- Natural (or canonical) exponential families

- Generalised Linear Models for data

- Fitting GLM’s to data
  MLE’s
  Iteratively Re-weighted Least Squares
Natural exponential families

Restrict observation model to $Y \sim f(y|\theta)$

$$f(y|\theta) = \exp \left( \frac{y\theta - \kappa(\theta)}{\phi} + c(y; \phi) \right), \quad y \in \Omega$$

$\phi = 1$ natural exponential family

$\phi > 0$ natural exponential dispersion family.

$$\exp(\kappa/\phi) = \int_{\Omega} \exp \left( \frac{y\theta}{\phi} + c(y) \right) \, dy$$

$\theta \in \{\theta; \kappa(\theta) < \infty\}$

$y$ natural observation

$\theta$ natural parameter

$\phi$ dispersion parameter

Note: discussion here restricted to natural exponential families with a scalar parameter. For general characterisation of EF's see recommended reading.
Normal: $Y \sim N(\mu, \sigma^2)$ is NEDF

$$f(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y - \mu)^2}{2\sigma^2} \right)$$

$$= \exp \left( \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{\log(2\pi\sigma^2)}{2} - \frac{y^2}{2\sigma^2} \right)$$

$$= \exp \left( \frac{y\theta - \kappa(\theta)}{\phi} + c(y; \phi) \right)$$

$\theta = \mu$, $y \in \mathbb{R}$, $\phi = \sigma^2$,
$\kappa = \mu^2/2$
$c(y; \phi) = -(1/2) \log(2\pi\phi) - y^2/2\phi$

Student’s-t: $Y \sim t(\nu)$ is not in this class

$$f(y|\nu) \propto (1 + y^2/\nu)^{-\left(\frac{\nu+1}{2}\right)}$$

cant factorise $y$ and $\nu$ in exponent.
\[ e^{\kappa/\phi} = \int_{\Omega} \exp \left( \frac{y\theta}{\phi} + c(y) \right) \, dy \]

\[ \kappa(\theta) \text{ cumulant GF (at } \phi = 1). \]

\[ \frac{\kappa'}{\phi} e^{\kappa/\phi} = \int_{\Omega} \left( \frac{y}{\phi} \right) \exp \left( \frac{y\theta}{\phi} + c(y) \right) \, dy \]

\[ \kappa' = E(Y) \]

\[ \left( \frac{(\kappa')^2}{\phi^2} + \frac{\kappa''}{\phi} \right) e^{\kappa/\phi} = \int_{\Omega} \left( \frac{y}{\phi} \right)^2 \exp \left( \frac{y\theta}{\phi} + c(y) \right) \, dy \]

\[ \left( \frac{(\kappa')^2}{\phi^2} + \frac{\kappa''}{\phi} \right) = E(Y^2)/\phi^2 \]

\[ \text{var}(Y) = E(Y^2) - E(Y)^2 \]

\[ = (\kappa')^2 + \phi \kappa'' - (\kappa')^2 \]

\[ = \phi \kappa''. \]

Let \( \mu_i = E(Y_i) \), \( \text{var}(Y_i) = \phi V(\mu_i) \).

**Exercise** \( \frac{d\mu_i}{d\theta_i} = V(\mu_i) \) so \( \mu_i \) increases with \( \theta_i \).
Modeling with GLM’s (intro)

Suppose we have a response $y$ and some (explanatory) variables $x$.

Let $\eta = x^T \beta$. In a NLM the response

$$y \sim N(\eta(\beta), \sigma^2)$$

depends on the parameters $\beta$ only through the linear combination $\eta$. Also, as $\eta$ gets bigger so does $E(Y)$, since $E(Y) = \eta$.

This makes it relatively easy to interpret the effect of the parameters $\beta$. If $x_i > 0$ then positive (increasing) $\beta_i$ is evidence for a positive (increasing) effect on the response.

Let’s generalise this but try and keep the interpretability.
Allow the response $Y \sim f(y|\theta)$ to have a distribution given by a function in an exponential family. $Y$ could be Poisson, Binomial, Gamma, ... lots.

Insist that the distribution of $Y$ depend on $\beta$ only through the linear combination $\eta(\beta)$, that is $\theta = \theta(\eta(\beta))$.

But, let the mean response $E(Y)$ be some smooth monotone function of the linear predictor which we can specify.

This defines a large but still tractable and interpretable class of models called GLM's.

In this course we will look at a few of the simplest and most useful examples ($\dim(\theta) = 1$, mostly). However, much of what we will learn carries across to other GLM's.
Modeling with GLM’s
three decisions fix the model.

Distribution: \( Y_i \sim f(y_i|\theta_i) \), iid for \( i = 1, 2, ..., n \).
given the EV, how is the response distributed?

