Eli Weissler

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.

Througout the problem I will be taking advantage of the recursive definition of the Gamma function:

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$$

$\boxed{\text{Mean}}$ To find the mean $\mu = E[P(\theta : a, b)]$, we take

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

$$= \frac{1}{B(a,b)}\int_0^1 \theta^a(1-\theta)^{b-1} d\theta \leftarrow$$ We know that $\frac{1}{B(x,y)}\int_0^1 \theta^{x-1}(1-\theta)^{y-1} d\theta = 1$ because it's normalized, Hence $\int_0^1 \theta^a(1-\theta)^{b-1} d\theta = B(a+1,b)$

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

$\boxed{\text{Mode}}$ If we assume $a, b > 1$, then we have a concave down distribution with a single peak, so we just need to find when the derivatie is 0.

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

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

$$\rightarrow \theta = \frac{a-1}{b-1}\cdot\frac{b-1}{a+b-2} = \boxed{\frac{a-1}{a+b-2}}$$

$\boxed{\text{variance}}$

$$Var = E[\theta^2] - E[\theta]^2$$ so we need to find $E[\theta^2]$

$$E[\theta^2] = \int_0^1 d\theta\, \theta^2 P(\theta; a,b) = \frac{1}{B(a,b)} \int_0^1 d\theta \cdot \theta^{a+1} (1-\theta)^{b-1} = \frac{B(a+2,b)}{B(a,b)}$$

$$= \frac{\Gamma(a+2)\cancel{\Gamma(b)}\Gamma(a+b)}{\Gamma(a+b+2)\Gamma(a)\cancel{\Gamma(b)}} = \frac{(a+1)a}{(a+b+1)(a+b)}$$

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

So

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

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

$P(y:\varphi) = b(y)\exp(\varphi^T T(y) - a(\varphi))$

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).

To show that it is in the exponential family, we

① Put in one equation

K choices for categories

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

x_i ← 1 if x_i is one; μ_i ← prob b/w 0 & 1

$$\vec{x} = \begin{pmatrix} I(x=1) \\ I(x=2) \\ \vdots \end{pmatrix}$$

Because $\sum \mu = 1$

$$\rightarrow \mathrm{Cat}(\vec{x}|\vec{\mu}) = \prod_{i=1}^{K} \mu_i^{x_i} = \prod_{i=1}^{K-1} \mu_i^{x_i} \cdot \left(1 - \sum_{j=1}^{K-1} \mu_j\right)^{1-\sum_{j=1}^{K-1} x_j}$$

← Because only one has stuff

② Force it into exponential family

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

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

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

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

Now if we let $\vec{\eta} = \begin{pmatrix} \log\frac{\mu_1}{\mu_K} \\ \vdots \\ \log\frac{\mu_i}{\mu_K} \\ \vdots \end{pmatrix}$, then $\mu_i = \mu_K e^{\eta_i}$

* Had to look at key to define η

& $\mu_K = 1 - \sum \mu_K e^{\eta_i}$

$\mu_K(1 + \sum e^{\eta_i}) = 1 \rightarrow \mu_K = \frac{1}{1+\sum e^{\eta_i}}$ & $\mu_i = \mu_K e^{\eta_i} = \frac{e^{\eta_i}}{1+\sum e^{\eta_i}}$

Thus we are left with

$$\mathrm{Cat}(\vec{x}|\vec{\mu}) = \exp\left(\eta^T x + \log\frac{1}{1+\sum_{i=1}^{K-1} e^{\eta_i}}\right)$$

So it is exponential family, with $b(\vec{x}) = 1$, $\vec{\varphi} = \vec{\eta}$ ∎

$T(\vec{x}) = \vec{x}$, $a(\vec{\varphi}) = \frac{1}{1+\sum e^{\varphi_i}}$

We also see that $\vec{\mu} = \begin{pmatrix} \frac{e^{\eta_i}}{1+\sum e^{\eta_i}} \\ \vdots \end{pmatrix}$ is the same as the softmax that we went over in class.