Generalized linear models for analyzing count data of rainfall occurrences

Sharmin Nahar Sumi¹ · Narayan Chandra Sinha² · M. Ataharul Islam³

Received: 3 June 2020 / Accepted: 4 March 2021 / Published online: 20 March 2021

© The Author(s) 2021 [OPEN](#)

Abstract

Having the adequate knowledge about the behavior of climatic variables on the occurrences of rainfall is needed to the country's economists and agriculturists for saving the country's people from the devastating natural hazards like flash flood, drought, heavy rainfall, etc. Therefore, the study has been taken initiative to identify the influence of climatic variables for the occurrences of rainfall. The study has been developed generalized linear models (GLMs) for Poisson distribution for weekly and fortnightly count data of daily rainfall occurrences for the summer and monsoon seasons for five regional rainfall stations of Bangladesh. For these models, minimum and maximum temperatures and relative humidity are considered as explanatory variables. For five regional rainfall stations, the model selection procedures AIC and BIC indicate that the GLMs for the Poisson distribution satisfactorily explain the influence of climatic variables for the fortnightly occurrences of rainfall in the summer and monsoon seasons. The GLMs for the summer season of fortnightly occurrences of rainfall indicate that if one unit of relative humidity increases, then the probability of rainy days will be increased by 12 percent in Feni station, 6 percent in Sylhet, Khulna and Rajshahi stations, and 7 percent in Dhaka station. Besides, the GLMs for the monsoon season of fortnightly occurrences of rainfall indicate that if one unit increases of minimum temperature, then the probability of rainy days will be increased by 22 percent, 19 percent, 24 percent, 17 percent and 19 percent in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively. Further, maximum temperature indicates negative influence on the occurrences of rainfall for all the stations and seasons of the period. The study indicates that the relative humidity for summer season and minimum temperature for monsoon season play remarkable role for changing fortnightly occurrences of rainfall in all the regions of the country.

Keywords GLMs for Poisson distribution · MLE · Test of hypothesis · Deviance of the models · AIC and BIC · Weekly and fortnightly count data of rainfall occurrences

1 Introduction

The count data for response variable (Y) contain non-negative integers which are generally explained in two ways, namely simple count and categorical data. The simple count data explain how many numbers of times an event occur within a specific period of time, such as

how many days rainfall occur within a week or fortnight in during several years. Furthermore, the categorical data are categorized on the basis of different criteria of each variable within a specific period of time. The distribution of count data usually follows exponential family under the assumption that the events are independently and identically distributed [25, 26]. Nelder and Wedderburn [33],

Professor M. Ataharul Islam died in December 2020.

✉ Narayan Chandra Sinha, ncsinha2002@yahoo.com; Sharmin Nahar Sumi, shumisharmin2128@gmail.com; M. Ataharul Islam, mataharul@yahoo.com | ¹Department of Statistics, East West University, Dhaka, Bangladesh. ²Dhaka School of Economics, Dhaka, Bangladesh. ³Institute of Statistical Research and Training, University of Dhaka, Dhaka, Bangladesh.

SN Applied Sciences (2021) 3:481 | <https://doi.org/10.1007/s42452-021-04467-x>

and McCullagh and Nelder [31, 32] developed generalized linear models (GLMs) for analyzing the effect of explanatory variables on count data considering the parameter of count data distribution as link function. Therefore, the count data models are explained the rate of change of occurrence of an event for per unit time or time interval under different covariates [11, 18, 20, 47].

To explain count data using GLMs, several authors have suggested Poisson, hurdle Poisson, negative binomial (NB), zero-inflated Poisson (ZIP) models as the most suitable for analyzing the violation of safety regulations, turnover in child care arrangements, labor mobility, market entry, recruitment rates, fish catching behavior, influencing factors on TV channel viewing, etc. [9, 15, 19, 20, 34, 46, 48, 49]. Vivekanandan [45] also conducted heavy rainfall study for Tohana, Maharashtra, India, and shown that the log Pearson type-3 provides satisfactory result compared to all other distributions. Srikanthan and McMahon [42] described Markov chain, hidden Markov chain, log normal distribution, etc., for explaining annual, monthly and daily rainfall data. These studies were not considered any atmospheric variables relating to the rainfall occurrences for analyzing the patterns of rainfall data.

For exploring the causes of extreme flood and climate change scenarios in the western Ireland, Chandler and Wheeler [14] obtained satisfactory result employing GLMs method for daily rainfall occurrences. Segond et al. [37] demonstrated Poisson cluster processes using GLMs for analyzing simulated hourly and daily rainfall data of Thames region, UK. Buishand et al. [10] employed GLMs considering logistic transformation as the link function for analyzing daily and monthly rainfall occurrences of Bern, Switzerland. Considering time function Fourier series as the link function, Coe and Stern [16] developed GLMs for Markov model for analyzing the patterns of daily rainfall data of Kharja, Jordan; Lunuwila, Sri Lanka; etc. Similar analysis is also described by Stern and Coe [41] for non-stationary Markov model for rainfall data of Morogoro, Tanzania.

Sinha et al. [38–40], Islam and Chowdhury [22] explained the patterns of daily rainfall occurrences in Bangladesh considering logistic regression for Markov chain model. Thyer and Kuczera [43, 44] formulated hidden Markov chain model for identifying long-term persistence of annual rainfall and dry spell, and drought risk for Sydney, Australia. These studies were not analyzed count data of daily rainfall occurrences as the response variable. The influences of climatic variables on count data of daily rainfall occurrences are very much effective for identifying the patterns of rainfall occurrences. To the agriculturalist and hydrologist, the results of rainfall pattern helpful for formulating their year or season-wise proper crop cultivation plans and activities, side by side it also important for

formulating the country's development programs implementation plans and activities. In this context, this study has taken initiative to develop generalized linear models (GLMs) for Poisson distribution using count data of daily rainfall occurrences.

The study has been organized into five sections. Following this section, in Sect. 2 has been discussed about the methodology of the study. In this section, the study has discussed about the derivation procedure of Poisson distribution from the exponential family distribution, formulation of generalized linear model (GLM) for Poisson distribution, interpretation of parameters of GLM, estimation of parameters of GLM, test of hypothesis for testing the significance of parameters, deviance of the models and model selection criteria (AIC and BIC). In Sect. 3, the behavior of data and regarding data collection has been discussed. The results and discussions of the study are discussed in Sect. 4. Finally, conclusions of the study are drawn in Sect. 5.

2 Methodology

2.1 Poisson distribution

Count data of response variable (Y) generally follow the exponential family distribution considering a parameter θ , which is demonstrated as [7, 18, 23, 32]

$$f(y; \theta, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{\alpha_i(\phi)} + c(y_i; \phi) \right\}, \quad y \in Z^+ \quad (1)$$

where $\alpha(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are the specific functions and Z^+ is the positive integer. Here θ_i indicates the canonical parameter or link function, $b(\theta_i)$ implies cumulant function, $\alpha(\phi)$ indicates scale parameter and $c(y_i; \phi)$ indicates the normalization factor. Then the function is defined as $E(y) = b'(\theta)$ and $v(y) = \alpha(\phi)b''(\theta)$, where $\alpha(\phi)$ indicates the dispersion parameter and $b''(\theta)$ indicates the variance function.

To formulate count data distribution of daily rainfall occurrences using Eq. (1), y_i explain count data, which implies within a week or fortnight the number of days are occurred rainfall. Further, the canonical parameter θ_i express $\log \lambda$ for the distribution of count data, and the cumulant function $b(\theta_i)$ indicates λ , which is the parameter of distribution. For this distribution, the normalizing factor $c(y_i, \phi)$ is defined as $-\log y!$ because y always indicated nonnegative value as count data. These expressions explain specific distribution form for the response variable (y) considering $\alpha(\phi) = 1$.

