Math189R SU17
Homework 3
Thursday, May 17, 2017

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.

First, we find the mean of the Beta distribution:

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

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

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

$$= \frac{a}{a+b}$$

Then, we find the variance of the Beta distribution:

$$\mathbb{E}[\theta^2] = \int_0^1 \theta^2 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} d\theta$$

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

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

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

$$var[P\theta; a, b)] = \mathbb{E}[\theta^2] - \mathbb{E}[\theta]^2$$

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

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

At last, we want to find the mode of the distribution. Given the definition of pdf, the mode appears when the probability is maximum. Thus we try to find the gradient of the distribution and see what value of θ can make probability maximum.

$$\nabla p(\theta; a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}((a-1)\theta^{a-2}\theta^{b-1} - \theta^{a-1}(b-1)(1-\theta)^{b-2})$$

Set it to zero we get:

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

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

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

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

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

■

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).

Given this is a multinoulli distribution, we realize that the sum of probabilities $\Sigma_{i=1}^{k}\mu_i = 1$. And as we assume x_i to be indicator functions, then $\Sigma_{i=1}^{k}x_i = 1$ Thus we have:

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

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

Then we transform the distribution in the form of exponential family:

$$Cat(\mathbf{x}|\mu) = \Pi_{i=1}^{k}\mu_i^{x_i}$$

$$= exp[log(\Pi_{i=1}^{k}\mu_i^{x_i})]$$

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

$$= exp(\Sigma_{i=1}^{K}x^i log(\mu_i)$$

$$= exp(\Sigma_{i=1}^{k-1}x_i log(\mu_i) + x_k log(\mu_k))$$

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

$$= exp\Sigma_{i=1}^{k-1}x_i log(\frac{\mu_i}{\mu_k}) + log(\mu_k)$$

Here, b(y) = 1, $\eta = \begin{bmatrix} log(\frac{\mu_1}{\mu_k}) \\ log(\frac{\mu_2}{\mu_k}) \\ ... \\ log(\frac{\mu_{k-1}}{\mu_k}) \end{bmatrix}$, T(y) = x, a($\eta$) = $-log(\mu_k)$ We also can find that $\mu_i = \mu_k exp(\eta_i)$. Thus, $\mu_k = 1 - \Sigma_{i=1}^{k-1}\mu_k exp(\eta_i) = \frac{1}{1+\Sigma^{k-1}i=1exp(\eta_i)}$. As a result, we can rewrite $a(\eta) = log(1 + \Sigma^{k-1}i = 1exp(\eta_i))$ And this exponential form is the same as softmax regression.

■