On the Confidence Band of Local Likelihood Estimates in Generalized Poisson Regression Model

Erni Tri Astuti[1], I Nyoman Budiantara[2], Sony Sunaryo[3], Mohamad Dokhi[4]

***Abstract*—Unlike it counterpart in parametric regression modeling, the development of nonparametric regression for count response are moving slowly. In this research we developed a new nonparametric regression approach for modeling count response using local polynomial smoothing. By assuming generalized Poisson distribution for count response, this model should robust for over dispersion problem that often occurred in count data modeling. Using maximum likelihood method for finding the estimator, we called it as local (maximum) likelihood estimator. In this paper we construct a confidence band of the unknown regression function, which is difficult to build in nonparametric regression context. The construction of the confidence band needs estimated bias and variance of local likelihood estimator that we have been derived earlier. We conducted some simulation to show the behavior of the estimator as well as the confidence band.**

***Index Terms*—generalized Poisson distribution, local polynomial smoothing, local likelihood, nonparametric regression, confidence band**

I. Introduction

In the context of parametric regression analysis, Poisson regression is a standard and baseline model for describing the relationship between count response with some covariates. Following its name, count response is assumed to follow the Poisson distribution which have restricted properties called equi-dispersion (i.e. mean should be equal to variance). This situation is hard to fulfill by observational data, and often the opposite situation where the observed variance exceeds the observed mean, called over-dispersion, is occurred. Fitting such data using Poisson regression model will seriously under estimated the variance and can lead to misleading conclusion in the inference [1]. As an alternative, there are others model such as: Negative Binomial regression model [2], *Poisson Log Normal* (PLN) model and *Poisson Inverse Gaussian* (PIG) model [3] and also Generalized Poisson regression model [4]. The last model is preferred because not only more general than Poisson regression (i.e. in special case it reduces to Poisson Regression), but it is simpler comparing to others

In many cases, the relationship between response and covariates cannot describe by simply fitting some parametric function such as linear, exponential or polynomial function. In such case, nonparametric regression seems to be a reliable and reasonable choice. The aim of nonparametric regression is to minimize the assumption about regression function and let the data seeking for the function itself [hardle]. In nonparametric regression, scatter plot smoothing is the simplest method to estimate regression function. There are several approaches for determined the regression function, such as kernel, spine and local polynomial technique. These techniques known as local fitting methods because the estimation of regression function is done locally around some interval of points.

Unlike it counterpart in parametric regression model, the development of nonparametric regression for count response with local fitting is moving slowly. There is not much research in this area, except [5], [6]. Local likelihood is a concept introduced by [7] and developed more intensively by [8]. This method extends the nonparametric regression analysis to maximum likelihood based regression model which also known as likelihood-based smoother. In this model, the mean of response variables are assumed depends on covariates with some nonlinear link function. Although, there are no presumed function for the regression curve itself.

In this research we develop a nonparametric regression model for count response using local polynomial approach for the estimation of regression function. The count response is assumed to have generalized Poisson distribution.. We called the estimator as local likelihood estimator because it is determined by local maximum likelihood method. Based on Taylor development of degree p and considering the generalized Poisson regression locally, in a neighborhood of some points of interest of the covariate, we also present the bias, the variance and the confidence band of the regression function. We also present some simulation result to show the behavior of the local likelihood estimator as well as the confidence band of the regression function.

II. Local Likelihood Estimator

Let Y be the response variable, which is a count, and x is a predictor variables. The distribution of Y_i (i=1,2,…,n) at given x_i is following the generalized Poisson distribution, with the probability density function given by:

$$f(y_i,\mu,\varphi)=\left(\frac{\mu_i}{1+\varphi\mu_i}\right)^{y_i}\frac{(1+\varphi y_i)^{y_i-1}}{y_i!}\exp\left(\frac{-\mu_i(1+\varphi y_i)}{1+\varphi\mu_i}\right)\quad, y_i=0,1,2,\cdots \quad (1)$$

With

[1] PhD student at Department of Statistics, Institut Teknologi Sepuluh, Surabaya, Indonesia and lecturer at Sekolah Tinggi Ilmu Statistik, Jakarta Indonesia, e-mail: erni@mhs.statistika.its.ac.id

[2,3] Lecturer at Department of Statistics, Institut Teknologi Sepuluh, Surabaya, Indonesia , e-mail: i_nyoman_b@statistika.its.ac.id , sonny_s@ statistika.its.ac.id

