Poisson Regression

- Recall: in binomial GLM, observe data

  $$(x_{i1}, x_{i2}, \ldots, x_{im}, y_i),\ i = 1, 2, \ldots, k$$
  $$\Leftrightarrow (\mathbf{x}_i, y_i) \Leftrightarrow (\mathbf{x}, y_{\mathbf{x}})$$

  where $y_{\mathbf{x}}$ is bounded by $n_{\mathbf{x}}$ = number of total trials at $\mathbf{x}$, i.e.,

  $0 \le y_{\mathbf{x}} \le n_{\mathbf{x}}$ and $n_{\mathbf{x}}$ is a *fixed* and *known* number

  - **Q**: what if the upper limit to $y_{\mathbf{x}}$ is infinite or effectively so?
  - Some examples of such $y_{\mathbf{x}}$:
    - number of incidents involving damage to ships over a give period of time
    - radiation counts as measured in, say, particles per second by a Geiger counter
    - number of species of tortoise found on 30 islands
  - For such $y_{\mathbf{x}}$, can use Poisson GLM or negative binomial GLM
  - If $y_{\mathbf{x}}$ is sufficiently large

    $\Rightarrow$ may fit a Normal linear model: $y_{\mathbf{x}} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

NTHU STAT 5230, 2025, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- If $Y_{\mathbf{x}}$ is Poisson with mean $\mu_{\mathbf{x}}$ (>0), then

  $$P(Y_{\mathbf{x}} = y_{\mathbf{x}}) = \frac{e^{-\mu_{\mathbf{x}}} \times \mu_{\mathbf{x}}^{y_{\mathbf{x}}}}{y_{\mathbf{x}}!}, \quad y_{\mathbf{x}} = 0, 1, 2, \ldots$$

  - $E(Y_{\mathbf{x}}) = \mu_{\mathbf{x}}$ and $Var(Y_{\mathbf{x}}) = \mu_{\mathbf{x}}$
  - $Y_i \sim$ Poisson(μ_i), $i=1,\ldots,t$ and independent, then $Y_1+\ldots+Y_t \sim$ Poisson($\mu_1+\ldots+\mu_t$)

    $\Rightarrow$ useful if only aggregated data is observed, say $y_{\mathbf{x}}$ is aggregated over x_m
  - Some rationales for using Poisson for count response
    - When $y_{\mathbf{x}} \sim B(n_{\mathbf{x}}, p_{\mathbf{x}})$ and $n_{\mathbf{x}}$ is large while $p_{\mathbf{x}}$ is small

      $$\binom{n_{\mathbf{x}}}{y_{\mathbf{x}}} p_{\mathbf{x}}^{y_{\mathbf{x}}} (1 - p_{\mathbf{x}})^{n_{\mathbf{x}} - y_{\mathbf{x}}} \approx \frac{e^{-n_{\mathbf{x}} p_{\mathbf{x}}} (n_{\mathbf{x}} p_{\mathbf{x}})^{y_{\mathbf{x}}}}{y_{\mathbf{x}}!} \Rightarrow \mu_{\mathbf{x}} = n_{\mathbf{x}} p_{\mathbf{x}}$$

      - example: incidence of rare cancer in a large sample
      - For small $p_{\mathbf{x}}$'s,

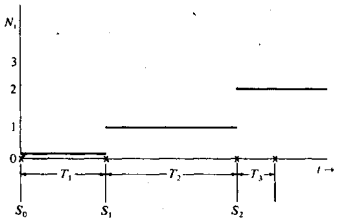
      $$\eta_{\mathbf{x}} = \text{logit}(p_{\mathbf{x}}) = \log\left(\frac{p_{\mathbf{x}}}{1 - p_{\mathbf{x}}}\right) \approx \log(p_{\mathbf{x}}) = \log(\mu_{\mathbf{x}}) - \log(n_{\mathbf{x}})$$

⇒ in this case, the use of the Poisson GLM with a log link is comparable to the binomial GLM with a logit link, especially when $n_{\mathbf{x}}$'s are similar.

