□ So far for GLIM {formula pages}:

$Y_i, \underset{\sim}{x}_i$

Y_i :- exponential family (θ_i)

$$\left[\theta_i \longleftrightarrow \mu_i, \operatorname{var}(Y_i) \right]$$

g :- link function

$$\left[g(\mu_i) = \underset{\sim}{x}_i^t \underset{\sim}{\beta} + c_i \right]$$

□ For MLE $\hat{\beta}$:

$$\frac{\partial L}{\partial \beta_j} = \cdots$$

$$\frac{\partial^2 L}{\partial \beta_k \partial \beta_j} = \cdots = \mathrm{I} + \underset{\uparrow \; \frac{\partial^2 \theta_i}{\partial \beta_k \partial \beta_j}}{\mathrm{II}}$$

□ Question:

when can ignore II in $\left(\frac{\partial^2 L}{\partial \beta_k \partial \beta_j} \right)$?

Answer: can ignore

$$w_i (y_i - \mu_i) \frac{\partial^2 \theta_i}{\partial \beta_k \partial \beta_j} \quad \text{in} \left(\frac{\partial^2 L}{\partial \beta_k \partial \beta_j} \right)$$

(A) if $\frac{\partial \theta_i}{\partial \beta_j} = x_{ij}$. If choose link function $g(\mu_i) = \underset{\sim}{x}_i^t \underset{\sim}{\beta}$ such that $g(\mu_i) = \theta_i$

then $\frac{\partial \theta_i}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} (\underset{\sim}{x}_i^t \underset{\sim}{\beta}) = x_{ij}$

{g is called a <u>canonical link function</u>}

<u>Examples</u> ($\phi = w_i = 1$):

$$\log f(y \mid \theta_i) = \underbrace{y \cdot \theta_i} - b(\theta_i) + c(y)$$

- $Y \sim N(\mu_i, 1)$: $\log f(y) = \underbrace{y \cdot \mu_i} - \mu_i^2/2$

$$\Rightarrow \theta_i = \mu_i \Rightarrow g(\mu_i) = \mu_i = \underset{\sim}{x}_i^t \underset{\sim}{\beta}$$

- $Y \sim \text{Pois}(\lambda_i)$; $f(y) = e^{-\lambda_i} \cdot \frac{\lambda_i^y}{y!}$, $y = 0, 1, \ldots$

$$\Rightarrow \log f(y) = \underbrace{y \cdot \log \lambda_i} - \lambda_i - \log(y!)$$

So $\theta_i = \lambda_i \Rightarrow g(\lambda_i) = \log \lambda_i = \underset{\sim}{x}_i^t \underset{\sim}{\beta}$

- $y \sim Ber(p_i)$

$$f(y) = p_i^{y} \cdot (1-p_i)^{1-y} \qquad (0 < p_i < 1)$$

$$\Rightarrow \quad \{ \log f(y|\theta_i) = y \cdot \theta_i - b(\theta_i) + c(y) \}$$

$$\log f(y) = y \log(p_i) + (1-y)\log(1-p_i)$$

$$= y\left[\log(p_i) - \log(1-p_i)\right] + \log(1-p_i)$$

$$= y \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)$$

$$\Rightarrow \quad \theta_i = \log\left(\frac{p_i}{1-p_i}\right)$$

☐ (B) Can also ignore $w_i (y_i - \mu_i) \dfrac{\partial^2 \theta_i}{\partial \beta_k \partial \beta_j}$

if use <u>Fisher Scoring Algorithm</u>

(replaces $\dfrac{\partial^2 L}{\partial \beta_k \partial \beta_j}$ by $E\left(\dfrac{\partial^2 L}{\partial \beta_k \partial \beta_j}\right)$) ;

$$E(\quad) = 0$$

□ Question: Since $g(\mu) = x^T \beta$, how interpret β_j?

