Sidhant Rastogi
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.

The beta function is the the integral $B(a,b) = \int_0^1 \theta^a(1-\theta)^{b-1}\,d\theta$. Using this info, we can see that the mean can be found by computing

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

The mode can be found by computing

$$\begin{aligned}
m = \arg\max_\theta \mathbb{P}(\theta; a, b) &= \arg\max_\theta \log \mathbb{P}(\theta; a, b) \\
&= \arg\max_\theta \log \theta^{a-1}(1-\theta)^{b-1} \\
&= \arg\max_\theta (a-1)\log\theta + (b-1)\log(1-\theta)
\end{aligned}$$

We can take the log because it is a monotonically increasing function. Now, we can evaluate this by computing the derivative of this function, then setting it equal to 0 as follows:

$$\begin{aligned}
\frac{\partial}{\partial\theta}[(a-1)\log\theta + (b-1)\log(1-\theta)] &= \frac{a-1}{\theta} - \frac{b-1}{1-\theta} = 0 \\
&\implies \frac{a-1}{\theta} = \frac{b-1}{1-\theta} \\
&\implies (a-1)(1-\theta) = (b-1)\theta \\
&\implies a - a\theta + \theta - 1 = b\theta - \theta \\
&\implies a - 1 = (b+a-2)\theta \\
&\implies \theta = \frac{a-1}{b+a-2} = \boxed{\frac{a-1}{(a-1)+(b-1)}}
\end{aligned}$$

The variance can be found by computing

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

However, we already compute μ, so we get that the variance is

$$\begin{aligned}
\sigma^2 &= \frac{a(a+1)}{(a+b)(a+b+1)} - (\frac{a}{a+b})^2 \\
&= \frac{a(a+1)(a+b) - a^2(a+b+1)}{(a+b)^2(a+b+1)} \\
&= \frac{a^3 + a^2b + a^2 + ab - a^3 - a^2b - a^2}{(a+b)^2(a+b+1)} = \boxed{\frac{ab}{(a+b)^2(a+b+1)}}
\end{aligned}$$

■

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).

The exponential family contains functions of the form

$$b(y)\exp\left(\eta^T T(y) - a(\eta)\right)$$

We see that we can write the multinoulli distribution as

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

However, we see that both $\mathbf{x}$ and $\boldsymbol{\mu}$ can be determined by only $K-1$ values, since

$$x_K = 1 - \sum_{i=1}^{K-1} x_i, \quad \mu_K = 1 - \sum_{i=1}^{K-1} \mu_i$$

Thus, we can break the sum into

$$\begin{aligned}\exp(\sum_{i=1}^{K-1} x_i \log \mu_i + x_K \log \mu_K) &= \exp(\sum_{i=1}^{K-1} x_i \log \mu_i + (1 - \sum_{i=1}^{K-1} x_i)\log \mu_K) \\ &= \exp(\sum_{i=1}^{K-1} x_i \log \frac{\mu_i}{\mu_K} + \log \mu_K)\end{aligned}$$

We can let

$$\boldsymbol{\eta} = (\log\frac{\mu_1}{\mu_K}, \ldots, \log\frac{\mu_{K-1}}{\mu_K})$$

Now, we see that we can rewrite μ_K in terms of the other μ_i as follows:

$$\mu_K = 1 - \sum_{i=1}^{K-1} \mu_i = 1 - \mu_K \sum_{i=1}^{K-1} e^{\eta_i} \implies \mu_K = \frac{1}{1 + \sum_{i=1}^{K-1} e^{\eta_i}}$$

$$e^{\eta_i} = \frac{\mu_i}{\mu_K} \implies \mu_j = \mu_K e^{\eta_j} = \frac{e^{\eta_j}}{1 + \sum_{i=1}^{K-1} e^{\eta_i}}$$

which we recognize as the softmax function applied to η_i. Therefore, this function is in the exponential family with the following parameters:

$$b(\mathbf{x}) = 1, \quad \boldsymbol{\eta} = (\log\frac{\mu_1}{\mu_K}, \ldots, \log\frac{\mu_{K-1}}{\mu_K})$$

$$T(\mathbf{x}) = \mathbf{x}, \quad a(\eta) = -\log(\mu_K) = \log(1 + \sum_{i=1}^{K-1} e^{\eta_i})$$

∎