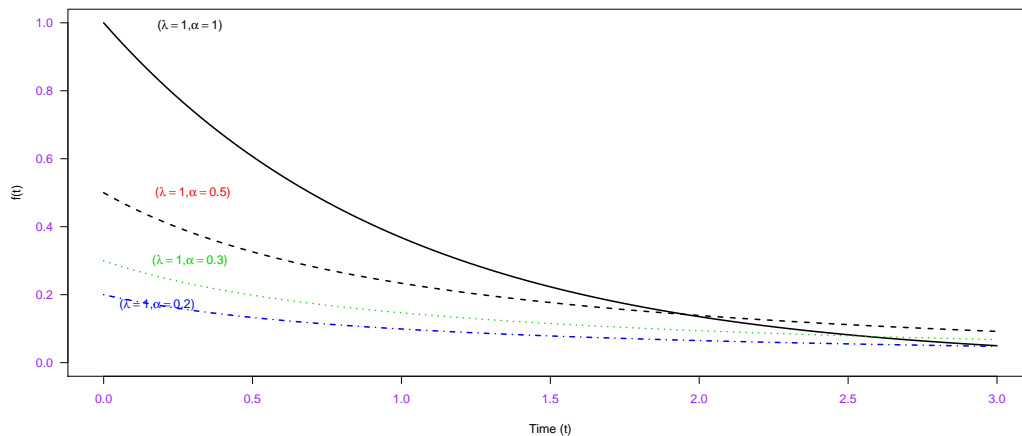


Figure 2. Probability density of EE Count model.

and its pdf given by:

$$f(t) = \alpha\lambda(1 + \lambda t)^{\alpha-1}e^{-(1+\lambda t)^\alpha}; \quad \alpha > 0, \lambda > 0, t > 0. \quad (2)$$

Its hazard function is given as

$$h(t) = \lambda\alpha(1 + \lambda t)^\alpha \quad \alpha > 0, \lambda > 0, t > 0. \quad (3)$$

which describe the general framework utilized to derive the model based upon the relationship between inter-arrival times and their count model equivalent. Let Y_n be the time from the measurement origin at which the n^{th} event occurs. Let $X(t)$ denote the number of events that have occurred up until time t . The relationship between inter-arrival times and the number of events is

$$Y_n \leq t \iff X(t) \geq n.$$

We therefore have the following relationships that allow us to derive our EE count model $C_n(t)$:

$$\begin{aligned} C_n(t) &= P(X(t) = n) \\ &= P(X(t) \geq n) - P(X(t) \geq n + 1) \\ &= P(Y_n \leq t) - P(Y_{n+1} \leq t). \end{aligned} \quad (4)$$

If we let $F_n(t)$ be the cumulative density function (cdf) of Y_n , then

$$C_n(t) = P(X(t) = n) = F_n(t) - F_{n+1}(t).$$

In the case where the measurement time origin (and thus the counting process) coincides with the occurrence of an event, then $F_n(t)$ is simply the n -fold convolution of the common inter-arrival time distribution which may or may not have a closed-form solution. Now we assume that the inter arrival times are independently and identically distributed, which follows the EE distribution. Then we derive our EE count model based upon a polynomial expansion of $F(t)$. Moreover, we use recursive relationship of the form $\int_0^t F(t-s)f(s)ds$ with a Taylor series expansion of the EE distribution.

In particular, the Taylor series approximations obtained by expanding the exponential pieces $(e^{(1+\lambda t)^\alpha} - 1)$ respectively, for both the cdf and pdf of the EE count model are

$$F(t) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1}((1 + \lambda t)^\alpha - 1)^j}{j!} \quad (5)$$

and

$$f(t) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \alpha \lambda (1 + \lambda t)^\alpha j ((1 + \lambda t)^\alpha - 1)^{j-1}}{j!} \tag{6}$$

when $0 < \alpha < 1$ then

$$f(t) = (1 + \alpha \lambda t) e^{-\alpha \lambda t}. \tag{7}$$

Utilizing that $C_n(t) = F_n(t) - F_{n+1}(t)$ we obtain the following recursive relationship that we utilize in deriving the EE count model:

$$\begin{aligned} C_n(t) &= \int_0^t F_{n-1}(t-s) f(s) ds - \int_0^t F_n(t-s) f(s) ds \\ &= \int_0^t C_{n-1}(t-s) f(s) ds. \end{aligned} \tag{8}$$

