Exponential Family

- Suppose $Y_1, \ldots, Y_n$ are independent random variables.
- Let $f(y_i; \theta_i, \phi)$ be the Probability Mass Function (PMF) or Probability Density Function (PDF) of Y_i, where ϕ is a scale parameter.
- If we can write

$$f(y_i; \theta_i, \phi) = \exp\left(y_i\theta_i - b(\theta_i)\frac{1}{a(\phi)} + c(y_i, \phi)\right),$$

then we call the PMF or the PDF $f(y_i; \theta_i, \phi)$ an exponential family.

Problem 1.

Find the form of GLM for the following distributions, and show the resonable link function:

1. Normal distributions
2. Inverse Gaussian
3. Binomial distribution
4. Poisson distribution
5. Gamma distribution
6. Beta

Normal Distribution

Assume $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$. Then, $E(Y_i) = \mu_i$ and σ is a scale parameter. The Probability Density Function (PDF) is given by

$$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y_i-\mu_i)^2}{2\sigma^2}} = \exp\left\{\frac{y_i\mu_i - \mu_i^2/2}{\sigma^2} - \left(\frac{1}{2}\log(2\pi\sigma^2) - \frac{y_i^2}{2\sigma^2}\right)\right\}.$$

Thus, use

- $\theta_i = \mu_i$,
- $b(\theta_i) = \frac{\theta_i^2}{2}$,
- $\phi = \sigma^2$,
- $a(\phi) = \phi$,
- $c(y_i, \phi) = -\frac{1}{2}\log(2\pi\phi) - \frac{y_i^2}{2\phi}$.

Inverse Gaussian Distribution

Let us rewrite the probability density function (pdf) of the Inverse Gaussian distribution with parameters μ_i and λ:

$$f(y_i;\mu_i,\lambda) = \left(\frac{\lambda}{2\pi y_i^3}\right)^{1/2} \exp\left\{-\frac{\lambda(y_i-\mu_i)^2}{2\mu_i^2 y_i}\right\}, \quad y>0$$

in the following form:

$$\begin{aligned} f(y_i;\mu_i,\lambda) &= \exp\left\{-\frac{\lambda(y_i-\mu_i)^2}{2\mu_i^2 y_i} + \frac{1}{2}\ln\left(\frac{\lambda}{2\pi y_i^3}\right)\right\} \\ &= \exp\left\{\frac{-\frac{1}{2\mu_i^2}y_i + \frac{1}{\mu_i}}{\frac{1}{\lambda}} + \left(\frac{1}{2}\ln\frac{\lambda}{2\pi y_i^3} - \frac{\lambda}{2y_i}\right)\right\} \end{aligned}$$

Now, let's identify the exponential family components:

- Canonical parameter: $\theta_i = -\frac{1}{2\mu_i^2}$
- $b(\theta_i) = -\frac{1}{\mu_i} = -(-2\theta_i)^{\frac{1}{2}}$
- $\phi = \lambda$
- $a(\phi) = \frac{1}{\phi}$
- $c(y_i,\phi) = \frac{1}{2}\ln\frac{\phi}{2\pi y_i^3} - \frac{\phi}{2y_i}$

Thus, the Inverse Gaussian distribution can be shown to be a member of the exponential family.

Binomial Distribution

Assume $Y_i \sim \text{Bin}(n_i,p_i)$. Then, $E(Y_i) = n_i p_i$. The Probability Mass Function (PMF) is given by

$$\binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i} = \exp\left\{y_i \log\left(\frac{p_i}{1-p_i}\right) + n_i\log(1-p_i) - \log\binom{n_i}{y_i}\right\}.$$

Thus,

- $\theta_i = \log\left(\frac{p_i}{1-p_i}\right)$,
- $b(\theta_i) = n_i\log(1+e^{\theta_i})$,
- $\phi = 1$, $a(\phi) = 1$,
- $c(y,\phi) = -\log\binom{n_i}{y_i}$.

Poisson Distribution

Assume $Y_i \sim \text{Poisson}(\lambda_i)$. Then, $E(Y_i) = \lambda_i$. The Probability Mass Function (PMF) is given by

$$\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} = \exp\left\{y_i \log(\lambda_i) - \lambda_i - \log(y_i!)\right\}.$$

Thus,

- $\theta_i = \log(\lambda_i)$,
- $b(\theta_i) = e^{\theta_i}$,
- $\phi = 1$, $a(\phi) = 1$,
- $c(y_i, \phi) = -\log(y_i!)$.

Gamma Distribution

Assume $x_i \sim \Gamma(\alpha, \beta_i)$, where β_i is unknown. Then, $E(x_i) = \frac{\alpha}{\beta_i}$. The Probability Mass Function (PMF) is given by

$$\frac{\beta_i^{\alpha} x_i^{\alpha-1} e^{-\beta_i x_i}}{\Gamma(\alpha)} = \exp\left\{\alpha \log x_i + \alpha \log(\beta_i) - \log(\Gamma(\alpha)) - \log(x_i) - \beta_i x_i\right\}.$$

Assuming α is known, if we choose $y_i = x_i$, then

- $\theta_i = -\beta_i$ $(\theta_i < 0)$,
- $b(\theta_i) = -\alpha \log(-\theta_i)$,
- $\phi = 1$, and $a(\phi) = 1$.

Remark: We can also choose $y_i = -x_i$ and $\theta_i = \beta_i$. In this case, $b(\theta_i) = -\alpha \log(\theta_i)$.

Problem 2.

