3.1 → Introduction

→ link function → $g(\mu_i) = x_i^T \beta$

3.2 → Exponential family of distributions

→ Let Y be a variable whose probability distribution depends on a single parameter θ. The distribution $\in$ exponential family if it can be written in the form

$$f(y|\theta) = s(y)\,t(\theta)\,e^{a(y)\,b(\theta)}$$

known functions

$$= \exp\{\ln[s(y)t(\theta)e^{a(y)b(\theta)}]\}$$

$$\boxed{= \exp\{a(y)\,b(\theta) + c(\theta) + d(y)\}}$$

$s(y) = e^{d(y)} \Rightarrow d(y) = \ln(s(y))$

$t(\theta) = e^{c(\theta)} \Rightarrow c(\theta) = \ln(t(\theta))$

→ terms

→ if $a(y) = y \Rightarrow$ canonical form (standard)

→ $b(\theta)$ = natural parameter

→ other parameter are nuissance parameters $\Rightarrow$ treated as known

→ Examples →

Dist	Natural parameter	c	d
Poisson	$\ln(\theta)$	$-\theta$	$-\ln(y!)$
Normal	μ/σ^2	$-\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)$	$-\frac{y^2}{2\sigma^2}$
Binomial	$\ln(\pi/1-\pi)$	$n\ln(1-\pi)$	$\ln\binom{n}{y}$

→ Poisson dist

→ $f(y|\theta) = \frac{e^{-\theta}\theta^y}{y!} = \exp\{\ln[e^{-\theta}\theta^y/y!]\}$

$$= \exp\{\underbrace{-\theta}_{c(\theta)} + \underbrace{y}_{a(y)}\underbrace{\ln(\theta)}_{b(\theta)} - \underbrace{\ln(y!)}_{d(y)}\}$$

$\Rightarrow$ canonical form

→ Poisson models count data

→ Real data often has higher variance $\Rightarrow$ overdispersed (see ch 7)

→ Normal dist

→ $f(y|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$

$$= \exp\{\ln[\langle\ldots\rangle]\}$$

$$= \exp\{-\tfrac{1}{2}\ln(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}(y-\mu)^2\}$$

$$= \exp\{-\tfrac{1}{2}\ln(2\pi\sigma^2) - \underbrace{\tfrac{1}{2\sigma^2}y^2}_{d(y)} + \underbrace{\tfrac{1}{\sigma^2}\mu}_{b(\mu)}\underbrace{y}_{a(y)} - \tfrac{1}{2\sigma^2}\mu^2\}$$

$\Rightarrow$ canonical form

$c(\mu)$ (could be included differently)

→ Normal models continuous, symmetric data

→ CLT $\Rightarrow$ even if skewed, average is $\approx$ normal

→ lots of theoretical properties $\Rightarrow$ want to transform to approx normal if possible

→ e.g. $y' = \sqrt{y}$ or $y' = \ln(y)$

→ Binomial dist

→ $f(y|\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$

$$= \exp\{\ln[\langle\ldots\rangle]\}$$

$$= \exp\{\ln\binom{n}{y} + y\ln(\pi) + (n-y)\ln(1-\pi)\}$$

$$= \exp\{\ln\binom{n}{y} + y\ln(\pi) + n\ln(1-\pi) - y\ln(1-\pi)\}$$

$$= \exp\{\underbrace{\ln\binom{n}{y}}_{d(y)} + \underbrace{y}_{a(y)}\underbrace{[\ln(\pi) - \ln(1-\pi)]}_{b(\pi)\,=\,\ln\left(\frac{\pi}{1-\pi}\right)} - \underbrace{n\ln(1-\pi)}_{c(\pi)}\}$$

→ canonical form

→ Binomial dist models processes w/ binary outcomes

→ etc.

3.3 → Properties of distributions in the exponential family

→ Need expressions for $E[a(y)]$ & $V[a(y)]$

→ needed results (assuming order of integration & derivation can be interchanged) (true for any dist)

