Available online www.unicrossjournals.com

JOURNAL OF CONTEMPORARY RESEARCH (JOCRES)

RESEARCH ARTICLE VOL. 1 (2) ISSN:2814-2241

Date Accepted: 30th Dec, 2022

Pages 139-160

ON GENERALIZED LINEAR MODEL (GLM) AND CLASSIFICATION OF STATISTICAL DISTRIBUTIONS

[1]Usen, John Effiong, [1]Egah, Friday Ogar

[1]Department of Statistics, University of Cross River State, Calabar
Corresponding author's e-mail: usenjohn@unicross.edu.ng

Abstract

This study on generalized linear model (GLM) aimed to classify a group of unclassified statistical distributions, namely, the: error, Gumbel, logistic, Pareto, and power function distributions into what is regarded as the exponential family of distributions. Conscious of how probability distributions (or density) functions are classified into this family, we were able to show that, indeed, the listed distributions all belong to the exponential family since, for a given parameter of interest θ it was possible to write their probability functions $f(x)$ in the form $\exp\{a(x)b(\theta)+c(\theta)+d(x)\}$. More so, in this study, the linear predictors of all distributions were derived, and observed not to be in canonical form as the term $a(x)\neq x$; their corresponding link functions were also deduced. These findings were postulated as theorems, with the results summarized in tables. However, our results also showed that when put in exponential forms, the error, Pareto and power function distributions all had $d(x)=0$; whereas, the Gumbel and logistic distributions had $d(x)\neq 0$. Finally, based on the study so far, it was suggested that scholars should explore the possibility of classifying other density or distribution functions not considered in this study, as this may lead to deeper practical and theoretical results. It was also suggested that the maximum likelihood estimates of the newly classified distributions should be obtained as this could make a new research direction to an interested scholar.

Key words: Linear Predictors, Link Functions, Exponential Family

1. Introduction

In statistics, and in experimental designs in particular, the concept of generalized linear models (GLMs) has become indispensably relevant by virtue of being a flexible generalization of the ordinary linear regression (OLR) in the sense that it accommodates response variables which have error distribution models other than a normal distribution (otherwise regarded as non-normal response variables). The GLM generalizes OLR by allowing the linear regression model to be related to the response variable via a link function, and also by allowing the magnitude of the variance of each measurement to be a function of the expected value (Annette, 2002; Altham, 2011; Montgomery, 2013).

Before the formulation of the GLM, various transformation techniques were used to tackle the problem of non-normal response variables. The most common and efficient transformation technique among such lot was the Box-Cox technique – a technique which showed how the parameter λ of the transformation could be obtained from the "power family" of transformation $y^* = y^\lambda$ (Cochran, 1992; Cox & Reid, 2000; Oehlert, 2010; Montgomery, 2013). In 1964, G. E. P. Box and G. M. Cox explained how the transformation parameter λ may be estimated simultaneously with the other parameters of the model such as: the overall mean and the treatment effects (Cox & Reid, 2000; Montgomery, 2013). The theory underlying their technique used the method of maximum likelihood, whereas the main computational steps consists of performing (for various values of λ) a standard analysis of variance on

$$y^{(\lambda)} = \begin{cases} \dfrac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y} \ln y & \lambda = 0 \end{cases}$$

where $\dot{y} = \ln^{-1}\left[\frac{\sum \ln y}{n}\right]$ is the geometric mean of the observations.

Although Box-Cox technique has remained one of the easy and efficient means for selecting the form of the transformation, recent studies have shown that the use of data transformation techniques could pose several difficulties. One difficulty is that most experimenters may be uncomfortable working with responses in the transformed scale (Cochran, 1992; Cox & Reid, 2000; Oehlert, 2010). Whereas a more serious difficulty is that a transformation can result in a nonsensical value for the response variable over some portion of the design factor space that is of interest to the experimenter (Cochran, 1992; Montgomery, 2013). A third difficulty is that there is no assurance that a transformation will effectively attain all of the objectives of transformation of a non-normal response variable simultaneously (Montgomery, 2013).

It was with a view to overcoming difficulties encountered using transformations, that GLM was formulated in 1972 by John Nelder and Robert Wedderburn as a technique which essentially unifies linear and nonlinear statistical models (including linear, logistic, and Poisson regressions) with both normal and non-normal responses. To achieve this, they proposed an iteratively reweighted least squares (IRLS) technique for maximum likelihood estimation (MLE) of the model parameters (McCullagh & Nelder, 1989; Lindsey, 1997; Heather, 2008; Marlene, 2004; Farzana, Borhan, & Ataharul, 2016).

The development of certain GLMs to obtain link functions, linear predictors, and maximum likelihood estimates of parameters, with observations which were distributed according to known distributions in the exponential family, had been established. So far, notable distributions in the exponential family are: Bernoulli, beta, binomial, chi-squared, Dirichlet, exponential, gamma, geometric, inverse Gaussian (Wald), logarithmic series, lognormal, multinomial, multi-normal, negative binomial, Gaussian, Poisson, Rayleigh, von Mises, Weibull, and Wishart distributions (McCullagh & Nelder, 1989; Lindsey, 1997; Lindsey & Jones, 1998; Rodriguez, 2001; Heather, 2008; Marlene, 2004; Mathew, Yuan, Yichong & Silun, 2016; Farzana, et al., 2016). This study gave a GLM-based approach to obtaining the linear predictors and link functions based on other distributions not yet classified as members of the exponential family.

In recent years, since its formulation in 1972, there have been many scholarly publications on the GLM, ranging from publications that describe general aspects of the GLM (e.g. Rodriguez, 2001; John, Patricia, Aaron & Jeffrey, 2007; Khuri, 2010; Murtala, Udokang, Raji & Bello, 2015, David, 2018, etc.), to real-life applications of the GLM (e.g. Lindsey, 1997; Jun, 2011; Mark, Anand & Dan, 2016; Moffat & Emmanuel, 2018, etc.), and then theoretical developments based on the GLM (e.g. Michael, 2003; Sourish, 2008; Sourish & Dipak, 2014, etc.). Such works were restricted to theories and applications based on different types of data from the binomial, Poisson and exponential distributions. Sadly, theoretical extensions and applications of the GLM based on data types from the beta, chi-squared, negative binomial, Dirichlet, gamma, geometric, logarithmic series, lognormal, multinomial, multi-normal, Gaussian, Rayleigh, von Mises, Weibull, and Wishart distributions appear to be noticeably scanty, if there are. Most popular among very recent publications is the work of Farzana, et al. (2016) which attempted a derivation of the GLM for the geometric distribution by showing its detailed estimation and test procedure with a view to discussing the possible challenges faced when applied to real life data. A major gap noticed from all these studies taken into consideration is that none of such studies have attempted to derive the GLM for other distributions not yet classified as being exponential; and even when the form of the link functions and the deviances can be stated for some of such distributions, no application had been shown yet. This study was an attempt to overcome a part of this gap. In this regard, the objectives were: (i) to show whether or not, each unclassified distribution were members of the exponential family of distributions, (ii) to obtain the linear predictors for the unclassified distributions via stated theorems, and (iii) to deduce the link functions for the unclassified distributions.

