Math189R SP20
Homework 3
Monday 2/17, 2020

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.

Mean:

For a continuous rv, $\mathbb{E}[X] = \int_X x p(x) dx$. In addition, Beta is in $[0,1]$.

Thus, $\mathbb{E}[\theta] = \int_0^1 \theta p(\theta) d\theta$

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

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

Note that $B(x,y) = \int_0^1 \theta^{x-1}(1-\theta)^{y-1} d\theta$

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

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

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

since $\Gamma(x+1) = x\Gamma(x)$

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

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

(Continue on next page)

■

variance

$$\sigma^2 = \mathbb{E}[\theta^2] - \mu^2$$

$$= \int_0^1 \theta^2 p(\theta)\, d\theta - \mu^2$$

substitute $B(a+2, b)$

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

$$= \frac{B(a+2,b)}{B(a,b)} - \mu^2$$

$\Gamma(a+2) = (a+1)a\Gamma(a)$

$$= \frac{(a+1)a\cancel{\Gamma(a)}\cancel{\Gamma(b)}}{\Gamma(a+2+b)} \cdot \frac{\Gamma(a+b)}{\cancel{\Gamma(a)}\cancel{\Gamma(b)}} - \mu^2$$

$\Gamma(a+b+2) = (a+b+1)(a+b)\Gamma(a+b)$

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

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

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

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

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

mode

Since at the mode, the probability is at maximum, we can calculate it by setting the derivative to 0.

$$0 = \frac{dp(\theta)}{d\theta}$$

$$0 = \frac{1}{B(a,b)}\left[\frac{d}{d\theta} \cdot (\theta^{a-1}(1-\theta)^{b-1})\right]$$

$$0 = \frac{1}{B(a,b)}\left[-\theta^{a-1}(b-1)(1-\theta)^{b-2} + (a-1)\theta^{a-2}(1-\theta)^{b-1}\right]$$

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

$$\theta = \frac{a-1}{b-1}(1-\theta) \Rightarrow \theta\left(1+\frac{a-1}{b-1}\right) = \frac{a-1}{b-1} \Rightarrow \theta = \boxed{\frac{a-1}{a+b-2}}$$

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).

Show in the exponential family:

$$Cat(x|u) = \exp\left(\log \prod_{i=1}^{k} u_i^{x_i}\right)$$

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

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

In order to have $a(\eta)$, we pull out the k^{th} term

$$Cat(x|u) = \exp\left(\sum_{i=1}^{k-1} x_i \log u_i + x_k \log u_k\right) \quad \text{since } \sum_{i=1}^{k} x_i = 1,$$

$$= \exp\left[\sum_{i=1}^{k-1} x_i \log u_i + \left(1 - \sum_{i=1}^{k-1} x_i\right)\log u_k\right] \quad x_k = 1 - \sum_{i=1}^{k-1} x_i$$

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

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

$$= \exp\left(\sum_{i=1}^{k-1} x_i \log\left(\frac{u_i}{u_k}\right) + \log u_k\right)$$

$$= \exp\left(\begin{bmatrix} \log\left(\frac{u_1}{u_k}\right) \\ \log\left(\frac{u_2}{u_k}\right) \\ \vdots \\ \log\left(\frac{u_{k-1}}{u_k}\right) \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \end{bmatrix} + \log u_k\right)$$

■

(continue on next page)

Since the exponential family is in the form

$p(x;\eta) = b(x)\exp(\eta^T T(x) - a(\eta))$, if we let

$$\vec{\eta} = \begin{bmatrix} \log(\frac{\mu_1}{\mu_k}) \\ \vdots \\ \log(\frac{\mu_{k-1}}{\mu_k}) \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{k-1} \end{bmatrix}, \quad a(\vec{\eta}) = -\log \mu_k, \quad b(y) = 1,$$

then $Cat(x;\mu)$ is in the exponential family.

Show GLM = softmax regression.

write μ_i in terms of μ_k:

$$\eta_i = \log\left(\frac{\mu_i}{\mu_k}\right)$$

$$\mu_i = e^{\eta_i} \cdot \mu_k \qquad (1)$$

Find μ_k in terms of η_i:

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

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

$$\mu_k = \frac{1}{1 + \sum_{i=1}^{k-1} e^{\eta_i}} \qquad (2)$$

write μ_i in terms of η_i:

$$\mu_i = e^{\eta_i} \mu_k \qquad \text{(from (1))}$$

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

$$= \text{softmax function } S(\eta)_i \qquad \blacksquare$$