**Far East Journal of Mathematical Sciences (FJMS)**

Published Online: September 2016
http://dx.doi.org/10.17654/MS100081215
Volume 100, Number 8, 2016, Pages 1215-1225 **ISSN: 0972-0871**

# DEVELOPING GENERALIZED LINEAR MODEL OF GAMMA-PARETO DISTRIBUTION

**Herlina Hanum[1], Aji Hamim Wigena[2], Anik Djuraidah[2] and I Wayan Mangku[3]**

[1]Department of Mathematics
Sriwijaya University
Indonesia
e-mail: linhanum@yahoo.com.au

[2]Department of Statistics
Bogor Agricultural University
Indonesia

[3]Department of Mathematics
Bogor Agricultural University
Indonesia

## Abstract

This paper discusses about developing generalized linear model (GLM) of Gamma-Pareto (G-P) distribution. It is shown that G-P is a member of exponential family distribution. Consequently, we can develop GLM based on G-P. With $\log(y/\min(y))$ as response variable, the GLM has inverse, identity, and log link functions. Dispersion parameter, deviance, and Akaike information criteria (AIC) of the model are derived in this paper. We also provide the iterative weighted least squares (IWLS) for estimating the model's parameters.

Received: May 15, 2016; Accepted: July 4, 2016
2010 Mathematics Subject Classification: 62J12.
Keywords and phrases: GLM, Gamma-Pareto, link function, IWLS.
Communicated by K. K. Azad

## 1. Introduction

Recently, some new distributions were developed. One of them is G-P distribution developed by Alzaatreh et al. [1]. Some applications of G-P were shown in Alzaatreh et al. [1]. Hanum et al. [5] used G-P to model and forecast extreme monthly rainfall.

The applications of G-P in those two papers used only one variable. Meanwhile, the variable may be affected by other variable(s). To explain the relationship, we need regression model based on G-P distribution. For non-normal response variable, the regression model is usually in the form of GLM. This paper discusses the development of GLM based on G-P distribution.

The objectives of this paper are: (1) to show that G-P is the member of exponential family as a condition to develop GLM based on G-P distribution, (2) to provide some attributes for developing GLM such as mean, variance, and dispersion parameters of response variable, (3) to define the link functions of GLM G-P, (4) to provide deviance and AIC for assessing the goodness of GLM G-P, and (5) to provide parameter estimation of GLM G-P.

## 2. Gamma-Pareto Distribution as a Member of Exponential Family

GLM is based on a distribution which belongs to a member of exponential family. To develop GLM G-P, we have to show that G-P is a member of exponential family distribution.

**Theorem 1.** *Gamma-Pareto distribution belongs to exponential family distribution.*

**Proof.** According to Alzaatreh et al. [1], the pdf of G-P distribution is

$$g(y) = \frac{\theta^{-1}}{\varrho^{\alpha}\Gamma(\alpha)}\left(\log\left(\frac{y}{\theta}\right)\right)^{\alpha-1}\left(\frac{y}{\theta}\right)^{(-(1/\varrho)-1)} \tag{1}$$

with $\alpha, \varrho, \theta > 0$ and $y > \theta$. In order to show that G-P belongs to

exponential family, the pdf of G-P should be able to be written in exponential form. Rewriting equation (1) in exponential form yields

$$g(y) = \exp\left\{\log\left(\frac{\theta^{-1}}{\varrho^{\alpha}\Gamma(\alpha)}\right) + \log\left(\log\left(\frac{y}{\theta}\right)\right)^{\alpha-1} + \log\left(\left(\frac{y}{\theta}\right)^{(-(1/\varrho)-1)}\right)\right\}$$

$$= \exp\left\{-\alpha\log\varrho - \log\Gamma(\alpha) + \log\theta^{-1}\right.$$

$$\left. + (\alpha-1)\log\left(\log\left(\frac{y}{\theta}\right)\right) + (-\varrho^{-1}-1)\log\left(\frac{y}{\theta}\right)\right\}$$

$$= \exp\left\{-\alpha\log\varrho - \log\Gamma(\alpha) + \log\theta^{-1}\right.$$

$$\left. + (\alpha-1)\log\left(\log\left(\frac{y}{\theta}\right)\right) - \varrho^{-1}\log\left(\frac{y}{\theta}\right) - \log\left(\frac{y}{\theta}\right)\right\}.$$