Therefore, for the count data of daily rainfall occurrences, considering time interval, Eq. (1) is explained as below

$$f(y; \lambda) = \exp \{y \log \lambda - \lambda - \log y!\}, \quad y \in \mathbb{Z}^+ \quad (2)$$

Here $E(y) = \lambda$ and $\text{var}(y) = \lambda$, i.e., the mean and variance of y indicate equal. Equation (2) can be written as

$$f(y; \lambda) = \lambda^y e^{-\lambda} \frac{1}{y!} = \frac{\lambda^y e^{-\lambda}}{y!}, \quad y = 0, 1, 2, \dots \quad \text{and} \quad \lambda \in \mathbb{R}^+ \quad (3)$$

which is the Poisson distribution of count data of rainfall occurrences with parameter λ , under the assumption that the occurrences of rainfall are occurred randomly over time. If the parameter lies $0 < \lambda < 1$, the function is strictly decreasing, and if it provides $\lambda > 1$, then the function indicates increasing.

2.2 GLM for Poisson distribution

The mean and variance of the response variable generally explain the behavior of data because the influence of covariates on the response variable highly affects the mean of data as well as variance. The count data of daily rainfall occurrences for the specified time period provide Poisson model with parameter λ . This parameter indicates the intensity of data, i.e., instantaneous rate of rainfall occurrences. Therefore, for count data, let us consider a variable y_i which consists of n independent observations, and the i th observation is defined as (y_i, x_i) . Here x_i indicates the vector of explanatory variables for which the occurrence of rainfall within the specified period of time is responsible, i.e., $x_i' = [x_{i1}, x_{i2}, \dots, x_{ik}]$; $j = 1, 2, \dots, k$. Then the Poisson distribution of y_i conditional x_i is defined as

$$f(y_i | x_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots \quad (4)$$

with $E(y_i | x_i) = \text{Var}(y_i | x_i) = \lambda_i$ which indicates that this variance is not constant. Therefore, the GLM for Poisson distribution explains the influence of covariates on the count data of daily rainfall occurrences. Then the mean function of the distribution is expressed under logarithm scale through explanatory variables, i.e., $\log(\lambda_i) = x_i' \beta$, $\Rightarrow \lambda_i = \exp(x_i' \beta)$ which indicates the canonical link function of GLM for Poisson distribution [33], where β indicates the parameter vector for explanatory variables. This transformation explains canonical form of y_i and canonical parameter of λ_i , and it indicates the constant elasticity of mean function in the model [47].

2.2.1 Interpretation of parameters for GLM under Poisson distribution

To formulate the GLM for count data of rainfall occurrences, the climatic variables like temperature, humidity,

etc., are employed as the covariates. For GLM, the influence of covariates on y_i is explained by link function, and then, it is stated as exponential form which indicates the mean function of count data of rainfall, i.e., $E(y_i | x_i) = \lambda_i = \exp(x_i' \beta)$. The influence of stated covariates on the occurrence of rainfall indicates the marginal effect of respective covariates. This marginal effect implies that for changing one unit individual climatic variable how much probability of rainfall occurrence will be changed. Furthermore, marginal effect indicates the instantaneous rate of change of probability of rainfall occurrences for one unit change of the specified climatic variable considering the effect of other variables is unchanged. Therefore, for count data model, the marginal effects of explanatory variables influence the average of the predicted counts of the model. Then the average marginal effects for binary explanatory variables indicate the partial effect on count variable [21]. Therefore, the marginal effect of respective covariate on count data under GLM is stated as

$$\frac{\partial E(y_i | x_i)}{\partial x_{ij}} = \frac{\partial \lambda_i}{\partial x_{ij}} = \beta_j \exp(x_i' \beta) \quad (5)$$

Then the expected or average marginal effect is defined as

$$E\left[\frac{\partial E(y_i | x_i)}{\partial x_{ij}}\right] = \frac{1}{n} \sum_{i=1}^n \beta_j \exp(x_i' \beta) \quad (6)$$

If the GLM includes intercept term, then the average marginal effect is expressed as

$$E\left[\frac{\partial E(y_i | x_i)}{\partial x_{ij}}\right] = \beta_j \bar{y} \quad (7)$$

because the first-order condition implies $\sum_{i=1}^n \exp(x_i' \beta) = \sum_{i=1}^n y_i$.

The β_j indicates the relative change in $E(y_i | x_i)$ associated with one unit change in x_{ij} ; then, it explains as

$$\frac{\partial E(y_i | x_i) / E(y_i | x_i)}{\partial x_{ij}} = \beta_j. \quad (8)$$

This parameter also indicates the semielasticity, i.e., it explains the percentage change of rainfall occurrences in a week or fortnight for one unit change of individual climatic variable. If x_i explain in logarithm form, then β_j interprets the elasticity of rainfall occurrences count for the individual climatic variables. Further, to assess the effect of discrete covariates on count data, i.e., the influence of one unit change in x_{ij} on the expected value of y_i , then there compares the expected value of y_i for x_{ij}

and $x_{ij} + 1$, respectively. In that situation, let us consider $\tilde{x}_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ij} + 1, \dots, x_{ik})$; then the relative change is defined as [47].

$$\frac{E(y_i|\tilde{x}_i, \beta) - E(y_i|x_i, \beta)}{E(y_i|x_i, \beta)} = \frac{\exp(x'_i\beta + \beta_j) - \exp(x'_i\beta)}{\exp(x'_i\beta)} = \exp(\beta_j) - 1 \quad (9)$$

This indicates the relative change in $E(y_i|x_i)$ due to one unit change in x_{ij} . Similarly, the relative influence of a dummy variable (takes the value 0 or 1) on the expected count data is defined as $\exp(\beta_j) - 1$.

2.2.2 Estimation of parameters for GLM

To analyze the count data of rainfall occurrences, the GLMs for Poisson distribution explain appropriate basis because it capable of identifying the instantaneous rainfall occurrences rate. For this purpose, need to identify the influence of climatic variables on the occurrence of daily rainfall. To identify the variables influence, the study needs to estimate the parameters of GLMs. For this purpose, the study has been considered maximum likelihood estimation (MLE) method [11, 12, 47]. The principle of MLE indicates that the influences of climatic variables on the rainfall occurrence under GLMs for Poisson distribution are estimated to maximize the probability of rainfall occurrences.

2.2.2.1 Maximum likelihood estimation for GLM Let us consider n pair of independent observations (y_i, x_{ij}) ; here $y_i = 0, 1, 2, \dots$ indicate the count data of rainfall occurrences and $x_{ij}, i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k$ indicate the climatic variables corresponding to the y_i . Then the GLM for Poisson distribution is defined as

$$f(y_i|x_{ij}, \beta) = \frac{[\exp(x'_{ij}\beta)]^{y_i} \exp[-\exp(x'_{ij}\beta)]}{y_i!} \quad (10)$$

For this model, the likelihood function of conditional distribution is stated as

$$L[f(y_i|x_{ij}; \beta)] = \prod_{i=1}^n f(y_i|x_{ij}; \beta)$$

Therefore, the log-likelihood function for the model is expressed as

$$\begin{aligned} \Rightarrow l(\beta; y, x) &= \sum_{i=1}^n \log f(y_i|x_{ij}; \beta) \\ &= \sum_{i=1}^n [y_i x'_{ij}\beta - \exp(x'_{ij}\beta) - \log(y_i!)] \end{aligned} \quad (11)$$

The logarithm transformation is the monotonic function. To estimate the parameters of the model, this logic indicates the maximization of log-likelihood function, $l(\beta; y, x)$. Under optimization condition to estimate the parameter $\hat{\beta}$ taking first derivatives on $l(\beta; y, x)$ and setting equal to zero, then it is written as

$$\frac{\partial l(\beta; y, x)}{\partial \beta} = \sum_{i=1}^n [y_i - \exp(x'_{ij}\beta)] x_{ij} = 0 \quad (12)$$

Thus the Hessian matrix of the model is defined as

$$\begin{aligned} H_n(\beta; y, x) &= \frac{\partial^2 l(\beta; y, x)}{\partial \beta \partial \beta'} \\ &= \frac{\partial \left[\sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n \exp(x'_{ij}\beta) \cdot x_{ij} \right]}{\partial \beta'} \\ &= - \sum_{i=1}^n \exp(x'_{ij}\beta) \cdot x_{ij} x'_{ij} \end{aligned} \quad (13)$$

That is, H_n is negative definite, so the stated function is globally concave and there well exist Newton–Raphson iteration method to estimate the unique solution of β .

