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.

distribution of θ over the range (a, b)?

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.

θ is a random continuous variable

$\theta \in [0, 1]$ as a given constraint of Beta functions

$$E(\theta) = \int_0^1 \theta P(\theta; a, b)\, d\theta$$

$$\mathbb{P}(\theta; a+1, b) = \frac{\Gamma(a+b+1)}{\Gamma(\quad - \quad \Gamma(b}\theta^a(1-\theta)^{b-1}$$

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

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

$\int_0 P(\theta; a+1, b) = 1$ because it's a probability density function of θ when $\theta \in [0, 1]$

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

$\Gamma(n) = (n-1)!$ for discrete parameter

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

mode of continuous random variable : pdf attains max value

$$P(\theta; a, b) = \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}_{\text{normalizing constant} \therefore \text{ignore}}\theta^{a-1}(1-\theta)^{b-1}$$

$\Rightarrow$ maximize $f = \theta^{a-1}(1-\theta)^{b-1}$

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

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

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

■

$$\theta = 0, 1, \frac{1-a}{2-a-b}$$

c) $\mathrm{Var}(x) = E[x^2] - \mu^2 = \int x^2 f(x)\,dx - \mu^2$ — definition of variance

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

$$= \frac{\Gamma(a+b)}{\Gamma(a)\cancel{\Gamma(b)}} \cdot \frac{\Gamma(a+2)\cancel{\Gamma(b)}}{\Gamma(a+b+2)} \underbrace{\int \frac{\Gamma(a+b+2)}{\Gamma(a+2)\Gamma(b)} \theta^{a+1}(1-\theta)^{b-1}\,d\theta}_{=1} - \frac{a^2}{(a+1)^2}$$

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

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).

form of exponential probability functions:

$$P(x;\mu) = b(x)\exp(\mu^T T(x) - a(\mu))$$

First transform $Cat(x|\mu)$ to exponential form as follows

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

$$= \exp\left[\sum_{i=1}^{k}\log\mu_i^{x_i}\right] = \exp\left[\sum_{i=1}^{k}x_i\log\mu_i\right]$$

$$= \exp\left[\sum_{i=1}^{k-1}x_i\log\mu_i + x_k\log(\mu_k)\right]$$ — checked answer key here

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

let $$\eta = \begin{bmatrix}\log\left(\frac{\mu_1}{\mu_k}\right)\\ \log\left(\frac{\mu_{k-1}}{\mu_k}\right)\end{bmatrix}$$ — checked answer key

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

now μ = soft max of η we defined earlier

■