

November 30, 2010

RT0923

Mathematics 17 pages

Research Report

Time-consistency of optimization problems

Takayuki Osogami and Tetsuro Morimura

IBM Research - Tokyo
IBM Japan, Ltd.
1623-14 Shimotsuruma, Yamato
Kanagawa 242-8502, Japan

Limited Distribution Notice

This report has been submitted for publication outside of IBM and will be probably copyrighted if accepted. It has been issued as a Research Report for early dissemination of its contents. In view of the expected transfer of copyright to an outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or copies of the article legally obtained (for example, by payment of royalties).



Time-consistency of optimization problems

Takayuki Osogami and Tetsuro Morimura
IBM Research - Tokyo
{osogami,tetsuro}@jp.ibm.com

November 28, 2010

Abstract

We study time-consistency of optimization problems, where we say that an optimization problem is time-consistent if the optimal solution, or the optimal policy for choosing actions, does not depend on when the optimization problem is solved. Time-consistency is a minimal requirement on an optimization problem for the decisions made based on the solution to the optimization problem to be rational. We show that the reward that we can gain by taking “optimal” actions selected by solving a time-inconsistent optimization problem can be surely dominated by the reward that we could gain by taking other actions that are suboptimal with respect to the time-inconsistent optimization problem. We establish sufficient conditions on the objective function and on the constraints for an optimization problem to be time-consistent. We also show when the sufficient conditions are necessary. Our results are relevant in stochastic settings particularly when the objective function is a risk measure other than expectation or when there is a constraint on a risk measure.

1 Introduction

We solve an optimization problem today and determine the actions to take today and tomorrow to maximize our benefit in future. We might solve an optimization problem tomorrow for the same purpose again, but the optimization problem to be solved tomorrow will be slightly different, because we will have obtained information that is uncertain today. Are the optimal solutions today “consistent” with the optimal solutions tomorrow? When can we guarantee that they are “consistent”? In this paper, we investigate the new concept, time-consistency of an optimization problem.

Roughly speaking, if an optimization problem is not time-consistent, the optimal actions suggested by the optimal solution today can become suboptimal (and sometimes worst) tomorrow simply because the time has passed and a piece of uncertain information is revealed. For instance, tomorrow might be sunny or rainy. Today, the weather being uncertain, the optimal solution to an optimization problem suggests that we should go picnic tomorrow, so we pack our baggage. Tomorrow, the weather will turn out to be sunny or rainy. If the optimization problem is not time-consistent, the optimal solution tomorrow can suggest that we should not go picnic no matter what the weather is. Time-consistency is thus a minimal requirement for optimization problems when they are solved at multiple periods in order for the decisions made based on their optimal solutions to be rational.

Time-consistency has been discussed in the literature of finance in the context of a risk measure (RM) [5, 6, 14]. In finance, an RM is a function that maps a random variable (risk) to a real number to quantify the riskiness of the random variable. Regulations based on a Basel accord require a bank to reserve the capital appropriate to the risk associated with its investment practices, and an RM is used to determine

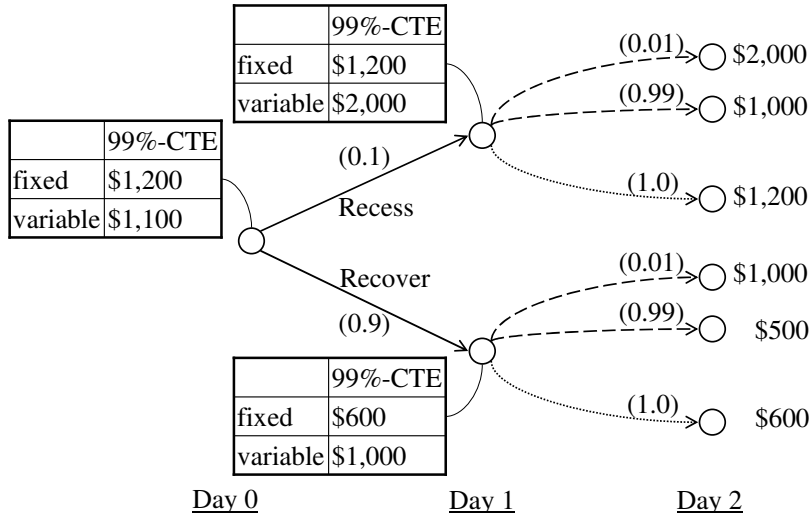


Figure 1: An example illustrating time-inconsistency of CTE. Each table in the figure shows the 99%-CTE of the amount of payment evaluated at the associated state.

the amount of the capital to be reserved. Time-consistency is widely considered to be a necessary property of such an RM, which will be discussed further in Section 2.

We find, however, that an optimization problem can be time-inconsistent (as defined in this paper) even if the objective function and the function that defines the feasible region are time-consistent (as defined in [5]), which will be discussed in Section 3. Notice that a time-consistent RM has the desirable property when it is used to quantify the risk. However, the goal of a financial institution is not simply to quantify the risk but to maximize the expected profit, or other functions of the profit, under the constraints on the risk given by regulations. For example, a regulation might require the financial institution to prove that the value of an RM is below a limit at the end of every month.

A primary contribution of this paper is a sufficient condition for an optimization problem to be time-consistent (see Section 5). We formally define time-consistency of an optimization problem (see Definition 3 and Definition 7) and prove that a certain form of an optimization problem is time-consistent (see Theorem 1 and Corollary 2). We also discuss the necessity of the sufficient conditions (see Lemma 1, Lemma 2, and Corollary 3).

2 Background: Time consistency of risk measures

Conditional tail expectation (CTE), also known as conditional value at risk, is a popular RM that has various desirable properties when it is used in a static setting (i.e., the risk is evaluated once) [4, 15], but CTE has been criticized for not being time-consistent [6]. We start by illustrating the undesirable property of a time-inconsistent RM. We will then state a formal definition of time-consistency and review examples of time-consistent RM.

Figure 1 illustrates the time-inconsistency of CTE. Suppose that, on Day 0, we purchase a product and must choose a way of payment, either “fixed” or “variable.” We know that the probability distribution of the amount of the payment is determined based on an economic condition on Day 1, which recovers with probability 0.9 and recesses otherwise. On the event of recovery, the amount of the “fixed” payment is \$600, and “variable” is \$1,000 with probability 0.01 and \$500 otherwise. On the event of recession,

“fixed” is \$1,200, and “variable” is \$2,000 with probability 0.01 and \$1,000 otherwise. Let us choose the payment-method that minimizes the 99%-CTE of the amount of payment. Evaluating the 99%-CTE on Day 0, we find that “variable” is the optimal choice, because the 99%-CTE is \$1,200 for “fixed” and \$1,100 for “variable.” On Day 1, we find out the economic condition. Given the economic condition, let us re-evaluate each payment-method to study whether the choice of “variable” was indeed optimal. If the economic condition has recovered, the 99%-CTE is \$600 for “fixed” and \$1,000 for “variable” payment (i.e., “fixed” is optimal). If the economic condition has recessed, the 99%-CTE is \$1,200 for “fixed” and \$2,000 for “variable” (i.e., “fixed” is optimal). Therefore, even though we found that “variable” was our optimal choice on Day 0, we find that “fixed” is optimal for any realization of the economic condition. Here, we assume that we cannot change the payment-methods on Day 1, so that we simply regret surely (with probability one) on Day 1 if we chose the the payment-method that was optimal on Day 0. The disagreement between the optimal choice on Day 0 and that on Day 1 is possible, because CTE is not time-consistent.