2.2.3 Test of hypotheses

To identify the significant effect of different climatic variables on the count data of rainfall occurrences under GLM for Poisson distribution, the study considers t_z statistic which is given below:

$$t_z = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}} \quad (14)$$

where $\text{var}(\hat{\beta}_j)$ is the j th diagonal element of the inverse of information matrix $l(\hat{\beta})^{-1}$, $l(\hat{\beta}) = -E[H_n(\hat{\beta}; y, x)] = -n^{-1} H_n(\hat{\beta})$; that is, diagonal element of $l(\hat{\beta})^{-1} = \text{var}(\hat{\beta}) = -[H_n(\hat{\beta})]^{-1}$ [47]. Here H_0 is rejected against $H_a : \beta_j \neq 0$ when $|t_z| > z_{\alpha/2}$, α indicates the level of significance. This is called t statistic because t_z follows normality assumption of the linear model. In general t_z does not follow t distribution for small samples. Further, for nonlinear count data models it asymptotically follows z statistic [11].

Therefore, to identify whether the effects of climatic variables on the count data of rainfall occurrences are insignificant or not, i.e., $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$, the study considers likelihood ratio test (LRT) [27]. Considering Eq. (11), this test is defined as

$$-2 \log D = 2(l(\lambda_i; y|\beta \neq 0) - l(\lambda_i; y|\beta = 0)). \quad (15)$$

Here $-2 \log D$ is asymptotically distributed as χ^2 with ϑ degrees of freedom, i.e., $\vartheta = [(\text{number of parameters of the model under alternative hypothesis}) - (\text{number of parameters of the model under null hypothesis})] = k - 0 = k$. For this study, the likelihood function under null hypothesis is explained as $l(\lambda_i; y | \beta = 0) = -n - \sum_{i=1}^n \log y_i!$ and the likelihood function under alternative hypothesis is stated as $l(\lambda_i; y | \beta \neq 0) = \sum_{i=1}^n y_i \log \lambda_i - \sum_{i=1}^n \lambda_i - \sum_{i=1}^n \log y_i!$. Then the H_0 is rejected when $\chi^2 > \chi^2_{\alpha/2, \vartheta}$, where ϑ explains the degrees of freedom.

2.2.4 Deviance for the models

Under normality assumption, deviance implies residual sum of squares for linear regression model. From this idea, deviance is used for GLM framework as the generalization of residual sum of squares because it generalizes the analysis of variance for comparing the sequences of nested GLMs. The count data for rainfall occurrences y_i are assumed independent and $y_i \sim \text{Poisson}(\lambda_i)$; then the log-likelihood function of GLMs for Poisson distribution is defined as

$$l(\beta; y, x) = \sum_{i=1}^n \left[y_i x'_{ij} \beta - \exp(x'_{ij} \beta) - \log(y_i!) \right],$$

which is stated as

$$l(\lambda_i; y) = \sum_{i=1}^n y_i \log \lambda_i - \sum_{i=1}^n \lambda_i - \sum_{i=1}^n \log y_i! \quad (16)$$

For saturated model, the estimates of $\hat{\lambda}_i$ are replaced by corresponding exact value of y_i ; then, the value of log-likelihood function (16) is expressed as

$$l(\lambda_{i(\max)}; y) = \sum_{i=1}^n y_i \log y_i - \sum_{i=1}^n y_i - \sum_{i=1}^n \log y_i! \quad (17)$$

Further, for maximum likelihood estimate $\hat{\lambda}_i$ of the model, the value of log-likelihood function (16) is computed as

$$l(\hat{\lambda}_i; y) = \sum_{i=1}^n y_i \log \hat{\lambda}_i - \sum_{i=1}^n \hat{\lambda}_i - \sum_{i=1}^n \log y_i! \quad (18)$$

Therefore, the deviance is stated as [30]

$$\begin{aligned} D &= 2[l(\lambda_{i(\max)}; y) - l(\hat{\lambda}_i; y)] \\ &= 2 \left[\sum_{i=1}^n y_i \log \left(\frac{y_i}{\hat{\lambda}_i} \right) - \sum_{i=1}^n (y_i - \hat{\lambda}_i) \right] \end{aligned} \quad (19)$$

which is twice the difference between the maximum log-likelihood achievable and the log-likelihood of the fitted model.

2.2.5 Model selection criteria

To identify the most effective model for explaining the occurrences of daily rainfall, several authors including McCullagh and Nelder [31] and Agresti [1] suggested various model selection procedures and they also identified some limitations and drawbacks. These selection procedures sometimes provide almost equal emphasis for several possible models; often procedures do not provide the best model among the models sufficiently for a true alternative hypothesis. For overcoming these problems, Akaike's information criterion (AIC) and Bayesian information criterion (BIC) procedures are employed in this study.

2.2.5.1 Akaike's information criterion (AIC) Akaike [5] developed AIC based on the extension of maximum likelihood principle. The likelihood approach developed by Bartlett [8] under the assumption that the frequency of counts is asymptotically normally distributed and also indicate that the frequency counts are occurred depending on the sequence of k explanatory variables along with considering MLE method for estimating the parameters of k explanatory variables. Further, to develop autoregressive model Akaike [6] employed final prediction error (FPE) considering the mean square prediction error of predictors. Considering the extension form of FPE procedure Akaike [3–5] formulated the following AIC form

$$\begin{aligned} \text{AIC} &= -2(\text{maximum log likelihood}) \\ &\quad 2(\text{number of estimable parameters in the model}) \end{aligned} \quad (20)$$

Following Kullback–Leibler information [28], this measure indicates the deviation of fitting model for the true structure. The minimum AIC estimate (MAICE) indicates the most effective model among the fitted models [35]. That is,

$$\begin{aligned} \text{AIC}(i) &< \text{AIC}(i+1) < \dots < \text{AIC}(i+s), \\ i &= 1, 2, \dots, \infty \quad \text{and} \quad s = 0, 1, 2, \dots, \infty \end{aligned} \quad (21)$$

where $(i+s)$ indicate the number of models and $\text{AIC}(i)$ provides the best model compared to others.

2.2.5.2 Bayesian information criterion (BIC) Schwarz [36] observed that the AIC procedure is not optimal and consistent, and argued Bayesian procedure (i.e., Bayesian information criterion, BIC) provides consistent estimator for determining the best effective model in several models under the assumption that observations come from an

exponential family distribution. He has also shown that BIC is not only consistent, but it is asymptotically optimal (i.e., minimizing the expected loss) under the specification form of prior distribution and the type of loss function. Further, Bayes estimator is asymptotically equivalent to MLE [27] considering large n . Therefore, Schwarz [36] defined following BIC form:

$$\begin{aligned} \text{BIC} = & -2(\text{maximum log likelihood}) \\ & + 2(\text{number of estimable parameters in the model}) \\ & \times \log n \end{aligned} \quad (22)$$

here n indicates sample size. For selecting best model, Sakamoto [35] suggested the minimum BIC estimate (MBICE) as the most effective model among the fitted models. That is,

$$\begin{aligned} \text{BIC}(i) < \text{BIC}(i+1) < \dots < \text{BIC}(i+s), \\ i = 1, 2, \dots \propto \quad \text{and} \quad s = 0, 1, 2, \dots \propto \end{aligned} \quad (23)$$

where $(i+s)$ indicate the number of models and $\text{BIC}(i)$ provides the best model compared to others.

