Haocheng Ye
Math189R SU18
Homework 3
Monday, February 18, 2018

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.

To calculate the mean:

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

Notice that

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

Plug this back to our calculation of the mean we have

$$\mathbb{E}[\theta] = \frac{B(a+1,b)}{B(a,b)} = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

Note that

$$\Gamma(z+1) = \int_0^\infty x^z \exp(-x)dx = -x^z\exp(-x)\Big|_0^\infty + \int_0^\infty zx^{z-1}\exp(-x)dx$$
$$= z\int_0^\infty x^{z-1}\exp(-x)dx = z\Gamma(z)$$

Using this result, we have

$$\mathbb{E}[\theta] = \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{a}{a+b}$$

Next we have to compute the variance. We can first calculate $\mathbb{E}[\theta^2]$ Similar to previous calculation, we have

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

Thus the variance is

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

Finally for the mode of a Beta distributed random variable should be the most likely value of the distribution, i.e. the θ corresponding to the maximum value of the PDF function. We can compute this by finding the $\nabla_\theta \mathbb{P}(\theta; a, b) = 0$. Note that $\frac{1}{B(a,b)}$ is just a normalizing constant, thus it does not contain θ. We have

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

Hence, the mode is

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

■

2 (**Murphy 9**) Show that the multinoulli 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 multinoulli logistic regression (softmax regression).

We can force this distribution to be in the exponential family by exp and log. We have,

$$\begin{aligned}
\mathrm{Cat}(\mathbf{x}|\boldsymbol{\mu}) &= \exp\left[\log\left(\prod_{i=1}^{K} \mu_i^{x_i}\right)\right] \\
&= \exp\left(\sum_{i=1}^{K} \log(\mu_i^{x_i})\right) \\
&= \exp\left(\sum_{i=1}^{K} x_i \log(\mu_i)\right)
\end{aligned}$$

Since this is a multinoulli distribution, we have $\sum_{i=1}^{K} \mu_i = 1$ and $\sum_{i=1}^{K} x_i = 1$, therefore we can write,

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

Hence, the vector $\eta^\top$ is

$$\eta^\top = \left[\log(\frac{\mu_1}{\mu_K}) \cdots \log(\frac{\mu_{K-1}}{\mu_K})\right]$$

Then we are left to substitute μ_K by η. Note that

$$\eta_i = \log(\frac{\mu_i}{\mu_K}) \Rightarrow \exp(\eta_i) = \frac{\mu_i}{\mu_K}$$

Hence,

$$\begin{aligned}\mu_K &= 1 - \sum_{i=1}^{K-1} \mu_i = 1 - \sum_{i=1}^{K-1} \mu_K \exp(\eta_i) \\ \left(1 + \sum_{i=1}^{K-1} \exp(\eta_i)\right) \mu_K &= 1 \\ \mu_K &= \frac{1}{1 + \sum_{i=1}^{K-1} \exp(\eta_i)}\end{aligned}$$

Thus, we can write the multinoulli distribution as a exponential family,

$$\mathrm{Cat}(\mathbf{x}|\boldsymbol{\mu}) = b(\mathbf{x}) \exp(\eta^\top T(\mathbf{x}) - a(\eta))$$

where $b(\mathbf{x}) = 1$, $\eta^\top = \left[\log(\frac{\mu_1}{\mu_K}) \cdots \log(\frac{\mu_{K-1}}{\mu_K})\right]$, $T(\mathbf{x}) = \mathbf{x}$, and
$a(\eta) = -\log\left(\frac{1}{1+\sum_{i=1}^{K-1} \exp(\eta_i)}\right) = \log\left(1 + \sum_{i=1}^{K-1} \exp(\eta_i)\right)$ ■