☐ Solution: Will formulate model using moments {formula page}:

$$E(Y_i) = \mu_i$$

$$\text{var}(Y_i) = \left(\frac{\phi}{w_i}\right) \cdot v(\mu_i) \quad \text{for some function } v$$

$$g(\mu_i) = x_i^t \beta \quad \text{for link function } g$$

and use quasi-likelihood to estimate β and ϕ

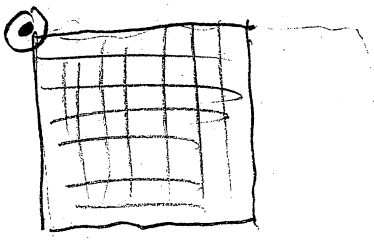
☐ Later; now look at Meron data {meron.pdf}

Y_i = tree. g_2 = binary response

☐ (x, y): geographical trend

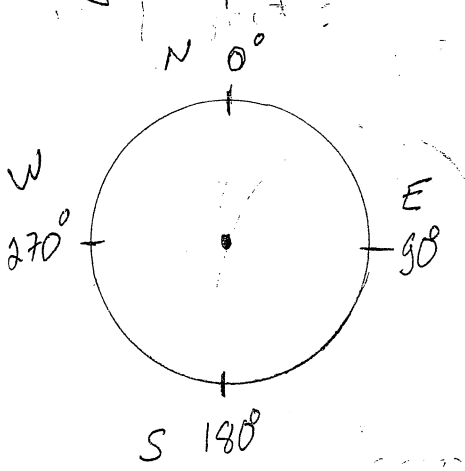
☐ grazing variables: →

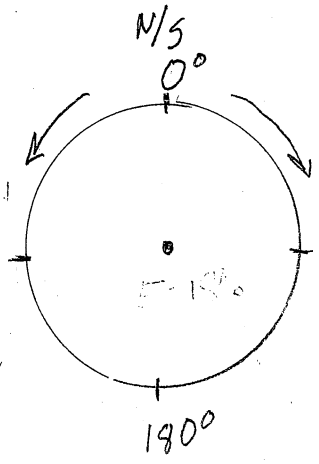
Kibbutz Sasa



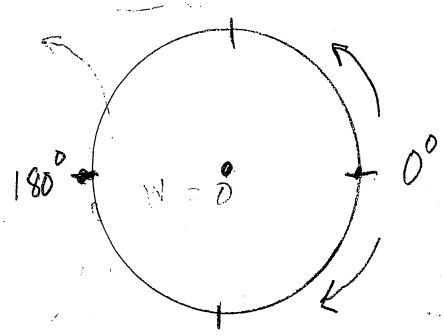
E/W

☐ Aspect:





South: more sun (⇒ fewer trees)

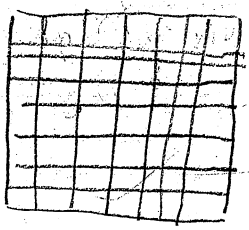


West: more moisture from Mediterranean

Remarks:

Full data set: $75 \times 80 = 6{,}000$ pixels

$\Rightarrow$ spatial dependence



- Original response:

$$y = \begin{cases} \%\text{ tree cover} \\ \%\text{ shrub cover} \\ \%\text{ grass} \end{cases}$$

- multivariate response

constraint: $\mu_1 + \mu_2 + \mu_3 = 1$

$\Rightarrow$ compositional data

- Could define vegetation index

$y = 1$ (all grass), ..., 7 (all trees)

ordered response

$\Rightarrow$ y is ordinal variable, ordered polychotomous regression

- Simplification

$$y = \begin{cases} 1 & \text{if trees dominant } (\%\text{tree} > \begin{cases} \%\text{shrub}, \ \%\text{grass} \\ \text{or} \\ \%\text{shrub} + \%\text{grass} \end{cases}) \\ 0 & \text{if not} \end{cases}$$

☐ Exploratory Analysis of Meron data {meron.preliminary.pdf}

{p.1}

— confounding of grazing type & intensity (like loc-host and south in medfly data)

{p.2}

— ew. aspect only up to 100 (obs. in a region ≈ sloping towards the east)

{p.3}

{p.4} Effect on trees:

— fewer trees (tree.82 = 0) when N/S aspect high

(N/S = 0 ⟺ north; so high N/S means ≈ south and more sun)

— more trees (tree.82 = 1) when E/W aspect high

(E/W = 0 ⟺ east, so high E/W aspect means ≈ west ⇒ wake from sea)

— higher altitude ≈ lower road distance (r = −0.73)

↓

fewer people

{p.5} ☐ linear regression results.

