Xingyao Chen
Math189R SP19
Homework 3
Monday, Feb 18, 2019

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.

Let us for solve for $\Gamma(x+1)$, which will be potentially useful later on.

$$\begin{aligned}
\Gamma(x+1) &= \int_0^\infty u^x e^{-u}\,du \\
&= -u^x e^{-u}\Big|_0^\infty + x\int_0^\infty u^{x-1}e^{-u}\,du \\
&= x\Gamma(x)
\end{aligned}$$

For the mean: Let $\mu = \mathbb{E}(\theta)$ be the mean.

$$\begin{aligned}
\mu &= \int_{\theta=0}^1 \theta p(\theta)\,d\theta \\
&= \int_{\theta=0}^1 \frac{\theta^a(1-\theta)^{b-1}}{B(a,b)}\,d\theta \\
&= \frac{1}{B(a,b)}\int_{\theta=0}^1 \theta^a(1-\theta)^{b-1}\,d\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_{\theta=0}^1 \theta^a(1-\theta)^{b-1}\,d\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\
&= \frac{a}{a+b}\cdot\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\
&= \boxed{\frac{a}{a+b}}
\end{aligned}$$

For the variance, let $\sigma = \text{Var}(\theta)$. Solve for σ

$$
\begin{aligned}
\sigma &= \int_0^1 \theta^2 p(\theta)\, d\theta - \mu^2 \\
&= \int_0^1 \frac{\theta^{a+1}(1-\theta)^{b-1}}{B(a,b)}\, d\theta - \mu^2 \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} - \mu^2 \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} - \mu^2 \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2} \\
&= \frac{a(a+1)(a+b)}{(a+b)(a+b+1)(a+b)} - \frac{a^2(a+b+1)}{(a+b)(a+b)(a+b+1)} \\
&= \frac{(a^3+a^2b+a^2+ab)-(a^3+a^2b+a^2)]}{(a+b)^2(a+b+1)} \\
&= \boxed{\frac{ab}{(a+b)^2(a+b+1)}}
\end{aligned}
$$

For the mode, we know that it occurs where the distribution reaches a maximum, i.e. where the derivative is 0. Solve for the mode

$$
\frac{d}{d\theta}\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\right]
$$

Since $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ is constant and does not affect the optimization of θ, we can drop it.

$$
\begin{aligned}
0 &= \left[(a-1)\theta^{a-2}(1-\theta)^{b-1} - \theta^{a-1}(b-1)(1-\theta)^{b-2}\right] \\
&= \left[\theta^{a-2}(1-\theta)^{b-2}\right]\left[(a-1)(1-\theta)-(b-1)(\theta)\right] \\
0 &= a - a\theta - 1 + \theta - b\theta + \theta \\
&= (a-1) - \theta(a+b-2)
\end{aligned}
$$

$$
\boxed{\theta* = \frac{a-1}{a+b+2}} = \arg\max(Beta(a,b))
$$

■

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 represent the multinoulli as a minimal exponential family as follows. Let $x_k = \mathbb{I}(x = k)$

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

Where $\mu^K = 1 - \sum_{i=1}^{K-1} \mu_i$. Now, we can write this in the exponential family as follows:

$$\mathrm{Cat}(x|\boldsymbol{\mu}) = \exp\left[\boldsymbol{\theta}^T \Phi(\mathbf{x}) - A(\boldsymbol{\theta})\right]$$

Where

$$\Phi(x) = [\mathbb{I}(x = 1), \ldots, \mathbb{I}(x = K - 1)]$$

We can recover the mean parameters from the canonical parameters using

$$\mu_k = 1 - \frac{\sum_{j=1}^{K} e^{\theta_j}}{1 + \sum_{i=1}^{K} e^{\theta_j}}$$

From this, we find

$$\mu_K = 1 - \frac{\sum_{j=1}^{K-1} e^{\theta_j}}{1 + \sum_{j=1}^{K-1} e^{\theta_j}} = \frac{1}{\sum_{j=1}^{K-1} e^{\theta_j}}$$

and hence

$$A(\boldsymbol{\theta}) = \log\left(1 + \sum_{i=1}^{K-1} e^{\theta_i}\right)$$

If we define $\boldsymbol{\theta}_K = 0$, we can write $\boldsymbol{\mu} = \mathcal{S}(\boldsymbol{\theta})$ and $A(\boldsymbol{\theta}) = \log \sum_{i=1}^{K} e^{\theta_i}$, where $\mathcal{S}$ is the softmax function. ∎