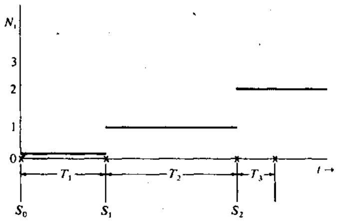
- Suppose that
  1. Probability of an event occurring once in a given (short) time interval is proportion to the length of the interval
  2. The numbers of events in two disjoint time intervals are independent

  Then, the number of events in a specified time interval will be Poisson distributed

  - Example: number of incoming telephone calls/earthquake in $[t_1, t_2]$



- Suppose that the times between any two adjacent events are i.i.d. as exponential distribution ⇒ number of events in a given time period is Poisson distributed.



NTHU STAT 5230, 2025, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- 3 components of Poisson GLM:
  - $y_{\mathbf{x}} \sim \text{Poisson}(\mu_{\mathbf{x}})$
  - $X\beta = \sum_{j=1}^{p} \beta_j \times h_j(X_1, \ldots, X_m) \equiv \eta_{\mathbf{x}}(\beta)$
    - For i=1, …, k, denote the ith row of X by $\mathbf{h}_i = (h_1(\mathbf{x}_i), \ldots, h_p(\mathbf{x}_i))^T \Rightarrow \eta_{\mathbf{x}_i} = \mathbf{h}_i{}^T\beta$
  - *link function* g: g monotone and differential, and $\eta_{\mathbf{x}} = g(\mu_{\mathbf{x}})$ [for Poisson GLM, g: $(0, \infty) \rightarrow (-\infty, \infty)$]
    - *Log-linear model*: $\eta_{\mathbf{x}} = \log(\mu_{\mathbf{x}}) \Leftrightarrow \mu_{\mathbf{x}} = \exp(\eta_{\mathbf{x}})$
- Log-likelihood of log-linear model:

$$\begin{aligned} l(\beta) &= log\left(\prod_{i=1}^{k} \frac{e^{-\mu_{\mathbf{x}_i}} \mu_{\mathbf{x}_i}{}^{y_i}}{y_i!}\right) \\ &= \sum_{i=1}^{k} \left[y_i \log(\mu_{\mathbf{x}_i}) - \mu_{\mathbf{x}_i} - \log(y_i!) \right] \\ &= \sum_{i=1}^{k} \left[y_i \, \eta_{\mathbf{x}_i} - \exp(\eta_{\mathbf{x}_i}) - \log(y_i!) \right] \\ &= \sum_{i=1}^{k} \left[y_i \, (\mathbf{h}_i{}^T\beta) - \exp(\mathbf{h}_i{}^T\beta) - \log(y_i!) \right] \end{aligned}$$

- Estimation (MLE) of β:
  - Partial derivative of $l(\beta)$: $$\frac{\partial}{\partial \beta_j} l(\beta) = \sum_{i=1}^{k} \left[y_i \mathbf{h}_{ij} - \exp(\mathbf{h}_i{}^T \beta)\, \mathbf{h}_{ij} \right] = \sum_{i=1}^{k} \left[y_i - \exp(\mathbf{h}_i{}^T \beta) \right] \mathbf{h}_{ij}$$
  - Set $\partial l(\beta)/\partial \beta_j = 0$, $\forall\, j$. The MLE $\hat{\beta}$ is the solution to: $$0 = \sum_{i=1}^{k} \mathbf{h}_{ij}\, y_i - \sum_{i=1}^{k} \mathbf{h}_{ij}\, \hat{\boldsymbol{\mu}}_i, \quad j = 1, \ldots, p.$$ $$\Leftrightarrow X^T Y - X^T \hat{\boldsymbol{\mu}} = 0 \Leftrightarrow X^T Y = X^T \hat{\boldsymbol{\mu}}$$ where $\hat{\boldsymbol{\mu}} = \exp(\hat{\boldsymbol{\eta}}) = \exp(X^T \hat{\beta})$
    - The link function having the property $X^T \hat{\boldsymbol{\mu}} = X^T Y$ is known as *canonical link*.
    - The canonical link for Normal linear model is identity link: $E(Y) = \boldsymbol{\mu} = \boldsymbol{\eta} = X\beta$, and for binomial GLM is logit link.
  - For Poisson GLM, *no explicit formula* for $\hat{\beta}$ $\Rightarrow$ must resort to numerical methods to find a approximated solution (same numerical method, IRWLS, as in binomial GLM, future lecture)

