Kai Kaneshiro

I worked w/ Colin.

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.

$$\mu = E[\theta] = \int_{-\infty}^{\infty} \theta \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} d\theta$$

$$\mu = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \underbrace{\int_{-\infty}^{\infty} \theta^{a}(1-\theta)^{b-1} d\theta}_{B(a+1,b)} \rightarrow B(a+1,b) = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}$$

$$\mu = \frac{\Gamma(a+b)}{\Gamma(a)\cancel{\Gamma(b)}} \cdot \frac{\Gamma(a+1)\cancel{\Gamma(b)}}{\Gamma(a+b+1)} \quad \Big| \quad \Gamma(x+1) = x\Gamma(x)$$

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

mode: max value of $\mathbb{P}(\theta | a, b)$; do ∇_θ + set $=0$: $\longrightarrow$

since the mode is the most common value, it has the highest probability

■

mode: $\vec{\nabla}_\theta P(\theta \mid a; b) = 0.$

$$\vec{\nabla}_\theta \left(\theta^{a-1} (1-\theta)^{b-1} \right) = 0$$

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

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

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

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

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

$$\theta(a+b-2) = a-1 \rightarrow \boxed{\text{mode} \quad \theta^* = \frac{a-1}{a+b-2}}$$

variance: $\mathrm{var}(\theta) = E[\theta^2] - E[\theta]^2$, $E[\theta] = \frac{a}{a+b} \rightarrow E[\theta]^2 = \frac{a^2}{(a+b)^2}$

$$E[\theta^2] = \int_{-\infty}^{\infty} \theta^2 \frac{1}{B(a,b)} (\theta^{a-1})(1-\theta)^{b-1} \, d\theta = \frac{1}{B(a;b)} \underbrace{\int_{-\infty}^{\infty} \theta^{a+1}(1-\theta)^{b-1} d\theta}_{B(a+2,\,b)}$$

$$E[\theta^2] = \frac{\Gamma(a+b)}{\Gamma(a)\cancel{\Gamma(b)}} \cdot \frac{\Gamma(a+2)\cdot\cancel{\Gamma(b)}}{\Gamma(a+b+2)} \quad \Big| \quad \Gamma(x+1) = x\,\Gamma(x)$$

$$E[\theta^2] = \frac{\Gamma(a+b)}{\Gamma(a)} \cdot \frac{(a+1)\,\Gamma(a+1)}{(a+b+1)\,\Gamma(a+b+1)} = \frac{\cancel{\Gamma(a+b)}}{\cancel{\Gamma(a)}} \cdot \frac{(a+1)a\,\cancel{\Gamma(a)}}{(a+b+1)(a+b)\,\cancel{\Gamma(a+b)}} = \frac{a^2+a}{(a+b+1)(a+b)}$$

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

$$= \frac{\cancel{a^3} + \cancel{ba^2} + \cancel{a^2} + ab - \cancel{a^3} - \cancel{ba^2} - \cancel{a^2}}{(a+b)^2(a+b+1)} \Rightarrow \boxed{\mathrm{var}[\theta] = \frac{ab}{(a+b)^2(a+b+1)}}$$

2 (**Murphy 9**) Show that the multinomial distribution

$$\mathrm{Cat}(\mathbf{x}|\boldsymbol{\mu}) = \prod_{i=1}^{K} \mu_i^{x_i}$$

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).

$$\mathrm{Cat}(x|\mu) = \exp\left(\log\left(\prod_{i=1}^{K} \mu_i^{x_i}\right)\right) = \exp\left(\log(\mu_1^{x_1}) + \dots + \log(\mu_K^{x_K})\right)$$

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

recall that $\sum_{i=1}^{K} \mu_i = 1$ + $\sum_{i=1}^{K} x_i = 1$

Thus we can rewrite as: and substitute

$$\mu_K + \sum_{i=1}^{K-1} \mu_i = 1 \rightarrow 1 - \sum_{i=1}^{K-1} \mu_i = \mu_K$$

$$x_K + \sum_{i=1}^{K-1} x_i = 1 \rightarrow 1 - \sum_{i=1}^{K-1} x_i = x_K$$

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

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

distribute $\log \mu_K$

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

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

exponential equation:

$$P(x|\eta) = b(x) \exp(\eta^T T(x) - a(\eta))$$

So in our case above:

$b(x) = 1$

$T(x) = x$

$a(\eta) = -\log(\mu_K) = -\left(1 - \sum_{i=1}^{K-1} \mu_i\right) = \sum_{i=1}^{K-1} \mu_i - 1$

$\eta_i = \log \frac{\mu_i}{\mu_K}$, $i = 1 : \dots K-1$

Solve for μ_i: $e^{\eta_i} = \frac{\mu_i}{\mu_K} \rightarrow \mu_i = \mu_K e^{\eta_i} = \left(1 - \sum_{i=1}^{K-1} \mu_i\right) e^{\eta_i}$

$$\mu_i = e^{\eta_i} - \sum_{i=1}^{K-1} \mu_i e^{\eta_i} \rightarrow \mu_i + \sum_{i=1}^{K-1} \mu_i e^{\eta_i} = e^{\eta_i} \rightarrow \mu_i \left(1 + \sum_{i=1}^{K-1} e^{\eta_i}\right) = e^{\eta_i}$$

$$\boxed{\mu_i = \frac{e^{\eta_i}}{1 + \sum_{i=1}^{K-1} e^{\eta_i}}}$$ ✓

matches softmax regression ✓