Exercises chapter 3

3.6 Exercises

3.1 The following relationships can be described by generalized linear models. For each one, identify the response variable and the explanatory variables, select a probability distribution for the response (justifying your choice) and write down the linear component.

(a) The effect of age, sex, height, mean daily food intake and mean daily energy expenditure on a person's weight.

(b) The proportions of laboratory mice that became infected after exposure to bacteria when five different exposure levels are used and 20 mice are exposed at each level.

(c) The relationship between the number of trips per week to the supermarket for a household and the number of people in the household, the household income and the distance to the supermarket.

3.2 If the random variable Y has the **Gamma distribution** with a scale parameter β, which is the parameter of interest, and a known shape parameter α, then its probability density function is

$$f(y;\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-y\beta}.$$

Show that this distribution belongs to the exponential family and find the natural parameter. Also using results in this chapter, find $\mathrm{E}(Y)$ and $\mathrm{var}(Y)$.

3.3 Show that the following probability density functions belong to the exponential family:

(a) Pareto distribution $f(y;\theta) = \theta y^{-\theta-1}$.
(b) Exponential distribution $f(y;\theta) = \theta e^{-y\theta}$.
(c) Negative Binomial distribution

$$f(y;\theta) = \binom{y+r-1}{r-1} \theta^r (1-\theta)^y ,$$

where r is known.

3.4 Use results (3.9) and (3.12) to verify the following results:

(a) For $Y \sim \mathrm{Po}(\theta)$, $\mathrm{E}(Y) = \mathrm{var}(Y) = \theta$.
(b) For $Y \sim \mathrm{N}(\mu, \sigma^2)$, $\mathrm{E}(Y) = \mu$ and $\mathrm{var}(Y) = \sigma^2$.
(c) For $Y \sim \mathrm{Bin}(n, \pi)$, $\mathrm{E}(Y) = n\pi$ and $\mathrm{var}(Y) = n\pi(1-\pi)$.

3.5 (a) For a Negative Binomial distribution $Y \sim \mathrm{NBin}(r, \theta)$, find $\mathrm{E}(Y)$ and $\mathrm{var}(Y)$.
(b) Notice that for the Poisson distribution $\mathrm{E}(Y) = \mathrm{var}(Y)$, for the Binomial distribution $\mathrm{E}(Y) > \mathrm{var}(Y)$ and for the Negative Binomial distribution $\mathrm{E}(Y) < \mathrm{var}(Y)$. How might these results affect your choice of a model?

3.6 Do you consider the model suggested in Example 3.5.3 to be adequate for the data shown in Figure 3.2? Justify your answer. Use simple linear regression (with suitable transformations of the variables) to obtain a model for the change of death rates with age. How well does the model fit the data? (Hint: Compare observed and expected numbers of deaths in each groups.)

3.7 Consider N independent binary random variables $Y_1, \ldots, Y_N$ with

$$P(Y_i = 1) = \pi_i \text{ and } P(Y_i = 0) = 1 - \pi_i.$$

The probability function of Y_i, the Bernoulli distribution $\mathrm{B}(\pi)$, can be written as

$$\pi_i^{y_i} (1-\pi_i)^{1-y_i} ,$$

where $y_i = 0$ or 1.

(a) Show that this probability function belongs to the exponential family of distributions.

(b) Show that the natural parameter is

$$\log\left(\frac{\pi_i}{1-\pi_i}\right).$$

This function, the logarithm of the **odds** $\pi_i/(1-\pi_i)$, is called the **logit** function.

(c) Show that $\mathrm{E}(Y_i) = \pi_i$.

(d) If the link function is

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \mathbf{x}^T\boldsymbol{\beta},$$

show that this is equivalent to modelling the probability π as

$$\pi = \frac{e^{\mathbf{x}^T\boldsymbol{\beta}}}{1+e^{\mathbf{x}^T\boldsymbol{\beta}}}.$$

(e) In the particular case where $\mathbf{x}^T\boldsymbol{\beta} = \beta_1 + \beta_2 x$, this gives

$$\pi = \frac{e^{\beta_1+\beta_2 x}}{1+e^{\beta_1+\beta_2 x}},$$

which is the **logistic function**.

