Optimal Control with Distorted Probability Distributions

Kerem Uğurlu

Tuesday 11th February, 2020

Abstract

We study a robust optimal control of discrete time Markov chains with finite terminal T and bounded costs using probability distortion. The time inconsistency of these operators and hence its lack of dynamic programming are discussed. Due to that, dynamic versions of these operators are introduced, and its availability for dynamic programming is demonstrated. Based on dynamic programming algorithm, existence of the optimal policy is justified and an application of the theory to portfolio optimization is also presented.

Keywords: Decision Science; Probability Distortion; Controlled Markov Chains; Risk Management; Mathematical Finance

1 Introduction

Dynamic programming [1] is one of the fundamental areas in operations research. Initially, dynamic programming models have used expected value as the performance criteria, but since in many real life scenarios, the expected value as the performance measure is not appropriate, models with risk aversion are represented via different approaches. The first approach is using concave utility functions modelling risk-aversion (see e.g. [2, 3, 4, 5, 6] and the reference therein). Another approach has been to use the so-called coherent/convex risk measures. Starting from the seminal work of Artzner et al. [7] dynamic coherent/convex risk measures have seen huge developments since then (see [8, 20, 10, 11, 12, 13, 14]).

The third approach that we will follow in this paper is so called probability distortion. Probability distortion is an approach that is used frequently in behaviour finance (see e.g.

[15, 16, 17]). It has been motivated by empirical studies in behavioural finance and aims to model the human tendency to exaggerate small probabilities of extreme events with respect to the underlying probability measure such as catastrophic ruin or the chance of winning lottery. This is characterized by an operator on random outcomes, where the underlying probability distribution is distorted by a weight function w satisfying some properties. The ability to capture human decision dynamics under uncertainty has strong empirical support ([18]).

Although, modelling random outcomes representing gains/losses using probability distortion goes back to at least 1970's ([15]), its axiomatic incorporation into multiperiod settings is still absent in the literature. There are few recents works in this direction. [19] studies a portfolio optimization problem in continuous time using probability distortion. [21] studies a discrete time controlled Markov chain in infinite horizon. [22] extends the probability distortion operator to multitemporal setting under some monotonicity assumptions of the cost/gain functions.

The reason for scarcity of using probability distortions in multitemporal settings is it does not satisy "Dynamic Programming Principle" (DPP) or "Bellman Optimality Principle". Namely, a sequence of optimization problems with the corresponding optimal controls is called time-consistent, if the optimal strategies obtained when solving the optimal control problem at time s stays optimal when the optimal control problem is solved at time $t > s$. We refer the reader to [26] for an extensive study on time-consistency.

In this paper, we introduce a dynamic version of probability distortion that does not suffer from time-inconsistency. Hence, DPP can be applied readily in our framework under controlled Markov chains, and additionally DPP gives the existence of the optimal policy under some reasonable assumptions on the model.

The rest of the paper is as follows. In Section 2, we describe probability distortion on random variables in static one period case first. Next, we introduce the concept of dynamic probability distortion on stochastic processes in multi-temporal discrete time setting. In Section 3, we introduce the controlled Markov chain framework that we are going to work on. In Section 4, we state and solve our optimal control problem by characterizing the optimal policy. In Section 5, we apply our results to a portfolio optimization problem and conclude the paper.

2 Probability Distortion

2.1 Probability Distortion on Random Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and denote by $L^\infty_+(\Omega, \mathcal{F}, \mathbb{P})$ the set of non-negative essentially bounded random variables on $(\Omega, \mathcal{F})$.

Definition 2.1.
- *A mapping $w : [0,1] \to [0,1]$ is called a distortion function, if it is continuous, strictly increasing, and satisfies $w(0) = 0$ and $w(1) = 1$.*
- *For any $\xi \in L^\infty_+(\Omega, \mathcal{F}, \mathbb{P})$, the operator with respect to the distortion function w is defined by*

$$\rho(\xi) \triangleq \int_0^\infty w(\mathbb{P}(\xi \geq z))dz \tag{2.1}$$

Lemma 2.1. *(i) Let $x, y, \alpha \in [0,1]$, $\xi \in L^\infty_+(\Omega, \mathcal{F}, \mathbb{P})$ and $w : [0,1] \to [0,1]$ be a distortion function that satisfies*

$$w(\alpha x + (1-\alpha)y) \geq \alpha w(x) + (1-\alpha)w(y). \tag{2.2}$$

Then, $\rho(\xi) \geq \mathbb{E}[\xi]$. Namely, for ξ representing the nonnegative bounded random losses, $\rho(\cdot)$ evaluates a bigger risk for ξ than $\mathbb{E}[\cdot]$ does.

(ii) Conversely, suppose w satisfies for $\alpha \in [0,1]$

$$w(\alpha x + (1-\alpha)y) \leq \alpha w(x) + (1-\alpha)w(y), \tag{2.3}$$

then $\rho(\xi) \leq \mathbb{E}[\xi]$. Namely, for ξ representing the nonnegative bounded random gains, $\rho(\cdot)$ evaluates a smaller gain for ξ than $\mathbb{E}[\cdot]$ does.