2. Empirical review of literature

Moffat and Emmanuel (2018) identified the problem associated with Ordinary Least Squares (OLS) in relation to the violation of assumptions of normality and constant variance. Mainly, the possible problem encountered when these assumptions are violated is the introduction of biases in the parameters of the fitted model thereby threatening the model's efficiency. The authors applied GLM to overcome such problems and to ensure the efficiency of the model parameters. The major reason they used GLM was that the GLM does not require transformation and assumptions of classical regression; instead, it employs a probabilistic approach in transforming the expected value of the dependent variable. The data used were obtained from the Central Bank of Nigeria Statistical Bulletin from 1981 to 2016, with each series consisting of 36 observations. The Gross Domestic Product (N' Billion) was considered as the dependent variable (Y_t) while Money Supply (X_{1t}), and Credit to Private Sector (X_{2t}) were considered as the independent variables (N' Billion). From the analysis, the result of the fitted regression model showed no significant relationship between the variables. The diagnosis on the residual series (using skewness, kurtosis, Jacque-Bera test and Breusch-Pagan-Godfrey test) provided sufficient evidence that both validity and efficiency of the model parameters are threatened. However, the results of the GLM procedure provided the much needed significance, validity, and efficiency of the model parameters. Further findings from GLM procedure revealed that the standard errors of the parameters of OLS were biased having been far larger in values than those of the GLM. Hence, for studies involving the regression of a discrete-time stochastic series such as GDP on Money Supply and Credit to Private Sector, the GLM was adjudged analytically tractable than the OLS.

Farzana, Borhan and Ataharul (2016) derived the GLM for the geometric distribution, estimation of parameters, and test procedures. An application was made to Bangladesh Demographic and Health Survey data to find the significant factors associated with the first occurrence of infant death in terms of birth order. Two different generalized linear models were fitted; one using the natural link function and the other one using the log link function. At the end, the results of both models were compared. It was found that the model fitted using the log link function had lower Akiake's information criteria (AIC) and deviance than the model fitted using the natural link function. This meant that the GLM for the geometric distribution using the log link function provided better result.

Alexander and Wolfgang (2015) questioned whether methods of the GLM, that is repeated measures of ANOVA and regression, could be used to estimate individual-specific parameters. Scenarios and corresponding design matrices were presented in which the shape of temporal trajectories of individuals is parameterized. Real world data examples and simulation results suggested that, for series of sufficient length, trajectories could be well described for individuals. In addition, scenarios were presented for the comparison of two individuals. Here again, trajectories could be well described and the statistical comparison of individuals were possible. However, in contrast to the power for the description of individual series, which was satisfactory, the power for the comparison of individuals was low (except when effect sizes were large). In all simulated scenarios, the power of tests increased only up to a certain number of observation points, and reached a ceiling at this number. The fact that all parameters could not always be estimated was also discussed, and options were presented that go beyond what standard general purpose software packages offer.

Muritala, et al. (2015) carried out an empirical study of GLM for count data. The authors noted, in particular, that the Poisson regression model which is known to be a generalized linear model for the Poisson error structure had been widely used in recent years; it was also used in modeling of count and frequency data. Quasi Poisson model was used for handling over-dispersion since the data was found to be over-dispersed, while the negative binomial regression model was used for handling over-dispersion. In this study, the two regression models were compared using the AIC, the model with minimum AIC showed the best which implied the Poisson regression model.

Sourish and Dipak (2014) obtained an estimator of the regression parameters for generalized linear models, using the Jacobian technique. They restricted themselves to the natural exponential family for the response variable and chose the conjugate prior for the natural parameter. Using the Jacobian of transformation, they obtained the posterior distribution for the canonical link function, and thereby obtained the posterior mode for the link. Under the full rank assumption for the covariate matrix, they then found an estimator for the regression parameters for the natural exponential family. The proposed estimator was specially derived for the Poisson model with logit link function. More so, the authors discussed extensions to the binomial response model when covariates were all positive. Finally, an illustrative real-life example was given for the Poisson model with log link. In order to estimate the standard error of their estimates, they used the Bemstein-von Mises theorem, and finally compared the results using their Jacobian technique with a maximum likelihood estimate for the regression parameters.

Jun (2011) studied the limited fluctuations credibility of the GLM estimators as well as in the extended case of GLMMs. The study showed how credibility depended on the sample size, the distribution of covariates and the link function. The study gave criteria for full credibility of the GLM estimators. This provided a mechanism to obtain confidence intervals for the GLM and GLMM estimators. If the full credibility criteria could not be satisfied, it was interesting to calculate the partial credibility matrix and the GLM estimators. Here, for a general link function the credibility matrix is not known explicitly. Under certain assumptions, numerical methods were developed to compute the GLM estimators and the credibility matrix. For some specific link functions, further properties were developed. For instance, Hachemeister's credibility regression model was one of such cases of his model, where the link function was linear. Loss reserving was a major challenge for casualty actuaries due to the frequently changing business environments. The author remarked that some aggregate loss reserving models had been extended to or developed by research actuaries within the framework of GLMs in recent times. The study therefore established a structural loss reserving model which combines the exposure and loss emergence patterns and the loss development pattern, again within the framework of a GLM. Discounted loss reserves could also be obtained from this model.

Heather (2008) presented an overview of generalized linear models (GLMs) which showed that the models extended the linear modeling framework to variables that were not normally distributed. Heather (2008) focused this overview on GLMs based on binary or count data because they were the most commonly used types.

Sourish (2008) developed an innovative approach to analyze the scientific studies using the GLM and beyond. In particular, the study developed the regression estimator, a new algorithm for fitting GLM and different model diagnostic technique for GLM. In the context of the longitudinal study, the study presented the Bayesian analysis of the generalized multivariate gamma distribution for the generalized multivariate analysis of variance (GMANOVA) model. The study demonstrated the method for modeling longitudinal studies as state space dynamic model, and this was accomplished using an introduction the power filter for dynamic generalized linear models (DGLMs). An information-processing optimality property of the power filter was presented, and with it the author established the relationship between the Kalman filter and the power filter as well. Sourish (2008) developed the Pareto regression model for analyzing the extreme drinking behavior of the alcohol dependence disorder patients.

John, Patricia, Aaron and Jeffrey (2007) compared GLMs and OLS in predicting individual patient costs in adults' intensive care units (ICUs) and sought to define the utility of the inverse Gaussian distribution family within

GLMs. A prospective "ground-up" utilization costing study was performed in three adult university associated ICUs, enrolling consecutive ICU admissions over a 6-month period in 1991. ICU utilization, patient demographic and ICU admission day data were recorded by dedicated data collectors. Model performance was assessed by prediction error (mean absolute error – MAE), root mean squared error – RMSE, and residual analysis. The cohort, 1098 patients surviving ICU, was of mean (SD) age 56 (19.5) years and 41 percent female. Patient costs per ICU episode (1991 AS) were A56311 with A595602. Prediction error for mean costs was minimal (MAE 4780; RMSE 8965) with OLS using heteroscedastic retransformation of log costs and GLM with Gaussian family and log link (MAE 4798; RMSE 8907). Residual analysis suggested optimal overall performance for the above two models and a GLM with inverse Gaussian family and log link. The authors concluded that traditional cost models of OLS with (log) transformation may be supplemented by approximately specified GLM which more closely model the error structure.

