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 θ.

Recall that the Beta function $\beta(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$, so we have that the mean of θ is:

$$\int_0^1 \left(\frac{1}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}\theta^{a-1}(1-\theta)^{b-1}\right)\theta \, d\theta =$$

$$\frac{1}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}\int_0^1 \theta^a(1-\theta)^{b-1}\, d\theta = \frac{\int_0^1 \theta^a(1-\theta)^{b-1}\, d\theta}{\int_0^1 t^{a-1}(1-t)^{b-1}\, dt}$$

Now let $N = \int_0^1 \theta^a(1-\theta)^{b-1}\, d\theta$, and $D = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt$, so the mean is $\frac{N}{D}$, and consider that by integration by parts:

$$N = \int_0^1 \theta^a(1-\theta)^{b-1}\, d\theta = \theta^a\left(-\frac{(1-\theta)^b}{b}\right)\Big|_0^1 + \int_0^1 \frac{(1-\theta)^b}{b}a\theta^{a-1}d\theta =$$

$$1\left(-\frac{0}{b}\right) - 0\left(\frac{1}{b}\right) + \frac{a}{b}\int_0^1 (1-\theta)^b\theta^{a-1}d\theta =$$

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

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

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

$$\frac{a}{b}(D-N) = N$$

$$\frac{aD}{b} = N\left(\frac{b+a}{b}\right)$$

$$\left(\frac{a}{b}\right)\left(\frac{b}{b+a}\right) = \frac{N}{D}$$

$$\frac{a}{b+a} = \text{mean}(\theta)$$

Next for the variance we do:

$$\int_0^1 \left(\frac{1}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}\theta^{a-1}(1-\theta)^{b-1}\right)\theta^2\, d\theta - \left(\frac{a}{b+a}\right)^2 =$$

$$\frac{\int_0^1 \theta^{a+1}(1-\theta)^{b-1}\, d\theta}{\int_0^1 t^{a-1}(1-t)^{b-1}dt} - \frac{a^2}{(a+b)(a+b)} =$$

Notice that the first term is:

$$\frac{\int_0^1 \theta^{a+1}(1-\theta)^{b-1}d\theta}{\int_0^1 t^{a-1}(1-t)^{b-1}dt} = \left(\frac{\int_0^1 \theta^{a+1}(1-\theta)^{b-1}d\theta}{\int_0^1 t^{a}(1-t)^{b-1}dt}\right)\left(\frac{\int_0^1 \theta^{a}(1-\theta)^{b-1}d\theta}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}\right)$$

Notice that from what we showed before about $\frac{N}{D}$ we can say that this term is $(\frac{a+1}{b+a+1})(\frac{a}{b+a}) = \frac{a^2+a}{(a+b+1)(a+b)}$. So we see that the variance is

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

$$\frac{a^3+a^2+a^2b+ab-a^3-a^2b-a^2}{(a+b+1)(a+b)^2} = \frac{ab}{(a+b+1)(a+b)^2}$$

This is the variance of θ. Lastly we show that the mode is the global maximum, so we find the critical points where

$$\frac{d}{d\theta}\left(\frac{\theta^{a-1}(1-\theta)^{b-1}}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}\right) = 0$$

$$\frac{\frac{d}{d\theta}(\theta^{a-1}(1-\theta)^{b-1})}{\int_0^1 t^{a-1}(1-t)^{b-1}dt} = 0$$

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

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

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

$$a-1+\theta(2-a-b) = 0$$

$$\theta = \frac{1-a}{2-a-b}$$

This is the mode of θ. We have now found the mean variance and mode of θ. ∎

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).

We want something in the exponential family $\mathbb{P}(\mathbf{x};\boldsymbol{\eta}) = b(\mathbf{x})\exp(\boldsymbol{\eta}^T T(\mathbf{x}) - a(\boldsymbol{\eta}))$, so consider the following:

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

$$\exp\log(\prod_{i=1}^{K} \mu_i^{x_i}) = \exp\sum_{i=1}^{K} \log(\mu_i)x_i = \exp(\log(\boldsymbol{\mu})^T\mathbf{x})$$

If we let $\boldsymbol{\eta} = \log(\boldsymbol{\mu})$, and $b(\mathbf{x}) = 1$, and $a(\boldsymbol{\eta}) = 0$, and $T(\mathbf{x}) = \mathbf{x}$, then we are able to write this in the form:

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

∎