Example 2 Generalyed linear model (GLM)

GLM extend the LM to accommodate non-normal responses. They were introduced in 1972 by Nelder and Wedderburn.

Denote data $(y_i, x_i)$ $i=1,..n$ $y_i \in \mathbb{R}$ response and $x_i \in \mathbb{R}^p$ covariate. Assume $y_i$ arises from model

(1) $Y_i \sim f(y_i, \theta_i, \phi)$; $f(y_i; \theta_i, \phi) = \exp\left\{ \dfrac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} \right\} h(y_i, \phi)$

$Y_i \sim EF(\theta_i, \phi)$ $\theta_i$ natural/canonical para

$\phi$ = dispersion para

$a_i(\cdot)$ = known fn.

(2) $g(EY_i) = x_i^T\beta$ where $g(\cdot)$ is known function which is monotone increasing

"$g$" is called link fn b/c links response to predictor
"$x_i^T\beta$" is called linear predictor

Obs There seem to be a lot of para! Identify the para of this model! (class exercise)

Recall $EY_i = b'(\theta_i)$ ; $Var\, Y_i = b''(\theta_i)\, a_i(\phi)$
(imagine $y_i \to y_i / a_i(\phi)$)

$g(b'(\theta_i)) = x_i^T\beta$ $\Rightarrow$ $\theta_i = b'^{-1}(g^{-1}(x_i^T\beta))$

Fortunately most cases use a simpler form of $g(\cdot)$, $g(EY_i)=\theta$
The fn $g(\cdot)$ s.t. $\underline{g(EY_i) = \theta_i}$ is called natural/canonical link fn.

Note This fn will be different according to the dist'n assumed for Y.

Eg If $\underline{Y_i \sim \text{Bernoulli}(p_i)}$ $\Rightarrow$ $\theta_i = \log \frac{p_i}{1-p_i}$ is the natural para. $EY = p = \frac{e^\theta}{1+e^\theta}$

$g$ is s.t. $g\left(\frac{e^\theta}{1+e^\theta}\right) = \theta$ $\Rightarrow g(\cdot) = \ldots\ldots ?$

$\frac{e^\theta}{1+e^\theta} = \mu \iff \theta = \log \frac{\mu}{1-\mu}$ $\qquad \underline{g(\mu) = \log \frac{\mu}{1-\mu}}$ "logit fn"

If $Y \sim \text{Normal}(\mu, \sigma^2)$

$$f(y; \mu, \sigma^2) = e^{-\frac{(y-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} = e^{-\frac{y^2}{2\sigma^2} + \frac{2\mu y}{2\sigma^2} - \frac{\mu^2}{2\sigma^2}} e^{-\frac{1}{2}\log 2\pi\sigma^2}$$

$$= e^{\frac{y\mu - \mu^2/2}{\sigma^2}} e^{-y^2/2\sigma^2 - 1/2 \log 2\pi\sigma^2}$$

$\Rightarrow b(\mu) = \mu^2/2$ ; $a(\phi) = \sigma^2$ $\quad \phi = \sigma^2$

$h(y, \phi) = \exp\left(-y^2/2\sigma^2 - \frac{1}{2}\log 2\pi\sigma^2\right)$

$\mu$ is the cannonical para.

The canonical link $g$ satisfies $g(EY) = \mu$; in this case $g(\mu) = \mu$ is the "identity fn".

Thus para is : $\beta$ (regression para) & $\phi$ (dispersion para)

How to write the density when $g$ is the canonical link?

$$f(y_i, \beta, \phi) = \exp\left\{ \frac{y_i x_i^T\beta - b(x_i^T\beta)}{a_i(\phi)} \right\} h(y_i, \phi)$$

Likel fn is $L(\beta, \phi) = \prod_{i=1}^{n} e^{\frac{y_i x_i^T\beta - b(x_i^T\beta)}{a_i(\phi)}} \; h(y_i, \phi)$

$$\ell(\beta, \phi) = \sum_{i=1}^{n} \frac{y_i x_i^T\beta - b(x_i^T\beta)}{a_i(\phi)} + \sum_{i=1}^{n} \log h(y_i; \phi)$$

$(\hat{\beta}_{MLE}, \hat{\phi}_{MLE})$ max the log-likel fn.

Unfortunately the analytical expression by using the usual route (taking derivatives, setting them to zero) cannot be obtained.

Instead the MLE in this case is obtained numerically and we'll discuss some techniques later in the chapter.

Read the text for:
- Generalized linear mixed model (GLMM
- Accelerated failure time model (ACT
- longit analysis
- survival analysis

Final remarks about the likelihood fn.

In statistical inference, the goal is to draw conclusions regarding the underlying dist'n of $Y$ on the basis of observing $Y=y$. In particular, for para models this is equivalent to drawing conclusions re the unknown model para, say $\theta$. As we discussed last time $L(\theta; y)$ measures how likely different values of $\theta$ are to be the true value of $\theta$.

$\boxed{\text{likel principle}}$ (strong version)

Suppose we have an observation $y$ from a statistical model $\{f_1(\cdot;\theta)\ \theta \in \Theta\}$. and an observation $x$ from model $\{f_2(\cdot;\theta), \theta \in \Theta\}$; here $\theta$ has the same meaning in the 2 models.

Define $L_1(\theta)$ the likel corresp to model 1
$L_2(\theta)$ — " — 2

If $L_1(\theta) = L_2(\theta)$ for all $\theta$ $\Rightarrow$ all conclusions of $\theta$ based on observing $y$ should be the same as the conclusions based on $x$.

This principle is considered "strong" and it was received with criticism. Here's why.

Example $\quad Y \sim \text{Binomial}(10, \theta) \quad y=5$

$$P(Y=y) = \binom{10}{y} \theta^y (1-\theta)^{10-y} \quad y=0,1,\dots,10$$

If $y=5 \Rightarrow L_1(\theta) = \theta^5 (1-\theta)^5$ by ignoring constant

$X \sim$ Negative Binomial $(r, \theta)$ $\qquad \theta =$ probab of success

$X$ - measures the number of successes until $r$ failures occur.

$$P(X = x; \theta) = \binom{x+r-1}{x} (1-\theta)^r \theta^x \qquad x = 0, 1, 2, \ldots$$

Take $r = 5$ and $x = 5$.

$$L_2(\theta) = (1-\theta)^5 \theta^5 \qquad \text{So that } L_1(\theta) = L_2(\theta).$$

According to the likel principle, our conclusions about $\theta$ should be the same, irrespective of how the data was collected (Bernoulli or Neg Binomial mechanism!)

Does this make sense?

This means that the dist'n of the quantity that is used to estimate the para is irrelevant!

This principle is in contrast with another principle - the repeated sampling principle. The repeated sampling principle states that the statistical procedures should be evaluated on the basis of their behavior in hypothetical repetitions of the experiment that generated the original data.

Idea: a stat procedure should be evaluated in terms of bias, variance, coverage probas etc and thus the model is important too! Not just the likel fn!

Info beyond that provided by the likelihood fn is necessary for proper statistical inference.

## (Weak) Likelihood principle (Severini 2000)

Suppose we have observations $y_1$, and $y_2$ from the statistical model $f(y;\theta)$. If $L(\theta;y_1) \equiv L(\theta;y_2)$ then inference for $\theta$ should be the same irrespective of whether $y_1$ or $y_2$ is observed.