Before proceeding to develop the general solution to the problem, we note that $F_0(t)$ is 1 for all t and $F_1(t) = F(t)$. Therefore, we have $C_0(t) = F_0(t) - F_1(t) = e^{-(1+(1+\lambda t)^\alpha)}$

$$= \sum_{j=0}^{\infty} \frac{(-1)^j ((1 + \lambda t)^\alpha - 1)^j}{j!}. \tag{9}$$

Using the recursive formula, we can compute $C_1(t)$:

$$\begin{aligned} C_1(t) &= \int_0^t C_0(t-s) f(s) ds \\ &= \int_0^t \sum_{j=0}^{\infty} \frac{(-1)^j ((1 + \lambda(t-s))^\alpha - 1)^j}{j!} \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \alpha \lambda (1 + \lambda t)^{\alpha-1} k ((1 + \lambda s)^\alpha - 1)^{k-1}}{k!} ds \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} \alpha \lambda k}{j! k!} \int_0^t ((1 + \lambda(t-s))^\alpha - 1)^j ((1 + \lambda s)^\alpha - 1)^{k-1} ds, \end{aligned} \tag{10}$$

and for small displacement i.e., $0 < \alpha < 1$

$$\begin{aligned} C_1(t) &= \int_0^t C_0(t-s) f(s) ds \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} \alpha \lambda k}{j! k!} \int_0^t (\alpha \lambda (t-s))^j (\alpha \lambda s)^{k-1} ds \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} \alpha^{j+k} \lambda^{j+k} k}{j! k!} \int_0^t (t-s)^j s^{k-1} ds \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} \alpha^{j+k} \lambda^{j+k} k}{j! k!} (t)^{j+k-1} \int_0^t \left(1 - \frac{s}{t}\right)^j \left(\frac{s}{t}\right)^{k-1} ds, \end{aligned} \tag{11}$$

now let $\frac{s}{t} = z$ then $ds = t dz$. When $s = 0, z = 0$ and when $s = t, z = 1$, the above equation can be written as

$$\begin{aligned} &= \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} (\alpha \lambda t)^{j+k} k}{j! k!} \int_0^1 (1-z)^j (z)^{k-1} ds \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^{j+k+1} (\alpha \lambda t)^{j+k}}{(j+k)!}. \end{aligned} \tag{12}$$

In order to get a closed form we use a change of variable $m = j$ and $l = m + k$, then we obtain

$$\begin{aligned} &= \sum_{l=1}^{\infty} \frac{(-1)^{l+1} (\alpha \lambda t)^l}{\Gamma(l+1)} \sum_{m=0}^{l-1} \frac{\Gamma(m+1) \Gamma(l-m+1)}{\Gamma(m+1) \Gamma(l-m+1)} \\ &= \sum_{l=1}^{\infty} \frac{(-1)^{l+1} (\alpha \lambda t)^l}{\Gamma(l+1)}. \end{aligned} \tag{13}$$

This suggests a general form for $C_n(t)$, namely $\sum_{l=n}^{\infty} \frac{(-1)^{l+1} (\alpha \lambda t)^l}{\Gamma(l+1)}$, which is confirmed by

$$\begin{aligned} C_{n+1}(t) &= \int_0^t C_n(t-s) f(s) ds \\ &= \int_0^t \sum_{l=n}^{\infty} \frac{(-1)^{l+1} (\alpha \lambda (t-s))^l}{\Gamma(l+1)} (1 + \lambda \alpha t) e^{-\alpha \lambda t} ds \\ &= (1 + \lambda \alpha t) e^{-\alpha \lambda t} \sum_{l=n}^{\infty} \frac{(-1)^{l+1} (\alpha \lambda)^l}{\Gamma(l+1)} \int_0^t (t-s)^{l+1-1} s^{1-1} ds \\ &= (1 + \lambda \alpha t) e^{-\alpha \lambda t} \sum_{l=n}^{\infty} \frac{(-1)^{l+1} (\alpha \lambda t)^l t}{\Gamma(l+2)} \end{aligned} \tag{14}$$

Therefore, we have the main result of this article, the EE count model.