Linear predictor: \( \eta_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + ... + x_{i,p}\beta_p \)
what variables \( x_1, x_2, ..., x_p \) should we use?

Link function \( g(\mu_i) = \eta_i \)
(increasing, continuous, differentiable function)
How does mean response increase with LP?

\[
\eta_i \rightarrow \mu_i \rightarrow \theta_i \rightarrow \text{dbn of } Y_i.
\]
All these steps work because the key relations are (invertible) smooth monotone functions.

NLM \( y_i = x_i\beta + \epsilon_i, \epsilon_i \sim N(0, \sigma^2) \)
Stochastic, \( Y_i \sim N(\mu_i, \sigma^2) \) jointly independent
Deterministic, \( \eta_i = x_i^T\beta, \eta = X\beta \)
Link, choice \( g(\mu_i) = \mu_i \) gives NLM \( E(Y_i) = x_i\beta \).
Log-Likelihood (GLM)

\[ \ell(\beta; y) = \sum_{i=1}^{n} \frac{y_i \theta_i - \kappa(\theta_i)}{\phi} + c(y_i; \phi), \]

with

\[ \theta_i = \theta(x_i \beta) \]

since \( \mu_i = \kappa'(\theta_i) \) and \( g(\mu_i) = \eta_i \) both invertible

Canonical link function (important special case)

If we choose link function

\[ g(\mu_i) = \kappa'^{-1}(\mu_i) \]

then \( \theta_i = \eta_i \) since \( g(\mu_i) = \theta_i \).

Log-Likelihood (canonical link, \( \phi = 1 \) or known)

\[ \ell(\beta; y) = \sum_{i=1}^{n} \frac{y_i \eta_i - \kappa(\eta_i)}{\phi} + c(y, \phi), \quad \eta_i = x_i \beta. \]
Example GLM for a Poisson response

\[ Y_i \sim \text{Poisson}(\lambda_i), \text{ independent } E(Y_i) = \lambda_i. \]

\[
\exp(-\lambda) \frac{\lambda^y}{y!} = \exp(y \log(\lambda) - \lambda - \log(y!))
\]

\[ = \exp \left( \frac{y\theta - \kappa(\theta)}{\phi} + c(y; \phi) \right) \]

\[ \theta = \log(\lambda), \phi = 1, \kappa(\theta) = \exp(\theta). \]

Check \( E(Y) \):

\[ \mu = \kappa'(\theta) = \exp(\theta) = \lambda. \]

Check \( \text{var}(Y) \):

\[ \phi V(\mu) = \kappa''(\theta) = \lambda = \mu. \]

Link function \( g(\mu) = \eta \) where \( \eta = x\beta \):

Canonical link \( g(\mu) = \kappa'^{-1}(\mu) = \log(\mu) \)

Log-likelihood

\[
\ell(\beta; y) = \sum_{i=1}^{n} y_i x_i \beta - e^{x_i \beta}
\]
$y \sim \text{Poisson}(\mu) \quad \mu = \exp(1+2x)$
Example GLM for $Y_i \sim \text{Binomial}(m, p_i)$

$$f(y|\theta) = C_y^m p^y (1-p)^{m-y}$$

$$= \exp(y \log \left( \frac{p}{1-p} \right) + m \log(1-p) + \log(C_y^m))$$

$$\theta = \log \left( \frac{p}{1-p} \right) \text{ (log odds)}, \ \phi = 1$$

$$\kappa(\theta) = -m \log(1-p) = m \log(1 + \exp(\theta))$$

Check $E(Y) = mp$:

$$\kappa'(\theta) = m \exp(\theta)/(1 + \exp(\theta)).$$

Exercise Check variance $\phi V = mp(1-p)$.

Canonical link $g(\mu) = \kappa'^{-1}(\mu)$

$$g(\mu) = \log \left( \frac{\mu/m}{1 - \mu/m} \right)$$

Log-likelihood

$$\ell(\beta; y) = \sum_{i=1}^{n} y_i x_i \beta - m \log(1 + e^{x_i \beta})$$
GLM MLEs

The MLEs \( \beta = \hat{\beta} \) for the GLM satisfy score equations

\[
\frac{\partial \ell(\beta; y)}{\partial \beta_i} = 0, \quad i = 1, 2, \ldots, p.
\]

\( \beta \in \mathbb{R}^p, \ell(\beta; y) : \mathbb{R}^p \rightarrow \mathbb{R} \).