Roughly speaking, if an RM is time-consistent, the optimal choice with respect to the RM at some time m must be optimal with respect to the RM at future time $n > m$ with nonzero probability. To formally state the definition of time-consistency, it is important to understand the RM, ρ , in a dynamic setting (i.e., ρ needs to be understood as a dynamic RM). Let X be a random variable, and let $\rho_n(X)$ be value of the RM of X evaluated at time n . Note that $\rho_n(X)$ is a random variable before time n and becomes deterministic at time n , because $\rho_n(X)$ depends on the state that is random before time n and becomes deterministic at time n . In this sense, $\rho_n(X)$ is called \mathcal{F}_n -measurable, which can be understood more precisely with measure theory. Throughout, let \mathbf{Z}_I denote the set of integers in the interval I , and $\mathbf{Z}_{[a,\infty]} \equiv \mathbf{Z}_{[a,\infty)} \cup \{\infty\}$ for $a < \infty$. Formally, a dynamic RM is defined as follows:

Definition 1 Consider a filtered probability space, (Ω, \mathcal{F}, P) , such that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_N = \mathcal{F}$, where $N \in \mathbf{Z}_{[1,\infty]}$, and, if $N = \infty$, \mathcal{F}_∞ is defined as the σ -field generated by $\cup_{\ell=0}^{\infty} \mathcal{F}_\ell$. Let Y be an \mathcal{F} -measurable random variable. We say that ρ is a dynamic RM if $\rho_\ell(Y)$ is \mathcal{F}_ℓ -measurable for each $\ell \in \mathbf{Z}_{[0,N]}$.

Now, the time-consistency of a dynamic RM is formally defined as follows:

Definition 2 A dynamic RM, ρ , is called time-consistent if, for any \mathcal{F} -measurable random variables X, Y and for any $0 \leq m < n \leq N$, it holds that $\rho_n(X) \leq \rho_n(Y)$ surely implies $\rho_m(X) \leq \rho_m(Y)$ surely.

Our definition follows closely with Artzner et al. [5]. For related definitions of time-consistency (also called dynamic consistency), see [14, 6].

In our example illustrated with Figure 1, we have $N = 2$ periods. Let X_F and X_V , respectively, be the amount of payment with “fixed” and “variable.” Notice that X_F and X_V are \mathcal{F}_2 -measurable, because their values become deterministic by Day 2. Let ρ be 99%-CTE. We have seen above that “fixed” has lower 99%-CTE than “variable” on Day 1 regardless of the economic condition (i.e., $\rho_1(X_F) \leq \rho_1(X_V)$ surely). If ρ were time-consistent, we must have $\rho_0(X_F) \leq \rho_0(X_V)$ surely on Day 0. However, with ρ_0 being 99%-CTE, we have $\rho_0(X_F) > \rho_0(X_V)$. Thus, indeed 99%-CTE is not time-consistent in the sense of Definition 2.

An example of a time-consistent dynamic RM is entropic risk measure, ERM_γ , which maps a \mathcal{F} -measurable random variable, X , to an \mathcal{F}_n -measurable random variable as follows:

$$\frac{1}{\gamma} \ln \mathbf{E} [\exp(\gamma X) \mid S_n],$$

where S_n is the state at time n , and γ is the parameter that specifies the sensitivity to riskiness [1, 10]. When $\gamma \rightarrow 0$, $\text{ERM}_\gamma[X \mid S_n]$ converges to $\mathbf{E}[X \mid S_n]$, so that the decisions based on ERM_γ become risk-

neutral. The decisions become risk-averse with $\gamma > 0$ and risk-seeking with $\gamma < 0$. Another example of a time-consistent RM is iterated conditional tail expectation [6].

We can verify the time-consistency of ERM in Figure 1. Let $\gamma = 1$. On Day 0, we find that $\text{ERM}_1[X_F] = \$1,197$ and $\text{ERM}_1[X_V] = \$1,993$, so that “variable” is riskier. Let S_1 denote the economic condition on Day 1. Then $\text{ERM}_1[X_F|S_1 = \text{recovery}] = \600 and $\text{ERM}_1[X_V|S_1 = \text{recovery}] = \995 . Also, $\text{ERM}_1[X_F|S_1 = \text{recess}] = \$1,200$ and $\text{ERM}_1[X_V|S_1 = \text{recess}] = \$1,995$. Hence, “variable” is surely riskier on Day 1, which is consistent with the result on Day 0.

3 Time-inconsistent optimization problem

Time-consistent RMs, including ERM and iterated conditional tail expectation, are considered to be more desirable than time-inconsistent RMs, including CTE, in determining the capital that a bank needs to reserve given the risk associated with its investment practices [5, 6, 14]. Now, we will see undesirable outcomes if a decision maker seeks to minimize expected loss (or equivalently to maximize expected profit) when there is a regulation that requires that the decision maker keep the value of the ERM of the loss below a threshold.

Specifically, consider the optimization problem of minimizing the expected loss, X , under the constraint on risk that $\text{ERM}_\gamma[X]$ is below a threshold, δ :

$$\begin{aligned} \min. \quad & \mathbb{E}[X] \\ \text{s.t.} \quad & \text{ERM}_\gamma[X] \leq \delta. \end{aligned} \tag{1}$$

We will see that solving the optimization problem (1) at multiple epochs can lead to contradicting decisions over time (i.e., (1) is not necessarily time-consistent), even though \mathbb{E} and ERM_γ are both time-consistent RMs.

To demonstrate why an optimization problem must be time-consistent, we return to the example illustrated by Figure 1 with the following modifications. We can now change the payment method on Day 1 from “fixed” to “variable” or vice versa. On the event of recession, however, a fee of $x = \$300$ is required to make the change. On the event of recovery, no fee is required for the change. Figure 2 illustrates this modified example. Because the amount of payment on Day 2 depends not only on the economic condition on Day 1 but also on the payment method chosen on Day 0, we have the following three states on Day 1:

- State a : “variable” is chosen on Day 0, and the economic condition has recessed,
- State b : “fixed” is chosen on Day 0, and the economic condition has recessed,
- State c : the economic condition has recovered.

We consider the optimization problem (1) with $\gamma = 0.01$ and $\delta = \$1,500$ in the setting of Figure 2, where X denotes the amount of payment including the fee for changing payment-methods. Solving (1) on Day 0, we find that the optimal policy, π_0 , is to choose “variable” on Day 0 and does not change the payment-method on Day 1 regardless of the economic condition. With π_0 , we have $\mathbb{E}^{\pi_0}[X] \approx \555 and $\text{ERM}_{0.01}^{\pi_0}[X] \approx \$1,309 \leq \delta = \$1,500$. Following π_0 , we choose “variable” on Day 0. Although π_0 suggests that we should not change the payment-method on Day 1, let us solve (1) on Day 1 to verify that we indeed should not change the payment-method. Suppose that the economic condition recesses and we transition into state $S_1 = a$, whose probability is 0.1 with π_0 . Surprisingly, we find that “variable” is infeasible (and so is π_0) from $S_1 = a$, because $\text{ERM}_{0.01}^{\pi_0}[X | S_1 = a] \approx \$1,539 > \$1,500$. Therefore, we must pay the fee and change the payment-method to “fixed.” That is, the optimal policy that we find by solving (1)

problem to be time-consistent.

