There are totally 3 free parameters in the COM-Poisson Skellam Distribution. Since each parameter is non-negative, we choose the log base to do the regression. Applying the vectorized regression method, I will give a detailed derivation and explanation of the VGLM (vector generalized linear model) Algorithm under the assumption of log-based regression. Although this algorithm already exist, there is no paper or reference about the details of it, and only the implementation in R software is available. So I tried to make a complete derivation and explanation of this algorithm used for my research.

Suppose in the general case, there are M parameter to estimate, namely, we can express the log-based regression formula in the following system of equations

 (1)

Each is the coefficient to estimate, is the kth independent variable (or explanation variable) of the ith sample supposing the sample size is n and each is the dependent variable. Then we can express the system of equations in (1) into the matrix form

 (2)

Here in the equation, we need to estimate the M\*(p+1) dimension matrix. We use the maximum likelihood method to estimate the parameter matrix. Namely, we need to find the matrix  to maximize the likelihood function

 (3)

Where n is again the sample size. To get the , obviously we need to employ some numerical method. Following the GLM Algorithm and extend it to the multivariate case, I use the Iterative Weighted Newton Method. Here in statistics, according to the way the information matrix is produced, this is named Newton-Raphson and Fisher Scoring.

For simplicity, we denote  as . From the basic formula of Newton method, we have

 (4)

Where means theth iteration。Multiply each side by, we have

 (5)

Taking the partial derivatives, we have

 (6)

Applying the chain rule, further we have：

 (7)

Where  is the block-diagonal matrix

 (8)

And

 (9)

From the Rao-Blackwell Theorem and its corollary in Statistics, if the relatively weak condition that the integration and differentiation can be exchanged is satisfied, we have

 (10)

From the Tonelli-Fubini Theorem in Real Analysis, this condition is easy to be satisfied. In addition, taking the expectation of (9), we will get a variance-covariance matrix, which is semi-positive-definite. Moreover, the vector we get by taking the first order derivative with respect to each coefficient during the iteration process, is always 0. Usually we call this the score vector. Then we have

 (11)

From (7) and (10), we have

 (12)

Where  (13)

From the part  in (13), we find that it is the solution of the weight least square solution in linear regression. More specifically, it is the solution when we try to minimize  with respect to.

The matrix in (9) is called the information matrix in statistics. The key part of this algorithm is to use the expected information matrix as a substitution for the observed information matrix. This method is sometimes called Fisher-Scoring in regression analysis. The advantage of computing the expected information matrix instead of the observed one is that usually, the matrix will have some good properties such as positive-definite, therefore making the computation easier. The disadvantage is that it will unavoidably bring about some bias. However, by increasing the sample size, we can avoid such deficiency. A lot of non-parametric method, such as Bootstrap, Jackknife and Permutations can be used to efficiently increase the sample size by simulation. Here, the method used in VGLM is to generate random numbers belong to the distribution with the parameters fitted in each iteration step. However, in practice, I found that if too large volume of random numbers is generated, the computation will be very slow. Therefore, we need to control that in the algorithm development. I write the random number generating program for the COM-Poisson Distribution. Please see the corresponding R file.

Now, I’m able to describe the VGLM Algorithm procedure,

1 Compute the maximum likelihood function for the distribution.

2 Pick an initial coefficient matrix.

3 Compute the dependent parameter vector, in the first iteration step,.

4 Compute the inverse link function, under the log-base, it is



5 From the vector computed in, compute the expected information matrix by generating random numbers belong to the distribution with this parameter vector. Notice that the random number is y is part in the distribution and from equation (10), we only need to compute



Then we can see there is “random part” in. By generating such random numbers, we can get the expected information matrix.

6 Get  from (13).

7 Get  from (12), here we can employ the straightforward method of using calculating the solution to the weighted least square linear regression problem.

8 Go back to step 3, calculate the vector, and based on this we can compute the value of the likelihood function. After this, we can easily get the difference of the likelihood functions in iteration step i and i+1. If the difference is smaller than 0, terminate the algorithm. If not, go on the iteration until the likelihood function converge.

Then after this algorithm is clear, I can derive the necessary properties of the COM-Poisson Skellam family function, and use this algorithm to do my own regression analysis.

References:

Peter MacCullagh and Nelder (1997). Generalized Linear Models, 2nd ed. Chapman &Hall/CRC, Yee, T. W. and Hastie, T. J. (2003) Reduced-rank vector generalized linear models. *Statistical Modelling*, **3**, 15–41.

Yee, T. W. and Wild, C. J. (1996) Vector generalized additive models. *Journal of the Royal Statistical Society, Series B, Methodological*, **58**, 481–493.