Exercises

Ryan Honea

9/8/2017

Exercise One

Question

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.

Solutions

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

Solution: The response is the weight, while all other variables are the explanatory. I would assume that the weight is likely distributed approximately normally with a slight right skew. It's far more likely for someone to be perhaps fifty pounds overweight than 50 pounds underwight. For example, if the average weight is 165 lb for some class, one could reasonably weigh 300 pounds but unreasonably weigh 0. In creating the linear component, I will assume general normality on the response. The linear component then would be

$$E[Y_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}; Y_i \sim \mathrm{N}(\mu_i, \sigma^2)$$

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

Solution: The levels of exposure are the explanatory variables, while the number infected is the response variable. Being infected or not infected are the possible outcomes so the response is likely distributed binomial. There are two ways the linear component could be created. One could have multiple x_{ij} acting as indicator variables for each level of the exposure, or the x_{ij} could be coded at different levels of exposure (i.e $x = \{1, 2, 3, 4, \text{ or } 5\}$). So,

$$E[Y_i] = \beta_0 + \beta_1 x_{i1}; Y_i \sim \mathrm{Bin}(20, p_i)$$

$$E[Y_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}; Y_i \sim \mathrm{Bin}(20, p_i)$$

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

Solution: The response here is the number of trips per week to the super-market while the others are explanatory variables. Because these are counts of trips, they are likely distributed Poisson.

$$E[Y_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}; Y_i \sim \mathrm{Po}(\lambda_i)$$

Exercise Three

Question

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

So, show that the following probability density functions can be written as

$$\exp[a(y)b(\theta) + c(\theta) + d(y)]$$

Solutions

(a): Pareto distribution $f(y_i; \theta) = \theta y^{-\theta-1} \mathbb{1}\{y \geq 1\}$

Solution: Note that $e^{\log x} = x$. We will do the same to $f(y_i; \theta)$.

$$
\begin{aligned}
f(y; \theta) &= \exp[\log(\theta y^{-\theta-1} \mathbb{1}\{y \geq 1\})] \\
&= \exp[\log \theta - \theta \log y - \log y + \log(\mathbb{1}\{y \geq 1\})]
\end{aligned}
$$

$$a(y) = -\log y \quad b(\theta) = \theta \quad c(\theta) = \log \theta \quad d(y) = -\log y + \log(\mathbb{1}\{y \geq 1\})$$

(b): Exponential distribution $f(y; \theta) = \theta e^{-y\theta} \mathbb{1}\{y \geq 0\}$

Solution:

$$
\begin{aligned}
f(y; \theta) &= \theta e^{-y\theta} \mathbb{1}\{y \geq 0\} \\
&= \exp[\log(\theta e^{-y\theta} \mathbb{1}\{y \geq 0\})] \qquad\qquad = \exp[\log \theta - y\theta + \log(\mathbb{1}\{y \geq 0\})]
\end{aligned}
$$

$$a(y) = -y \quad b(\theta) = \theta \quad c(\theta) = \log \theta \quad d(y) = \log(\mathbb{1}\{y \geq 0\})$$

(c): Negative binomial distribution

$$f(y; \theta) = \binom{y+r-1}{r-1} \theta^r (1-\theta)^y \mathbb{1}\{r \in \mathbb{Z}^+, y \in (r+0, r+1, r+2, ...)\},$$

where r is known.

Solution:

$$
\begin{aligned}
f(y; \theta) &= \binom{y+r-1}{r-1} \theta^r (1-\theta)^y \mathbb{1}\{r \in \mathbb{Z}^+, y \in (r+0, r+1, r+2, ...)\} \\
&= \exp\left[\log\left(\binom{y+r-1}{r-1} \theta^r (1-\theta)^y \mathbb{1}\{r \in \mathbb{Z}^+, y \in (r+0, r+1, r+2, ...)\}\right)\right] \\
&= \exp\left[\log\binom{y+r-1}{r-1} + r\log\theta + y\log(1-\theta) + \log(\mathbb{1}\{r \in \mathbb{Z}^+, y \in (r+0, r+1, r+2, ...)\})\right]
\end{aligned}
$$

$$a(y) = y \quad b(\theta) = \log(1-\theta) \quad c(\theta) = r\log\theta \quad d(y) = \log\binom{y+r-1}{r-1} + \log(\mathbb{1}\{r \in \mathbb{Z}^+, y \in (r+0, r+1, r+2, ...)\})$$

Exercise Four

Question

Use results

$$E[a(Y)] = -c'(\theta)/b'(\theta) \quad \text{and} \quad \text{var}[a(Y)] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}$$

to verify the following results:

Solutions

(a): For $Y \sim \text{Po}(\theta)$, $E(Y) = \text{var}(Y) = \theta$.

Solutions: The canonical form of the Poisson distribution is the following:

$$\exp[y \log \theta - \theta - \log y!]$$

where $a(y) = y$, $b(\theta) = \log \theta$, $c(\theta) = -\theta$, and $d(y) = -\log y!$.

Because this is in canonical form, $E[y] = E[a(y)]$.

$$
\begin{aligned}
E[y] &= \frac{-c'(\theta)}{b'(\theta)} \\
&= \frac{1}{1/\theta} \\
&= \theta
\end{aligned}
\qquad\qquad
\begin{aligned}
\text{var}[y] &= \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} \\
&= \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} \\
&= \frac{1/\theta^2}{1/\theta^3} \\
&= \theta
\end{aligned}
$$

(b): For $Y \sim \text{N}(\mu, \sigma^2)$, $E(Y) = \mu$ and $\text{var}(Y) = \sigma^2$.