4 Time-consistency of an optimization problem

In this section, we define the time-consistency of an optimization problem. Suppose that the goal of a decision maker is to maximize the worthiness of a random quantity, X , while keeping the riskiness associated with X at an acceptable level. The value of X will be determined in future, on or before time N , but is random before that. To achieve the goal, the decision maker solves an optimization problem at time 0 to select the policy, π_0 , that determines the action to take for each state at each time $n \in \mathbf{Z}_{[0,N]}$. At a future time $\ell \in \mathbf{Z}_{[1,N]}$, the decision maker might solve an optimization problem that can depend on the state, S_ℓ , at time ℓ , where S_ℓ can include the belief of the decision maker, conditions of the environment, and other information available to the decision maker by time ℓ . We want the policy selected by solving the optimization problem at time ℓ to be consistent with π_0 .

We allow N to be either finite or infinite. There are multiple ways to interpret X when $N = \infty$. For example, we can see X as a random quantity, X_τ , that is determined at a stopping time, τ , when S_τ becomes a particular state, where τ is finite almost surely but unbounded. Notice that S_n can include some information about X_n . Alternatively, we can see X as a limiting value. For example, $X \equiv \lim_{n \rightarrow \infty} X_n$ for $X_n \equiv (R_0 + \dots + R_{n-1})/n$, where R_ℓ is a random reward that the decision maker receives at time ℓ .

4.1 Process of optimization problem

More formally, for $n \in \mathbf{Z}_{[0,N]}$, let $\mathbf{P}_n(s)$ be the optimization problem that a decision maker solves at time n under the condition that $S_n = s \in \mathbf{S}_n$, where \mathbf{S}_n is the set of all of the possible states at time n . Specifically, we consider the optimization problem, $\mathbf{P}_n(s)$, of the following form:

$$\mathbf{P}_n(s) : \begin{array}{ll} \max_{\pi \in \Pi_n} & f_n(X^\pi(s, n)) \\ \text{s.t.} & g_n(X^\pi(s, n)) \in B_n, \end{array} \quad (2)$$

where $X^\pi(s, n)$ denotes the conditional random variable, X , given that $S_n = s$ and π is the policy used on and after time n . The decision maker seeks to find the optimal policy from the candidate set, Π_n , where $|\Pi_n| > 1$. A policy, $\pi_n \in \Pi_n$, determines an action, a_ℓ , to take for each state, $s_\ell \in \mathbf{S}_n$, at each time, $\ell \in \mathbf{Z}_{[n,N]}$. For two distinct policies, $\pi_n, \pi'_n \in \Pi_n$, we assume that the actions taken with π_n and those with $\pi'_n \neq \pi_n$ differ for at least one state, $s_\ell \in \mathbf{S}_\ell, \ell \in \mathbf{Z}_{[n,N]}$.

In (2), the objective function, f_n , is an RM that maps a conditional random variable, $X^\pi(s, n)$, to a real number. For example, setting $f_n(X^\pi(s, n)) = \mathbf{E}[X^\pi(s, n)]$, the decision maker can select the optimal policy from Π_n that maximizes the expected value of the random quantity, X , given that $S_n = s$.

The constraint in (2) specifies the acceptable riskiness at time n . Here, g_n is a multidimensional function that maps $X^\pi(s, n)$ to real numbers, and B_n specifies the feasible region in the codomain of g_n . For example, the constraint could be specified with an entropic risk measure such that

$$\text{ERM}_\gamma[-X^\pi(s, n)] \leq b_n,$$

where b_n denotes the upper bound of the acceptable risk. Here, the value of ERM represents a magnitude of a loss, so that the negative sign is appended to the reward X . Recall, however, that the optimization problem is not necessarily time-consistent even when the objective function is expectation and the constraint is on the entropic risk measure.

Observe that the optimization problems that the decision maker is solving can be seen as a stochastic process, $\text{POP}(X, \mathbf{S}, p)$, which we will refer to as a Process of Optimization Problems (POP):

$$\text{POP}(X, \mathbf{S}, p) : \begin{array}{ll} \max_{\pi \in \Pi} & f(X^\pi) \\ \text{s.t.} & g(X^\pi) \in B. \end{array} \quad (3)$$

Here, X^π denotes the conditional random variable of X given that policy π is used. For simplicity, we assume that the initial state S_0 at time 0 is known to be $s_0 \in \mathbf{S}_0$ (i.e., $|\mathbf{S}_0| = 1$), but it is trivial to extend our results to the case with $|\mathbf{S}_0| > 1$. For $n \in \mathbf{Z}_{[0, N)}$, the decision maker solves $P_n(s)$ at time n under the condition that $S_n = s \in \mathbf{S}_n$. Note that the optimal policy at time n depends on how the state transitions after time n (i.e., the set of transition probabilities, $\{p_m \mid \ell \in \mathbf{Z}_{[n, N)}\}$, where

$$p_m \equiv \left\{ p_\ell^\pi(s' \mid s) \mid s \in \mathbf{S}_\ell, s' \in \mathbf{S}_{\ell+1}, \ell \in \mathbf{Z}_{[m, N)}, \pi \in \Pi_n \right\},$$

and $p_\ell^\pi(s' \mid s)$ denotes the probability that $S_{\ell+1} = s' \in \mathbf{S}_{\ell+1}$ under the condition that $S_\ell = s \in \mathbf{S}_\ell$ and $\pi \in \Pi_n$ is used to decide actions to take). Namely, the decision maker knows or assumes the set of transition probabilities, $\{p_m \mid \ell \in \mathbf{Z}_{[n, N)}\}$, when he solves $P_n(s)$. In (3), $\mathbf{S} \equiv \{\mathbf{S}_n \mid n \in \mathbf{Z}_{[0, N)}\}$ represents the state space, and $p \equiv \{p_n \mid n \in \mathbf{Z}_{[0, N)}\}$ represents the set of transition probabilities that the decision maker uses in solving the optimization problems.

For simplicity, we assume that the state $s_n \in \mathbf{S}_n$ includes all of the information about the history of prior states, $s_m \in \mathbf{S}_m$ for $m \in \mathbf{Z}_{[0, n)}$, and the actions taken before time n . Then we limit Π to be the set of Markovian policies, with which the action to take at $s \in \mathbf{S}_n$ depends only on s (i.e., the action is conditionally independent of the history of the prior states and actions given s).

We assume that the decision maker knows the state at any moment, so that he solves $P_n(s)$ if he knows that the state is $s \in \mathbf{S}_n$ at time $n \in \mathbf{Z}_{[0, N)}$. Notice, however, that the state might just be the belief of the decision maker, and in that case he only knows what he believes and the relevant history. Alternatively, the state might represent the conditions of the environment, and in that case our assumption implies that the decision maker knows the exact conditions of the environment.

4.2 Time-consistent process of optimization problems

Our goal is to determine whether the optimization problems that the decision maker is solving lead to contradicting decisions over time. Toward that end, we define time-consistency of a POP. We start with the case where the distribution of X and p , which the decision maker is using in solving the optimization problems, are known to a verifier who determines whether a POP is time-consistent. In Section 5.3, we will show that the results for this case can be easily translated into those for the case where the distribution of X and p are unknown to the verifier.

