Exercise 1

The exponential family of distributions have a probability density given as follows, where θ is the canonical and ϕ is the dispersion parameter:

$$f(y|\theta, \phi) = exp\{\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\}$$

Show that the normal, binomial, and Poisson distributions belong to the exponential family.

Normal:

$$\begin{aligned} f(y|\mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma}} exp\{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2\} \\ &= exp(\ln \frac{1}{\sqrt{2\pi\sigma}})exp(-\frac{1}{2}\frac{y^2 - 2\mu y + \mu^2}{\sigma^2}) \\ &= exp(\frac{\mu y - \frac{1}{2}\mu^2}{\sigma^2} - \frac{1}{2}\frac{y^2}{\sigma^2} + \ln \frac{1}{\sqrt{2\pi\sigma^2}}) \\ &= exp\{\frac{\mu y - b(\mu)}{\sigma^2} + c(y, \sigma^2)\}, \\ \text{where } \theta = \mu,\ \phi = \sigma^2,\ b(\theta) &= \frac{1}{2}\theta^2,\ c(y, \phi) = \ln \frac{1}{\sqrt{2\pi\phi}} - \frac{1}{2}\frac{y^2}{\phi} \end{aligned}$$

Poisson:

$$p(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} = exp(-\lambda)exp(\ln \frac{\lambda^k}{k!}) = exp(\frac{k\lambda - \lambda^2}{\lambda} + \ln \frac{\lambda^2}{k!} - k),$$
$$\text{where } \theta = \lambda, \phi = \lambda, b(\theta) = \theta^2, c(y, \phi) = \ln \frac{\phi^y}{y!} - y$$

Binomial:

$$\begin{aligned} p(k|n, p) &= \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \\ &= exp\{\ln(\frac{p}{1-p})^k + \ln(1-p)^n + \ln\binom{n}{k}\} \\ &= exp\{k \ln \frac{p}{1-p} - \ln \frac{1}{(1-p)^n} + \ln\binom{n}{k}\} \\ &= exp\{k \ln \frac{p}{1-p} - \ln(1 + \frac{p}{1-p})^n + \ln\binom{n}{k}\} \\ &= exp\{k \ln \frac{p}{1-p} - \ln(1 + e^{\ln \frac{p}{1-p}})^{-n} + \ln\binom{n}{k}\}, \\ \text{where } \theta = \ln \frac{p}{1-p},\ \phi = 1,\ b(\theta) &= \ln(1 + e^\theta)^{-n},\ c(y, \phi) = \ln\binom{n}{y} \end{aligned}$$

Exercise 2

Identify the link function, distribution and linear predictor components in a standard linear regression model.

Linear predictor: $\eta_i = \sum_{k=1}^{p} \beta_k x_{ik} + \epsilon_i$
Distribution: $Y \sim \mathcal{N}(\mu, \sigma^2)$
Link function: $l(\mu) = \eta$

Exercise 3

Write down a Poisson regression model. Identify the distribution and linear predictor, and a suitable link function. Do the same for a binomial regression model.

Poisson:

$$\text{Distribution: } Y \sim Poisson(\mu), \text{ where } \mu > 0$$
$$\text{Link function: Need f mapping } (0, \infty) to (-\infty, \infty).\ l(\mu) = \ln(\mu) \text{works.}$$

Binomial:

$$\text{Distribution: } Y \sim Binom(\mu), \text{ i.e. } P(y|\mu) = \binom{n}{ny} \mu^{ny}(1-\mu)^{n(1-y)}$$
$$\text{Link function: Need f mapping } (0, 1) \text{to } (-\infty, \infty).\ l(\mu) = \ln \frac{p}{1-p} = \eta \text{ works.}$$

Exercise 4

Consider the linear regression model $Y = X\beta + \epsilon$. Obtain the OLS estimators $\hat{\beta}$ analytically (make suitable assumptions on the errors as needed).

$$\begin{aligned}
\arg\min_{\beta} RSS(\beta) &= \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta) \\
&= \arg\min_{\beta} (Y^T Y - \beta^T X^T Y - Y^T X\beta + \beta^T X^T X\beta) \\
&= \arg\min_{\beta} (Y^T Y - 2Y^T X\beta + (X\beta)^T X\beta) \\
\frac{\delta RSS}{\delta \beta} &= (X\beta)^T X + (X\beta)^T X - 2Y^T X = 2(X\beta)^T X - 2Y^T X \\
&\implies (X\hat{\beta})^T X - Y^T X = 0 \iff (X\hat{\beta})^T X = Y^T X \\
&\iff (\hat{\beta}^T X^T X)^T = (Y^T X)^T \iff X^T X\hat{\beta} = X^T Y \iff \hat{\beta} = (X^T X)^{-1} X^T Y
\end{aligned}$$

Exercise 5

Write down the log likelihood function for a linear regression model. Analytically, maximize the log likelihood function to obtain the ML estimators of a linear regression.

$$\text{Assume } \epsilon \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d. Then } Y = X\beta + \epsilon \sim \mathcal{N}(X\beta, \sigma^2) \text{ i.i.d., so}$$

$$f(y|x, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} exp\{-\frac{1}{2}(\frac{y - x\beta}{\sigma})^2\}$$

$$\mathcal{L}(\beta) = f(y_1|x_1, \beta, \sigma^2) \dots f(y_n|x_n, \beta, \sigma^2)$$

$$\implies \ln \mathcal{L}(\beta) = \sum_{i=1}^{n} \ln f(y_i|x_i, \beta, \sigma^2) = \sum_{i=1}^{n} \{\ln \frac{1}{\sqrt{2\pi\sigma}} - \frac{1}{2} \frac{(y_i - x_i\beta)^2}{\sigma^2}\}$$

$$\implies \arg\max_{\beta} \ln \mathcal{L} = \arg\max_{\beta} -\frac{1}{2}\{(Y - X\beta)^T(Y - X\beta)\} = \arg\min_{\beta} (Y - X\beta)^T(Y - X\beta),$$

equivalent to minimising $RSS(\beta)$.