UBE samples PriorMDP ($t = 500$, $\zeta = 0.1$, $N_s = 4$, $N_a = 2$)

Sample mean — ✕ (red)   True $Q^*_{\mathbf{s}, \mathbf{a}}$ — ✕ (black)