$$Y \sim N(\mu, \sigma^2), \qquad \mu = \beta_1 x_1 + \dots + \beta_p x_p \qquad [g(\mu) = \mu]$$

$$\Rightarrow E(Y \mid x_1+\Delta, x_2, \dots, x_p) - E(Y \mid x_1, x_2, \dots, x_p) = \beta_1 \cdot \Delta$$

for any $x_1, \dots, x_p$

$$\text{"} \frac{\partial E(Y)}{\partial x_1} = \beta_1 \text{"}$$

$$[g(\mu) = \log(\mu)]$$

□ $Y \sim Pois(\lambda)$, $\log(\lambda) = \beta_1 x_1 + \dots + \beta_p x_p$

$$\Rightarrow \log[E(Y \mid x_1+\Delta, x_2, \dots, x_p)] - \log[E(Y \mid x_1, \dots, x_p)] = \beta_1 \cdot \Delta$$

$$\Rightarrow \frac{E(Y \mid x_1+\Delta, \dots)}{E(Y \mid x_1, \dots)} = e^{\beta_1 \cdot \Delta} \quad \text{for any } x_1, \dots, x_p$$

change in $E(Y)$, on multiplicative scale

□ $Y \sim Ber(p)$, $\log\left(\frac{p}{1-p}\right) = \beta_1 x_1 + \dots + \beta_p x_p$

$$\left[g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)\right]$$

? ?

□ If $p = 3/4$, the <u>odds</u> for $Y=1$

are $3:1 = \frac{3}{1}$

$$= \frac{3/4}{1/4} = \frac{3/4}{1-3/4}$$

$$= p/(1-p)$$

□ Thus if p_2 is p for $x_1+\Delta, x_2, \ldots, x_p$

and p_1 — — — — — $x_1, x_2, \ldots, x_p$,

$$\frac{p_2/(1-p_2)}{p_1/(1-p_1)}$$ is the <u>odds ratio</u>

and $\log[p_2/(1-p_2)] - \log[p_1(1-p_1)]$ is the log-odds ratio

$$\parallel$$

$$\beta_1\Delta$$

□ So $\beta_1\Delta$ gives the change in the log-odds-ratio

due to a change of Δ in x_1, for any $x_1, \ldots, x_p$

$$\frac{\text{odds of } p \text{ for } x_1+\Delta}{\text{odds of } p \text{ for } x_1} = e^{\beta_1\Delta}, \text{ any } x_1, x_2, \ldots, x_p$$

□ $\log\left(\frac{p}{1-p}\right)$ is the <u>logistic transformation</u>

□ Distribution of $\hat{\beta}$?

□ From formulas page:

NR: $\beta_{(m+1)} = \beta_{(m)} + (X^t C_m^{-1} X)^{-1} X^t A_m^{-1} r_m$

where $\boxed{A = \text{diag}\left(g'(\mu_i)^2 \, \text{var}(y_i)\right)}$

$r = \left(g'(\mu_i)(y_i - \mu_i)\right)$

$$C_m = \begin{cases} A_m & \text{(canonical link)} \\ A_m + D_m & \text{(not)} \end{cases}$$

□ If $\beta_{(m)} \to \hat{\beta}$ then $\beta_{(m+1)} \approx \beta_{(m)} \; (= \hat{\beta})$ and thus

$$(X^t \hat{C}^{-1} X)^{-1} X^t \hat{A}^{-1} \hat{r} = \beta_{(m+1)} - \beta_{(m)} \approx 0$$

$$\boxed{\Rightarrow \circledast \quad X^t \hat{A}^{-1} \hat{r} \approx 0}$$

{whether use D_m or not}

□ $\circledast$ is an example of an estimating equation, which implicitly defines $\hat{\beta}$. [via $\hat{\mu}_i$ and possibly $\hat{\text{var}}(y_i)$]

EG normal: $g(\mu) = \mu$, $A = \sigma^2 I$, $r = y - \mu$

$\Rightarrow \hat{r} = y - X\hat{\beta}$

so $\circledast$ becomes

$$X^t(\sigma^2 I)^{-1}(y - \hat{\mu}) = 0$$

$$\Rightarrow \quad X^t(y - X\hat{\beta}) = 0$$

$$\Rightarrow \quad \hat{\beta} = (X^t X)^{-1} X^t y \quad \Rightarrow \ldots \hat{\beta} \sim N(\beta, \sigma^2 (X^t X)^{-1})$$

☐ So since (*) $X^t \hat{A}^{-1} \hat{\underset{\sim}{r}} = \underset{\sim}{0}$

we need $\hat{\underset{\sim}{r}}$ in terms of $\hat{\underset{\sim}{\beta}}$.

