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

10

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

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.

a) mean $= \mathbb{E}[x] = \int_{\mathbb{R}} x f(x)\,dx$

So $\mathbb{E}[\theta] = \int_{\mathbb{R}} \frac{1}{B(a,b)} \theta^{a} (1-\theta)^{b-1}\, d\theta$.

We know that

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

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

If the above is true, then

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

Plugging this into $\mathbb{E}[\theta]$, we get

$$\int_{\mathbb{R}} \frac{1}{B(a,b)} \cdot B(a+1,b)\,d\theta$$

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

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

$*$ Note: $\Gamma(a) = (a-1)!$

b) The mode of a continuous probability density function occurs when the density function is at its maximum value. So, we need to take the derivative of $\mathbb{P}$ and set it equal to zero.

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

$$0 = \theta^{a-1}\theta^{b-1}\left[(b-1)(1-\theta)^{-1} - (a-1)\theta^{-1}\right]$$

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

$$0 = \theta(b-1) + (a-1)(1-\theta)$$

$$0 = b\theta - \theta - a + a\theta + 1 - \theta$$

$$0 = \theta(b-1+a-1) - a + 1$$

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

■

c). We know that $\mathrm{var}[x] = \mathbb{E}[(x-\mu)^2] = \mathbb{E}[x^2] - (\mathbb{E}[x])^2$

$$\mathrm{var}[x] = \int_{\mathbb{R}} \theta^2 \frac{1}{B(a,b)} \cdot \theta^{a-1} (1-\theta)^{b-1} d\theta - \left(\frac{a}{a+b}\right)^2$$

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

From part a, we know that this is equivalent to

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

$$= \frac{(a+b-1)!}{(a-1)!(b-1)!} \cdot \frac{(a+1)!(b-1)!}{(a+b+1)!} - \left(\frac{a}{a+b}\right)^2$$

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

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

$$= \frac{a}{(a+b+1)(a+b)^2}\left[\cancel{a^2} + \cancel{ab} + \cancel{a} + b - \cancel{a^2} - \cancel{ab} - \cancel{a}\right]$$

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

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).

First, we will show $\text{Cat}(x|\mu) = \prod_{i=1}^{K} \mu_i^{x_i}$ is in the exponential family:

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

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

We know that $\sum_{i=1}^{K} x_i = 1$ and $\sum_{i=1}^{K} \mu_i = 1$ so, we don't need to include μ_K or x_K because they can be solved for in terms of $\mu_1, \mu_2, \ldots \mu_{K+1}$ and $x_1, x_2, \ldots x_{K+1}$ because

$\mu_K = 1 - (\mu_1 + \mu_2 + \mu_3 + \ldots + \mu_K)$

$x_K = 1 - (x_1 + x_2 + x_3 + \ldots + x_K)$

So now, we have

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

$$= \exp\left[(x_1 \log \mu_1 + x_2 \log \mu_2 + \ldots + x_{K-1} \log \mu_{K-1}) + (1 - x_{K-1} - \ldots - x_2 - x_1)(\log \mu_K)\right]$$

$$= \exp\left[\log \mu_K + x_1 \log \frac{\mu_1}{\mu_K} + x_2 \log \frac{\mu_2}{\mu_K} + \ldots + x_{K-1} \log \frac{\mu_{K-1}}{\mu_K}\right]$$

so,

$$= \exp\left[\eta^T T(y) - a(\eta)\right]$$

where $\eta = \begin{bmatrix} \log \frac{\mu_1}{\mu_K} \\ \vdots \\ \log \frac{\mu_{K-1}}{\mu_K} \end{bmatrix}$ $\quad T(X) = \begin{bmatrix} x_1 \\ \vdots \\ x_{K-1} \end{bmatrix}$ $\quad a(\eta) = -\log(\mu_K)$

$b(X) = 1$

Now, we want to find μ_i.

First, we will note that $\theta^T x = \frac{\mu_i}{\mu_k}$

$$① \quad \mu_i = e^{\theta^T x} \mu_k$$

Now, we have $\mu_k = 1 - \sum_{i=1}^{k-1} \mu_i$ ← Looked at solution for this step

$$= 1 - \sum_{i=1}^{k-1} e^{\theta^T x_i} \mu_k$$

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

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

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

Plugging this into ①, we get

$$\boxed{\mu_i = \frac{e^{\eta_i}}{1 + \sum_{l=1}^{k-1} e^{\eta_i} \mu_k} \quad \text{as desired}}$$