→ $\int f(y|\theta)\,dy = 1$

→ differentiate both sides

$$\frac{d}{d\theta}\int f(y|\theta)\,dy = \frac{d}{d\theta}1 = 0$$

→ reverse order of derivative & integral

$$\int \frac{d}{d\theta} f(y|\theta)\,dy = 0$$

→ differentiate twice

$$\int \frac{d^2}{d\theta^2} f(y|\theta)\,dy = 0$$

→ These results can now be used for distributions in the exponential family

→ $f(y|\theta) = \exp\{a(y)b(\theta) + c(\theta) + d(y)\}$

$\Rightarrow \frac{d f(y|\theta)}{d\theta} = [a(y)b'(\theta) + c'(\theta)]\underbrace{f(y|\theta)}_{= e^{a(y)b(\theta)+c(\theta)+d(y)}}$ → $\left(\frac{d}{dx}e^{g(x)} = g'(x)e^{g(x)}\right)$

$$\Rightarrow 0 = \int \frac{d}{d\theta}f(y|\theta)\,dy = \int [a(y)b'(\theta) + c'(\theta)]f(y|\theta)\,dy$$

$$= \int (a(y)b'(\theta)f(y|\theta) + c'(\theta)f(y|\theta))\,dy$$

$$= b'(\theta)\underbrace{\int a(y)f(y|\theta)\,dy}_{= E(a(y))} + c'(\theta)\underbrace{\int f(y|\theta)\,dy}_{=1}$$

$$= b'(\theta)E[a(y)] + c'(\theta)$$

$$\Rightarrow \boxed{E[a(y)] = \frac{-c'(\theta)}{b'(\theta)}}$$

→ Similarly, $0 = \int \frac{d^2}{d\theta^2} f(y|\theta)\,dy$ → product rule $\frac{d}{dx}g(x)f(x) = g'(x)f(x) + f'(x)g(x)$

$$= \int\left[(a(y)b''(\theta) + c''(\theta))f(y|\theta) + [\underbrace{a(y)b'(\theta) + c'(\theta)}_{= a(y)b'(\theta) - b'(\theta)E(a(y))}]^2 f(y|\theta)\right]dy$$

$$= b''(\theta)\int a(y)f(y|\theta) + c''(\theta)\int f(y|\theta) + \int[(b'(\theta))^2(a(y) - E(a(y)))^2]f(y|\theta)\,dy$$

$$= \text{— — —} \quad (b'(\theta))^2\int(a(y) - E(a(y)))^2 f(y|\theta)\,dy$$

$$= b''(\theta)E[a(y)] + c''(\theta) + (b'(\theta))^2 V[a(y)]$$

$$\Rightarrow V(a(y)) = \frac{-b''(\theta)E[a(y)] - c''(\theta)}{(b'(\theta))^2}$$

$$= \frac{b''(\theta)\frac{c'(\theta)}{b'(\theta)} - c''(\theta)}{(b'(\theta))^2}$$

$$\boxed{= \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{(b'(\theta))^3}}$$

→ Can use $\boxed{\text{equations}}$ to verify formulas for normal, Poisson, binomial, etc.

→ Expressions are also needed for expected value & variance of derivatives of the log-likelihood function

for a distribution in the exponential family, the log likelihood is

→ $\ell(\theta|y) = \ln[\exp\{a(y)b(\theta) + c(\theta) + d(y)\}] = a(y)b(\theta) + c(\theta) + d(y)$

→ derivative of $\ell(\theta|y)$ w/ respect to θ

$$\boxed{U = \frac{d\ell(\theta|y)}{d\theta} = a(y)b'(\theta) + c'(\theta)}$$

↳ score statistic

→ depends on $Y \Rightarrow$ it is a RV

$U = a(y)b'(\theta) + c'(\theta)$

→ $E(U) = E[a(y)b'(\theta) + c'(\theta)]$

$= b'(\theta)E(a(y)) + E(c'(\theta))$

$= -b'(\theta)\frac{c'(\theta)}{b'(\theta)} + c'(\theta)$

$\boxed{= 0}$

→ $V(U) = V[a(y)b'(\theta) + c'(\theta)]$

$= (b'(\theta))^2 V(a(y))$

$= (b'(\theta))^2\left[\frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{(b'(\theta))^3}\right]$

$$\boxed{= \frac{b''(\theta)c'(\theta)}{b'(\theta)} - c''(\theta) = \mathcal{J} = \underline{\text{Information}}}$$

→ The score statistic is used for inference about parameter values in glms

→ Property of score statistic → $V(U) \overset{①}{=} E(U^2) \overset{②}{=} -E(U')$

① $E(U^2) - \underbrace{(E(U))^2}_{=0}$

② $U' = \frac{dU}{d\theta} = \frac{d}{d\theta}[a(y)b'(\theta) + c'(\theta)]$

$= a(y)b''(\theta) + c''(\theta)$

$E(U') = E(\langle\ldots\rangle)$

$= b''(\theta)E[a(y)] + c''(\theta)$

$= -b''(\theta)\frac{b'(\theta)}{c'(\theta)} + c''(\theta)$

$= -V(U) = -\mathcal{J}$

3.4 → GLMs

→ GLM definition → $Y_1, \ldots, Y_N$ are $\perp\!\!\!\perp \sim$ Exponential family