Proof. We will only prove the first part. By $w(0) = 0$ and (2.2), we have $w(\alpha x) \geq \alpha w(x)$ for any $\alpha \in [0,1]$. In particular, for $x = 1$, we get $w(\alpha) \geq \alpha$ for any $\alpha \in [0,1]$. Thus, $w(\mathbb{P}(\xi \geq z)) \geq \mathbb{P}(\xi \geq z)$ for any $z \in \mathbb{R}$. By taking integrals on both sides, we conclude the result. □

Remark 2.1. *Lemma 2.1 implies that* (2.2)*, respectively* (2.3)*, is an appropriate property of the distortion function w for modelling risk averse behaviour towards random costs, respectively risk seeking behaviour, towards random profits.*

Lemma 2.2. *Let $\rho : L^\infty_+(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be the distortion operator as in* (2.1)*. Then*

(i) *ρ is positively translation invariant, i.e., $\rho(\xi + c) = \rho(\xi) + c$ for $c \geq 0$. In particular, $\rho(c) = c$ for any $c \geq 0$.*

(ii) *ρ is positively homogeneous, i.e. $\rho(\lambda\xi) = \lambda\rho(\xi)$ for $\lambda \geq 0$.*

(iii) *ρ is monotone, i.e. $\rho(\xi_1) \leq \rho(\xi_2)$ for $\xi_1, \xi_2 \in L^\infty_+(\Omega, \mathcal{F}, \mathbb{P})$ and $\xi_1 \leq \xi_2$.*

Proof. (i)

$$\begin{aligned}
\rho(\xi + c) &= \int_0^\infty w(\mathbb{P}(\xi + c \geq z))dz \\
&= \int_0^\infty w(\mathbb{P}(\xi \geq z - c))dz \\
&= \int_{-c}^0 w(\mathbb{P}(\xi \geq z))dz + \int_0^\infty w(\mathbb{P}(\xi \geq z))dz \\
&= c + \rho(\xi)
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\rho(0) &= \int_0^\infty w(P(0 \geq z))dz \\
&= \int_{\{0\}} w(P(0 \geq z))dz \\
&= \int_{\{0\}} w(1)dz = 0
\end{aligned}$$

Hence, by the first equality above, we have $\rho(0 + c) = c$ for $c \geq 0$.

(ii)

$$\begin{aligned}
\rho(\lambda\xi) &= \int_0^\infty w(\mathbb{P}(\lambda\xi \geq z))dz \\
&= \lambda \int_0^\infty w(\mathbb{P}(\xi \geq \frac{z}{\lambda}))d\frac{z}{\lambda} \\
&= \lambda\rho(\xi)
\end{aligned}$$

(iii) Since $\xi_1 \leq \xi_2$ and w is monotone, we have for any $z \geq 0$

$$\begin{aligned}
\mathbb{P}(\xi_1 \geq z) &\leq \mathbb{P}(\xi_2 \geq z) \\
w(\mathbb{P}(\xi_1 \geq z)) &\leq w(\mathbb{P}(\xi_2 \geq z))
\end{aligned}$$

Thus, we have $\rho(\xi_1) \leq \rho(\xi_2)$.

□

2.2 Dynamic Probability Distortion on Stochastic Processes

The main issue occurs when one tries to extend (2.1) to the multiperiod setting. In particular, it is not clear what the "conditional version" of distortion operator is. Hence, we first construct the corresponding operator for multitemporal dynamic setting.

Fix $T \in \mathbb{N}$ and denote $\mathcal{T} \triangleq [0, 1, \ldots, T]$ and $\widetilde{\mathcal{T}} \triangleq [0, 1, \ldots, T-1]$. Let Ω be the sample space with $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \mathcal{F}_T$ being the filtration and $\mathbb{P}$ being the probability measure on Ω such that $(\Omega, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ is the stochastic basis. Let $\xi = (\xi_t)_{t\in\mathcal{T}}$ be a discrete time non-negative stochastic process that is adapted to the filtration $(\mathcal{F}_t)_{t\in\mathcal{T}}$ and uniformly bounded that is $\sup_{t\in\mathcal{T}} \operatorname{ess\,sup}(\xi_t) < \infty$. We denote in that case $\xi \in L_+(\Omega, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$. Then, if we define for $t \in \mathcal{T}$

$$\rho_t(\xi_T) \triangleq \int_0^\infty w(\mathbb{P}(\xi_T \geq z|\mathcal{F}_t))dz,$$

we do not necessarily have

$$\rho(\xi) = \rho(\rho_t(\xi))$$

In particular, the "tower property" of expectation operator fails in distortion operators. In the context of stochastic optimization, this implies that the optimization problem becomes "time-inconsistent", i.e. the "Dynamic Programming Principle" (DPP) does not hold. On the other hand, for $w(x) = x$, the distortion operator (2.1) reduces to expectation operator, where for $\mathbb{E}_t[\xi_T] \triangleq \mathbb{E}[\xi_T|\mathcal{F}_t]$, we have $\mathbb{E}[\xi_T] = \mathbb{E}[\mathbb{E}_t[\xi_T]]$ with the towering property and DPP holds (see Example 2.1 below.)

Thus, we define first *dynamic distortion mappings* on a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\in\mathcal{T}})$ analogous to Definition 2.1 in multitemporal setting as follows.

