Journal of Environmental Statistics
February 2013, Volume 4, Issue 5. http://www.jenvstat.org

A Counting Process with Gumbel Inter-arrival Times for Modeling Climate Data

K.K. Jose,
St.Thomas College, Pala

Bindu Abraham,
B.P.C.II College, Piravom

Abstract

Changes in temperature and rainfall will lead to frequent occurrence of floods, droughts as well as heat and cold waves. In this paper we introduce a new generalized counting process with Gumbel inter-arrival time distribution. This count model can be applied to analyze the climate variability due to the changes in temperature and rainfall. The decreasing hazard function of the new model leads to over dispersion, whereas increasing hazard function leads to underdispersion. Thus this new Gumbel count model can model both over and under dispersed count data. The new model has many nice features such as its closed form nature, computational simplicity, ability to nest Poisson, existence of moments etc. The use of the model is illustrated in two applications with respect to a data on the monthly rainfall of Kerala, the southern state of India which crossed the extreme level during 2005-2008 (under dispersed) and the data on the daily temperature in Kerala which attained the maximum level in Kerala during 2005-2008 (over dispersed).

Keywords: Inter-arrival times, Gumbel distribution, counting process, hazard function, over dispersion, under dispersion.

1. Introduction

The untimely rain in Kerala, which hit the entire region has caused crop damage and flooding. It is estimated that farmers could not harvest paddy worth about Rs.128 crores(1280 million rupees) due to unexpected flooding in the Kuttanad fields extending to 2000 hectares which is quite unusual with the normal summer rain. Experts suggest that this untimely rain is a clear evidence of climate change.

In a report in the news daily, *The Hindu*, it is noted that the “abrupt changes in precipitation and temperature are the main characteristics of climate change. Rising global temperatures will bring changes in weather patterns, rising sea levels and increased frequency and intensity of extreme weather". The summer rain usually comes to the state as a relief to the inhabitants in the month of March and April as it may ease the water shortage. Unlike in temperature trends, rainfall trends are uncertain at several locations see (Kumar et al. 2002, Krishnakumar et al. 2008 and 2009, Rao et al., 2008). It is also a matter of concern that the untimely rain may cause an outbreak of infectious diseases. The state had witnessed infectious diseases in the last few years during rainy season, particularly with Chickungunya and Dengue fever. There had been reports of jaundice attack at many places in the state.(Miguel et al. 2008) studied the influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003-2006. (Pascual et al. 2002) analyzed the quantitative evidence on cholera and climate and (Michael et al. 2008) studied the seasonality of cholera from 1974 to 2005. As this is a clear indication of climate change, both the government and the people must be vigilant to cop up with untimely eventualities which may be a regular phenomenon in the coming years. To study the climate change it is necessary to know the pattern of changes in the level of temperature and rainfall at different time points. Here we take the series of time points which crosses the extreme temperature as well as rainfall . Then by using this time series we developed count models to predict the future climate change.

There is a straight forward connection between the count model and a timing process. But out of many of the count models that have been developed over the years, (see Wimmer and Altman, 1999), very few share this relationship. Poisson count model is truly valid only in the case where the data of interest support the restrictive assumptions of equi-dispersion (i.e., the mean and variance of the data are equal). Statisticians recognized this limitation for many years and now use models that admits over dispersion (i.e., the data sets marked by a fatter, longer right tail than the Poisson can accommodate). A heterogeneous Gamma-Poisson(Negative binomial) model is the first count model invoked for this common situation. Next problem is how to accommodate the data sets with under dispersion. Statisticians have acknowledged and addressed this issue in different ways (see King, 1989; Cameron and Johansson, 1997; Trivedi and Cameron, 1996). But with the possible exception of a count model featuring gamma distributed inter arrival times (see Winklemann, 1995a,b), none of these under dispersed count models offers the conceptual elegance and usefulness of the Poisson exponential connection.

An important entity for the analysis of durations, used to capture duration dependence is the hazard rate h(t) which gives the instantaneous exit probability conditional on survival. Hence we consider, $h(t) = \frac{f(t)}{1-F(t)} = \frac{d}{dt} logF(t)$ where f(t) and F(t) are the density function and cumulative distribution function of the inter arrival time respectively. 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 to as positive duration dependence. No duration dependence corresponds to the case of a con stant hazard. The hazard is a constant if and only if the distribution of waiting times is exponential.

We assume that the waiting times between the events are independent but not exponential (which would lead to the Poisson distribution for counts). Instead they follow some other distribution with a nonconstant 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 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. Events are dependent in the sense that the occurrence of at least one event (in contrast to none) up to time t influences the probability of a further occurrence in $t + \Delta t$.

