Math 189R problem set 3

Adam Guo 2020-02-17

1. **(Murphy 2.16)** Suppose $\theta \sim Beta(a, b)$ such that

$$\mathbb{P}(\theta; a, b) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function and $\Gamma(x)$ is the gamma function. Derive the mean, mode, and variance of θ.

Solution:

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

Mean of θ is $\frac{a}{a+b}$.

To find mode, we want to find the maximum of $\mathbb{P}(\theta; a, b)$.

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{P}(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(-\theta^{a-1}(b-1)(1-\theta)^{b-2} + (a-1)\theta^{a-2}(1-\theta)^{b-1})$$

Set the derivative to 0,

$$\begin{aligned}
0 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(-\theta^{a-1}(b-1)(1-\theta)^{b-2} + (a-1)\theta^{a-2}(1-\theta)^{b-1}) \\
0 &= -\theta^{a-1}(b-1)(1-\theta)^{b-2} + (a-1)\theta^{a-2}(1-\theta)^{b-1} \\
0 &= -\theta(b-1) + (a-1)(1-\theta) \\
0 &= -b\theta + \theta + a - a\theta - 1 + \theta \\
0 &= \theta(2-a-b) + a - 1 \\
\theta &= \frac{1-a}{2-a-b}
\end{aligned}$$

Mode of θ is $\frac{1-a}{2-a-b}$.

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

Variance of θ is $\frac{ab}{(a+b)^2(a+b+1)}$.

2. **(Murphy 9)** Show that the multinomial distribution

$$Cat(x \mid \mu) = \prod_{i=1}^{K} \mu_i^{x_i}$$

is in the exponential family and show that the generalised linear model corresponding to this distribution is the same as multinomial logistic regression (softmax regression).

Solution:

By Murphy, the exponential family is in the form

$$p(x \mid \theta) = h(x) \exp\big(\theta^T \phi(x) - A(\theta)\big)$$

In the multinomial distribution,

$$x_k = \mathbb{1}(x = k)$$

Expressing $Cat(x \mid \mu)$ in log form:

$$\begin{aligned}
p(x \mid \mu) &= \prod_{i=1}^{K} \mu_i^{x_i} \\
&= \exp\left(\ln\left(\prod_{i=1}^{K} \mu_i^{x_i}\right)\right) \\
&= \exp\left(\sum_{i=1}^{K} x_i \ln(\mu_i)\right)
\end{aligned}$$

Since μ is a vector of probabilities for the value of x, $\sum_{i=1}^{K} \mu_i = 1$ and $\mu_K = 1 - \sum_{i=1}^{K-1} \mu_i$. We also have $x_K = 1 - \sum_{i=1}^{K-1} x_i$.

$$\begin{aligned}
p(x \mid \mu) &= \exp\left(\sum_{i=1}^{K-1} x_i \ln(\mu_i) + (1 - \sum_{i=1}^{K-1} x_i) \ln(\mu_K)\right) \\
&= \exp\left(\sum_{i=1}^{K-1} x_i \ln(\mu_i) + \ln(\mu_K) - \sum_{i=1}^{K-1} x_i(\ln(\mu_K))\right) \\
&= \exp\left(\sum_{i=1}^{K-1} x_i \ln\left(\frac{\mu_i}{\mu_K}\right) + \ln(\mu_K)\right)
\end{aligned}$$

Hence, we can write

$$p(x \mid \mu) = h(x) \exp\big(\theta^T \phi(x) - A(\theta)\big)$$

where

$$h(x) = 1, \quad \theta_i = \ln\left(\frac{\mu_i}{\mu_K}\right) \text{ for } 1 \leq i \leq K-1, \quad \phi(x) = x, \quad A(\theta) = -\ln(\mu_K)$$

Solving for μ_i in terms of θ_i,

$$\begin{aligned}
\theta_i &= \ln\left(\frac{\mu_i}{\mu_K}\right) \\
\mu_i &= e^{\theta_i} \mu_K
\end{aligned}$$

Solving for μ_K,

$$\begin{aligned}
\mu_K &= 1 - \sum_{i=1}^{K-1} \mu_i \\
&= 1 - \sum_{i=1}^{K-1} e^{\theta_i} \mu_K \\
0 &= 1 - \mu_K (1 + \sum_{i=1}^{K-1} e^{\theta_i}) \\
\mu_K &= \frac{1}{1 + \sum_{i=1}^{K-1} e^{\theta_i}}
\end{aligned}$$

Substituting back into μ_i,

$$\mu_i = \frac{e^{\theta_i}}{1 + \sum_{i=1}^{K-1} e^{\theta_i}}$$

Thus,

$$p(x \mid \mu) = \exp\bigl(\theta^T \phi(x) - A(\theta)\bigr)$$

where

$$A(\theta) = -\ln(\mu_K) = \ln\left(1 - \sum_{i=1}^{K-1} e^{\theta_i}\right)$$

and we can write

$$\mu = S(\theta)$$

where S is the softmax function. Thus, $Cat(x \mid \mu)$ is in the exponential family, and the generalised linear model is the same as softmax regression.