Consider a decision maker who takes actions based on the optimal policy for a Markov decision process (MDP). The decision maker might want to verify that the POP associated with the problem of finding the optimal policy for the MDP is time-consistent. Because the decision maker is a verifier, the verifier knows the distribution of X and p that the decision maker is using.

At time 0, the decision maker finds the optimal policy, π_0^* , for the MDP that starts from s_0 . At time 0, π_0^* is most appealing to the decision maker, because the constraint, $g_0(X^{\pi_0^*}) \in B_0$, is satisfied, and the value of the objective function cannot be made greater than $f_0(X^{\pi_0^*})$ by any feasible policy $\pi \in \Pi_0$. Notice that π_0^* can be used to determine the action that the decision maker should take for any $s_n \in \mathbf{S}_n$ at any time $n \in \mathbf{Z}_{[0, N)}$. We expect that π_0^* continues to be the most appealing policy to the decision maker at any time $n \in \mathbf{Z}_{[1, N)}$.

Suppose, however, that the decision maker reevaluates the optimality of π_0^* at time 1. The state at time 1 can be any state $s_1 \in \mathbf{S}_1$ such that $p_0^{\pi_0^*}(s_1 | s_0) > 0$. The decision maker finds the optimal policy, $\pi_1^*(s_1)$, for the MDP that starts from an s_1 at time 1. The constraint, $g_1(X^{\pi_1^*(s_1)}(s_1, 1)) \in B_1$, is satisfied, and no feasible policy can make the objective function greater than $f_1(X^{\pi_1^*(s_1)}(s_1, 1))$. Notice that the optimization problem that the decision maker solved at time 0 is different from that the decision maker solves at time 1, so that the optimal policy found at time 1 could be different from that found at time 0. Our expectation is, however, that π_0^* is one of the optimal policies for the MDP that starts at time 1 from the s_1 , if the associated POP is time-consistent.

We formally define time-consistency of the POP as follows:

Definition 3 *We say that $\text{POP}(X, \mathbf{S}, p)$ is time-consistent if the following property is satisfied. For any $n \in \mathbf{Z}_{[1, N]}$, if π^* is optimal from $s \in \mathbf{S}_{n-1}$ (i.e., π^* is an optimal solution to $\mathbf{P}_{n-1}(s)$), then π^* is optimal from any $s' \in \mathbf{S}_n$ such that $p_{n-1}^{\pi^*}(s' | s) > 0$.*

Observe that Definition 3 matches with our intuition about optimizing a standard MDP, where the optimal policy found at time 0 is optimal at any time $n \in \mathbf{Z}_{[0, N]}$.

We now revisit the optimization problem studied in Section 3. The optimal policy, π_0 , that we find by solving the optimization problem at time 0 becomes infeasible (hence not optimal) for the optimization problem at time 1 if the state at time 1 is a . The transition probability to state a is 0.1, which is strictly positive. Hence, the optimization problem is indeed time-inconsistent in the sense of Definition 3.

5 Conditions for time-consistent optimization problem

In this section, we provide conditions that the objective function and the constraints should satisfy so that the POP is time-consistent.

5.1 Definitions about dynamic risk measures

Given π , the objective function can be seen as a dynamic RM (recall Definition 1). In our context, X^π is a random variable before time N , but its value becomes deterministic by time N , so that X^π is \mathcal{F}_N -measurable. Because S_n is random before time n , the value of the objective function, $f_n(X^\pi(S_n, n))$, is random before it becomes deterministic at time n . Hence, the value of the objective function to be evaluated at time n is \mathcal{F}_n -measurable, so that the objective function is a dynamic RM.

We will use the following definition to provide a sufficient condition for an objective function so that the associated POP is time-consistent:

Definition 4 *A dynamic RM, ρ , is called optimality-consistent if the following property is satisfied: for any \mathcal{F} -measurable random variables, Y and Z , and for any $n \in \mathbf{Z}_{[1, N]}$, if $\Pr(\rho_n(Y) \leq \rho_n(Z)) = 1$ and $\Pr(\rho_n(Y) < \rho_n(Z)) > 0$, then $\Pr(\rho_{n-1}(Y) < \rho_{n-1}(Z)) > 0$. Also, we say that ρ is optimality-consistent for a particular n if the above property is satisfied for the n .*

We can define a dynamic RM, $h(X^\pi)$, associated with the constraints, $g(X^\pi) \in B$, in a POP. Specifically, for each $n \in \mathbf{Z}_{[0, N]}$, let $h_n(X^\pi) = 1$ if $g_n(X^\pi) \in B_n$ and $h_n(X^\pi) = 0$ otherwise. Observe that $h(X^\pi)$ is a dynamic RM, because $g_n(X^\pi)$ and B_n are random before time n but becomes deterministic by time n . We will use the following definition to provide a sufficient condition for constraints so that the associated POP is time-consistent:

Definition 5 Let Y be an \mathcal{F} -measurable random variable. A dynamic RM, ρ , is called non-decreasing if $\Pr(\rho_{n-1}(Y) \leq \rho_n(Y)) = 1$ for any $n \in \mathbf{Z}_{[1,N]}$. Also, we say that ρ is non-decreasing for a particular n if the above property is satisfied for the n .

5.2 Sufficient condition

We are now ready to state a sufficient condition for an POP to be time-consistent. Let $\mathbb{I}\{\cdot\}$ be an indicator random variable.

Theorem 1 If f is an optimality-consistent dynamic RM and $h(\cdot) \equiv \mathbb{I}\{g(\cdot) \in B\}$ is a non-decreasing dynamic RM, then $\text{POP}(X, \mathbf{S}, p)$ as defined with (3) is time-consistent.

Proof: We will prove the contrapositive of the property of the time-consistent POP: if π is not optimal from $s_n \in \mathbf{S}_n$, then π is not optimal from any s_{n-1} such that $p_{n-1}^\pi(s_n|s_{n-1}) > 0$. Suppose that π is not optimal from $s_n \in \mathbf{S}_n$. Notice that π might be either infeasible or feasible but not optimal.

We first consider the case where π is infeasible from s_n . If $p_{n-1}^\pi(s_n|s_{n-1}) > 0$ for an $s_{n-1} \in \mathbf{S}_{n-1}$, then π must be infeasible from the s_{n-1} ; this is because $h(\cdot)$ is a non-decreasing dynamic RM and $h_n(X^\pi(s_n, n)) = 0$, so that $h_{n-1}(X^\pi(s_{n-1}, n-1)) = 0$. Hence, the contrapositive of the property of the time-consistent POP holds.

In the rest of the proof, we consider the case where π is feasible but not optimal from s_n . Then there exists an optimal policy, $\pi^* \neq \pi$, from state s_n , and we have

$$f_n(X^\pi(s_n, n)) < f_n(X^{\pi^*}(s_n, n)).$$

Now, consider a policy π' that assigns the same actions as those assigned by π^* for all of the the states reachable from s_n . For all of the other states, π' assigns the same actions as those assigned by π . Notice that a state $s' \in \mathbf{S}_\ell$ for $\ell \in \mathbf{Z}_{[n+1,N]}$ is not reachable from more than one $s \in \mathbf{S}_n$, because a state includes the information about the history of the prior states.

