- Recall: $\mu_{\mathbf{x}_0} \underset{g^{-1}}{\overset{g}{\rightleftarrows}} \eta_{\mathbf{x}_0} = \mathbf{h}_0{}^T\beta$, and under log-linear model,

MLE of β: $\hat{\beta} \overset{a}{\sim} N(\beta, \Sigma)$

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

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

$\hat{\eta}_{\mathbf{x}_0} \overset{a}{\sim} N(\mathbf{h}_0^T\beta, \mathbf{h}_0^T\Sigma\mathbf{h}_0)$

$[\underline{\eta}_{\mathbf{x}_0}, \overline{\eta}_{\mathbf{x}_0}] \equiv \hat{\eta}_{\mathbf{x}_0} \pm z(\alpha/2) \times se(\hat{\eta}_{\mathbf{x}_0})$ — $z(\alpha/2)$ from N(0,1); $se = \sqrt{\hat{h}_0^T\hat{\Sigma}\hat{h}_0}$

($\hat{\Sigma}$ = estimated covariance matrix of $\hat{\beta}$)

- $\hat{\mu}_{\mathbf{x}_0} = \exp(\hat{\eta}_{\mathbf{x}_0})$ $= (X^T\hat{W}X)^{-1}$ from IRWLS (need to adjust if overdispersion (LNp.4-10))

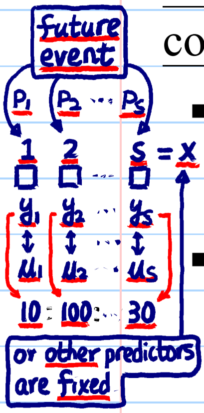
$[\exp(\underline{\eta}_{\mathbf{x}_0}), \exp(\overline{\eta}_{\mathbf{x}_0})]$

check LNp 3-27 cf.

- The construction of confidence intervals for $\eta_{\mathbf{x}_0}$ and $\mu_{\mathbf{x}_0}$ follows a similar manner as in binomial GLM.

(LNp 2-26) For $i=1,\dots,s$, $Y_i \overset{indep}{\sim} Poisson(\mu_i)$ $\Rightarrow (Y_1,\dots,Y_s \mid \sum_i Y_i = n) \sim multinomial(n, p_1,\dots,p_s)$, $p_i = \mu_i/(\mu_1+\dots+\mu_s)$

- Suppose that there is only an s-level categorical covariate (or other covariates are held fixed)
  - Denote the s categories by $1,\dots,s$; their count observations by $y_1,\dots,y_s$; their count parameters by $\mu_1,\dots,\mu_s$ — Poisson
  - 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 $y_1,\dots,y_s$ were collected under the same value of a size variable (next slide)



check ★ in LNp.4-3 — What if different

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

Rate Model

count response — interested in

e.g., time length (1-dim), total # of trials (1-dim), area (2-dim), volume (3-dim) (check example in LNp.4-1&3)

- Data: observe the number of events $y_{\mathbf{x}}$ at $\mathbf{x}$, and assume $y_{\mathbf{x}} \sim \text{Poisson}(\mu_{\mathbf{x}})$, (expected average count) 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

y_x's ↔ t_x's in data

- Some examples: $y_x \ll t_x$
  - $y_{\mathbf{x}}$ = 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 ☒

◉ difference in μ_x's cf. ◉ difference in μ_x/t_x's — equivalent if t_x's is a constant over x's

Note. If data with count response does not contain t_x's, should always exam whether their t_x's are comparable in size

- For such cases, we may be interested in the relationship between:

$$[0,1] \ni, [0,a] \ni, [0,\infty) \ni \mu_{\mathbf{x}}/t_{\mathbf{x}} \text{ (the rate at } \mathbf{x}) \leftrightarrow \eta_{\mathbf{x}} = X\beta \in (-\infty, \infty)$$

how X influence rate

instead of $\mu_{\mathbf{x}} \leftrightarrow \eta_{\mathbf{x}} = X\beta$ — how X influence μ_x in Poisson GLM

The meanings of parameters β are different.

