Comparıson Study of The Poisson Regression Model Parameters Estımated With Different tow Methods(statistical study)

Muthanna Ali Hussein

muthanaalii@yahoo.com

Iraqi Ministry of Education/ General Directorate of Educational Planning/Iraq

ABSTRACT

The aim of this study is comparative examination of the estimation methods where can be employed to estimate Poisson regression model parameters. Occurrence number of any events that takes place within a specified time period as a result of conducted experiments can be expressed as count data. Poisson regression model is employed as an important data interpretation tool to analyze this kind of count data. Poisson regression models are regarded as a sub-branch of generalized linear models.

The following tow methods are used for parameters estimation: 1)Maximum Likelihood Estimation (MLE),2) linear least squares(OLS). MATLAB packaged software is used for generation of simulation data and for parameter estimates. Poisson regression model parameters were estimated and models were generated by using of Monte Carlo simulation with sample sizes of 30, 60, 90 and 100 in accordance with Poisson distribution.

Mean square error (MSE) and mean absolute percentage error (MAPE) criteria were used for comparison of estimated parameters in terms of their effectiveness

Mean square error (MSE) and mean absolute percentage error (MAPE) criteria were used for comparison of estimated parameters in terms of their effectiveness. As a result of comparison, it was shown that MLE gives better results than other method OLS.

Keywords: Generalized linear model, , , Maximum Likelihood Method,linear least squares

Poisson Distribution

The Poisson distribution in discrete distributions is very useful in many statistical applications.
is important. The Poisson distribution is random over a given time interval (or space).
is the discrete probability distribution for counts of events occurring. Y specific range
The average number of events per interval, if treated as the number of events in(λ)
The first person to derive this function was a Frenchman named Simeon Poisson. Poisson found that the derivative of this distribution function was close to the binary binomial distribution, and in 1873published the distribution he derived in

$$P_r(Y/\mu)=\frac{e^{-\lambda}\ \lambda^{Y}}{Y!}, \qquad \lambda>0,\ Y_i=0,1,2,...$$

Here,

e is the natural logarithm constant, e ≅ 2.718282

λ: is the only parameter of the distribution and is always greater than zero (λ > 0)

$$S_1^2=\frac{1}{N}\sum_{t=1}^{N}\hat{\varepsilon}_t^2+2\sum_{t=1}^{l}\left(1-\frac{i}{l+1}\right)\frac{1}{N}\sum_{t=i+1}^{N}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}$$

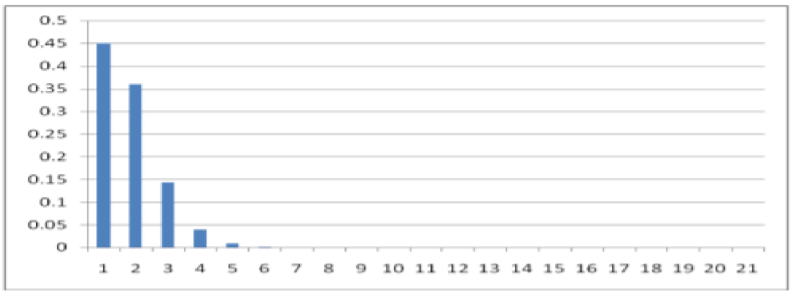
Properties of the Poisson Distribution

A- The average or expected value of a certain event in a certain time period.
the number of occurrences parameter distribution, as well as the arithmetic mean and
shows that the variance is equal to each other

$$E(Y)=\sum_{k=0}^{\infty}ke^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}=e^{-\lambda}\lambda\sum_{k=0}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=e^{-\lambda}\lambda\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

$$E(Y^2)=\lambda^2+\lambda$$

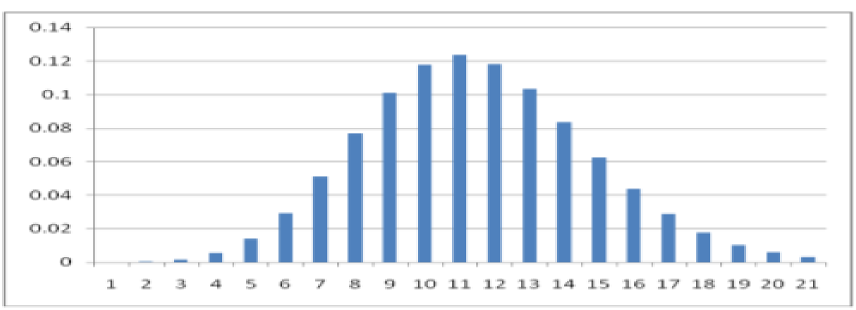
B- The histogram of the Poisson distribution is skewed to the right