According to Lindsey and Jones (1998), when testing effect or a difference among groups, the distributional assumptions made about the response variable may have critical impact on the conclusions drawn since controversy could arise over transformations of the response. Hence, an alternative approach was to use some member of the family of generalized linear models. However, this raised the issue of selecting the appropriate member, a problem of testing non-nested hypotheses. Standard model selection criteria (e.g. AIC) were proposed by the authors to be used to resolve problems. These procedures for comparing generalized linear models were applied to checking for difference in T4 cell counts between two disease groups. The authors concluded that appropriate model selection criteria should be specified in the protocol for any study, including clinical trials, in order for optimal inferences to be drawn about treatment differences.

3. Materials and method

3.1. The exponential family of distributions

Most of the commonly used statistical distributions are members of what is called "the exponential family". This family is a very rich and flexible collection of distributions applied in many experimental situations. At present, the family includes: Bernoulli, beta, binomial, chi-squared, Dirichlet, exponential, gamma, geometric, logarithmic series, lognormal, multinomial, multi-normal, negative binomial, Gaussian, Poisson, Rayleigh, von Mises, Weibull, and Wishart distributions (Christian, 2007; Penzer, 2011; Catherine, Mervan, Nicholas & Brian, 2011).

Let X be single random variable whose probability distribution function depends on a single parameter θ. The probability distribution function of X is said to belong to the exponential family if it can be written in the form

$$f(x;\theta) = s(x)t(\theta)e^{a(x)b(\theta)} \quad (2)$$

where a, b, s and t are known functions. Equation (2) may be rewritten as:

$$f(x;\theta) = \exp\{a(x)b(\theta) + c(\theta) + d(x)\} \quad (3)$$

Where
$s(x) = \exp d(x)$, and
$t(\theta) = \exp c(\theta)$.

If $a(x) = x$, the distribution is said to be in "canonical (that is, standard) form" and $b(\theta)$ is sometimes called the "natural parameter" of the distribution. However, if there are other parameters, in addition to the parameter of interest, θ, they are regarded as "nuisance parameters" forming parts of the functions a, b, c and d, and they are treated as though they are known. Many of the well-known distributions belong to the exponential family, and can all be written in the canonical form.

3.2. Generalized linear model (GLM)

The unity of many statistical methods was demonstrated in 1972 by Nelder and Wedderburn using the idea of a GLM. The GLM is defined in terms of a set on independent random variables $X_1, \ldots, X_N$ each with a distribution from the exponential family and the following properties:

i. The distribution of each X_i has the canonical form and depends on a single parameter θ_i (the θ_i's do not all have to be same), thus:

$$f(x_i;\theta_i) = \exp\{x_i b_i(\theta_i) + c_i(\theta_i) + d_i(x_i)\} \quad (4)$$

ii. The distributions of all the X_i's are of the same form (e.g., all Normal or all binomial) so that the subscripts on b, c and d are not needed. Thus, the joint probability density function of $X_1, \ldots, X_N$ is:

$$f(x_1,\ldots,x_N;\theta_1,\ldots,\theta_N) = \left\{ \begin{array}{l} \prod_{i=1}^{N} \exp\{x_i b_i(\theta_i) + c_i(\theta_i) + d_i(x_i)\} \\ \exp\left\{ \sum_{i=1}^{N} x_i b_i(\theta_i) + \sum_{i=1}^{N} c_i(\theta_i) + \sum_{i=1}^{N} d_i(x_i) \right\} \end{array} \right\} \quad (5)$$

The parameters θ_i are typically not of direct interest (since there may be one for each observation). For model specification we are usually interested in a smaller set of parameters $\beta_1, \ldots, \beta_p$ (where $p < N$).

Suppose that $\mu_i = E(Y_i)$ where μ_i is some function of θ_i. For a generalized linear model there is a transformation of μ_i such that:

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad (6)$$

In equation (6), g is a monotone differentiable function called the "link function"; $\mathbf{x}_i$ is a $p \times 1$ vector of explanatory variables (covariates and dummy variables for levels of factors),

$$\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$$
$$\Rightarrow \mathbf{x}_i^T = (x_{i1}, \ldots, x_{ip})$$

and $\boldsymbol{\beta}$ is the $p \times 1$ vector of parameter $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$. The vector $\mathbf{x}_i$ is the i^{th} column of the design matrix $\mathbf{X}$.

Thus, a generalized linear model has three components:

1. Response variables $Y_1, \ldots, Y_N$ which are assumed to share the same distribution from the exponential family;
2. A set of parameters $\boldsymbol{\beta}$ and explanatory variables

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{N1} & \cdots & x_{Np} \end{pmatrix}$$

3. A monotone link function g such that (where $\mu_i = E(Y_i)$)

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

3.3. Statistical distributions for classification

A total of five (5) distributions were considered for classification. They are: error, extreme value (Gumbel), logistic, Pareto, and power function distributions. Each of these distributions may be characterized by features such as: variate, range, location parameter, scale parameter, distribution function, probability density function, characteristic function, inverse distribution function, moments, cumulants, mode, median, etc. However, out of the features that characterize these distributions, the primary focus of this study was on the probability density function.

3.3.1. Error distribution

The error distribution is also known as the exponential power distribution or the general error distribution.

Range $-\infty < x < \infty$

Location parameter $-\infty < \alpha < \infty$

Scale parameter $\theta > 0$

Shape parameter $\beta > 0$

Alternative parameter $\lambda = 2/\beta$

Probability density function

$$\frac{\exp\left\{-\left(|x-\alpha|/\theta\right)^{\frac{2}{\beta}}/2\right\}}{\theta\left(2^{\frac{\beta}{2}+1}\right)\Gamma(1+\beta/2)}$$

Mean α

Median α

Mode α

rth Moment about the mean

$$\begin{cases} \theta^r\, 2^{r\beta/2}\,\dfrac{\Gamma((r+1)\beta/2)}{\Gamma(\beta/2)}, & r \text{ even} \\ 0, & r \text{ odd} \end{cases}$$

Variance $\dfrac{2^{\beta}\,\theta^2\,\Gamma(3\beta/2)}{\Gamma(\beta/2)}$

Mean deviation $\dfrac{2^{\beta/2}\,\theta\,\Gamma(\beta)}{\Gamma(\beta/2)}$

Coefficient of skewness 0

Coefficient of kurtosis $\dfrac{\Gamma(5\beta/2)\Gamma(\beta/2)}{[\Gamma(3\beta/2)]^2}$

3.3.2. Extreme value (Gumbel) distribution

The extreme value distribution was developed as the distribution of the largest of a number of values and was originally applied to the estimation of flood levels. It has since been applied to the estimation of the magnitude of earthquakes. The distribution may also be applied to the study of athletic and other records.

