Exponential Families

a.) $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

$$= \exp\left\{\frac{-1}{2\sigma^2}(x-\mu)^2 + \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\right\} = \exp\left\{\frac{-1}{2\sigma^2}x^2 + \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \frac{x\mu - \frac{1}{2}\mu^2}{\sigma^2}\right\}$$

$$\boxed{\begin{array}{l} a(\sigma^2) = \sigma^2 \\ b(\mu) = \frac{1}{2}\mu^2 \\ c(x;\sigma^2) = \frac{-1}{2\sigma^2}x^2 + \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) \end{array}}$$

$Y = Z/N$ where $Z \sim \text{bin}(N, p)$ $\quad Ny = Z$

$$f_Y(y) = f_Z(Ny)\left|\frac{1}{N}\right| \quad = \binom{N}{Ny} p^{Ny}(1-p)^{N-Ny}\frac{1}{N}$$

$$= \binom{N}{Ny}\frac{1}{N} p^{Ny}(1-p)^{N-Ny}$$

$$= \exp\left\{\log\binom{N}{Ny} - \log(N) + Ny\log p + N\log(1-p) - NY\log(1-p)\right\} = \exp\left\{y\left(N\log\left(\frac{p}{1-p}\right)\right) - N\log(1-p) - \log(N) + \log\binom{N}{Ny}\right\}$$

$$\boxed{\begin{array}{l} \theta = N\log\left(\frac{p}{1-p}\right) \\ b(\theta) = N\log(1-p) \\ a(\phi) = 1 \\ c(y;\phi) = \log\binom{N}{Ny} - \log(N) \end{array}}$$

$Y \sim \text{Poisson}(\lambda)$

$$P(Y=k) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{\exp\{k\log\lambda - \lambda\}}{\exp\{\log(k!)\}} = \exp\{k\log\lambda - \lambda - \log(k!)\}$$

$$\boxed{\begin{array}{l} \theta = \log\lambda \\ a(\phi) = 1 \\ b(\theta) = \lambda \\ c(y;\phi) = -\log(y!) \end{array}}$$

b.) $$s(\theta) = \frac{\partial}{\partial\theta}\log L(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f(y_i;\theta)$$

Show that $E[s(\theta)] = 0$. Let $f(x;\theta)$ be the likelihood of the vector of n observations.

$$E[s(\theta)] = \int_{\chi}\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta)\,dx$$

$$= \int_{\chi}\left(\frac{1}{f(x;\theta)}\frac{\partial}{\partial\theta}f(x;\theta)\right) f(x;\theta)\,dx \quad \text{by chain rule.}$$

$$= \int_{\chi}\frac{\partial}{\partial\theta}f(x;\theta)\,dx \quad \text{then interchange integral and partial.}$$

$$= \frac{\partial}{\partial\theta}\int_{\chi}f(x;\theta)\,dx = \frac{\partial}{\partial\theta}(1) = 0.$$

Show that $I(\theta) \equiv Var[s(\theta)] = -E[H(\theta)]$ where $H(\theta)$ is Hessian matrix of log-likelihood

$$0 = E[s(\theta)]$$

$$0 = \frac{\partial}{\partial\theta'}E[s(\theta)] = \frac{\partial}{\partial\theta^T}\int_{\chi}\frac{\partial}{\partial\theta}\log f(\theta;x)\, f(x;\theta)\,dx$$

$$= \int_{\chi}\left(\frac{\partial^2}{\partial\theta\partial\theta^T}\log f(\theta;x)\right)f(x;\theta) + \frac{\partial}{\partial\theta}\log f(\theta;x)\frac{\partial}{\partial\theta^T}f(x;\theta)\,dx$$

$$= \int_{\chi}\frac{\partial^2}{\partial\theta\partial\theta^T}\log f(\theta;x) f(x;\theta)\,dx + \int\frac{\partial}{\partial\theta}\log f(\theta;x)\frac{\partial}{\partial\theta^T}\log f(\theta;x)\, f(x;\theta)\,dx$$

bc $$\frac{\partial\log f(x;\theta)}{\partial\theta^T}f(x;\theta) = \frac{1}{f(x;\theta)}\frac{\partial}{\partial\theta^T}f(x;\theta)\, f(x;\theta) = \frac{\partial}{\partial\theta^T}f(x;\theta)$$

$$= E\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\log f(x;\theta)\right] + E\left[\underbrace{\frac{\partial}{\partial\theta}\log f(\theta;x)}_{s(\theta)}\underbrace{\frac{\partial}{\partial\theta^T}\log f(x;\theta)}_{s(\theta)^T}\right]$$

$$= E[H(\theta)] + E[s(\theta)s(\theta)^T]$$ where $H(\theta)$ = 2nd derivative of log likelihood.

$$\Rightarrow \text{var}[s(\theta)] = -E[H(\theta)]$$

c.) $f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi)\right\}$

$$s(\theta) = \frac{\partial}{\partial\theta}\log f(y;\theta,\phi)$$
$$= \frac{\partial}{\partial\theta}\log\exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi)\right\}$$
$$= \frac{\partial}{\partial\theta}\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi)\right)$$
$$= \frac{y - b'(\theta)}{a(\phi)}$$