Because of the way how π' is constructed, we observe that π' is feasible from s_n . Hence, π' is feasible from s_{n-1} if $p_{n-1}^\pi(s_n|s_{n-1}) > 0$, because $h(\cdot)$ is non-decreasing.

We will show that π is not optimal from s_{n-1} if $p_{n-1}^\pi(s_n|s_{n-1}) > 0$ by establishing that

$$f_{n-1}(X^\pi(s_{n-1}, n-1)) < f_{n-1}(X^{\pi'}(s_{n-1}, n-1))$$

for any s_{n-1} such that $p_{n-1}^\pi(s_n|s_{n-1}) > 0$. Observe that

$$f_n(X^\pi(s_n, n)) < f_n(X^{\pi'}(s_n, n)), \tag{4}$$

and

$$f_n(X^\pi(s, n)) = f_n(X^{\pi'}(s, n)) \tag{5}$$

for all $s \in \mathbf{S}_n$ such that $s \neq s_n$. Now, if $p_{n-1}^\pi(s_n|s_{n-1}) > 0$, then, by the optimality-consistency of the objective function, we must have

$$f_{n-1}(X^\pi(s_{n-1}, n-1)) < f_{n-1}(X^{\pi'}(s_{n-1}, n-1)), \tag{6}$$

which can be formally proved as follows. Let $Y \equiv X^\pi(s_{n-1}, n-1)$ and $Z \equiv X^{\pi'}(s_{n-1}, n-1)$. Notice that $f_n(Y)$ is a random variable that takes value $f_n(X^\pi(s, n))$ with probability $p_{n-1}^\pi(s|s_{n-1})$ for all $s \in \mathbf{S}_n$, and

$f_n(Z)$ is a random variable that takes value $f_n(X^{\pi'}(s, n))$ with probability $p_{n-1}^{\pi'}(s|s_{n-1})$ for all $s \in \mathbf{S}_n$. From the observations made with (4) and (5), we find $\Pr(f_n(Y) \leq f_n(Z)) = 1$ and $\Pr(f_n(Y) < f_n(Z)) > 0$. Hence, the optimality-consistency of the objective function implies that $\Pr(f_{n-1}(Y) < f_{n-1}(Z)) > 0$. However, $\Pr(f_{n-1}(Y) < f_{n-1}(Z))$ is 0 or 1. Therefore, (6) is established. ■

We elaborate on the sufficient conditions in the rest of Section 5.2.

5.2.1 Optimality-consistent objective functions

First, we remark that expectation and ERM are optimality-consistent.

Proposition 1 *Expectation and ERM are optimality-consistent.*

Proof: To verify that expectation is optimality-consistent for an arbitrary $n \in \mathbf{Z}_{[1, N]}$, consider \mathcal{F} -measurable random variables, Y and Z , such that

$$\Pr(\mathbb{E}[Y|S_n] \leq \mathbb{E}[Z|S_n]) = 1 \quad \text{and} \quad \Pr(\mathbb{E}[Y|S_n] < \mathbb{E}[Z|S_n]) > 0,$$

where S_n denotes the state at time n . It suffices to show $\Pr(\mathbb{E}[Y|S_{n-1}] < \mathbb{E}[Z|S_{n-1}]) > 0$. Letting $D_n \equiv \mathbb{E}[Z|S_n] - \mathbb{E}[Y|S_n]$, we can claim that there exist $\varepsilon > 0$ and $\delta_\varepsilon > 0$ such that

$$\Pr(0 \leq D_n < \varepsilon) = 1 - \delta_\varepsilon \quad \text{and} \quad \Pr(D_n \geq \varepsilon) = \delta_\varepsilon.$$

Then we have

$$\begin{aligned} \mathbb{E}[Z|S_{n-1}] - \mathbb{E}[Y|S_{n-1}] &= \mathbb{E}[\mathbb{E}[Z|S_n]|S_{n-1}] - \mathbb{E}[\mathbb{E}[Y|S_n]|S_{n-1}] \\ &= \mathbb{E}[D_n|S_{n-1}] \\ &\geq \mathbb{E}[\varepsilon \mathbb{I}\{\varepsilon \geq D_n\} | S_{n-1}] \\ &= \varepsilon \mathbb{E}[\mathbb{I}\{\varepsilon \geq D_n\} | S_{n-1}], \end{aligned}$$

which is positive with positive probability (i.e., $\Pr(\mathbb{E}[Y|S_{n-1}] < \mathbb{E}[Z|S_{n-1}]) > 0$), because

$$0 < \delta_\varepsilon = \Pr(D_n \geq \varepsilon) = \mathbb{E}[\mathbb{I}\{D_n \geq \varepsilon\}] = \mathbb{E}[\mathbb{E}[\mathbb{I}\{D_n \geq \varepsilon\} | S_{n-1}]].$$

We can verify that ERM is optimality-consistent analogously to expectation. For an arbitrary n and for $\gamma \neq 0$, we will show $\Pr(\text{ERM}_\gamma[Y|S_{n-1}] < \text{ERM}_\gamma[Z|S_{n-1}]) > 0$ for \mathcal{F} -measurable random variables, Y and Z , that satisfy

$$\Pr(\text{ERM}_\gamma[Y|S_n] \leq \text{ERM}_\gamma[Z|S_n]) = 1 \quad \text{and} \quad \Pr(\text{ERM}_\gamma[Y|S_n] < \text{ERM}_\gamma[Z|S_n]) > 0. \quad (7)$$

Letting $D_n \equiv \text{ERM}_\gamma[Z|S_n] - \text{ERM}_\gamma[Y|S_n]$, we have

$$\begin{aligned} \text{ERM}_\gamma[Z|S_{n-1}] - \text{ERM}_\gamma[Y|S_{n-1}] &= \text{ERM}_\gamma[\text{ERM}_\gamma[Z|S_n]|S_{n-1}] - \text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n]|S_{n-1}] \\ &= \text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n] + D_n|S_{n-1}] - \text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n]|S_{n-1}]. \quad (8) \end{aligned}$$

where the first equality follows from the recursiveness of ERM (specifically, $\text{ERM}_\gamma[V|S_{n-1}] = \text{ERM}_\gamma[\text{ERM}_\gamma[V|S_n]|S_{n-1}]$ for any random variable V), and the second equality follows from the definition of D_n . Now, consider a random variable, D'_n , that is dominated by D_n in such a way that

$$D'_n = \begin{cases} 0 & \text{if } 0 \leq D_n < \varepsilon \\ \varepsilon & \text{if } D_n \geq \varepsilon. \end{cases}$$