$\mu = 0.8$

A special case where the arithmetic mean and variance parameters are equal to 10.5.

in cases (2 σ Y =λ=μ=10.5) the poisson distribution approaches the normal distribution

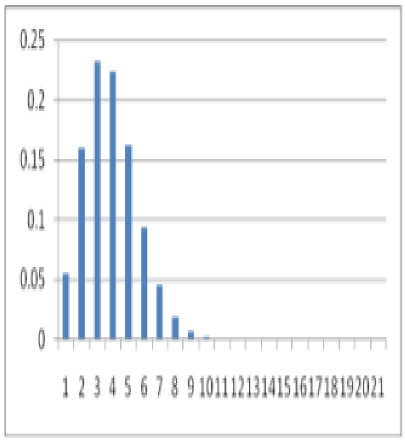


$\mu = 10.5$

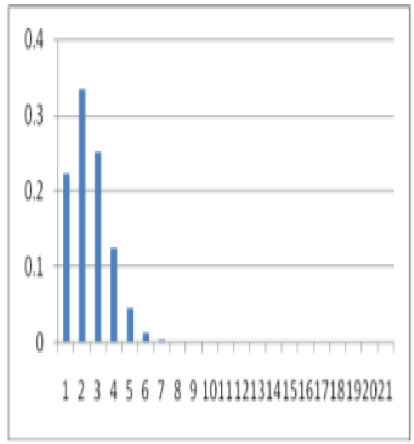
The probability of parameter value distribution starting from zero is low μ = 0.8

The probability of zero is P(0)=0.449, P(0)=0.055 when μ=2.9, and P(0)=0.00003 when μ=10.5.

This so the mass distribution tends to the normal distribution



$\mu = 2.9$



$\mu = 1.5$

Generalized Linear Models

Thus far our focus has been on describing interactions or associations between two or three categorical variables mostly via single summary statistics and with significance testing. Models can handle more complicated situations and analyze the simultaneous effects of multiple variables, including mixtures of categorical and continuous variables. For example, the Breslow-Day *statistics* only works for 2 × 2 × *K* tables, while log-linear models will allow us to test of homogeneous associations in *I* × *J* × *K* and higher-dimensional tables. We will focus on a special class of models known as the *generalized linear models (GLIMs or GLMs* in Agresti*)*.

The structural form of the model describes the patterns of interactions and associations. The model parameters provide measures of strength of associations. In models, *the focus is on estimating the model parameters*. The basic inference tools (e.g., point estimation, hypothesis testing, and confidence intervals) will be applied to these parameters. When discussing models, we will keep in mind

start clear up some potential misunderstandings about terminology. The term *general linear model* (GLM) usually refers to conventional linear regression models for a continuous response variable given continuous and/or categorical predictors. It includes multiple linear regression, as well as ANOVA and ANCOVA (with fixed effects only). The form is $y_i \sim N(x_i^T\beta,\sigma^2)$, where x_i contains known covariates and β contains the coefficients to be estimated.

Assumptions:

- The data $Y_1,Y_2,...,Y_n$ are independently distributed, i.e., cases are independent.
- The dependent variable Y_i does NOT need to be normally distributed, but it typically assumes a distribution from an exponential family (e.g. binomial, Poisson, multinomial, normal,...)
- GLM does NOT assume a linear relationship between the dependent variable and the independent variables, but it does assume linear relationship between the transformed response in terms of the link function and the explanatory variables; e.g., for binary logistic regression $\text{logit}(\pi)=\beta_0+\beta X$.
- Independent (explanatory) variables can be even the power terms or some other nonlinear transformations of the original independent variables.
- The homogeneity of variance does NOT need to be satisfied. In fact, it is not even possible in many cases given the model structure, and *overdispersion* (when the observed variance is larger than what the model assumes) maybe present.
- Errors need to be independent but NOT normally distributed.
- It uses maximum likelihood estimation (MLE) rather than ordinary least squares (OLS) to estimate the parameters, and thus relies on large-sample approximations.
- Goodness-of-fit measures rely on sufficiently large samples, where a heuristic rule is that not more than 20% of the expected cells counts are less than 5

