Math189R SP19
Homework 3
Monday, Feb 18, 2019

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 $\theta$.

First finding the mean of $\theta$:

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

Note:

$$B(a,b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Therefore we have

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

Now finding the variance we have:

$$\text{var}[\theta] = \mathbb{E}[\theta^2] - \mathbb{E}[\theta]^2$$

We need to compute $\mathbb{E}[\theta^2]$

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

So now we have:

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

The mode of $\theta$ will be the value of $\theta$ when $\nabla_\theta \mathbb{P}(\theta; a, b)$ is equal to 0. Since the $\frac{1}{B(a,b)}$ term in constant in this case and we are setting the gradient to 0, we can ignore it in the computation.

$$
\begin{aligned}
\nabla_\theta \mathbb{P}(\theta; a, b) &= \nabla_\theta [\theta^{a-1}(1-\theta)^{b-1}] \\
&= (a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-1}
\end{aligned}
$$

Setting this equal to 0 and solving for $\theta$:

$$0 = (a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2}$$

$$(b-1)\theta^{a-1}(1-\theta)^{b-2} = (a-1)\theta^{a-2}(1-\theta)^{b-1}$$

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

Therefore the mode is $\frac{a-1}{a+b-2}$.

■

**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).

First we want to express the function as a exponent, so it would look like the following:

$$\prod_{i=1}^{K} \mu_i^{x_i} = \exp(\log \prod_{i=1}^{K} \mu_i^{x_i})$$

From log rules we can change this expression to be a sum of logarithms rather than the logarithm of a product:

$$\begin{aligned} \exp(\log \prod_{i=1}^{K} \mu_i^{x_i}) &= \exp(\sum_{i=1}^{K} \log \mu_i^{x_i}) \\ &= \exp(\sum_{i=1}^{K} x_i \log \mu_i) \end{aligned}$$

Therefore we can allow $\eta$ to be $\eta = \begin{bmatrix} \log \mu_i \\ \vdots \\ \log \mu_K \end{bmatrix}$.

Now we have

$$\text{Cat}(\mathbf{x}|\boldsymbol{\mu}) = \exp(\eta^T \mathbf{x} - a(\eta))$$

where $a(\eta) = 0$, and the distribution is in the exponential family.

■