By (7), we can claim that there exist $\varepsilon > 0$ and $\delta_\varepsilon > 0$ such that $\Pr(D'_n = 0) = 1 - \delta_\varepsilon$ and $\Pr(D'_n = \varepsilon) = \delta_\varepsilon$. By the monotonicity of ERM (specifically, $\text{ERM}_\gamma[V] \leq \text{ERM}_\gamma[W]$ if W stochastically dominates V), we have $\text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n] + D_n] \geq \text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n] + D'_n]$. Hence, by (8), it suffices to show $\text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n] + D'_n] > \text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n]]$ to establish $\Pr(\text{ERM}_\gamma[Y|S_{n-1}] < \text{ERM}_\gamma[Z|S_{n-1}]) > 0$. Now,

$$\begin{aligned} \text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n] + D'_n] &= \frac{1}{\gamma} \ln \left(\delta \mathbb{E} \left[e^{\gamma(\text{ERM}_\gamma[Y|S_n] + \varepsilon)} \mid D'_n = \varepsilon \right] + (1 - \delta) \mathbb{E} \left[e^{\gamma \text{ERM}_\gamma[Y|S_n]} \mid D'_n = 0 \right] \right) \\ &= \frac{1}{\gamma} \ln \left(\delta e^{\gamma \varepsilon} \mathbb{E} \left[e^{\gamma \text{ERM}_\gamma[Y|S_n]} \mid D'_n = \varepsilon \right] + \mathbb{E} \left[e^{\gamma \text{ERM}_\gamma[Y|S_n]} \right] \right) \\ &> \frac{1}{\gamma} \ln \left(\mathbb{E} \left[e^{\gamma \text{ERM}_\gamma[Y|S_n]} \right] \right) \\ &= \text{ERM}_\gamma[\text{ERM}_\gamma[Y|S_n]], \end{aligned}$$

where the equality follows from $\delta > 0$ and strict monotonicity of \ln . \blacksquare

Our definition of the optimality-consistency of a dynamic RM should be compared against the definition of the time-consistency of a dynamic RM (Definition 2). In particular, we find that a POP with a time-consistent objective-function and no constraints is not necessarily time-consistent.

Specifically, we consider an iterated conditional tail expectation (ICTE) in the context of the example illustrated with Figure 2. ICTE is a dynamic RM that is known to be time-consistent [6]. The 99%-ICTE of the amount of payment X with π_0 is calculated as follows, where recall that, with π_0 , we choose “variable” on Day 0 and do not change the payment-method on Day 1 regardless of the economic condition. When there is only one remaining period (i.e., on Day 1), ICTE is equivalent to CTE (i.e., $\text{ICTE}_{0.99}^{\pi_0}[X|S_1] = \text{CTE}_{0.99}^{\pi_0}[X|S_1]$). In particular, $\text{ICTE}_{0.99}^{\pi_0}[X|S_1 = a] = \$2,000$ and $\text{ICTE}_{0.99}^{\pi_0}[X|S_1 = c] = \$1,000$. On Day 0, S_1 is random, so that $\text{ICTE}_{0.99}^{\pi_0}[X|S_1]$ is a random variable taking value \$2,000 with probability 0.1 and \$1,000 with probability 0.9. Hence, on Day 0, we can evaluate the riskiness of the random variable $\text{ICTE}_{0.99}^{\pi_0}[X|S_1]$ with CTE: namely, $\text{CTE}_{0.99}^{\pi_0}[\text{ICTE}_{0.99}^{\pi_0}[X|S_1]]$, which is defined to be the 99%-ICTE of X when it is evaluated on Day 0 (i.e., $\text{ICTE}_{0.99}^{\pi_0}[X] \equiv \text{CTE}_{0.99}^{\pi_0}[\text{ICTE}_{0.99}^{\pi_0}[X|S_1]]$). In particular, $\text{ICTE}_{0.99}^{\pi_0}[X] = \$2,000$. Notice that ICTE is different from CTE, because $\text{CTE}_{0.99}^{\pi_0}[X] = \$1,100$.

Now, let the exchange fee be $x = \$800$ to illustrate that an optimization problem whose objective function is ICTE is not necessarily time-consistent. Recall that π'_0 is the policy where we chose “variable” on Day 0 and change the payment-method to “fixed” on Day 1 if and only if the economic condition recesses. Then we have $\text{ICTE}_{0.99}^{\pi'_0}[X] = \$2,000$, which is equal to $\text{ICTE}_{0.99}^{\pi_0}[X]$. Because π_0 and π'_0 are equally appearing with respect to $\text{ICTE}_{0.99}[X]$ on Day 0, we could choose π'_0 as our optimal policy. However, if the economic condition recesses on Day 1, the action suggested by π'_0 is clearly inferior to that suggested by π_0 , because the amount of payment with π'_0 is \$2,000 surely, while π_0 results in paying \$1,000 with probability 0.99 and \$2,000 with probability 0.01.

Formally, ICTE is a particular iterated RM, which is defined as follows:

Definition 6 Consider a filtered probability space, (Ω, \mathcal{F}, P) , such that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_N = \mathcal{F}$, where $N \in [1, \infty)$. Let Y be an \mathcal{F} -measurable random variable. We say that $\bar{\rho}$ is an iterated RM if $\bar{\rho}_N[Y] = Y$ and $\bar{\rho}_n[Y] = \rho_n[\bar{\rho}_{n+1}[Y]]$, where ρ_n is a conditional RM that maps an \mathcal{F}_{n+1} -measurable random variable to an \mathcal{F}_n -measurable random variable, for $n \in [0, N)$.

Notice that expectation is an iterated RM, where $\rho_n[\cdot] = \mathbb{E}[\cdot|S_n]$. Also, ERM is an iterated RM, where $\rho_n[\cdot] = \text{ERM}[\cdot|S_n]$. Now, the proof of Proposition 1 implies the following sufficient condition for an iterated

RM to be optimality-consistent:

Corollary 1 *An iterated RM, as defined in Definition 6, is optimality-consistent for a particular n if the following property is satisfied: for any \mathcal{F} -measurable random variables, Y and D , such that $\Pr(D \geq 0) = 1$ and $\Pr(D > 0) > 0$, we have $\rho_n[Y + D] > \rho_n[Y]$.*

Proof: We will show $\Pr(\bar{\rho}_{n-1}[Y] < \bar{\rho}_{n-1}[Z]) > 0$ for \mathcal{F} -measurable random variables, Y and Z , that satisfy

$$\Pr(\bar{\rho}_n[Y] \leq \bar{\rho}_n[Z]) = 1 \quad \text{and} \quad \Pr(\bar{\rho}_n[Y] < \bar{\rho}_n[Z]) > 0. \quad (9)$$

Letting $D_n \equiv \bar{\rho}_n[Z] - \bar{\rho}_n[Y]$, we have

$$\begin{aligned} \bar{\rho}_{n-1}[Z] - \bar{\rho}_{n-1}[Y] &= \bar{\rho}_{n-1}[\rho_n[Z]] - \bar{\rho}_{n-1}[\rho_n[Y]] \\ &= \bar{\rho}_{n-1}[\rho_n[Y] + D_n] - \bar{\rho}_{n-1}[\rho_n[Y]] \\ &> 0, \end{aligned}$$

where the first equality follows from the definition of iterated RM, the second equality follows from the definition of D_n , and the last inequality follows from the assumption of the corollary and (9). ■

5.2.2 Non-decreasing constraints

Simple examples of the constraints that make a POP time-consistent include $\max(X^\pi) \leq \delta$ and $\min(X^\pi) \geq \delta$, where $\max(X^\pi)$ (respectively, $\min(X^\pi)$) denote the maximum (respectively, minimum) value that X^π can take with positive probability. Notice that $\max(X^\pi)$ is non-increasing over time for any sample path, because we obtain more information about X^π (specifically, about the maximum possible value of X^π) as time passes. Therefore, $\mathbb{I}\{\max(X^\pi) \leq \delta\}$ is non-decreasing. Analogous argument holds for $\mathbb{I}\{\min(X^\pi) \geq \delta\}$.

