1)

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

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

now calculate mean of θ

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

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

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

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

$$Var[\theta) = E(\theta - E(\theta))^2 = E(\theta^2] - E(\theta)^2$$

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

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

$$\left(\frac{\Gamma(a+2)\,\Gamma(b)}{\Gamma(a+b+2)}\right)\left(\frac{\Gamma(a+b)}{\Gamma(a)+\Gamma(b)}\right)$$

$$\downarrow$$

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

so variance $= E(\theta^2) - E(\theta)^2$

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

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

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

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

mode : $\nabla_\theta\, p(\theta; a, b) = 0$

$$\nabla_\theta\, p(\theta; a, b) = \nabla_\theta\left(\theta^{a-1}(1-\theta)^{b-1}\right) = 0$$

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

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

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

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

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

2. Exponential family

$$f(x|\theta) = h(x)\exp(\eta(\theta)^T T(x) - A(\theta))$$

$$\mathrm{Cat}(x|u) = \prod_{i=1}^{K} u_i^{x_i} = \exp\left(\log\left(\prod_{i=1}^{K} u_i^{x_i}\right)\right)$$

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

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

Since $\sum_{i=1}^{K} u_i = 1$ and $\sum_{i=1}^{K} x_i = 1$, so we only need to specify first $K-1$ of these terms, since x_K and u_K would be clarified from the result.

$$u_K = 1 - \sum_{i=1}^{K-1} u_i$$

$$x_K = 1 - \sum_{i=1}^{K-1} x_i$$

$$\mathrm{Cat}(x|u) = \exp\left(\sum_{i=1}^{K} x_i \log(u_i)\right) = \exp\left(\sum_{i=1}^{K-1} x_i \log(u_i) + x_K \log(u_K)\right)$$

$$= \exp\left(\sum_{i=1}^{K-1} x_i \log(u_i) + \left(1 - \sum_{i=1}^{K-1} x_i\right)\log(u_K)\right)$$

$$= \exp\left(\sum_{i=1}^{K-1} x_i \left(\log(u_i) - \log(u_K)\right) + \log(u_K)\right)$$

$$= \exp\left(\sum_{i=1}^{K-1} x_i \log\left(\frac{u_i}{u_K}\right) + \log(u_K)\right)$$

so $\vec{\eta} = \vec{\eta} = \begin{bmatrix} \log\left(\frac{u_1}{u_k}\right) \\ \vdots \\ \log\left(\frac{u_{k-1}}{u_k}\right) \end{bmatrix} \Rightarrow u_i = u_k e^{\vec{\eta}_i}$

$$u_k = 1 - \sum_{i=1}^{k-1} u_i = 1 - \sum_{i=1}^{k-1} u_k e^{\eta} = 1 - u_k \sum_{i=1}^{k-1} e^{\vec{\eta}_i} = \frac{1}{1 + \sum_{i=1}^{k-1} e^{\vec{\eta}_i}}$$

so $u_i = u_k e^{\vec{\eta}_i} = \dfrac{e^{\vec{\eta}_i}}{1 + \sum_{i=1}^{k-1} e^{\vec{\eta}_i}}$

writing the distribution in the form of exponential family as

$Cat(x|u) = \exp(\eta^T x - a(\eta))$, $b(\vec{\eta}) = 1$ $T(\vec{x}) = x$

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

therefore, $Cat(x|u)$ is in the exponential family. and $u = S(\vec{\eta})$, which is the softmax function, which implies the generalized linear model of this distribution is the same as softmax regression.