3 Data

Bangladesh is situated in subtropical region, and it is characterized by wide seasonal variations due to heavy rainfall, high temperature and high humidity. Considering these variations of the country the Bangladesh Bureau of Statistics (BBS) subdividing the country into three seasons (from March to June for summer, from July to October for monsoon, and from November to February for winter). In this country, more than 90 percent annual rainfall are occurred in summer and monsoon seasons. Within these seasons, maximum rainfall is occurred in monsoon season [2, 29]. Bangladesh is agro-based country; the crop plantation, the crop growing and harvesting are basically depended on the behavior of country's annual rainfall patterns. To analyze the rainfall occurrence patterns, the daily rainfall recording stations of the country have been subdivided into five regions such as southeast region (Feni station), northeast region (Sylhet station), southwest region (Khulna station), northwest region (Rajshahi station) and mid-region (Dhaka station). The daily rainfall occurrences basically depend on the behavior of climatic variables like temperature (maximum and minimum) and relative humidity [38]; these variables have been identified as covariates in the study. The study has been collected data for daily rainfall, temperature and relative humidity from January, 1975 to December, 2014 (40 years) from the Bangladesh Meteorological Department (BMD) for each rainfall region. The missing values of these data have

been substituted by using the Single Best Estimator (SBE) method considering the reference stations (within 100 km of regional rainfall station) [24].

3.1 Count data of daily rainfall occurrences

To transfer daily rainfall data into count data, the study considers binary variable; when rainfall occurred within a day greater or equal to 1 mm, then the observation is considered as 1, otherwise 0. The summer and monsoon seasons, each of them, are formed by four months. Each month contains four weeks plus additional two or three days. That is, summer season of each year contains 17 weeks plus 3 additional days. These three additional days are excluded from summer season and included in monsoon season (for the continuation of time series data); then, the monsoon season of each year is formed by 18 weeks. Similarly, for these seasons each month contains two fortnights or two fortnights plus one additional day. So summer season contains 8 fortnights plus two additional days in each year. These two additional days are excluded from summer season and considered in monsoon season. Then the monsoon season is formed by 8 fortnights plus 5 additional days in each year. Therefore, to perform the study, the monsoon season is considered by 8 fortnights and the last 5 days of October month are excluded as the 5 additional days in each year. So for weekly count data of daily rainfall occurrences, this study $(17 \times 7 \times 40) = 4760$ days and $(17 \times 40) = 680$ weeks for summer season and $(18 \times 7 \times 40) = 5040$ days and $(18 \times 40) = 720$ weeks for monsoon season are considered. Similarly, for fortnightly count data, this study $(8 \times 15 \times 40) = 4800$ days and $(8 \times 40) = 320$ fortnights are considered for each of summer season and monsoon season.

4 Results and discussions

To explain the influence of climatic variables on the patterns of rainfall occurrences for the summer and monsoon seasons, the GLMs for Poisson distribution have been employed considering weekly and fortnightly count data of rainfall occurrences. Further, deviance of the fitted models has been employed to identify the goodness of fit of the models, and AIC and BIC have been considered to select the most effective model among the models. Similarly, to identify the significant effect of climatic variables on the count data of daily rainfall occurrences, different test statistics are used. Therefore, to formulate GLMs for weekly and fortnightly count data of daily rainfall occurrences for the summer and monsoon seasons, the study considers maximum temperature, minimum temperature and relative humidity as the covariates. The observation

of each covariate considers weekly or fortnightly mean for performing GLMs, i.e., for weekly data the i th observation of j th covariate indicates $x_i : \bar{x}_{i1}, \bar{x}_{i2}, \dots, \bar{x}_{ik}$; here $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k$ [10]. To identify the patterns of weekly and fortnightly daily rainfall occurrences for the summer and monsoon seasons, the result of GLMs under the Poisson distribution is discussed below.

4.1 Analysis of weekly and fortnightly count data of daily rainfall occurrences for the summer and monsoon seasons

The frequency distribution of the summer and monsoon seasons for weekly and fortnightly count data of daily rainfall occurrences for five regions of Bangladesh is shown in Appendix 1, 2 and 3. The distribution of weekly rainfall occurrences for summer season indicates that within 26.2 percent, 9.6 percent, 27.9 percent, 28.2 percent, 21.6 percent weeks are not occurred any rainfall in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively. Similarly, the distribution of weekly rainfall occurrences for monsoon season indicates that within 8.6 percent, 7.6 percent, 8.8 percent, 11.5 percent and 7.1 percent weeks are not occurred any rainfall in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively. Further, the distribution of fortnightly rainfall occurrences for summer season explains that within 15.3 percent, 2.8 percent, 12.5 percent, 12.5 percent and 7.2 percent weeks are not occurred any rainfall in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively. Similarly, the distribution of fortnightly occurrences of rainfall for monsoon season indicates that within 1.9 percent, 1.3 percent, 1.6 percent, 2.2 percent and 1.3 percent weeks are not occur any rainfall in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively.

In summer season of Feni station within the highest week (18.1 percent), one-day rainfall is occurred, and in the highest fortnight (9.4 percent), six days of rainfall are occurred. Further, in this season, within the lowest week (3.5 percent) seven days of rainfall are occurred, and within the lowest fortnight (0.3 percent) fourteen days or fifteen days of rainfall are occurred. In monsoon season of this station within the highest week (18.1 percent), four days of rainfall are occurred, and within the highest fortnight (12.8 percent), ten days of rainfall are occurred. Again, in this season, within the lowest week (6.5 percent) and within the lowest fortnight (1.5 percent) one-day rainfall is occurred.

In summer season of Sylhet station, within highest week (16.3 percent), five days of rainfall are occurred, and in the highest fortnight (9.7 percent) eleven days of rainfall are occurred. In monsoon season of the station, within highest week (28.8 percent) seven days of rainfall are occurred, and in the highest fortnight (15.6 percent), twelve days of rainfall are occurred. Further, in this season, within the lowest week

(5.1 percent) one-day rainfall is occurred, and in the lowest fortnight (0.6 percent) two-day rainfall is occurred.

In summer season of Khulna station within the highest week (20.6 percent) and in the highest fortnight (12.8 percent), one-day rainfall is occurred. In monsoon season of the station, within the highest week (19.6 percent) six days of rainfall are occurred, and in the highest fortnight (12.8 percent), twelve days of rainfall are occurred. Again, in this season within the lowest week (6.3 percent), and in the lowest fortnight (1.9 percent) one-day rainfall is occurred.

In summer season of Rajshahi station, within the highest week (18.5 percent) two days of rainfall are occurred, and in the highest fortnight (14.1 percent) one-day rainfall is occurred. In monsoon season of the station, within the highest week (18.1 percent) four days of rainfall are occurred, and in the highest fortnight (13.8 percent), eight days of rainfall are occurred.

In summer season of Dhaka station within the highest week (15.6 percent) two days of rainfall are occurred, and in the highest fortnight (11.3 percent), three days of rainfall are occurred. In monsoon season of the station within the highest week (20.8 percent), five days of rainfall are occurred, and in the highest fortnight (14.7 percent) eleven days of rainfall are occurred. Further, in this season within the lowest week (6.8 percent) one-day rainfall is occurred, and in the lowest fortnight (0.9 percent) fifteen-day rainfall is occurred.