We have seen with Figure 2 that the POP of formulation (1) is not necessarily time-consistent. We can now see that the inconsistency is due to the constraint in (1), because the objective function in (1) is optimality-consistent. Observe that $\text{ERM}_{0,01}^{\pi_0}[X]$ increases from \$1,310 on Day 0 to \$1,540 on Day 1 if the state transition to $S_1 = a$ on Day 1. Hence, $\mathbb{I}\{\text{ERM}_{0,01}[\cdot] \leq \$1,500\}$ is not non-decreasing.

A way to modify (1) into a time-consistent POP is to add additional constraints on the ERM to be evaluated on each day:

$$\begin{aligned} \min. \quad & \mathbb{E}[X] \\ \text{s.t.} \quad & \text{ERM}_\gamma[X \mid S_\ell = s_\ell] \leq \delta, \quad \forall s_\ell \in \mathbf{S}, \forall \ell \in \mathbf{Z}_{[0,N]}, \end{aligned} \quad (10)$$

Then π_0 becomes infeasible for the optimization problem to be solved on Day 0, which resolves the issue of the time-inconsistency.

5.3 Extensions for an unknown environment

In the rest of Section 5, we consider the case where the verifier does not know the distribution of X and p that the decision maker is using. This case is relevant when the decision maker estimates p , but his estimation is unknown to the verifier. If the verifier does not know p , then she does not know the distribution of X , because X depends on p . In fact, there might be multiple decision makers who solve

$P(X, \mathbf{S}, p)$ but with different p . Because the verifier knows nothing about p , any assumption about p should seem reasonable to the verifier. Alternatively, it might be the case that a decision maker chooses the optimal policy for the MDP, where he estimates p . Depending on how p is estimated, the decision maker solves the MDP having varying p . The verifier want to know whether the MDP is time-consistent before the decision maker estimates p .

When p is unknown to the verifier, we will say that $\text{POP}(\cdot, \mathbf{S}, \cdot)$ is time-consistent, if $\text{POP}(X, \mathbf{S}, p)$ is time-consistent for any X and p in the sense of Definition 3.

Definition 7 *We say that the POP, $\text{POP}(\cdot, \mathbf{S}, \cdot)$, is time-consistent if $\text{POP}(X, \mathbf{S}, p)$ is time-consistent for any X and p .*

Theorem 1 provided in Section 5.2 gives a sufficient condition for $\text{POP}(X, \mathbf{S}, p)$ to be time-consistent, but notice that the sufficient condition does not depend on X and p . Therefore, this sufficient condition also guarantees that $\text{POP}(\cdot, \mathbf{S}, \cdot)$ is time-consistent:

Corollary 2 *If f is an optimality-consistent dynamic RM and $h(\cdot) \equiv \mathbb{1}\{g(\cdot) \in B\}$ is a non-decreasing dynamic RM, then $\text{POP}(\cdot, \mathbf{S}, \cdot)$ as defined in Definition 7 is time-consistent.*

Similarly, all the results for the necessary conditions provided in Section 5.4 do not depend on X and p . Thus, these results are also true for $\text{POP}(\cdot, \mathbf{S}, \cdot)$.

5.4 Necessary conditions

In this section, we study necessity of the sufficient condition provided in Theorem 1 and Corollary 2. We will show, with a technical condition on \mathbf{S} , that the sufficient condition on the constraints is necessary for $\text{POP}(\cdot, \mathbf{S}, \cdot)$ to be time-consistent for any objective function. Also, the sufficient condition on the objective function is necessary for $\text{POP}(\cdot, \mathbf{S}, \cdot)$ to be time-consistent for any constraints.

The technical condition on \mathbf{S} requires that there exists n such that $|\mathbf{S}_n| \geq 2$. However, this technical condition is trivially satisfied for all MDPs of practical interest. Recall that $|\mathbf{S}_n|$ is non-decreasing with n , because our state includes the history of states and actions, and the state space increases over time. At time 0, we have $|\mathbf{S}_0| = 1$, but there must be an $\ell \in \mathbf{Z}_{[1, N]}$ such that $|\mathbf{S}_\ell| \geq 2$, because there must be multiple possible actions to take before N , as $|\Pi| > 1$. In the following, we use $=_d$ to denote equality in distribution.

Lemma 1 *Let ℓ be the minimum n such that $|\mathbf{S}_n| \geq 2$. If $\text{POP}(\cdot, \mathbf{S}, \cdot)$ as defined in Definition 7 is time-consistent for any objective function f , then $h_n(\cdot) \equiv \mathbb{1}\{g_n(\cdot) \in B_n\}$ must be non-decreasing for $n \geq \ell$.*

Proof: Suppose that $\mathbb{1}\{g(\cdot) \in B\}$ is not non-decreasing at an $n \in \mathbf{Z}_{[\ell, N]}$. We will show that there exist an objective function f , a distribution of X , and p such that $\text{POP}(X, \mathbf{S}, p)$ is not time-consistent.

Because $\mathbb{1}\{g(\cdot) \in B\}$ is not non-decreasing at n , there exists Y such that $\Pr(\mathbb{1}\{g_{n-1}(Y) \in B_{n-1}\} > \mathbb{1}\{g_n(Y) \in B_n\}) > 0$. That is, with non-zero probability, we have $g_{n-1}(Y) \in B_{n-1}$ and $g_n(Y) \notin B_n$. Hence,

$$q \equiv \Pr(g_n(Y) \notin B_n \mid g_{n-1}(Y) \in B_{n-1}) > 0.$$

Consider an $s_{n-1} \in \mathbf{S}_{n-1}$, an $s_n \in \mathbf{S}_n$, and a $\pi' \in \Pi$. We define p so that $p_{n-1}^{\pi'}(s_n | s_{n-1}) = q$. Setting $p_{n-1}^{\pi'}(s_n | s_{n-1})$ at an arbitrary number in $[0, 1]$ is possible, because $|\mathbf{S}_n| \geq 2$. We can define X so that

$X^{\pi'} =_d Y$, and the following two properties are satisfied:

$$g_{n-1}(X^{\pi'}(s_{n-1}, n-1)) \in B_{n-1} \quad (11)$$

$$g_n(X^{\pi'}(s_n, n)) \notin B_n. \quad (12)$$

Notice that (11) and (12) are possible because of the way how we define Y and p (in particular, $p_{n-1}^{\pi'}(s_n|s_{n-1}) = q$). Further, we can define f and X so that, for all $\pi \in \Pi$ such that $\pi \neq \pi'$, we have

$$\begin{aligned} f_{n-1}(X^\pi(s_{n-1}, n-1)) &\leq f_{n-1}(Y) \\ g_n(X^\pi(s_n, n)) &\in B_n. \end{aligned}$$

Then π' is one of the optimal policies from s_{n-1} . However, π' is infeasible from s_n , which can be reached from s_{n-1} with probability $q > 0$ with π' . Then π' is not optimal from s_n , because there is a policy, $\pi \in \Pi$, feasible from s_n . Therefore, $\text{POP}(X, \mathbf{S}, p)$ is not time-consistent with the objective function f . ■

Lemma 2 *Let ℓ be the minimum n such that $|\mathbf{S}_n| \geq 2$. If $\text{POP}(\cdot, \mathbf{S}, \cdot)$ as defined in Definition 7 is time-consistent for any constraints $g(\cdot) \in B$, then f must be optimality-consistent for all $n \geq \ell$.*

Proof: Suppose that f is not optimality-consistent at an $n \in \mathbf{Z}_{[\ell, N]}$. We will construct a distribution of X and p such that $\text{POP}(X, \mathbf{S}, p)$ with no constraint is not time-consistent.

