MATH189HW3

Shu Bin

February 18, 2019

[1] Murphy 2.16

$$
\begin{aligned}
E(\theta) &= \int_0^1 \theta \mathbb{P}(\theta; a, b) d\theta \\
&= \int_0^1 \theta \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1}
\end{aligned}
$$

Plugging in $B(a+1,b) = \int_0^1 \theta^a (1-\theta)^{b-1} d\theta$, we have:

$$
\begin{aligned}
E(\theta) &= \frac{B(a+1,b)}{B(a,b)} \\
&= \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{a}{a+b}
\end{aligned}
$$

Then:

$$
\begin{aligned}
Var(\theta) &= E\big[(\theta - E(\theta)^2)\big] \\
&= E(\theta^2) - E(\theta)^2 \\
&= E(\theta)^2 - \left(\frac{a}{a+b}\right)^2 \\
E(\theta^2) &= \int_0^1 \theta^2 \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1} d\theta \\
&= \frac{1}{B(a,b)} \int_0^1 \theta^{a+1}(1-\theta)^{b-1} d\theta \\
&= \frac{B(a+2,b)}{B(a,b)} \\
&= \frac{(a)(a+1)\Gamma(a)\Gamma(b)}{(a+b)(a+b+1)\Gamma(a+b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} \\
Var(\theta) &= \frac{a(a+1)}{(a+b)(a+b+1)} - \left(\frac{a}{a+b}\right)^2 \\
&= \frac{(a^2+a)(a+b) - (a^3+a^2)}{(a+b)^2(a+b+1)} \\
&= \frac{a^3 + a^2b + a^2 + ab - a^3 - a^2}{(a^2+2ab+b^2)(a+b+1)} \\
&= \frac{a^3 + a^2b + a^2 + ab - a^3 - a^2}{a^3 + a^2b + a^2 + 2a^2b + 2ab^2 + 2ab + ab^2 + b^3 + b^2} \\
&= \frac{a^2b + ab}{a^3 + b^3 + a^2 + b^2 + 3a^2b + 3ab^2 + 2 + 2ab}
\end{aligned}
$$

[2] Murphy 9

$$
\begin{aligned}
Cat(x|\mu) &= exp\Bigg(log\Bigg(\prod_{i=1}^{K}\mu_i^{x_i}\Bigg)\Bigg) \\
&= exp\Bigg(\sum_{i=1}^{K} x_i log(\mu_i)\Bigg) \\
&= exp\Bigg(\sum_{i=1}^{K-1} x_i log(\mu_i) + x_k log(\mu_k)\Bigg) \\
&= exp\Bigg(\sum_{i=1}^{K-1} x_i log(\mu_i) + \Bigg(1 - \sum_{i=1}^{K-1} x_i\Bigg) log(\mu_k)\Bigg) \\
&= exp\Bigg(\sum_{i=1}^{K-1} x_i log(\mu_i) + log(\mu k) - \Bigg(\sum_{i=1}^{K-1} x_i\Bigg) log(\mu_k)\Bigg) \\
&= exp\Bigg(\sum_{i=1}^{K-1} x_i \big(log(\mu_i) - log(\mu_k)\big) + log(\mu_k)\Bigg) \\
&= exp\Bigg(\sum_{i=1}^{K-1} x_i log\Bigg(\frac{\mu_i}{\mu_k}\Bigg) + log(\mu_k)\Bigg)
\end{aligned}
$$

Let

$$
\vec{v} = \begin{pmatrix} log(\frac{\mu_1}{\mu_k}) \\ \cdots \\ log(\frac{\mu_{k-1}}{\mu_k}) \end{pmatrix}
$$

Then

$$
\begin{aligned}
\mu_i &= \mu_k e^{\vec{v}_i} \\
\mu_k &= 1 - \sum_{i=1}^{K-1} \mu_i \\
&= 1 - \sum_{i=1}^{K-1} \mu_k e^{\vec{v}_i} \\
&= 1 - \mu_k \sum_{i=1}^{K-1} e^{\vec{v}_i} \\
&= \frac{1}{1 + \sum_{i=1}^{K-1} e^{\vec{v}_i}} \\
\mu_i &= \frac{1}{1 + \sum_{i=1}^{K-1} e^{\vec{v}_i}} e^{\vec{v}_i} \\
&= \frac{e^{\vec{v}_i}}{1 + \sum_{i=1}^{K-1} e^{\vec{v}_i}}
\end{aligned}
$$

Thus

$$
\begin{aligned}
Cat(x|\mu) &= exp(\vec{v}^T x - a(\vec{v})) \\
b(\vec{v}) &= 1 \\
T(x) &= x \\
a(\vec{v}) &= -log(\mu_k) \\
&= log\left(1 + \sum_{i=1}^{K-1} e^{\vec{v}_i}\right)
\end{aligned}
$$

Let $S(\vec{v})$ be the softmax regression, then $\mu = S(\vec{v})$. Thus $Cat(x|\mu)$ is in the exponential family and the generalized linear model of $Cat(x|\mu)$ is the same as softmax regression.