Statistical Modeling SDS 383D: Excercise 2

Khai Nguyen - kbn397

April 24, 2022

Exponential Family :

(A) Let start with $Y \sim \mathcal{N}(\mu, \sigma^2)$ for known σ^2

$$
\begin{aligned}
f\left(y \mid \mu, \sigma^2\right) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(y^2-2y\mu+\mu^2\right)\right\} \cdot \exp\left\{\log\left(\left(2\pi\sigma^2\right)^{-1/2}\right)\right\} \\
&= \exp\left\{\frac{y^2}{-2\sigma^2}+y\frac{\mu}{\sigma^2}-\frac{\mu^2}{2\sigma^2}-\frac{1}{2}\log\left(2\pi\sigma^2\right)\right\} \\
&= \exp\left\{\frac{y\mu-\frac{1}{2}\mu^2}{\sigma^2}+\left(-\frac{y^2}{2\sigma^2}-\frac{1}{2}\log\left(2\pi\sigma^2\right)\right)\right\}
\end{aligned}
$$

So, $\theta = \mu, a(\phi) = \sigma^2, b(\theta) = \frac{1}{2}\mu^2,$ and $c(y \mid \phi) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)$.

Next, $Y = Z/N$ where $Z \sim Binom(N, P)$ with known N.

$$
\begin{aligned}
P\left(Y=\frac{Z}{N}\right) &= P(Z=NY) \cdot \left|\frac{1}{N}\right| \\
&= \left(\begin{array}{c} N \\ NY \end{array}\right) P^{NY}(1-P)^{N-NY} \cdot \frac{1}{N} \\
&= \exp\left\{\log\left[\left(\begin{array}{c} N \\ NY \end{array}\right) P^{NY}(1-P)^{N-NY} \cdot \frac{1}{N}\right]\right\} \\
&= \exp\left\{\log\left[\left(\begin{array}{c} N \\ NY \end{array}\right)\right] + NY\log(P) + N\log(1-P) - NY\log(1-P) - \log(N)\right\} \\
&= \exp\left\{Y\left[N\log\left(\frac{P}{1-P}\right)\right] - N\log\left(\frac{1}{1-P}\right) + \log\left[\left(\begin{array}{c} N \\ NY \end{array}\right)\right] - \log(N)\right\}
\end{aligned}
$$

So, $\theta = N\log\left(\frac{P}{1-P}\right), b(\theta) = N\log\left(\frac{1}{1-P}\right), a(\phi) = 1,$ and $c(y \mid \phi) = \log\left[\left(\begin{array}{c} N \\ NY \end{array}\right)\right] - \log(N)$.

Next, $Y \sim Poisson(\lambda)$:

$$
\begin{aligned}
f(y \mid \lambda) &= \frac{e^{-\lambda}\lambda^y}{y!} \\
&= \exp\left\{\log\left[\frac{e^{-\lambda}\lambda^y}{y!}\right]\right\} \\
&= \exp\{-\lambda + y\log(\lambda) - \log(y!)\} \\
&= \exp\{y\log(\lambda) - \lambda + (-\log(y!))\}
\end{aligned}
$$

So, $\theta = \log(\lambda), a(\phi) = 1, b(\theta) = \lambda$, and $c(y \mid \phi) = -\log(y!)$.
(B) We have

$$
\begin{aligned}
E[s(\theta)] &= \int_{\mathcal{Y}} \frac{\frac{\partial}{\partial \theta} f(y \mid \theta)}{f(y \mid \theta)} \cdot f(y \mid \theta) dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} f(y \mid \theta) dy \\
&= \frac{\partial}{\partial \theta} \int_{\mathcal{Y}} f(y \mid \theta) dy \\
&= \frac{\partial}{\partial \theta}(1) = 0
\end{aligned}
$$

Next,

$$
\begin{aligned}
\frac{\partial}{\partial \theta^T} E(s(\theta)) &= \frac{\partial}{\partial \theta^T} \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \log L(\theta) f(y \mid \theta) dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta^T} \left[\frac{\partial}{\partial \theta} \log L(\theta) f(y \mid \theta) \right] dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \log L(\theta) \cdot \frac{\partial}{\partial \theta^T} f(y \mid \theta) + f(y \mid \theta) \frac{\partial^2}{\partial \theta^T \theta} \log L(\theta) dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \log L(\theta) \cdot \frac{\partial}{\partial \theta^T} L(\theta) dy + \int_{\mathcal{Y}} f(y \mid \theta) \frac{\partial^2}{\partial \theta^T \theta} \log L(\theta) dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \log L(\theta) \cdot \frac{\partial}{\partial \theta^T} \log L(\theta) \cdot f(y \mid \theta) dy + E\left[\frac{\partial^2}{\partial \theta^T \theta} \log L(\theta) \right] \\
&= E\left[\frac{\partial}{\partial \theta} \log L(\theta) \cdot \frac{\partial}{\partial \theta^T} \log L(\theta) \right] + E[H(\theta)] \\
&= E\left[s(\theta) s(\theta)^T \right] + E[H(\theta)] \\
&= \frac{\partial}{\partial \theta^T}(0) = 0
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\operatorname{var}[s(\theta)] &= E\left[s(\theta) s(\theta)^T \right] - (E[s(\theta)])^2 \\
&= E\left[s(\theta) s(\theta)^T \right] \\
&= -E[H(\theta)]
\end{aligned}
$$