① Now $\hat{r}_i = g'(\hat{\mu}_i)(y_i - \hat{\mu}_i)$

$$= g'(\hat{\mu}_i)(y_i - \mu_i) + g'(\hat{\mu}_i)(\mu_i - \hat{\mu}_i).$$

② Since $g(\mu_i) \approx g(\hat{\mu}_i) + g'(\hat{\mu}_i)(\mu_i - \hat{\mu}_i)$, $\{\hat{\mu}_i \approx \mu_i\}$

get $\hat{r}_i \approx g'(\hat{\mu}_i)(y_i - \mu_i) + g(\mu_i) - g(\hat{\mu}_i)$

$$= g'(\hat{\mu}_i)(y_i - \mu_i) + \underset{\sim}{x}_i^t(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}) \qquad \{g(\mu) = \underset{\sim}{x}^t \underset{\sim}{\beta}\}$$

③ In matrix form:

$$\hat{\underset{\sim}{r}} \approx \hat{G}(\underset{\sim}{y} - \underset{\sim}{\mu}) + X(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}), \qquad \hat{G} = \mathrm{diag}(g'(\hat{\mu}_i)).$$

☐ Thus estimating equation (*) $X^t \hat{A}^{-1} \hat{\underset{\sim}{r}} = \underset{\sim}{0}$

becomes

$$X^t \hat{A}^{-1} [\hat{G}(\underset{\sim}{y} - \underset{\sim}{\mu}) + X(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}})] \approx \underset{\sim}{0}$$

$$\Rightarrow \cancel{X}^t \hat{A}^{-1} \hat{G}(\underset{\sim}{y} - \underset{\sim}{\mu}) \approx (X^t \hat{A}^{-1} X) \cdot (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})$$

$$\Rightarrow (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \approx (X^t A^{-1} X)^{-1} X^t A^{-1} G(\underset{\sim}{y} - \underset{\sim}{\mu})$$

(remove hats)

Since $Y_i - \mu_i$ are independent, CLT $\Rightarrow$

$$\hat{\underset{\sim}{\beta}} \approx N(\underset{\sim}{\beta}, W)$$

where

$$W = (X^t A^{-1} X)^{-1} X^t A^{-1} G \cdot cov(\underset{\sim}{Y}) \cdot G^t A^{-t} X (X^t A^{-1} X)^{-1}$$

for

$cov(\underset{\sim}{Y}) \equiv V = diag\,(var(Y_i))$

$G = diag\,(g'(\mu_i))$

$A = diag\,[g'(\mu_i)^2 \cdot var(Y_i)] \equiv GVG$ (p. 20)

{p. 22; formulas page}

{G, A symmetric}

Thus

$$W = (X^t A^{-1} X)^{-1} X^t A^{-1} G \cdot V \cdot G\, A^{-1} X (X^t A^{-1} X)^{-1}$$

$$= (X^t A^{-1} X)^{-1} X^t (GVG)^{-1} \underbrace{GVG\, (GVG)^{-1}}\, X (X^t A^{-1} X)^{-1}$$

$$= (X^t A^{-1} X)^{-1} X^t (\overset{A}{GVG})^{-1} X (X^t A^{-1} X)^{-1}$$

$$= (X^t A^{-1} X)^{-1} \cdot X^t A^{-1} X \cdot (X^t A^{-1} X)^{-1}$$

$$= (X^t A^{-1} X)^{-1}$$

Note: Since $A = diag\,[g'(\mu_i)^2 \cdot var(Y_i)]$,

W is "bigger" when $var(Y_i)$ are bigger

☐ Now: Application to 3 datasets

{see Week 2}:

medfly

meron

mismatch

☐ medfly.pdf

(1)

(2)

(3) traps because ... was near the

(4)