→ 1) each Y_i has canonical form & depends on single parameter θ_i (don't have to be the same)

$\Rightarrow f(y_i|\theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)]$

→ 2) distributions of all Y_i's are of the same form (e.g.) all normal, or Binomial)

$\Rightarrow$ subscripts on a, b, c & d are not needed

→ Joint pdf

$$f(y_1, \ldots, y_N|\theta_1, \ldots, \theta_N) = \prod_{i=1}^{N}\exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)]$$

$$\boxed{= \exp\left[\sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right]}$$

→ parameters θ_i are usually not of direct interest (could be one for each obs).
For model specification, usually interested in a smaller set of parameters $\beta_1, \ldots, \beta_p$ ($p < N$)
Let $E(Y_i) = \mu_i$ where μ_i is some function of θ_i

→ for a GLM, there is a transformation of μ_i such that

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

→ ☆ This equation $\Rightarrow$ 3 parts of glm model

→ 1) Response variable $Y_i \overset{iid}{\sim} \underbrace{f(\theta)}$ ↳ exponential family

2) Set of parameters & explanatory variables $\beta = \begin{bmatrix}\beta_1\\ \vdots \\ \beta_p\end{bmatrix}$ & $x_i = \begin{bmatrix}x_{i1}\\ \vdots \\ x_{ip}\end{bmatrix}$

$$X = \begin{bmatrix}x_1^T\\ \vdots\\ x_N^T\end{bmatrix} = \begin{bmatrix}x_{11} & \cdots & x_{1p}\\ \vdots & & \vdots\\ x_{N1} & \cdots & x_{Np}\end{bmatrix}$$

↳ design matrix

→ link function g is monotone & differentiable such that

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

↳ $= E(Y_i)$

3.5 → Examples

→ Normal linear model → $E(Y_i) = \mu_i = x_i^T\beta$, $Y_i \overset{\perp\!\!\!\perp}{\sim} N(\mu_i, \sigma^2)$

↳ $g(\mu_i) = \mu_i$ (identity)

$\Rightarrow \underbrace{Y = X\beta}_{\text{signal}} + \underbrace{\varepsilon}_{\text{noise}}$ ↳ $\overset{iid}{\sim} N(0, \sigma^2)$

→ MLR, ANOVA, & ANCOVA are all this form

→ Bernoulli model

→ $Y_i = \begin{cases}1\\0\end{cases}$

$P(Y_i = 1) = e^{-\theta t} = \pi$ (parameter)

$\Rightarrow P(Y_i = 0) = 1 - e^{-\theta t} = 1 - \pi$

→ Interested in θ rather than π

$\Rightarrow g(\pi) = \ln(\pi) = \ln(e^{-\theta t}) = -\theta t$

↳ $= E(Y_i)$

$\Rightarrow g(\pi)$ is linear in parameter θ & $x_i = [-t]$, $\beta = [\theta]$

→ Poisson model

→ $f(y|\mu) = \frac{e^{-\mu}\mu^y}{y!}$, $y = 0, 1, 2, \ldots$

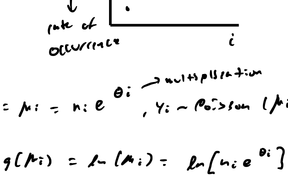
→ $E(Y) = \mu$

→ parameter μ depends on various characteristics of the population & the population size

$\Rightarrow E(Y) = \mu = n\underbrace{\lambda}(x^T\beta)$

↳ function depends on

→ currently unknown ☆



Linear on log scale

$\Rightarrow$ exponential growth

$Y = e^{f(x)}$

Figured it out ☆

→ $E(Y_i) = \mu_i = n_i e^{\theta_i}$ (→ multiplication), $Y_i \sim \text{Poisson}(\mu_i)$

→ $g(\mu_i) = \ln(\mu_i) = \ln[n_i e^{\theta_i}] = \ln n_i + \theta_i$

$\Rightarrow$ linear component $x_i^T\beta$ → $= \begin{bmatrix}1\\ \theta\end{bmatrix}$

↳ $= [\ln n_i \;\; i]$

Same process as Bernoulli model

$\Rightarrow E(Y) = \mu =$ some function of θ

$g(\mu) =$ linear comb. of θ