There is a link between duration dependence and dispersion. If we denote the mean of the interarrival distribution by μ, the variance by $\sigma,^2$ then we say that the distribution has negative duration dependence if $\frac{d}{dt}h(t) < 0$ and positive duration dependence if $\frac{d}{dt}h(t) > 0$. If the hazard function is monotonic, then we have if $\frac{d}{dt}h(t) > 0$, then $\frac{\sigma}{\mu} < 1$; if $\frac{d}{dt}h(t) = 0$, then $\frac{\sigma}{\mu} = 1$; if $\frac{d}{dt}h(t) < 0$, then $\frac{\sigma}{\mu} > 1$, (see Bradlow et. al, 2002). These three cases correspond to count data characterized by under dispersion, equi dispersion and over dispersion respectively. It is shown that negative duration dependence (asymptotically) causes over-dispersion and positive duration dependence causes under-dispersion. The Poisson process can be taken as a sequence of independently and identically exponentially distributed waiting times (see Cox, 1972) . To derive a generalized model we replace the exponential distribution with a less restrictive non negative distribution. Possible candidates known from the duration literature are the Weibull (see McShane et.al, 2008), the gamma(including generalized gamma) (see Winkelmannn, 1995a,b, 2008), and the log normal distributions (see Bradlow et. al, 2002; Everson and Bradlow, 2002; Miller et al., 2006). Both Weibull and gamma nest the exponential distribution and both allow for a monotone hazard rate function that is duration dependent. Recently some authors developed (see Jose and Bindu, 2011) a count model with Mittag-Leffler inter-arrival time distribution.

In this paper we replace the exponential distribution by Type II Gumbel distribution where Weibull and Frechet are the special cases. A corresponding count model is formulated which nest Poisson process as a special case. The Gumbel inter arrival time model is richer than the exponential, because it allows for nonconstant hazard rates (duration dependence). We can derive the model by using a polynomial expansion.The remainder of this article is as follows. In section 2, a description about Type II Gumbel distribution is given. Section 3 contains the derivation of the Type II Gumbel count model, focusing on the polynomial expansion that leads to the closed form benefits and its prperties. In section 4 the Gumbel count model is used in two real applications in climate data.

2. Gumbel Distribution

The Gumbel distribution was first developed by a German mathematician Emil Gumbel (1891-1966). Gumbel's focus was primarily on applications of extreme value theory to engineering problems, in particular modeling of meteorological phenomena such as annual flood flows. The Gumbel distribution, also known as the Extreme Value Type I distribution, is unbounded (defined on the entire real axis). The cumulative distribution function is

$$F(x; a, b) = 1 - e^{-bx^{-a}}.$$

The moments $E[X^k]$ exist for $k < a$. We can define Type II Gumbel distribution, as one whose cumulative distribution function [c.d.f] has the polynomial expansion of the form

$$F(x; a, b) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1}(bx^{-a})^j}{\Gamma(j+1)} \tag{1}$$

and p.d.f is given by

$$f(x;a,b) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1}(-a)jb^j(x^{-aj-1})}{\Gamma(j+1)}. \tag{2}$$

The type II Gumbel distribution admits a closed form expression for the hazard function as

$$h(t) = \frac{f(t)}{1-F(t)} = -abx^{-a-1}.$$

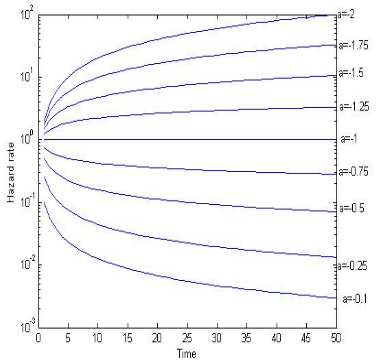

Figure 1

Hazard rate of Type II Gumbel distribution for different values of a and b=0.7.

which is monotonically increasing for $a < -1$, monotonically decreasing for $a > -1$, and constant when $a = -1$. Figure 1 supports this intuitive fact.

Gumbel has shown that the maximum value (or first order statistic) in a sample of a random variable following an exponential distribution approaches closer and closer to the Gumbel distribution with increasing sample size. In hydrology, therefore, the Gumbel distribution is used to analyze such variables as monthly and annual maximum values of daily rainfall and river discharge volumes.