4.2 GLMs for Poisson distribution for analyzing count data of rainfall occurrences

In order to identify the effect of climatic variables on weekly and fortnightly count data of rainfall occurrences for five regions (southeast (Feni station), northeast (Sylhet station), southwest (Khulna station), northwest (Rajshahi station) and mid-region (Dhaka station)) of Bangladesh, the study formulates generalized linear models (GLMs) for Poisson distribution considering maximum temperature (x_{i1}), minimum temperature (x_{i2}) and relative humidity (x_{i3}) as covariates. The link function (λ) of GLMs can be stated as $\lambda_i = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})$, $i = 1, 2, \dots, n$. Here λ indicates the mean of the Poisson distribution and β indicates the parameter of the corresponding variable, i.e., this parameter implies the relative change or semi-elasticity of weekly/fortnightly rainfall occurrences. To analyze the effect of climatic variables on count data of rainfall occurrences, the study has formulated four models, within these models two for weekly data of summer and monsoon seasons, and two for fortnightly data of summer and monsoon seasons. In these models, y_i indicates the weekly or fortnightly count data of rainfall occurrences. For performing the analysis, the study has considered MLE method under Newton–Raphson iteration procedure to

estimate the parameters of these models and the t_z statistic is employed to identify the significance of parameter. Using the parameter of respective climatic variable, the study has identified the average marginal effect for the occurrence of rainfall within weekly or fortnightly. For each region, the estimated parameters of the summer and monsoon seasons model for weekly and fortnightly count data of rainfall occurrences are stated in Table 1. The estimated average marginal effect of each climatic variable for these regions is also shown in this table.

Table 1 indicates that the minimum temperature and relative humidity are explained positive effect on weekly or fortnightly daily rainfall occurrences in summer and monsoon seasons for all the regions. The negative influence is found for maximum temperature on the weekly or fortnightly occurrences of rainfall for summer and monsoon seasons. However, the influence of these climatic variables on the occurrences of daily rainfall within weekly or fortnightly is found significant in all the regions.

In southeast region (Feni station) for summer season, the influence of relative humidity is found highest among the covariates on the occurrence of daily rainfall in weekly and fortnightly. That is, within week or fortnight one-unit increase in relative humidity indicates the likelihood of the occurrence of rainfall will be increased by 12 percent, considering other variables influence as usual. For this season, one-unit increase in maximum temperature indicates the probability of rainy days will be decreased by 7 percent within week and 9 percent within fortnight. Further, average marginal effect of this period indicates that if one-unit increase in relative humidity, then the expected rainy days will increase 0.30 within week and 0.59 within fortnight. For this region in monsoon season, within week or fortnight one-unit increase in minimum temperature indicates that the probability of rainy days will be increased by 22 percent. Further, the average marginal effect indicates that if there is one-unit increase in minimum temperature, then the expected rainy days will be increased by 0.86 within week and 1.90 within fortnight. Besides, this effect indicates that if one unit maximum temperature getting increases, then the expected rainy days will be reduced by 0.27 within week and 0.88 within fortnight.

In summer season of northeast region (Sylhet station), if one unit of relative humidity increases, then the likelihood of the rainy days occurrence will be increased by 8 percent within week and 6 percent within fortnight, considering other covariates effect fixed. In monsoon season of this region, if one unit of minimum temperature increases, then the occurrences of daily rainfall will be increased by 18 percent within week and 19 percent within fortnight. Further, if one unit of maximum temperature increases, then the occurrence of daily rainfall will be reduced by 4 percent within week and 3 percent within fortnight.

In summer season of southwest region (Khulna station), the GLMs indicate that if one unit of relative humidity rises, then the occurrences of daily rainfall will be increased by 9 percent within week. Again, if one unit of minimum temperature increases, then the occurrences of daily rainfall will be increased by 10 percent within fortnight considering other climatic variables effect as usual. For monsoon season of this region, GLMs are indicate that if one unit of minimum temperature increases, then the occurrences of daily rainfall will be increased by 23 percent within week and 25 percent within fortnight. Further, if one unit of maximum temperature increases, then the occurrences of daily rainfall will be reduced by 8 percent within week and 10 percent within fortnight. In this season, the marginal effect indicates that if one unit of minimum temperature getting rises, then the expected rainy days will increase 0.43 within week and 2.23 within fortnight.

For summer season of northwest region (Rajshahi station), the GLMs indicate that if one unit of relative humidity increases, then the occurrences of daily rainfall will be increased by 8 percent within week and 6 percent within fortnight. For this region and season if one unit of maximum temperature increases, then the occurrences of daily rainfall will be reduced by 3 percent within week and 4 percent within fortnight. In monsoon season of the region, GLMs indicate that considering other climatic variables effect as usual, if one unit of minimum temperature increases, then the occurrences of daily rainfall will be increased by 17 percent within week and within fortnight. Further, for this region and season, if one unit of maximum temperature increases, then the occurrences of daily rainfall will be reduced by 5 percent within week and 6 percent within fortnight.

In summer season of mid-region (Dhaka station), the GLMs indicate that if one unit of relative humidity increases, then the occurrences of daily rainfall will be increased by 9 percent within week and 7 percent within fortnight. In monsoon season of this region, it indicates that if one unit of minimum temperature increases, then the occurrences of daily rainfall will be increased by 17 percent within week and 19 percent within fortnight. For this season, the marginal effect of minimum temperature indicates that if one unit of the variable increases, then the expected rainy days will increase 0.10 within week and 1.66 within fortnight. Similarly, the marginal effect of maximum temperature indicates that if one unit of the variable rises then the expected rainy days will reduce 0.04 within week and 0.68 within fortnight.

To test the jointly significance of all climatic variables influence on the occurrence of rainfall within weekly or fortnightly for the summer and monsoon seasons of five regions, the study has considered likelihood ratio test (LRT). Therefore, according to Eq. (15) the null hypothesis is defined as $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. Further, to test the goodness of fit of the models the study has considered

Table 1 Estimated parameters of GLMs for summer and monsoon seasons for weekly and fortnightly count data of rainfall occurrences

Regions (rainfall station)	Seasons	Variables	GLMs for weekly count data				GLMs for fortnightly count data			
			Estimated value of parameters	t _z -value	p-value	Average marginal effect of variable	Estimated value of parameters	t _z -value	p-value	Average marginal effect of variable
Southeast (Feni station)	Summer	Maximum temperature	-0.071883 (0.029597)	-2.953	0.0135*	-0.03692	-0.088271 (0.030512)	-1.910	0.0466*	-0.05043
		Minimum temperature	0.045804 (0.014857)	4.658	0.00351*	0.08749	0.093291 (0.005808)	6.169	3.21e-07*	0.05752
		Relative humidity	0.124633 (0.008869)	14.053	2e-16*	0.2951	0.116224 (0.010837)	10.725	2e-16*	0.5942
	Monsoon	Maximum temperature	-0.067932 (0.028816)	-2.357	0.0184*	-0.2656	-0.10218 (0.03473)	-2.942	0.00326*	-0.8826
		Minimum temperature	0.220547 (0.024422)	9.031	2e-16*	0.8623	0.21939 (0.02883)	7.610	2.75e-14*	1.895
		Relative humidity	0.078123 (0.009796)	7.975	1.52e-15*	0.3054	0.06747 (0.01241)	5.437	5.43e-08*	0.5828
Northeast (Sylhet station)	Summer	Maximum temperature	-0.066372 (0.021805)	-3.044	0.00234*	-0.2576	-0.026203 (0.026571)	-1.086	0.02717*	-0.1712
		Minimum temperature	0.077747 (0.021653)	3.591	0.00033*	0.3017	0.025332 (0.026015)	3.150	0.00163*	0.86
		Relative humidity	0.080432 (0.006394)	12.580	2e-16*	0.3121	0.064163 (0.007495)	8.560	2e-16*	0.4051
	Monsoon	Maximum temperature	-0.035729 (0.023833)	-1.499	0.03717*	-0.2191	-0.02792 (0.02882)	-2.969	0.00333*	-0.2954
		Minimum temperature	0.179416 (0.022674)	7.913	2.51e-15*	0.2118	0.19498 (0.02625)	7.426	1.12e-13*	2.063
		Relative Humidity	0.084514 (0.008192)	10.316	2e-16*	0.5366	0.07851 (0.00977)	8.036	9.26e-16*	0.8307
Southwest (Khulna station)	Summer	Maximum temperature	-0.087954 (0.021933)	-4.010	6.07e-05*	-0.1806	-0.143373 (0.025114)	-5.709	1.14e-08*	-0.313
		Minimum temperature	0.028131 (0.017663)	2.593	0.00111*	0.05775	0.097254 (0.021004)	4.630	3.65e-06*	0.9421
		Relative humidity	0.089018 (0.007005)	12.707	2e-16*	0.1827	0.064843 (0.007969)	8.137	4.05e-16*	0.3618
	Monsoon	Maximum temperature	-0.077242 (0.022159)	-3.486	0.00049*	-0.6362	-0.100171 (0.027362)	-3.661	0.00025*	-0.8975
		Minimum temperature	0.232468 (0.022986)	10.113	2e-16*	0.4316	0.248991 (0.027226)	9.145	2e-16*	2.231
		Relative humidity	0.089281 (0.007551)	11.824	2e-16*	0.2877	0.081094 (0.009104)	8.907	2e-16*	0.7266

