Generalized Linear Models (GLM)

(Beyond the standard GLMs)

R. Bellio & N. Torelli

Spring 2018

University of Udine & University of Trieste

More on some specific GLMs

Beyond GLMs

More on some specific GLMs

Offset in Poisson regression/modelling rates

There are many cases where the observed counts should be interpreted as relative to some baseline.

- The parameter λ of a Poisson regression model can be interpreted with reference to a specific unit of time or space. And the number of cases y_i in a process is then Poisson($e_i\lambda$). λ is the rate of the process and e_i is the exposure.
- Suppose we want to model the number of those with a specific disease within an geographical area, it depends clearly on the rate and on the number of units living in that area.
- One could then model the rate λ_i/e_i. In this case

$$\frac{\lambda_i}{e_i} = exp(\boldsymbol{x}_i^T\hat{\beta}) \rightarrow log(\frac{\lambda_i}{e_i}) = \boldsymbol{x}_i^T\hat{\beta} \rightarrow log(\lambda_i) = \boldsymbol{x}_i^T\hat{\beta} + log(e_i)$$

- Then y_i is again modelled as in Poisson rgeression by specifying its mean as λ_i=exp$X_i\beta$ but also $log(e_i)$ is introduced into the model.
- $log(e_i)$ is included in the model as a predictor whose coefficient is fixed to 1 and it is called the offset (in `R` introduce in `glm  the optionoffset`=log(....)‘.

Overdispersed count data

- Poisson regression models in a GLM context imply that the dispersion parameter is fixed to 1.
- Variance function is then functionally related to the mean function (it is actually the same).
- For a Poisson models the standardized residuals are

$$z_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$$

  where $\hat{\mu}_i = \exp^{x_i^T \hat{\beta}}$
- If the Poisson model holds the z_i are approximately independent and will have mean equal to 0 and variance equal to 1. Approximately $\sum_{i=1}^{n} z_i^2$ is a χ^2_{n-k} distribution if the model holds. This can be used for detecting overdispertion.
- Considering a model for counts with overdispersion will be more realistic in many practical cases

Logistic regression

Binomial regression with logit link is by far the most popular.
some reasons are:

1. it is the canonical link
2. it can be interpreted as log-odds of probability of success
3. statistical analysis is simplified
4. appropriate for data collected in a retrospective study or when oversampling one of the classes.

The last property deserve some more words since it could be relevant when this model is used for classification (prediction).

It states that if we we oversample one of the two classes (also typically done in retrospective studies) the estimates of the β_j, $(j = 1, 2, \ldots, p-1)$ parameters are unchanged with the exception of the intercept β_0.

This can bias, but in a predictable direction, the estimated probability of success.

Overdispersion in binomial regression

- Also in the case of Binomial regression the mean $\mu = np$ completely defines the variance function $v(\mu) = np(1-p)$.
- Also in this context data can appear to have more variance than expected under binomial variation.
- Again, one can use standardized residuals to reveal this.
- The simplest and more common mechanism that give rise to overdispersion is clustering in the population. In the case of a binomial response assume data are clustered and that cluster size k is fixed. Since we have m individuals in the sample, there are m/k clusters. Now if we assume that in each cluster the number of successes Z_i follows a $Bi(k, \pi_i)$ which varies across clusters. The response $Y = Z_1 + Z_2 + \cdots + Z_{m/k}$.

Now let's $E(\pi_i) = \pi$ and $var(\pi_i) = \tau^2\pi(1-\pi)$ the mean and the variance of Y are equal to $E(Y) = m\pi \quad var(Y) = m\phi\pi(1-\pi)$

- Note that overdispersion cannot arise in case of a Bernoulli model. And then can be an issue to take into account only with grouped data.

Quasi-likelihood

- For LMs the method of LS allows to obtain estimates of the regression parameters without the specification of a probabilistic model.

- The method of LS requires only the specification of the relation between the expected value of the response variable and the linear predictor, and the specification of the variance of the error term, which is not related to the expected value:
$$E(Y_i) = \mu_i = \eta_i \quad var(Y_i) = \sigma^2$$
- Also for the GLMs it is possible to specify only these two relations (assuming that the variance function $V(\mu_i)$ is known).
- Indeed, the likelihood equation for β
$$\sum_{i=1}^{n} \frac{y_i - \mu_i}{V(\mu_i)g'(\mu_i)} x_{ij} = 0 \ , \quad j = 1, \dots, p \ ,$$
is an unbiased estimating equation provided that $E(Y_i) = \mu_i = g^{-1}(\eta_i)$.
- In other words, this means that the parametric assumption $Y_i \sim EF(\cdot, \phi)$ could not even be satisfied. Only the assumption about expectations is essential:$\mu_i = E(Y_i) = g^{-1}(\eta_i)$
- the only distributional feature that must be known in order to calculate the estimating equation is the variance function $V(\mu)$.

Quasi-likelihood

- Under suitable regularity conditions, the likelihood equations for a GLM give estimates for the coefficients β which maintain several properties, also if the parametric assumptions of Y_i are substituted with weaker **second order assumptions**:
    1. $g(\mu_i) = g(E(Y_i)) = \eta_i, \quad i = 1, \ldots, n,$
    2. $var(Y_i) = \phi V(\mu_i), \quad i = 1, \ldots, n,$
    3. $cov(Y_i, Y_j) = 0$, if $i \neq j$.
- The semi-parametric statistical model specified by assumptions 1–3 is called **quasi-likelihood model**.
- If $V(\mu) = 1$ and $g(\mu) = \mu$, the assumptions 1–3 match the usual second order assumptions of the classical LM.
- On the other hand, if $V(\mu) = \mu^2$ we obtain a multiplicative model, $Y_i = \mu_i \epsilon_i$, with $E(\epsilon_i) = 1$ and $var(\epsilon_i) = \phi$.