3. Gumbel count model

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 upto time t. The relationship between interarrival times and the number of events is

$$Y_n \le t \quad \Leftrightarrow \quad X(t) \ge n.$$

We can derive the type II Gumbel count model by using the following relationship

$$\begin{aligned} G_n(t) = P[X(t) = n] &= P[X(t) \ge n] - P[X(t) \ge n+1] \\ &= P[Y_n \le t] - P[Y_{n+1} \le t]. \end{aligned} \tag{3}$$

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

$$G_n(t) = P[X(t) = n] = F_n(t) - F_{n+1}(t). \tag{4}$$

In the case where the measurement time origin (and thus counting process) coincides with the occurrence of an event, $F_n(t)$ is simply the n-fold convolution of the common interarrival time distribution which may or may not have a closed form solution. Based upon (3) we can derive our Type 2 Gumbel count model using the polynomial expansion of F(t).

$$G_n(t) = \int_0^t F_{n-1}(t-s)f(s)ds - \int_0^t F_n(t-s)f(s)ds = \int_0^t G_{n-1}(t-s)f(s)ds. \qquad (5)$$

Before proceeding to develop the general solution to the problem, we note that $F_0(t)$ is 1 and $F_1(t) = F(t)$ for every t. Therefore we have

$$G_0(t) = F_0(t) - F_1(t) = 1 - F(t) = e^{-bx^{-a}} = \sum_{j=0}^{\infty} \left[\frac{(-1)^j (bx^{-a})^j}{\Gamma(j+1)} \right]. \qquad (6)$$

By equation (5) we can derive

$$\begin{aligned} G_1(t) &= \int_0^t G_0(t-s)f(s)ds \\ &= \int_0^t \left(\sum_{j=0}^{\infty} \frac{(-1)^j (b(t-s)^{-a})^j}{\Gamma(j+1)} \right) \times \left(\sum_{k=1}^{\infty} \frac{(-1)^{k+1} - akb^k s^{-ak-1}}{\Gamma(k+1)} \right) ds \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} b^j b^k}{\Gamma(j+1)\Gamma(k+1)} \int_0^t -ak(t-s)^{-aj} s^{-ak-1} ds \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} b^j b^k}{\Gamma(j+1)\Gamma(k+1)} \times \frac{t^{-ak} t^{-aj} \Gamma(-aj+1)\Gamma(-ak+1)}{\Gamma(-ak+-aj+1)}. \end{aligned}$$

Then by using a change of variables m=j and l=m+k we obtain

$$\begin{aligned} G_1(t) &= \sum_{l=1}^{\infty} \left(\sum_{m=0}^{l-1} \frac{(-1)^m (-1)^{l-m+1} b^m b^{l-m}}{\Gamma(m+1)\Gamma(l-m+1)} \times \frac{t^{-a(l-m)} t^{-am} \Gamma(-am+1)\Gamma(-a(l-m)+1)}{\Gamma(-a(l-m)-am+1)} \right) \\ &= \sum_{l=1}^{\infty} \left(\sum_{m=0}^{l-1} \frac{(-1)^{l+1} (bt^{-a})^l \Gamma(-am+1)\Gamma(-al+am+1)}{\Gamma(m+1)\Gamma(l-m+1)\Gamma(-al+1)} \right) \\ &= \sum_{l=1}^{\infty} \frac{(-1)^{l+1} (bt^{-a})}{\Gamma(-al+1)} \left(\sum_{m=0}^{l-1} \frac{\Gamma(-am+1)\Gamma(-al+am+1)}{\Gamma(m+1)\Gamma(l-m+1)} \right) \\ &= \sum_{l=1}^{\infty} \frac{(-1)^{l+1} (bt^{-a}) \delta_m^l}{\Gamma(-al+1)}, \text{ where } \delta_m^l = \sum_{m=0}^{l-1} \frac{\Gamma(-am+1)\Gamma(-al+am+1)}{\Gamma(m+1)\Gamma(l-m+1)}. \end{aligned}$$

$$\text{Similarly } G_2(t) = \sum_{l=2}^{\infty} \frac{(-1)^{l+2} (bt^{-a}) \delta_l^2}{\Gamma(-al+1)}.$$

We use the method of mathematical induction to derive $G_n(t)$. Thus we have

$$G_n(t) = \sum_{l=n}^{\infty} \frac{(-1)^{l+n}(bt^{-a})\delta_l^n}{\Gamma(-al+1)}.$$