Theorem. *If the interval times are independently and identically distributed as EE distribution, then the count model probabilities are given by*

$$C_n(t) = P(X(t) = n) = \sum_{l=n}^{\infty} \frac{(-1)^{l+1} (\alpha \lambda t)^l}{\Gamma(l+1)}, n = 0, 1, 2, \dots \tag{15}$$

3 Properties

- I. The EE count model generalizes the most commonly used model such as exponential distribution. When $\alpha = 1$,

$$C_n(t) = P(X(t) = n) = \sum_{l=n}^{\infty} \frac{(-1)^{l+1} (\lambda t)^l}{\Gamma(l+1)}.$$

- II. The expectation of EE count model is

$$\begin{aligned} E(N) &= \sum_{n=1}^{\infty} n C_n(t) \\ &= \sum_{n=1}^{\infty} \sum_{l=n}^{\infty} n \frac{(-1)^{l+1} (\alpha \lambda t)^l}{\Gamma(l+1)} \end{aligned}$$

and variance of the EE count model given as

$$V(n) = E(N^2) - (E(N))^2 = \sum_{n=1}^{\infty} \sum_{l=n}^{\infty} n^2 \frac{(-1)^{l+1} (\alpha \lambda t)^l}{\Gamma(l+1)} - \left(\sum_{n=1}^{\infty} \sum_{l=n}^{\infty} n \frac{(-1)^{l+1} (\alpha \lambda t)^l}{\Gamma(l+1)} \right)^2.$$

III. The moment generating function (MGF) is given as:

$$M_I(u) = E(e^{iu}) = \sum_{i=0}^{\infty} \sum_{l=i}^{\infty} \frac{e^{iu} (-1)^{l+1} (\lambda \alpha t)^l}{\Gamma(l+1)} \tag{16}$$

and the corresponding moment is

$$\begin{aligned} \frac{d^n}{du^n} M_I(u) &= \frac{d^n}{du^n} \sum_{i=0}^{\infty} \sum_{l=i}^{\infty} \frac{e^{iu} (-1)^{l+1} (\lambda \alpha t)^l}{\Gamma(l+1)} \\ &= \sum_{i=n}^{\infty} \sum_{l=i}^{\infty} \frac{i^n e^{iu} (-1)^{l+1} (\lambda \alpha t)^l}{\Gamma(l+1)}. \end{aligned} \tag{17}$$

Hence, on the basis of our polynomial expansion, we obtain closed-form expression for the density as well as its moments.

IV. The hazard function is

$$h(t) = \frac{f(t)}{1 - F(t)}. \tag{18}$$

4 Numerical Illustration

In this section we study some numerical findings based on extensive simulated (because the result is unavailable in closed form) data. Table 1 represents the probabilities of EE count model for different values of the parameter α and $\lambda = 1$ at $t = 1, 2$ and 3 . By using Metropolis-Hastings algorithm we simulate the EE count model and verify that for $0 \leq \alpha < 1$ the conditional expectation exceeds the conditional variance which indicates under dispersion (value of conditional mean is greater than conditional variance). Thus Figure 3 indicates that the count model can be used to represent the under dispersed data, which is also shown in Table 2.

Table 1. The mean and variance of the EE Count model probabilities for different values of the parameter α with $\lambda = 1$ at $t = 1, 2, 3$.

