Bennett Mountain
Math189R SP19
Homework 3
Monday, Feb 18, 2019

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 $\theta \sim \mathrm{Beta}(a,b)$ 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.

We know $B(a,b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\,d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, thus $\Gamma(x+1) = x\Gamma(x)$.

The expectation, or mean is

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

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

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

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

$$\boxed{= \frac{a}{a+b}}$$ is our mean.

For the variance, we know $\mathrm{var}(\theta) = E(\theta - E(\theta))^2 = E(\theta^2) - E(\theta)^2$

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

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

$$= \frac{a(a+1)}{(a+b)(a+b+1)}$$; Since we now have $E(\theta^2)$, we can find our variance.

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

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

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

■

For the mode, we know that occurs when $\nabla_\theta \mathbb{P}(\theta; a, b) = 0$ on $[0,1]$.

$\nabla_\theta \mathbb{P}(\theta; a, b) = \nabla_\theta \left(\frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} \right) = 0$ $\quad$ $\frac{1}{B(a,b)}$ has no θ in it (it's a constant term), so we can get rid of it

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

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

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

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

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

2. (**Murphy 9**) Show that the multinoulli distribution

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

is in the exponential family and show that the generalized linear model corresponding to this distribution is the same as multinoulli logistic regression (softmax regression).

We know the exponential family is $\mathbb{P}(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta))$

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

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

since $\sum_{i=1}^{K} \mu_i = 1$ and $\sum_{i=1}^{K} x_i = 1$, we only need consider the first $K-1$ terms, like

$\mu_K = 1 - \sum_{i=1}^{K-1} \mu_i$ and $x_K = 1 - \sum_{i=1}^{K-1} x_i$

Thus

$$\mathrm{Cat}(x|\mu) = \exp(\sum_{i=1}^{K} x_i \log(\mu_i)) = \exp(\sum_{i=1}^{K-1} x_i \log(\mu_i) + x_K \log(\mu_K))$$

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

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

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

so, we can call $\eta = \begin{bmatrix} \log(\frac{\mu_1}{\mu_K}) \\ \vdots \\ \log(\frac{\mu_{K-1}}{\mu_K}) \end{bmatrix}$, which means $\mu_i = \mu_K e^{\eta_i}$

so $\mu_K = 1 - \sum_{i=1}^{K-1} \mu_i = 1 - \sum_{i=1}^{K-1} \mu_K e^{\eta_i}$

$$= 1 - \mu_K \sum_{i=1}^{K-1} e^{\eta_i}$$

$$= \frac{1}{1 + \sum_{i=1}^{K-1} e^{\eta_i}}, \text{ so } \mu_i = \mu_K e^{\eta_i} = \frac{e^{\eta_i}}{1 + \sum_{i=1}^{K-1} e^{\eta_i}}$$

■

In the exponential family's form, $cat(x|\lambda) = \exp(\eta^T x - a(\eta))$.

Thus, $b(\eta) = 1$ and $T(x) = x$, and $a(\eta) = -\log(\lambda_n) = \log(1 + \sum_{i=1}^{n-1} e^{\eta_i})$

meaning $cat(x|\lambda)$ is in the exponential family.