[4]Lecturer at Department of Applied Statistics, Sekolah Tinggi Ilmu Statistik, Jakarta, Indonesia , e-mail:dokhi@stis.ac.id

$$E(Y_i \mid x_i) = \mu_i$$

and

$$V(Y_i \mid x_i) = \mu_i(1+\varphi\mu_i)^2$$

The parameter φ plays as dispersion parameter. When $\varphi = 0$, it will reduce to Poisson probability density. When $\varphi < 0$ this model is under dispersed, and when $\varphi > 0$ it will over disperse relative to Poisson distribution respectively [9]. In the local generalized Poisson regression model, instead of considering some specified regression function, the dependence of mean response with a covariate is describe by a smooth nonparametric regression function s:

$$\mu_i = \exp(s(x_i)) \quad (2)$$

Assume that the function s has a $(p+1)^{th}$ continuous derivative at the point x_0. For data points x_i in a neighborhood of x_0 or $x_i \in (x_0 - h, x_0 + h)$, with h is a bandwidth, we approximate $s(x_i)$ via a Taylor expansion by a polynomial of degree p:

$$s(x_i) \approx s(x_0) + s'(x_0)(x_i - x_0) + \cdots + \frac{s^p(x_0)}{p!}(x_i - x_0)^p = \mathbf{x}_i^T\boldsymbol{\beta} \quad (3)$$

where $\mathbf{x}_i = (1, (x_i - x_0), \cdots, (x_i - x_0)^p)^T$, $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)^T$ with

$$\beta_j = \frac{s^j(x_0)}{j!},\ j = 0,1,\cdots,p \quad (4)$$

For data points (X_i, Y_i) in a neighborhood of x_0, the contribution to the log likelihood function is weighted by some kernel function $K_h(\cdot) = K(\cdot/h)/h$. By assuming generalized Poisson distribution for response variable Y_i , these considerations yield the conditional local kernel weighted log-likelihood:

$$L_{p,h}(\boldsymbol{\beta},\varphi,x_0) = \sum_{i=1}^{n}\left\{ y_i \ln\left(\frac{\mu_i}{1+\varphi\mu_i}\right) + (y_i - 1)\ln(1+\varphi y_i) - \frac{\mu_i}{1+\varphi\mu_i} - \ln(y!) \right\} \times K_h(x_i - x_0) \quad (5)$$

where $\mu_i(x) = \exp\left[\mathbf{x}_i^T\boldsymbol{\beta}\right]$ and $K_h(\cdot) = K(\cdot/h)/h$ is a Kernel weight. The choice of the kernel function is not a crucial issues, because the result is almost similar for any kind of kernel function including Epachnecnikov, Gaussian or Boxcar Kernels [10]. The estimator for regression function, is the solution of $(p+2)$ equation :

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i(x)}{(1+\varphi\mu_i(x))^2} K_h(x_i - x_0)(x_i - x_0)^v = 0 \quad ,j = 0,1,\cdots,p \quad (6)$$

$$\frac{\partial L}{\partial \varphi} = \sum_{i=1}^{n}\left\{ -\frac{y_i\mu_i(x)}{1+\varphi\mu_i(x)} + \frac{y_i(y_i-1)}{1+\varphi y_i} - \frac{\mu_i(x)(y_i - \mu_i(x))}{(1+\varphi\mu_i(x))^2} \right\} K_h(x_i - x_0) \quad (7)$$

The solution of the system which is called local (maximum) likelihood estimator can be solved by iterative procedure such as Newton Raphson Methods. The log-likelihood function above depends on two quantities, the smoothing parameter (h) and the order of polynomial (p). The model complexity is effectively controlled by the bandwidth h. As h increases from 0 to $+\infty$, the model runs from the most complex model (interpolation) to the simplest model and [10] stated that a too large bandwidth under parameterizes the regression function causing a large modeling bias, while too small bandwidth over parameterizes the unknown function and result in noisy estimates. Ideal or optimal model is lying between the two models, which can be obtained by different criteria's, one such criteria is cross validation (CV) [11].