α	t=1		t=2		t=3	
	mean	Var	mean	Var	mean	Var
0.1	0.4967628	0.07805997	0.50747201	0.08939391	0.5063129	0.0901227
0.2	0.5055814	0.0776719	0.50789464	0.09131011	0.51348644	0.09006574
0.3	0.50303158	0.07717554	0.51348644	0.09006574	0.50719775	0.09488594
0.4	0.49971719	0.07910678	0.50628037	0.09368348	0.49967383	0.09536808
0.5	0.49778192	0.07912192	0.50690388	0.09493151	0.51369941	0.09195346
0.6	0.50097433	0.07896364	0.49967383	0.09536808	0.5120975	0.09628886
0.7	0.50431318	0.07847249	0.50509274	0.09233733	0.55258469	0.09955615
0.8	0.50126	0.0801147	0.51239046	0.09429889	0.500207	0.1014056

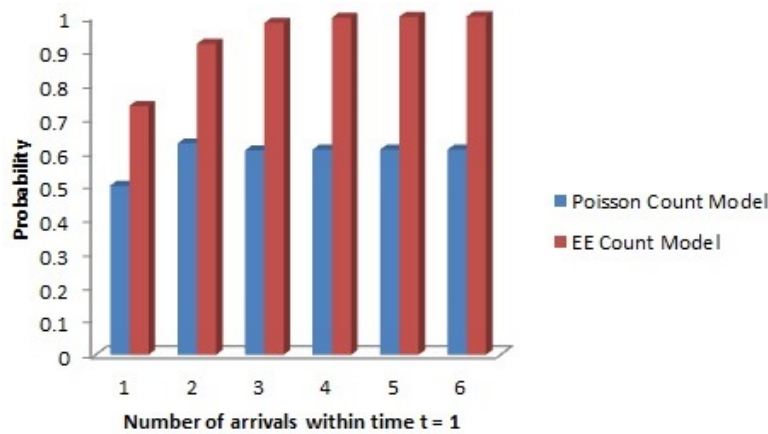


Figure 3. Probability histogram of EE Count model ($\alpha = 0.5, \lambda = 1$) and Poisson model ($\lambda = 1$)

Table 2. The EE count model probability for different values of the parameter α with $\lambda = 1$ at $t = 1, 2, 3$.

α	t=1			t=2			t=3		
	C1(t)	C2(t)	C3(t)	C1(t)	C2(t)	C3(t)	C1(t)	C2(t)	C3(t)
0.1	0.9048374	0.8187308	0.7408182	0.9048374	0.8187308	0.7408182	0.9048374	0.8187308	0.7408182
0.2	0.8187308	0.67032	0.5488116	0.8187308	0.67032	0.7408182	0.8187308	0.67032	0.5488116
0.3	0.7408182	0.5488116	0.4065697	0.7408182	0.5488116	0.4065697	0.7408182	0.5488116	0.4065697
0.4	0.67032	0.449329	0.3011942	0.67032	0.449329	0.3011942	0.67032	0.449329	0.3011942
0.5	0.6065307	0.3678794	0.2231302	0.6065307	0.3678794	0.2231302	0.6065307	0.3678794	0.2231302
0.6	0.5488116	0.3011942	0.1652989	0.5488116	0.3011942	0.1652989	0.9048374	0.3011942	0.1652989
0.7	0.4965853	0.246597	0.1224564	0.4965853	0.246597	0.1224564	0.4965853	0.246597	0.1224564
0.8	0.44932896	0.20189652	0.09071795	0.44932896	0.20189652	0.09071795	0.4965853	0.20189652	0.09071795

5 Application to a Real Data Set

For illustrative on purpose, we have considered a data set of an engineer, observing a nuclear reaction, measuring time interval between emissions of beta particles (Shuda and Purohit[9]). The following data are inter arrival times:

0.894, 0.235, 0.071, 0.459, 0.1, 0.991, 0.424, 0.159, 0.431, 0.919, 0.061, 0.216, 0.082, 0.092, 0.9, 0.186, 0.579, 1.653, 0.83, 0.093, 0.311, 0.429, 2.01, 1.718, 0.041, 0.817, 0.612, 0.158, 0.099, 0.712, 2.267, 0.143, 0.527, 0.162, 0.994, 0.091, 0.055, 1.033, 0.076, 0.149, 0.139, 0.752, 2.863, 0.107, 0.866, 0.083, 0.188, 0.365, 0.278, 0.054.

