Luke Fu

I was guided by solution's guide

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

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

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

where $B(a,b) = \Gamma(a)\Gamma(b) / \Gamma(a+b)$

& $\Gamma(x)$ is the gamma function.

Derive the mean, mode, & variance of θ

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

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

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

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

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

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

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

$$= \boxed{\frac{a}{a+b}} \quad \text{Mean}$$

By def:

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

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

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

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

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

$$E[\theta^2] - E[\theta]^2$$

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

$$= \frac{a^3 + a^2 b + a^2 + ab - a^3 - a^2 b - \cancel{a^2}}{(a+b)^2(a+b+1)}$$

$$= \boxed{\frac{ab}{(a+b)^2(a+b+1)}} \quad \text{Variance}$$

$$\nabla_\theta\, p(\theta; a,b) = 0 \text{ on interval } [0,1]$$

$$= \nabla_\theta \left[\theta^{a-1}(1-\theta)^{b-1}\right] = 0$$

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

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

$$= (b-1)\theta$$

$$= a-1$$

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

2.

$$\mathrm{Cat}(x|\mu) = \prod_{i=1}^{k} \mu_i^{x_i}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

$b(\eta) = 1$

$T(x) = x$

$a(\eta) = -\log(\mu_k) = \log\left(1 + \sum_{i=1}^{k-1} e^{\eta_i}\right)$ $\therefore$ GLM of this distr. same as softmax regression

$\mu = S(\eta)$ where $S(\eta)$ is softmax func.