Colin Adams
Math189R SU20
Homework 3
June 2020

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.

The mean is given by

$$\begin{aligned}
\mu &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta\theta^{a-1}(1-\theta)^{b-1}\,\mathrm{d}\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^{a}(1-\theta)^{b-1}\,\mathrm{d}\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\
&= \frac{\Gamma(a+1)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+1)} \\
&= \frac{a\Gamma(a)\Gamma(a+b)}{(a+b)\Gamma(a)\Gamma(a+b)} \\
&= \frac{a}{a+b}.
\end{aligned}$$

The variance is given by

$$\begin{aligned}
\sigma^2 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^2\theta^{a-1}(1-\theta)^{b-1}\,\mathrm{d}\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^{a+1}(1-\theta)^{b-1}\,\mathrm{d}\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)}\frac{\Gamma(a+2)}{\Gamma(a+b+2)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)}\frac{(a+1)\Gamma(a+1)}{(a+b+1)\Gamma(a+b+1)} \\
&= \frac{a(a+1)}{(a+b+1)(a+b)}.
\end{aligned}$$

Since $\mathbb{P} = p$ is greater than or equal to zero, the mode is found by taking solving for θ when

$$0 = \frac{\mathrm{d}}{\mathrm{d}\theta} p(\theta; a, b) = a\theta^{a-2}(1-\theta)^{b-1} - \theta^{a-2}(1-\theta)^{b-1} - b\theta^{a-1}(1-\theta)^{b-2} + \theta^{a-1}(1-\theta)^{b-2}$$

If we restrict $0 < \theta < 1$, then the mode can be solved for when

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

And, if we divide both sides by θ^{a-2} and $(1-\theta)^{b-2}$ we can rearrange and solve for θ, giving us

$$\Rightarrow \theta = \frac{a-1}{a+b-2} = \text{mode}.$$

$\blacksquare$

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

So the goal here is to rewriteCat in the form $p(\mathbf{y}|\boldsymbol{\eta}) = b(\mathbf{y})\exp\left[\boldsymbol{\eta}\cdot T(\mathbf{y}) - a(\boldsymbol{\eta})\right]$. So let's do it. First, recall that $\sum_i x_i = 1$ and $\sum_i \mu_i = 1$ for this distribution. Since we know

$$\begin{aligned}
\text{Cat}(\mathbf{x}|\boldsymbol{\mu}) &= \prod_{i=1}^{K} \mu_i^{x_i} \\
&= \exp\left[\log\left(\prod_{i=1}^{K} \mu_i^{x_i}\right)\right] \\
&= \exp\left[\sum_{i=1}^{K} x_i \log \mu_i\right] \\
&= \exp\left[\sum_{i=1}^{K-1} x_i \log \mu_i + x_k \log \mu_k\right] \\
&= \exp\left[\sum_{i=1}^{K-1} x_i \log \mu_i + (1 - x_1 - \ldots - x_{K-1}) \log \mu_k\right] \quad \text{(from constraint)} \\
&= \exp\left[\sum_{i=1}^{K-1} x_i \log\left(\frac{\mu_i}{\mu_k}\right) + \log \mu_k\right]
\end{aligned}$$

If we define $\boldsymbol{\eta} = [\log(\mu_1/\mu_k), \cdots, \log(\mu_{K-1}/\mu_k)]^T$ and $\mathbf{x} = [x_1, \cdots, x_{K-1}]^T$, then we can simplify our function and write

$$\text{Cat}(\mathbf{x}|\boldsymbol{\mu}) = \exp\left[\boldsymbol{\eta}\cdot\mathbf{x} + \log \mu_k\right]$$

which implies that $b(\mathbf{x}) = 1$ and $T(\mathbf{x}) = \mathbf{x}$. However, we need to get the $\log \mu_k$ term in terms of $\boldsymbol{\eta}$ to find $a(\boldsymbol{\eta})$. Since we know that

$$\begin{aligned}
\mu_k = 1 - \mu_1 - \ldots - \mu_{K-1} &= 1 - \sum_{i=1}^{K-1} \mu_k \exp\left[\log(\mu_i/\mu_k)\right] \\
&= 1 - \mu_k \sum_{i=1}^{K-1} \exp\left[\eta_i\right]
\end{aligned}$$

If we solve for μ_k we have

$$\mu_k = \frac{1}{1 + \sum_i^{K-1} \exp\left[\eta_i\right]}, \qquad \Rightarrow a(\boldsymbol{\eta}) = -\log \mu_k = \log\left(1 + \sum_{i=1}^{K-1} \exp\left[\eta_i\right]\right).$$

Thus, the multinomial distribution is part of the exponential family.

■