From Figure 4 it is clearly identified that inter arrival times are positively skewed. Here we also observe that the conditional mean is greater than the conditional variance (mean = 0.54948 and variance = 0.38961). Thus this dataset is under dispersed and hence we apply the EE Count model.

To test whether there is a significant difference between an observed inter arrival time distribution and the EE model distribution, the fitting of EE model was checked using CDF(Cumulative distribution function)-plot and PP(Probability Probability)-plot given in Figure (5-6) and Kolmogorov-Smirnov(KS) test. Value of the test statistics $0.1041684 < 0.21496$ (KS(Tabulated)), which shows that EE model provides a satisfactory fit to the considered data set. Hence, the EE model assumption for interarrival times is valid. To estimate the number of emission of beta particles, we use the EE count model. Figure 7 supports above discussion clearly.

5.1 Model Suitability for Real Data

Use the classical inferential procedure from Nadaraja [8] and Singh et al[10], which were used as estimates for the maximum likelihood approach for model parameter. For comparing model fitting, Akaike [1], namely $AIC = -2 * \hat{l}(\cdot) + 2 * k$ is adopted, where k is the number of parameters in the model under

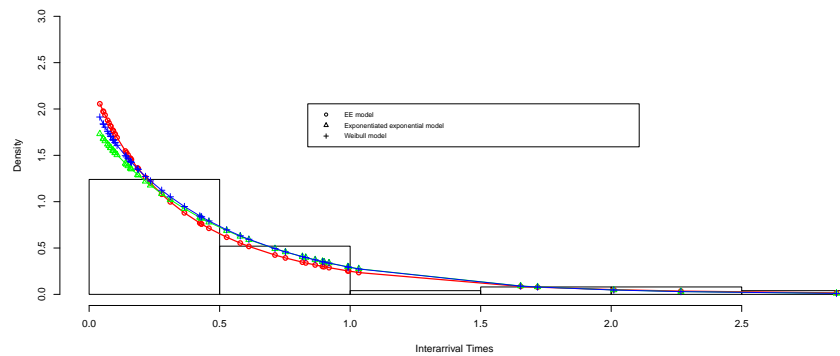


Figure 4. The histogram of real data with superimposed EE distribution.

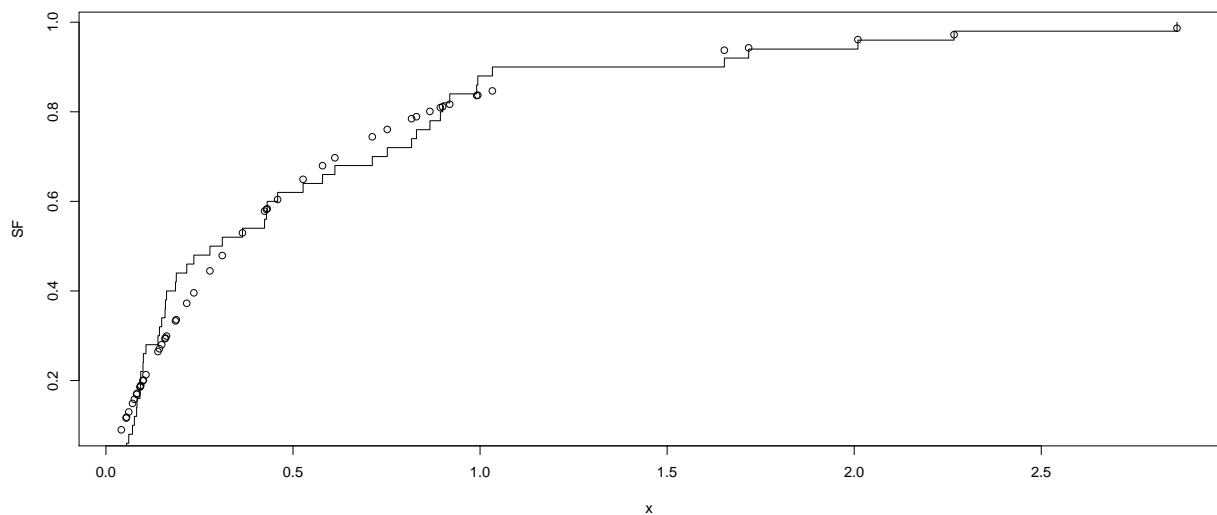


Figure 5. The CDF Plot for the real data set.