This study considers the distribution of the largest extreme; although a reversal of the sign of x gives the distribution of the smallest extreme. This is the Type I, the most common of the three extreme value distributions, known as the Gumbel distribution.

Variate $V : \alpha, \theta$

Range $-\infty < x < +\infty$

Location parameter α, the mode

Scale parameter $\theta > 0$

Distribution function $\exp\left\{-\exp\left[-\frac{(x-\alpha)}{\theta}\right]\right\}$

Probability density function

$$\left(\frac{1}{\theta}\right)\exp\left[-\frac{(x-\alpha)}{\theta}\right]\times\exp\left\{-\exp\left[-\frac{(x-\alpha)}{\theta}\right]\right\}$$

Inverse distribution function

$$\alpha-\theta\log\left[\log\left(\frac{1}{1-\mu}\right)\right]$$

Inverse survival function

$$\alpha-\theta\log\left[\log\left(\frac{1}{1-\mu}\right)\right]$$

Hazard function

$$\frac{\exp\left[-\frac{(x-\alpha)}{\theta}\right]}{\theta\left(\exp\left\{\exp\left[-\frac{(x-\alpha)}{\theta}\right]\right\}-1\right)}$$

Moment generating function

$$\exp(\alpha t)\Gamma(1-\theta t), \quad t < 1/\theta$$

Characteristic function

$$\exp(i\alpha t)\Gamma(1-i\theta t)$$

Mean $\alpha-\theta\Gamma'(1)$

$\Gamma'(1) = -0.57722$ is the first derivative of the $\Gamma'(n)$ with respect to n at $n = 1$

Variance $\frac{\theta^2\pi^2}{6}$

Coefficient of skewness 1.139547

Coefficient of kurtosis 5.4

Mode α

Median $\alpha-\theta\log(\log 2)$

3.3.3. Logistic distribution

The distribution function of the logistic is used as a model for growth. For example, with a new product we often find that growth is initially slow, then gains momentum, and finally slows down when the market is saturated or some form of equilibrium is reached.

Applications include the following:

- Market penetration of a new product
- Population growth
- The expansion of agricultural production
- Weight gain in animals

Range: $-\infty < x < \infty$

Location parameter: α , the mean

Scale parameter: $\theta > 0$

Alternative parameter $k = \frac{\pi\theta}{3^{\frac{1}{2}}}$, the standard deviation

Distribution function

$$1-\left[1+\exp\left(\frac{x-\alpha}{\theta}\right)\right]^{-1} = \left\{1+\exp\left[-\left(\frac{x-\alpha}{\theta}\right)\right]\right\}^{-1}$$

$$=\frac{1}{2}\left\{1+\tanh\left[\frac{1}{2}\left(\frac{x-\alpha}{\theta}\right)\right]\right\}$$

Probability density function

$$\frac{\exp\left[-\left(\frac{x-\alpha}{\theta}\right)\right]}{\theta\left\{1+\exp\left[-\left(\frac{x-\alpha}{\theta}\right)\right]\right\}^2}=\frac{\exp\left[\left(\frac{x-\alpha}{\theta}\right)\right]}{\theta\left[1+\exp\left(\frac{x-\alpha}{\theta}\right)\right]^2}$$

$$=\frac{\sec h^2\left(\frac{x-\alpha}{2\theta}\right)}{4\theta}$$

Inverse distribution function

$$\alpha+\theta\log\left(\frac{\mu}{1-\mu}\right)$$

Survival function

$$\left[1+\exp\left(\frac{x-\alpha}{\theta}\right)\right]^{-1}$$

Inverse survival function

$$\alpha+\theta\log\left(\frac{1-\mu}{\mu}\right)$$

Hazard function

$$\left\{\theta\left[1+\exp\left[-\left(\frac{x-\alpha}{\theta}\right)\right]\right]\right\}^{-1}$$

Cumulative hazard function

$$\log\left[1+\exp\left(\frac{x-\alpha}{\theta}\right)\right]$$

Moment generating function

$$\exp(\alpha t)\Gamma(1-\theta t)\Gamma(1+\theta t)=\frac{\pi\theta t\exp(\alpha t)}{\sin(\pi\theta t)}$$

Characteristic function

$$\frac{\exp(i\alpha t)\pi\theta it}{\sin(\pi\theta it)}$$

Mean	α
Variance	$\frac{\pi^2\theta^2}{3}$
Mode	α
Median	α
Coefficient of skewness	0
Coefficient of kurtosis	4.2
Coefficient of variation	$\frac{\pi\theta}{3^{\frac{1}{2}}\alpha}$

3.3.4. Pareto distribution

The Pareto distribution is often described as the basis of 80/20 rule. For example, 80 percent of customer complaints regarding a make of vehicle typically arise from 20 percent of components. Other applications include the distribution of income and the classification of stock in a warehouse on the basis of frequency of movement.

Range	$\alpha \le x < \infty$
Location parameter	$\alpha > 0$
Shape parameter	$\theta > 0$

Distribution function	$1-\left(\frac{\alpha}{x}\right)^{\theta}$
Probability density function	$\frac{\theta a^{\theta}}{x^{\theta+1}}$
Inverse distribution function	$\alpha(1-\mu)^{-\frac{1}{\theta}}$
Survival function	$\left(\frac{\alpha}{x}\right)^{\theta}$
Inverse survival function	$\alpha\mu^{-\frac{1}{\theta}}$
Hazard function	$\frac{\theta}{x}$
Cumulative hazard function	$\theta\log\left(\frac{x}{\alpha}\right)$

rth Moment about the mean

$$\frac{\theta\alpha^r}{\theta-r},\theta>r$$

Mean $\frac{\theta\alpha}{\theta-r},\ \theta>1$

Variance $\dfrac{\theta\alpha^2}{(\theta-1)^2(\theta-2)}$, $\theta > 2$

Mode α

Median $2^{\frac{1}{\theta}}\alpha$

Coefficient of variation

$$[\theta(\theta-2)]^{-\frac{1}{2}}, \theta > 2$$

3.3.5. Power function distribution

Range $0 \le x \le \alpha$

Shape parameter θ

Scale parameter $\alpha > 0$

Distribution function $\left(\dfrac{x}{\alpha}\right)^{\theta}$

Probability density function $\dfrac{\theta x^{\theta-1}}{\alpha^{\theta}}$

Inverse distribution function $\alpha\mu^{\frac{1}{\theta}}$

Hazard function $\dfrac{\theta x^{\theta-1}}{\alpha^{\theta}-x^{\theta}}$

Cumulative hazard function

$$-\log\left[1-\left(\frac{x}{\alpha}\right)^{\theta}\right]$$

rth Moment about the origin $\dfrac{\alpha^r\theta}{\theta+r}$

Mean $\dfrac{\alpha\theta}{\theta+1}$

Variance $\dfrac{\alpha^2\theta}{(\theta+2)(\theta+1)^2}$

Mode α for $\theta > 0$, 0 for $\theta < 1$

Median $\dfrac{\alpha}{2^{\frac{1}{\theta}}}$

Coefficient of skewness

$$\frac{2(1-\theta)(2+\theta)^{\frac{1}{2}}}{(3+\theta)\theta^{\frac{1}{2}}}$$

Coefficient of kurtosis

$$\frac{3(\theta+2)\left\{2(\theta+1)^2+\theta(\theta+5)\right\}}{\theta(\theta+3)(\theta+4)}$$

Coefficient of variation

$$[\theta(\theta+2)]^{-\frac{1}{2}}$$

3.4. Methodology

The procedure for performing the proposed classification via GLM would need a step-by-step procedure with which attempts will be made to: show, whether or not, each unclassified distribution belongs to the exponential family of distributions; and to obtain the link functions and linear predictors for each of the five (5) described distributions.

