HW # 3

1.

$\theta \sim \text{Beta}(a, b)$ s.t.

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

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

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(x+1) = x\,\Gamma(x)$$

Mean of θ:

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

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

$$= \left[\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \right] \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right] = \left[\frac{a\,\cancel{\Gamma(a)}\cancel{\Gamma(b)}}{\cancel{\Gamma(a+b)}\Gamma(a+b)} \right] \left[\frac{\cancel{\Gamma(a+b)}}{\cancel{\Gamma(a)\Gamma(b)}} \right]$$

$$= \frac{a}{a+b}$$

We know $\text{Var}[\theta] = E[\theta^2] - E[\theta]^2$

$$\mathbb{E}[\theta^2] = \int_0^1 \theta^2 \left(\frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} \right) d\theta$$

$$= \frac{B(a+2,b)}{B(a,b)} = \left[\frac{\Gamma(a+2)\,\Gamma(b)}{\Gamma(a+b+2)} \right] \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]$$

$$\vdots$$

$$= \frac{a(a+1)}{(a+b)(a+b+1)}$$

$$\mathrm{Var}[\theta] = \mathbb{E}[\theta^2] - \mathbb{E}[\theta]^2$$

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

$$\vdots$$

$$= \frac{ab}{(a+b)^2(a+b+1)}$$

For mode, we want to find when
$\nabla_\theta P(\theta; a, b) = 0$ on $[0,1]$.

$$\nabla_\theta P(\theta;\ a,b) = \nabla_\theta [\theta^{a-1}(1-\theta)^{b-1}] = 0$$

$$= (a-1)\,\theta^{a-2}(1-\theta)^{b-1} - (b-1)\,\theta^{a-1}(1-\theta)^{b-2} = 0$$

So, $(a-1)(1-\theta) = (b-1)\,\theta$

$$(a+b-2)\,\theta = (a-1)$$

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

2. Show $\quad Cat(x|\mu) = \prod_{i=1}^{k} \mu_i^{x_i}$

Recall...

$$P(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta))$$

$$Cat(x|\mu) = \prod_{i=1}^{k} \mu_i^{x_i} = \exp\left[\log\left(\prod_{i=1}^{k} \mu_i^{x_i}\right)\right]$$

$$= \exp\left(\sum_{i=1}^{k} x_i \log(\mu_i)\right)$$

Since $\sum_{i=1}^{k} \mu_i = 1$ $\quad \sum_{i=1}^{k} x_i = 1$ we need to specify $k-1$ of these since x_k & μ_k will be determined at the end:

$$\mu_k = 1 - \sum_{i=1}^{k-1} \mu_i$$

$$x_k = 1 - \sum_{i=1}^{k-1} x_i$$

$$\therefore \quad Cat(x|\mu) = \exp\left(\sum_{i=1}^{k} x_i \log(\mu_i)\right)$$

$$= \exp\left(\sum_{i=1}^{k-1} x_i \log(\mu_i) + x_k \log(\mu_k)\right)$$

$$= \exp\left[\sum_{i=1}^{k-1} x_i \log(\mu_i) + \left(1 - \sum_{i=1}^{k-1} x_i\right) \log(\mu_k)\right]$$

$$= \exp\left[\sum_{i=1}^{k-1} x_i \left(\log(\mu_i) - \log(\mu_k)\right) + \log(\mu_k)\right]$$

$$= \exp\left[\sum_{i=1}^{k-1} x_i \log\left(\frac{\mu_i}{\mu_k}\right) + \log(\mu_k)\right]$$

$\therefore$ let vector η be:

$$\eta = \begin{bmatrix} \log\left(\frac{\mu_1}{\mu_k}\right) \\ \cdots \\ \log\left(\frac{\mu_{k-1}}{\mu_k}\right) \end{bmatrix}$$

we can see that $\mu_i = \mu_k e^{\eta}$

$$\therefore \mu_k = 1 - \sum_{i=1}^{k-1} \mu_i = 1 - \sum_{i=1}^{k-1} \mu_k e^{\eta_i}$$

$$= 1 - \mu_k \sum_{i=1}^{k-1} e^{\eta_i}$$

$$= \frac{1}{1 + \sum_{i=1}^{k} e^{\eta_i}}$$

$$\therefore \mu_i = \mu_k e^{\eta_i} = \frac{e^{\eta_i}}{1 + \sum_{i=1}^{k-1} e^{\eta_i}}$$

In form of exponential family, we get:

$b(\eta) = 1 \qquad T(x) = x$

$$a(\eta) = -\log(\mu_k) = \log\left(1 + \sum_{i=1}^{k-1} e^{\eta_i}\right)$$

Therefore, $Cat(x|\mu)$ is in the exponential family. $\mu = S(\eta)$ is softmax function which implies generalized linear model.