For counting data, we are all familiar with the Poisson distribution, and its mass function is:



The distribution of the difference of the two independent Poisson distribution samples is called Skellam distribution in statistics. Suppose, we say with the mass function



Where  is a modified Bessel function of first kind in PDE.

As us well known by statisticians, the assumption of Poisson distribution, where the mean must equal to the variance, is rarely met. To overcome this deficiency, negative-binomial distribution can be used to model the situation where the variance is greater than the mean. However, under some circumstances the sample variance is smaller than the mean. Professor Shumeli proposed a distribution with the mass function



Where is called the normalizing constant.

This is called the Com-Poisson distribution. is called the dispersion parameter. Some properties of this distribution is useful for my research. Namely, then are







Since this distribution overcomes the deficiency of Poisson distribution, my research advisor, prof. Sellers proposed the distribution to overcome the deficiency of Skellam distribution based on this.

Assuming that, where

She calls this distribution the Com-Poisson Skellam distribution. The mass function is



Where  is a generalized form of the Bessel function of the first kind.

For all the details and properties of these distributions, please read the reference I list at the end of this writing.

My work in the research, with my advisor prof. Sellers is to find a method to do the regression analysis, when the response variable group can be considered the difference of two groups of counting data with same dispersion parameter but different means.

For the Com-Poisson Skellam distribution, there are totally 3 parameters. To model these three parameters, I find that a good way is to use the vector generalized linear regression method.

Since I’ve explained the detail and give the steps in implementing VGLM algorithm in the other scratch writing, I only need to give the properties I derived from the Com-Poisson Skellam distribution. By applying or “embedding” these properties into VGLM algorithm, I can get implement the desired regression analysis.

The likelihood function of the Com-Skellam mass function is



Then I need to derive the expression for the score vector. The first element is



Here I used the property of Com-Poisson distribution that



Also



Similarly, we get the expression for the second element in the score vector



For the third element, we have



Here in the derivation I used the property mentioned in prof. Seller’s paper 

Finally, we have the expression for the score vectoras



Until now, I have the expression for the likelihood function and the score vector expression of the Com-Poisson-Skellam distribution. Only these two, and the random number generating program is need to estimate the coefficient in the Com-Poisson Skellam regression.

I write the R family function comskellam.R and other supporting functions. It belongs to the family function of “VGAM” package in R.

I will give a sample program of doing the Com-Poisson Skellam regression by generated data in R:

> sdata <- data.frame(x2 = runif(nn <- 20))

> sdata <- transform(sdata, mu1 = exp(1+x2), mu2 = exp(x2),mu3=1)

> sdata <- transform(sdata, y = rcomskellam(nn, mu1, mu2))

> fit1 <- vglm(y ~ x2, comskellam, sdata, trace = TRUE, crit = "c")

VGLM linear loop 1 : coefficients =

0.34760104164, -0.92172896832, 0.03727165097, 2.34909383275,

3.25104683721, 0.00183513074

VGLM linear loop 2 : coefficients =

0.634873817, -0.274779297, 0.190708836, 1.614285643, 1.708767730,

-0.148772258

VGLM linear loop 3 : coefficients =

0.276239411, -1.118843830, 0.410244214, 2.584974944, 3.581778420,

-0.349445965

VGLM linear loop 4 : coefficients =

1.830397467, 1.048701394, 0.599621637, 0.528020034, 0.407706516,

-0.539755704

> fit1 <- vglm(y ~ x2, comskellam, sdata, trace = TRUE)

VGLM linear loop 1 : loglikelihood = -48.7773

VGLM linear loop 2 : loglikelihood = -48.2352

Taking a modified step..

VGLM linear loop 2 : loglikelihood = -47.5262

VGLM linear loop 3 : loglikelihood = -47.0608

Taking a modified step...........

Warning message:

In eval(expr, envir, enclos) :

iterations terminated because half-step sizes are very small

<environment: namespace:VGAM>

References:

J.G Skellam, the frequency distribution between two Poisson variates belonging to different populations, J.Royal Stat. Soc. 109(3) (1946), 296.

Kimberly F Sellers and Galit Shmueli. (2010). A flexible regression model for count data. The Annals of Applied Statistics. 2010. Vol.4, No2, 943-961

Kimberly F Sellers.(2012). A distribution describing differences in count data containing common distribution levels. Advances and Applications in Statistical Sciences. 2012, under review.

Kimberly F Sellers. (2010) Supplementary Material for “a Flexible Regression Model for Count Data”. Advances and Applications in Statistical Sciences. 2012, under review.