Regression and Simulation Methods

Week 6: Generalised Linear Models

Clara Panchaud, 17th November 2023

What has been covered over the last week?

- The Poisson distribution
- Logistic regression
- Exponential family
- GLMs

Exponential Family

Definition

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

a(y) - Canonical link function
b(theta) - natural parameter

Expectation and Variance

Proposition 6.2. For a random variable Y in exponential family form (6.3),

(i) $\mathbb{E}\big(a(Y)\big) = \frac{-c'(\theta)}{b'(\theta)}$

(ii) $\operatorname{var}\big(a(Y)\big) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}$

Example

The Gaussian distribution

The (univariate) Gaussian density can be written as follows (where the underlying measure is Lebesgue measure):

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \tag{8.19}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}\mu^2 - \log\sigma\right\}. \tag{8.20}$$

This is in the exponential family form, with:

b(theta)

$$\eta = \begin{bmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{bmatrix} \tag{8.21}$$

a(y)

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \tag{8.22}$$

c(theta)

$$A(\eta) = \frac{\mu^2}{2\sigma^2} + \log\sigma = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \tag{8.23}$$

NOT d(y)

$$h(x) = \frac{1}{\sqrt{2\pi}}. \tag{8.24}$$

GLMs

Definition

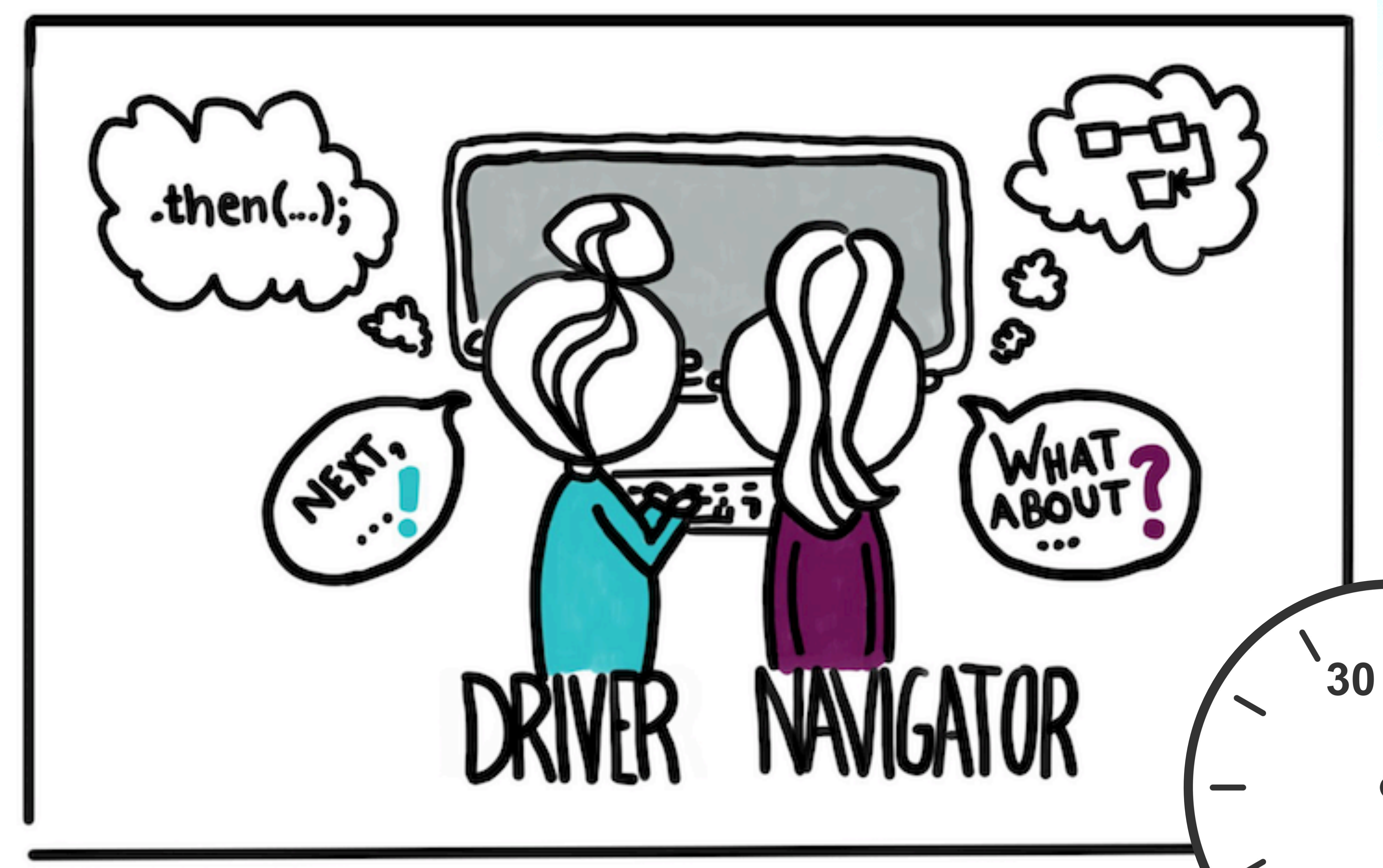
Definition 2: Generalised Linear Model