☐ (5) "Solution" regress $Y_i = A_i$ on $dist_i$

"proves" they migrate

☐ (6) Problems: - outlier

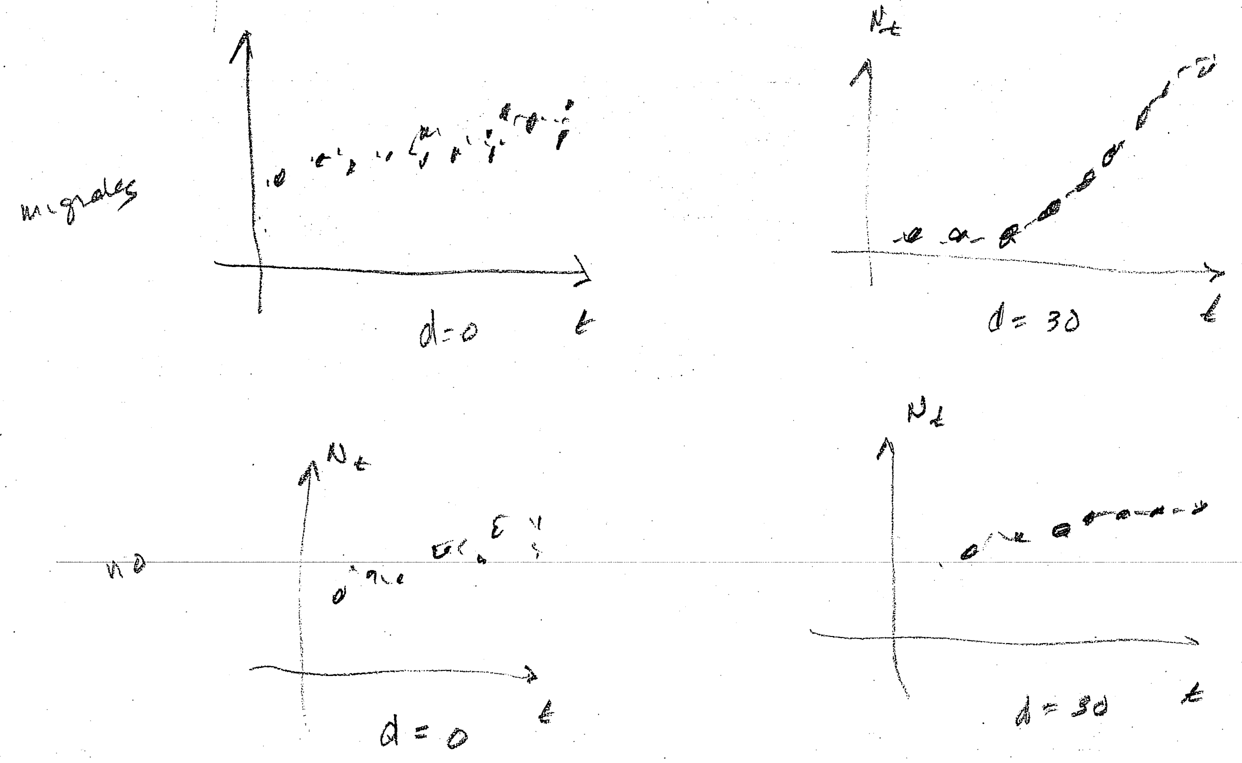
- $var(Y_i) \propto E(Y_i)$

{ Y_i = # flies ~ Poisson, $var(Y) = E(Y)$ }

Also:

Original data:

N_t = # of flies in week t, $t = 1, \ldots, 27$



d = distance to source

Problem: Time series

Solution: Use summary stats A, W

Note: — should analyze (A, W) (mvar response)
— dist.? of W (~ Γ ? discrete)
— could be spatial correlation

☐ First: medfly.preliminary.pdf 25)

{ Do the your data sets:

- $\log_{10}$, not log   ($\log y = 4 \Rightarrow y = ?$)

- (1) $alt_i < 0$?   A skewed $\Rightarrow$ log (A)

- (2) A, dist - MC ?   (better with log (A))

- (3) table of south, loc-host;   (South = 0)

problem w $Y = \dots + \text{south} + \text{loc-host} + \dots$

☐ EG $(E(Y)) = \alpha \cdot \text{South} + \beta \cdot \text{Loc-host}$;

S \ L	0	1
0	0	β
1	α	$\alpha+\beta$

— based on only 2 observations!