Taking $\alpha$ and $\theta$ as nuisance parameters, G-P's pdf can be written as

$$g(y) = \exp\left\{-\varrho^{-1}\log\left(\frac{y}{\theta}\right) - \alpha\log\varrho - \left[\log y\Gamma(\alpha) + (\alpha-1)\log\left(\log\left(\frac{y}{\theta}\right)\right)\right]\right\}. \quad (2)$$

Equation (2) confirms the pdf exponential family distribution in Dobson [4], that is,

$$g(y:\tau) = \exp\{a(y)b(\tau) + c(\tau) + d(y)\}. \quad (3)$$

In equation (2), $a(y) = \log\left(\frac{y}{\theta}\right)$, $b(\tau) = \varrho^{-1}$, $c(\tau) = \alpha\log\varrho$, and

$$d(y) = \left[\log y\Gamma(\alpha) + (\alpha-1)\log\left(\log\left(\frac{y}{\theta}\right)\right)\right]. \quad \square$$

## 3. Mean, Variance and Dispersion Parameters of $\log\left(\frac{y}{\theta}\right)$ in G-P Distribution

Mean, variance and dispersion parameters are important parameters

in developing GLM. These parameters are determined based on $a(y)$ in equation (3).

**Lemma 1.** *The* $\log\left(\frac{y}{\theta}\right)$ *in G-P distribution has mean and variance equal to* $E(a(y)) = \alpha\varrho$ *and* $var(a(y)) = \alpha\varrho^2$, *respectively*.

**Proof.** Dobson [4] noted that for *Y* which has pdf in the form of equation (2), the mean and variance are $E(a(y)) = \frac{-c'(\tau)}{b'(\tau)}$ and

$$var(a(y)) = \frac{b''(\tau)c'(\tau) - c''(\tau)b'(\tau)}{(b'(\tau))^3},$$

respectively. The first and second derivatives of $b(\tau)$ and $c(\tau)$, respectively, are $b'(\tau) = \varrho^{-2}$, $b''(\tau) = -\varrho^{-3}$, $c'(\tau) = -\alpha\varrho^{-1}$ and $c''(\tau) = \alpha\varrho^{-2}$. So the mean and variance of G-P with $a(y) = \log\left(\frac{y}{\theta}\right)$ are $E(a(y)) = \alpha\varrho$ and $var(a(y)) = \alpha\varrho^2$. □

**Lemma 2.** *The dispersion parameter for GLM-GP is* $\varphi = \frac{1}{\alpha}$.

**Proof.** From equation (2) with $a(y) = \log\left(\frac{y}{\theta}\right)$, $var(a(y))$ is

$$var(a(y)) = \alpha\varrho^2 = (1/\alpha)[E(a(y))]^2 = \varphi[E(a(y))]^2. \quad \square$$

The dispersion parameter is estimated by

$$\hat{\varphi} = \sum_{i=1}^{N} \left( \frac{\left(\log\left(\frac{y_i}{\theta}\right)\right) - \hat{\mu}_i}{\hat{\mu}_i} \right)^2 /(n - p) = \chi^2/(n - p).$$

## 4. GLM of Gamma-Pareto Distribution

GLM with link function $g$ is $g(\mu_i) = X_i^T\beta = \eta_i$, where $\mu_i = E(a(y_i))$.

GLM of certain distribution may have some link functions. The link function may be based on the pdf or another property of the distribution. In order to obtain the link function from pdf, first we reparameterized the pdf with $\mu$, the mean of $a(y)$. The link function is $b(\mu)$ in the form of (3).