NTHU STAT 5230, 2025, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- Deviance of log-linear model:
  - For a saturated model L^*, the MLE of $\mu_{\mathbf{x}}$ is $y_{\mathbf{x}}$;
  - For a model S with s ($\leq k$) parameters, denote its MLE of $\mu_{\mathbf{x}}$ by $\hat{\mu}_{\mathbf{x},S}$. Then, the deviance of S is (considering the likelihood ratio test statistic of H_0:S vs. H_1:$L^*\backslash S$): $$D_S = 2\left(l_{L^*} - l_S \right) = 2 \sum_{i=1}^{k} \left[y_i \log\left(\frac{y_i}{\hat{\mu}_{i,S}} \right) - (y_i - \hat{\mu}_{i,S}) \right]$$
  - $D_S \overset{a}{\sim} \chi^2_{k-s}$ when S is the true model

    $\Rightarrow$ can do goodness-of-fit test (Note. it is because $\mu_{\mathbf{x}}=E(y_{\mathbf{x}})=Var(y_{\mathbf{x}})$)

    (**Q**: what should tend to ∞ for this asymptotics?)
  - The D is known as G^2-statistic in the analysis of contingency table (future lecture)

- Testing for log-linear model:
  - Consider two nested models S (s parameters) $\subset L$ (l parameters)
  - Test statistic and null distribution D for H_0:S vs. H_1:$L\backslash S$

$$0 \leq D_S - D_L \Rightarrow \overset{a}{\sim} \chi^2_{df_S - df_L} \text{ (under } H_0\text{)}$$

where $df_S = k - s$ and $df_L = k - l$

(cf., same asymptotic properties as for binomial GLM)

$\Leftrightarrow$ profile likelihood confidence interval for individual parameter

  - An alternative for testing significance of individual parameter: Wald test statistic for H_0: $\beta_i = c$ vs. H_1: $\beta_i \neq c$

$$z_i = (\hat{\beta}_i - c)/se(\hat{\beta}_i) \Rightarrow \overset{a}{\sim} N(0,1) \text{ (under } H_0\text{)}$$

$\Leftrightarrow$ $100(1-\alpha)\%$ confidence interval: $\hat{\beta}_i \pm z(\alpha/2) \times se(\hat{\beta}_i)$

- Pearson X^2 statistic of S (an alternative goodness-of-fit measure):

$$X^2_S = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{(y_i - \hat{\mu}_{i,S})^2}{\hat{\mu}_{i,S}}$$

NTHU STAT 5230, 2025, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

  - Pearson X^2 typically close in size to the deviance because:

$$\begin{aligned} D &= 2\sum_{i=1}^{k}[y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)] \\ &= 2\sum_{i=1}^{k}\{[(y_i - \hat{\mu}_i) + (y_i - \hat{\mu}_i)^2/(2\hat{\mu}_i) + \cdots] - (y_i - \hat{\mu}_i)\} \\ &\approx \sum_{i=1}^{k} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} = X^2 \end{aligned}$$