$$\begin{aligned}
G_{n+1}(t) &= \int_0^t G_n(t-s)f(s)ds \\
&= \int_0^t \left(\sum_{j=n}^{\infty} \frac{(-1)^{j+n}(bt^{-a})\delta_j^n}{\Gamma(-aj+1)} \right) \times \left(\sum_{k=1}^{\infty} \frac{(-1)^{k+1} - akb^k s^{-ak-1}}{\Gamma(k+1)} \right) ds \\
&= \sum_{j=n}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^{j+n}(-1)^{k+1} b^j b^k \delta_j^n}{\Gamma(-aj+1)\Gamma(k+1)} \times \frac{t^{-ak} t^{-aj} \Gamma(-aj+1)\Gamma(-ak+1)}{\Gamma(-ak+-aj+1)} \\
&= \sum_{l=n+1}^{\infty} \frac{(-1)^{l+n+1}(bt^{-a})^l}{\Gamma(-al+1)} \left(\sum_{m=n}^{l-1} \delta_m^n \frac{\Gamma(-al+am+1)}{\Gamma(l-m+1)} \right) \\
&= \sum_{l=n+1}^{\infty} \frac{(-1)^{l+n}(bt^{-a})^l \delta_l^{n+1}}{\Gamma(-al+1)}, \text{ where } \delta_l^{n+1} = \sum_{m=n}^{l-1} \delta_m^n \frac{\Gamma(-al+am+1)}{\Gamma(l-m+1)}.
\end{aligned}$$

Theorem 3.1 *If the interarrival times are independently and identically distributed as Type II Gumbel distribution, then the count model probabilities are given by*

$$G_n(t) = P[X(t) = n] = \sum_{j=n}^{\infty} \frac{(-1)^{j+n}(bt^{-a})^j \delta_j^n}{\Gamma(-aj+1)}, a < 0, n = 0, 1, 2... \tag{7}$$

$$where\ \delta_j^0 = \frac{\Gamma(-aj+1)}{\Gamma(j+1)}, j = 0, 1, 2, ...$$

$$and\ \delta_j^{n+1} = \sum_{m=n}^{j-1} \delta_m^n \frac{\Gamma(-aj+am+1)}{\Gamma(j-m+1)}, for\ n = 0, 1, 2... for\ j = n+1, n+2, n+3, ...$$

3.1. Characteristics of the Gumbel count model

1. The model handles both over-dispersed and under-dispersed data

   Through extensive simulations we have verified that the hazard function of the Type II Gumbel distribution is a decreasing function of time when $a > -1$, so that the distribution displays negative duration dependence which causes over-dispersion, whereas $a < -1$ the hazard function is monotonically increasing function of time, so that the distribution displays positive duration dependence which causes under-dispersion. A lack of duration dependence occurs when $a = -1$ which leads to the Poisson distribution with equal mean and variance.

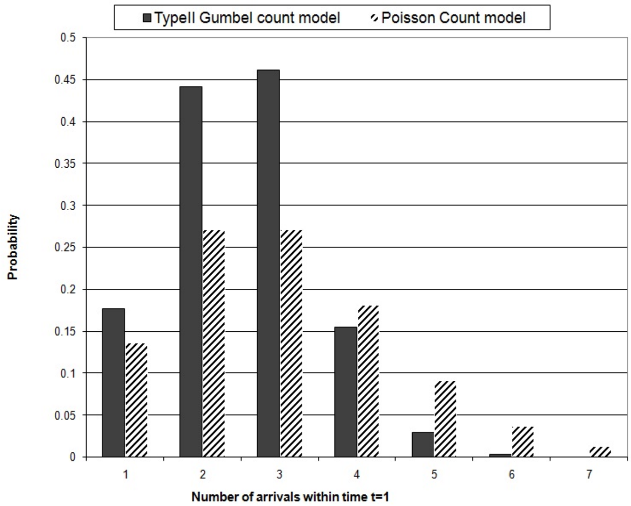

Figure 2
Probability histogram of Type II Gumbel count model ($a = -1.5$) and Poisson model (λ=2) displaying under-dispersion.

Figures 2 and 3 display probability histograms for the Type II Gumbel and Poisson models with different parameter values. Both the Gumbel and Poisson have identical means, but their dispersion is quite different. In Figure 2, probability histogram for an under-dispersed Type II Gumbel count model with parameters $a = -1.5$ and $b = 1.845$ and a Poisson with $\lambda = 2$. The variance of the Gumbel count model in this case is 0.388 which is smaller than the mean, as expected. In Figure 3, we have the probability histogram for an over-dispersed Type II Gumbel count model with parameters $a = -0.5$ and $b = 0.89$, and again follow the Poisson with $\lambda = 2$. The variance of the Gumbel count model in this case is 2.17 which is greater than the mean, as expected.

