Automatica 97 (2018) 1–6

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Probabilistically distorted risk-sensitive infinite-horizon dynamic programming ☆

Kun Lin [a], Cheng Jie [b], Steven I. Marcus [a,*]

[a] Electrical and Computer Engineering and Institute for Systems Research, 2466 A.V. Williams Bldg. University of Maryland College Park, MD 20742, United States

[b] Mathematics Department, 4424 Kirwan Hall. University of Maryland College Park, MD 20742, United States

ARTICLE INFO

Article history:
Received 12 March 2017
Received in revised form 25 February 2018
Accepted 10 July 2018

Keywords:
Cumulative prospect theory
Risk measures
Dynamic programming

ABSTRACT

Historically, the study of risk-sensitive criteria has focused on their normative applications — i.e., what should be done. The classic example is expected utility functions which produce deterministic policies. More recently, the literature on dynamic coherent risk measures has broadened the choices for risk-sensitive performance evaluation. However, coherent risk measures must be convex. This paper presents an alternative to both the expected utility and coherent risk measure approaches. This new approach, inspired by cumulative prospect theory (CPT), is nonconvex and has substantial empirical evidence supporting its descriptive power for human decisions, i.e., what is actually done. A key unique feature of the CPT-based approach, essential for modeling human decisions, is probabilistic distortion. Hence, CPT should be used instead of both expected utility and coherent risk measures when modeling human decisions, which requires a higher level of expressiveness than allowed by previous work. In addition, although both coherent risk measures and CPT produce randomized policies, which are more robust against inaccurate probabilistic descriptions of systems, CPT generates policies that are significantly different from those of coherent risk measures.

1. Introduction

Dynamic programming, introduced by Bellman (1952), has been the subject of intense research in the past decades. Dynamic optimization problems modeled by controlled Markov processes and solved via dynamic programming are commonly referred to as Markov decision processes (MDPs).

An important class of risk-sensitive criteria is the class of coherent risk measures, which are convex risk measures with the additional property of positive homogeneity (see Föllmer & Schied, 2008, Def. 2.3). Prominent examples include mean–semideviation and conditional value-at-risk (Artzner, Delbaen, Eber, & Heath, 1999; Delbaen & Hochschule, 2002). Recently, dynamic coherent risk measures have received much attention in the literature (Cheridito, Delbaen, & Kupper, 2004; Riedel, 2004). In particular, Ruszczyński in Ruszczyński (2010) concludes that time-consistent coherent risk measures (Ruszczyński & Shapiro, 2006) satisfy dynamic programming equations and are suitable for solving the dynamic optimization problem.

In problems involving a human decision maker, it is desirable to use criteria that are beyond expected utility and coherent risk measures. A well-known example of a non-coherent performance measure is suggested by Tversky and Kahneman in their cumulative prospect theory (CPT) (Tversky & Kahneman, 1992). Unlike both expected utility and coherent risk measures, which are normative approaches, CPT-based criteria have risen from the search for a powerful descriptive model for human decision making. Their ability to capture human decision dynamics under uncertainty (e.g., lotteries) has strong empirical support (Wakker, 2010). Although CPT had its beginning in the 1990s, its incorporation into dynamic systems is still nascent. Recently, He and Zhou (2011) have studied a portfolio choice problem using a CPT-based approach. The problem maximizes the terminal wealth of a self-financing portfolio, a constraint on the action space of the MDP, driven by a financial market that is uncontrollable from the perspective of the investor (see He & Zhou, 2011, Eq. 3). These results become more difficult, if not impossible, to obtain if these assumptions are eliminated. The motivation of this paper is to widen the application of CPT-based criteria to more general

☆ This work was supported in part by the National Science Foundation (NSF) under Grants CNS-0926194, CMMI-0856256, CNS-1446665, and CMMI-1362303, and by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-15-10050. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Valery Ugrinovskii under the direction of Editor Ian R. Petersen.