$\Rightarrow$ Pearson X^2 often used in the same manner as the deviance

- Overdispersion under a model S, i.e., $Var(y_{\mathbf{x},S}) \gg E(y_{\mathbf{x},S})$
  - Overdispersion $\Rightarrow$ large D_S (goodness-of-fit test rejected)
  - Other possible reasons causing large D_S should be examined before concluding there is an overdispersion. For example: (1) outliers, (2) wrong $\boldsymbol{X\beta}$ structure, …
  - A mechanism that can explain why overdispersion appears:

$$y_{\mathbf{x}}|\lambda_{\mathbf{x}} \sim \text{Poisson}(\lambda_{\mathbf{x}}), \text{ but } \lambda_{\mathbf{x}} \text{ is a random variable}$$

    - Example: tendency to fail for a product may vary from batch to batch even though they have same $\mathbf{x}$.

- Suppose that $\lambda_{\mathbf{x}} \sim$ gamma distribution with $E(\lambda_{\mathbf{x}})=\mu_{\mathbf{x}}$ and $Var(\lambda_{\mathbf{x}})=\mu_{\mathbf{x}}/\phi_{\mathbf{x}}$
- Then (exercise), $y_{\mathbf{x}} \sim$ negative binomial, $y_{\mathbf{x}} = 0, 1, 2, \ldots,$ with

  $E(y_{\mathbf{x}})=\mu_{\mathbf{x}}$ but $Var(y_{\mathbf{x}})=\mu_{\mathbf{x}}\times(1+\phi_{\mathbf{x}})/\phi_{\mathbf{x}} \geq \mu_{\mathbf{x}}=E(y_{\mathbf{x}})$

➢ For overdispersion cases, we can model $y_{\mathbf{x}}$ as:

- When mechanism known ⇒ can model $y_{\mathbf{x}}$ as a negative binomial response (or other more flexible distribution)
- When mechanism unknown ⇒ can add a dispersion parameter σ^2, which is an unknown constant for all $\mathbf{x}$'s, such that:

$$Var(y_{\mathbf{x}}) = \sigma^2 \times E(y_{\mathbf{x}}) = \sigma^2 \mu_{\mathbf{x}}$$

  - ▫ $\sigma^2 = 1$ ⇒ the regular Poisson GLM; $\sigma^2 > 1$ ⇒ overdispersion; $\sigma^2 < 1$ ⇒ underdispersion
  - ▫ Estimation of σ^2:

$$\hat{\sigma}^2 = \frac{X^2}{k-s} = \frac{\sum_i \left[(y_i - \hat{\mu}_i)^2/\hat{\mu}_i\right]}{k-s}$$

NTHU STAT 5230, 2025, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- ◆ Note: the estimation of β is unaffected by σ^2 ⇒ choosing a dispersion parameter other than 1 has no effect on the estimation procedure (IRWLS) of β (cf., similar to Normal LM and binomial GLM)
- ◆ The standard error of $\hat{\beta}_i$ should be adjusted for

  ⇒ scale up the standard error by a factor of $\hat{\sigma}$
- ◆ The z-statistic (Wald test) and its corresponding confidence interval should use the scaled $se(\hat{\beta}_i)$

▫ When comparing two nested models, an F-test statistic:

$$\frac{(D_S - D_L)/(df_S - df_L)}{\hat{\sigma}_L^2} \Rightarrow \overset{a}{\sim} F_{df_S - df_L, df_L} \text{ (under } H_0{:}S)$$

should be used, rather than the chi-square test.

- ◆ The F-test is more reliable than the z-statistic

▫ No goodness-of-fit test is possible.