Bias and Variance of the Estimator

The estimator $\hat{\boldsymbol{\beta}}$ is biased because there is an approximation error in Taylor expansion (3). By considering a further expansion with $(p+a)$ degree for approximate $s(x)$, the estimated bias for local likelihood estimator is given as in [12]

$$\hat{\mathbf{b}}_p(\hat{\boldsymbol{\beta}}) = -\mathbf{L}''^{*}_{p,h}(\hat{\boldsymbol{\beta}},x_0)^{-1}\mathbf{L}'^{*}_{p,h}(\hat{\boldsymbol{\beta}},x_0) \quad (8)$$

where $\mathbf{L}'^{*}_{p,h}(\boldsymbol{\beta},x_0)$ and $\mathbf{L}''^{*}_{p,h}(\boldsymbol{\beta},x_0)$ are the gradient vector and Hessian matrix of the local likelihood given by

$$L^{*}_{p,h}(\boldsymbol{\beta},\varphi,x_0) = \sum_{i=1}^{n}\left\{ y_i \ln\left(\frac{\mu_i^*}{1+\varphi\mu_i^*}\right) + (y_i - 1)\ln(1+\varphi y_i) - \frac{\mu_i^*}{1+\varphi\mu_i^*} - \ln(y!) \right\} \times K_h(x_i - x_0) \quad (9)$$

with

$$\mu_i^* = \exp(\mathbf{x}_i^T\boldsymbol{\beta} + r_i)$$

and

$$r_i = \beta_{p+1}(x_i - x_0)^{p+1} + \beta_{p+2}(x_i - x_0)^{p+2} + \cdots + \beta_{p+a}(x_i - x_0)^{p+a} \quad (10)$$

For example, if we set p=1 and a=2, then (8) can be expressed

$$\hat{\mathbf{b}}_1(\hat{\boldsymbol{\beta}}) = -\left(\sum_{i=1}^{n}\frac{\hat{\mu}_i^*}{(1+\hat{\varphi}\hat{\mu}_i^*)^3}(1+2\hat{\varphi}y_i - \hat{\varphi}\hat{\mu}_i^*)K_h(x_i - x_0)\mathbf{x}_i\mathbf{x}_i^T\right)^{-1} \times \left(\sum_{i=1}^{n}\frac{y_i - \hat{\mu}_i^*}{(1+\hat{\varphi}\hat{\mu}_i^*)^2}K_h(x_i - x_0)\mathbf{x}_i\right) \quad (11)$$

where $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p$ and $\hat{\varphi}$ is the solution of (6),(7) and $\hat{r}_i$ is the solution of (9).

On the other hands the estimated variance of the estimator can be computed by

$$\hat{\mathbf{V}}_p(\hat{\boldsymbol{\beta}}) = V\left(\ell'(s(x_0),y_i)\right)\left(\mathbf{L}''_{p,h}(\hat{\boldsymbol{\beta}},x_0)\right)^{-1}\mathbf{S_n}\left(\mathbf{L}''_{p,h}(\hat{\boldsymbol{\beta}},x_0)\right)^{-1} \quad (12)$$

where $\mathbf{L}''_{p,h}(\boldsymbol{\beta},x_0)$ is the Hessian matrix of (5) and

$$\mathbf{S_n} = \sum_{i=1}^{n} K_h^2(x_i - x_0)\mathbf{x}_i\mathbf{x}_i^T$$

And for p=1 and a=2, the estimated variance is

$$\hat{\mathbf{V}}_1(\hat{\boldsymbol{\beta}}) = \frac{\exp(\hat{\beta}_0)}{\left(1+\hat{\varphi}\exp(\hat{\beta}_0)\right)}\left(\mathbf{L}''_{1,h}(\hat{\boldsymbol{\beta}},x_0)\right)^{-1}\mathbf{S_n}\left(\mathbf{L}''_{1,h}(\hat{\boldsymbol{\beta}},x_0)\right)^{-1} \quad (13)$$

with $\mathbf{L}''_{1,h}(\hat{\boldsymbol{\beta}},x_0)$ is Hessian matrix evaluated at $\hat{\boldsymbol{\beta}}$.

Confidence Band of Regression Function

The confidence interval is an important tool for evaluating he estimator precision. But in nonparametric regression context, constructing such confidence interval is difficult because of non-negligible bias. However with our estimated bias and variance defined previously, we can construct a confidence interval or confidence band for regression function. Because the estimated bias and variance involves of higher order derivative curve, whose estimation can be unstable, they need to be averaged to prevent from abrupt change [8]. So define

$$\hat{b}_p^A(\hat{\beta}_j) = \sum_{i=1}^{n}\hat{b}_p(\hat{\beta}_j)K_i(x_i - x_0) \quad (14)$$

$$\hat{V}_p^A(\hat{\beta}_j) = \sum_{i=1}^{n} \hat{V}_p(\hat{\beta}_j) K_i(x_i - x_0) \qquad (15)$$

with

$$K_i(x_i - x_0) = \frac{K_h(x_i - x_0)}{\sum_{i=1}^{n} K_h(x_i - x_0)}$$

Under some regularity condition [13], the asymptotic distribution of the local likelihood estimator $\hat{\beta}_j$ at a point $x=x_0$

$$\frac{\beta_j - \hat{\beta}_j - \hat{b}_p{}^A(\hat{\beta}_j)}{\sqrt{\hat{V}_p{}^A(\hat{\beta}_j)}} \xrightarrow{D} N(0,1) \qquad (16)$$