3.4.1. Linear predictor

The linear predictor is the quantity which incorporates the information about the independent variables into the model. The symbol η denotes the linear predictor. It is related to the expected value of the data through the link function. η is expressed as linear combinations of unknown parameters $\boldsymbol{\beta}$. The coefficients of the linear combination are represented as the matrix of independent variables $\mathbf{x}_i$. η can thus be expressed as

$$\eta = \mathbf{x}_i\,\boldsymbol{\beta} \tag{7}$$

Where $\mathbf{x}_i$ and $\boldsymbol{\beta}$ are as defined in equation (6).

3.4.2. Link functions

The link function $g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ specifies the link between random systematic components. It says how the expected value of the response relates to the linear predictor of explanatory variables. For instance, $g(E(Y_i)) = E(Y_i)$ for linear regression, or will be $\eta = \text{logit}(\pi)$ for logistic regression. Where μ_i, g, $\mathbf{x}_i$, and $\boldsymbol{\beta}$ are as defined in equation (6).

3.4.3. Maximum likelihood estimates of parameters for the GLM

Consider independent random variables $Y_1, \ldots, Y_N$ satisfying the properties of GLM. We wish to estimate parameters $\boldsymbol{\beta}$ which are related to the Y_i's through $E(Y_i) = \mu_i$ and $g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$.

For each Y_i, the log-likelihood function is:

$$l_i = y_i\, b(\theta_i) + c(\theta_i) + d(y_i) \qquad (8)$$

where the functions b, c and d are as defined in equation (3). Also,

$$E(Y_i) = \mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)} \qquad (9)$$

$$\text{var}(Y_i) = -\frac{\{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)\}}{\{b'(\theta_i)\}^3} \qquad (10)$$

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \qquad (11)$$

where $\mathbf{x}_i$ is a vector with elements x_{ij}, $j = 1, \ldots, p$.

The log-likelihood function for all the Y_i's is

$$l = \sum_{i=1}^{N} l_i = \sum x_i\, b(\theta_i) + \sum c(\theta_i) + \sum d(x_i)$$

To obtain the maximum likelihood estimator for the parameter β_j we need

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N} \left\{ \frac{\partial l_i}{\partial \theta_i} . \frac{\partial \theta_i}{\partial \mu_i} . \frac{\partial \mu_i}{\partial \beta_i} \right\} \qquad (12)$$

using the chain rule for differentiation. We will consider each term on the right hand side of equation (11). First,

$$\frac{\partial l_i}{\partial \theta_i} = x_i\, b'(\theta_i) + c'(\theta_i) = b'(\theta_i)(x_i - \mu_i)$$

by differentiating equation (8) and substituting (9). Next

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\left\{ \frac{\partial \mu_i}{\partial \theta_i} \right\}}$$

Differentiation of equation (9)

$$\frac{\partial \mu_i}{\partial \theta_i} = -\frac{c''(\theta_i)}{b'(\theta_i)} + \frac{c'(\theta_i)b''(\theta_i)}{\{b'(\theta_i)\}^2} = b'(\theta_i)\text{var}(Y_i)$$

from equation (10). Finally, from equation (11)

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} . \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij}$$

Hence the score, given in equation (12), is

$$U_j = \sum_{i=1}^{N} \left\{ \frac{(y_i - \mu_i)}{\operatorname{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right\} \quad (13)$$

The variance-covariance matrix of the U_j's has terns

$$\Im_{jk} = E\{U_j U_k\}$$

which form the "informative matrix $\Im$". From equation (13)

$$\Im_{jk} = E\left\{ \sum_{i=1}^{N} \left\{ \frac{(Y_i - \mu_i)}{\operatorname{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right\} \sum_{l=1}^{N} \left\{ \frac{(Y_l - \mu_l)}{\operatorname{var}(Y_l)} x_{lk} \left(\frac{\partial \mu_l}{\partial \eta_l} \right) \right\} \right\}$$
$$= \sum_{l=1}^{N} \frac{E\{(Y_i - \mu_i)^2\} x_{ij} x_{ik}}{\{\operatorname{var}(Y_i)\}^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \quad (14)$$

Because $E\{(Y_i - \mu_i)(Y_l - \mu_l)\} = 0$ for $i \neq l$ since all the Y_i's are independent. Using $E\{(Y_i - \mu_i)^2\} = \operatorname{var}(Y_i)$, equation (14) can be simplified to

$$\Im_{jk} = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{\operatorname{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \quad (15)$$

Now, we recall that one estimating equation for the method of scoring may be given by

$$\theta^{(m)} = \theta^{(m-1)} + \frac{U^{(m-1)}}{\Im^{(m-1)}} \quad (16)$$

This estimating equation as given by equation (16) for the method of scoring generalizes to

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + \left\{ \Im^{(m-1)} \right\}^{-1} \mathbf{U}^{(m-1)} \quad (17)$$

where $\mathbf{b}^{(m)}$ is the vector of estimates of the parameters $\beta_1, \dots, \beta_p$ at the m^{th} iteration. In equation (17), $\left\{ \Im^{(m-1)} \right\}^{-1}$ is the inverse of the information matrix with elements $\Im_{jk}$ given by equation (15) and $\mathbf{U}^{(m-1)}$ is the vector of elements given by (13), all evaluated at $\mathbf{b}^{(m-1)}$. If both sides of equation (17) are multiplied by $\Im^{(m-1)}$ we obtain

$$\Im^{(m-1)} \mathbf{b}^{(m)} = \Im^{(m-1)} \mathbf{b}^{(m-1)} + \mathbf{U}^{(m-1)} \quad (18)$$

From (15) $\Im$ can be written as

$$\Im = \mathbf{X}^T \mathbf{W} \mathbf{X}$$

where $\mathbf{W}$ is the $N \times N$ diagonal matrix with elements

$$w_{ii} = \frac{1}{\operatorname{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \quad (19)$$

The expression on the right-hand side of equation (18) is the vector with elements

$$\sum_{k=1}^{p} \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{\operatorname{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 b_k^{(m-1)} + \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{\operatorname{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

evaluated at $\mathbf{b}^{(m-1)}$; this follows from equations (15) and (13). Thus the right-hand side of equation (18) can be written as

$$\mathbf{X}^T \mathbf{W} \mathbf{z}$$

where $\mathbf{z}$ has elements

$$z_i = \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \quad (20)$$

with μ_i and $\frac{\partial \eta_i}{\partial \mu_i}$ evaluated at $\mathbf{b}^{(m-1)}$.