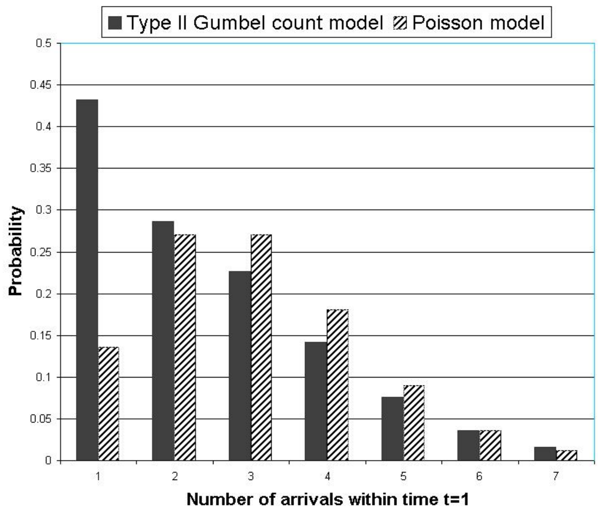

Figure 3
Probability histogram of Type II Gumbel count model ($a = -0.5$) and Poisson model (λ=2) displaying over-dispersion.

2. The model generalizes the most commonly used model such as the Poisson as special case when $t = 1$ and $a = -1$. Then

$$G_n(t) = P(X(t) = n) = \sum_{j=n}^{\infty} \frac{(-1)^{j+n} b^j \delta_j^n}{\Gamma(j+1)}.$$

This is a Poisson process with unit rate.

3. The mean and variance of the Type II Gumbel count model exist,when $a < 0$

$$\text{Mean} = E[X(t)] = \sum_{n=1}^{\infty} \sum_{j=1}^{\infty} n \frac{(-1)^{j+n} (bt^{-a})^j \delta_j^n}{\Gamma(-aj+1)}.$$

$$\text{Variance} = Var[X(t)] = \sum_{n=1}^{\infty} \sum_{j=1}^{\infty} n^2 \frac{(-1)^{j+n} (bt^{-a})^j \delta_j^n}{\Gamma(-aj+1)} - \left(\sum_{n=1}^{\infty} \sum_{j=1}^{\infty} n \frac{(-1)^{j+n} (bt^{-a})^j \delta_j^n}{\Gamma(-aj+1)} \right)^2.$$

Table 1

Values of probabilities, mean and variance of the Type II Gumbel count model for different values of the parameter a when b=0.7 at t=1

a	$G_0(t)$	$G_1(t)$	$G_2(t)$	$G_3(t)$	$G_4(t)$	$G_5(t)$	$\cdots$	μ	σ^2	relation
-2	0.6999	0.4722	0.0700	0.0035	0.0004	0.0001	$\cdots$	0.6230	0.3969	under dispersion
-1.75	0.6350	0.4531	0.0899	0.0069	0.0003	0.0000	$\cdots$	0.6548	0.4509	under dispersion
-1.5	0.5728	0.4262	0.1127	0.0132	0.0009	0.0000	$\cdots$	0.6950	0.5285	under dispersion
-1.25	0.5218	0.3913	0.1364	0.0240	0.0028	0.0002	$\cdots$	0.7487	0.6439	under dispersion
-1	0.4903	0.3516	0.1578	0.0413	0.0077	0.0011	$\cdots$	0.8282	0.8246	equi dispersion
-0.75	0.4826	0.3136	0.1733	0.0651	0.0191	0.0047	$\cdots$	0.9614	1.1273	overdispersion
-0.5	0.4966	0.2857	0.1821	0.0931	0.0409	0.0160	$\cdots$	1.2074	1.6556	overdispersion
-0.25	0.5259	0.2745	0.1884	0.1202	0.0725	0.0417	$\cdots$	1.6490	2.4248	overdispersion
-0.1	0.5483	0.2776	0.1938	0.1334	0.0906	0.0609	$\cdots$	1.9758	2.7823	overdispersion

Table 1 gives the probabilities of Type II Gumbel count model for different values of the parameter a when $b = 0.7$. By simulation of the Type II Gumbel count model we verified that for $a > -1$, the conditional variance exceeds conditional expectation which represents the over-dispersion and for $a < -1$, the conditional expectation exceeds conditional variance which represents the under-dispersion but for $a = -1$ conditional mean equals conditional variance which means the equi-dispersion. Thus Type II Gumbel count model can be used to represent overdispersed, under-dispersion as well as equidispersed real data. Table 1 supports this intuitive fact.

4. If the inter arrival times of the data set are Type II Gumbel distributed then, we have a corresponding counting model to use. The model (7) is derived from Type II Gumbel timing model, the link between the timing model and its counting model equivalent is maintained. Hence in those cases where an analysis of the inter-arrival times suggests that a more flexible timing model is needed, it can now be incorporated via its count model equivalent. Furthermore, in those cases where one only has count data, but would like to make forecasts of the next arrival time, this can be done given the timing and count model link that is now achieved.

5. We can simulate the Type II Gumbel count model. The model is computationally feasible to work with and it is estimable without requiring a formal programming language or time consuming simulation based methods.

