The Weibull distribution

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

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

w.r.t. the Lebesgue measure.

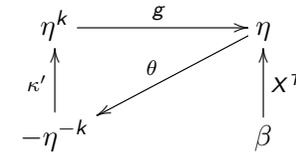For fixed k (a nuisance parameter) the transformation $Y = Z^k$ gives an exponential family with

$$\theta(\eta) = -\eta^{-k}, \quad c(\eta) = k \log \eta, \quad \frac{\mathrm{d}\nu}{\mathrm{d}m} = kz^{k-1}$$

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

The Weibull distribution

The link function is the inverse of μ; $g(\mu) = \mu^{1/k}$.

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

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

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

The score statistic is

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

and

$$E_\eta U(\eta) = 0.$$

The Fisher information

Definition

The Fisher information,

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

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

Observe that

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

Hence

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

Note that it holds that

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

Replications

With independent $Y_1, \ldots, Y_n$ observations 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

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

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 measure ν – the structure of the model, e.g. discrete or continuous data.
- Any transformation of data prior to the modeling.
- The function θ that relates the linear predictor, the canonical parameter and the mean of Y. The choice is mostly made implicitly by choosing the link function g.

We cannot choose c (or ϕ or κ) – it is given by the other choices.

Nuisance parameters may appear in prior transformations or in the map θ as well as in c or ν.

The Weibull distribution

With i.i.d. Weibull distributed observations $Z_1, \ldots, Z_n$ then $Y_1 = Z_1^k, \ldots, Y_n = Z_n^k$ are i.i.d. from the exponential family.

This is a generalized linear model with $\mathbf{X} = \mathbf{1}$ the $n \times 1$ matrix of ones, and

$$\boldsymbol{\eta} = \mathbf{1}\eta.$$

The score equation is

$$0 = U(\eta) = \sum_{i=1}^{n} \theta'(\eta) y_i - c'(\eta) = \frac{k}{\eta^{k+1}} \sum_{i=1}^{n} y_i - \frac{nk}{\eta}.$$

The Newton algorithm

We replace the (nonlinear) score equation by the linearization around η_1

$$U(\eta_1) + U'(\eta_1)(\eta - \eta_1) = 0 \qquad (1)$$

relying on $U(\eta) \simeq U(\eta_1) + U'(\eta_1)(\eta - \eta_1)$.

The solution to (1) is

$$\eta = \eta_1 - \frac{U(\eta_1)}{U'(\eta_1)}.$$

The Newton algorithm is obtained by iteration;

$$\eta_{m+1} = \eta_m - \frac{U(\eta_m)}{U'(\eta_m)}$$

with η_1 the initial value or start guess.

Fisher scoring

Replacing $-U'(\eta)$ by its expectation $J(\eta)$ we get the Fisher scoring algorithm

$$\eta_{m+1} = \eta_m + \frac{U(\eta_m)}{J(\eta_m)}.$$

For the Weibull example the Fisher information is

$$J(\eta) = n\theta'(\eta)^2 V_\eta(Y) = \frac{nk^2}{\eta^2}.$$

Generalized linear models

The p-dimensional score equation for a GLM reads

$$U(\beta) = \mathbf{X}^T U(\boldsymbol{\eta}) = 0.$$

Given β_1 and corresponding $\eta_{1,i} = X_i^T \beta_1$ let

$$\mathbf{W}_1^{\text{obs}} = - \begin{pmatrix} U_1'(\eta_{1,1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_n'(\eta_{1,n}) \end{pmatrix}.$$

The linearization of the score equation is

$$\mathbf{X}^T U(\boldsymbol{\eta}_1)^T - \mathbf{X}^T \mathbf{W}_1^{\text{obs}} \mathbf{X}(\beta - \beta_1) = 0,$$

whose solution is

$$\beta = \beta_1 + (\mathbf{X}^T \mathbf{W}_1^{\text{obs}} \mathbf{X})^{-1} \mathbf{X}^T U(\boldsymbol{\eta}_1).$$

Fisher scoring

Replacing $\mathbf{W}_m^{\text{obs}}$ by its expectation

$$\mathbf{W}_m = \frac{1}{\psi} \begin{pmatrix} \mu'(\eta_{m,1})\theta'(\eta_{m,1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu'(\eta_{m,n})\theta'(\eta_{m,n}) \end{pmatrix}$$

we get by iteration the Fisher scoring algorithm.

$$\begin{aligned} \beta_{m+1} &= \beta_m + (\mathbf{X}^T \mathbf{W}_m \mathbf{X})^{-1} \mathbf{X}^T U(\boldsymbol{\eta}_m) \\ &= (\mathbf{X}^T \mathbf{W}_m \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_m \underbrace{\left(\mathbf{X}\beta_m + \mathbf{W}_m^{-1} U(\boldsymbol{\eta}_m) \right)}_{\mathbf{z}_m}. \end{aligned}$$

Iterative Weighted Least Squares

The vector $\mathbf{Z}_m$ is known as the working response and

$$Z_{m,i} = X_i^T \beta_m + \frac{Y_i - \mu(\eta_{m,i})}{\mu'(\eta_{m,i})} \tag{2}$$

In terms of the working response, the vector β_m is the minimizer of the weighted sum of squares

$$(\mathbf{Z}_m - \mathbf{X}\beta)^T \mathbf{W}_m (\mathbf{Z}_m - \mathbf{X}\beta). \tag{3}$$

This is a standard weighted least squares problem.

Iterative Weighted Least Squares

The Fisher scoring algorithm for GLMs is known as IWLS due to the iterative solution of a weighted least squares problem: Given β_1 we iterate until convergence

- Compute the working response vector $\mathbf{Z}_m$ based on β_m using (2).
- Compute the weights

$$w_{m,ii} = \frac{(\mu'_{m,i})^2}{\mathcal{V}(\mu_{m,i})} = \frac{\mu'(\eta_{m,i})^2}{\mathcal{V}(\mu(\eta_{m,i}))}.$$

- Minimize the weighted sum of squares (3).

Computations rely only on the mean value function μ, its derivative μ' and the variance function $\mathcal{V}$.

Score function with canonical link

With the canonical link function $\theta'(\eta) = 1$ and the score function is

$$U(\beta) = \sum_{i=1}^n (Y_i - \mu(\eta_i))X_i = t - \tau(\beta)$$

whose second derivative equals the Fisher information

$$D_\beta U(\beta) = \mathcal{J}(\beta) = \mathbf{X}^T \mathbf{W} \mathbf{X}$$

with $\mathbf{W}_{ii} = \mu'(\eta_i) = \mathcal{V}(\mu(\eta_i))$. We assume throughout that $\mu'(\eta) > 0$ for all η.

Define

$$\tau(\beta) = \sum_{i=1}^n \mu(\eta_i)X_i \quad \text{and} \quad t = \sum_{i=1}^n Y_i X_i$$

The score equation

The score equation is

$$\tau(\beta) = t. \tag{4}$$

Theorem

If $\mathbf{X}$ has full rank p the map $\tau : \mathbb{R}^p \to \mathbb{R}^p$ is one-to-one. With $C := \tau(\mathbb{R}^p)$ there is a unique solution to (4) if and only if $t \in C$.

Lemma

If $t_0 = \sum_{i=1}^n \mu_i X_i$ with $\mu_i \in J := \mu(\mathbb{R})$ there is a solution to the equation $\tau(\beta) = t_0$.

On this slide it is assumed that $H = \mathbb{R}$.