Poisson Regression

of covariate classes

- Recall: in binomial GLM, observe data

covariates $(x_{i1}, x_{i2}, \ldots, x_{im}, y_i)$, $i = 1, 2, \ldots, k$ ← response

$\Leftrightarrow (\mathbf{x}_i, y_i) \Leftrightarrow (\mathbf{x}, y_{\mathbf{x}})$ ← $\sim$ binomial$(n_{\mathbf{x}}, p_{\mathbf{x}})$

where $y_{\mathbf{x}}$ is bounded by $n_{\mathbf{x}}$ = number of total trials at $\mathbf{x}$, i.e., ← Bernoulli (0 or 1) trials

count response: $0 \le y_{\mathbf{x}} \le n_{\mathbf{x}}$ and $n_{\mathbf{x}}$ is a *fixed* and *known* number ← i.e., $n_{\mathbf{x}}$'s are given in data

$y_{\mathbf{x}}$: non-negative integers

➢ **Q**: what if the upper limit to $y_{\mathbf{x}}$ is infinite or effectively so?

➢ Some examples of such $y_{\mathbf{x}}$: What's its upper limit? $0 \le y_{\mathbf{x}} < \infty$ ← assume; e.g., finite upper limit, but do not know what it is

count response without explicit upper limit:

- 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

some event observed over a time interval (1-dimensional) or over a geographic region (2-dimensional)

LNp.2-24~26, LNp 2-20~21

➢ For such $y_{\mathbf{x}}$, can use Poisson GLM or negative binomial GLM

➢ If $y_{\mathbf{x}}$ is sufficiently large → (LNp.2-26) $Y_{\mathbf{x}} \sim$ Poisson$(\lambda_{\mathbf{x}}) \approx$ Normal$(\lambda_{\mathbf{x}}, \lambda_{\mathbf{x}})$ when $\lambda_{\mathbf{x}}$'s are large & Normal approximation to count data in LNp 3-3

$\Rightarrow$ may fit a Normal linear model: $y_{\mathbf{x}} = X\beta + \varepsilon$ ← non-constant variance

- If $Y_{\mathbf{x}}$ is Poisson with mean $\mu_{\mathbf{x}}$ (>0), then ← as a function of covariates

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

pmf of Poisson

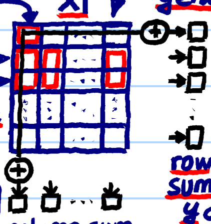
variance structure is controlled by mean structure ⇒ know mean, variance known ⇒ can do goodness-of-fit test ⇒ But, might need to check for over-dispersion

LNp 2-25~26

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

$n_{\mathbf{x}}$'s must be available in binomial GLM

$\Rightarrow$ 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 → then, $y_{\mathbf{x}} \ll n_{\mathbf{x}}$ in most cases

binomial pmf → ; binomial GLM ; LNp 2-24 ; Poisson pmf

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

$E(y_{\mathbf{x}}) = n_{\mathbf{x}} p_{\mathbf{x}}$; no need to know $n_{\mathbf{x}}$'s in Poisson GLM

But, Poisson GLM studies $\mathbf{x} \leftrightarrow \mu_{\mathbf{x}}$, rather than $\mathbf{x} \leftrightarrow p_{\mathbf{x}}$ in binomial GLM ($\because \mu_{\mathbf{x}} = n_{\mathbf{x}} p_{\mathbf{x}} \ne p_{\mathbf{x}}$)

- example: incidence of rare cancer in a large sample ← $p_{\mathbf{x}}$ is small ← $n_{\mathbf{x}}$ is large
- For small $p_{\mathbf{x}}$'s, ← canonical link of binomial GLM

canonical link of Poisson GLM (LNp.4-4)

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

close to 1 when $p_{\mathbf{x}} \approx 0$; check rate model (LNp 4-12)

Their β's have similar meanings & should have similar estimated values

⇒ 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_x's are similar.