Intuitive explanation of maximum likelihood estimation

Estimation of Poisson regression parameters is based on the maximum likelihood estimation (MLE) method. It seeks to answer the question of what values the regression coefficients can take so that the maximum likelihood estimation data will yield results. The maximum likelihood estimation depends on a likelihood function. This function describes the probability of viewing the data as a function of the parameter set, the poisson regression uses the poisson distribution as the probability model, and the regression coefficients define the parameters that determine the mean structure of the data. The purpose of the maximum likelihood method is to estimate the regression coefficients that maximize the likelihood function. This is possible by equating the first derivative of the likelihood equation to zero and solving for the regression coefficients.

In the most practical situations, maximum likelihood estimation requires iterative processes. This adds extra complexity to these models. In particular, complex models with many parameters and small sample sizes prevent the process from converging. Ultimately, the results of the maximum likelihood estimation yield asymptotic standard errors for the regression coefficient. To discuss the maximum likelihood calculation for the Poisson regression, let (μi) be the mean of the ith outcome variable, with i = 1,2 ,...,n. The mean of the outcome variable; Since X1, X2, ..., Xk are assumed to be a function of the set of explanatory variables, the notation μ(Xi, β); used to associate the mean (μi) with Xi (the explanatory variable value for case i) and β (regression coefficients)

Probability density function (pdf) is shown f(y|θ) with a set of θ parameters for the random variable y. This function defines the data generation process. This process forms the basis of the observed sample data; same

It also provides a mathematical explanation of the data that the process will create. The joint density of the n independent variables and ideally distributed observations in this process is the product of the individual intensities.

$$f(y_1, \ldots\ldots, y_n \mid \theta) = \prod f(y_i / \theta) = L(\theta \mid y)$$

This is the joint density likelihood function. It is a function defined by the unknown parameter vector θ. Here, y is used to denote the aggregation of sample data (Myung, 2002). Let's consider the following poisson regression model for estimation.

$$\mu = \mu(x, \beta) = e^{(x_i, \beta)}$$

$$P(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$P(9, 9.5, 11; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(9-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(9.5-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(11-\mu)^2}{2\sigma^2}\right)$$

We just have to figure out the values of μ and σ that results in giving the maximum value of the above expression.

Explaining the Simulation Experience

Initially, the random error variable is transferred to the poisson distribution with the following parameter.
according to the Poisson regression model.(μi)
Also, in accordance with the formula
In the poisson regression model, the values of a dependent variable (Yi) are calculated:

$$Y_i = \mu_1 = EXP\left\{ \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip} \right\}$$

i=1,2, …, n gözlem sayısını temsil eder

j=1,2, …, p bağımlı değişken sayısını gösterir

Model	Dist.	KolmogorovSmirnov		AndersonDarling	
(1)		Statistic	Rank	Statistic	Rank
1	D. Uniform	0.2	2	8.8135	5
2	Geometric	0.39673	4	3.5247	3
3	Logarithmic	0.49522	5	6.2388	4
4	Neg. Binomial	0.20193	3	0.865	1
5	Poisson	0.14257	1	0.96911	2
6	Bernoulli	No fit (data max > 1)			
7	Binomial	No fit			
(2)					
8	D. Uniform	0.15	1	4.1414	4
9	Geometric	0.28741	4	2.3433	3
10	Logarithmic	0.40605	5	5.0328	5
11	Neg. Binomial	0.19095	3	0.85734	1
12	Poisson	0.16556	2	1.9358	2
13	Bernoulli	No fit (data max > 1)			
14	Binomial	No fit			
(3)					
15	D. Uniform	0.36728	3	7.9234	5
16	Geometric	0.32676	2	3.4715	3
17	Logarithmic	0.45261	5	6.0602	4
18	Neg. Binomial	0.38153	4	2.7271	1
19	Poisson	0.23636	1	3.1162	2
20	Bernoulli	No fit (data max > 1)			
21	Binomial	No fit			