So by invoking asymptotic normality the point wise confidence interval with $(1-\alpha)$ coverage probability β_j falls in random interval

$$\hat{\beta}_j - \hat{b}_p{}^A(\hat{\beta}_j) \pm z_{1-\alpha/2}\sqrt{\hat{V}_p{}^A(\hat{\beta}_j)} \qquad (17)$$

From (4) we have $s^j(x_0) = \beta_j j!,\ j = 0,1,\cdots,p$, so equivalently the confidence band for the regression function

$$\hat{s}^j(x_0) - \hat{b}_p{}^A(\hat{s}^j) \pm z_{1-\alpha/2}\sqrt{\hat{V}_p{}^A(\hat{s}^j)} \qquad (18)$$

However according to [8], the coverage probability of (17) or (18) can converge slowly to the nominal level $(1-\alpha)$. There are two reason for this. One is that the number of data point used to estimate the regression function at a particular point can be much smaller than n and the other is that the bias can possibly be non-negligible. It will show in our simulation next.

III. SIMULATION RESULT

We conducted some simulation with some purposes. First is to show the behaviour of the local likelihood estimator as the bandwidth parameter *h* and polynomial degree p are increased. Second the behaviour of confidence band of regression function before and after averaging process. And finnaly to show the coverage probability of the confidence band at nominal level 0.95. For that, we use sample of size n=100, 200 and 500. We generate *x* from Uniform distribution on [-1,1]. And from each x_i we generate the count response from generalized Poisson distribution with 3 different regression function

$$s_1(x) = 1 - 2x$$
$$s_3(x) = 3\sin(2x)$$

We also use Epanechnikov kernel for weight and the dispersion parameter is set to 0.2. Fig.(1) shows the behavior of the estimator when we increased the bandwidth parameter from h=0.005 to h=0.5 for regression function $s_1(x)$.

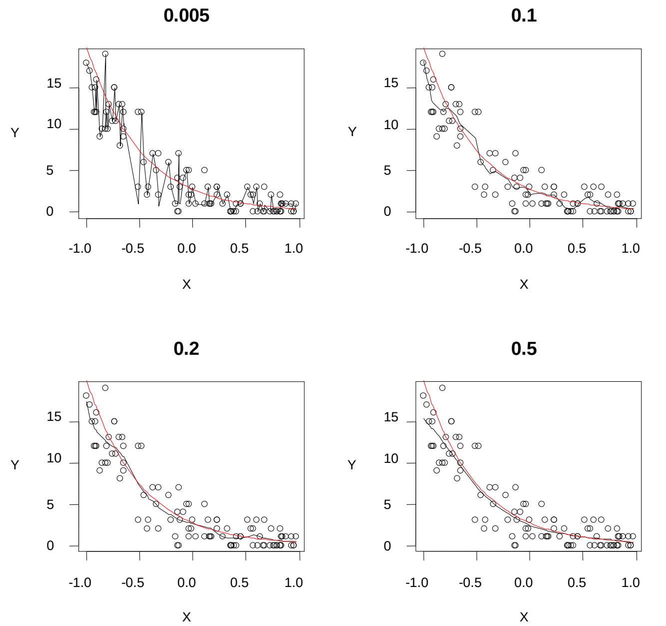

Fig 1. Estimated regression function (black) with h=0.005, h=0.1, h=0.2 and h=0.5 and the true regression function (red)

As we can see the estimated curve runs from the complex model (interpolation) to more simplest model. The ideal bandwidth or ideal model can be select by considering the value of CV which is minimum. The influence of the polynomial degree *p* can be seen in Fig. 2 as we use $s_3(x)$ for true regression function.