e.g. (unit: 次/單位時間) × (時間大小) = (多少次), 時間大小 = t_x, 多少次 = μ_x

- **Q**: how to model the rate data?
  - use binomial GLM if $t_{\mathbf{x}}$ represents the total number of trials (an integer) — n_x in LNp.3-5
    - Then, $p_{\mathbf{x}} = \mu_{\mathbf{x}}/t_{\mathbf{x}} \in [0,1]$ — rate becomes a probability, i.e., prob. of getting one count, and μ_x is still $E(y_x)$, but y_x ✗ Poisson
    - Example: burglary example ⇒ a household regarded as a trial

① $y_x \sim B(t_x, p_x)$, $\overset{g}{\downarrow}$ $X\beta = \eta_x$

cf no offset

LNp.4-2
n_x: large
p_x small

- If the proportions μ_x/t_x are small and t_x are large, Poisson GLM is an effective approximation to the binomial GLM

similar meaning of β as ① (p.4-3)

- **Q**: can t_x always be regarded as the total number of trials?
  - In customer service example, the t_x is not a count

t_x: time length

if t_x's (or $\log(t_x)$'s) close in size

  - In burglary example, some households may be affected more than once

a household is not a Bernoulli trial

~Normal

$\mu_x^* = t_x p_x = t_x(\mu_x/t_x) = \mu_x$

$y_x \sim$ Poisson(μ_x^*) ② $\xrightarrow{\log}$ $x\beta = \eta_x$

- Fit a Normal linear model: $\overset{\text{or taking } \log(\cdot)}{y_x/t_x} \overset{③}{=} X\beta + \varepsilon$

∵ $E(y_x/t_x) = \mu_x/t_x$ interested parameters

cf When is it appropriate?

(LNp.3-3) $y_x \sim B(n_x, p_x)$, model y_x/n_x using LM

  - However, there are often difficulties with normality and unequal variance, particularly if y_x are small ⇔ μ_x small

$E(y_x/t_x) = \mu_x/t_x$ might be close to its lower bound, i.e., 0

using offset

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

$$\underbrace{\log(\mu_x)}_{\text{unknown}} - \underbrace{\log(t_x)}_{\text{known}} = \log(\mu_x/t_x) = \underbrace{\eta_x}_{\geq 0} = \underbrace{X\beta}_{規} \;(\Leftrightarrow \log(\mu_x) = \underbrace{1}_{\text{not a parameter}}\times\log(t_x) + X\beta \equiv \eta_x')$$

β explains log rate (cf. ③)

  - $\log(t_x)$ 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*

i.e., its β interpret change in rate, not change in μ_x (over X)

  - A term on the covariate side having no parameter attached is called an *offset*

$y_x \sim$ Poisson(μ_x) ④ $\xrightarrow{\log}$ $\log(t_x) + X\beta = \eta_x'$

Its β might be different from what in
- ①, when R's large
- ②, if t_x's different
- ③, ∵ log link

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

Recall LM, LNp.4-14

Negative Binomial Response

Recall overdispersion in LNp.4-9

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

$$\Leftrightarrow (\mathbf{x}_i, y_i) \Leftrightarrow (\mathbf{x}, \underbrace{y_\mathbf{x}}_{\text{count response}})$$

$E(Y_x) = r_x/p_x$
$Var(Y_x) = r_x(1-p_x)/p_x^2$

where $y_x \sim$ negative binomial, i.e.,

\# of Bernoulli trials

might be not appropriate to model count response ∵ $y_x \geq r_x$

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

$\frac{(y'+r-1)!}{(r-1)!\,y'!} = \frac{(y-1)!}{(r-1)!\,(y-r)!}$

now, r_x might treated as a parameter & $r_x \in [0,\infty)$

- Negative binomial distribution can arise naturally in several ways:
  - ➢ y_x = the number of $\underset{\text{Bernoulli}}{\text{trials}}$ until the r_xth success;

short time period ⇒ a hit or no hit

gamma distribution cf

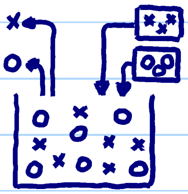
  - Example: a system can withstand r_x hits; the prob. of a hit in a given time period is p_x ⇒ y_x = the lifetime of the system
  - ➢ The generalization of Poisson response: $y_x|\lambda_x \sim$Poisson(λ_x), where λ_x is gamma distributed