4. Application to the real data sets

Although Kerala lies close to the equator, its proximity with the sea and the presence of the fort like Western Ghats, provides it with an equable climate which varies little from season to season. The temperature varies from 28^0 to 32^0C. Southwest Monsoon and Retreating Monsoon (Northeast Monsoon) are the main rainy seasons. The thermo-sensitive crops like black pepper, cardamom, tea, coffee and cocoa will be badly accepted as temperature range (the difference between maximum and minimum temperatures) is likely to increase and rainfall is likely to decline.

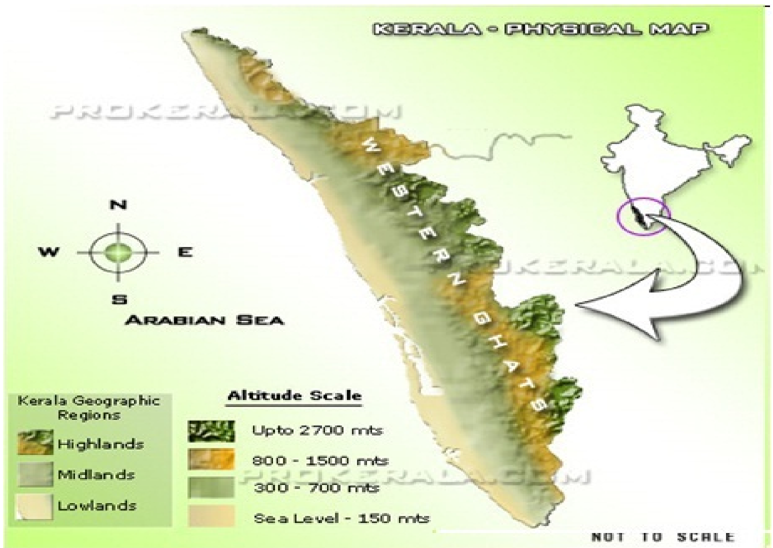

Figure 4
The location of Kerala

4.1. Monthly Rainfall Data in Kerala(Under dispersed case)

In this section we apply the model to a data on the monthly rainfall of Kerala which crosses the extreme level. The data collected concern monthly distribution of Normal and Actual rainfall of Kerala state in India from 1994-2003. The total annual rainfall in the state varies from 360cm over the extreme northern parts to about 180cm in the southern parts. The southwest monsoon (June to October) is the principal rainy season in the state. The thunderstorm rains in the pre monsoon months of April and May and that of monsoon months are locally known as '*Edavapathi*'. Rainfall during northeast monsoon season is known as '*Thulavarsham*' in local language. The southwest monsoon sets over the southern parts of the State by about 1^{st} June and extends over the entire State by 5^{th} June. June and July are the rainiest months, each accounting individually to about 23 of annual rainfall. In each of these months number of rainy days (with daily rainfall of at least 2.5 mm) varies from 27 in the north to 15 in the south.

The time interval between the months having extreme rainfall had the conditional mean 1.5217 and conditional variance 1.3662. Thus the data set is underdispersed and hence we can apply the Gumbel count model with estimates of parameters $\hat{a} = -1$ and $\hat{b} = 0.4450$.

To test whether there is significant difference between an observed interarrival time distribution and the Gumbel distribution we use Kolmogorov-Smirnov [K.S] test for H_0 : Gumbel distribution with parameter $a = -1$ and $b = 0.4450$ is a good fit for the given data. Here the calculated value of the K.S. test statistic is 0.0869 and the critical value corresponding to the significance level 0.01 is 0.2403 showing that the Gumbel assumption for inter arrival times is valid.

To estimate the number of months having extreme rainfall in a class, we use the Gumbel Count model. The Figure 5 shows that the Gumbel count model can be applied to underdispersed data.