Definition 2.2. *Let $t \in \widetilde{\mathcal{T}}$ and $\xi_{t+1} \in L_+^\infty(\Omega, \mathcal{F}_{t+1}, \mathbb{P})$. $w(\cdot)$ is the distortion function as in Definition 2.1.*

- *A one-step dynamic distortion mapping $\varrho_{t+1|t} : L_+^\infty(\Omega, \mathcal{F}_{t+1}, \mathbb{P}) \to L_+^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$ is defined as*

$$\varrho_{t+1|t}(\xi_{t+1}) \triangleq \int_0^\infty w(\mathbb{P}(\xi_{t+1} \geq z_{t+1}|\mathcal{F}_t))dz_{t+1}$$

- *A mapping $\varrho_t : L_+^\infty(\Omega, \mathcal{F}_T, \mathbb{P}) \to L_+^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$ is called a dynamic distortion mapping, if it is composition of one step dynamic distortion mappings of the form*

$$\varrho_t \triangleq \varrho_{t+1|t} \circ \ldots \circ \varrho_{T|T-1}$$

Remark 2.2. *Definition 2.2 is well defined. Indeed, let $\xi_T \in L_+^\infty(\Omega, \mathcal{F}_T, \mathbb{P})$, going backwards iterative, by properties of w and construction of ϱ_t, uniform boundedness and $\mathcal{F}_s$ measurability*

at each $s \in [t, \dots, T]$ are preserved, such that $\varrho_t(\xi_T)$ maps ξ_T to $L_+^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$. Furthermore, by construction $\varrho_s(\cdot) = \varrho_s(\varrho_t(\cdot))$ for $0 \le s \le t \le T$. In particular, it is a time-consistent operator.

Lemma 2.3. *For $t \in \mathcal{T}$, let $\varrho_t : L_+^\infty(\Omega, \mathcal{F}_T, \mathbb{P}) \to L_+^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$ be the dynamic distortion operator as in Definition 2.2 and $\xi, \xi_1, \xi_2 \in L_+^\infty(\Omega, \mathcal{F}_T, \mathbb{P})$. Then*

(i) ϱ_t is positively translation invariant, i.e., $\varrho_t(\xi + c) = \varrho_t(\xi) + c$ $\mathbb{P}$-a.s., if c is nonnegative and $\mathcal{F}_t$ measurable.

(ii) ϱ_t is positively homogeneous, i.e. $\varrho_t(\lambda\xi) = \lambda\varrho_t(\xi)$ $\mathbb{P}$-a.s. for any scalar $\lambda \ge 0$.

(iii) ϱ_t is monotone, i.e. $\varrho_t(\xi_1) \le \varrho_t(\xi_2)$ $\mathbb{P}$-a.s. for $\xi_1 \le \xi_2$, $\mathbb{P}$-a.s..

Proof. The proof is a simple modification of Lemma 2.2. □

Next, we illustrate the failure of towering property that causes time inconsistency via the following example.

Example 2.1. *Let X and Y be two i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}(X = 1) = \mathbb{P}(Y = 2) = \frac{1}{2}$ and $w(x) = x^{1/2}$. Let $\xi_1 = X$ and $\xi_2 = X + Y$, with $\mathcal{F}_1 = \sigma(X)$ and $\mathcal{F}_2 = \sigma(X, Y)$. Then*

$$\begin{aligned} \rho_{2|1}(\xi_2) &= X + \rho_{1|0}(Y) \\ &= X + w(1) + w(\frac{1}{2}), \end{aligned}$$

where we use Lemma 2.3 i) in the first equality, above. Similarly, we have

$$\begin{aligned} \rho_{1|0} \circ \rho_{2|1}(\xi_2) &= \rho(X) + w(1) + w(\frac{1}{2}) \\ &= 2w(1) + 2w(\frac{1}{2}) \\ &= 2 + 2(\frac{1}{2})^{1/2}. \end{aligned}$$

On the other hand, we have

$$\begin{aligned}\rho(\xi_2) &= \int_0^\infty w(P(\xi_2 \geq z))dz \\ &= \int_{[0,2]} w(P(\xi_2 \geq z))dz + \int_{(2,3]} w(P(\xi_2 \geq z))dz + \int_{(3,4]} w(P(\xi_2 \geq z))dz \\ &= 2w(1) + \int_{(2,3]} w(3/4)dz + \int_{(3,4]} w(1/4)dz \\ &= 2 + w(3/4) + w(1/4) \\ &= 2 + (\frac{3}{4})^{1/2} + (\frac{1}{4})^{1/2}\end{aligned}$$

Hence, $\rho_{1|0} \circ \rho_{2|1}(\xi_2) > \rho(\xi_2)$ by strict concavity of w. We further note that the two expressions would be equal to each other, if $w(x) = x$.

3 Controlled Markov Chain Framework

In this section, we are going to introduce the necessary background on controlled Markov chain (see e.g. [25]) that we are going to work on with dynamic probability distortion. We take the control model $\mathcal{M} = \{\mathcal{M}_t, t \in \mathcal{T}\}$, where we have

$$\mathcal{M}_t := (S_t \times X_t, A_t, \mathbb{K}_t, Q_t, F_t, r_t)$$

with the following components:

- $S_t \times X_t$ and A_t denote the state and action (or control) space, respectively, which are assumed to be complete seperable metric spaces with their corresponding Borel σ-algebras $\mathcal{B}(S_t \times X_t)$ and $\mathcal{B}(A_t)$. We emphasize here that the state space is composed of two spaces, S_t and X_t, where both are subsets of Polish spaces.

