ENTROPY CALCULATION FOR EFs:

$$H(p) = -\int p(x) \ln p(x) =$$

$$= -\int p(x) \left[\theta\, s(x) - A(\theta) + h(x) \right] dx =$$

$$= -\int p(x)\, \theta\, s(x) dx + \int p(x) A(\theta) dx - \int p(x) h(x)\, dx =$$

$$= -\theta^T \underbrace{\mu}_{\text{(moment param.)}} + A(\theta) - K.$$

→ To get moment parameters:

* **GAMMA DISTRIBUTION:**

$$H(\gamma) = -\begin{pmatrix} \alpha - 1 \\ -\beta \end{pmatrix} \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} - \alpha \ln\beta + \ln\Gamma(\alpha) =$$

$$= -(\alpha-1)\mu_0 + \beta\mu_1 - \alpha\ln\beta + \ln\Gamma(\alpha)$$

Must be equal to (from Wikipedia)

$$H(\gamma) = \alpha - \ln\beta + \ln(\Gamma(\alpha)) + (1-\alpha)\Psi(\alpha)$$

$$(1-\alpha)\mu_0 + \beta\mu_1 - \alpha\ln\beta = \underline{(1-\alpha)\Psi(\alpha) - \ln\beta + \alpha}$$

if $\boxed{\mu_0 = \Psi(\alpha) - \ln\beta}$ & $\boxed{\mu_1 = \frac{\alpha}{\beta}}$ $\Rightarrow$

$$(1-\alpha)(\Psi(\alpha) - \ln\beta) + \cancel{\beta}\frac{\alpha}{\cancel{\beta}} - \alpha\ln\beta =$$

$$= \Psi(\alpha) - \alpha\Psi(\alpha) - \ln\beta + \cancel{\alpha\ln\beta} + \alpha - \cancel{\alpha\ln\beta} =$$

$$= \underline{(1-\alpha)\Psi(\alpha) - \ln\beta + \alpha} \quad //.$$

* INVERSE GAMMA DISTRIBUTION:

$$H(\gamma^{-1}) = -\begin{pmatrix} -\alpha-1 \\ -\beta \end{pmatrix}\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} + \ln\Gamma(\alpha) - \alpha\ln\beta =$$

$$= -(-\alpha-1)\mu_0 + \beta\mu_1 + \ln\Gamma(\alpha) - \alpha\ln\beta =$$

$$= (\alpha+1)\mu_0 + \beta\mu_1 + \ln\Gamma(\alpha) - \alpha\ln\beta$$

Must be equal to (from Wikipedia):

$$H(\gamma^{-1}) = \alpha + \ln(\beta\Gamma(\alpha)) - (1+\alpha)\psi(\alpha)$$

$$(\alpha+1)\mu_0 + \beta\mu_1 - \alpha\ln\beta = \underline{\alpha + \ln\beta - (1+\alpha)\psi(\alpha)}$$

if $\boxed{\mu_0 = -\psi(\alpha) + \ln\beta}$ & $\boxed{\mu_1 = \frac{\alpha}{\beta}}$

$$(\alpha+1)[-\psi(\alpha) + \ln\beta) + \beta\mu_1 - \alpha\ln\beta =$$

$$= -\alpha\psi(\alpha) - \psi(\alpha) + \cancel{\alpha\ln\beta} + \ln\beta + \cancel{\beta}\frac{\alpha}{\cancel{\beta}} - \cancel{\alpha\ln\beta} =$$

$$= \underline{-(1+\alpha)\psi(\alpha) + \ln\beta + \alpha} \quad //$$

* **FROM MOMENT TO NATURAL PARAMETERS:**

$\boxed{\text{Gamma dist.}}$

$$\theta^* = \arg\max_{\theta} \overbrace{\theta\,\mu - A(\theta)}^{f(\alpha,\beta)} =$$

$$= \arg\max_{\substack{\alpha,\beta \\ \alpha>0,\,\beta>0}} \begin{pmatrix} \alpha-1 \\ -\beta \end{pmatrix}\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} - \ln\Gamma(\alpha) + \alpha\ln\beta =$$

