The arbitrary parametrization

Exponential families may be parametrized by the canonical parameter, by the mean value parameter or by a third parameter

In terms of an arbitrary parameter η the density for the exponential dispersion model will be

$$e^{\frac{\theta(\eta)y - c(\eta)}{\psi}},$$

where $c(\eta) = \kappa(\theta(\eta))$.

We can always find κ from c via $\kappa = c \circ \theta^{-1}$, but it can also be convenient to express computations of moments directly in terms of $\theta(\eta)$ and $c(\eta)$ as in Corollary 5.10.

The Weibull distribution

The Weibull distribution with shape parameter $k > 0$ and scale parameter η has density

$$\frac{kz^{k-1}}{\eta^k} e^{-\left(\frac{z}{\eta}\right)^k} = e^{-z^k \eta^{-k} - k\log(\eta)} kz^{k-1}$$

w.r.t. the Lebesgue measure on $(0, \infty)$.

For fixed k (a nuisance parameter) define the transformation $t(z) = z^k$. Then the distribution of $Y = t(Z) = Z^k$ is an exponential family with

$$\theta(\eta) = -\eta^{-k}, \quad c(\eta) = k \log \eta, \quad \nu = t(kz^{k-1} \cdot m_{(0,\infty)}) = m_{(0,\infty)}$$

$$\mu(\eta) = E(Y) = \frac{c'(\eta)}{\theta'(\eta)} = \eta^k, \quad V(Y) = \frac{\mu'(\eta)}{\theta'(\eta)} = \eta^{2k}$$

Generalized linear models

If X denotes a vector of predictors we define the linear predictor as

$$\eta = X^T \beta$$

for a parameter vector $\beta \in \mathbb{R}^p$.

In a generalized linear model we let the predictor enter into the distribution via the linear predictor η and a link function g such that

$$g(\mu) = g(E(Y \mid X)) = \eta = X^T \eta.$$

Or with $\mu = g^{-1}$

$$E(Y \mid X) = \mu(\eta) = \mu(X^T \beta).$$

The link function

The linear predictor is related to the mean via the choice of θ through the expression

$$\mu(\eta) = \frac{c'(\eta)}{\theta'(\eta)}.$$

This function is the inverse link function.

Definition

The link function, g, is the inverse of μ;

$$\eta = g(\mu).$$

Note the formula

$$\mathcal{V}(\mu(\eta)) = \frac{\mu'(\eta)}{\theta'(\eta)}$$

for the variance in terms of η.

The link function

$J = \kappa'(I)$ is the range of possible mean values.

$$
\begin{tikzcd}
J \arrow[rr, "g"] & & H \arrow[dll, "\theta"'] \\
I \arrow[u, "\kappa'"] & & \mathbb{R}^p \arrow[u, "X^T"']
\end{tikzcd}
\qquad
\begin{tikzcd}
\mu \arrow[rr, "g"] & & \eta \arrow[dll, "\theta"'] \\
\theta \arrow[u, "\kappa'"] & & \beta \arrow[u, "X^T"']
\end{tikzcd}
$$

The mean, μ, is related to the linear predictor, $X^T\beta$, via the link function, g:

$$g(\mu) = \eta = X^T\beta.$$

If $H \neq \mathbb{R}$ the parameter space $\mathbb{R}^p$ of β is effectively reduced to

$$\{\beta \mid X^T\beta \in H\}.$$

Link functions

Definition

The canonical link function is the inverse of the canonical mean value function κ';

$$g = (\kappa')^{-1}.$$

Note that the link function is canonical if and only if $\theta = \eta$, and for generalized linear models this means that the linear predictor equals the canonical parameter.

For the Weibull example the link function is $g(\mu) = \mu^{1/k}$.

$$\begin{tikzcd}
\eta^k \arrow[r, "g"] & \eta \arrow[dl, "\theta"'] \\
-\eta^{-k} \arrow[u, "\kappa'"] & \beta \arrow[u, "X^T"']
\end{tikzcd}$$

This is not the canonical link as θ is not the identity.

Deviance

Definition (Unit deviance)

With g the canonical link,

$$d(Y,\mu) = 2\left(\sup_{\mu_0 \in J}\left\{g(\mu_0)Y - \kappa(g(\mu_0))\right\} - g(\mu)Y + \kappa(g(\mu))\right)$$