Quasi-likelihood

- Gauss-Markov (BLUE) optimality of LS extends to quasi-likelihood estimates and it has minimum asymptotic variance among estimating equations that are linear (in Y) and unbiased
- Indeed, the likelihood equation for β

$$q(y;\beta) = \sum_{i=1}^{n} q(y_i;\beta) = \sum_{i=1}^{n} \frac{y_i - \mu_i}{V(\mu_i)g'(\mu_i)} x_{ij} = 0 \ , \quad j = 1, \ldots, p \ ,$$

behaves like a score vector. Specifically:

$$E(q(Y;\beta)) = 0, \quad \text{and} \quad var(q(Y;\beta)) = -E(\partial q(Y;\beta)/\partial\beta) \ .$$

- Quasi likelihood estimators shares many properties of a proper likelihood : the quasi-MLE β is asymptotically normal, the quasi-likelihood ratio statistic has a null chi-squared distribution.

Quasi-likelihood and overdispersion

- The assumptions 1–3 offer an increase in flexibility with respect to the usual parametric specifications based, respectively, on the Poisson, binomial or exponential distributions.
- In practice, there are situations in which the dispersion parameter does not agree with the assumed exponential family.
- For example, for the binomial or Poisson distributions we have $\phi = 1$, but data could show agreement with $\phi > 1$.
- In this case we have *overdispersion*, i.e. the variance of Y is greater than its theoretical value, and it is more plausible to assume $var(Y_i) = \phi V(\mu_i)$, with $\phi > 1$. For example, for proportions, it can be assumed that $var(Y) = \phi n\pi(1-\pi) > n\pi(1-\pi)$, with $\phi > 1$, where $n\pi(1-\pi)$ is the variance of a binomial distribution.
- In general, the quasi-likelihood approach allows to deal with *overdispersion problems*: it is possible to specify $var(Y_i)$ so that there is more variability with respect to the exponential family.
- The case of *underdispersion*, i.e. $\phi < 1$, is less important in applications, but can be dealt with under the quasi-likelihood model

using quasi-likelihood in ‘glm

- When estimating a GLM by using quasi-likelihood one can use the same variance function deriving from a Binomial or from a Poisson model and using the canonical link for those models. In R this leads to a specification of the `family` that is called `quasibinomial` or `quasipoisson`.
- Estimates of the β are the same since the estimating equations do not change
- But standard errors of estimates will change since a value different from 1 is estimated for ϕ. In `quasipoisson` one should take into account that variance is modelled as $Var(y_i) = \phi\mu_i$
- The parameter ϕ can be also estimated as

$$\hat{\phi} = \frac{1}{n-p} \sum \frac{(y_1 - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

- In those cases also the Deviance of the model has to be corrected because it is computed assuming $\phi = 1$. The deviance reported has to be divided by $\hat{\phi}$

Beyond GLMs

Negative binomial regression

- It is an alternative model that can be considered when data exhibits overdispersion. Its probability function is

$$Pr(Z = z) = \binom{z-1}{k-1} p^k (1-p)^{z-k} \qquad z = k, k+1, \ldots$$

where $E(Z) = k(1-p)/p$ and $Var(Z)=k(1-p)/p^2

- compared with Poisson
  - since it has an extra parameter it proves to be more flexible
  - mean is larger than variance and then it accommodates overdispersion
  - Poisson is a limiting case of negative binomial (if $p \to 1$ e $k \to 0$ then $rp \to \lambda$)
- Recall that negative binomial emerges as a mixture of Poisson when each unit Y is Poisson with mean λ and λ are drown from a *Gamma* distribution.

Negative Binomial regression

- When building a model for Negative Binomial a different parametrization is more appropriate, by defining $Y = Z - k$ and $p = \frac{1}{1+\alpha}$
- $$Pr(Y = y) = \binom{y+k-1}{k-1} \frac{\alpha^y}{(1+\alpha)^{y+k}} \qquad y = 0, 1, \ldots$$
- Then
  - $E(Y) = \mu = k\alpha$
  - e che $Var(Y) = k\alpha + k\alpha^2 = \mu + \mu^2/k$
- and the following link can be used $\log \frac{\alpha}{1+\alpha} = \log \frac{\mu}{k+\mu}$
- Note that BiN is not a member of the EF. Then it is not a proper GLM. Classical IWLS cannot be used as it is.
- In `R` a specific function has to be used: `glm.nb(...)` included in the package MASS

Zero inflated Poisson

- Zero inflation means that we have far more zeros than what would be expected for a Poisson or BiN distribution
- Ignoring zero inflation can have two consequences;
  - the estimated parameters and standard errors may be biased, and + the excessive number of zeros can cause overdispersion
- A possible model hypotizes that the observed counts derive from a mixture of two populations:
  - for a part of the population (with probability p) Y can only be 0
  - for the remaining part (with probability $1 - p$) Y is distributed as a Poisson or a BiN.
- Distribution of counts is then, in case of Poisson

$$\begin{aligned} P(y_i = 0) = & \quad p_i + (1 - p_i)e^{-\mu_i} \\ P(y_i = y_i | y_i > 0) = & \quad (1 - p_i)\frac{\mu_i^{y_i} e^{\mu_i}}{y_i!} \end{aligned}$$

- Covariates can be introduced, like in GLM, for modelling p_i and μ_i