

# Subsampling in dispersed count model

March 23, 2021

Poisson model is widely used to model the count data. However, the Poisson model assumes the mean is equal to variance, which might not be the case for the real data in some cases. When the dispersion of the data exceeds that implied by the Poisson model, i.e.  $Var(Y) > E(Y)$ , we say the data is overdispersed. On the other hand, if  $Var(Y) < E(Y)$ , the data is underdispersed.

One natural way to handle overdispersion is change the Poisson model to the negative binomial model. A second way is to use the overdispersed Poisson model, which breaks the mean/variance relationship. Both negative binomial model and overdispersed Poisson model have their own limitations. The negative binomial model can only handle overdispersion case. Although, the overdispersed Poisson model can handle both overdispersed and underdispersed model, the distribution does not truly correspond to an exponential family. Conway-Maxwell Poisson model (CMP) is another way to handle both over- and underdispersion, and it corresponds to an exponential family p.m.f. However, normalization constant has no closed form in general.

In this manuscript, we will consider subsampling in these three models, ie. subsampling in 1) negative binomial model, 2) overdispersed Poisson model and 3) Conway-Maxwell Poisson model. So far, the nuisance parameter in each model is treated as a constant nuisance parameter. (Although we can jointly model the mean and dispersion parameter, as I did before)

To facilitate the presentation, denote the full data matrix by  $\mathcal{F}_n = (\mathbf{X}, \mathbf{y})$ , where  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$  is the covariate matrix and  $\mathbf{y} = (y_1, \dots, y_n)^T$  is the response vector. Assume  $(\mathbf{x}_i, y_i)$ 's are independent. Denote the sub-sampling distribution as  $\pi_i$  for all data points.

## 1 Negative Binomial Model

The p.m.f. for negative binomial of  $y_i$ :

$$f(y_i|\nu, \mu_i) = \frac{\Gamma(\nu + y_i)}{\Gamma(\nu)y_i!} \left( \frac{\mu_i}{\nu + \mu_i} \right)^{y_i} \left( \frac{\nu}{\nu + \mu_i} \right)^\nu$$

where the nuisance parameter  $\nu$  can be estimated as a nuisance parameter. The  $\mu_i$  is modeled by a log-linear model, i.e.  $\log \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ . Then, the log-likelihood for each  $y_i$  is:

$$l_i(\boldsymbol{\beta}) = \log \Gamma(\nu + y_i) - \log \Gamma(\nu) - \log y_i! + y_i \log \frac{\mu_i}{\nu + \mu_i} + \nu \log \frac{\nu}{\nu + \mu_i}$$

Therefore, the score function and the observed information for each  $y_i$  are:

$$\begin{aligned} \frac{\partial l_i}{\partial \boldsymbol{\beta}} &= (y_i - \mu_i) \frac{\nu \mathbf{x}_i}{\nu + \mu_i} \\ -\frac{\partial^2 l_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= \frac{\nu(\nu + y_i) \mu_i}{(\nu + \mu_i)^2} \mathbf{x}_i \mathbf{x}_i^T \end{aligned}$$

Denote  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_{MLE}$  and  $\hat{\mu}_i = \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$ . Let  $M_X = n^{-1} \sum_{i=1}^n -\frac{\partial^2 l_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$ , then under A-optimality criterion, the sub-sampling probabilities (SSPs)  $\pi_i^{mMSE} \propto \|M_X^{-1} \frac{\partial l_i}{\partial \boldsymbol{\beta}}\|$ , while under L-optimality criterion,  $\pi_i^{mVc} \propto \|\frac{\partial l_i}{\partial \boldsymbol{\beta}}\|$ . In other words,

$$\begin{aligned} \pi_i^{mMSE} &= \frac{|y_i - \hat{\mu}_i| \|M_x^{-1} \frac{\nu \mathbf{x}_i}{\nu + \hat{\mu}_i}\|}{\sum_{j=1}^n |y_j - \hat{\mu}_j| \|M_x^{-1} \frac{\nu \mathbf{x}_j}{\nu + \hat{\mu}_j}\|} \\ \pi_i^{mVc} &= \frac{|y_i - \hat{\mu}_i| \|\frac{\nu \mathbf{x}_i}{\nu + \hat{\mu}_i}\|}{\sum_{j=1}^n |y_j - \hat{\mu}_j| \|\frac{\nu \mathbf{x}_j}{\nu + \hat{\mu}_j}\|} \end{aligned}$$