consideration. The AIC specifies that the model that best fits the data is the one with the smallest AIC value. Table 3 shows parameter estimators for distributions EE, exponentiated exponential and Weibull distributions using maximum likelihood (ML) approach and the corresponding Akaike information criterion (AIC). For these data, AIC shows a better fit for the EE model. Figure 4 reveals model fitting for the three models, and Figure 5 compares the distribution functions for the models with the empirical distribution function.

6 Conclusion

In this article we have introduced a new count model based upon EE inter arrival time process. More importantly, the model provides a sizeable improvement over the traditional Poisson process. One important advantage of the new model is used in underdispersion data set for considered parameters. The new model has closed form and the computation is possible using R. This new model can be applied to real data sets where the assumption of underdispersion is considered.

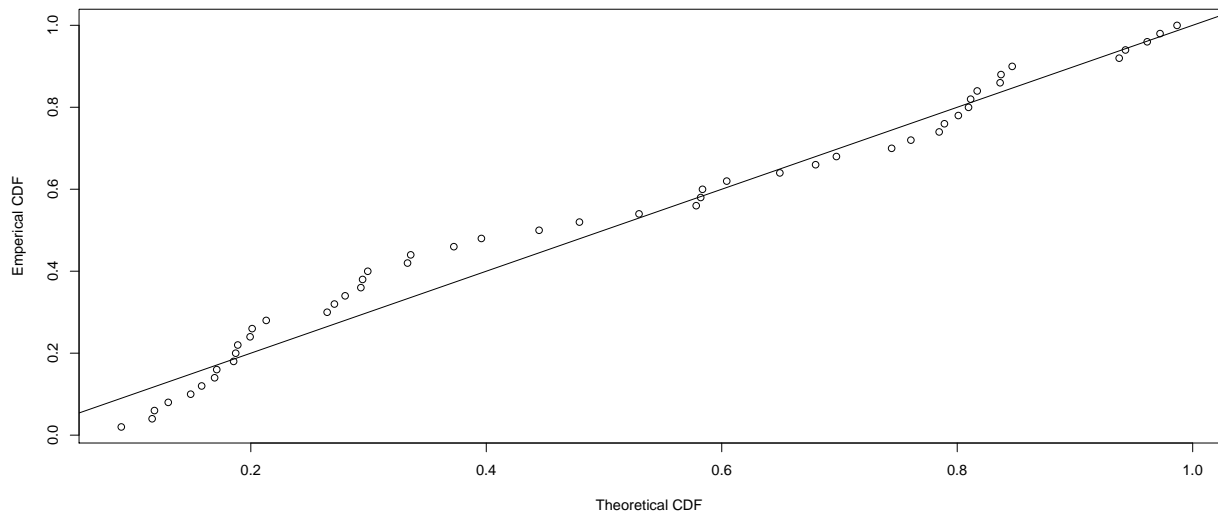


Figure 6. The Probability Plot for the real data set.