- For each $(s, x) \in S_t \times X_t$, let $A_t(s, x) \subset A_t$ be the set of all admissible controls in the state (s, x). We assume that $A_t(s, x)$ is compact for $t \in \mathcal{T}$ and denote

  $$\mathbb{K}_t := \{(s, x, a) : (s, x) \in S_t \times X_t,\ a \in A_t(s, x)\}$$

  as the set of feasible state-action pairs.

- We define the system function as

  $$F_t(s_t, x_t, a_t, \eta_t) \triangleq (s_{t+1}, x_{t+1}), \tag{3.1}$$

for all $t \in \widetilde{\mathcal{T}}$ with $x_t \in X_t$ and $a_t \in A_t$ with i.i.d. random variables $(\eta_t)_{t\in\widetilde{\mathcal{T}}}$ on a probability space $(Y, \mathcal{B}(Y), P^\eta)$ with values in Y that are complete separable Borel spaces. We assume that the mapping $(s, x, a) \to F(s, x, a, y)$ in (3.1) is continuous on $S_t \times X_t \times A_t$ for every $y \in Y$ at every $t \in \widetilde{\mathcal{T}}$.

- Let
$$\Omega \triangleq \otimes_{t=0}^{T}(S_t \times X_t)$$
and for $t \in \mathcal{T}$, and
$$\mathcal{F}_t = \sigma(S_0, X_0, A_0, \ldots, S_{t-1}, X_{t-1}, A_{t-1}, S_t, X_t)$$
be the filtration of increasing σ-algebras.

- Let $\mathbb{F}_t$ be the family of measurable functions and $\pi_t \in \mathbb{F}_t$ with $\pi_t : S_t \times X_t \to A_t$ for $t \in \widetilde{\mathcal{T}}$. A sequence $(\pi_t)_{t\in\widetilde{\mathcal{T}}}$ of functions $\pi_t \in \mathbb{F}_t$ is called a control policy (or simply a policy), and the function $\pi_t(\cdot, \cdot)$ is called the decision rule or control at time t. We denote by Π the set of all control policies. We denote by $\mathfrak{P}(A_t(s_t, x_t))$ as the set of probability measures on $A_t(s_t, x_t)$ for each time $t \in \widetilde{\mathcal{T}}$. A randomized Markovian policy $\pi = (\pi_t)_{t\in\widetilde{\mathcal{T}}}$ is a sequence of measurable functions such that $\pi_t(s_t, x_t) \in \mathfrak{P}(A_t(s_t, x_t))$ for all $(s_t, x_t) \in S_t \times X_t$, i.e. $\pi_t(s_t, x_t)$ is a probability measure on $A_t(s_t, x_t)$. $(\pi_t)_{t\in\widetilde{\mathcal{T}}}$ is called a deterministic policy, if $\pi_t(s_t, x_t) = a_t$ with $a_t \in A_t(s_t, x_t)$.

- Let $r_t(s_t, x_t, a_t) : S_t \times X_t \times A_t \to \mathbb{R}_+$ for $t \in \widetilde{\mathcal{T}}$ and $r_T : S_T \times X_T \to \mathbb{R}_+$ be the non-negative real-valued reward-per-stage and terminal reward function, respectively. For $(\pi_t)_{t\in\widetilde{\mathcal{T}}} \in \Pi$, we write
$$\begin{aligned} r_t(s_t, x_t, \pi_t) &\triangleq r_t(s_t, x_t, \pi_t(s_t, x_t)) \\ &\triangleq r_t(s_t, x_t, a_t). \end{aligned}$$

- For a fixed $\pi \in \Pi$ and given $x_0 \in X_0$ with $\pi_t(x_t, s_t) = a_t$, we aggregate the cumulative reward at time $t \in [1, \ldots, T-1]$ as
$$x_t \triangleq x_0 + \sum_{i=0}^{i=t-1} r_i(s_i, x_i, a_i)$$

- Let $\pi \in \Pi$ and $x_0 \in X_0$ be given. Then, there exists a unique probability measure $\mathbb{P}^\pi$ on $(\Omega, \mathcal{F})$ such that given $(s_t, x_t) \in S_t \times X_t$, a measurable set $B_{t+1} \subset S_{t+1} \times X_{t+1}$ and $(s_t, x_t, a_t) \in \mathbb{K}_t$, for any $t \in \widetilde{\mathcal{T}}$, we have
$$Q_{t+1}(B_{t+1}|s_t, x_t, a_t) \triangleq \mathbb{P}^\pi_{t+1}(x_{t+1} \in B_{t+1}|s_t, x_t, a_t, \ldots, x_0).$$