$$= \arg\max_{\alpha>0,\beta>0} (\alpha-1)\mu_0 - \beta\mu_1 - \ln\Gamma(\alpha) + \alpha\ln\beta$$

$$\frac{\partial f(\alpha,\beta)}{\partial\alpha} = \mu_0 - \frac{\partial \ln\Gamma(\alpha)}{\partial\alpha} + \ln\beta = 0;$$

$$\mu_0 - \Psi(\alpha) + \ln\beta = 0; \quad \boxed{\alpha = \Psi^{-1}(\mu_0 + \ln\beta)}$$

$$\frac{\partial f(\alpha,\beta)}{\partial\beta} = -\mu_1 + \alpha\cdot\frac{1}{\beta} = 0; \quad \boxed{\beta = \frac{\alpha}{\mu_1}}$$

Coordinate ascent

$\boxed{\text{Inv-gamma dist.}}$

$$\theta^* = \arg\max_{\alpha>0,\beta>0} \begin{pmatrix} -\alpha-1 \\ -\beta \end{pmatrix}\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} - \ln\Gamma(\alpha) + \alpha\ln\beta =$$

$$= \arg\max (-\alpha-1)\mu_0 - \beta\mu_1 - \ln\Gamma(\alpha) + \alpha\ln\beta$$

$$\frac{\partial f(\alpha,\beta)}{\partial\alpha} = -\mu_0 - \Psi(\alpha) + \ln\beta = 0; \quad \boxed{\alpha = \Psi^{-1}(\ln\beta - \mu_0)}$$

$$\frac{\partial f(\alpha,\beta)}{\partial\beta} = -\mu_1 + \frac{\alpha}{\beta}; \quad \boxed{\beta = \frac{\alpha}{\mu_1}}$$

Coord. ascent

* A good starting point for coordinate ascent can be:

$$\alpha = e^{\mu_0 + \ln\beta} \quad \text{for gamma dist. \&}$$

$$\alpha = e^{\ln\beta - \mu_0} \quad \text{for inv-gamma dist}$$

* **FROM NATURAL TO MOMENT PARAMETERS:**

$\boxed{\text{Gamma dist.}}$ $\quad \alpha - 1 = \theta_1 \;;\; \alpha = \theta_1 + 1$

$\quad -\beta = \theta_2 \;;\; \beta = -\theta_2$

$$\boxed{\mu_1} = \frac{\partial A(\theta_1,\theta_2)}{\partial\theta_1} = \partial/\partial\theta_1 \; \ln\Gamma(\theta_1+1) - (\theta_1+1)\ln(-\theta_2) =$$

$$= \Psi(\theta_1+1) - \ln(-\theta_2) = \boxed{\Psi(\alpha) - \ln(\beta)}$$

$$\boxed{\mu_2} = \frac{\partial A(\theta_1,\theta_2)}{\partial\theta_2} = -(\theta_1+1)\frac{-1}{-\theta_2} = -(\alpha)\frac{-1}{\beta} = \boxed{\frac{\alpha}{\beta}}$$

$\boxed{\text{Inv-gamma dist.}}$ $\quad -\alpha - 1 = \theta_1 \;;\; \alpha = -1 - \theta_1$

$\quad -\beta = \theta_2 \;;\; \beta = -\theta_2$

$$\boxed{\mu_1} = \frac{\partial A(\theta_1,\theta_2)}{\partial\theta_1} = \partial/\partial\theta_1 \; \ln\Gamma(-1-\theta_1) - (-1-\theta_1)\ln(-\theta_2) =$$

$$= -\Psi(-1-\theta_1) + \ln(-\theta_2) = \boxed{\ln(\beta) - \Psi(\alpha)}$$

$$\boxed{\mu_2} = \frac{\partial A(\theta_1,\theta_2)}{\partial\theta_2} = (1+\theta_1)\frac{-1}{-\theta_2} = (1-\alpha-1)\frac{1}{-\beta} = \boxed{\frac{\alpha}{\beta}}$$