Online Supplement to Risk-Sensitive Control of Branching Processes

by Pelin G. Canbolat

Risk-Seeking Exponential Utility and Power Utility

In this supplement, we extend the results obtained for risk-averse exponential utility in the paper titled *Risk-Sensitive Control of Branching Processes* to risk-seeking exponential utility and to (risk-averse and risk-seeking) power utility with multiplicative rewards. To start with, we express the risk-seeking exponential utility of a reward w as $u(w) = e^{\lambda w}$ for some parameter $\lambda > 0$ and define the *utility operator* $\mathcal{T}_a$ of action a as

$$\mathcal{T}_a U = \sum_{x \in \mathcal{X}} q(x|a) e^{\lambda r(x,a)} U^x,$$

for any scalar $U \geq 0$ (with $U^0 = 1$ for $U \geq 0$). Theorem 1 of the main paper remains valid after replacing $\mathcal{D}_a$ by $\mathcal{T}_a$, and the expected disutility C_a by the expected utility U_a.

We define the *optimal utility operator* $\mathcal{T}$ as

$$\mathcal{T} U = \max_{a \in A} \left\{ \sum_{x \in \mathcal{X}} q(x|a) e^{\lambda r(x,a)} U^x \right\},$$

for $U \geq 0$. Like the optimal disutility operator $\mathcal{D}$, the optimal utility operator $\mathcal{T}$ is continuous and increasing. The following theorem is the risk-seeking counterpart of Theorem 2 of the main paper.

Theorem 5. *The maximum expected N-period utility $\mathcal{T}^N 1$ converges monotonically to $0 \leq U = \sup_{\pi \in A^\infty} U_\pi$ where $A^\infty = \{(a_n)_{n \geq 1} : a_n \in A\}$. If $U < \infty$, then there exists a stationary maximum-expected-utility policy $\pi^* = (a_n)_{n \geq 1}$ with $a_n = a^*$ for all $n \geq 1$ and $a^* \in A$ such that $U = \mathcal{T}_{a^*} U$.*

Proof. The optimal-utility operator $\mathcal{T}$ is increasing, implying that either $\mathcal{T}1 \leq 1$ and $\mathcal{T}^N 1$ converges down to $U < 1$, or $\mathcal{T}1 > 1$ converges up to $U \geq 1$. Since $\mathcal{T}^N 1 \geq 0$ for every N, $U \geq 0$ in both cases. Also, $\mathcal{T}^N 1 \geq U_\pi^N$ for all π, taking limit superior of both sides ensures that $U \geq U_\pi$ for all π. Now suppose $U < \infty$. The finiteness of the action set guarantees the existence of $a^* \in A$ with $U = \mathcal{T}_{a^*} U$. As $U \geq U_\pi$ for all π, $U \geq U_{a^*}$.

If $U \leq 1$, then $U = \mathcal{T}_{a^*}^N U \leq \mathcal{T}_{a^*}^N 1$ and taking limit gives $U \leq U_{a^*}$, hence $U = U_{a^*}$. On the other hand, if $U > 1$, then $\mathcal{T}1 > 1$ and U is the least of the fixed points of $\mathcal{T}$ that are greater than or equal to 1. To see why the latter statement holds, first note that the continuity of $\mathcal{T}$ implies that $U = \mathcal{T}U$; secondly, for any other fixed point $U' \geq 1$, the monotonicity of $\mathcal{T}$ implies $\mathcal{T}^N 1 \leq \mathcal{T}^N U'$, and taking limit of both sides leads to $U \leq U'$. The optimal utility operator $\mathcal{T}$ is maximum of polynomials with nonnegative coefficients, so it is convex on $(0, \infty)$. The function $\mathcal{T}V - V$ must be positive and decreasing on the interval $[1, U)$, since $\mathcal{T}1 - 1 > 0 = \mathcal{T}U - U$ and U is the smallest value greater than one satisfying $\mathcal{T}V - V = 0$. Then for a sufficiently small $\epsilon > 0$, $\mathcal{T}V = \mathcal{T}_{a^*}V$ for all $V \in [U - \epsilon, U]$, so $\mathcal{T}_{a^*}V - V$ is positive and decreasing on the interval $[U - \epsilon, U)$. The convexity of $\mathcal{T}_{a^*}$ implies $\mathcal{T}_{a^*}V - V > 0$ for all $V < U$. Hence U must be the smallest nonnegative fixed point of $\mathcal{T}_{a^*}$ and is greater than 1, so $U_{a^*} = U$. $\square$

The results with risk-seeking exponential utility allow extending the theory to power utility functions. To show this, we consider the power utility function $u(w) = w^\lambda$ with parameter $\lambda > 0$ for $w > 0$ and multiplicative rewards such that the reward collected over periods $n = 1, \ldots, N$ is

$$\prod_{n=1}^{N} \prod_{k=1}^{Z_n} r(X_{n+1}^k, \pi_n^k),$$

where π_n^k represents the action assigned to individual k in period n. The expected power utility of these rewards given the initial population size $Z_1 = z$ is then

$$U_\pi^N(z) = \mathrm{E}\left[\left(\prod_{n=1}^{N} \prod_{k=1}^{Z_n} r(X_{n+1}^k, \pi_n^k)\right)^\lambda \mid Z_1 = z, \pi\right] = \mathrm{E}\left[e^{\lambda \sum_{n=1}^{N} \sum_{k=1}^{Z_n} \ln r(X_{n+1}^k, \pi_n^k)} \mid Z_1 = z, \pi\right],$$

meaning that the expected power utility of multiplicative rewards equals the expected risk-seeking exponential utility of additive logarithmic rewards. Consequently, we can apply the results stated in this section to the problem with power utility.