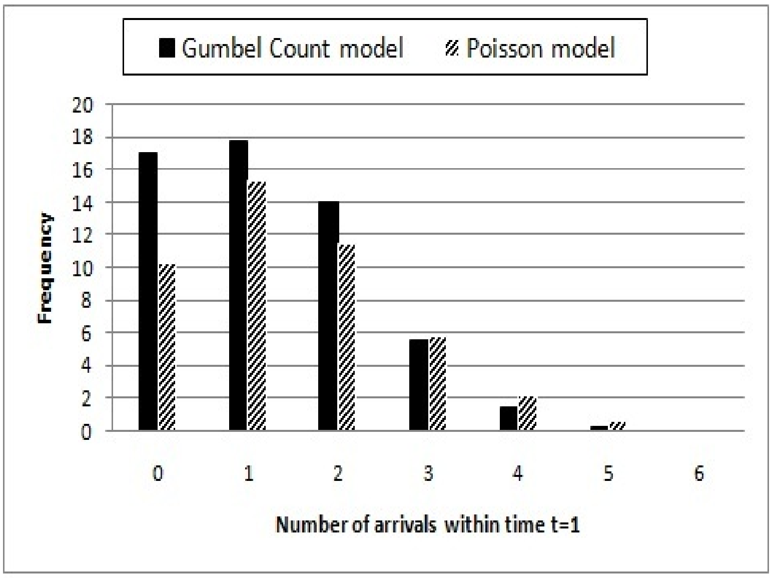

Figure 5

Probability histogram of the expected number of months having extreme rainfall in Kerala during 1994-2003 according to Gumbel count model and Poisson model

4.2. Data on Maximum Temperature in Kerala (Over dispersed data)

In this section we apply the model to a data on the daily maximum temperature which attains extreme during 2005-2008 in Kerala. The data collected is from the meteorological observations recorded at Kayamkulam station during the period 2005-2008.

Day temperatures are more or less uniform over the plains throughout the year except during monsoon months when these temperatures drop down by about 3 to 5^0C. Both day and night temperatures are lower over the plateau and at high level stations than over the plain. Day temperatures of coastal places are less than those of interior places. March is the hottest month with a mean maximum temperature of about 33^0C. Mean maximum temperature is minimum in the month of July when the State receives plenty of rainfall and the sky is heavily clouded. It is 28.5^0C for the State as a whole in July, varying from about 28^0C in the north to about 29^0C in the South. The night temperature is minimum in January when clouding is also minimum. For the State as a whole it is about 22.5^0 C in January, varying from 22^0C in the north to 22.6^0C in the South). The time interval between (interarrival) the days having maximum temperature had the conditional mean 1.7553 and the conditional variance 6.2178.

Thus the data set is overdispersed, and hence we can apply the Gumbel Count model with estimates of parameters $\hat{a} = -0.93$ and $\hat{b} = 0.5950$.

To test whether there is significant difference between an observed interarrival time distribution and the Gumbel distribution, we use Kolmogorov-Smirnov [K.S] test for H_0 : Gumbel distribution with parameters $a = -0.93$ and $b = 0.5950$ is a good fit for the given data. Here the calculated value of the K.S. test statistic is 0.1711 and the critical value corresponding to the significance level 0.01 is 0.1869 showing that the Gumbel assumption for inter arrival times is valid.

To estimate the number of days having maximum temperature in a class, we use the Gumbel Count model. The Figure 6 shows that the Gumbel count model can be applied to overdispersed data.

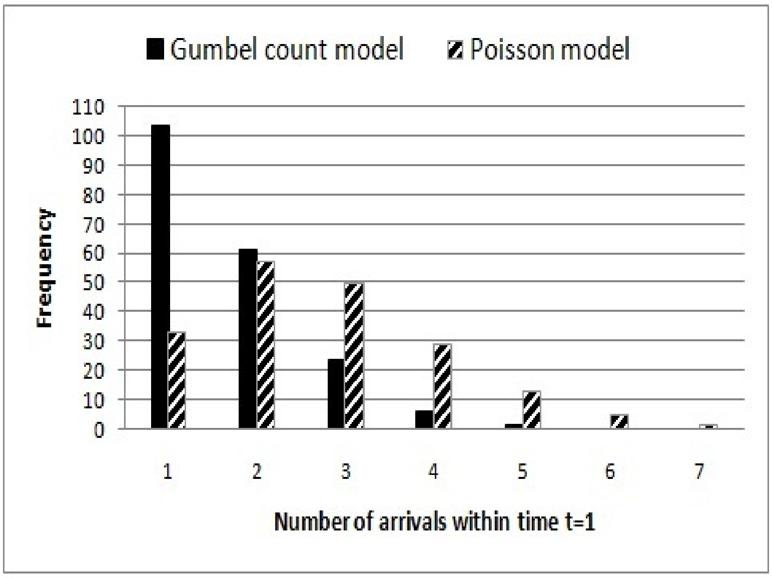

Figure 6
Probability histogram of the expected number of days having maximum temperature in Kerala during 2005-2008 according to Gumbel count model and Poisson model

5. Conclusions

In this article we have introduced a new count model based upon Gumbel inter arrival time process. More importantly, the model provided a sizeable improvement over the more traditional Poisson. The new model can be used to predict the extreme changes in temperature and rainfall which causes the climate change. One important advantage of the new model is that it removed the artificial symmetry between overdispersion and equidispersion, a violation of the constant hazard assumption underlying the Poisson model. This new model can be treated as a generalization of the Poisson distribution. The new model has closed form nature and computation is possible using Matlab. This new model can be applied to real data sets where the assumption of equidispersion is violated.