Table 1 (continued)

Regions (rainfall station)	Seasons	Variables	GLMs for weekly count data			GLMs for fortnightly count data				
			Estimated value of parameters	t _z -value	p-value	Average marginal effect of variable	Estimated value of parameters	t _z -value	p-value	Average marginal effect of variable
Northwest (Rajshahi station)	Summer	Maximum temperature	-0.028794 (0.017817)	-3.616	0.00106*	-0.05115	-0.039012 (0.022438)	-2.439	0.00821*	-0.172
		Minimum temperature	0.006321 (0.019583)	1.523	0.0347*	0.01123	0.019254 (0.022203)	1.867	0.03859*	0.5701
	Monsoon	Relative humidity	0.076052 (0.004763)	15.968	2e-16*	0.1351	0.064857 (0.005115)	12.679	2e-16*	0.4024
		Maximum temperature	-0.05028 (0.01596)	-3.151	0.00163*	-0.1496	-0.057008 (0.020128)	-2.832	0.00462*	-0.4324
		Minimum temperature	0.16664 (0.02444)	6.819	9.19e-12*	0.07383	0.169647 (0.028592)	5.933	2.97e-09*	1.287
Mid-region (Dhaka station)	Summer	Relative humidity	0.11762 (0.00748)	15.725	2e-16*	0.2487	0.094464 (0.009444)	10.003	2e-16*	0.7165
		Maximum temperature	-0.025720 (0.024509)	-2.049	0.00939*	-0.06517	-0.007326 (0.028234)	-1.259	0.07953*	-0.2608
	Monsoon	Minimum temperature	0.072590 (0.019466)	3.729	0.00019*	0.1839	0.017601 (0.022997)	1.765	0.04441*	0.6886
		Relative humidity	0.088512 (0.005744)	15.410	2e-16*	0.2243	0.072387 (0.006514)	11.112	2e-16*	0.2372
		Maximum temperature	-0.06590 (0.02672)	-2.466	0.0137*	-0.04006	-0.077186 (0.034943)	-2.209	0.02718*	-0.6754
		Minimum temperature	0.17395 (0.02230)	7.802	6.09e-15*	0.09626	0.190061 (0.025837)	7.356	1.89e-13*	1.663
		Relative humidity	0.05991 (0.00650)	9.217	2e-16*	0.3959	0.041957 (0.008002)	5.243	1.58e-07*	0.3671

Parenthesis indicates the standard error of parameter. *indicates the level of significance (p-value less than 0.05)

the deviance of the models, which is discussed in Eq. (19). Besides, to select the appropriate model for explaining the weekly and fortnightly rainfall occurrences in the summer and monsoon seasons for five regions, the study has considered AIC and BIC procedure (Sect. 2.2.5). The results of LRT, deviance of the models, AIC and BIC are shown in Table 2.

Form Table 2, the LRT indicates that for the occurrences of rainfall within weekly or fortnightly in summer and monsoon seasons, the influence of climatic variables (maximum and minimum temperature, and relative humidity) is found jointly significant for all the regions. The value of LRT is found lower for fortnightly data compared to weekly data for respective season and region. Further, in all the regions this table also indicates that the value of deviance of the models for the fortnightly data in summer and monsoon seasons provides very much low value than that of weekly data. For all the regions, the GLMs for the Poisson distribution provide good fit for the fortnightly count data than that of weekly count data of rainfall occurrences for the summer and monsoon seasons.

Further, for all the regions the value of AIC and BIC for the fortnightly data in the summer and monsoon seasons is found minimum than that of weekly data. Therefore, for climatic variables maximum and minimum temperatures and relative humidity, the GLMs for the Poisson

distribution explain satisfactory information for the fortnightly rainfall occurrences in summer and monsoon seasons for all the regions of the country.

5 Conclusions

Bangladesh is an agro-based country. The country is always suffered by the devastating natural disaster, e.g., flash flood, drought, heavy rainfall, etc. For saving country's people from these sufferings, country's economists and agriculturists are needed having appropriate knowledge about the patterns of daily rainfall occurrences. For this purpose, the study develops GLMs for Poisson distribution considering count data of rainfall occurrences. In these models the maximum and minimum temperatures and relative humidity as explanatory variables are considered. To perform the study, weekly and fortnightly count data of daily rainfall for five regions of the country of the summer and monsoon seasons have been considered as response variable. The GLMs indicate that the effect of explanatory variables on the count data of daily rainfall occurrences is significant. For both seasons, the influence of maximum temperature on the weekly and fortnightly occurrences of rainfall is found negative. Besides, for both seasons the influence of minimum temperature and relative humidity

Table 2 Value of Chi-square statistic for LRT, deviance of the models, AIC and BIC

Regions (rainfall station)	Period of count data	Season	Value of LRT	<i>p</i> -value of LRT	Deviance of the models	Value of AIC	Value of BIC
Southeast (Feni station)	Weekly	Summer	732.4279	1.950854e−158*	778.4754	2236.294	2254.383
		Monsoon	456.3767	1.353947e−98*	581.2251	2712.948	2731.265
	Fortnightly	Summer	606.7415	3.482751e−131*	503.8378	1450.777	1465.85
		Monsoon	263.7144	7.067889e−57*	318.171	1558.898	1573.971
Northeast (Sylhet station)	Weekly	Summer	702.7861	5.222932e−152*	409.746	2391.938	2410.026
		Monsoon	572.8741	7.650508e−124*	477.5316	2765.455	2783.772
	Fortnightly	Summer	553.0918	1.485128e−119*	211.4741	1421.793	1436.867
		Monsoon	358.9575	1.714069e−77*	225.4015	1537.552	1552.626
Southwest (Khulna station)	Weekly	Summer	704.087	2.727798e−152*	682.875	2053.789	2071.877
		Monsoon	538.7403	1.91631e−116*	508.7504	2663.66	2681.977
	Fortnightly	Summer	590.8144	9.87827e−128*	387.9696	1310.201	1325.275
		Monsoon	326.7133	1.64228e−70*	230.1346	1486.486	1501.559
Northwest (Rajshahi station)	Weekly	Summer	702.236	6.873737e−152*	480.824	1791.628	1809.716
		Monsoon	608.8456	1.218327e−131*	485.7024	2484.154	2502.466
	Fortnightly	Summer	557.02	2.0907e−120*	259.62	1148.121	1163.194
		Monsoon	304.6007	1.004692e−65*	208.4305	1408.502	1423.575
Mid-region (Dhaka station)	Weekly	Summer	781.6573	4.114104e−169*	542.968	2107.543	2125.631
		Monsoon	392.7914	8.064028e−85*	513.4418	2683.262	2701.579
	Fortnightly	Summer	588.6782	2.869307e−127*	254.1672	1286.686	1301.76
		Monsoon	198.9398	7.148948e−43*	228.1384	1487.945	1503.018

For LRT Chi-square is explained with 3 degrees of Freedom. Further, *indicates the level of significance (*p*-value less than 0.05)

is found positive in weekly and fortnightly daily rainfall occurrences for all the regions. Using AIC and BIC, the GLMs for Poisson distribution are explained satisfactory results for fortnightly occurrences of daily rainfall for both seasons and all stations.