* Corresponding author.
E-mail addresses: klin@umd.edu (K. Lin), cjie@math.umd.edu (C. Jie), marcus@umd.edu (S.I. Marcus).

https://doi.org/10.1016/j.automatica.2018.07.028
0005-1098/

dynamic problems, paying attention to the structure of optimal policies obtained. The finite-horizon case was investigated in Lin and Marcus (2013).

1.1. Generalizability and challenges

In addition to CPT's ability to model human decisions and provide robust policies, it also generalizes both the expected utility and coherent risk measure approaches. Let X be a Bernoulli random variable that takes the value 1 with probability p and 0 otherwise. Since $E[u(X)] = pu(1)$, its expected utility is always linear in p. Using a typical CPT weighting function $\frac{p^\delta}{(p^\delta+(1-p)^\delta)^{\frac{1}{\delta}}}$, when $\delta = 1$, the linear case is recovered. Convex risk measures can also be recovered with appropriate weighting functions.

The expressiveness of CPT induced by probability distortion is also the source of its greatest technical challenge: in particular, the nonconvexity of the resulting risk measure; this is indeed a stark contrast to both the expected utility and coherent risk measures. The reason why previous work in this area has insisted on the convexity of the risk measure is because of the diversification principle: the fact that a portfolio is less risky than its individual parts. While this constraint makes sense when asking what a rational agent should do, it falls short when we are trying to model the way humans make decisions.

The paper is organized as follows. In Section 2, we introduce cumulative prospect theory and demonstrate the properties of CPT-based decisions. In Section 3, CPT-based criteria are applied to general dynamic problems. Our focus is on proving the suitability of dynamic programming for solving CPT-based risk-sensitive problems. In particular, we are interested in the case of discounted and transient infinite-horizon problems. Our proof strategy and conclusion have many parallels with that of Çavuş and Ruszczyński (2014).

2. Background

2.1. Cumulative prospect theory (CPT)

Prospect theory was suggested in the 1970s by Kahneman and Tversky (1979). They were unsatisfied with the theory and suggested its improved version, cumulative prospect theory (CPT), in the 1990s (Tversky & Kahneman, 1992). CPT asserts that the human decision making process can be modeled by a criterion with the following characteristics: (1) The utility function has a reference point against which gains and losses are measured; (2) The utility function is concave on gains and convex on losses (i.e., horizontal S-shape); (3) A probability weighting function (cf. Definition 1) that transforms the distribution of a probability measure such that a small probability is inflated and a large probability is deflated.

Definition 1. A probability weighting function, w, is a monotonically non-decreasing continuous function from $[0, 1]$ to $[0, 1]$ with $w(0) = 0$ and $w(1) = 1$.

Let Z be a real random variable defined on an appropriate probability space $(\Omega, \mathcal{F}, P)$; then its CPT value is calculated according to the equation

$$\rho(Z) = \int_0^\infty w_+\left(P\left(u_+\left((Z-B)_+\right) > z\right)\right) dz - \int_0^\infty w_-\left(P\left(u_-\left((Z-B)_-\right) > z\right)\right) dz, \quad (1)$$

where $u_+, u_- : \mathbf{R}^+ \to \mathbf{R}^+$ are utility functions, w_+, w_- are probability weighting functions, and B is a random variable. B is interpreted as the benchmark against which the outcomes are compared. In addition, the notations $(\cdot)_+$ and $(\cdot)_-$ denote $\max(\cdot, 0)$ and $-\min(\cdot, 0)$, respectively. Appropriate integrability assumptions are assumed.

3. Dynamic programming

In this section, CPT-based criteria will be analyzed in a general dynamic setting. The standard expected value case is time-consistent and can be rewritten as $E[g(x_0, a_0, \delta_0) + E[g(x_1, a_1, \delta_1) + \cdots |x_1] |x_0]$ where x_k, a_k, δ_k are the state, control and disturbance at time k. Here, the system evolves according to the dynamics $x_{k+1} = f(x_k, a_k, \delta_k)$, and $g(x, a, \delta)$ is the per-step cost for taking action a at state x under disturbance δ.

