Lecture 1. Exponential families, deviances.

Course Designer: Jesper Rydén, SLU, Uppsala, Sweden

Course Lecturer: Raazesh Sainudiin, UU, Uppsala, Sweden

Generalized Linear Models 1MS369 • Autumn 2018

GLM and the variance function

Consider a response variable Y_i with mean μ_i. Its variance is assumed to be proportional to the **variance function**

$$\mathsf{V}[Y_i] = \sigma^2 V(\mu_i), \quad i = 1, 2, \ldots, n$$

where $\sigma^2 > 0$ is the **dispersion parameter**, assumed to be common for all units.

The variance function describes how the variance of the response Y_i varies as a function of the mean μ_i.

GLM and the link function

The role of the **link function** g is to transform the mean μ_i onto a scale where the model is linear:

$$g(\mu_i) = \beta_1 x_{i1} + \cdots + \beta_k x_{ik}, \quad i = 1, 2, \ldots, n$$

where $\beta_1, \ldots, \beta_k$ are unknown regression coefficients, to be estimated.

Some remarks

NOTE 1: For the traditional model of multiple linear regression,

$$V(\mu) = 1, \quad g(\mu) = \mu.$$

NOTE 2: Often the variance assumption is generalised as follows:

$$\mathrm{V}[Y_i] = \frac{\sigma^2}{w_i} V(\mu_i), \quad i = 1, 2, \ldots, n$$

where $w_1, \ldots, w_n$ are known weights.

For instance, weights may be sample sizes if Y_i are group averages.

On choices of link and variance functions

Common choices for certain situations (motivated later in the course)

- **Data on the real line**, μ a location parameter with unbounded domain to the left and right; identity link $g(\mu) = \mu$ and constant variance function $V(\mu) = 1$.
- **Strictly positive data**, with $\mu > 0$; log link $g(\mu) = \ln \mu$ together with square variance function $V(\mu) = \mu^2$, or alternatively power-link functions $g(\mu) = \mu^q$ and power-variance functions $V(\mu) = \mu^p$.
- **Non-negative data**, in particular counts; log link $g(\mu) = \ln \mu$ and linear variance function $V(\mu) = \mu$.
- **Proportions**, satisfying $0 \leq Y \leq 1$, where $0 < \mu < 1$; common choice is logit link $g(\mu) = \ln(\mu/(1-\mu))$ with variance function $V(\mu) = \mu(1-\mu)$.

Exponential dispersion model (ED)

DEFINITION. A family of probability densities which can be written on the form

$$f_Y(y;\theta) = c(y,\lambda)\exp(\lambda\{\theta y - \kappa(\theta)\})$$

is called an **exponential dispersion family** of distributions.

- The parameter $\lambda > 0$ is called the **precision parameter** or **index parameter**.
- The parameter θ is the **canonical parameter**.
- The function $\kappa(\theta)$ is called the **cumulant generator**.

Motivation: Exponential dispersion family

Idea: Separate

mean-value related distributional properties described by the cumulant generator $\kappa(\theta)$

from

dispersion features as sample size, common variance or common over-dispersion not related to the mean value.

Natural exponential family

DEFINITION. Natural exponential family of distributions:

$$f_Y(y;\theta) = c(y)\exp(\theta y - \kappa(\theta)), \quad \theta \in \Omega$$

Here:

θ	Canonical parameter
$\kappa(\theta)$	Cumulant generator
Ω	Parameter space: a subset of the real line

REMARK. When λ is known, the natural exponential family arises from the ED family as a special case. Alternatively, view the ED family as an indexed set of natural families, indexed by λ.

Note: other forms of “exponential families”

Kendall’s Advanced Theory of Statistics, 5th ed (1987):

$$f(y;\theta) = \exp\big[A(\theta)B(y) + C(y) + D(\theta)\big]$$

The exponential family of distributions.

The subclass of the family with $B(y) = y$

$$f(y;\theta) = \exp\big[yA(\theta) + C(y) + D(\theta)\big]$$

is called the natural exponential family.

Other forms, ctd.

Nelder and Wedderburn (1972):

$$f(y;\theta,\phi) = \exp\big[\alpha(\phi)(y\theta - g(\theta) + h(y)) + \beta(\phi, y)\big]$$

With $\alpha(\phi) > 0$, for fixed ϕ we have an exponential family. Here, ϕ is called a nuisance parameter.

Dobson and Barnett (2008):

$$f(y;\theta) = s(y)t(\theta)\exp\big[a(y)b(\theta)\big]$$

rewritten as

$$f(y;\theta) = \exp\big[a(y)b(\theta) + c(\theta) + d(y)\big]$$

If $a(y) = y$, the distribution is said to be in canonical form.

Common distributions, found in the ED family

We study closer Poisson distribution, normal distribution and binomial and Bernoulli distributions, identifying

- canonical parameters,
- cumulant generators,
- precision parameters.

Blackboard

Two results from likelihood theory

Consider the score function

$$S(\theta; Y) = \frac{\partial}{\partial \theta} \ell(\theta; Y).$$

ML result I.

$$\mathrm{E}[S(\theta; Y)] = 0$$

ML result II.

$$\mathrm{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = -\mathrm{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right]$$

Proof: See e.g. Liero and Zwanzig (2012), Section 3.2.