Sketch the graph of π against x in this case, taking β_1 and β_2 as constants. How would you interpret this graph if x is the dose of an insecticide and π is the probability of an insect dying?

3.8 Is the **extreme value** (**Gumbel**) **distribution**, with probability density function

$$f(y;\theta) = \frac{1}{\phi}\exp\left\{\frac{(y-\theta)}{\phi} - \exp\left[\frac{(y-\theta)}{\phi}\right]\right\}$$

(where $\phi > 0$ is regarded as a nuisance parameter) a member of the exponential family?

3.9 Suppose $Y_1, \ldots, Y_N$ are independent random variables each with the Pareto distribution and

$$\mathrm{E}(Y_i) = (\beta_0 + \beta_1 x_i)^2.$$

Is this a generalized linear model? Give reasons for your answer.

3.10 Let $Y_1, \ldots, Y_N$ be independent random variables with

$$\mathrm{E}(Y_i) = \mu_i = \beta_0 + \log(\beta_1 + \beta_2 x_i); \quad Y_i \sim \mathrm{N}(\mu, \sigma^2)$$

for all $i = 1, \ldots, N$. Is this a generalized linear model? Give reasons for your answer.

3.11 For the Pareto distribution, find the score statistics U and the information $\mathfrak{I} = \mathrm{var}(U)$. Verify that $\mathrm{E}(U) = 0$.

3.12 Some more relationships between distributions—see Figure 3.3.

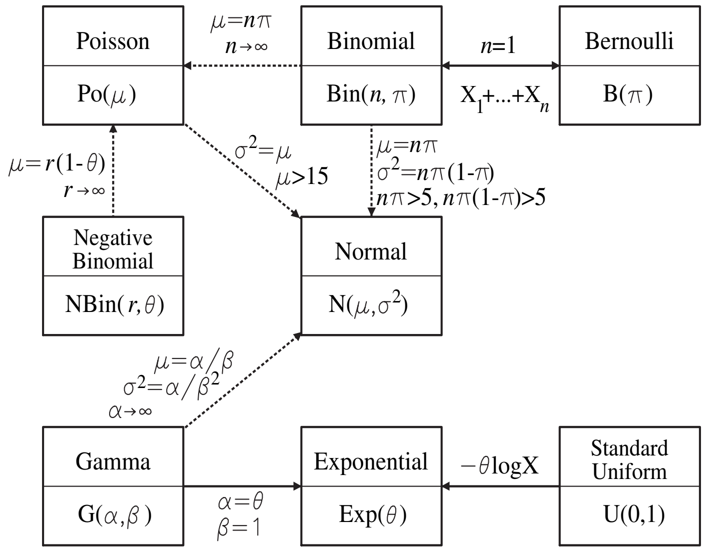

Figure 3.3 *Some relationships among distributions in the exponential family. Dotted lines indicate an asymptotic relationship and solid lines a transformation.*

(a) Show that the Exponential distribution $\text{Exp}(\theta)$ is a special case of the Gamma distribution $\text{G}(\alpha, \beta)$.

(b) If X has the Uniform distribution $\text{U}[0,1]$, that is, $f(x) = 1$ for $0 < x < 1$, show that $Y = -\theta \log X$ has the distribution $\text{Exp}(\theta)$.

(c) Use the moment generating functions (or other methods) to show

  i. $\text{Bin}(n, \pi) \to \text{Po}(\lambda)$ as $n \to \infty$.
  ii. $\text{NBin}(r, \theta) \to \text{Po}(r(1-\theta))$ as $r \to \infty$.

(d) Use the Central Limit Theorem to show

  i. $\text{Po}(\lambda) \to \text{N}(\mu, \mu)$ for large μ.
  ii. $\text{Bin}(n, \pi) \to \text{N}(n\pi, n\pi(1-\pi))$ for large n, provided neither $n\pi$ nor $n\pi(1-\pi)$ is too small.
  iii. $\text{G}(\alpha, \beta) \to \text{N}(\alpha/\beta, \alpha/\beta^2)$ for large α.