Here, $Q_{t+1}(B_{t+1}|s_t, x_t, a_t)$ is the stochastic kernel (see e.g. [23]). Namely, for each pair $(s_t, x_t, a_t) \in \mathbb{K}_t$, $Q_{t+1}(\cdot|s_t, x_t, a_t)$ is a probability measure on $S_{t+1} \times X_{t+1}$, and for each $B_{t+1} \in \mathcal{B}_{t+1}(S_{t+1} \times X_{t+1})$, $Q_{t+1}(B_{t+1}|\cdot, \cdot, \cdot)$ is a measurable function on $\mathbb{K}_t$. We remark that at each $t \in \mathcal{T}$, the stochastic kernel depends only on (s_t, x_t, a_t) rather than the whole history $(x_0, a_0, x_1, a_1, s_1, \ldots, s_t, a_t, x_t)$. By (3.1), we have

$$Q_{t+1}(B_{t+1}|s_t, x_t, a_t) = \int_Y I_{B_{t+1}}[F(s, x, a, y)]dP^\eta(y),\ B_{t+1} \in \mathcal{B}(S_{t+1} \times X_{t+1}),$$

where $I_{B_{t+1}}$ denotes the indicator function of B_{t+1}.

Assumption 3.1.

- *The reward functions $r_t(s_t, x_t, a_t)$ for $t \in \widetilde{\mathcal{T}}$ and $r_T(s_T, x_T)$ are non-negative, continuous in their arguments and uniformly bounded i.e. $0 \leq r_t(s_t, x_t, a_t) < \infty$ and $0 \leq r_T(s_T, x_T) < \infty$.*
- *The multi-function (also known as a correspondence or point-to-set function) $(s_t, x_t) \to A_t(s_t, x_t)$ is upper semi-continuous (u.s.c.). That is, if $\{s_t^m, x_t^m\} \subset S_t \times X_t$ and $\{a_t^m\} \subset A_t(s_t^m, x_t^m)$ are sequences such that $(s_t^m, x_t^m) \to (\bar{s}_t, \bar{x}_t)$, and $a_t^m \to \bar{a}_t$, then $\bar{a}_t \in A_t(\bar{s}_t, \bar{x}_t)$ for $t \in \widetilde{\mathcal{T}}$.*
- *For every state $(s, x) \in S_t \times X_t$, the admissible action set $A_t(s, x)$ is compact for $t \in \widetilde{\mathcal{T}}$.*

4 Optimal Control Problem

4.1 Main Result

For every $t \in \widetilde{\mathcal{T}}$, $(s_t, x_t) \in S_t \times X_t$ and $\pi \in \Pi$, let

$$V_t(s_t, x_t, \pi) \triangleq \varrho_t(\sum_{i=t}^{T-1} r_i(s_i, x_i, \pi_i) + r_T(s_T, x_T))$$

be the performance evaluation from time $t \in \widetilde{\mathcal{T}}$ onwards using the policy $\pi \in \Pi$ given the initial condition $(s, x) \in S_t \times X_t$. The corresponding optimal (i.e. maximal) value is then

$$V_t^*(s_t, x_t) \triangleq \sup_{\pi \in \Pi} V_t(s_t, x_t, \pi) \tag{4.1}$$

A control policy $\pi^* = (\pi_t^*)_{t \in \widetilde{\mathcal{T}}}$ is said to be optimal, if it attains the maximum in (4.1), that is

$$V_t^*(s, x) = V_t(s, x, \pi^*) \text{ for all } (s, x) \in S_t \times X_t \text{ and for } t \in \widetilde{\mathcal{T}}. \tag{4.2}$$

Thus, the optimal control problem is to find an optimal policy and the associated optimal value (4.2) for all $t \in \mathcal{T}$. We now present the main result of the paper.

Theorem 4.1. *The optimization problem* (4.1) *obeys dynamic programming principle and has an admissible policy* $\pi^* \in \Pi$. *Furthermore,* $V_t^*(s_t, x_t)$ *is continuous in its arguments.*

4.2 Proof of Theorem 4.1

To prove Theorem 4.1, we need the following key lemma.

Lemma 4.1. *Let* $\mathbb{K}$ *be defined as*

$$\mathbb{K} := \{(s, x, a) : \ \ (s, x) \in S \times X, a \in A\},$$

where $S \times X$ *and* A *are complete separable metric Borel spaces. Suppose further that* A *is compact. Let* $V : \mathbb{K} \to \mathbb{R}$ *be a nonnegative continuous function. For* $(s, x) \in S \times X$, *define*

$$V^*(s, x) \triangleq \sup_{a \in A} V(s, x, a),$$

then for any $(s, x) \in S \times X$, *there exists a* $\mathcal{B}(S \times X)$ *measurable mapping* $\pi^* : S \times X \to A$ *such that*

$$V^*(s, x) = V(s, x, \pi^*(s, x)) \tag{4.3}$$

and $V^* : S \times X \to \mathbb{R}$ *is continuous.*

Proof. By [24], we have that there exists $\mathcal{B}(S \times X)$ measurable mapping $\pi^* : S \times X \to A$ such that (4.3) holds and $V^*(s, x)$ is upper semi-continuous. But, since $V(\cdot, \cdot, \cdot)$ is continuous, $\sup_{a \in A} V(s, x, a)$ is lower semi-continuous in $(s, x) \in S \times X$, as well. Hence, $V^*(\cdot, \cdot)$ is continuous in its arguments. □