Acknowledgements

The first author thanks the International Statistical Institute, Netherlands for providing travel support under World Bank Fund Scheme to attend and present this paper in the TIES 2012 International Conference held at Hyderabad, India. The authors are also grateful to Bovas Abraham, University of Waterloo, Canada and the editors for their valuable comments and suggestions.

References

Bradlow ET, Hardie BGS, Fader PS (2002). "Bayesian Inference for the Negative Binomial Distribution via Polynomial Expansions". *Journal of Computational and Graphical Statistics*, **11**, 189-201.

Cameron AC, Johansson P (1997). "Count Data Regression Using Series Expansion: With Applications". *Journal of Applied Econometrics*, **12**, 203-223.

Cox DR (1972). "Regression Models and Life Tables". *Journal of the Royal Statistical Society*, Ser.B, **34**,187-220.

Everson PJ, Bradlow ET (2002). "Bayesian Inference for the Beta binomial distribution via polynomial expansion". *Journal of Computational and Graphical Statistics*, **11**, 202-207.

Gumbel EJ (1954). "Statistical theory of extreme values and some practical applications". *Applied Mathematics Series*, **33**.

Jose KK, Bindu A (2011). "A count model based on Mittag-Leffler inter arrival times". *Statistica*, anno LXXI, **4**, 501-514.

King G (1989). "Variance Specifications in Event Count Models: From Restrictive Assumptions to a Generalized Estimator". *American Journal of Political Science*, **33**, 762-784.

Krishnakumar KN, Rao HS, Gopakumar CS (2008). "Climate change at selected locations in the humid tropics". *Journal of Agrometeorology.*, **10**(1), 59-64.

Krishnakumar KN, Rao HS, Gopakumar CS (2009). "Rainfall trends in twentieth century over Kerala, India". *Atmospheric Environment* **43** 1940-1944.

Kumar KR, Kumar KK, Ashrit RG, Patwardhan SK, Pant GB (2002). "Climate change in India: Observation and Model Projections". *Climate change and India-Issues, Concern and Opportunities. (eds. P.R. Shukla, K.S. Sharma, and P.V. Ramana). Tata McGraw-Hill Publishing Company Limited, New Delhi,* pp 24-75.

Mcshane B, Adrian M, Bradlow ET, Fader PS (2008). "Count models based on Weibull interarrival times". *Journal of Business and Economic Statistics*, **26**(3), 369-378.

Michael E, Caryl F, Islam MS, Mohammad A (2008). "Seasonality of cholera from 1974 to 2005: a review of global patterns". *International Journal of Health Geographics*, **7**(31).

Miguel LF, Bauernfeind A, Jimenez JD, Gil CL, Omeiri NE, Guibert DH (2008). "Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003-2006: analysis of a time series". *Transactions of Royal Society of Tropical Medicine and Hygiene*, doi:10.1016/j.trstmh.2008.07.017.

Miller SJ, Bradlow ET,Dayaratna K (2006). "Closed form Bayesian inferences for the logit model via polynomial expansions". *Quantitative Marketing and Economics*, **4**, 173-206.

Pascual M, Menno JB, Andrew PD (2002). "Cholera and climate: revisiting the quantitative evidence". *Microbes and Infection*, **4**, 237-245.

Rao HS, Ram M, Gopakumar CS, Krishnakumar KN (2008). "Climate Change and cropping systems over Kerala in the humid tropics". *Journal of Agrometeorology, special issue* Part 2: 286-291.

Trivedi PK, Cameron AC (1996). *Applications of Count Data Models to Financial Data.* Handbook of Statistics, North Holland, Chap. **12**, pp. 363-391.

Wimmer G, Altmann G (1999). *Thesaurus of Univariate Discrete Probability Distributions*, Germany, Stamm Verlag.

Winkelmann R (1995a). "Duration dependence and dispersion in count data models". *Journal of Business and Economic Statistics*, **13**, 467-474.

Winkelmann R (1995b). "Recent developments in count data modeling theory and applications". *Journal of Economic Survey*, **9**, 1-24.

Winkelmann R (2008). *Econometric Analysis of Count Data*, Springer 5th Edition.

Affiliation:

K. K. Jose,
Professor, Department of Statistics, St.Thomas College, Pala, Mahatma Gandhi University, Kottayam, Kerala-686574 India.
Email: (kkjstc@gmail.com)

Bindu Abraham,
Assistant Professor, Department of Statistics, Baselios Poulose II Catholicose College, Piravom, Mahatma Gandhi University, Kottayam, Kerala-686664 India.
Email: (babpc@rediffmail.com)