A

Mean

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

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

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

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

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

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

Varince

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

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

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

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

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

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

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

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

mode

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

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

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

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

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

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

2.

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

$$= \exp[\log(\prod_{i=1}^{k} \mu_i^{x_i})]$$

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

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

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

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

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

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

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

$$\mu_i = \mu_k e^{\eta_i}$$

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

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

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

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

$\rightarrow$

$$Cat(x|\mu) = \exp(\eta^T x - a(\eta))$$

$$b(\eta) = 1$$

$$T(x) = x$$

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

$\therefore$ $cat(x|\mu)$ is in exponential family & $\mu = S(\eta)$ which is softmax func distribution