Julianne Lin

Math189R SU17
Homework 3
Wednesday, May 24, 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.

By def $B(a,b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\,d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma(x+1) = x\Gamma(x)$. checked soln for this.

a) Find mean:

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

$B(a,b)$ idpt of θ & combine θ & θ^{a-1}

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

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

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

b) Find variance:

By def $\text{Var}[\theta] = \mathbb{E}[(\theta-\mathbb{E}[\theta])^2] = \mathbb{E}[\theta^2] - \mathbb{E}[\theta]^2$ (We have $\mathbb{E}[\theta]^2$ from above but not $\mathbb{E}[\theta^2]$)

$$\text{so, } \mathbb{E}[\theta^2] = \int_0^1 \theta^2\left(\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\right)d\theta$$

(if $\Gamma(x+1) = x\Gamma(x)$ then $\Gamma(x+2) = (x+1)\Gamma(x+1) = x(x+1)\Gamma(x)$)

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

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

$$\mathbb{E}[\theta]^2 = \left(\frac{a}{a+b}\right)^2$$

■

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

(cont'd) $\dfrac{-a^2(a+b+1) + a(a+1)(a+b)}{(a+b)^2(a+b+1)} = \dfrac{-(\cancel{a^3}+a^2b+\cancel{a^2}) + (\cancel{a^3}+\cancel{a^2b}+\cancel{a^2}+ab)}{(a+b)^2(a+b+1)}$

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

c) Find mode : Calc when $\nabla_\theta P(\theta; a,b) = 0$ on $[0,1]$.

$\Rightarrow$ disregard constant $\dfrac{1}{B(a,b)}$ term :

$$\nabla_\theta P(\theta; a,b) = \nabla_\theta \left[\theta^{a-1}(1-\theta)^{b-1}\right] = 0$$

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

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

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

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

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

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

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

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

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

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

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).

Exponential family : $\mathbb{P}(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta))$

① Rewrite distribution w/ exp & log!

b/c $\log(ab) = \log(a) + \log(b)$

$$\text{Cat}(x|\mu) = \prod_{i=1}^{K} \mu_i^{x_i} = \exp\left[\log\left(\prod_{i=1}^{K} \mu_i^{x_i}\right)\right] = \exp\left(\sum_{i=1}^{K} \log(\mu_i^{x_i})\right)$$

b/c $\log(a^b) = b\log a$

$$= \exp\left(\sum_{i=1}^{K} x_i \log(\mu_i)\right)$$

(★ By def , $\sum_{i=1}^{K} \mu_i = 1$ & $\sum_{i=1}^{K} x_i = 1$. So, $\mu_K = 1 - \sum_{i=1}^{K-1} \mu_i$ & $x_k = 1 - \sum_{i=1}^{K} x_i$.)

Thus, from above :

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

honestly, following similar-ish ex in class

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

(checked sol'n for these cuz tbh, it's late)

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

$$= \underbrace{1}_{b(y)} \exp\left[\sum_{i=1}^{K-1} \underbrace{x_i}_{T(y)} \log\left(\frac{\mu_i}{\mu_K}\right) + \log(\mu_k)\right]$$

So, according to the exp family form :

$$\eta = \begin{bmatrix} \log\left(\frac{\mu_1}{\mu_K}\right) \\ \vdots \\ \log\left(\frac{\mu_{K-1}}{\mu_K}\right) \end{bmatrix}$$

■

Then $\mu_i = \mu_k e^{\eta_i}$ (b/c $\mu_k e^{\eta_i} = \mu_k e^{\left(\log \frac{\mu_i}{\mu_k}\right)} = \cancel{\mu_k} \left(\frac{\mu_i}{\cancel{\mu_k}}\right) = \mu_i$)

Substitute

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

So $\mu_i = \mu_k e^{\eta_i} = \dfrac{1}{1 + \sum_{i=1}^{k-1} e^{\eta_i}} e^{\eta_i}$ Softmax form!

Thus

$$b(\eta) = 1$$
$$T(\vec{x}) = \vec{x}$$
$$a(\vec{\eta}) = -\log(\mu_k) = \log\left(1 - \sum_{i=1}^{k-1} e^{\eta_i}\right)$$

exponential family form

$\blacksquare$