Arjun Natarajan
Math189R SP19
Homework 3
Monday, February 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.

So, the Beta function is defined by

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

and that $\Gamma(\theta+1) = \theta\Gamma(\theta)$.

Then, the mean

$$\mu = \mathbb{E}[\theta] = \int_0^1 \theta P(\theta; a, b)\,d\theta = \int_0^1 \frac{\theta}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\,d\theta$$

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

$$= \frac{B(a+1,b)}{B(a,b)}$$

$$= \frac{\dfrac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}}{\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}}$$

$$= \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$\boxed{= \frac{a}{(a+b)}}.$$

■

So, the mode occurs when $\nabla P(\theta; a, b) = 0$, as this represents a local maximum.

Then,

$$\nabla_\theta P(\theta; a, b) = \frac{\partial}{\partial\theta} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}$$

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

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

$$(a-1)(1-\theta) = \theta(b-1)$$

$$a - a\theta - 1 + \theta = \theta b - \theta$$

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

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

Then, the variance of θ is defined by

$$Var[\theta] = \mathbb{E}[(\theta - \mathbb{E}[\theta])^2] = \mathbb{E}[\theta^2] - \mathbb{E}[\theta]^2.$$

Then, we have $\mathbb{E}[\theta]$ and $\mathbb{E}[\theta^2] = \int_0^1 \theta^2 \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1} d\theta$

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

$$= \frac{1}{B(a,b)} B(a+2, b)$$

$$= \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+1+1)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$= \frac{(a+1)\Gamma(a+1)}{(a+b+1)\Gamma(a+b+1)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)}$$

$$= \frac{a(a+1)\Gamma(a)}{(a+b+1)(a+b)\Gamma(a+b)} \frac{\Gamma(a+b)}{\Gamma(a)}$$

$$= \frac{a(a+1)}{(a+b)(a+b+1)}.$$

Then,

$$\text{var}[\theta] = \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2}$$

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

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

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

$$\boxed{= \frac{ab}{(a+b)^2(a+b+1)}}.$$

2 (**Murphy 9**) Show that the multinomial 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 multinomial logistic regression (softmax regression).

So, we can rewrite this distribution as

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

$$= \exp \sum_{i=1}^{k} \log(\mu_i^{x_i})$$

$$= \exp \sum_{i=1}^{k} x_i \log \mu_i \,.$$

So, since $\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} \mu_k = 1$, we can say that

$$x_k = 1 - \sum_{i=1}^{k-1} x_i$$

$$\mu_k = 1 - \sum_{i=1}^{k-1} \mu_i \,.$$

Then,

$$\text{Cat}(\mathbf{x}|\boldsymbol{\mu}) = \exp\left(\sum_{i=1}^{k-1} x_i \log(\mu_i) + x_k \log(\mu_k)\right)$$

$$= \exp\left(\sum_{i=1}^{k-1} x_i \log(\mu_i) + \left(1 - \sum_{i=1}^{k-1} x_i\right)\log(\mu_k)\right)$$

$$= \exp\left[\left(\sum_{i=1}^{k-1} x_i \log(\mu_i) - x_i \log(\mu_k)\right) + \log(\mu_k)\right]$$

1) $$= \exp\left[\left(\sum_{i=1}^{k-1} x_i \log\frac{\mu_i}{\mu_k}\right) + \log(\mu_k)\right].$$

■

Then, we can set

$$\vec{\boldsymbol{\eta}} = \begin{bmatrix} \log \frac{\mu_1}{\mu_k} \\ \vdots \\ \vdots \\ \log \frac{\mu_{k-1}}{\mu_k} \end{bmatrix},$$

and also recognize that

$$\mu_i = \mu_k e^{\log \frac{\mu_i}{\mu_k}} = \mu_k e^{\eta_i}.$$

Then, from

$$\mu_k = 1 - \sum_{i=1}^{k-1} \mu_i$$

$$= 1 - \sum_{i=1}^{k-1} \mu_k e^{\eta_i}$$

$$\mu_k = 1 - \mu_k \sum_{i=1}^{k-1} e^{\eta_i}$$

$$\mu_k \left(1 + \sum_{i=1}^{k-1} e^{\eta_i}\right) = 1$$

$$\mu_k = \frac{1}{1 + \sum_{i=1}^{k-1} e^{\eta_i}}$$

So,

$$\mu_i = \frac{e^{\eta_i}}{1 + \sum_{i=1}^{k-1} e^{\eta_i}}.$$

Plugging into 1),

$$\text{Cat}(\mathbf{x} | \boldsymbol{\mu}) = \exp\left(\boldsymbol{\eta}^T \mathbf{x} + \log(\mu_k)\right).$$

Thus this is in standard form of a Generalized Linear model, where

$b(\boldsymbol{\eta}) = 1,$

$T(\boldsymbol{\eta}) = \mathbf{x}$

$a(\boldsymbol{\eta}) = -\log(\mu_k).$

So $\text{Cat}(\mathbf{x}|\boldsymbol{\mu})$ is in the exponential family, and note that the softmax regression of $\boldsymbol{\eta}$ is equal to $\boldsymbol{\mu}$, implying the two are the same.