Mellowmax

Shayan Shafquat

July 2024

1 Background

$$\mathrm{mm}_\omega(\mathbf{X}) = \frac{\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\omega x_i}\right)}{\omega} \tag{1}$$

$$\pi_{\mathrm{mm}}(a \mid s) = \frac{e^{\beta \hat{Q}(s,a)}}{\sum_{a \in A} e^{\beta \hat{Q}(s,a)}} \quad \forall a \in A,$$

where β is a value obtained after optimizing the below equation:

$$\sum_{a \in A} e^{\beta\left(\hat{Q}(s,a) - \mathrm{mm}_\omega \hat{Q}(s,.)\right)} \left(\hat{Q}(s,a) - \mathrm{mm}_\omega \hat{Q}(s,.)\right) = 0. \tag{2}$$

2 Mellowmax operator for patch-foraging task

Value Functions of staying/leaving at nth timestep:

$$Q(n, \text{stay}) = r(n) = r \qquad\qquad Q(n, \text{leave}) = 0$$

From (1), we have

$$\mathrm{mm}_\omega(\hat{Q}(n,.)) = \frac{\log\left(\frac{1}{n}(1 + e^{\omega r(n)})\right)}{\omega} = mm_\omega$$

Probability of staying/leaving at nth timestep given by:

$$\pi_{mm}(\text{stay} \mid n) = \frac{e^{\beta r(n)}}{1 + e^{\beta r(n)}} \qquad\qquad \pi_{mm}(\text{leave} \mid n) = \frac{1}{1 + e^{\beta r(n)}}$$

Optimization step: Using the above value functions of staying and leaving to (2) we obtain,

$$e^{\beta(-\mathrm{mm}_\omega(\hat{Q}(n,.)))}(-\mathrm{mm}_\omega(\hat{Q}(n,.))) + e^{\beta(r(n) - \mathrm{mm}_\omega(\hat{Q}(n,.)))}(r(n) - \mathrm{mm}_\omega(\hat{Q}(n,.))) = 0$$

$$\Rightarrow -e^{-\beta \mathrm{mm}_\omega}\mathrm{mm}_\omega + e^{\beta(r - \mathrm{mm}_\omega)}(r - \mathrm{mm}_\omega) = 0$$

$$\Rightarrow r e^{\beta(r - \mathrm{mm}_\omega)} = \mathrm{mm}_\omega\left[e^{-\beta \mathrm{mm}_\omega} + e^{\beta(r - \mathrm{mm}_\omega)}\right]$$

Solving the above equation we obtain $\hat{\beta}$, which is used as the stochastic choice model,

$$\pi_{mm}(\text{leave} \mid n) = \frac{1}{1 + e^{\hat{\beta} r(n)}}$$