Lemma 4.2. *Suppose Assumption 3.1 holds true. Then, supremum is attained at* (4.1) *for some* $\mathcal{B}(S_t \times X_t)$ *measurable mapping* $\pi_t^*(s_t, x_t) = a_t^*$ *for* $t \in \widetilde{\mathcal{T}}$. *Furthermore, each* V_t^* *is continuous in its arguments.*

Proof. We will show only the case for $t = T - 1$. The others follow going backwards iterative down to $t = 0$. We first show that

$$(s_{T-1}, x_{T-1}, a_{T-1}) \to \int_0^\infty w(P^\eta(x_{T-1} + r_T(F(x_{T-1}, a_{T-1}, \eta_{T-1})) \geq z_T | s_{T-1}, x_{T-1}) dz_T$$

is continuous in its arguments. Let $(s_{T-1}^m, x_{T-1}^m, a_{T-1}^m) \to (s_{T-1}, x_{T-1}, a_{T-1})$ as $m \to \infty$.

Then, we have

$$\begin{aligned}
&\lim_{m\to\infty} V_{T-1}(s_{T-1}^m, x_{T-1}^m, a_{T-1}^m) \\
&= \lim_{m\to\infty} \int_0^\infty w(P^\eta(x_{T-1}^m + r_T(F(s_{T-1}^m, x_{T-1}^m, a_{T-1}^m, \eta_{T-1})) \geq z_T | (s_{T-1}^m, x_{T-1}^m))dz_T \\
&= \int_0^\infty \lim_{m\to\infty} w(P^\eta(x_{T-1}^m + r_T(F(s_{T-1}^m, x_{T-1}^m, a_{T-1}^m, \eta_{T-1})) \geq z_T | (s_{T-1}^m, x_{T-1}^m))dz_T \\
&= \int_0^\infty w(\lim_{m\to\infty} P^\eta(x_{T-1}^m + r_T(F(s_{T-1}^m, x_{T-1}^m, a_{T-1}^m, \eta_{T-1})) \geq z_T | (s_{T-1}^m, x_{T-1}^m))dz_T \\
&= \int_0^\infty w(P^\eta(\lim_{m\to\infty} x_{T-1}^m + r_T(F(s_{T-1}^m, x_{T-1}^m, a_{T-1}^m, \eta_{T-1})) \geq z_T | (s_{T-1}^m, x_{T-1}^m))dz_T \\
&= \int_0^\infty w(P^\eta(x_{T-1} + r_T(F(s_{T-1}, x_{T-1}, a_{T-1}, \eta_{T-1})) \geq z_T | (x_{T-1}, x_{T-1}))dz_T
\end{aligned}$$

The second equality follows by boundedness of $r_T()$, $w(\cdot)$ and Lebesgue dominated convergence theorem. The third equality follows by continuity of $w(\cdot)$, the fourth equality follows by continuity of probability measure, and the fifth equality follows by continuity of transition $F(\cdot,\cdot,\cdot)$ as in (3.1). Hence, $V_{T-1}(\cdot,\cdot,\cdot)$ is continuous in its arguments. The result follows by Lemma 4.1. □

Now, we are ready to prove Theorem 4.1.

Proof of Theorem 4.1. We have

$$\begin{aligned}
V_{T-1}^*(s_{T-1}, x_{T-1}) &= \sup_{\pi\in\Pi} \int_0^\infty w(P^\eta(x_{T-1} + r_T(F(x_{T-1}, \pi_{T-1}(s_{T-1}, x_{T-1}), \eta_{T-1})) \geq z_T | s_{T-1}, x_{T-1})dz_T \\
&= \int_0^\infty w(P^\eta(x_{T-1} + r_T(F(x_{T-1}, \pi_{T-1}^*(s_{T-1}, x_{T-1}), \eta_{T-1})) \geq z_T | s_{T-1}, x_{T-1})dz_T.
\end{aligned}$$

By Lemma (4.2), there exists a $\mathcal{B}(S_{T-1} \times X_{T-1})$ measurable mapping $\pi \in \Pi$ such that $\pi_{T-1}^*(s_{T-1}, x_{T-1}) = a_{T-1}^*$, and $V_{T-1}^*(s_{T-1}, x_{T-1})$ is continuous in (s_{T-1}, x_{T-1}). Hence, for $t = T-2$, we have

$$\begin{aligned}
V_{T-2}^*(s_{T-2}, x_{T-2}) = \sup_{\substack{a_{T-2}\in A_{T-2}(x_{T-2}, x_{T-2}) \\ a_{T-1}\in A_{T-1}(s_{T-1}, x_{T-1})}} \Bigg\{ \int_0^\infty w(P^\eta(V_{T-1}(x_{T-2} \\
+ r_{T-1}(F(x_{T-2}, a_{T-2}, \eta_{T-2}), a_{T-1}), F(x_{T-2}, a_{T-2}, \eta_{T-2})) \geq z_{T-1} \\
| s_{T-2}, x_{T-2})dz_{T-1} \Bigg\}
\end{aligned}$$