In terms of the gradient operator

\[
\frac{\partial \ell}{\partial \beta} = \left( \frac{\partial \ell}{\partial \beta_1}, \frac{\partial \ell}{\partial \beta_2}, \ldots, \frac{\partial \ell}{\partial \beta_p} \right)^T
\]

\[
\frac{\partial \ell}{\partial \beta} = 0 \quad \text{at} \quad \beta = \hat{\beta}.
\]

Hessian operator is

\[
\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} = \begin{pmatrix}
\frac{\partial^2 \ell}{\partial \beta_1^2} & \frac{\partial^2 \ell}{\partial \beta_1 \beta_2} & \cdots & \frac{\partial^2 \ell}{\partial \beta_1 \beta_p} \\
\frac{\partial^2 \ell}{\partial \beta_2 \beta_1} & \frac{\partial^2 \ell}{\partial \beta_2^2} & \cdots & \frac{\partial^2 \ell}{\partial \beta_2 \beta_p} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 \ell}{\partial \beta_p \beta_1} & \frac{\partial^2 \ell}{\partial \beta_p \beta_2} & \cdots & \frac{\partial^2 \ell}{\partial \beta_p^2}
\end{pmatrix}
\]
Observed information

\[ J(y) = -\partial^2 \ell / \partial \beta \partial \beta^T \]

Expected information

\[ I = -\mathbb{E}(\partial^2 \ell / \partial \beta \partial \beta^T) \]

\[ \hat{\beta} \xrightarrow{D} \mathcal{N}(\beta, I^{-1}) \quad \text{with } n. \]

Compute \( I \)

\[
\frac{\partial \ell}{\partial \beta_i} = \sum_{k=1}^{n} \frac{\partial \ell}{\partial \eta_k} \frac{\partial \eta_k}{\partial \beta_i} \\
= \frac{\partial \ell}{\partial \eta} \frac{\partial \eta}{\partial \beta_i} \\
\frac{\partial \ell}{\partial \beta^T} = \frac{\partial \ell}{\partial \eta} \frac{\partial \eta}{\partial \beta^T} \\
= \frac{\partial \ell}{\partial \eta^T} X \\
\frac{\partial^2 \ell}{\partial \beta \beta^T} = \frac{\partial \eta^T}{\partial \beta} \frac{\partial}{\partial \eta} \frac{\partial \ell}{\partial \beta^T} \\
= X^T \frac{\partial^2 \ell}{\partial \eta \partial \eta^T} X 
\]
Let
\[ W = -E \left( \frac{\partial^2 \ell}{\partial \eta \partial \eta^T} \right) \]
so that \( I = X^T W X \) and
\[ \hat{\beta} \overset{D}{\to} N(\beta, (X^T W X)^{-1}). \]

\( W \) is diagonal since \( \ell = \sum_k \ell_k \) with \( \ell_k = \ell_k(\eta_k) \).

\[
W_{kk} = -E \left( \frac{\partial^2 \ell}{\partial \eta_k^2} \right) = -E \left( \left[ \frac{d \theta_k}{d \eta_k} \right]^2 \frac{\partial^2 \ell}{\partial \theta_k^2} \right)
\]

\[
\eta_k = g(\kappa'(\theta_k))
\]

\[
1 = \frac{d \theta_k}{d \eta_k} g'(\mu_k) \kappa''(\theta_k)
\]

\[
\frac{d \theta_k}{d \eta_k} = \frac{1}{g'(\mu_k)V(\mu_k)}
\]

\[
\frac{\partial^2 \ell}{\partial \theta_k^2} = -\kappa''/\phi
\]

\[
W_{kk} = \frac{1}{g'(\mu_k)^2 \phi V(\mu_k)}
\]
Example: Poisson, canonical link

\[ \phi = 1, \, \kappa(\theta) = \exp(\theta), \, V(\mu) = \mu. \]

Link \( g(\mu) = \eta \), canonical link \( \log(\mu) = \eta \).

Log-likelihood

\[ \ell(\beta; y) = \sum_{i=1}^{n} y_i \eta_i - e^{\eta_i} \]

\[ W_{kk} = -E \left( \frac{\partial^2 \ell}{\partial \eta_k^2} \right) \]
\[ = e^{\eta_k} \]
\[ = \mu_k \]
\[ = \frac{1}{g'(\mu_k)^2 \phi V(\mu_k)} \]

\[ I = J \]
\[ = X^T \text{diag}(\mu_1, \ldots, \mu_n) X \]
Iteratively Re-weighted Least Squares

Newton-Raphson iteration \( z^* : f(z^*) = 0 \)

Given \( z : f(z) \approx 0 \), improve estimate via

\[
    f(z') \approx f(z) + \frac{df}{dz}(z' - z)
\]

to \( z' : f(z') \approx 0 \)

\[
    z' = z - \left(\frac{df}{dz}\right)^{-1}f(z)
\]

and iterate. For stationary point \( z^* : f'(z^*) = 0 \)

\[
    z' = z - \left(\frac{d^2f}{dz^2}\right)^{-1}f'(z), \quad i = 0, 1, 2, ...
\]