**Lemma 3.** *GLM G-P has inverse, identity and log link functions.*

**Proof.** Since $E(a(y)) = \mu = \alpha\varrho$, we take $\mu = \alpha\varrho$ in order to reform $g(y)$ in $\mu$. Then we obtain

$$\begin{aligned} g(y) &= \exp\left\{\alpha\left(\frac{-1}{\mu}\right)\log\left(\frac{y}{\theta}\right) - \alpha\log\left(\frac{\mu}{\alpha}\right)\right. \\ &\quad \left. - \left[\log(y\Gamma(\alpha)) + (\alpha-1)\log\left(\log\left(\frac{y}{\theta}\right)\right)\right]\right\} \\ &= \exp\left\{\alpha\left[\left(\frac{-1}{\mu}\right)\log\left(\frac{y}{\theta}\right) - \log(\mu)\right]\right. \\ &\quad \left. - \left[\alpha\log(\alpha) + \log(y\Gamma(\alpha)) + (\alpha-1)\log\left(\log\left(\frac{y}{\theta}\right)\right)\right]\right\}. \end{aligned}$$

Taking $\varphi = \frac{1}{\alpha}$, pdf G-P becomes

$$g(y) = \exp\left\{\frac{\left[\left(\frac{-1}{\mu}\right)\log\left(\frac{y}{\theta}\right) - \log(\mu)\right]}{\varphi}\right.$$

$$\left. - \left[\frac{1}{\varphi}\log\left(\frac{1}{\varphi}\right) + \log\left(y\Gamma\left(\frac{1}{\varphi}\right)\right) + \left(\frac{1-\varphi}{\varphi}\right)\log\left(\log\left(\frac{y}{\theta}\right)\right)\right]\right\}. \quad (4)$$

Based on (4), we may use link function $\left(\frac{-1}{\mu}\right)$ with the model

$$g(\mu) = -\frac{1}{\mu} = X'\beta \text{ or } \frac{1}{\mu} = X'(-\beta) \text{ or } \frac{1}{\mu} = X'(\beta^*) \text{ or } \mu = \frac{1}{X'(\beta^*)}.$$

According to McCullagh and Nelder [6], the simple model for inverse link function is

$$\mu = \frac{x_i}{b_1 + b_0 x_i} \text{ or } \frac{1}{\mu} = b_0 + b_1/x_i.$$

Identity link function is used as the inverse of GLM with inverse link function. The GLM with identity link function is $\mu = b_0 + b_1/x_i$ (Balajari [2]).

Log link function is used based on the relationship between variance and mean of $a(y)$. It is showed that $var(a(y)) = \mu^2\phi$. For each observation, the variance is $var(a(y_i)) = \phi\mu_i^2$. According to McCullagh and Nelder [6], the transformation that can stabilize the variance, when $var(y_i) = \phi\mu_i^2$, is logarithmic. Using $z_i = \log(y_i)$, we obtain $var(z_i) = \phi$ which is constant. GLM G-P with log link function is

$$g(\mu) = \log(\mu) = X'\beta \text{ or } \mu = e^{X'\beta}. \qquad \square$$

## 5. Deviance and Akaike Information Criteria

In order to assess the goodness of the model, we need some criteria. The criteria are also used to compare some models. For GLM, deviance and AIC are usually used. Both are based on the likelihood function of response variable. Furthermore, for comparing models, a model with less AIC and/or deviance is better.

The log likelihood for $i$th observation in G-P is the log of equation (2) for $i = 1, 2, 3, ..., N$,

$$l(\mu_i : y_i) = \alpha\left[-\frac{1}{\mu_i}\log\left(\frac{y_i}{\theta}\right) - \log\mu_i\right]$$

$$-\left[\log y_i\Gamma(\alpha) + (\alpha - 1)\log\left(\log\left(\frac{y_i}{\theta}\right)\right)\right].$$

Then for all observations, the log likelihood function is

$$l(\mu) = \sum_{i=1}^{N} \alpha\left[-\frac{1}{\mu}\log\left(\frac{y_i}{\theta}\right) - \log\mu\right]$$

$$- \sum_{i=1}^{N}\left[\log y_i\Gamma(\alpha) + (\alpha-1)\log\left(\log\left(\frac{y_i}{\theta}\right)\right)\right]. \quad (5)$$

**Lemma 4.** *The deviance and AIC of GLM G-P are*

$$D(y, \hat{\mu}) = -2\alpha\sum_{i=1}^{N}\left\{\log\left(\left(\log\left(\frac{y_i}{\theta}\right)\right)\Big/\hat{\mu}_i\right) - \left(\frac{\left(\log\left(\frac{y_i}{\theta}\right)\right) - \hat{\mu}_i}{\hat{\mu}_i}\right)\right\},$$

$$AIC = 2\alpha\sum_{i=1}^{N}\left(\frac{\log\left(\frac{y_i}{\theta}\right)}{\mu_i} + \log\mu_i\right) + 2p.$$