- Prediction of $\mu_{\mathbf{x}}$ at $\mathbf{x}_0 = (x_{01}, \ldots, x_{0m})^T$ under a log-linear model:

➢ Denote $\mathbf{h}_0 = (h_1(\mathbf{x}_0), \ldots, h_p(\mathbf{x}_0))^T$

- Recall: $\mu_{\mathbf{x}_0} \xrightarrow{g} \eta_{\mathbf{x}_0} = \mathbf{h}_0{}^T\beta$, and under log-linear model,

$$\eta_{\mathbf{x}_0} = \log(\mu_{\mathbf{x}_0}) \Leftrightarrow \mu_{\mathbf{x}_0} = \exp(\eta_{\mathbf{x}_0})$$

- $\hat{\eta}_{\mathbf{x}_0} = \mathbf{h}_0{}^T\hat{\beta}$ and $\hat{Var}(\hat{\eta}_{\mathbf{x}_0}) = \mathbf{h}_0{}^T\hat{\Sigma}\,\mathbf{h}_0$

  ($\hat{\Sigma}$ = estimated covariance matrix of $\hat{\beta}$)
- $\hat{\mu}_{\mathbf{x}_0} = \exp(\hat{\eta}_{\mathbf{x}_0})$
- The construction of confidence intervals for $\eta_{\mathbf{x}_0}$ and $\mu_{\mathbf{x}_0}$ follows a similar manner as in binomial GLM.
- Suppose that there is only an s-level categorical covariate (or other covariates are held fixed)
  - Denote the s categories by $1,\ldots,s$; their count observations by $y_1,\ldots,y_s$; their count parameters by $\mu_1,\ldots,\mu_s$
  - Can predict the probability that an observed event falls in the i^{th} category by $\hat{\mu}_i/(\hat{\mu}_1 + \hat{\mu}_2 + \cdots + \hat{\mu}_s)$, given $y_1,\ldots,y_s$ were collected under the same value of a size variable (next slide)

❖ **Reading**: Faraway (2006, 1st ed.), 3.1

Rate Model

- Data: observe the number of events $y_{\mathbf{x}}$ at $\mathbf{x}$, and assume $y_{\mathbf{x}} \sim \text{Poisson}(\mu_{\mathbf{x}})$, where $y_{\mathbf{x}}$ (or $\mu_{\mathbf{x}}$) depends on a size variable $t_{\mathbf{x}}$ (which is *observed*) that determines the number of opportunities for the events to occur
  - Some examples:
    - $y_{\mathbf{x}}$ = number of burglaries reported in different cities
      $t_{\mathbf{x}}$ = number of households in these cities
    - $y_{\mathbf{x}}$ = number of customers served by sales workers
      $t_{\mathbf{x}}$ = the amounts of time these workers spent
  - For such cases, we may be interested in the relationship between:

$$\mu_{\mathbf{x}}/t_{\mathbf{x}} \text{ (the rate at } \mathbf{x}) \leftrightarrow \eta_{\mathbf{x}} = X\beta,$$

  instead of $\mu_{\mathbf{x}} \leftrightarrow \eta_{\mathbf{x}} = X\beta$
- **Q**: how to model the rate data?
  - use binomial GLM if $t_{\mathbf{x}}$ represents the total number of trials
    - Then, $p_{\mathbf{x}} = \mu_{\mathbf{x}}/t_{\mathbf{x}}$
    - Example: burglary example ⇒ a household regarded as a trial

- If the proportions $\mu_{\mathbf{x}}/t_{\mathbf{x}}$ are small and $t_{\mathbf{x}}$ are large, Poisson GLM is an effective approximation to the binomial GLM
- **Q**: can $t_{\mathbf{x}}$ always be regarded as the total number of trials?
  - In customer service example, the $t_{\mathbf{x}}$ is not a count
  - In burglary example, some households may be affected more than once

➢ Fit a Normal linear model: $y_{\mathbf{x}}/t_{\mathbf{x}} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

- However, there are often difficulties with normality and unequal variance, particularly if $y_{\mathbf{x}}$ are small