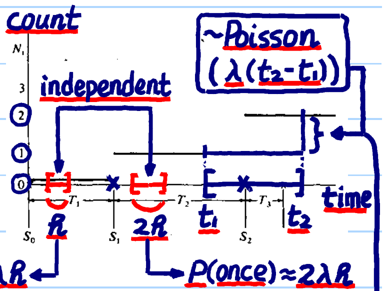
Recall β under different links has different meanings

a stochastic process for count, called Poisson Process

- Suppose that $P(\text{once on } [t, t+h]) \approx \lambda h \ (\leq 1)$, where h is small
  length of time interval
  occurance rate (unit: 次/單位時間)
  1. Probability of an event occurring once in a given (short) time interval is proportion to the length of the interval ★ can be changed to a (geographic) region or space
  2. The numbers of events in two disjoint time intervals are independent

Should exam whether these assumptions are clearly violated in the data (e.g., λ is not a constant over time)

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

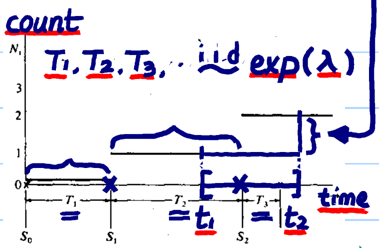


- Example: number of incoming telephone calls/earthquake in $[t_1, t_2]$

check the examples in LNp.4-1

- Suppose that the times between any two adjacent events are i.i.d. as exponential distribution ← with mean $1/\lambda$ (unit: 單位時間/次)
  waiting time
  ⇒ number of events in a given time period is Poisson distributed.

memoryless distribution



- 3 components of Poisson GLM: cf. binomial GLM LNp.3-5

parameters: $y_i \xleftarrow{\text{pmf}} \mu_i \underset{\text{log}}{\overset{\text{exp}}{\rightleftarrows}} \eta_i \xleftarrow{h_i^T\beta} x_i$

response 隨機 ① $y_x \sim \text{Poisson}(\mu_x)$
random, indep.
parameter $\in [0,\infty)$ $= E(y_x)$: expected average count

covariates 規律 ② $X\beta = \sum_{j=1}^{p} \beta_j \times h_j(X_1,\ldots,X_m) \equiv \eta_{\mathbf{x}}(\beta) \in (-\infty,\infty)$

sample size cf. # of covariate classes in binomial GLM

▪ For $i=1, \ldots, k$, denote the ith row of X by

$h_j(x_i) \to (h_{ij}) \equiv \mathbf{h}_i = (h_1(\mathbf{x}_i), \ldots, h_p(\mathbf{x}_i))^T \Rightarrow \eta_{\mathbf{x}_i} = \mathbf{h}_i^T\beta$

usually, the intercept term

③ *link function* g: g monotone and differential, and $\eta_{\mathbf{x}} = g(\mu_{\mathbf{x}})$
g^{-1} exists
$\mu_x = g^{-1}(\eta_x)$

build the link btwn parameters in y_x & $X\beta$

[for Poisson GLM, g: $(0, \infty) \to (-\infty, \infty)$]

▪ *Log-linear model*: $\eta_{\mathbf{x}} = \log(\mu_{\mathbf{x}}) \Leftrightarrow \mu_{\mathbf{x}} = \exp(\eta_{\mathbf{x}})$ ← canonical link
check Δ in LNp.4-2

- Log-likelihood of log-linear model:

X^TY is the suff. stat. of β
vector of response data (check LNp.4-5)

$$l(\beta) = \log\left(\prod_{i=1}^{k} \frac{e^{-\mu_{\mathbf{x}_i}} \mu_{\mathbf{x}_i}^{y_i}}{y_i!}\right)$$

joint pmf of $y_1, \ldots, y_k$

$$(*) \quad = \sum_{i=1}^{k} \left[y_i \log(\mu_{\mathbf{x}_i}) - \mu_{\mathbf{x}_i} - \log(y_i!) \right] \leftarrow \text{as a function of } \mu_i\text{'s}$$

$$= \sum_{i=1}^{k} \left[y_i\, \eta_{\mathbf{x}_i} - \exp(\eta_{\mathbf{x}_i}) - \log(y_i!) \right] \leftarrow \text{as a function of } \eta_i\text{'s}$$

$$= \sum_{i=1}^{k} \left[y_i\, (\mathbf{h}_i^T\beta) - \exp(\mathbf{h}_i^T\beta) - \log(y_i!) \right] \leftarrow \text{as a function of } \beta$$

$\sum_{j=1}^{p} h_{ij}\beta_j$

p. 4-5

- Estimation (MLE) of β:
  - Partial derivative of $l(\beta)$: $\frac{\partial}{\partial \beta_j} l(\beta) = \sum_{i=1}^{k} [\, y_i \mathbf{h}_{ij} - \exp(\mathbf{h}_i{}^T\beta)\, \mathbf{h}_{ij} \,] = \sum_{i=1}^{k} [\, y_i - \exp(\mathbf{h}_i{}^T\beta) \,]\, \mathbf{h}_{ij}$

by chain rule: $\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{k} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{k} \frac{\partial l_i}{\partial \mu_i} \times \frac{\partial \mu_i}{\partial \eta_i} \times \frac{\partial \eta_i}{\partial \beta_j}$ (log-likelihood of ith obs. y_i)