**Proof.** Based on (5), we obtain the log likelihood function of G-P $(\alpha, \mu, \theta)$ for known $\alpha$ and $\theta$ is proportional to

$$l(y_i) = \sum_{i=1}^{N}\alpha\left(-\frac{\log\left(\frac{y_i}{\theta}\right)}{\mu_i} - \log\mu_i\right).$$

This log likelihood function achieves its maximum at $\mu = \log\left(\frac{y_i}{\theta}\right)$ with the value

$$\alpha\sum_{i=1}^{N}\left(-\frac{\log\left(\frac{y_i}{\theta}\right)}{\log\left(\frac{y_i}{\theta}\right)} - \log\left(\log\left(\frac{y_i}{\theta}\right)\right)\right) = -\alpha\sum_{i=1}^{N}\left(1 + \log\left(\log\left(\frac{y_i}{\theta}\right)\right)\right).$$

The deviance is proportional to 2 times of difference between the log likelihood of the model and maximum value of log likelihood (McCullagh

and Nelder [6]). So the deviance for GLM G-P is

$$D(y,\,\hat{\mu}) = 2\left\{\sum_{i=1}^{N}\alpha\left(-\frac{\log\left(\frac{y_i}{\theta}\right)}{\hat{\mu}_i} - \log\hat{\mu}_i\right) + \alpha\sum_{i=1}^{N}\left(1 + \log\left(\log\left(\frac{y_i}{\theta}\right)\right)\right)\right\}$$

$$= 2\alpha\sum_{i=1}^{N}\left\{\left(-\frac{\log\left(\frac{y_i}{\theta}\right)}{\hat{\mu}_i} - \log\hat{\mu}_i\right) + \left(1 + \log\left(\log\left(\frac{y_i}{\theta}\right)\right)\right)\right\}$$

$$= 2\alpha\sum_{i=1}^{N}\left\{\left(-\log\hat{\mu}_i + \log\left(\log\left(\frac{y_i}{\theta}\right)\right)\right) + \left(1 - \frac{\log\left(\frac{y_i}{\theta}\right)}{\hat{\mu}_i}\right)\right\}$$

$$= -2\alpha\sum_{i=1}^{N}\left\{\left(\log\left(\log\left(\frac{y_i}{\theta}\right)\right)\Big/\hat{\mu}_i\right) + \left(\frac{\log\left(\frac{y_i}{\theta}\right) - \hat{\mu}_i}{\hat{\mu}_i}\right)\right\}.$$

AIC takes the form $AIC = -2\log(\cdot\,|\,y) + 2p$ (Burnham and Anderson [3]). Using the proportional of likelihood function in (5), AIC for GLM G-P is

$$AIC = -2\sum_{i=1}^{N}\alpha\left(-\frac{\log\left(\frac{y_i}{\theta}\right)}{\mu_i} - \log\mu_i\right) + 2p$$

$$= 2\alpha\sum_{i=1}^{N}\left(\frac{\log\left(\frac{y_i}{\theta}\right)}{\mu_i} + \log\mu_i\right) + 2p. \qquad \square$$

## 6. Parameter Estimation of GLM Gamma-Pareto

Estimation of parameter $\beta_j$ using maximum likelihood is done by first determining the derivative of likelihood function with respect to $\beta_j$

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N}\left[\frac{\partial l_i}{\partial \beta_j}\right] = \sum_{i=1}^{N}\left[\frac{\partial l_i}{\partial \tau_i}\frac{\partial \tau_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j}\right] \text{ with}$$

$$\frac{\partial l_i}{\partial \tau_i} = a(y)b'(\tau) + c'(\tau) = \varrho^{-2}\left(\log\left(\frac{y_i}{\theta}\right) - \mu_i\right),$$

$$\frac{\partial \tau_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \tau_i}} = \frac{1}{\frac{\partial \alpha\varrho}{\partial \varrho}} = \frac{1}{\alpha},$$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_j}x_{ij},$$

where $\frac{\partial \mu_i}{\partial \eta_j}$ depends on the link function of the GLM. So the score for $\beta_j$ in GLM Gamma-Pareto is

