INTRO TO GLMs (GENERALIZED LINEAR MODELS) (CHAPTER 4)

MODEL COMPONENTS

SUPPOSE DENSITY OF R.V. IS OF THE FORM:

$$f(y;\theta) = \underbrace{a(\theta)}\underbrace{b(y)}\underbrace{\exp\{y\,Q(\theta)\}}$$

HERE $Q(\theta)$ IS "NATURAL PARAMETER"

EXAMPLES: $Y \sim \text{POIS}(\theta)$

$$f(y;\theta) = \frac{\exp\{-\theta\}\theta^y}{y!} = \frac{1}{y!}\exp\{-\theta\}\exp\{\log\theta^y\} =$$

$$= \underbrace{\frac{1}{y!}}_{b}\underbrace{\exp\{-\theta\}}_{a}\underbrace{\exp\{y\log\theta\}}_{y\cdot Q(\theta)}$$

HERE $Q(\theta) = \log\theta$ IS NATURAL PARAMETER.

EXAMPLE $Y \sim \text{BERN}(\pi)$

$$f(y;\pi) = \pi^y(1-\pi)^{1-y} = (1-\pi)\left(\frac{\pi}{1-\pi}\right)^y =$$

$$= (1-\pi)\exp\left\{\log\left(\frac{\pi}{1-\pi}\right)^y\right\} =$$

$$= (1-\pi)\exp\left\{y\log\left(\frac{\pi}{1-\pi}\right)\right\}$$

$$Q(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$ $\longrightarrow$ LOOK FAMILIAR?

$Q(\pi)$ IS THE LOG ODDS.

LET $Y_i \sim f(y_i; \theta_i) = a(\theta_i)\, b(y_i) \exp\{y_i Q(\theta_i)\}$

& $\mu_i = \mathbb{E} Y_i$ (FOR $i = 1, \dots, N$) & Y_i ARE IND.

LET $x_{i1}, \dots, x_{ik}$ BE (MEASUREMENT, COVARIATES, EXPLANATORY, $\boxed{\text{~~INDEPENDENT~~}}$)

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

↳ * Y_i ARE THE RANDOM COMPONENT
* η_i ARE THE SYSTEMATIC COMPONENT

THE LINK FUNCTION g IS A MAPPING BETWEEN μ_i & η_i, (g MUST BE MONOTONE, DIFFERENTIABLE)

$$g(\mu_i) = \eta_i$$

DEFINITIONS:
- $g(\mu_i) = \mu_i$ (IDENTITY LINK)
- $g(\mu_i) = Q(\theta_i)$ (CANONICAL LINK)

IDENTITY LINK IN BERNOULLI MODEL:

$$g(\pi_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_K x_{iK}$$

PROBLEM: GET VALUES $(-\infty, 0) \cup (1, \infty)$
(THAT DON'T MAKE VALID PROBABILITIES)

FOR CANONICAL LINK:

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_K x_{iK}$$

NOW, π_i IS STILL CONSTRAINED TO $[0,1]$. ALSO, WE HAVE MODELED "LOG ODDS" AS LINEAR FUNCTION OF COVARIATES.

IN GLMS, WE ARE CHOOSING NON-IDENTITY LINKS.

AN IMPLICATION IS THAT SQUARED RESIDUALS

DON'T MEAN SAME THING (R^2, ...)

↳ DEVIANCE.

LET $\underset{\sim}{y} = (y_1, \ldots, y_N)^T$, $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_N)$

$$L(\underset{\sim}{\mu}; \underset{\sim}{y}) \quad \text{(log likelihood)}$$

FIND UNCONSTRAINED MLE

$$\max_{\underset{\sim}{\mu}} L(\underset{\sim}{\mu}; \underset{\sim}{y}) = L(\underset{\sim}{y}; \underset{\sim}{y})$$

NOW, INSTEAD DEFINE RESTRICTED MLE, μ_i IS

OF THE η_i. CALL THIS $\hat{\underset{\sim}{\mu}}$.

$$\text{DEVIANCE}(\hat{\underset{\sim}{\mu}}) \doteq -2\left(L(\hat{\underset{\sim}{\mu}}, \underset{\sim}{y}) - L(\underset{\sim}{y}, \underset{\sim}{y}) \right)$$

↳ LIKELIHOOD RATIO TEST.

LOGISTIC REGRESSION

$$Y \mid X = x \sim \text{BERN}(\pi(x)),$$

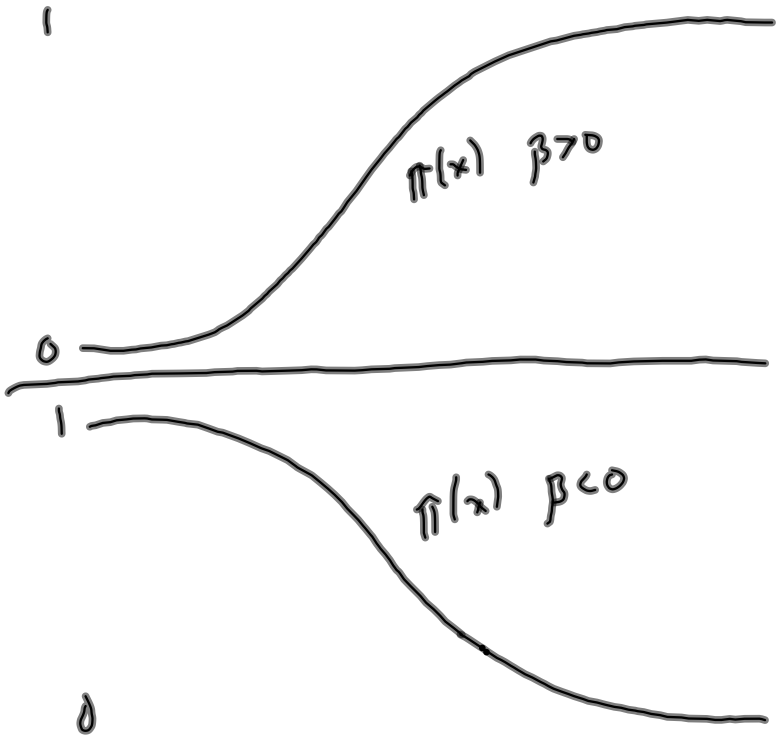
$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \Leftrightarrow \pi(x) = \underbrace{\frac{\exp\{\alpha+\beta x\}}{1+\exp\{\alpha+\beta x\}}}$$

NOTES:

1) MODEL IS GLM FOR CANONICAL LINK
2) log ODDS $Y=1 \mid X=x$ IS LINEAR IN x
3) $\beta > 0$: $x \to \infty$, $\pi(x) \to 1$

   $\beta < 0$: $x \to \infty$, $\pi(x) \to 0$

   $\beta = 0$: $\pi(x) \equiv \pi$



$$\frac{\pi(x)}{1-\pi(x)} = \exp\{\alpha+\beta x\} =$$

$$= \exp\{\alpha\}\underbrace{\exp\{\beta x\}}$$

x, $x' = x+1$.

$$\frac{\exp\{\beta(x+1)\}}{\exp\{\beta x\}} = \exp\{\beta\}$$

"A ONE UNIT CHANGE IN x, LEADS TO A MULTIPLICATIVE CHANGE IN THE ODDS (OF $Y=1$) BY $\exp\{\beta\}$"

AS OPPOSED TO:

LINEAR REG.

"A ONE UNIT CHANGE IN x LEADS TO AN ADDITIVE CHANGE IN Y BY β."