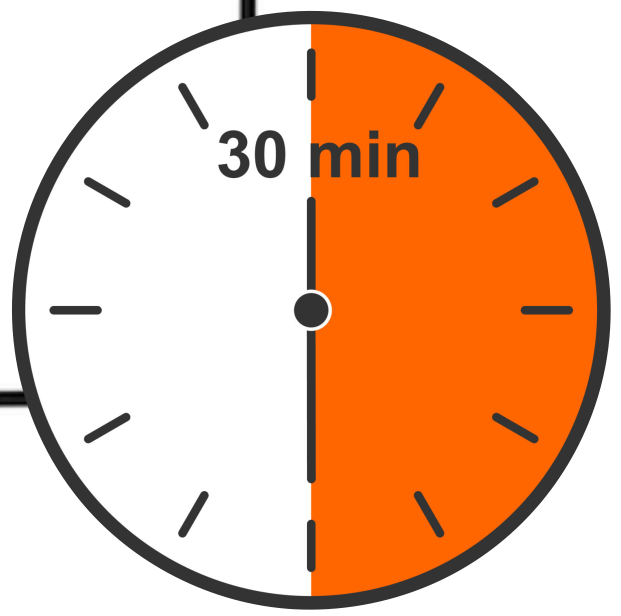
A generalised linear model (GLM) consists of the following three components:

1. A (common) distribution from the exponential family for the independent response variables $Y_1, ..., Y_n$.
2. A linear predictor
$$\eta_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}.$$
3. A differentiable, strictly monotone link function such that,
$$g(\mu_i) = \eta_i,$$
where $\mu_i = \mathbb{E}[Y_i]$

Rest of the tutorial...

- In pairs work on the sixth notebook. Paired programming will continue!





Rest of the week...

Once you have finished the R-Script go ahead and answer 6-1 through 6-4 from the SMSTC resources.

Additional Questions

1. The gamma pdf of a random variable Y is given by:

$$f(y) = (s^a \Gamma(a))^{-1} y^{a-1} e^{-y/s}$$

where $y \geq 0$, s is the *scale* parameter, a the *shape* parameter, and Γ the gamma function. (If a is integer then $\Gamma(a) = (a-1)!$)

   (a) Reparameterise this pdf by setting $a = 1/\phi$ and $s = \mu\phi$, and hence show that it is a member of the exponential family.

   (b) Deduce that the canonical link for the gamma is $\theta = \frac{1}{\mu} = \eta = \mathbf{X}\beta$.

   (c) Using the result above, compute $b''(\theta)$.

   (d) Compute $\mathrm{V}[Y]$ (hint: you can use Proposition 6.2).

   (e) How does this last result help if you want to assess whether data are best modelled as a normal, overdispersed Poisson or gamma distribution?