A Proof of Lemma 1

We show that $\zeta^{\mathcal{L}}(s,a;D_{s,a},k)$ defined in Eq. (7) decays to 0 at a rate of $1/n_{s,a}^2$. We present the proof for $k=1$. The extension to $k>1$ is straightforward.

In finite MDPs, we have to learn separate transition probability tables $\mathcal{T}(s' \mid s,a)$ for each s,a. For simplicity, we focus on one fixed (s,a) and investigate how to estimate the distribution $\mathcal{T}(s')$. We consider a Dirichlet learner where α denotes the Dirichlet posterior based on the data $D_{s,a} = \{s'_i\}_{i=1}^{n_{s,a}}$ and α' be the posterior based on the reduced data set $D_{s,a}^{k=-1}$, that is, $\alpha'_c + 1 = \alpha_c$ where c is the outcome of the missing experience in $D_{s,a}^{k=-1}$, and $\alpha'_j = \alpha_j$ for all $j \neq c$. Given Dirichlet parameters α, the MAP model $\hat{\mathcal{T}}_\alpha(s')$ is given by the vector $\alpha/\bar{\alpha}$, $\bar{\alpha} = \sum_i \alpha_i$, and we estimate ζ using Eq. (7). The log-likelihood of the data $D_{s,a}$ under $\hat{\mathcal{T}}_\alpha$ is

$$\begin{aligned} L^+ &:= \log P(D_{s,a}|\hat{\mathcal{T}}_\alpha) \\ &= \log \prod_{i=1}^{n_{s,a}} \frac{\alpha_{s'_i}}{\bar{\alpha}} = \sum_{i=1}^{n_{s,a}} \log \alpha_{s'_i} - n_{s,a} \log \bar{\alpha} \end{aligned} \tag{14}$$

The likelihood of the data $D_{s,a}$ under $\hat{\mathcal{T}}_{\alpha'}$ is

$$\begin{aligned} L^- &:= \log P(D_{s,a}|\hat{\mathcal{T}}_{\alpha'}) \\ &= \log\Big(\prod_{i=1,s'_i\neq c}^{n_{s,a}} \frac{\alpha_{s'_i}}{\bar{\alpha}-1}\Big)\Big(\prod_{i=1,s'_i=c}^{n_{s,a}} \frac{\alpha_c - 1}{\bar{\alpha}-1}\Big) \\ &= \sum_{i=1,s'_i\neq c}^{n_{s,a}} \log \alpha_{s'_i} + \sum_{i=1,s'_i=c}^{n_{s,a}} \log(\alpha_c - 1) - n_{s,a}\log(\bar{\alpha}-1) \end{aligned} \tag{15}$$

The average difference is

$$\begin{aligned} \zeta^{\mathcal{L}}(s,a;k=1) &= \frac{1}{n_{s,a}}|L^+ - L^-| \\ &= \frac{1}{n_{s,a}} \sum_{i=1,s'_i=c}^{n_{s,a}} (\log\alpha_c - \log(\alpha_c - 1)) - \log\bar{\alpha} + \log(\bar{\alpha}-1) \\ &= \frac{n_{s,a,c}}{n_{s,a}} \log\frac{1}{1-\frac{1}{\alpha_c}} + \log(1-\frac{1}{\bar{\alpha}})\ . \end{aligned} \tag{16}$$

Since $\bar{\alpha} \propto n_{s,a}$, by taking the derivative of the expected value of $\zeta^{\mathcal{L}}(s,a;k=1)$ we can verify that $E_{D_{s,a}}(\zeta(s,a;D_{s,a},k=1)) = O\left(\frac{1}{n_{s,a}^2}\right)$. $\square$