Solution
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.

Recall the definition of the Beta Function and the property of Gamma function:

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

$$\Gamma(x+1) = x\,\Gamma(x)$$

Now we can compute the mean of θ, we have

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

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

$$= \frac{B(a+1,b)}{B(a,b)}$$

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

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

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

We have obtained the mean as we desired. we know

$$Var[\theta] = \mathbb{E}[(\theta - \mathbb{E}[\theta])^2] = \mathbb{E}[\theta^2] - \mathbb{E}[\theta]^2$$

To compute variance, we need to first compute $\mathbb{E}[\theta^2]$ using the same method as above, and we get

$$\begin{aligned}
\mathbb{E}[\theta^2] &= \int_0^1 \theta^2 \left(\frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1} \right) d\theta \\
&= \frac{1}{B(a,b)} \int_0^1 \theta^{a+1}(1-\theta)^{b-1}\, d\theta = \frac{B(a+2,b)}{B(a,b)} \\
&= \left[\frac{\Gamma(a+2)\,\Gamma(b)}{\Gamma(a+b+2)} \right] \left[\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \right] \\
&= \left[\frac{a(a+1)\,\Gamma(a)\,\Gamma(b)}{(a+b)(a+b+1)\Gamma(a+b)} \right] \left[\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \right] \\
&= \frac{a(a+1)}{(a+b)(a+b+1)}
\end{aligned}$$

Thus we can compute the variance, and we have

$$\begin{aligned}
Var[\theta] &= \mathbb{E}[\theta^2] - \mathbb{E}[\theta]^2 \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2} \\
&= \frac{a(a+1)(a+b) - a^2(a+b+1)}{(a+b)^2(a+b+1)} \\
&= \frac{a^3 + a^2 b + a^2 + ab - a^3 - a^2 b - a^2}{(a+b)^2(a+b+1)} \\
&= \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned}$$

To compute the mode, we want to find when $\nabla_\theta \mathbb{P}(\theta; a, b) = 0$ on the interval $[0, 1]$. Because a constant term won't change the optimizing value, we can work with the unnormalized distribution (ignoring the $1/B(a,b)$ term). This gives

$$\begin{aligned}
\nabla_\theta \mathbb{P}(\theta; a, b) &= \nabla_\theta \left[\theta^{a-1}(1-\theta)^{b-1} \right] = 0 \\
&= (a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2} = 0
\end{aligned}$$

where we can now compute the mode

$$\begin{aligned}(a-1)\theta^{a-2}(1-\theta)^{b-1} &= (b-1)\theta^{a-1}(1-\theta)^{b-2}\\ (a-1)(1-\theta) &= (b-1)\theta\\ (a+b-2)\theta &= a-1\\ \theta^{\star} &= \frac{a-1}{a+b-2}\end{aligned}$$

We have obtained mean, variance and mode of θ as desired.

■

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).

Recall that exponential family is all in the form:

$$\mathbb{P}(\mathbf{y};\boldsymbol{\eta}) = b(\mathbf{y}) \exp\left(\boldsymbol{\eta}^\top T(\mathbf{y}) - a(\boldsymbol{\eta})\right)$$

Also recall the property of logarithm:

$$\log(ab) = \log(a) + \log(b)$$
$$\log(a^b) = b\log(a)$$

To show that the multinomial distribution is in the exponential family, we simply need to rewrite the distribution to include an exponential and logarithm:

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

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

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

Now observe that since $\sum_{i=1}^{K} \mu_i = 1$ and $\sum_{i=1}^{K} x_i = 1$, we then need only specify the first $K-1$ of these terms, since the term x_K and μ_K will be automatically be deterined at the end:

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

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

We can therefore split up our summation above into

$$\begin{aligned}
\text{Cat}(\mathbf{x}|\boldsymbol{\mu}) &= \exp\left(\sum_{i=1}^{K} x_i \log(\mu_i)\right) = \exp\left(\sum_{i=1}^{K-1} x_i \log(\mu_i) + x_K \log(\mu_K)\right) \\
&= \exp\left[\sum_{i=1}^{K-1} x_i \log(\mu_i) + \left(1 - \sum_{i=1}^{K-1} x_i\right) \log(\mu_K)\right] \\
&= \exp\left[\sum_{i=1}^{K-1} x_i \left(\log(\mu_i) - \log(\mu_K)\right) + \log(\mu_K)\right] \\
&= \exp\left[\sum_{i=1}^{K-1} x_i \log\left(\frac{\mu_i}{\mu_K}\right) + \log(\mu_K)\right]
\end{aligned}$$

Therefore, let the vector $\boldsymbol{\eta}$ be

$$\boldsymbol{\eta} = \begin{bmatrix} \log\left(\frac{\mu_1}{\mu_K}\right) \\ \cdots \\ \log\left(\frac{\mu_{K-1}}{\mu_K}\right) \end{bmatrix}$$

From which we can see that $\mu_i = \mu_K e^{\boldsymbol{\eta}_i}$, and if we make the substitution, we have

$$\begin{aligned}
\mu_K &= 1 - \sum_{i=1}^{K-1} \mu_i = 1 - \sum_{i=1}^{K-1} \mu_K e^{\boldsymbol{\eta}_i} \\
&= 1 - \mu_K \sum_{i=1}^{K-1} e^{\boldsymbol{\eta}_i} \\
&= \frac{1}{1 + \sum_{i=1}^{K-1} e^{\boldsymbol{\eta}_i}}
\end{aligned}$$

Hence, we also get

$$\mu_i = \mu_K e^{\boldsymbol{\eta}_i} = \frac{e^{\boldsymbol{\eta}_i}}{1 + \sum_{i=1}^{K-1} e^{\boldsymbol{\eta}_i}}$$

Writing the distribution in the form of exponential family as $\text{Cat}(\mathbf{x}|\boldsymbol{\mu}) = \exp(\boldsymbol{\eta}^\top \mathbf{x} - a(\boldsymbol{\eta}))$ we conclude that

$$
\begin{aligned}
b(\boldsymbol{\eta}) &= 1 \\
T(\mathbf{x}) &= \mathbf{x} \\
a(\boldsymbol{\eta}) &= -\log(\mu_K) = \log\left(1 + \sum_{i=1}^{K-1} e^{\boldsymbol{\eta}_i}\right)
\end{aligned}
$$

From which we can conclude that the distribution $\text{Cat}(\mathbf{x}|\boldsymbol{\mu})$ is in the exponential family. And $\boldsymbol{\mu} = S(\boldsymbol{\eta})$, where $S(\boldsymbol{\eta})$ is the softmax function, which implies the generalized linear model of this distribution is the same as softmax regression.

■