Hence the iterative equation (18), can be written as

$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}^T\mathbf{W}\mathbf{z} \qquad (21)$$

This is the same form as the normal equations for a linear model obtained by weighted least squares, except that it has to be solved iteratively because, in general, $\mathbf{z}$ and $\mathbf{W}$ depend on $\mathbf{b}$. Thus for GLMs, MLEs are obtained by the IWLS procedure.

Most statistical packages that include procedures for fitting GLMs have efficient algorithm based on equation (21). They begin by using some initial approximation $\mathbf{b}^{(0)}$ to evaluate $\mathbf{z}$ and $\mathbf{W}$, then equation (21) is solved to give $\mathbf{b}^{(1)}$ which in turn is used to obtain better approximations for $\mathbf{z}$ and between successive approximations $\mathbf{b}^{(m-1)}$ and $\mathbf{b}^{(m)}$ is sufficiently small, $\mathbf{b}^{(m)}$ is taken as the maximum likelihood estimate.

4. Implementation

4.1. Classifying the statistical distributions

To classify the error, Gumbel, logistic, Pareto and power function distributions the we must be able to express their respective probability distribution (density) functions in the form of equation 3, with an appropriate choice of the parameter θ. The same process is required for obtaining the linear predictors/deducing the link functions. In this study we have shown that the five (5) distributions all belong to the exponential family because they can all be expressed in the form:

$$f(x) = \exp\{a(x)b(\theta) + c(\theta) + d(x)\}$$

4.2. Obtaining the linear predictors

The linear predictors are obtained as follows.

1. State the probability distribution (or density) function.
2. Obtain the logarithm of the probability distribution (or density) function.
3. Take the exponent of the logarithm of the probability distribution (or density) function, noting that the coefficients of the $a(x)$ term, in the exponential form, are the linear predictors of the distribution.

4.2.1. Linear predictor for the error distribution

Theorem 1:

The linear predictor for the error distribution with probability density function $f(x)$, range $-\infty < x < \infty$, location parameter $-\infty < \alpha < \infty$ (the mean), scale parameter $\theta > 0$ and shape parameter $\beta > 0$ is $\theta^{-\left(\frac{2}{\beta}\right)}$.

Proof 1:

Given the probability density function of the error distribution

$$f(x) = \frac{\exp\left[-\left(\frac{|x-\alpha|}{\theta}\right)^{\frac{2}{\beta}} \Big/ 2\right]}{\theta\left(2^{\frac{\beta}{2}+1}\right)\Gamma\left(1+\frac{\beta}{2}\right)}$$

We take the logarithm of the probability density function to get:

$$\Rightarrow \ln f(x) = \ln\left\{\frac{\exp\left[-\left(\frac{|x-\alpha|}{\theta}\right)^{\frac{2}{\beta}} \Big/ 2\right]}{\theta\left(2^{\frac{\beta}{2}+1}\right)\Gamma\left(1+\frac{\alpha}{2}\right)}\right\}$$

$$\Rightarrow \ln f(x) = \ln\left\{\exp\left[-\left(\frac{|x-\alpha|}{\theta}\right)^{\frac{2}{\beta}} \Big/ 2\right]\right\} - \ln\left\{\theta\left(2^{\frac{\beta}{2}+1}\right)\Gamma\left(1+\frac{\beta}{2}\right)\right\}$$

$$\Rightarrow \ln f(x) = \left\{-\left(\frac{|x-\alpha|}{\theta}\right)^{\frac{2}{\beta}} \Big/ 2\right\} - \ln\left\{\theta\left(2^{\frac{\beta}{2}+1}\right)\Gamma\left(1+\frac{\beta}{2}\right)\right\}$$

$$\Rightarrow \ln f(x) = -\frac{\left(\frac{|x-\alpha|}{\theta}\right)^{\frac{2}{c}}}{2} - \ln\theta - \ln\left(2^{\frac{\beta}{2}+1}\right) - \ln\Gamma\left(1+\frac{\beta}{2}\right)$$

Now, restricting the domain of definition for the random variable from $-\infty < x < \infty$ to $x > 0$; and also restricting the domain of definition of the location parameter (that is, the mean) from $-\infty < \alpha < \infty$ to $\alpha > 0$, we have that:

$$\Rightarrow \ln f(x) = -\frac{\left(\frac{x-\alpha}{\theta}\right)^{\frac{2}{\beta}}}{2} - \ln\theta - \ln\left(2^{\frac{\beta}{2}+1}\right) - \ln\Gamma\left(1+\frac{\beta}{2}\right)$$

$$\Rightarrow \ln f(x) = -\frac{\left(\frac{x-\alpha}{\theta}\right)^{\frac{2}{\beta}}}{2} - \ln\theta - \ln\left(2^{\frac{\beta}{2}} \times 2\right) - \ln\Gamma\left(1+\frac{\beta}{2}\right)$$

$$\Rightarrow \ln f(x) = -\frac{\left(\frac{x-\alpha}{\theta}\right)^{\frac{2}{\beta}}}{2} - \ln\theta - \ln\left(2^{\frac{\beta}{2}}\right) - \ln 2 - \ln\Gamma\left(1+\frac{\beta}{2}\right)$$

$$\Rightarrow \ln f(x) = -\frac{1}{2}\left(\frac{x-\alpha}{\theta}\right)^{\frac{2}{\beta}} - \ln\theta - \frac{\beta}{2}\ln 2 - \ln 2 - \ln\Gamma\left(1+\frac{\beta}{2}\right)$$

Now, writing the last result here in the form of equation (3) gives:

$$\Rightarrow f(x) = \exp\left\{-\frac{1}{2}(x-\alpha)^{\frac{2}{\beta}}\theta^{-\left(\frac{2}{\beta}\right)} - \ln\left[\theta\left(2^{\frac{\beta}{2}+1}\right)\Gamma\left(1+\frac{\beta}{2}\right)\right]\right\}$$

Here,

$$a(x) = -\frac{1}{2}(x-\alpha)^{\frac{2}{\beta}},$$

$$b(\theta) = \theta^{-\left(\frac{2}{\beta}\right)},$$

$$c(\theta) = \ln\left[\theta\left(2^{\frac{\alpha}{2}+1}\right)\Gamma\left(1+\frac{\alpha}{2}\right)\right], \text{ and}$$

$$d(x) = 0$$

4.2.2. Linear predictor for the Gumbel distribution

Theorem 2:

The linear predictor for and Gumbel distribution with probability density function $f(x)$, range $-\infty < x < \infty$, location parameter (mode) α , and scale parameter $\theta > 0$, is given as $\frac{1}{\theta}$.