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N}\alpha^{-1}\varrho^{-2}\left(\log\left(\frac{y_i}{\theta}\right) - \mu_i\right)\frac{\partial \mu_i}{\partial \eta_j}x_{ij}.$$

Finally, we have the *j*th score

$$U_j = \sum_{i=1}^{N}\left[var\left(\log\left(\frac{y_i}{\theta}\right)\right)\right]^{-1}\left[\log\left(\frac{y_i}{\theta}\right) - \mu_i\right]x_{ij}\frac{\partial \mu_i}{\partial \eta_i}.$$

The variance of $U_j$ is

$$var(U_j) = \Im_{jk} = \sum_{i=1}^{N}\frac{x_{ij}x_{ik}}{\left[var\left(\log\left(\frac{y_i}{\theta}\right)\right)\right]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = X^T W X,$$

where

$$W = \frac{1}{\left[var\left(\log\left(\frac{y_i}{\theta}\right)\right)\right]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.$$

Since the estimator of $\beta_j$ is not in close form, Dobson [4] suggested the iterative weighted least squares (IWLS) for estimating $\beta_j$. The IWLS is

$$X^T W X b^{(m)} = X^T W z.$$

Using $W$ and $var(U_j)$ for G-P, we obtain the iteration for $\beta_j$ as

$$X^T W X b^{(m)} = \sum_{k=1}^{p} \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{\left[ var\left( \log\left( \frac{y_i}{\theta} \right) \right) \right]} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 b_k^{(m-1)}$$

$$+ \sum_{i=1}^{N} \frac{\left( \log\left( \frac{y_i}{\theta} \right) - \mu_i \right) x_{ij}}{\left[ var\left( \log\left( \frac{y_i}{\theta} \right) \right) \right]} \left( \frac{\partial \mu_i}{\partial \eta_i} \right),$$

where

$$z_i = \sum_{i=1}^{N} x_{ik} b_k^{(m-1)} + \left( \log\left( \frac{y_i}{\theta} \right) - \mu_i \right) \left( \frac{\partial \mu_i}{\partial \eta_i} \right).$$

## 7. GLM G-P and GLM Gamma

Alzaatreh et al. [1] mentioned mathematical relationship between G-P and gamma. GLM G-P may also have relationship with GLM gamma.

**Theorem 2.** *GLM G-P with response variable* $\log\left( \frac{y_i}{\theta} \right)$ *is similar to GLM gamma with response variable* $u = \log\left( \frac{y_i}{\theta} \right)$.

**Proof.** Lemma 1 to Lemma 4 show the similarity. □

## 8. Conclusions

Regression modeling based on G-P distribution can be developed in the form of GLM. GLM G-P may use inverse, identity, and log link functions.

Parameters estimation in GLM G-P may use iterative weighted least squares. GLM G-P is similar to GLM gamma.

## References

[1] A. Alzaatreh, F. Famoye and C. Lee, Gamma-Pareto distribution and its application, Journal of Modern Applied Statistical Methods 11(1) (2012), 78-94, Article 7.

[2] Balajari, GLM with a gamma-distributed dependent variable, 2013.

http://civil.colorado.edu/~balajir/CVEN6833/lectures/GammaGLM-01.pdf.

[3] K. P. Burnham and D. R. Anderson, Multinomial inference, understanding AIC and BIC in model selection, Sociol. Methods Res. 33(2) (2004), 261-304. DOI:10.1177/0049124104268644/2004.

[4] A. J. Dobson, An Introduction to Generalized Linear Models, Second Edition, Chapman & Hall/CRC, Florida, 2002.

[5] H. Hanum, A. H. Wigena, A. Djuraidah and I. W. Mangku, Fitting extreme rainfall with Gamma-Pareto distribution, Appl. Math. Sci. 9(121) (2015), 6029-6039. http://dx.doi.org/10.12988/ams.2015.57489.

[6] P. McCullagh and J. A. Nelder, Generalized Linear Models, Second Edition, Chapman & Hall, London, 1989.