$E\left[\frac{y - b'(\theta)}{a(\phi)}\right] = 0$ then $E(Y) = b'(\theta)$

$$\text{var}\left[\frac{y - b'(\theta)}{a(\phi)}\right] = -E[H(\theta)]$$
$$\frac{\text{Var}(y)}{a(\phi)^2} = \frac{b''(\theta)}{a(\phi)}$$
$$\text{Var}(y) = a(\phi)b''(\theta)$$

$$H(\theta) = \frac{\partial}{\partial\theta^2}\log f(y;\theta)$$
$$= \frac{\partial}{\partial\theta}\frac{y - b'(\theta)}{a(\phi)}$$
$$= \frac{-b''(\theta)}{a(\phi)}$$

d.) $E(Y) = b'(\theta) = \frac{\partial}{\partial\mu}\frac{1}{2}\mu^2 = \mu$

$$\text{Var}(Y) = a(\phi)b''(\theta) = \sigma^2\left(\frac{\partial^2}{\partial\mu^2}\frac{1}{2}\mu^2\right)$$
$$= \sigma^2\left(\frac{\partial}{\partial\mu}\mu\right)$$
$$= \sigma^2$$

Generalized Linear Models

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

$\mu_i = b'(\theta)$

$$g(b'(\theta_i)) = X_i'\beta$$
$$b'(\theta_i) = g^{-1}(X_i'\beta)$$
$$\theta_i = (b')^{-1}(g^{-1}(X_i'\beta))$$

$$\text{var}(Y_i) = a(\phi)b''(\theta) = \frac{\phi}{w_i}b''(\theta) = \frac{\phi}{w_i}b''((b')^{-1}(\underbrace{g^{-1}(X_i'\beta)}_{\mu_i})) = \frac{\phi}{w_i}b''((b')^{-1}(\mu_i))$$
$$= \frac{\phi}{w_i}V(\mu_i) \text{ then } V = b''(b'^{-1})$$

b.) $V(\mu) = \lambda$

$$b(\theta) = \lambda = \exp(\theta)$$
$$b = b' = b'' = \exp\theta$$
$$V = \exp\{\log(\mu)\} = \mu$$

$V(\mu) = \mu(1-\mu)$

$$b(\theta) = \log(1 + \exp\theta)$$
$$b'(\theta) = \frac{\exp(\theta)}{1 + \exp(\theta)}$$
$$b''(\theta) = \frac{\exp(\theta)}{(1 + \exp(\theta))^2}$$
$$(b')^{-1} = \frac{\exp\theta}{1 + \exp\theta}$$

Plug into variance function

$$V(\mu) = \frac{\exp\{\log(\frac{\mu}{1-\mu})\}}{(1 + \exp\{\log(\frac{\mu}{1-\mu})\})^2}$$
$$= \mu(1-\mu)$$

c) (1.) canonical link: $g(\mu) = (b')^{-1}(\mu)$

$b(\mu) = e^{\mu} = b'(\mu)$

$\downarrow$ invert this

$g(\mu) = \log \mu$

(2.) $b'(\mu) = \dfrac{\exp(\mu)}{1+\exp(\mu)}$

$(b')^{-1}(\mu) = \log\left(\dfrac{\mu}{1-\mu}\right)$

$y = \dfrac{\exp(\mu)}{1+\exp(\mu)}$

$y + y\exp(\mu) = \exp(\mu)$

$y = \exp(\mu)(1-y)$

$\dfrac{y}{1-y} = \exp(\mu)$

$\mu = \log\left(\dfrac{y}{1-y}\right)$

$\theta = \log \lambda$

$e^{\theta} = \lambda$

Fitting GLMs

a.) $s(\beta,\phi) \equiv \nabla_\beta \log(\beta,\phi) = \nabla_\beta \log\left[\prod_{i=1}^{n} \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i;\phi/w_i)\right\}\right]$

$$= \frac{\partial}{\partial\beta}\sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i;\phi/w_i)$$

$\theta_i = (b')^{-1}(\mu_i)$

$\frac{\partial\theta_i}{\partial\mu_i} = ((b')^{-1})'(\mu_i)$

$= \frac{1}{b''((b')^{-1}(\mu_i))}$

$\mu_i = g^{-1}(x_i'\beta)$

$\frac{\partial\mu_i}{\partial\beta} = (g^{-1})'(x_i'\beta)(x_i)$

$= \frac{x_i}{g'(g^{-1}(x_i'\beta))}$

$$= \sum_{i=1}^{n}\left(\frac{y_i - b'(\theta_i)}{\phi/w_i}\right)\left(\frac{1}{b''((b')^{-1}(\mu_i))}\right)\left(\frac{x_i}{g'(g^{-1}(x_i'\beta))}\right)$$