## 2 Overdispersed Poisson Model

Define the quasi-likelihood for overdispersed Poisson for  $y_i$  as:

$$Q_i(\lambda_i, y_i) = \int_{y_i}^{\lambda_i} \frac{y_i - t}{\phi t} dt$$

where dispersion parameter  $\phi$  is nuisance. The  $\lambda_i$  is modeled by a log-linear model, i.e.  $\log \lambda_i = \mathbf{x}_i^T \boldsymbol{\beta}$ . Then,

$$\begin{aligned} \frac{\partial Q_i}{\partial \boldsymbol{\beta}} &= \phi^{-1} (y_i - \lambda_i) \mathbf{x}_i \\ -\frac{\partial^2 Q_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= \phi^{-1} \lambda_i \mathbf{x}_i \mathbf{x}_i^T \end{aligned}$$

Denote  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_{MLE}$  and  $\hat{\lambda}_i = \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$ . Let  $M_X = n^{-1} \sum_{i=1}^n \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i^T$ , then SSPs under A-optimality and L-optimality criteria are:

$$\begin{aligned} \pi_i^{mMSE} &= \frac{|y_i - \hat{\lambda}_i| \|M_x^{-1} \mathbf{x}_i\|}{\sum_{j=1}^n |y_j - \hat{\lambda}_j| \|M_x^{-1} \mathbf{x}_j\|} \\ \pi_i^{mVc} &= \frac{|y_i - \hat{\lambda}_i| \|\mathbf{x}_i\|}{\sum_{j=1}^n |y_j - \hat{\lambda}_j| \|\mathbf{x}_j\|} \end{aligned}$$

### 3 Conway-Maxwell Poisson (COM-Poisson, CMP) Model

The p.m.f. for Conway-Maxwell Poisson (CMP) for  $y_i$  is:

$$f(y_i|\nu, \lambda_i) = \frac{\lambda_i^{y_i}}{(y_i!)^\nu} \frac{1}{Z(\lambda_i, \nu)}$$

where  $\nu$  is the nuisance parameter, and  $Z(\lambda_i, \nu) = Z_i$  is the normalization constant, i.e.  $Z_i = \sum_{y=0}^{\infty} \frac{\lambda_i^y}{(y!)^\nu}$ , which doesn't have closed form in general. The domain for parameters is  $\lambda_i, \nu > 0$  and  $0 < \lambda_i < 1, \nu = 0$ . The parameter  $\nu$  controls the dispersion: 1) when  $\nu = 1$ , the CMP is Poisson distribution, 2) when  $\nu < 1$ , the distribution is over-dispersed and 3) when  $\nu > 1$ , the distribution is under-dispersed. When  $\nu \rightarrow \infty$ , the CMP approaches a Bernoulli distribution, while  $\nu = 0$ , it reduces to a geometric distribution.

The  $\lambda_i$  is modeled by a log-linear model, i.e.  $\log \lambda_i = \mathbf{x}_i^T \boldsymbol{\beta}$ . Then, the log-likelihood for each  $y_i$  is:

$$l_i(\boldsymbol{\beta}) = y_i \log(\lambda_i) - \log(y_i!)^\nu - \log(Z_i)$$

Therefore, the score function and observed information for each  $y_i$  are:

$$\begin{aligned} \frac{\partial l_i}{\partial \boldsymbol{\beta}} &= (y_i - E(Y_i)) \mathbf{x}_i \\ -\frac{\partial^2 l_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= \text{Var}(Y_i) \mathbf{x}_i \mathbf{x}_i^T \end{aligned}$$

where  $E(Y_i) = \frac{\partial \log(Z_i)}{\partial \log(\lambda_i)}$  and  $\text{Var}(Y_i) = \frac{\partial^2 \log(Z_i)}{\partial \log(\lambda_i)^2}$ . The expectation and variance has no closed form, and they are approximated by truncated summation in this manuscript.

Denote  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_{MLE}$ ,  $\hat{\lambda}_i = \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$ ,  $\widehat{E(Y_i)} = E(Y_i|\mathbf{x}_i, \hat{\boldsymbol{\beta}})$  and  $\widehat{\text{Var}(Y_i)} = \text{Var}(Y_i|\mathbf{x}_i, \hat{\boldsymbol{\beta}})$ . Denote  $M_x = n^{-1} \sum_{i=1}^n \text{Var}(Y_i) \mathbf{x}_i \mathbf{x}_i^T$ . Then the SSPs under A-optimality and L-optimality criteria are:

$$\begin{aligned} \pi_i^{mMSE} &= \frac{|y_i - \widehat{E(Y_i)}| \|\mathbf{M}_x^{-1} \mathbf{x}_i\|}{\sum_{j=1}^n |y_j - \widehat{E(Y_j)}| \|\mathbf{M}_x^{-1} \mathbf{x}_j\|} \\ \pi_i^{mVc} &= \frac{|y_i - \widehat{E(Y_i)}| \|\mathbf{x}_i\|}{\sum_{j=1}^n |y_j - \widehat{E(Y_j)}| \|\mathbf{x}_j\|} \end{aligned}$$

When  $\nu = 1$ , the SSPs are reduced to Poisson and overdispersed Poisson SSPs.