(C) We have

$$
\begin{aligned}
E[s(\theta)] &= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta}\left[\frac{\theta}{a(\phi)} \sum_{i=1}^{n} y_i - \frac{nb(\theta)}{a(\phi)} + \sum_{i=1}^{n} c\left(y_i \mid \phi\right)\right] f(y \mid \theta) dy \\
&= \int_{\mathcal{Y}} \left[\frac{1}{a(\phi)} \sum_{i=1}^{n} y_i - \frac{nb'(\theta)}{a(\phi)}\right] f(y \mid \theta) dy \\
&= E\left[\frac{\sum_{i=1}^{n} y_i}{a(\phi)} - \frac{nb'(\theta)}{a(\phi)}\right] \\
&= \frac{1}{a(\phi)} \sum_{i=1}^{n} E(Y) - \frac{nb'(\theta)}{a(\phi)} \\
&= 0
\end{aligned}
$$

Therefore, $E[Y] = b'(\theta)$. For the variance,

$$
\begin{aligned}
\operatorname{var}(s(\theta)) &= \operatorname{var}\left[\frac{1}{a(\phi)} \sum_{i=1}^{n} y_i - \frac{nb'(\theta)}{a(\phi)}\right] \\
&= \frac{1}{a(\phi)^2} \sum_{i=1}^{n} \operatorname{var}(Y) \\
&= -E[H(\theta)]
\end{aligned}
$$

We also have:

$$
\begin{aligned}
-E[H(\theta)] &= -E\left[\frac{\partial}{\partial \theta^T}\left(\frac{1}{a(\phi)} \sum_{i=1}^{n} y_i - \frac{nb'(\theta)}{a(\phi)}\right)\right] \\
&= -\int_{\mathcal{Y}} \frac{\partial}{\partial \theta^T}\left(\frac{1}{a(\phi)} \sum_{i=1}^{n} y_i - \frac{nb'(\theta)}{a(\phi)}\right) f(y \mid \theta) dy \\
&= \int_{\mathcal{Y}} \frac{nb''(\theta)}{a(\phi)} f(y \mid \theta) dy \\
&= E\left[\frac{nb''(\theta)}{a(\phi)}\right] \\
&= \frac{nb''(\theta)}{a(\phi)}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{1}{a(\phi)^2} \sum_{i=1}^{n} \operatorname{var}(Y) &= \frac{nb''(\theta)}{a(\phi)} \\
\Longrightarrow \operatorname{var}(Y) &= a(\phi) b''(\theta)
\end{aligned}
$$

(D) We have

$$\begin{aligned}
E(Y) &= b'(\theta) \\
&= \frac{\partial}{\partial \mu}\left(\frac{1}{2}\mu^2\right) \\
&= \mu, \\
\operatorname{var}(Y) &= a(\phi) b''(\theta) \\
&= \sigma^2 \frac{\partial^2}{\partial \mu^2}\left(\frac{1}{2}\mu^2\right) \\
&= \sigma^2 \frac{\partial}{\partial \mu}(\mu) \\
&= \sigma^2,
\end{aligned}$$

Generalized Linear Model :

(A) We have

$$\mu_i = E[Y_i; \theta_i, \phi] = b'(\theta)$$

Since we know that $g(\mu_i) = x_i^\top \beta$, we have

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

Combining the above two equations, we have

$$b'(\theta) = g^{-1}(x_i^\top \beta) \iff \theta = (b')^{-1}(g^{-1}(x_i^\top \beta))$$

We know that

$$Var[Y_i; \theta_i, \phi] = a(\phi) b''(\theta_i)$$

since $a(\phi) = \frac{\phi}{w_i}$ and $b''(\theta_i) = b''((b')^{-1}(\mu)) = V(\mu)$

(B) (1) We have the exponential family form of Poisson distribution is:

$$\begin{aligned}
f(y_i; \theta_i, \phi) &= \exp\left\{\frac{y_i \theta_i - b(\theta_i)}{\phi / w_i} + c(y_i; \phi / w_i)\right\} \\
&= \exp\{y \log \lambda - \lambda - \log(y!)\}
\end{aligned}$$

In this case, $\theta = \log \lambda$ and $b(\theta) = e^\theta$. Hence $b'(\theta) = e^\theta$,$b''(\theta) = \mu$.