➢ Use a Poisson GLM with the link (cf., link in log-linear model):

$$\log(\mu_{\mathbf{x}}/t_{\mathbf{x}}) = \eta_{\mathbf{x}} = \boldsymbol{X\beta} \ (\Leftrightarrow \log(\mu_{\mathbf{x}}) = 1\times\log(t_{\mathbf{x}}) + \boldsymbol{X\beta} \equiv \eta_{\mathbf{x}}')$$

- $\log(t_{\mathbf{x}})$ is regarded as a term with coefficient fixed as 1
- This approach models the rate while still maintaining the count response for the log-linear model ⇒ called *rate model*
- A term on the covariate side having no parameter attached is called an *offset*

❖ **Reading**: Faraway (2006, 1st ed.), 3.2

NTHU STAT 5230, 2025, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

Negative Binomial Response

- Data: observe $(x_{i1}, x_{i2}, \ldots, x_{im}, y_i)$, $i = 1, 2, \ldots, k$

$$\Leftrightarrow (\mathbf{x}_i, y_i) \Leftrightarrow (\mathbf{x}, y_{\mathbf{x}})$$

where $y_{\mathbf{x}} \sim$ negative binomial, i.e.,

$$P(Y_{\mathbf{x}} = y_{\mathbf{x}}) = \binom{y_{\mathbf{x}}-1}{r_{\mathbf{x}}-1} p_{\mathbf{x}}^{r_{\mathbf{x}}}(1-p_{\mathbf{x}})^{y_{\mathbf{x}}-r_{\mathbf{x}}}, \quad y_{\mathbf{x}} = r_{\mathbf{x}}, r_{\mathbf{x}}+1, \ldots$$

- Negative binomial distribution can arise naturally in several ways:
  - ➢ $y_{\mathbf{x}}$ = the number of trials until the $r_{\mathbf{x}}$th success;
    - Example: a system can withstand $r_{\mathbf{x}}$ hits; the prob. of a hit in a given time period is $p_{\mathbf{x}}$ ⇒ $y_{\mathbf{x}}$ = the lifetime of the system
  - ➢ The generalization of Poisson response: $y_{\mathbf{x}}|\lambda_{\mathbf{x}} \sim \text{Poisson}(\lambda_{\mathbf{x}})$, where $\lambda_{\mathbf{x}}$ is gamma distributed
  - ➢ A limiting distribution for urn schemes that can be used to model contagion
- We now re-parameterize the negative binomial as follows:

$$y_{\mathbf{x}}' = y_{\mathbf{x}} - r_{\mathbf{x}} \text{ and } \alpha_{\mathbf{x}} = (1/p_{\mathbf{x}}) - 1 \ \ (\Leftrightarrow p_{\mathbf{x}} = (1+\alpha_{\mathbf{x}})^{-1})$$

- The pmf becomes

$$P(Y'_{\mathbf{x}} = y'_{\mathbf{x}}) = \frac{\Gamma(y'_{\mathbf{x}} + r_{\mathbf{x}})}{\Gamma(r_{\mathbf{x}})\,\Gamma(y'_{\mathbf{x}} + 1)} \times \frac{\alpha_{\mathbf{x}}^{y'_{\mathbf{x}}}}{(1+\alpha_{\mathbf{x}})^{y'_{\mathbf{x}}+r_{\mathbf{x}}}}, \quad y'_{\mathbf{x}} = 0, 1, 2, \ldots$$

  where $r_{\mathbf{x}} \in (0, \infty)$ and $\alpha_{\mathbf{x}} \in (0, \infty)$.
- $E(y_{\mathbf{x}}') = r_{\mathbf{x}}\alpha_{\mathbf{x}} \equiv \mu_{\mathbf{x}}$
- $Var(y_{\mathbf{x}}') = r_{\mathbf{x}}\alpha_{\mathbf{x}} + r_{\mathbf{x}}\alpha_{\mathbf{x}}^2 = \mu_{\mathbf{x}} + (\mu_{\mathbf{x}}^2/r_{\mathbf{x}})$