Two key results for GLM

GLM result I.

$$\mathsf{E}[Y] = \kappa'(\theta) = \mu_i.$$

GLM result II.

$$\mathsf{V}[Y] = \frac{1}{\lambda}\kappa''(\theta) = \frac{1}{\lambda}V(\mu_i).$$

Blackboard, proofs

Results for standard distributions

By simple derivation of $\kappa(\theta)$, we may find mean values and variances for the previously considered distributions.

For instance, mean values:

Distribution	$\kappa(\theta)$	$\kappa'(\theta) = \mathsf{E}[Y]$
Poisson, $Y \sim \mathsf{Po}(\mu)$	$\mathrm{e}^{\theta} = \mathrm{e}^{\ln \mu}$	μ
Normal, $Y \sim \mathsf{N}(\mu, \sigma^2)$	$\theta^2/2 = \mu^2/2$	μ

Compare with well-known results.

More on the derivative of the cumulant generator: the mean-value mapping

We have, GLM result I, that

$$\mu = \kappa'(\theta).$$

The function

$$\tau(\theta) = \kappa'(\theta)$$

gives a relation between the canonical parameter θ and the expectation parameter μ. It is called the **mean-value mapping**.

It defines a one-to-one mapping $\mu = \tau(\theta)$ of the parameter space Ω for the canonical parameter θ on a subset $\mathcal{M}$ of the real line, called the *mean-value space*.

Note. For some distributions, the mean-value space is a subset of the real line, for instance, for the binomial distribution: $0 < \mu < 1$.

The (unit) variance function

We have GLM result II:

$$\mathsf{V}[Y] = \frac{1}{\lambda}\kappa''(\theta).$$

By introducing the variance function $V(\mu) = \kappa''(\theta)$ and using $\theta = \tau^{-1}(\mu)$, we get for the natural exponential family

$$V(\mu) = \kappa''(\tau^{-1}(\mu))$$

Distinguish between the *variance operator* V[.] and the *variance function* (the variance as a function of the mean).

Unit deviance function

DEFINITION. For a given variance function V, we define the **unit deviance function** by

$$d(y;\mu) = 2\int_{\mu}^{y} \frac{y-z}{V(z)}\,\mathrm{d}z$$

which is strictly positive except for $y=\mu$ (where it is zero).

The unit deviance may be interpreted as a measure of squared distance between y and μ.

In particular, for the normal distribution, we find $d(y;\mu) = (y-\mu)^2$.

Blackboard

Mean-value space, unit variance and unit deviance

Table for some common distributions:

Family	$\mathcal{M}$	$V(\mu)$	$d(y;\mu)$
Normal	$\mathbb{R}$	1	$(y-\mu)^2$
Poisson	$(0,\infty)$	μ	$2\left[y\ln(y/\mu)-(y-\mu)\right]$
Binomial (proportion)	$(0,1)$	$\mu(1-\mu)$	$2\left[y\ln(y/\mu)+(1-y)\ln((1-y)/(1-\mu))\right]$

ED families and convolution

Assume that $Y_1, \ldots, Y_n$ are independent and

$$Y_i \sim \mathrm{ED}(\mu, \sigma^2/w_i), \quad i = 1, \ldots, n$$

for given numbers $w_1, \ldots, w_n$. Let

$$w_\bullet = w_1 + \cdots + w_n.$$

It can then be shown that

$$\sum_{i=1}^{n} w_i Y_i / w_\bullet \sim \mathrm{ED}(\mu, \sigma^2/w_\bullet).$$

Special case: $Y_1, \ldots, Y_n$ iid $\mathrm{ED}(\mu, \sigma^2)$:

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \sim \mathrm{ED}(\mu, \sigma^2/n).$$

Jørgensen (1992). *The theory of exponential dispersion models and analysis of deviance.*

Example: Energy expenditure data

The energy expenditure for human subjects at rest for a 24-hour period is investigated (response variable y). Regressors are x_1: mass of tissue and x_2: mass of fat-free tissue.

Suppose we divide the tissues of subject i into k compartments (homogeneous with respect to energy expenditure).

Let $w_{i1}, \ldots, w_{ik}$ denote the masses of the k compartments.
Let $Y_{i1}, \ldots, Y_{ik}$ be the corresponding energy expenditures.

Total body mass of subject i:

$$w_i = w_{i1} + \cdots + w_{ik}$$

Total energy expenditure of subject i:

$$Y_i = Y_{i1} + \cdots + Y_{ik}$$

$\boxed{\text{Blackboard: further modelling}}$

Exponential families and statistical models

Assume independent observations $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^\mathsf{T}$ from an ED family with cumulant generator $\kappa(.)$ and where the precision parameter is a known weight, $\lambda_i = w_i$.

Joint density: With respect to the canonical parameter, we find

$$f(\mathbf{y}; \boldsymbol{\theta}) = \exp\left[\sum_{i=1}^{n} w_i(\theta_i y_i - \kappa(\theta_i))\right] \prod_{i=1}^{n} c(y_i, w_i)$$

Formulating in terms of mean-value parameter $\mu = \tau(\theta)$, and unit deviance, we find

$$\begin{aligned} g(\mathbf{y}; \boldsymbol{\mu}) &= \prod_{i=1}^{n} g_y(y_i; \mu_i, w_i) \\ &= \exp\left[-\frac{1}{2}\sum_{i=1}^{n} w_i d(y_i; \mu_i)\right] \prod_{i=1}^{n} c(y_i, w_i) \end{aligned}$$

Blackboard: loglikelihoods, score functions