(2) We have the exponential family form of Bionmial distribution with $\theta = N \log\left(\frac{p}{1-p}\right)$ and $b(\theta) = -N\log(1 - \frac{e^{\theta/N}}{1+e^{\theta/N}}) = N\log(1 + e^{\theta/N})$. We have

$$\begin{aligned}
b'(\theta) &= \frac{e^{\theta/N}}{1 + e^{\theta/N}} = \mu \\
b''(\theta) &= \frac{\frac{1}{N}e^{\theta/N}(1 + e^{\theta/N}) - \frac{1}{N}e^{\theta/N}e^{\theta/N}}{(1 + e^{\theta/N})^2} = \mu(1-\mu)
\end{aligned}$$

(C) (1) It's easy to see that $b'(\theta) = e^\theta = \mu$ and $\theta = (b')^{-1}(\mu)$ hence $(b')^{-1}(\mu) = \log(\mu)$.
(2) We have $b'(\theta) = \frac{e^{\theta/N}}{1+e^{\theta/N}} = \mu$ which implies

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

Fitting GLMS :
(A) We have

$$\frac{\partial}{\partial\beta}L(\beta,\phi) = \sum_{i=1}^{N}\frac{\partial}{\partial\beta}L_i(\beta,\phi)$$

By the chain rule

$$\frac{\partial}{\partial\beta}L_i(\beta,\phi) = \frac{\partial L_i}{\partial\theta_i}\cdot\frac{\partial\theta_i}{\partial\mu_i}\cdot\frac{\partial\mu_i}{\partial\beta}$$

In detialm

$$\begin{aligned}
\frac{\partial L_i}{\partial\theta_i} &= \frac{\partial}{\partial\theta_i}\left(\frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i;\phi/w_i)\right) = \frac{w_i(y_i-\mu)}{\phi} \\
\frac{\partial\theta_i}{\partial\mu_i} &= \frac{\partial}{\partial\mu_i}(b')^{-1}(\mu_i) = \frac{1}{b''((b')^{-1}(\mu_i))} = \frac{1}{V(\mu_i)} \\
\frac{\partial\mu_i}{\beta} &= \frac{\partial}{\partial\beta}g^{-1}(x_i^\top\beta) = \frac{x_i}{g'(g^{-1}(x_i^\top\beta))} = \frac{x_i}{g'(\mu_i)}
\end{aligned}$$

Combining them we get the target equation.

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

(B) Using the result from (A), we plug $g'(\mu) = 1/V(\mu)$ in

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

(C) We have

$$\log\left(\frac{p_i}{1-p_i}\right) = x_i^\top\beta$$
$$\iff p_i = \frac{e^{x_i^\top\beta}}{1+e^{x_i^\top\beta}}$$

So the log-likelihood function is

$$\begin{aligned}
L(\beta) &= \sum_{i=1}^{M}\left(y_i\log(\frac{e^{x_i^\top\beta}}{1+e^{x_i^\top\beta}}) + (N_i-y_i)\log(\frac{1}{1+e^{x_i^\top\beta}})\right) + constant \\
&= \sum_{i=1}^{M}\left(y_ix_i^\top\beta - N_i\log(1+\exp\left(x_i^\top\beta\right))\right) + constant
\end{aligned}$$

Taking the gradient respect to β:

$$\nabla_\beta = \sum_{i=1}^{n}(Y_i - Np_i)x_i = \sum_{i=1}^{n}(Y_i - N_i\frac{e^{x_i^\top \beta}}{1+e^{x_i^\top \beta}})x_i$$

Using gradient descent using the stopping criteria:

$$\frac{|L(\beta^t) - L(\beta^{t-1}|}{|L(\beta^{t-1}|} \leq 1e-8$$

I plot the log-likelihood over iterations and the result of Sklearn package in Figure 1.

