BigDataHW3

Alex Fay

February 2020

1 Questions

Suppose θ $Beta(a,b)$ such that $P(\theta;a,b) = \frac{1}{B(a,b)*\theta^{b-1}=}$
$\tau(a)\tau(b) * \theta^{a-1}(1-\theta)^{b-1}$

Derive the mean, mode, and variance of theta where B(a,b) is the Beta function and $\tau(x)$ is the Gamma Function.

Mean: The mean can be used using the mean value theorem.

$Mean = \frac{1}{a-b}P(\theta)\theta d\theta$ where a is 1 and b is 0
Substituting our knowns above,

$\int_0^1 (\frac{\theta^{a-1}*((1-\theta)^{b-1}))}{B(a,b)} * \theta d\theta)$
Since theta B(a,b) then we know:

$\tau(a+1*\tau(b) * \frac{1}{\tau(a+b+1)} * \frac{\tau(a+b)}{\tau(a)\tau(b)}$
Thus the mean is $= \frac{a}{a+b}$

Mode:
From finding the gradient vector, we can use this to find the maximum or minimum of the function when the slope is equal to 0.

$0 = \nabla[\theta^{a-1} * (1-\theta^{b-1})]$
$= (a-1) * \theta^(a-2)(1-\theta)^{b-1} - (b-1) * \theta^{a-1} * (1-\theta)^{b-2}$

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

Variance:
$Var[\theta] = E[\theta] - E[\theta]^2$
$E[\theta] = \frac{a}{a+b}$ given above
$E[\theta]^2 = \frac{a^2}{(a+b)^2}$

$E[\theta^2] = \int_0^1 \theta^2 * P(\theta, a, b) d\theta$
$= \frac{1}{B(a,b) \int_0^1 \theta^{a+1}(1-\theta)^{b-1} d\theta}$
$= \frac{B(a+2,b)}{B(a,b)}$

Using similar method to part A:
$= \frac{a(a+1)}{(a+b)(a+b+1)}$
$Var[\theta] = \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2}$

From simplifying the above we find the Variance is:
$= \frac{ab}{(a+b)^2(a+b+1)}$

2. Show that the multinomial distribution is in the exponential family and that the generalized linear model corresponding to this distribution is the same as multinomial log regression.

$Cat(x|\mu) = \prod_i = 1_K \mu_i^{xi}$

$= e^{log(\prod_i = 1_K \mu_i^{xi})}$
$= e^{\sigma_{i=1}^{K} x_i * log(\mu_i)}$

Since $\Sigma_{i=1}^{K} \mu_i = 1 = \Sigma_{1K} x$ then:

$= e^{\Sigma_{i=1}^{K} - 1 x_i log(\mu_i) + (1 - \Sigma_{i=1K} - 1 x_i) * log(\mu_k)}$

$1 - \sigma_{i=1}^{K} - 1 x_i$ becomes $log \mu_k$

Thus from properties of logs:

$e^{\Sigma_{i=1}^{K} - 1 x_i * log(\frac{\mu_i}{\mu_k) + log(\mu_K)}}$
Since from the above, $\mu_i = \mu_k * e^n then n = [log(\frac{\mu_i}{\mu_k}); log(\frac{\mu_{k-1}}{\mu_k})]$

b(n) = x and a(n) = $-log(\Sigma_{i=1}^{K} - 1 e^n)$
Thus, we know the multinomial distribution is part of the exponential family and since the mu is the multinomial log function then it has a multinomial log/softmax regression.