The canonical form of the normal distribution is

$$\exp\left[-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right]$$

where $a(y) = y$, $b(\mu) = \frac{\mu}{\sigma^2}$, $c(\mu) = -\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$, and $d(y) = -\frac{y^2}{2\sigma^2}$.

Because this is in canonical form, $E[y] = E[a(y)]$.

$$
\begin{aligned}
E[y] &= \frac{-c'(\mu)}{b'(\mu)} \\
&= \frac{(\mu/\sigma^2)}{(1/\sigma^2)} \\
&= \mu
\end{aligned}
\qquad\qquad
\begin{aligned}
\text{var}[y] &= \frac{b''(\mu)c'(\mu) - c''(\mu)b'(\mu)}{[b'(\mu)]^3} \\
&= \frac{0*(-\mu/\sigma^2) - (-1/\sigma^2)*(1/\sigma^2)}{1/\sigma^6} \\
&= \frac{1/\sigma^4}{1/\sigma^6} \\
&= \sigma^2
\end{aligned}
$$

(c): For $Y \sim \text{Bin}(n, \pi)$, $E(Y) = n\pi$ and $\text{var}(Y) = n\pi(1 - \pi)$

Solutions: The canonical form of the Binomial distribution is

$$\exp\left[y \log\left(\frac{\pi}{1 - \pi}\right) + n\log(1 - \pi) + \log\binom{n}{y}\right]$$

where $a(y) = y$, $b(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$, $c(\pi) = n\log(1-\pi)$, and $d(y) = \log\binom{n}{y}$.

$$\mathrm{E}[y] = \frac{-c'(\pi)}{b'(\pi)} \qquad\qquad \mathrm{var}[y] = \frac{b''(\pi)c'(\pi) - c''(\pi)b'(\pi)}{[b'(\pi)]^3}$$

$$= -\frac{n/(\pi-1)}{1/(\pi-\pi^2)} \qquad\qquad = \frac{\frac{2\pi-1}{(\pi-1)^2\pi^2}\frac{n}{\pi-1} + \frac{n}{(\pi-1)^2}\frac{1}{\pi-\pi^2}}{\frac{1}{(\pi-\pi^2)^3}}$$

$$= \frac{n(\pi^2-\pi)}{\pi-1} \qquad\qquad = \frac{\frac{n-\pi n}{(1-\pi)^3\pi^2}}{\frac{1}{(\pi-\pi^2)^3}}$$

$$= \frac{n\pi(\pi-1)}{\pi-1} \qquad\qquad = \frac{\frac{\pi n-\pi^2 n}{(\pi-\pi^2)^3}}{\frac{1}{(\pi-\pi^2)^3}}$$

$$= n\pi \qquad\qquad = n\pi - \pi^2 n$$

$$\qquad\qquad\qquad\qquad\qquad = n\pi(1-\pi)$$

Exercise Five

Questions and Solutions

(a): For a Negative Binomial distribution $Y \sim \mathrm{NBin}(r, \theta)$, find $\mathrm{E}(Y)$ and $\mathrm{var}(Y)$.

Solutions: The canonical form of the binomial distribution is listed in Exercise Three.

$$\mathrm{E}[y] = \frac{-c'(\theta)}{b'(\theta)} \qquad\qquad \mathrm{var}[y] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}$$

$$= \frac{r/\theta}{1/(1-\theta)} \qquad\qquad = \frac{-r\theta^{-1}(1-\theta)^{-2} - r\theta^{-2}(1-\theta)^{-1}}{(-(1-\theta)^{-1})^3}$$

$$= \frac{r(1-\theta)}{\theta} \qquad\qquad = \frac{r\theta^{-1}(1-\theta)^{-1}((1-\theta)^{-1} + \theta^{-1})}{(1-\theta)^{-3}}$$

$$\qquad\qquad\qquad\qquad = \frac{r\theta^{-2}(1-\theta)^{-2}}{(1-\theta)^{-3}}$$

$$\qquad\qquad\qquad\qquad = \frac{n(1-\theta)}{\theta^2}$$

(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?

Solutions:

Exercise Seven

Question

Consider N independent binary random variables $Y_1, ..., Y_N$ with

$$P(Y_i = 1) = \pi_1 \text{ and } P(Y_i = 0) = 1 - \pi_i.$$

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

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

where $y_i = 0$ or 1.

Solutions

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

$$\begin{aligned}
f(y_i; \pi_i) &= \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \\
&= \exp[\log(\pi_i^{y_i}(1 - \pi_i)^{1-y_i})] \\
&= \exp[y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i)] \\
&= \exp[y_i \log \pi_i + \log(1 - \pi_i) - y_i \log(1 - \pi_i)] \\
&= \exp\left[y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + \log(1 - \pi_i)\right]
\end{aligned}$$

where $a(y_i) = y_i, b(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right), c(\pi_i) = \log(1 - \pi_i)$, and $d(y) = 0$

(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.

By definition, $b(\pi_i)$ is the natural parameter and it is shown above.

(c): Show that $E(Y_i) = \pi_i$.

$$\begin{aligned}
E[y] &= \frac{-c'(\pi_i)}{b'(\pi_i)} \\
&= \frac{(1/(1 - \pi_i)}{1/(\pi_i - pi_i^2)}
\end{aligned}$$

(d): If the link function is

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

show that this is equivalent to modelling the probability π as

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

(e): In the particular case where $\vec{x}^T\vec{\beta} = \beta_1 + \beta_2 x$, 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?

Exercise Twelve

Questions and Solutions

(Note: See Figure 3.3 for more relationships of distributions)

(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 Uniford 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.