Plugging $\pi^*_{T-1}(x_{T-1}, s_{T-1})$ in $V^*_{T-2}(s_{T-2}, x_{T-2})$, we have

$$\begin{aligned}
V^*_{T-2}(s_{T-2}, x_{T-2}) = & \sup_{a_{T-2}\in A_{T-2}(s_{T-2},x_{T-2})} \Big\{ \int_0^\infty w(P^\eta(V^*_{T-1}(x_{T-2} \\
& + r_{T-1}(F(s_{T-2}, x_{T-2}, a_{T-2}, \xi_{T-2}), a^*_{T-1}), F(s_{T-2}, x_{T-2}, a_{T-2}, \eta_{T-2})) \geq z_{T-1} \\
& |s_{T-2}, x_{T-2}) dz_{T-1} \Big\} \\
= & \sup_{a_{T-2}\in A_{T-2}(s_{T-2},x_{T-2})} \Big\{ \int_0^\infty w(P^\eta(V^*_{T-1}(x_{T-2} \\
& + r_{T-1}(F(s_{T-2}, x_{T-2}, a_{T-2}, \eta_{T-2}), a^*_{T-1}), F(s_{T-2}, x_{T-2}, a_{T-2}, \eta_{T-2})) \geq z_{T-1} \\
& |s_{T-2}, x_{T-2}) dz_{T-1} \Big\}
\end{aligned}$$

By Lemma (4.2) again, it admits an optimal policy $a^*_{T-2} \in A_{T-2}$ such that

$$\begin{aligned}
V^*_{T-2}(s_{T-2}, x_{T-2}) = & \Big\{ \int_0^\infty w(P(V^*_{T-1}(x_{T-2} \\
& + r_{T-1}(F(s_{T-2}, x_{T-2}, a^*_{T-2}, \eta_{T-2}), a^*_{T-1}), F(s_{T-2}, x_{T-2}, a^*_{T-2}, \eta_{T-2})) \geq z_{T-1} \\
& |s_{T-2}, x_{T-2}) dz_{T-1} \Big\}
\end{aligned}$$

Going backwards iterative, we conclude that dynamic programming holds, (4.1) admits an optimal policy $\pi^* \in \Pi$ attaining supremum that depends only on on s_t and x_t at each $t \in \widetilde{\mathcal{T}}$. Furthermore, $V^*_t(\cdot, \cdot)$ is continuous again by Lemma (4.2). Hence, we conclude the proof. □

5 An Application to Portfolio Optimization

Suppose an investor has a portfolio of n stocks. The price of n stocks at $t \in \mathcal{T}$ are denoted by

$$S_t \triangleq (S^1_t, \ldots, S^n_t).$$

The price of stock $i \in [1, 2, , \ldots, n]$ at time $t \in \widetilde{\mathcal{T}}$, denoted by S^i_t, has dynamics

$$S^i_{t+1} = (1 + \delta^i) S^i_t \text{ with probability } p^i, \tag{5.1}$$

where $-1 < \delta^i < 1$ is the proportional decrement/increment of price of ith stock S^i_t. Let $P^\eta(\cdot)$ denote the joint probability mass function of S_t for $t \in \mathcal{T}$. Let $\pi = (\pi_t)_{t\in\mathcal{T}}$ be the

policy of the investor that stands for the number of shares of n stocks investor is holding at time $t \in \widetilde{\mathcal{T}}$ with

$$\pi_t : S_t \times X_t \to \mathbb{R}^n,$$

where π_t is $\mathcal{B}(S_t \times X_t)$ measurable. Here, X_t is the total wealth of the portfolio at t with $x_0 > 0$ being the initial wealth. We assume that the investor has a capacity to be in the long or short position. Namely, we take that $\|\pi_t(s,x)\| \leq C$ for some $C > 0$ for all $(s,x) \in S_t \times X_t$ and $t \in \widetilde{\mathcal{T}}$. We denote all those strategies $(\pi_t)_{t\in\widetilde{\mathcal{T}}}$ that are $\mathcal{B}(X_t, S_t)$ measurable and uniformly bounded by C by Π. We take that the market is self-financing in the sense

$$X_{t+1} = \pi_t^\intercal S_{t+1} \text{ for } t \in \widetilde{\mathcal{T}},$$

with S_{t+1} being the n-dimensional vector as defined in (5.1) such that denoting $x_{-1} \triangleq x_0$,

$$\Delta X_t \triangleq X_t - x_{t-1},$$

is the difference of the total wealth between time t and $t-1$ for $t \in \mathcal{T}$. Hence, the reward function at $t \in \mathcal{T}$ reads as

$$\begin{aligned} r_t(s_t, x_t, \pi_t) &= \Delta X_t \\ r_T(s_T, x_T) &= \Delta X_T \\ X_t &= x_{t-1} + r_t(s_t, x_t, \pi_t), \end{aligned}$$

Let $w(x) = x^2$ for $x \in [0,1]$ be the distortion of the probability function such that for a fixed π_{T-1} given $X_{T-1} = x_{T-1}$ and $S_{T-1} = s_{T-1}$, the performance measure is defined by

$$\begin{aligned} \varrho_{T-1}(X_T) &\triangleq \int_0^\infty \left(P^\eta(x_{T-1} + \pi_{T-1}^\intercal (S_T - s_{T-1}) \geq z_T | x_{T-1}, s_{T-1}) \right)^2 dz_T \\ &\triangleq V_{T-1}(s_{T-1}, x_{T-1}, \pi) \end{aligned}$$

such that

$$V_{T-1}^*(s_{T-1}, x_{T-1}) = \max_{\pi\in\Pi} V_{T-1}(s_{T-1}, x_{T-1}, \pi).$$