① 隨. $= y_i/\mu_i - 1$ ② link. $= \exp(\eta_i) = \mu_i$ ③ 規. $= h_{ij}$

ith row of X; μ_i

  - Set $\partial l(\beta)/\partial \beta_j = 0$, $\forall\, j$. The MLE $\hat\beta$ is the solution to: (cf. substitute $\hat\beta$)

$$0 = \sum_{i=1}^{k} \mathbf{h}_{ij}\, y_i - \sum_{i=1}^{k} \mathbf{h}_{ij}\, \hat{\boldsymbol{\mu}}_i, \quad j = 1, \ldots, p.$$

vector of response data

$$\Leftrightarrow X^T Y - X^T \hat{\boldsymbol{\mu}} = 0 \Leftrightarrow X^T Y \overset{\star}{=} X^T \hat{\boldsymbol{\mu}}$$

$X^T Y$: suff stat of β

where $\hat{\boldsymbol{\mu}} = \exp(\hat{\boldsymbol{\eta}}) = \exp(X^T\hat\beta)$ normal equation

Q: Can we just set $\hat\mu = Y$ to obtain the MLE $\hat\beta$?
- Note. $\dim(\mu) = \dim(\beta) = p$. But. $\dim(Y) = k\ (\geq p)$
- Yes when saturated ($k = p$) under which X^{-1} exists & $\hat\mu = Y$

  - The link function having the property $X^T\hat{\boldsymbol{\mu}} = X^T Y$ is known as *canonical link*.

$X^T X\hat\beta = X^T Y \Rightarrow \hat\beta = (X^T X)^{-1} X^T Y$

  - 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.

隨; 規; link function; check LNp.3-7

  - 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)

$\widehat{Var}(\hat\beta) = (X^T\hat{W}X)^{-1}\hat\sigma^2$

p. 4-6

- Deviance of log-linear model: ← Recall. Deviance analogous to RSS in LM (LNp.3-8~9)

compare log-likelihood of

  - For a saturated model L^*, the MLE of $\mu_{\mathbf{x}}$ is $y_{\mathbf{x}}$; ($\hat\mu_{\mathbf{x},L^*}$)

$\dim(\beta) = \dim(\eta) = k \Rightarrow \dim(\mu) = k$, $\mu \in [0,\infty)^k$
$\Rightarrow$ each μ_i can freely vary in $[0,\infty)$ & the MLE of one Poisson obs. y_i is $\hat\mu_i = y_i$ (also. check LNp.3-9)

β; sample size

  - 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:$L^*\backslash S$):

$\sum_i y_i - \sum_i \hat\mu_i = 0$ by $\star$ in LNp.4-5 if S contains the intercept

$= \exp(\hat\eta_{\mathbf{x},S}) = \exp(\mathbf{h}_{\mathbf{x}}^T\hat\beta_S)$

$\hat\mu_S$ is a k-dim vector. But. $\dim(\hat\mu_S) = s$

$$D_S = 2\,(l_{L^*} - l_S) = -2\log\left[\mathcal{L}(\hat\beta_S)/\mathcal{L}(\hat\beta_{L^*})\right]$$

$$= 2\sum_{i=1}^{k}\left[y_i \log\left(\frac{y_i}{\hat\mu_{i,S}}\right) - (y_i - \hat\mu_{i,S})\right]$$

likelihood; MLE; check LNp.3-8; ratio; difference; $y_{i,S}$

By (*) in LNp.4-4

When $S = L^*$ $\Rightarrow D_S = 0$

(exercise) Express this equation in terms of $\hat\eta$ or $\hat\beta$ (check LNp.4-4)

  - $D_S \overset{a}{\sim} \chi^2_{k-s}$ when S is the true model

test stat: D_S; null dist.: χ^2_{k-s}

$\Rightarrow$ 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?)

1. In binomial case, asymptotics hold as n_x's $\to \infty$.
2. In the approximation of Poisson to binomial (LNp4-2), $\mu_x \approx n_x p_x$. $\Rightarrow \mu_x \to \infty$ as $n_x \to \infty \Rightarrow$ asymptotics hold as μ_x's $\to \infty$ or large
3. But. μ_x's are unknown. Q: What should we check in data? Note. $\mu_x = E(y_x)$ $\Rightarrow y_x$'s large (say, ≥ 5)

  - The D is known as G^2-statistic in the analysis of contingency table (future lecture) cf. Pearson X^2