Paper Summarization

The part 1 of this page talks about:

- **Sliced inverse regression (SIR)**: A novel data-analytic tool for reducing the dimension of the input variable x without fitting any parametric or nonparametric model[1][1]. It explores the inverse view of regression, where x is regressed against y, and uses a simple step function to estimate the inverse regression curve[2][2]. - **Effective dimension reduction (e.d.r.) space**: The linear space generated by the unknown row vectors 3k (k = 1, ..., K) in the model y = f(3lx, ..., 3Kx, e), where f is an arbitrary unknown function. The goal is to estimate this space, which captures all the information about y from x. - **Inverse regression curve**: The curve E(x I y) that connects the conditional mean of x given y as y varies. Under certain conditions, this curve falls into the e.d.r. space. A principal component analysis on the covariance matrix of the estimated inverse regression curve can locate its main orientation, yielding the estimates for e.d.r. directions[3][3]. - **Sampling properties of SIR**: The output of SIR provides root n consistent estimates for the e.d.r. directions under a design condition on the distribution of x [4]. The eigenvalues of the covariance matrix can be used to assess the number of components in the model and the effectiveness of SIR. - **Simulation results**: SIR is demonstrated to be effective in reducing the dimension of x from 10 to 2 for a data set with 400 observations. The spin-plot of y against the projected variables obtained by SIR mimics the spin-plot of y against the true directions very well [5]. A chi-squared statistic is proposed to test whether a direction found by SIR is spurious [6].

Problem 3.

Describe

1. how to conduct the EM (Expectation-Maximization) algorithm
2. how to conduct MCMC

Denoting

$$Q(q|q_0, x) = \mathbb{E}_{q_0}\left[\log L_c(q|x, Z)\right],$$

the EM algorithm indeed proceeds "iteratively" by maximizing $Q(q|q_0, x)$ at each iteration, and, if $q^{(\hat{1})}$ is the value of q maximizing $Q(q|q_0, x)$, by replacing q_0 by the updated value $q^{(\hat{1})}$. In this manner, a sequence of estimators $\{q^{(\hat{j})}\}_j$ is obtained, where

$$Q(q^{(\hat{j})}|q^{(j\hat{-}1)})$$

Pick a starting value $q^{(\hat{0})}$ and set $m = 0$. Repeat

1. Compute (the E-step)
$$Q(q|q^{(\hat{m}),x}) = \mathbb{E}_{q^{(\hat{m})}}\left[\log L_c(q|x, Z)\right],$$
where the expectation is with respect to $k(z|q^{(\hat{m}),x})$.
2. Maximize $Q(q|q^{(\hat{m}),x})$ in q and take (the M-step)
$$q^{(m\hat{+}1)} = \arg\max_q Q(q|q^{(\hat{m}),x})$$
and set $m = m + 1$.

until a fixed point is reached; i.e., $q^{(m\hat{+}1)} = q^{(\hat{m})}$.

. . . .

Gibbs Sampling of MCMC (Markov Chain Monte Carlo)

Gibbs Sampling approach for GLM with random effects - Consider with ij and T.

The likelihood function of ij is:

$$\begin{aligned} L(\beta, \gamma, \sigma^2|y, X, Z) &= f(y|X, Z, \beta, \gamma, \sigma^2) \\ &= \prod_{ij} f_{ij}(y_{ij}|x_{ij}, z_{ij}, \beta, \gamma, \sigma^2), \end{aligned}$$

where

$$f_{ij}(y_{ij}|x_{ij}, z_{ij}, \beta, \gamma, \sigma^2) = B(y_{ij}|\mu_{ij}, \sigma^2).$$

The linear predictor μ_{ij} for random effects is:

$$\mu_{ij} = x_{ij}^T\beta + z_{ij}^T\gamma.$$

The conditional distributions for Gibbs sampling are:

$$\beta|\gamma, \sigma^2, y, X, Z \sim \text{Normal}(\hat{\beta}, V_\beta),$$

$$\gamma|\beta, \sigma^2, y, X, Z \sim \text{Normal}(\hat{\gamma}, V_\gamma),$$

$$\sigma^2|\beta, \gamma, y, X, Z \sim \text{Inverse-Gamma}(a_{\sigma^2}, b_{\sigma^2}).$$

Here, $\hat{\beta}$ and V_β are the posterior mean and variance for β, $\hat{\gamma}$ and V_γ are the posterior mean and variance for γ, and a_{σ^2} and b_{σ^2} are the parameters for the Inverse-Gamma distribution.

The M-step updates for the random effects are:

$$\hat{\mu}_{ij}^{(k+1)} = \frac{1}{2}\left(\hat{\beta}^{(k+1)} x_{ij} + \hat{\gamma}^{(k+1)} z_{ij}\right),$$

where k is the current iteration.

The linear predictor for the link function is:

$$\eta_{ij} = g(\mu_{ij}),$$

where $g(\cdot)$ is the link function.

The conditional distribution for y_{ij} is:

$$y_{ij}|x_{ij}, z_{ij}, \beta, \gamma, \sigma^2 \sim f_{ij}(y_{ij}|x_{ij}, z_{ij}, \beta, \gamma, \sigma^2).$$

Problem 4.

Under the regularity conditions,

$$n^{-\frac{1}{2}}\left[\begin{pmatrix}\hat{\beta}\\ \hat{\alpha}\end{pmatrix} - \begin{pmatrix}\beta\\ \alpha\end{pmatrix}\right] \xrightarrow{d} N(0, \Sigma),$$

where $Sigma$ can be estimated By