Proof 2:

Given the probability density function of the Gumbel distribution

$$f(x) = \left(\frac{1}{\theta}\right)\exp\left[-\frac{(x-\alpha)}{\theta}\right]\exp\left\{-\exp\left[-\frac{(x-\alpha)}{\theta}\right]\right\}$$

We take the logarithm of the probability density function to get:

$$\ln f(x) = \ln\left\{\left(\frac{1}{\theta}\right)\exp\left[-\frac{(x-\alpha)}{\theta}\right]\exp\left\{-\exp\left[-\frac{(x-\alpha)}{\theta}\right]\right\}\right\}$$

$$\Rightarrow \ln f(x) = \ln\left(\frac{1}{\theta}\right) + \left[-\frac{(x-\alpha)}{\theta}\right] + \left\{-\exp\left[-\frac{(x-\alpha)}{\theta}\right]\right\}$$

$$\Rightarrow \ln f(x) = \ln\left(\frac{1}{\theta}\right) - \frac{(x-\alpha)}{\theta} - \exp\left[-\frac{(x-\alpha)}{\theta}\right]$$

$$\Rightarrow \ln f(x) = -\frac{(x-\alpha)}{\theta} + \ln\left(\frac{1}{\theta}\right) - \exp\left[-\frac{(x-\alpha)}{\theta}\right]$$

$$\Rightarrow \ln f(x) = -(x-\alpha)\left(\frac{1}{\theta}\right) + \ln\left(\frac{1}{\theta}\right) - \exp\left[-(x-\alpha)\left(\frac{1}{\theta}\right)\right]$$

Now, writing the last result here in the form of equation (3) gives

$$\Rightarrow f(x) = \exp\left\{-(x-\alpha)\left(\frac{1}{\theta}\right) + \ln\left(\frac{1}{\theta}\right) - \exp\left[-(x-\alpha)\left(\frac{1}{\theta}\right)\right]\right\}$$

Here,

$$a(x) = -(x-\alpha),$$

$$b(\theta) = \frac{1}{\theta},$$

$$c(\theta) = \ln\left(\frac{1}{\theta}\right), \text{ and}$$

$$d(x) = -\exp\left[-(x-\alpha)\left(\frac{1}{\theta}\right)\right].$$

4.2.3. Linear predictor for the logistic distribution

Theorem 3:

The linear predictor for the logistic distribution with probability density function $f(x)$, range $-\infty < x < \infty$, location parameter α (the mean), and scale parameter $\theta > 0$ is given as $\frac{1}{\theta}$.

Proof 3:

Given the probability density function of the logistic distribution:

$$f(x) = \frac{\exp\left[\left(\frac{x-\alpha}{\theta}\right)\right]}{\theta\left[1+\exp\left(\frac{x-\alpha}{\theta}\right)\right]^2}$$

We take the logarithm of the probability density function to get:

$$\ln f(x) = \ln\left\{\frac{\exp\left[\left(\frac{x-\alpha}{\theta}\right)\right]}{\theta\left[1+\exp\left(\frac{x-\alpha}{\theta}\right)\right]^2}\right\}$$

$$\Rightarrow \ln f(x) = \ln\left\{\exp\left[\left(\frac{x-\alpha}{\theta}\right)\right]\right\} - \ln\theta - \ln\left[1+\exp\left(\frac{x-\alpha}{\theta}\right)\right]^2$$

$$\Rightarrow \ln f(x) = \frac{x-\alpha}{\theta} - \ln\theta - \ln\left[1+\exp\left(\frac{x-\alpha}{\theta}\right)\right]^2$$

$$\Rightarrow \ln f(x) = (x-\alpha)\frac{1}{\theta} - \ln\theta - \ln\left\{1+\exp\left[(x-\alpha)\frac{1}{\theta}\right]\right\}^2$$

Now, writing the last result here in the form of equation (3) gives

$$\Rightarrow f(x) = \exp\left\{(x-\alpha)\frac{1}{\theta} - \ln\theta - \ln\left\{1+\exp\left[(x-\alpha)\frac{1}{\theta}\right]\right\}^2\right\}$$

Here,

$$a(x) = (x-\alpha),$$

$$b(\theta) = \frac{1}{\theta},$$

$$c(\theta) = -\ln\theta, \text{ and}$$

$$d(x) = -\ln\left\{1+\exp\left[(x-\alpha)\left(\frac{1}{\theta}\right)\right]\right\}^2$$

4.2.4. Linear predictor for the Pareto distribution

Theorem 4:

The linear predictor for the Pareto distribution with probability density function $f(x)$, range $\alpha \le x < \infty$, location parameter $\alpha > 0$, and shape parameter $\theta > 0$ is given as $\theta + 1$

Proof 4:

Given the probability density function of the Pareto distribution:

$$f(x)=\frac{\theta\alpha^{\theta}}{x^{\theta+1}}$$

We take the logarithm of the probability density function to get:

$$\ln f(x)=\ln\left\{\frac{\theta\alpha^{\theta}}{x^{\theta+1}}\right\}$$

$$\Rightarrow \ln f(x)=\ln\{\theta\alpha^{\theta}\}-\ln\{x^{\theta+1}\}$$

$$\Rightarrow \ln f(x)=-(\theta+1)\ln x+\ln\theta+\ln\alpha^{\theta}$$

$$\Rightarrow \ln f(x)=-(\theta+1)\ln x+\ln\{\theta\alpha^{\theta}\}$$

Now, writing the last result here in the form of equation (3) gives

$$\Rightarrow f(x)=\exp\{-(\theta+1)\ln x+\ln\{\theta\alpha^{\theta}\}\}$$

Here,

$a(x)=-\ln x$,

$b(\theta)=\theta+1$,

$c(\theta)=\ln\{\theta\alpha^{\theta}\}$, and

$d(x)=0$.

4.2.5. Linear predictor for the power function distribution

Theorem 5:

The linear predictor for the power function distribution with probability density function $f(x)$, range $0\le x\le\alpha$, scale parameter $\alpha>0$, and shape parameter θ is given as $\theta-1$.

Proof 5:

Given the probability density function of the power function distribution:

$$f(x)=\frac{\theta\, x^{\theta-1}}{\alpha^{\theta}}$$

We take the logarithm of the probability density function to get:

$$f(x)=\ln\left\{\frac{\theta\, x^{\theta-1}}{\alpha^{\theta}}\right\}$$

$$\Rightarrow \ln f(x)=\ln\{\theta\, x^{\theta-1}\}+\ln\alpha^{\theta}$$

$$\Rightarrow \ln f(x)=\ln x^{\theta-1}+\ln\theta+\ln\alpha^{\theta}$$

$$\Rightarrow \ln f(x)=(\theta-1)\ln x+\ln\{\theta\,\alpha^{\theta}\}$$

Now, writing the last result here in the form of equation (3) gives

$$\Rightarrow f(x)=\exp\{(\theta-1)\ln x+\ln(\theta\,\alpha^{\theta})\}$$

Here,

$a(x)=\ln x$,

$b(\theta)=\theta-1$,

$c(\theta)=\ln\{\theta\,\alpha^{\theta}\}$, and

$d(x)=0$.

