Aimee Co
Math189 SP24
Homework 3
2/12/24

Feel free to work with other students, but make sure you write up the homework and code on your own (no copying homework *or* code; no pair programming). Feel free to ask students or instructors for help debugging code or whatever else, though.

1 (**Murphy 2.16**) Suppose $\theta \sim \text{Beta}(a, b)$ 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.

Given that $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$
The mean is given by:

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

The mode, for $a > 1$ and $b > 1$, is found by solving:

$$\begin{aligned}
\nabla_\theta \mathbb{P}(\theta; a, b) &= \frac{d}{d\theta}\left(\theta^{a-1}(1-\theta)^{b-1}\right) = 0 \\
&= (a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2} = 0 \\
(a-1)\theta^{a-2}(1-\theta)^{b-1} &= (b-1)\theta^{a-1}(1-\theta)^{b-2} \\
(a-1)(1-\theta) &= (b-1)\theta \\
(a+b-2)\theta &= a-1 \\
\text{Mode}[\theta] &= \frac{a-1}{a+b-2}
\end{aligned}$$

To compute variance, first calculate $\mathbb{E}[\theta^2]$ and then subtract $(\mathbb{E}[\theta])^2$:

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

$$\begin{aligned}
Var[\theta] &= \mathbb{E}[\theta^2] - (\mathbb{E}[\theta])^2 \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2} \\
&= \frac{a^3 + a^2 b + a^2 + ab - a^3 - a^2 b - a^2}{(a+b)^2 (a+b+1)} \\
&= \frac{ab}{(a+b)^2 (a+b+1)}
\end{aligned}$$

■

2 (**Murphy 9**) Show that the multinoulli distribution

$$\text{Cat}(\mathbf{x}|\boldsymbol{\mu}) = \prod_{i=1}^{K} \mu_i^{x_i}$$

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

Rewriting the multinomial distribution as a summation function:

$$\begin{aligned}
\text{Cat}(x|\mu) &= \prod_{i=1}^{K} \mu_i^{x_i} \\
&= \exp\left[\log\left(\prod_{i=1}^{K} \mu_i^{x_i}\right)\right] \\
&= \exp\left(\sum_{i=1}^{K} \log(\mu_i^{x_i})\right) \\
&= \exp\left(\sum_{i=1}^{K} x_i \log(\mu_i)\right)
\end{aligned}$$

Since $\sum_{i=1}^{K} \mu_i = 1$ and $\sum_{i=1}^{K} x_i = 1$, we only need to consider the first $K-1$ terms. Thus,

$$\begin{aligned}
\mu_K &= 1 - \sum_{i=1}^{K-1} \mu_i \\
x_K &= 1 - \sum_{i=1}^{K-1} x_i
\end{aligned}$$

The summation expression becomes:

$$\begin{aligned}
\text{Cat}(x|\mu) &= \exp\left(\sum_{i=1}^{K} x_i \log(\mu_i)\right) \\
&= \exp\left(\sum_{i=1}^{K-1} x_i \log(\mu_i) + x_K \log(\mu_K)\right) \\
&= \exp\left[\sum_{i=1}^{K-1} x_i \log(\mu_i) + \left(1 - \sum_{i=1}^{K-1} x_i\right) \log(\mu_K)\right] \\
&= \exp\left[\sum_{i=1}^{K-1} x_i \log\left(\frac{\mu_i}{\mu_K}\right) + \log(\mu_K)\right]
\end{aligned}$$

Let vector η be

$$\eta = \begin{bmatrix} \log\left(\frac{\mu_1}{\mu_K}\right) \\ \dots \\ \log\left(\frac{\mu_{K-1}}{\mu_K}\right) \end{bmatrix}$$

Substitute $\mu_i = \mu_K e^{\eta_i}$ into the expression for μ_K:

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

Comparing expression to $\mathrm{Cat}(x|\mu) = \exp(\eta^\top x - a(\eta))$, we get:

$$\begin{aligned} &b(\eta) = 1 \\ &T(x) = x \\ &a(\eta) = -\log(\mu_K) = \log\left(1 + \sum_{i=1}^{K-1} e^{\eta_i}\right) \end{aligned}$$

Thus, the distribution $\mathrm{Cat}(x|\mu)$ is in the exponential family. $\mu = S(\eta)$ (softmax function), showing that the generalized linear model of the multinomial distribution is the same as the softmax regression. ■