REINFORCE P-T (σ=0.6)
action hist for p0 in]0.8, 0.6]