- interpretation of slope, aspect, slope * aspect coeff's - later
- cattle at low intensity??
  - fertilizer ?
  - {p.3: cattle.low ≈ ⟷ { alt. high; E/W aspect ≈ high (W: more moisture)}

☐ Before fitting GLM to Meron, another link function

☐ So far- formula page

Canonical links $\{g(\mu_i) = \theta_i\}$

$x_i^T \beta = g(\mu_i) = \mu_i$ (normal)

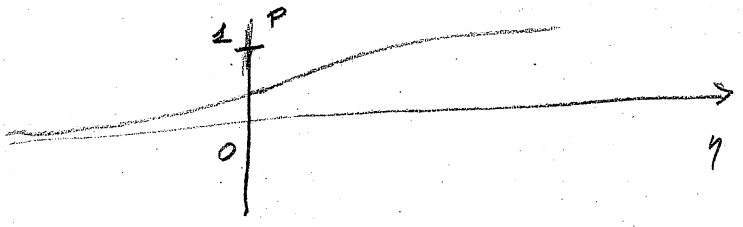
$= g(\lambda_i) = \log(\lambda_i)$ (Poisson)

$= g(p_i) = \log \frac{p_i}{1-p_i}$ (Bernoulli)

☐ Another link function for Bernoulli:

Want $g: (0,1) \to (-\infty, \infty)$

$g^{-1}: (-\infty, \infty) \to (0,1)$



$g^{-1}(\eta) = \Phi(\eta)$, Φ = N(0,1) CDF

$g(p) = \Phi^{-1}(p)$, probit link function

("probability unit")

☐ Interpretation of logit: $\log\left(\frac{p}{1-p}\right)$ = log "odds"

Interpretation of probit?

$$y_i \sim \text{Ber}(p_i),$$

$$p_i = \Phi(\eta_i), \qquad \eta_i = \tilde{x}_i^t \tilde{\beta}$$

$$= P(Z_i \le \eta_i), \qquad Z_i \sim N(0,1).$$

Interpretation / motivation:

$y = 1$ when underlying latent variable W exceeds an unknown threshold c.

($Z = -W$:

$W > c$

$\Leftrightarrow$

$Z < -c$)

□ EG W = "heart risk index"

$$\sim N(\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2, \sigma^2)$$

where x_1 = BMI (γ_i probably > 0)

x_2 = cholesterol level

$y = 1$ if has heart attack.

□ Model: $y = 1\{W > c\}$ where c is unknown threshold

Then

$$P(y=1) = P(W > c)$$

$$= P(\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \sigma\varepsilon > c), \quad \varepsilon \sim N(0,1)$$

$$= P\left(-\varepsilon \leq \frac{\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 - c}{\sigma}\right)$$

$$= P\left(Z < \left(\frac{\gamma_0 - c}{\sigma}\right) + \left(\frac{\gamma_1}{\sigma}\right) x_1 + \left(\frac{\gamma_2}{\sigma}\right) x_2\right), \quad Z = -\varepsilon \sim N(0,1)$$

(β_0, β_1, β_2)

⇒ Probit model

□ Note: $\gamma_j > 0 \iff \beta_j > 0$

□ logit or probit? Saw {p. 21, notes}

- that for any $\underset{\sim}{x} = (x_1, \underbrace{x_2, \ldots x_p}_{\underset{\sim}{x}_{(2)}^t})^t$,

and $\frac{p}{1-p}$ = "odds for $y=1$",

$$\frac{\text{odds}(x_1+\Delta;\ \underset{\sim}{x}_{(2)})}{\text{odds}(x_1, \underset{\sim}{x}_{(2)})} = e^{\beta_1 \Delta}$$

Q: How interpret β_1?

$$\Rightarrow \frac{\text{odds}(x_1+\Delta, \underset{\sim}{x}_{(2)}) - \text{odds}(x_1, \underset{\sim}{x}_{(2)})}{\Delta} = \text{odds}(x_1, \underset{\sim}{x}_{(2)}) \frac{[e^{\beta_1 \Delta} - 1]}{\Delta}$$

$\Delta \to 0$

$$\Rightarrow \frac{\partial}{\partial x_1} \text{odds}(x_1, \underset{\sim}{x}_{(2)}) = \text{odds}(x_1, \underset{\sim}{x}_{(2)}) \cdot \underbrace{\frac{d}{dx} e^{\beta_1 x}\Big|_{x=0}}_{\beta_1}$$

□ So β_j gives change in odds as fn of x_j on a multiplicative scale,

for any $\underset{\sim}{x}$.