☐ (4) Instead, only consider effect of Loc-host for traps in south; define variable

- all β_i make sense
- assumptions justified

☐ Linear Regression results;

- $\hat{\beta}_i$ make sense

p-values are too small for ? obs (n = 89)

(p-values ??)

□ Poisson regression for medfly data

medfly.poisson.rslts.pdf

$\log_{10}$ vs $\ln$:

$$z = 10^{\log_{10}(z)} = e^{\ln(10)\cdot\log_{10}(z)} = e^{\ln(z)}$$

$\Rightarrow$

$$\ln(z) = \underbrace{\ln(10)}_{2.3} \cdot \log_{10}(z)$$

□ Thus if $\log_{10}(A) \approx \beta_0 + \beta_1 x$

then
$$\begin{aligned} \ln(A) &= \ln(10)\cdot\log_{10}(A) \\ &\approx \ln(10)\cdot(\beta_0 + \beta_1 x) \\ &= \gamma_0 + \gamma_1 x \end{aligned}$$

where

$$\gamma_i = \ln(10)\cdot\beta_i$$

□ [Poisson $\hat{\gamma}_i$] $\approx 2.3 \times$ [regression $\hat{\beta}_i$]

But PROBLEM:

$\hat{\beta}_i$ too significant!!!

$\Rightarrow$ perhaps A_i is not Poisson ?

(which $\Rightarrow$ $var(A_i) = E(Y_i)$) ??

☐ I.e. {formula page}

$$(*) \quad \mathrm{cov}(\hat{\beta}) \approx W = \left\{ X^t \left[\mathrm{diag}\left(g'(\mu_i)^2 \cdot \mathrm{var}(A_i)\right) \right]^{-1} X \right\}^{-1}$$

– link function g is OK {linear regression of $\log(A_i)$}

– But if true $\mathrm{var}(A_i)$ is $\gg \lambda_i = E(Y_i)$, then using λ_i in $(*)$ gives too small a value.

☐ Crude check:

$$\widehat{\mathrm{var}}(A) / \hat{E}(A) \approx 1{,}000$$

(assuming $\lambda_i \equiv \lambda$; later will see how to check using residuals from GLIM fit)

☐ Possible solution: use

Negative Binomial dist.

☐ Negative Binomial

Repeated trials, independent $\quad p = P(\text{success})$

$1-p = P(\text{failure})$

r parameter

$Y = \#$ of successes until get r failures

☐ $Y = y \iff$ in $y + (r-1)$ trials get y successes (in random order) and then get a failure

$$= \binom{y+r-1}{y} p^y (1-p)^{r-1} \cdot (1-p)$$

$$= \binom{y+r-1}{y} p^y (1-p)^r , \quad y = 0, 1, 2 \ldots$$

☐ $\underbrace{\overbrace{S\,S\,F\,S\,S\,S}^{\text{random order}}\,F}_{r=2}\,F \ldots \quad y = 5$

$1 + p + p^2 \ldots = \frac{1}{1-p}$

$Y \sim NB(r, p) \qquad \{ r = 1 : \text{Geometric dist} \}$

☐ $E(Y) = r\left(\frac{p}{1-p}\right) = \mu$

$$\text{var}(Y) = \frac{rp}{(1-p)^2}$$

☐ Problem w NB for weekly data

$$\mathrm{var}(y) = \frac{rp}{(1-p)^2}$$

$$= \frac{rp}{(1-p)} \cdot \frac{1}{(1-p)}$$

$$= \mu \cdot \left(\frac{1}{1-p}\right)$$

$$= \mu \left[1 + \frac{p}{1-p}\right]$$

$$= \mu \left[1 + \frac{1}{r} \cdot \frac{rp}{1-p}\right]$$

$$= \mu \cdot \left[1 + \frac{1}{r} \cdot \mu\right].$$

☐ largest $\mathrm{var}(y)$ (for $r=1$) is $\mu(1+\mu)$.

If $\mu = 400 = \bar{A}$

$\mu(1+\mu) \approx 160{,}000 << 400{,}000 = \mathrm{var}(A)$.