Multivariate case, expand log-likelihood at \( \beta \)

\[
    \frac{\partial \ell}{\partial \beta}\bigg|_{\beta=\beta'} \approx \frac{\partial \ell}{\partial \beta} + \frac{\partial^2 \ell}{\partial \beta \partial \beta^T} (\beta' - \beta).
\]

and solving for LHS=0,

\[
    \beta' = \beta - \left(\frac{\partial^2 \ell}{\partial \beta \partial \beta^T}\right)^{-1} \frac{\partial \ell}{\partial \beta}.
\]
Scoring

\[ \beta' = \beta + (X^T W X)^{-1} \frac{\partial \ell}{\partial \beta} \]

\[ \frac{\partial \ell}{\partial \beta} = X^T \frac{\partial \ell}{\partial \eta} \]

\[ = X^T \frac{\partial \theta^T}{\partial \eta} \frac{\partial \ell}{\partial \theta} \]

\[ \beta' = (X^T W X)^{-1} X^T W \left( X\beta + W^{-1} \frac{\partial \theta^T}{\partial \eta} \frac{\partial \ell}{\partial \theta} \right) \]

\[ \frac{\partial \ell}{\partial \theta} = \frac{(y - \mu)}{\phi} \]

\[ \frac{\partial \theta^T}{\partial \eta} = \text{diag} \left( \frac{1}{g'(\mu_1) V(\mu_1)} \ldots \frac{1}{g'(\mu_n) V(\mu_n)} \right) \]

\[ W^{-1} = \text{diag} \left( g'(\mu_1)^2 \phi V(\mu_1) \ldots g'(\mu_n)^2 \phi V(\mu_n) \right) \]

The IRLS iteration is

\[ z = X\beta + g'(\mu)(y - \mu) \]

\[ \beta' = (X^T W X)^{-1} X^T W z \]
Example: Poisson, canonical link

\[ \phi = 1, \kappa(\theta) = \exp(\theta), V(\mu) = \mu. \]
Link \( g(\mu) = \eta, \) canonical link \( \log(\mu) = \eta. \)

\[ W = \text{diag}(\mu_1, \ldots, \mu_n) \]

\[ z = X\beta + \frac{(y - \mu)}{\mu} \]

\[ \beta' = (X^T W X)^{-1} X^T W z \]

\[ \mu' = \exp(X\beta') \]

\[ z' = X\beta' + \frac{(y - \mu')}{\mu'} \]

\[ \cdot \]
\[ \cdot \]
\[ \cdot \]
$y \sim \text{Poisson}(\mu) \quad \mu = \exp(1+2x)$
y~Poisson(mu)    mu=exp(1+2x)
> #10 iterations of IRLS
> its<-10;
> for (i in 1:its) {
+   beta<-solve(t(X)%*%W%*%X)%*%t(X)%*%W%*%z;
+   eta<-X%*%beta;
+   mu<-exp(eta);
+   W<-diag(c(mu))
+   z<-(X%*%beta+(y-mu)/mu);
+ }

y~Poisson(mu)     mu=exp(1+2x)
> # MLE (if converged)
> beta
>   [,1]
> 1.131908
> 1.769498
> # inverse Fisher Information $I = X^T W X$
> Iinv <- solve(t(X) %*% W %*% X)
> # standard errors
> sqrt(diag(Iinv))
> 0.1654835 0.2277604
> # compare glm()
> summary(glm(y ~ 1 + x, family='poisson'))
> ...
> Coefficients:
>                       Estimate Std. Error z value Pr(>|z|)
> (Intercept)          1.1319     0.1655   6.840  7.91e-12
> x                    1.7695     0.2278   7.769  7.90e-15
> ...
> AIC: 150.09
> Number of Fisher Scoring iterations: 4
Iteratively Re-weighted Least Squares.

\[ \mu^{(0)} = y \text{ so } X\beta^{(0)} = \eta^{(0)} = g(\mu^{(0)}) = g(y) \]
\[ z^{(0)} = g(y) \text{ and } W^{(0)} = \text{diag}(g'(y)^2\phi V(y))^{-1}. \]

For \( i = 0, 1, 2, \ldots, \)

(a) set \( \beta^{(i+1)} = (X^T W^{(i)} X)^{-1} X^T W^{(i)} z^{(i)} \)
Regress \( z^{(i)} \) on \( X \) with weights \( \text{diag}(W) \).

(b) \( \eta^{(i+1)} = X \beta^{(i+1)} , \mu^{(i+1)} = g^{-1}(\eta^{(i+1)}) \),
\[ z^{(i+1)} = \eta^{(i+1)} + g'(\mu^{(i+1)})(y - \mu^{(i+1)}) \]
and
\[ W^{(i+1)} = \text{diag}\left(\frac{1}{g'(\mu^{(i+1)})^2\phi V(\mu^{(i+1)})}\right). \]