So:

- Motivation for probit:

possible latent variable

- Motivation for logit:
  - "log odds" interpretation, of β's for any $\underset{\sim}{x}$
  - also useful in retrospective sampling.

□ Y = rare disease = $\begin{cases} 1 \\ 0 \end{cases}$, X = risk factor = $\begin{cases} 1 \\ 0 \end{cases}$

Want $P(Y=1|X=1)$ vs $P(Y=1|X=0)$

Prospective sample:

100 individuals w $X=0$

100 — " — $X=1$ observe Y & compare

Problem: might get all $Y_i = 0$

□ Retrospective Sample

100 with $Y=0$ compute X & compare

100 — $Y=1$

Problem Can only estimate $P(X|Y)$; want $P(Y|X)$

□ Sol'n: Use data to estimate odds ratio

$$\frac{\text{odds}(X=1|Y=1)}{\text{odds}(X=1|Y=0)} = \frac{P(X=1|Y=1)/[1-P(X=1|Y=1)]}{P(X=1|Y=0)/[1-P(X=1|Y=0)]}$$

$$= \frac{P(X=1|Y=1)\cdot P(X=0|Y=0)}{P(X=1|Y=0)\cdot P(X=0|Y=1)} \cdot \frac{P(Y=1)}{P(Y=1)} \cdot \frac{P(Y=0)}{P(Y=0)}$$

$$= \frac{P(Y=1, X=1) \cdot P(Y=0, X=0)}{P(Y=0, X=1) \cdot P(Y=1, X=0)}$$

$$= \frac{P(Y=1, X=1) / P(Y=0, X=1)}{P(Y=1, X=0) / P(Y=0, X=0)} \cdot \frac{P(X=1)/P(X=1)}{P(X=0)/P(X=0)}$$

$$= \frac{P(Y=1 | X=1) / P(Y=0 | X=1)}{P(Y=1 | X=0) / P(Y=0 | X=0)}$$

$$= \frac{\text{odds}(Y=1 | X=1)}{\text{odds}(Y=1 | X=0)}$$

☐ Example ... logistic & probit regression

(1) Reproduce ... logit ...

(2) Probit ≈ logit. ... more variables are significant

(3) Compare logit & probit curves

... slope ?

$(\hat{\beta} > 0)$

☐ Now: GLIM for Meron {meron.glim.fit.pdf}

Linear regression (from before) significant $\hat{\beta}_i$:

alt, water.dist, cattle.low $\hat{\beta}_i > 0$

stream.dist, slope:ns.aspect $\hat{\beta}_i < 0$

Logistic regression

alt, water.dist $\hat{\beta}_i > 0$

stream.dist, slope: ns.aspect $\hat{\beta}_i < 0$

Probit regression

alt, water.dist, ns.aspect $\hat{\beta}_i > 0$

stream.dist, slope: ns.aspect $\hat{\beta}_i < 0$

☐ ☐ - Deviance Residuals

- Dispersion parameter ($= \phi$)

- Null, Residual deviance

later

later

- AIC

☐ Question (probit results):

A = N/S aspect {A = 0: North}

S = slope

$\hat{\beta}_A = 0.2$

$\hat{\beta}_S = 4.8$

$\hat{\beta}_{A \times S} = -0.18$

Is facing north (A = 0) good or bad for trees?

☐ Answer depends on slope

$$\Phi^{-1}(p) \approx c + 4.8\,S + 0.02\,A - 0.18\,A \times S$$

$$= C + 0.02\,A + [4.8 - 0.18\,A] \cdot S$$

- Need a graph {meron.interaction.graphs.pdf}
- For $n = 6{,}000$, assuming Y_i independent

solid line is top curve

☐ Top graph: facing North is good for trees and

Bottom graph: facing West " " " "

IF the slope is large enough