In the summer season, other climatic variables influence considering usual; within fortnight, one-unit increase in relative humidity indicates that the probability of rainy days will be increased by 12 percent in Feni station, 6 percent in Sylhet, Khulna and Rajshahi stations, and 7 percent in Dhaka station. Further, within fortnight in this season one unit increase in maximum temperature indicates that the probability of rainy days will be decreased by 9 percent, 3 percent, 14 percent, 4 percent and 0.7 percent in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively. In the summer season within this period, average marginal effect indicates that if one unit increases of relative humidity, then the expected rainy days will be increased by 0.59 in Feni station. Similarly, if one unit increases of minimum temperature, then the expected rainy days will be increased by 0.86, 0.94, 0.57 and 0.69 in Sylhet, Khulna, Rajshahi and Dhaka stations, respectively.

In monsoon season within fortnight, one-unit increase in minimum temperature indicates the probability of rainy days will be increased by 22 percent, 19 percent, 25 percent, 17 percent and 19 percent in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively. Further, in this season one-unit increase in maximum temperature indicates the probability of rainy days will be decreased by 10 percent, 2 percent, 10 percent, 6 percent, 8 percent in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively. For this season and period, the average marginal effect indicates that if one unit minimum temperature increases then the

expected rainy days will be increased by 1.90, 2.06, 2.23, 1.29 and 1.66 in Feni, Sylhet, Khulna, Rajshahi and Dhaka stations, respectively. Therefore, the relative humidity for summer season and the minimum temperature for monsoon season play key role for changing fortnightly rainy days in all the regions of the country. Therefore, for the development of rural economy these GLMs may helpful to formulate effective plan for cultivating agricultural crops, vegetable, etc., and to save fisheries, wildlife, human asset, etc.

Compliance with ethical standards

Conflict of interest There is no any type of conflict of interest among the authors (financial, personal, etc.) to prepare the article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Frequency distribution of weekly and fortnightly count data of rainfall occurrences for summer and monsoon seasons of Feni and Sylhet stations

Number of day	Summer season								Monsoon season							
	Weekly frequency		Percent of total frequency		Fortnightly frequency		Percent of total frequency		Weekly frequency		Percent of total frequency		Fortnightly frequency		Percent of total frequency	
	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet
0	178	65	26.2	9.6	49	9	15.3	2.8	62	55	8.6	7.6	6	4	1.9	1.3
1	123	61	18.1	9.0	24	12	7.5	3.8	47	37	6.5	5.1	5	7	1.5	2.2
2	85	73	12.5	10.7	29	11	9.0	3.4	73	40	10.1	5.6	11	2	3.4	0.6
3	81	88	11.9	12.9	24	17	7.5	5.3	107	51	14.9	7.1	13	9	4.1	2.8
4	79	85	11.6	12.5	28	16	8.7	5.0	130	65	18.1	9.0	17	11	5.3	3.4
5	70	111	10.3	16.3	22	15	6.9	4.7	108	118	15.0	16.4	13	11	4.1	3.4
6	40	106	5.9	15.6	30	29	9.4	9.1	110	147	15.3	20.4	19	14	5.9	4.4
7	24	91	3.5	13.4	22	16	6.9	5.0	83	207	11.5	28.8	28	13	8.7	4.1
8	–	–	–	–	20	26	6.2	8.1	–	–	–	–	26	15	8.1	4.7
9	–	–	–	–	21	29	6.6	9.1	–	–	–	–	29	19	9.1	5.9
10	–	–	–	–	14	29	4.4	9.1	–	–	–	–	41	17	12.8	5.3
11	–	–	–	–	20	31	6.3	9.7	–	–	–	–	37	18	11.5	5.6

Number of day	Summer season								Monsoon season							
	Weekly frequency		Percent of total frequency		Fortnightly frequency		Percent of total frequency		Weekly frequency		Percent of total frequency		Fortnightly frequency		Percent of total frequency	
	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet	Feni	Sylhet
12	–	–	–	–	7	20	2.2	6.3	–	–	–	–	29	50	9.1	15.6
13	–	–	–	–	8	26	2.5	8.1	–	–	–	–	28	47	8.8	14.7
14	–	–	–	–	1	22	0.3	6.9	–	–	–	–	13	41	4.1	12.8
15	–	–	–	–	1	12	0.3	3.8	–	–	–	–	5	42	1.6	13.1
Total	680	680	100	100	320	320	100	100	720	720	100	100	320	320	100	100.0

Appendix 2: Frequency distribution of weekly and fortnightly count data of rainfall occurrences for summer and monsoon seasons of Khulna and Rajshahi stations

Number of day	Summer season								Monsoon season							
	Weekly frequency		Percent of total frequency		Fortnightly frequency		Percent of total frequency		Weekly frequency		Percent of total frequency		Fortnightly frequency		Percent of total frequency	
	Khulna	Rajshahi	Khulna	Rajshahi	Khulna	Rajshahi	Khulna	Rajshahi	Khulna	Rajshahi	Khulna	Rajshahi	Khulna	Rajshahi	Khulna	Rajshahi
0	190	192	27.9	28.2	40	40	12.5	12.5	63	83	8.8	11.5	5	7	1.6	2.2
1	140	154	20.6	22.6	41	45	12.8	14.1	45	57	6.3	7.9	6	4	1.9	1.3
2	96	126	14.1	18.5	36	44	11.3	13.8	64	98	8.9	13.6	7	16	2.2	5.0
3	96	107	14.1	15.7	31	38	9.7	11.9	92	116	12.8	16.1	13	15	4.1	4.7
4	62	47	9.1	6.9	31	34	9.7	10.6	108	130	15.0	18.1	14	15	4.4	4.7
5	60	37	8.8	5.4	32	31	10.0	9.7	129	108	17.9	15.0	16	26	5.0	8.1
6	24	11	3.5	1.6	26	24	8.1	7.5	141	97	19.6	13.5	16	22	5.0	6.9
7	12	6	1.8	0.9	19	20	5.9	6.3	78	31	10.8	4.3	21	34	6.6	10.6
8	–	–	–	–	15	18	4.7	5.6	–	–	–	–	27	44	8.4	13.8
9	–	–	–	–	12	10	3.8	3.1	–	–	–	–	31	39	9.7	12.2
10	–	–	–	–	14	8	4.4	2.5	–	–	–	–	37	37	11.6	11.6
11	–	–	–	–	10	5	3.1	1.6	–	–	–	–	37	33	11.6	10.3
12	–	–	–	–	8	3	2.5	0.9	–	–	–	–	41	18	12.8	5.6
13	–	–	–	–	4	0	1.3	0	–	–	–	–	28	6	8.8	1.9
14	–	–	–	–	0	0	0.0	0	–	–	–	–	14	4	4.4	1.3
15	–	–	–	–	1	0	0.3	0	–	–	–	–	7	0	2.2	0
Total	680	680	100	100	320	320	100	100	720	720	100	100	320	320	100	100.0

Appendix 3: Frequency distribution of weekly and fortnightly count data of rainfall occurrences for summer and monsoon seasons of Dhaka station

