Sam Mikes
Math189R SU20.1
Homework 3
Thursday, June 18, 2019

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.

Mean

$$\begin{aligned}
E(x) &= \int \theta \mathbb{P}(\theta; a, b) d\theta \\
&= \frac{1}{B(a,b)} \int \theta\theta^{a-1}(1-\theta)^{b-1} d\theta \\
&= \frac{1}{B(a,b)} \int \theta^{a}(1-\theta)^{b-1} d\theta \\
&= \frac{1}{B(a,b)} \int (1-\theta)^{b-1}\theta^{a} d\theta \\
\text{(by parts)} &= \frac{1}{B(a,b)}[(1-\theta)^{b-1}\frac{\theta^{a+1}}{a+1}) + \frac{b-1}{a+1}\int \theta^{a+1}(1-\theta)^{b-2} d\theta]
\end{aligned} \tag{1}$$

Or maybe mathematica. Closed form:

$$\int tB_t(t, a, b)dt = \frac{1}{2}(t^2 B_t(t, a, b) - B_t(t, 2+a, b)) \tag{2}$$

since $B(0, a, b) = 0$ and $B(1, a, b)$ is a constant in terms of the Gamma function

$$\begin{aligned}
E(x) &= \frac{1}{2}(B(a,b) - B(2+a,b)) \\
&= \frac{1}{2}(\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} - \frac{\Gamma(2+a)\Gamma(b)}{\Gamma(a+b+1)}) \\
&= \frac{\Gamma(b)}{2}(\frac{\Gamma(a)}{\Gamma(a+b)} - \frac{\Gamma(2+a)}{\Gamma(2+a+b)})
\end{aligned} \tag{3}$$

Mode

The Mode occurs where $B(t,a,b)$ has a maximum, so

$$\begin{aligned}
\frac{\partial}{\partial t}B(t,a,b) &= 0 \\
\frac{\partial}{\partial t}t^{(a-1)}(1-t)^{b-1} &= 0 \\
(a-1)t^{a-2}(1-t)^{b-1} - (b-2)t^{a-1}(1-t)^{b-2} &= 0 \\
(a-1)(1-t) - (b-2)t &= 0 \\
a + t - at - 1 - bt + 2t &= 0 \\
(a+b-2)t - (a-1) &= 0 \\
\text{when } t^* &= \frac{a-1}{a+b-2}
\end{aligned} \tag{4}$$

can be evaluated by subsituting t^* into the definition of B_t

Variance

Variance is $E[x^2] - (E[x])^2$, again with Mathematic

$$\begin{aligned}
E[x^2] &= \int \theta^2 \mathbb{P}(\theta; a, b) d\theta \\
&= \frac{1}{B(a,b)} \int \theta^2 \theta^{a-1}(1-\theta)^{b-1} d\theta \\
&= \frac{1}{3}(t^3 B_t(t,a,b) - B_t(t, 3+a, b))
\end{aligned} \tag{5}$$

so

$$E[x^2] = \frac{1}{3}(B(a,b) - B(3+a,b)) \tag{6}$$

From Mean, we have

$$\begin{aligned}
E[x]^2 &= (\frac{1}{2}(B(a,b) - B(2+a,b)))^2 \\
&= \frac{1}{4}(B(a,b)^2 + B(2+a,b)^2 - 2B(a,b)B(2+a,b))
\end{aligned} \tag{7}$$

and Variance is

$$\begin{aligned}
Variance &= E[x^2] - E[x]^2 \qquad = \\
&= \frac{1}{3}(B(a,b) - B(3+a,b)) - \frac{1}{4}(B(a,b)^2 + B(2+a,b)^2 - 2B(a,b)B(2+a,b))
\end{aligned} \tag{8}$$

■

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

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

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).

To be in the exponential family, we must show:

$$f_X(x|\theta) = h(x)\exp[\eta(\theta)\cdot T(x) - A(\theta)] \tag{9}$$

$$\begin{aligned} h(x) &= 1 \\ \eta(\theta) &= [\ldots \log \mu_i \ldots] \\ T(x) &= [\ldots x_i \ldots] \\ A(x) &= 0 \end{aligned} \tag{10}$$

NLL of the above is linear in $x_i \log \mu_i$

■