LNp.4-9

  - ➢ A limiting distribution for urn schemes that can be used to model contagion

S_n: # of o until nth draw

$\lim_{n\to\infty} P(S_n = k) = $ NB pmf



Pólya urn process

- We now re-parameterize the negative binomial as follows:

\# of o's (failures)
$y_x' = 0, 1, 2, \ldots$
better for count response

$$y_x' = y_x - r_x \text{ and } \underbrace{\alpha_x}_{1/\text{odds}_x,\ \in[0,\infty)} = (1/p_x) - 1 \quad (\Leftrightarrow \underbrace{p_x}_{\in[0,1]} = (1+\alpha_x)^{-1}) \quad \alpha_x\uparrow\ p_x\downarrow$$

$P_x = 1/1+\alpha_x$, $1-P_x = \alpha_x/1+\alpha_x$

$y_x = y_x' + r_x$

- The pmf becomes

gamma function: generalization of factorial

$\Gamma(u)=\int_0^\infty t^{u-1}e^{-t}dt$

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

more appropriate to model count response

allow r_x to be non-integer

could be a parameter

where $r_x \in (0, \infty)$ and $\alpha_x \in (0, \infty)$.

$\mu_x^2/r_x \to 0$, as $r_x \to \infty$ & μ_x fixed

$\Leftrightarrow \phi_x$ (LNp.4-9) $\to \infty \Leftrightarrow \lambda_x$ has var 0

$\Leftrightarrow$ negative binomial $\to$ Poisson

- $E(y_x') = r_x\alpha_x \equiv \mu_x$ $\frac{r_x}{P_x} - r_x$

$r_x \times (1+\alpha_x)^2 \times \frac{\alpha_x}{1+\alpha_x}$

- $Var(y_x') \overset{\uparrow}{=} r_x\alpha_x + r_x\alpha_x^2 = \mu_x + (\mu_x^2/r_x)$

$\geq 0 \Rightarrow$ overdispersion

$\mu_x + \frac{\mu_x^2}{r_x} = \mu_x(1+\frac{\mu_x}{r_x})$

$= \mu_x(1+1/\phi_x)$

$\Rightarrow \phi_x = r_x/\mu_x = 1/\alpha_x$

odds$_x$

[cf., (LNp.4-9) alternative expression: $Var(y_x') = \mu_x \times (1+\phi_x)/\phi_x$]

4/14

- The log-likelihood in terms of r_x's and α_x's is:

parameterization for NB

$(r, p), (r, \alpha), (r, \mu), (\mu, \phi)$

from gamma λ

$\because$ independent

$$-l(\boldsymbol{\alpha}, \boldsymbol{r}) = \sum_{i=1}^{k}\Big\{ y'_i \log\Big(\underbrace{\frac{\alpha_i}{1+\alpha_i}}_{ⓐ}\Big) + \underbrace{\sum_{j=0}^{y'_i-1} \log(j + r_i)}_{ⓑ}$$

$\because \Gamma(u+1) = u\Gamma(u)$

$= \theta_i$: canonical parameter

ⓐ a function of data & parameter (α_i)

ⓑ a function of data & parameter (r_i)

$$\underbrace{- r_i \log(1 + \alpha_i)}_{ⓒ} - \underbrace{\log[\Gamma(y'_i + 1)]}_{ⓓ} \Big\}$$

ⓒ a function of parameters only

ⓓ a function of data only

- Statistical modeling: $y_x' \overset{\text{indep.}}{\sim} NB(r_x, \alpha_x)$ and $\eta_x = X\beta$

隨 規

- a convenient way to link $\mu_x = E(y_x') = r_x\alpha_x$ and η_x is: $(\mu_i, r_i) \underset{?}{\overset{g}{\rightleftarrows}} \eta_i$

unit: log(probability)

note. not logit

$\in (-\infty, 0)$ not $(-\infty, \infty)$

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

probability of failure (0)

$\because \alpha_x = \mu_x / r_x$

a parameter r_i is involved in the link in addition to mean μ_i (cf., other GLM)