GLMS - Exponential Families and Link Functions Stat 495

The text presents exponential families via a formula relating any two densities in the family via a renormalized exponential tilt. There are other ways to recognize that a density is in an exponential family.

Suppose X is a random variable with density given by: $f(x \mid \theta) = a(\theta)b(x)\exp[c(\theta)d(x)]$, where a() and c() are functions only of the parameter theta, and b() and d() are functions of x (the data). If X's density can be written in this form, then the density is in an exponential family.

For example, for the Bernoulli distribution, some re-writing enables us to see that:

$$f(x \mid p) = p^x(1-p)^{1-x} = (1-p)\left(\frac{p}{1-p}\right)^x = (1-p)\exp\left[x\log\left(\frac{p}{1-p}\right)\right].$$

Here, $a(p) = 1 - p, b(x) = 1, c(p) = \log\left[\frac{p}{1-p}\right], d(x) = x.$ Hint: $r = \exp(\log(r))$.

The link function relates the parameters of the distribution to a linear function of X. The link function is c(). What is the link function for the Bernoulli distribution? Where have you seen this before?

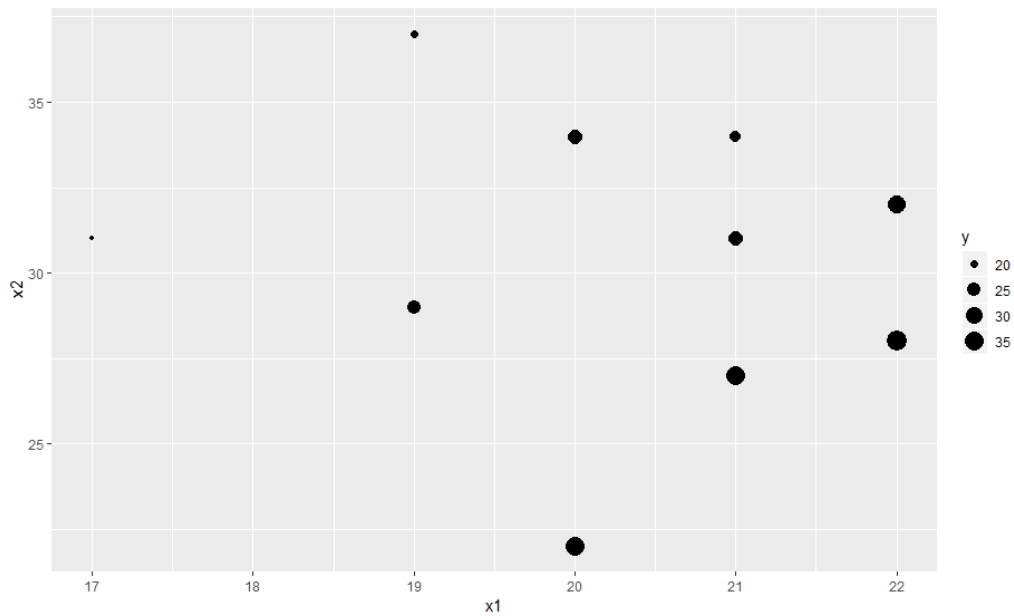
Verify that the Poisson density can be written in this form.

Poisson $\mathrm{Poi}(\mu)$ $\dfrac{e^{-\mu}\mu^x}{x!}$

What is the link function for the Poisson distribution? Your textbook calls these the natural parameters (the lambdas).

Regression trees. Use the data set provided to make your own regression tree, with a maximum depth of 3.

x1	x2	y
17	31	19
19	37	20
19	29	25
20	22	34
20	34	27
21	34	22
21	31	26
21	27	35
22	28	37
22	32	33



(You can aim for this to be a “good” tree, but don’t try to maximize any particular criteria through computation. Guestimate / eye-ball the cutoffs and write down what your tree would be.)