REINFORCE P-T-V (σ=0.4)
action hist for p0 >= 0.8