  [cf., (LNp.4-9) alternative expression: $Var(y_{\mathbf{x}}') = \mu_{\mathbf{x}} \times (1+\phi_{\mathbf{x}})/\phi_{\mathbf{x}}$]
- The log-likelihood in terms of $r_{\mathbf{x}}$'s and $\alpha_{\mathbf{x}}$'s is:

$$l(\boldsymbol{\alpha}, \boldsymbol{r}) = \sum_{i=1}^{k} \Big\{ y_i' \log\Big(\frac{\alpha_i}{1+\alpha_i}\Big) + \sum_{j=0}^{y_i'-1} \log(j + r_i) - r_i \log(1+\alpha_i) - \log[\Gamma(y_i' + 1)] \Big\}$$

- Statistical modeling: $y_{\mathbf{x}}' \sim NB(r_{\mathbf{x}}, \alpha_{\mathbf{x}})$ and $\eta_{\mathbf{x}} = \boldsymbol{X\beta}$
  - a convenient way to link $\mu_{\mathbf{x}} = E(y_{\mathbf{x}}') = r_{\mathbf{x}}\alpha_{\mathbf{x}}$ and $\eta_{\mathbf{x}}$ is:

$$\eta_{\mathbf{x}} = \log\Big(\frac{\alpha_{\mathbf{x}}}{1+\alpha_{\mathbf{x}}}\Big) = \log\Big(\frac{\mu_{\mathbf{x}}}{\mu_{\mathbf{x}} + r_{\mathbf{x}}}\Big)$$

NTHU STAT 5230, 2025, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- Analysis of negative binomial responses assuming $r_1 = r_2 = \ldots = r_k \equiv r$:
  - $r_1 = r_2 = \ldots = r_k \;\Leftrightarrow\; r_{\mathbf{x}} = \mu_{\mathbf{x}}\phi_{\mathbf{x}} = [E(\lambda_{\mathbf{x}})]^2/Var(\lambda_{\mathbf{x}})$ (LNp.4-9) is a constant for all $\mathbf{x}$
    - cf., assuming $\phi_{\mathbf{x}} = r_{\mathbf{x}}/\mu_{\mathbf{x}} = E(\lambda_{\mathbf{x}})/Var(\lambda_{\mathbf{x}})$ is a constant for all $\mathbf{x}$
    - cf., dispersion parameter method: $Var(y_{\mathbf{x}}) = \sigma^2 E(y_{\mathbf{x}}) = \sigma^2 \mu_{\mathbf{x}}$
    - it results in a different structure from the dispersion-parameter approach for $E(y_{\mathbf{x}})$ and $Var(y_{\mathbf{x}})$
  - Treating r as a *known* constant
    - It is an ordinary GLM with canonical link

$$\eta_{\mathbf{x}} = \log\big(\mu_{\mathbf{x}}/(\mu_{\mathbf{x}} + r)\big)$$

    - Data analysis techniques for ordinary GLM can be adopted
  - Treating r as an *unknown* parameter
    - Not anymore an ordinary GLM

- MLE of β and r can be obtained by
  1) using a Newton-Raphson routine on all parameters simultaneously, or
  2) evaluating the profile likelihood for various fixed r, or
  3) an approach alternates between a) using IRWLS to estimate β with $\hat{r}$ fixed, b) using Newton-Raphson method to estimate r with $\hat{\beta}$ fixed
- $\hat{\beta}$ and $\hat{r}$ are asymptotically independent
- The usual inferential techniques in GLM can then be used to compare negative binomial models (regarded as inferences conditional on the estimate of r)

➢ Analysis of negative binomial responses allowing parameters $r_1, r_2, \ldots, r_k$ to have different values

- Outside the framework of GLM (out of the scope of this course)
- Manton et al. (1981), *Biometrics* **37**, p.259-269

❖ **Reading**: Faraway (2006, 1st ed.), 3.3

NTHU STAT 5230, 2025, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)