Hence, going backwards iterative, we have at each time $t \in \widetilde{\mathcal{T}}$

$$\begin{aligned} V_t(s_t, x_t, \pi) &= \int_0^\infty \left(P^\eta(x_t + V_{t+1}(S_{t+1}, X_{t+1}, \pi) \geq z_{t+1} | x_t, s_t) \right)^2 dz_{t+1} \\ V_t^*(s_t, x_t) &= \max_{\pi\in\Pi} V_t(s_t, x_t, \pi). \end{aligned} \tag{5.2}$$

Then, by Theorem 4.1, the dynamic programming yields an optimal strategy $(\pi_t^*)_{t\in\widetilde{\mathcal{T}}} \in \Pi$ attaining (5.2) along with the ptimal value V_t^* at each time $t \in \widetilde{\mathcal{T}}$. An important observation is that our scheme dictates that the investor should take both his current wealth x_t and the current stock price s_t into consideration while making his decision π_t on investing into the assets at $t \in \widetilde{\mathcal{T}}$.

References

[1] Bellman, R. (1952). On the theory of dynamic programming, . Proc. Natl. Acad. Sci 38, 716

[2] Chung,K., Sobel,M.J. (1987). *Discounted MDP's: Distribution functions and exponential utility maximization*. SIAM J. Control Optim. 25, 49-62.

[3] Fleming, W.H., Sheu, S.J. (1999). *Optimal long term growth rate of expected utility of wealth*. Ann. Appl. Probab. 9, 871-903.

[4] Fleming, W.H., Sheu, S.J. (2000). *Risk-sensitive control and an optimal investment model*. Math. Finance 10, 197-213.

[5] Jaquette, S.C.(1973).*Markov decision processes with a new optimality criterion: Discrete time* (1976). Ann. Stat. 1, 496-505

[6] Jaquette, S.C. (1976).*Autility criterion for Markov decision processes*. Manag.Sci.23,43-49

[7] Artzner, P., Delbaen, F., Eber, J.M., Heath, D. (1999). *Coherent measures of risk*, Math. Finance 9, 203-228.

[8] Artzner, P., Delbaen, F., Eber, J.M., Heath, D., Ku, H. (2007). *Coherent multiperiod risk adjusted values and Bellmans principle*. Ann. Oper. Res. 152, 5-22.

[9] Ruszczynski, A. (2010). *Risk-averse dynamic programming for Markov decision processes,* Math. Program. B 125 (2010) 235-261.

[10] Cheridito, P., Delbaen, F., Kupper, M (2006). *Dynamic monetary risk measures for bounded discrete-time processes*. Electron. J. Probab. 11, 57-106.

[11] Eichhorn, A. Romisch, W. (2005). *Polyhedral risk measures in stochastic programming*. SIAM J. Optim. 16, 69-95.

[12] Follmer,H.,Penner,I. (2006). *Convex risk measures and the dynamics of their penalty functions*. Stat.Decis. 24, 6196

[13] Fritelli,M.,Rosazza Gianin,E.(2002). *Putting order in risk measures*. J.Bank.Finance26,1473-1486.

[14] Frittelli,M., Rosazza Gianin,E. (2005). *Dynamic convex risk measures* (2005). Risk Measures for the 21st Century, 227-248. Wiley, Chichester

[15] Kahneman, D. and Tversky, A. (1979), *Prospect Theory: An Analysis of Decision Under Risk* Econometrica, 47, 263-292.

[16] Kahneman, D. and Tversky, A. (1992), *Advances in prospect theory: Cumulative representation of uncertainty.* Journal of Risk and Uncertainty, 5, 297-323.

[17] Zhou, X. (2010), *Mathematicalising Behavioural Finance,* (2010). Proceedings of the International Congress of Mathematicians Hyderabad, India.

[18] Wakker, P. (2010), *Prospect theory: For risk and ambiguity*, Cambdridge University Press.

[19] He, X. D., Zhou, X. Y. (2011), *Portfolio choice via quantiles.* Mathematical Finance, 21(2), 203-231.

[20] Ruszczynski, A. (2010), *Risk-averse dynamic programming for Markov decision processes*, Math. Program. B 125 (2010), 235-261.

[21] Kun, L., Cheng J., Marcus, S. I (2018), *Probabilistically distorted risk-sensitive infinite-horizon dynamic programming*, Automatica (97), 1-6.

[22] Ma, J., Wong T., Zhang, J. *Time-consistent Conditional Expectation under Probability Distortion*, preprint.

[23] Hernandez-Lerma, O.(1989), *Adaptive Markov Control Processes*, Springer-Verlag. New York.

[24] Rieder, U. (1978). *Measurable Selection Theorems for Optimisation Problems*, Manuscripta Mathematica, 24, 115-131.

[25] Hernandez-Lerma, O., Lasserre, J.B. (1996). *Discrete-time Markov control processes, in: Basic Optimality Criteria*, Springer, New York.

[26] Bjork,T., Khapko, M., Murgoci, A. (2017). *On Time-inconsistent Stochastic control in Continuous Time*, Finance and Stochastics, 21, 331-360.