4.3. Deducing the link functions

Recall that if $b(\theta)=\mathbf{x}^{T}\boldsymbol{\beta}$ defines the linear predictor of a distribution $f(x)$ then the link function of the distribution will be defined by specific instances $b(\theta_i)=\mathbf{x}_i^{T}\boldsymbol{\beta}$ such that $\theta=b^{-1}(\mathbf{x}^{T}\boldsymbol{\beta})$, where θ is the parameter of interest. Table 1 illustrates this.

4.4. Discussion of findings

The results in this study have shown that the error, Gumbel, logistic, Pareto, and power function distributions all belong to the exponential family since, for a given parameter of interest θ it was possible to write their probability functions $f(x)$ in the form $\exp\{a(x)b(\theta)+c(\theta)+d(x)\}$. The linear predictors of all distributions have been obtained in this study, and are observed not to be in canonical form especially because the term $a(x)\neq x$; the corresponding link functions of the newly classified distributions have been also deduced from the linear predictors. Based on the linear predictors we have observed that the error, Pareto and power function distributions all have $d(x)=0$; whereas, the Gumbel and logistic distributions have $d(x)\neq 0$.

TABLE 1
Newly classified members of the exponential family

Distribution	$a(x)$	**Link function**	$c(\theta)$	$d(x)$	**Linear predictor or Natural parameter** $b(\theta)$
Error	$-\frac{1}{2}(x-\alpha)^{\frac{2}{\beta}}$	$\theta_i^{-\left(\frac{2}{\beta}\right)}$	$\ln\left[\theta\left(2^{\frac{\alpha}{2}+1}\right)\Gamma\left(1+\frac{\alpha}{2}\right)\right]$	0	$\theta^{-\left(\frac{2}{\beta}\right)}$
Gumbel	$-(x-\alpha)$	$\frac{1}{\theta_i}$	$\ln\left(\frac{1}{\theta}\right)$	$-\exp\left[-(x-\alpha)\left(\frac{1}{\theta}\right)\right]$	$\frac{1}{\theta}$
Logistic	$x-\alpha$	$\frac{1}{\theta_i}$	$-\ln\theta$	$-\ln\left\{1+\exp\left[(x-\alpha)\left(\frac{1}{\theta}\right)\right]\right\}^2$	$\frac{1}{\theta}$
Pareto	$-\ln x$	θ_i+1	$\ln\{\theta\alpha^{\theta}\}$	0	$\theta+1$
Power function	$\ln x$	θ_i-1	$\ln\{\theta\alpha^{\theta}\}$	0	$\theta-1$

5. Conclusion

In conclusion, this study showed that the error, Gumbel, logistic, Pareto, and power function distributions are members of the exponential family, although not in canonical form; hence, they should be treated as such. This study has also established the link functions and linear predictors for the newly established probability distributions (or density) functions with certain properties as shown in Table 1 above. It was therefore the submission of this study that software meant for theoretical and practical computations on the generalized linear model (GLM) should adopt and upgrade to accommodating the findings of this study as tendencies abound that this could widen the scope of its application.

For further studies in this research direction, we have suggested that prospective scholars should attempt the classification of more unclassified statistical distributions in order to unravel the nature of their linear predictors and link functions. Further suggestions are that attempts should be made at obtaining the maximum likelihood estimates (via GLM) of the newly classified distributions in this research.

References

Alexander, V. E. & Wolfgang, W. (2015). Generalized linear models for the analysis of single subject data and for the comparison of individuals. *Journal for Person-Oriented Research, 1*(1), 56 - 71.

Altham, P. M. E. (2011). *Introduction to generalized linear modeling*. London: University of Cambridge Press.

Annette, J. D. (2002). *Generalized linear models* (2nd ed.). London: Chapman and Hall/CRC.

Catherine, F., Mervan, E., Nicholas, H. & Brian, P. (2011). *Statistical distributions* (4th ed.) Canada: John Wiley & Sons.

Christian, W. (2007). *Hand-book on statistical distributions for experimentalists*. University of Stockholm: Particle Physics Group Fysikum.

Cochran, W. G. & Cox. G. M. (1992). Experimental designs (2nd ed.). USA: Wiley Classic Library.

Cox, D. R. & Reid, N. (2000). *The theory of the design of experiments*. Boca Raton: Chapman and Hall/CRC.

David, A. S. (2018). MATH 523 – *Generalized linear model*. USA: McGill University.

Farzana, J., Borhan, S. & Ataharul, I. M. (2016). An application of the generalized linear model for the geometric distribution. *Journal of Statistics: Advances in Theory and Applications, 16*(1), 45 - 65.

Heather, T. (2008). *Introduction to generalized linear models*. United Kingdom: University of Warwick Press.

John, L. M., Patricia, J. S., Aaron, R. P. & Jeffrey, M. (2007). New models for old questions: generalized linear models for cost prediction. *Journal of Evaluation in Clinical Practice, 5*(4), 1 - 9.

Jun, Z. (2011). Theory and application of generalized linear models in insurance (PhD mathematics). Department of Mathematics and Statistics, Concordia University, Canada.

Khuri, A. I. (2010). Response surface methodology. *Computational Statistics, 2*(2), 23 - 34.

Lindsey, J. K. (1997). *Applying generalized linear models*. New York: Springer.

Lindsey, J. K. & Jones, B. (1998). Choosing among generalized linear models applied to medical data. *Statistics in Medicine, 17*(1), 59 – 68.

Mark, G., Anand, K. & Dan. T. (2016). *Generalized linear models for insurance rating*. Virginia: Casualty Actuarial Society.

Mathew, G., Yuan, L., Yichong, X., Silun, W. (2016). Exponential family and generalized linear models. Probabilistic Graphical Models. USA: Spring.

Marlene, M. (2004). Generalized linear model. *Fraunhofer Institute for Industrial Mathematics (ITWM), 3*(5), 1 - 24.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.

Michael, I. J. (2003). *An introduction to probabilistic graphical models*. Berkeley: University of Carlifornia.

Moffat, U. I. & Emmanuel, A. A. (2018). A probabilistic application of generalized linear model in discrete-time stochastic series. *Journal of Scientific Research and Reports, 19*(3), 1 – 10.

Montgomery, D. C. (2013). *Design and analysis of experiment* (8th ed.). USA: John Wiley and Sons, Inc.

Murtala, A., Udokang, A. E., Raji, S. T. & Bello, L. K. (2015). An empirical study of generalized linear model for count data. *Applied and Computational Mathematics, 2*(3), 4 – 5.

Oehlert, W. O. (2010). *A first course in design and analysis of experiments*. USA: Library of Congress.

Penzer, J. (2011). *Advanced statistics: distribution theory*. London: University of London.

Rodriguez, G. (2001). *Generalized linear model theory*. USA: Appendix B.

Sourish, D. (2008). *Generalized linear models and beyond: An innovative approach from Bayesian perspective* (PhD dissertation). University of Connecticut, United States.

Sourish, D. & Dipak, D. (2014). On Bayesian analysis of generalized linear models using the Jacobian technique. *The American Statistician*, *60*(3), 1 – 5.