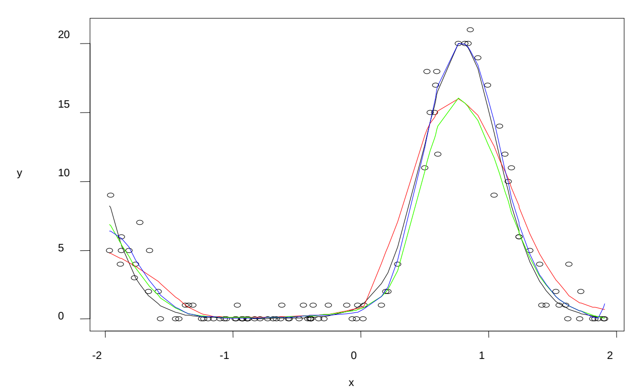

Fig 2. Estimated regression function with p=0 (red), p=1 (green), p=2 (blue) and the true regression function (black)

As we can see that the higher the degree of polynomial (p=2) then the estimator can reaches peak or valleys of the data better than p=0 or p=1, and can approximate the true regression function nicely.

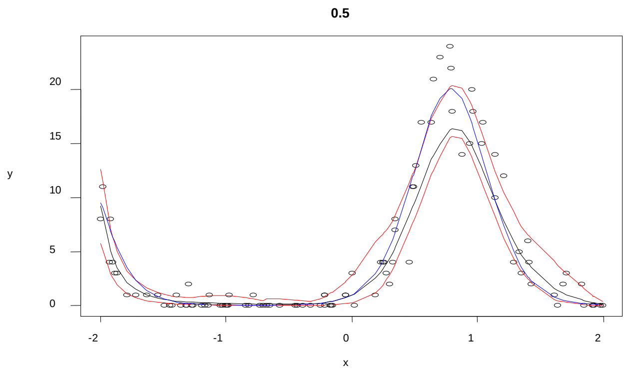

Fig. 3 The local likelihood estimator (black), 95% confidence band of true regression function (red) and the true regression function (blue)

Figure 3 is an example of 95% confidence band for true regression function with coverage probability 0.91. This coverage probability means that 91% of points in true regression function are included in the confidence band. The behavior of this coverage probability is that the estimator can reach

IV. Conclusion

A new approach of nonparametric regression for count response has been developed using local polynomial technique. We also derived estimated bias and variance of the estimator and constructing a confidence band for the unknown regression function. Simulation result shows that the performance of the estimator depends on the choice of bandwidth parameter h and polynomial degrees p. The confidence band of the regression function shows coverage probability near the nominal level 0.95 as expected

References

[1] E.T. Astuti and T. Yanagawa , "Testing Trend for Count Data with Extra-Poisson Variability", *Biometrics*,58, 2002, pp. 398-402.

[2] A. Ferrari, G. Letac and J.Y. Tourneret, J.Y, (2007), "Exponential families of mixed Poisson distributions", *Journal of Multivariate Analysis*, 98, 2007, pp. 1283 – 1292

[3] J.P. Boucher, M.Denuit and M. Guillen, M., "Risk Classification For Claim Counts: A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models", *North American Actuarial Journal*, 11:4, 2007, 110-131

[4] F.Famoye, "Restricted Generalized Poisson Regression", *Communication in Statistics-Theory and Methods*, 33, 2002, pp.1135-1154.

[5] J.A. Santos and M.M. Neves," A Local Maximum Likelihood Estimator for Poisson Regression", *Metrika*, 68,2008, pp. 257-270.

[6] Y.Terzi and M.A. Cengiz,M.A.," Using of Generalized Additive Model for Model Selection in Multiple Poisson Regression for Air Pollution Data", *Scientific Research and Essay*, 4:9, 2009, pp. 867-871.

[7] R. Tibshirani and T. Hastie," Local Likelihood Estimation", *Journal of the American Statistical Association*, 82:398,1987, pp. 559-567.

[8] J. Fan, M. Farmen and I. Gijbels, I, "Local Maximum Likelihood estimation and Inference", *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 60:3, 1998, pp. 591-608 .

[9] P. C. Consul and G. C. Jain, " A Generalization of the Poisson Distribution", *Technometrics*, 15:4, 1973, pp. 791-799.

[10] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Application*,. Chapman and Hall. London, 1997.

[11] L. Gyorfi, M. Kohler, A. Krzyzak and H.Walk, *A Distribution-Free Theory of Nonparametric Regression* , Springer, New York, 2002.

[12] E.T. Astuti, I.N. Budiantara, S. Sunaryo and M. Dokhi, "Bias dan Varians dari Estimator Linier Lokal pada Model Regresi Poisson Tergeneralisir", presented at Seminar Nasional Pascasarjana XII- ITS, Surabaya, July 12, 2012.

[13] J. Fan, N.E. Heckman and M.P. Wand, M.P.," Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions", *Journal of the American Statistical Association,* 90:429, 1995, pp. 141-150.