	Parametr		0.3553	0.2491	0.3462	0.6916	0.3423	0.3104
N	mothed		B_0	B_1	B_2	B_3	B_4	B_5
30	MLE	**Parametre**	0.556089	0.357241	0.695224	0.349376	0.313179	0.005278
		MSE	0.254036	0.076815	0.183000	0.175068	0.062991	0.152776
		$MAPE$	3.003680	1.699130	0.527796	9.170388	4.323978	6.603551
	ols	**Parametre**	0.510980	0.362191	0.718242	0.358956	0.321708	0.005002
		MSE	0.296333	0.090946	0.217032	0.18145	0.073005	0.163909
		$MAPE$	4.014366	2.412184	0.661648	5.641141	5.898058	6.617107
60	MLE	**Parametre**	0.586987	0.352138	0.687247	0.350634	0.313829	-0.00374
		MSE	0.112826	0.02823	0.135293	0.134097	0.01936	0.116851
		$MAPE$	0.679997	2.573485	0.477586	2.88272	1.386082	22.97172
	ols	**Parametre**	0.573059	0.356402	0.69281	0.354095	0.31352	-0.00305
		MSE	0.117307	0.031242	0.14167	0.134198	0.021146	0.118671

		MAPE	1.132394	0.977011	0.478259	2.749133	1.080811	25.63029
90	MLE	**Parametre**	0.684905	0.349495	0.594562	0.353515	0.31654	0.000505
		MSE	0.093058	0.021391	0.132098	0.131138	0.010772	0.105725
		MAPE	0.630805	0.400043	0.489832	1.295059	0.368347	35.11
	ols	**Parametre**	0.568593	0.352412	0.701355	0.347058	0.319925	0.001167
		MSE	0.096748	0.023712	0.139116	0.130336	0.012276	0.106672
		MAPE	0.856344	1.557413	0.492725	1.43252	0.432786	87.1599
100	MLE	**Parametre**	0.601159	0.345862	0.661671	0.391088	0.408597	0.00547
		MSE	0.088618	0.017836	0.127763	0.131624	0.009312	0.102069
		MAPE	0.432488	0.308429	0.490216	1.22151	0.353804	20.70656
	ols	**Parametre**	0.591106	0.346807	0.655365	0.32424	0.329764	0.004749

	MSE	0.089825	0.019489	0.131399	0.130758	0.010092	0.102255
	$MAPE$	0.417502	0.326342	0.491766	1.281704	0.379563	45.6552

Al – Nasir, A. M & Rashid, D. H (1988). Statistical Inference, Baghdad University, Higher Education Printing Press, Iraq, Baghdad.

Atkins, D.C., Gallop, R.J., (2007). Rethinking How Family Researchers Model Infrequent Outcomes: A Tutorial On Count Regression And Zero-Inflated Models, Journal Of Family Psychology, Vol. 21, No. 4, Pp. 726 – 735.

Batah, F. S (2010). A New Estimator By Generalized Modefied Jackknife Ridge Regression Estimator, Jornal Of Basrah Researches (Sciences), Vol. 37, No. 4, Pp. 138-149.

Binjie, G.,Feng, P., (2013). Modified Gravitational Search Algorithm With Particle Memory Ability And Its Application, Jiangnan University, China.

Brent, R.P. (1973). Algorithms For Minimization Without Derivatives, Englewood Cliffs, NJ: Prenticehall, Cambridge University Press, USE, P. 78.

Cameron, A. C.,& Trivedi, P. K. (2013). Regression Analysis Of Count Data (Vol. 53). Cambridge University Press.

Cameron, A.C., Trivedi, P .K., (1998). Essentials of Count Data Regression, Cambridge University Press, New York, USA

Chan, Y.H. (2005). Log-Linear Models: Poisson Regression. Singapore Med. J. 46(8), Pp. 377– 386.

De Jong, P., And Heller, G. Z. (2008). Generalized Linear Models For Insurance Data (Vol. 136). Cambridge: Cambridge UniversityPress.

Dobson,J,A.,(1990), An Introduction generalized Linear Models, New South Wales, Journal Of Computational And Graphical Statistics Australia, Pp.30-33

Fallah, N., Gu, H., Mohammad, K., Seyyedsalehi, S. A., Nourijelyani, K., & Eshraghian, M. R. (2009). Nonlinear Poisson regression using neural networks: a simulation study. Neural Computing and Applications, 18(8),939-943.