We are interested in nonempty Borel spaces X and A of states and controls such that for each $x \in X$ there is a nonempty feasible control Borel set $A(x) \subset A$. We denote the set of probability measures over A equipped with the Prohorov metric by $\mathcal{P}(A)$. We denote by $\mathcal{S}$ the set of all measurable functions $\mu : X \to \mathcal{P}(A)$ satisfying $\mu(x) \in \mathcal{P}(A(x))$, $\forall x \in X$, which we refer to as policies. The nonempty Borel space of disturbances is denoted by Δ, and given a state-action pair $(x_k, a_k) \in X \times A$, an element $\delta_k \in \Delta(x_k, a_k) \subset \Delta$ drives the system to its next state through a measurable function $f : X \times A \times \Delta \to X$ by $x_{k+1} = f(x_k, a_k, \delta_k)$. At each time k, a per-step cost is accumulated and denoted by a measurable function $g : X \times A \times \Delta \to \mathbf{R}$. The stochastic kernel $P(\cdot|x, a)$ is defined over $\Delta(x, a)$. Furthermore, we denote both the realization and the random variable disturbance at time k by δ_k. We denote by $R(X)$ the set of real-valued measurable functions $J : X \to \mathbf{R}$. A nonstationary Markov policy is denoted by $\pi = \{\mu_0, \mu_1, \mu_2, \dots\}$, where $\mu_k \in \mathcal{S}$ and Π denotes the set of all feasible non-stationary Markov policies.

Given an element $\bar{J} \in R(X)$, we minimize the cost over all non-stationary Markov policies, i.e.,

$$J^*(x) = \inf_{\pi \in \Pi} J_\pi(x), \text{ where}$$

$$J_\pi(x) = \limsup_{k\to\infty} \left(T_{\mu_0} T_{\mu_1} T_{\mu_2} \cdots T_{\mu_k} \bar{J}\right)(x), \quad (2)$$

for all $x \in X$, and $T_\mu : R(X) \to R(X)$ is a problem dependent operator. We define a mapping $H : X \times \mathcal{P}(A) \times R(X) \to \mathbf{R}$ such that for each policy $\mu \in \mathcal{S}$ it satisfies $\left(T_\mu J\right)(x) = H(x, \mu(x), J)$, $\forall x \in X$. We define the operator T by $(TJ)(x) = \inf_{a\in\mathcal{P}(A(x))} H(x, a, J) = \inf_{\mu\in\mathcal{S}} \left(T_\mu J\right)(x)$, $\forall x \in X$.

Eq. (2) highlights the nested operator mapping form of dynamic programming, which enables the application of Bertsekas's abstract dynamic programming (Bertsekas, 2013). Hence, the problem simplifies into proving properties of the H operator. Note that Eq. (2) is also the objective function for time-consistent coherent risk measures, albeit with a different class of H operators. We will analyze two infinite horizon problems, namely discounted and transient, using Eq. (2). For both cases, our objective is to satisfy the monotonicity and contraction assumptions in abstract dynamic programming (Bertsekas, 2013), which will yield the strongest results for dynamic programming: value and policy iteration converge to a unique value function and an optimal policy can be attained. Furthermore, according to Prop. 2.1.2 of Bertsekas (2013), such a value function attained by the optimal policy can be approximated within arbitrary accuracy by a stationary policy. We reiterate below the monotonicity and contraction assumptions for the reader's convenience (see Bertsekas, 2013, Assumptions 2.1.1 and 2.1.2).

Assumption 2 (*Monotonicity*). If $J, J' \in R(X)$ and $J \le J'$, then $H(x, a, J) \le H(x, a, J')$, $\forall x \in X$, $a \in \mathcal{P}(A(x))$.