REINFORCE L-T-V (σ=0.4)
action hist for p0 in]0.6, 0.4]