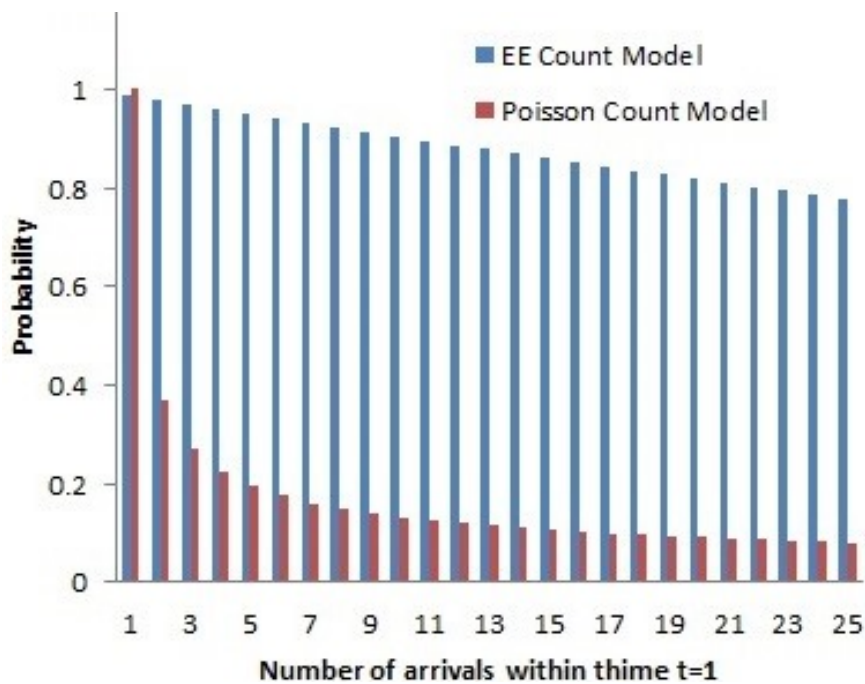


Figure 7. Probability of the predicted number of emission of particles according to EE count model and Poisson model.

Acknowledgment

The authors wish to thanks the referee and associate editor for going through the paper carefully and providing valuable comments which led to an improvement in the presentation of the paper.

Table 3. Parameter estimates for EE, exponentiated eponential and Weibull distributions for a nuclear reaction data set, measures time (inter arrivals) interval between emissions of beta particles.

Parameter estimates	EE model	Exponentiated Eponential	Weibull Distribution
α	0.7141213	0.9753652	0.940298
λ	3.2800951	1.7901338	1.877093
<i>AIC</i>	34.78671	36.10387	35.78125

References

1. H Akaike. A new look at statistical model identification. *IEEE Transaction on Automatic Control*, AC-19(6):716-723, 1974.
2. E T Bradlow, B G S Hardie, and P S Fader. Bayesian inference for the negative binomial distribution via polynomial expansions. *Journal of Computational and Graphical Statistics*, 11:189-201, 2002.
3. D R Cox. Regression model and life-tables. *Journal of the Royal Statistical Society, Ser. B*, 34:187-220, 1972.
4. P J Everson and E T Bradlow. Bayesian inference for the beta-binomial distribution via polynomial expansions. *Journal of Computational and Graphical Statistics*, 11:202-207, 2002.
5. K K Jose and B Abraham. A count model based on mittag-leffler. *Statistica*, 04:501-514, 2011.
6. Blake Mcshane, M Adrian, E T Bradlow, and P S Fader. Count models based on weibull interarrival time. *Journal of Business and Economic Statistics*, 26(3):369-378, 2008.
7. S J Miler, E T Bradlow, and K Dayaratna. Close-form bayesian inferences for the logit model via polynomial expansions. *Quantitative Marketing and Economics*, 4:173-206, 2006.
8. S Nadarajaha and F Haghghi. An extension of the exponential distribution. *Statistics*, 45(6):543-558, 2011.
9. S G Purohit, S D Gore, and S R Deshmukh. *Statistics Using R*. Narosa Publishing House Pvt. Ltd, Daryaganj, New Delhi-02, 2008.
10. S K Singh, U Singh, M Kumar, and P K Vishwakarma. Classical and bayesian inference for an extension of the exponential distribution under progressive type-ii censored data with binomial removals. *Journal of Statistics Applications and Probability Letters*, 1(3):75-86, 2014.
11. R Winkelmann. Duration dependence and dispersion in count data model. *Journal of Business and Economic Statistics*, 13:467-474, 1995.