for $Y \in \overline{J}$ and $\mu \in J$

For $Y \in J$,

$$d(Y,\mu) = 2\left(Y(g(Y) - g(\mu)) - \kappa(g(Y)) + \kappa(g(\mu))\right).$$

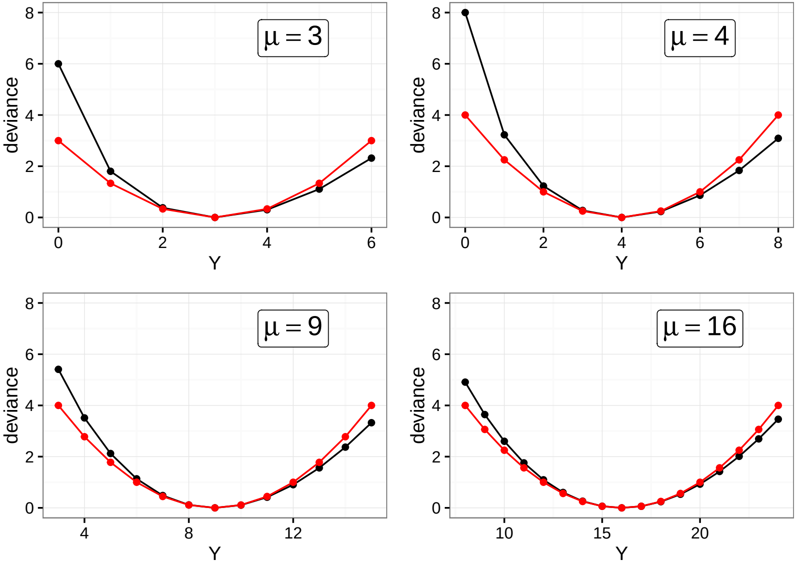
It holds that $d(Y,Y) = 0$ and $d(Y,\mu) > 0$ for $\mu \neq Y$.

Theorem

$$d(Y,\mu) = \frac{(Y-\mu)^2}{\mathcal{V}(\mu)} + o((Y-\mu)^2)$$

for $Y, \mu \in J$.

Poisson deviance (black) and quadratic approx.



Generalized linear model assumptions

GA1: The conditional expectation of Y given X is

$$E(Y \mid X) = \mu(\eta),$$

with $\mu : \mathbb{R} \to J$ the mean value function and $\eta = X^T\beta$ the linear predictor.

GA2: The conditional variance of Y given X is

$$V(Y \mid X) = \psi\mathcal{V}(\mu(\eta)),$$

with $\mathcal{V} : J \to (0, \infty)$ the variance function and $\psi > 0$ the *dispersion parameter*.

GA3: The conditional distribution of Y given X is the $(\theta(\eta), \nu_\psi)$-exponential dispersion distribution,

$$Y \mid X \sim \mathcal{E}(\theta(\eta), \nu_\psi).$$

The log-likelihood and the score statistic

The density for the exponential dispersion model is

$$e^{\frac{\theta(\eta)y - c(\eta)}{\psi}}$$

and the log-likelihood function is

$$\ell(\eta) = \frac{\theta(\eta)y - c(\eta)}{\psi}.$$

The score function is

$$\ell'(\eta) = \frac{\theta'(\eta)y - c'(\eta)}{\psi} = \frac{\theta'(\eta)\Big(y - \mu(\eta)\Big)}{\psi}.$$

The score statistic is

$$U(\eta) = \frac{\theta'(\eta)\Big(Y - \mu(\eta)\Big)}{\psi}$$

and

$$EU(\eta) = 0.$$

The Fisher information

Definition

The Fisher information,

$$\mathcal{J}(\eta) = -E(U'(\eta)),$$

is the expectation of the derivative of minus the score statistic.

Observe that

$$\psi U'(\eta) = \theta''(\eta)\Big(Y - \mu(\eta)\Big) - \theta'(\eta)\mu'(\eta).$$

Hence

$$\mathcal{J}(\eta) = \frac{\theta'(\eta)\mu'(\eta)}{\psi}.$$

Note that it holds that

$$\mathcal{J}(\eta) = V(U(\eta)).$$

Replications

With independent observations $Y_1, \ldots, Y_n$ (conditionally on $\mathbf{X}$) the log-likelihood is

$$\ell(\boldsymbol{\eta}) = \sum_{i=1}^{n} \ell(\eta_i) = \frac{1}{\psi} \sum_{i=1}^{n} \theta(\eta_i) Y_i - c(\eta_i).$$

The score statistic is

$$U(\boldsymbol{\eta}) = \nabla_{\boldsymbol{\eta}} \ell(\boldsymbol{\eta}) = (U_1(\eta_1), \ldots, U_n(\eta_n))^T$$

and the Fisher information is

$$\mathcal{J}(\boldsymbol{\eta}) = \left(\begin{array}{ccc} \mathcal{J}_1(\eta_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \mathcal{J}_n(\eta_n) \end{array} \right).$$

Generalized linear models

With $\mathbf{X}$ an $n \times p$ matrix and

$$\boldsymbol{\eta} = \mathbf{X}\beta$$

the linear predictor, we have the score statistic

$$U(\beta) = \nabla_\beta \ell(\beta) = \mathbf{X}^T U(\boldsymbol{\eta})$$

and Fisher information

$$\mathcal{J}(\beta) = \mathbf{X}^T \underbrace{\mathcal{J}(\boldsymbol{\eta})}_{\mathbf{W}} \mathbf{X}.$$

The entries in the diagonal weight matrix $\mathbf{W}$ are

$$w_{ii} = \frac{\mu'(\eta_i)\theta'(\eta_i)}{\psi} = \frac{\mu'(\eta_i)^2}{\psi \mathcal{V}(\mu(\eta_i))}.$$

Modeling choices

Choices that are up to us:

- The dispersion model family / the variance function / the unit structure measure ν. Either choice determines the structure of the model.
- The link function that relates the linear predictor to the mean of Y. The choice implicitly chooses the map θ that relates the linear predictor to the canonical parameter.
- Any transformation of data prior to the modeling.

Nuisance parameters may appear in transformations, in the link function or in ν / the variance function.