REINFORCE P-T (σ=0.6)
action hist for p0 >= 0.8