Number of day	Summer Season				Monsoon Season			
	Weekly frequency	Percent of total frequency	Fortnightly frequency	Percent of total frequency	Weekly frequency	Percent of total frequency	Fortnightly frequency	Percent of total frequency
0	147	21.6	23	7.2	51	7.1	4	1.3
1	103	15.1	28	8.8	49	6.8	4	1.3
2	106	15.6	24	7.5	63	8.8	4	1.3
3	96	14.1	36	11.3	98	13.6	12	3.8
4	87	12.8	26	8.1	140	19.4	11	3.4
5	90	13.2	29	9.1	150	20.8	13	4.1
6	35	5.1	30	9.4	112	15.6	22	6.9
7	16	2.4	28	8.8	57	7.9	23	7.2
8	–	–	26	8.1	–	–	42	13.1
9	–	–	26	8.1	–	–	41	12.8
10	–	–	14	4.4	–	–	40	12.5
11	–	–	12	3.8	–	–	47	14.7
12	–	–	10	3.1	–	–	28	8.8
13	–	–	6	1.9	–	–	17	5.3
14	–	–	1	0.3	–	–	9	2.8
15	–	–	1	0.3	–	–	3	0.9
Total	680	100	320	100	720	100	320	100

References

- Agresti A (1984) Analysis of ordinal categorical data. Wiley, New York
- Ahasan MN, Chowdhary AM, Quadir DA (2010) Variability and trends of summer monsoon rainfall over Bangladesh. *J Hydrol Meteorol* 7(1):1–17. <https://doi.org/10.3126/jhm.v7i1.5612>
- Akaike H (1974) A new look at the statistical identification model. *IEEE Trans Autom Control* 19(6):716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Akaike H (1972) Use of information theoretic quantify for statistical model identification. In: Proceedings of the 5th Hawaii international conference system sciences, Western Periodicals, pp 249–250.
- Akaike H (1972) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Second international symposium on information theory. Akademiai Kiado, Budapest, pp 267–281
- Akaike H (1970) Statistical predictor identification. *Ann Inst Statist Math* 22:203–217
- Andersen EB (1970) Asymptotic properties of conditional maximum likelihood estimators. *J R Stat Soc Ser B* 32:283–301
- Bartlett MS (1951) The frequency goodness of fit test for probability chains. *Proc Cambridge Philos Soc* 47:86–95
- Blau DM, Robins PK (1991) Turnover in child care arrangements. *Rev Econ Stat* 73:152–157
- Buishanand TA, Shabalova MV, Brandsma T (2004) On the choice of the temporal aggregation level for statistical downscaling of precipitation. *J Clim Am Meteorol Soc* 17:1816–1827
- Cameron AC, Trivedi PK (1998) Regression analysis of count data. Cambridge University Press, Cambridge
- Cameron AC, Trivedi PK (1986) Econometric models based on count data: comparisons and applications of some estimators. *J Appl Econom* 1:29–53
- Caskey JE (1963) A Markov chain for the probability of precipitation occurrence in intervals of various lengths. *Month Weath Rev* 91:298–301
- Chandler RE, Wheeler HS (2002) Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. *Water Resour Res*. <https://doi.org/10.1029/2001WR000906>
- Chappell WF, Kimenyi MS, Mayer WJ (1990) A Poisson probability model of entry and market structure with an application to US industries during 1972–77. *South Econ J* 56:918–927
- Coe R, Stern RD (1982) Fitting models to rainfall data. *J Appl Meteorol Soc* 21:1024–1031
- Cox DR, Isham V (1994) Stochastic models of precipitation. In: Barnett V, Turkman KF (eds) Statistics for the environment 2, water issues. Wiley, New York, pp 3–18
- Dobson AJ (2002) An introduction to generalized linear models, 2nd edn. Chapman & Hall, New York
- Feinstein JS (1989) The safety regulations of US nuclear power plants: violations, inspections, and abnormal occurrences. *J Polit Econ* 97:115–154
- Hausman J, Hall BH, Griliches Z (1984) Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* 52:909–938
- Hilbe JM (2011) Negative binomial regression, 2nd edn. Cambridge University Press, Cambridge

22. Islam MA, Chowdhury RI (2006) A higher order Markov model of analyzing covariate dependence. *Appl Math Model* 30:477–488
23. Islam MA, Chowdhury RI (2017) Analysis of repeated measures data. Springer Nature Singapore sPte Ltd, 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore. <https://doi.org/10.1007/978-981-10-3794-8>
24. Jahan F, Sinha NC, Rahman MM, Rahman MM, Mondal SHM, Islam MA (2019) Comparison of missing value estimation techniques in rainfall data of Bangladesh. *Theor Appl Climatol* 136:1115–1131. <https://doi.org/10.1007/s00704-018-2537-y>
25. Johnson NL, Kotz S, Kemp AW (1992) Univariate discrete distributions, 2nd edn. Wiley, New York
26. Johnson NL, Kotz S (1969) Distributions in statistics: discrete distributions. Wiley, New York
27. Kendall MG, Stuart A (1973) The advance theory of statistics, vol II. Griffin, London
28. Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22:79–86
29. Mannan MA, Chowdhury MAM, Karmakar S (2016) Prediction of rainfall over southeastern part of Bangladesh during monsoon season. *Int J Integ Sci Technol* 2:73–82
30. McCullagh P (1986) The conditional distribution of goodness-of-fit statistics for discrete data. *J Am Stat Assoc* 81:104–107
31. McCullagh P, Nelder JA (1983) Generalized linear models. Chapman and Hall, London
32. McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, New York
33. Nelder JA, Wedderburn RWM (1972) Generalized linear models. *J R Stat Soc Ser A* 135:370–384
34. Saffari SE, Adnan R (2010) Zero-inflated Poisson regression models with right censored count data. *MATEMATIKA* 27(1):21–29
35. Sakamoto Y (1991) Categorical data analysis by AIC. Kluwer Academic Publishers, New York
36. Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464
37. Segond ML, Onof C, Wheeler HS (2006) Spatial-temporal disaggregation of daily rainfall from a generalized linear model. *J Hydrol* 331:674–689
38. Sinha NC, Islam MA, Ahmed KS (2011) Logistic regression models for higher order transition probabilities of Markov chain for analyzing the occurrences of daily rainfall data. *J Mod Appl Stat Methods* 10(1):337–348
39. Sinha NC, Islam MA, Ahmed KS (2010) Order determination for the transition probabilities Markov chain based on logistic regression model. *J Appl Stat Sci* 17(4):519–540
40. Sinha NC, Islam MA, Ahmed KS (2006) Chain dependence and stationarity test for transition probabilities of Markov chain under logistic regression model. *J Korean Stat Soc* 35(4):355–376
41. Stern RD, Coe R (1984) A model fitting analysis of daily rainfall data. *J R Stat Soc Ser A* 147:1–34
42. Srikanthan R, McMahon TA (2001) Stochastic generation of annual, monthly and daily climate data: a review. *Hydrol Earth Syst Sci* 5(4):653–670
43. Thyer MA, Kuczera G (2000) Modelling long-term persistence in hydro-climatic time series using a hidden state Markov model. *Water Resour Res* 36:3301–3310
44. Thyer MA, Kuczera G (1999) Modelling long-term persistence in rainfall time series, Sydney rainfall case study. *Hydrol Water Resour Symp Instit Eng Aust*. pp 550–555.
45. Vivekanandan N (2016) Intercomparison of probability distributions for extreme value analysis of rainfall under missing data scenario. *Develop Earth Sci* 4:1–7. <https://doi.org/10.14355/des.2016.04.001>
46. Wedel M, Böckenholt U, Kamakura WA (2003) Factor models for multivariate count data. *J Multivar Anal* 87:356–369
47. Winkelmann R (2008) Econometric analysis of count data, 5th edn. Springer, New York
48. Winkelmann R, Zimmermann KF (1995) Recent developments in count data modeling: theory and application. *J Econ Surv* 9(1):1–24
49. Winkelmann R (1994) Count data models-econometric theory and an application to labour mobility. *Lecture Notes on Economics Mathematics System* vol. 410, Heidelberg, New York: Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.