$$= \sum_{i=1}^{n}\left(\frac{w_i(y_i-\mu_i)}{\phi}\right)\left(\frac{x_i}{g'(\mu_i)}\right)\left(\frac{1}{b''((b')^{-1}(\mu_i))}\right)$$ because $b'(\theta_i)=\mu_i$ and $\mu_i = g^{-1}(x_i'\beta)$

$$= \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_i}{\phi\, V(\mu_i) g'(\mu_i)}$$ bc $V = b''(b')^{-1}$

b.) $g(\mu) = (b')^{-1}(\mu)$

$$g'(\mu) = \frac{1}{b''((b')^{-1}(\mu))} = \frac{1}{V(\mu)}$$

$$s(\beta,\phi) = \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_i}{\phi V(\mu_i) g'(\mu_i)}$$ substitute $g'(\mu) = \frac{1}{V(\mu)}$

$$= \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_i}{\phi V(\mu_i)\frac{1}{V(\mu_i)}}$$

$$= \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_i}{\phi}$$

c.) canonical link: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$

$\phi = w_i = 1$

$\mu_i = g^{-1}(x_i'\beta)$

$g^{-1}(\theta) = \frac{e^\theta}{1+e^\theta}$

$\mu_i = \frac{\exp\{x_i'\beta\}}{1+\exp\{x_i'\beta\}}$

$y = \log\left(\frac{\mu}{1-\mu}\right)$

$e^y = \frac{\mu}{1-\mu}$

$e^y(1-\mu) = \mu$

$e^y - \mu e^y = \mu$

$e^y = \mu(1+e^y)$

$\mu = \frac{e^y}{1+e^y}$

$$s(\beta,\phi) = \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_i}{\phi}$$

$\log L(\beta,\phi) = \log\left(\prod \exp\{\ \ \}\right)$ → constant; only y's

$= \sum_{i=1}^{n}\frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + \cancel{c(y_i;\phi/w_i)}$

$= \sum_{i=1}^{n} y_i(x_i'\beta) - b(x_i'\beta)$

$b(\theta) = \log(1+e^\theta)$

d.) $$S(\beta,\phi)=\sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_i}{\phi}$$

$$H(\beta,\phi)=\frac{\partial}{\partial\beta'}S(\beta,\phi)$$

$$=\frac{\partial}{\partial\beta'}\sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)x_i}{\phi} \qquad \frac{\partial}{\partial\mu}\cdot\frac{\partial\mu}{\partial\beta'}$$

$$=\sum_{i=1}^{n}\left(\frac{-w_i\,x_i}{\phi}\right)\frac{x_i'}{g'(g^{-1}(x_i'\beta))} \qquad \mu_i=g^{-1}(x_i'\beta) \qquad g'(\mu_i)$$

$$=-\sum_{i=1}^{n}\frac{-w_i x_i x_i'}{\phi g'(\mu_i)}=-X'WX \quad \text{where } \boxed{W=\text{diag}\left(\frac{w_i}{\phi g'(\mu_i)}\right)}$$

e.) $$f(\beta)=f(\beta_0)+\underbrace{\frac{\partial f(\beta_0)}{\partial\beta}}_{S(\beta_0)}(\beta-\beta_0)+\frac{1}{2}(\beta-\beta_0)'H(\beta_0)(\beta-\beta_0)$$

$$f(\beta_0)\to\log L(\beta_0)$$

$$\frac{\partial f(\beta_0)}{\partial\beta'}=\sum\underbrace{\frac{w_i\,b''(x_i'\beta)}{\phi}}_{W}\cdot\underbrace{\frac{(y_i-\mu_i)x_i}{b''(x_i'\beta)}}_{\tilde{z}}x_i'$$

$$=\tilde{z}'WX$$

$$=\tilde{z}_0'WX(\beta-\beta_0)-\frac{1}{2}(\beta-\beta_0)'X'WX(\beta-\beta_0)+c^*$$

$$=\tilde{z}_0'WX\beta-\frac{1}{2}\beta'X'WX\beta+\beta_0'X'WX\beta+c^*$$

$$=\tilde{y}'WX\beta-\frac{1}{2}\beta'X'WX\beta+c^* \quad \text{if } \boxed{\tilde{y}'=\tilde{z}_0'+\beta_0'X'} \text{ and}$$

$$=-\frac{1}{2}\tilde{y}'W\tilde{y}-\frac{1}{2}\beta'X'WX\beta+\tilde{y}WX\beta+c^* \quad \text{this only changes the constant}$$

$$=-\frac{1}{2}(\tilde{y}-X\beta)'W(\tilde{y}-X\beta) \quad \text{complete square}$$

f.) $$X_{n+1}=X_n-(H)^{-1}*\text{gradient}$$

$$\text{diag}\left(\frac{1}{g'(\mu_i)}\right) \quad \text{Hessian:} \quad \frac{\exp(\theta)}{1+\exp(\theta)} \qquad \theta_i=X_i'\beta$$

$$p\cdot p \qquad X: n\cdot p$$

$$\underset{(p\,n)(n\,n)(n\cdot p)}{X'WX}$$

f.) Newton's method

find roots of f

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$