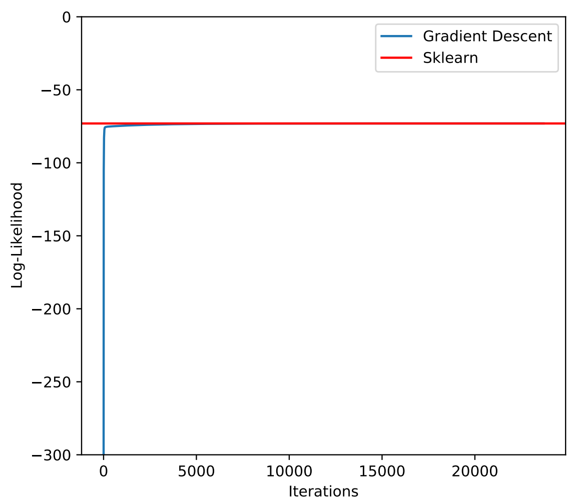

Figure 1: Gradient descent for Logistic regression.

(D) We already have the

$$\nabla_\beta \log L(\beta, \phi) = \sum_{i=1}^{n} \frac{w_i\,(Y_i - \mu_i)\,x_i}{\phi}$$

Hence

$$\begin{aligned}
H(\beta, \phi) &= \frac{\partial^2}{\partial \beta \partial \beta^T} \log L(\beta, \phi) \\
&= \frac{\partial}{\partial \beta^\top} \sum_{i=1}^{n} \frac{w_i \left(Y_i - \mu_i\right) x_i}{\phi} \\
&= \frac{\partial}{\partial \beta^\top} \sum_{i=1}^{n} \frac{w_i \left(Y_i - \mu_i\right) x_i}{\phi} \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \beta^\top} \frac{w_i \left(Y_i - \mu_i\right) x_i}{\phi} \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \beta^\top} \frac{w_i \left(Y_i - g^{-1}\left(x_i^T \beta\right)\right) x_i}{\phi} \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \beta^\top} \frac{w_i Y_i x_i - w_i g^{-1}\left(x_i^T \beta\right) x_i}{\phi} \\
&= \sum_{i=1}^{n} \frac{-w_i x_i x_i^\top}{\phi g'\left(g^{-1}(x_i^T \beta)\right)} = \sum_{i=1}^{n} \frac{-w_i x_i x_i^\top V(\mu_i)}{\phi}
\end{aligned}$$

(E) We have the high dimensional Taylor Expansion:

$$f(\mathbf{x}) \approx f\left(\mathbf{x}_0\right) + \nabla^T \left(\mathbf{x} - \mathbf{x}_0\right) + \frac{1}{2}\left(\mathbf{x} - \mathbf{x}_0\right)^T \mathbf{H}\left(\mathbf{x} - \mathbf{x}_0\right)$$

Hence

$$\begin{aligned}
\log L(\beta, \phi) \approx \log L(\beta_0, \phi) &+ \sum_{i=1}^{n} \left(\frac{w_i \left(Y_i - \mu_i\right) x_i}{\phi}\right)^\top (\beta - \beta_0) \\
&- \frac{1}{2}(\beta - \beta_0)^\top \left(\sum_{i=1}^{n} \frac{-w_i x_i x_i^\top V(\mu_i)}{\phi}\right)(\beta - \beta_0) \\
&= \log L(\beta_0, \phi) + \left(X^\top W_1 (Y - \mu)\right)^\top (\beta - \beta_0) - \frac{1}{2}(\beta - \beta_0)^\top X^\top W X (\beta - \beta_0)
\end{aligned}$$

with $W_1 = diag(w_1/\phi, \ldots, w_p/\phi)$ and $W = diag(w_1 V(\mu_1)/\phi, \ldots, w_p V(\mu_p)/\phi)$. Then, we have

$$x^T M x - 2b^T x = \left(x - M^{-1} b\right)^T M \left(x - M^{-1} b\right) - b^T M^{-1} b$$

Then

$$\begin{aligned}
\log L(\beta, \phi) &\approx -\frac{1}{2}\left((\beta - \beta_0)^\top X^\top W X (\beta - \beta_0) - 2\left(X^\top W_1 (Y - \mu)\right)^\top (\beta - \beta_0)\right) \\
&\approx -\frac{1}{2}(\beta - \beta_0 - (X^\top W X)^{-1}\left(X^\top W_1 (Y - \mu)\right))^\top X^\top W X (\beta - \beta_0 - (X^\top W X)^{-1}\left(X^\top W_1 (Y - \mu)\right))
\end{aligned}$$

(F) I got

$$\begin{aligned}
\beta = [&-0.48701675, 7.21550165, -1.65330142, 1.73610268, -13.99253364, \\
&-1.07400828, 0.07716665, -0.67452961, -2.59059481, -0.445864, 0.48206004]]
\end{aligned}$$

and the result of Skleanr is

$$\begin{aligned}\beta = [&-0.48678711, 7.20590997, -1.6533028, 1.74444156, -13.99050961,\\ &-1.07393186, 0.07669252, -0.67463619, -2.59065684, -0.44585117, 0.48209165\end{aligned}$$