Because f is not optimality-consistent at n , there exist $Y^{(1)}$ and $Y^{(2)}$ such that $\Pr(f_n(Y^{(1)}) \leq f_n(Y^{(2)})) = 1$ and $\Pr(f_n(Y^{(1)}) < f_n(Y^{(2)})) > 0$, but $\Pr(f_{n-1}(Y^{(1)}) < f_{n-1}(Y^{(2)})) = 0$. Let

$$q \equiv \Pr(f_n(Y^{(1)}) < f_n(Y^{(2)})) > 0.$$

Consider an $s_{n-1} \in \mathbf{S}_{n-1}$, an $s_n \in \mathbf{S}_n$, and two distinct policies, $\pi^{(1)}, \pi^{(2)} \in \Pi$. We define p so that $p_{n-1}^{\pi^{(i)}}(s_n|s_{n-1}) = q^{(i)}$ for $i = 1, 2$, and $q^{(1)}q^{(2)} = q$. For $i = 0, 1$, setting $p_{n-1}^{\pi^{(i)}}(s_n|s_{n-1})$ at an arbitrary number in $[0, 1]$ is possible, because $|\mathbf{S}_n| \geq 2$. We can define X so that $X^{\pi^{(1)}} =_d Y^{(1)}$, $X^{\pi^{(2)}} =_d Y^{(2)}$, and the following two properties are satisfied:

$$f_{n-1}(X^{\pi^{(1)}}(s_{n-1})) \geq f_{n-1}(X^{\pi^{(2)}}(s_{n-1})) \quad (13)$$

$$f_n(X^{\pi^{(1)}}(s_n)) < f_n(X^{\pi^{(2)}}(s_n)) \quad (14)$$

Notice that (13) and (14) are possible because of the way how we define $Y^{(1)}$, $Y^{(2)}$, and p (in particular, $p_{n-1}^{\pi^{(1)}}(s_n|s_{n-1})p_{n-1}^{\pi^{(2)}}(s_n|s_{n-1}) = q$). Further, we can define X so that $f_\ell(X^\pi(s_{n-1}, n-1)) \leq f_\ell(X^{\pi^{(1)}}(s_{n-1}, n-1))$ for any $\pi \in \Pi$. Then $\pi^{(1)}$ is optimal from s_{n-1} but not from s_n , and $p_{n-1}(s_n|s_{n-1}) > 0$. Thus, $\text{POP}(X, \mathbf{S}, p)$ with no constraints is not time-consistent. ■

The results in Section 5 lead to the following necessary and sufficient condition:

Corollary 3 *Let ℓ be the minimum n such that $|\mathbf{S}_n| \geq 2$. Suppose there exists $\gamma \equiv \min_{n \in \mathbf{Z}_{[0, N]}, s_n \in \mathbf{S}_n} f_n(X^\pi(s_n, n))$. Then $\text{POP}(\cdot, \mathbf{S}, \cdot)$ is time-consistent iff $(f(\cdot) - \gamma) \mathbf{1}\{\cdot \in B\}$ is optimality-consistent for all $n \in \mathbf{Z}_{[\ell, N]}$.*

Proof: Consider an POP with objective function, $(f(\cdot) - \gamma)I\{\cdot \in B\}$, and no constraints. Recall that we have proved Lemma 1 by showing that the condition is necessary when there are no constraints. Hence, the POP is time-consistent only if $(f(\cdot) - \gamma)I\{\cdot \in B\}$ is optimality-consistent. The sufficiency of $(f(\cdot) - \gamma)I\{\cdot \in B\}$ for the POP to be time-consistent can be shown by following the argument in the proof of Theorem 1 ■

6 Related work

Time-consistency was first discussed in deterministic settings regarding how future cost and profit should be discounted. In particular, Strotz shows that exponential discounting is necessary for time-consistency [18]. As a result, exponential discounting is standard for decision making with discounted expected utility [13]. The necessity of time-consistency for rational decision making is illuminated by the following quote from [2] (p.30-31): “if a hyperbolic discounter engaged in trade with someone who used an exponential curve, she’d soon be relieved of her money. Ms. Exponential could buy Ms. Hyperbolic’s winter coat cheaply every spring, for instance, because the distance to the next winter would depress Ms. H’s evaluation of it more than Ms. E’s. Ms. E could then sell the coat back to Ms. H every fall when the approach of winter sent Ms. H’s valuation of it into a high spike.”

Although time-consistency has been discussed in the context of risk measures [5, 6, 14], there has not been a unified discussion about time-consistency in the context of decision making under uncertainties. However, time-inconsistency has been reported for a few models of MDPs. In the prior work, when optimization of an MDP is time-inconsistent, it has been stated in various ways, including “the optimal policy changes over time,” “the optimal policy is nonstationary,” or “the principle of optimality is not satisfied.”

For example, a constrained MDP requires to minimize an expected utility of one type of cost, while keeping an expected utility of another type of cost below a threshold [3]. Haviv has pointed out that Bellman’s principle of optimality is not necessarily satisfied by an optimal policy of a constrained MDP [7]. Similar observations have been made for example in [8, 16, 17]. Counter-examples have been constructed for the case where the constrained MDP is a multi-chain over an infinite horizon, which is closely related to time-inconsistency over a finite horizon discussed in this paper. Notice that a multi-chain can be seen as a Markov chain over a finite horizon, where a state in the last period corresponds to an ergodic Markov chain. Our results for a finite horizon have implications to the multi-chain MDP over an infinite horizon. It has also been pointed out that the Bellman’s principle of optimality is satisfied if the constraint must be satisfied for every sample path [7, 16]. This can also be explained with our results, because the indicator random variable associated with the constraint on every sample-path is non-decreasing. Notice that our results apply not only to the constrained MDP studied in the prior work but also to the MDP with iterated RMs. See [11, 12] for the significance of the MDP with iterated RMs.

Also, it has been pointed out that the optimal policy of an MDP can change over time when its objective is to maximize or minimize, at each moment, an expected exponential-utility of the cumulative cost that will be incurred after the moment, where the cost is discounted over time [9, 20, 11].

7 Conclusion

We have shown that making decisions based on an optimal solution to an optimization problem can lead to inconsistent decision-making over time when there are uncertainties. Specifically, we have constructed an example of an optimization problem whose objective is to minimize the expected cost, and there is a constraint that the entropic risk measure of the cost is below a threshold. We have shown that a decision maker who takes actions based on the optimal solution to this optimization problem pays at least as much as and sometimes (with positive probability) pays more than those who make decisions in other ways. That is, making decisions based on a time-inconsistent optimization problem is irrational. The time-inconsistency of the above optimization problem might be counter-intuitive, because the functions consisting of the objective (i.e., expectation) and the constraints (i.e., entropic risk measure) are risk measures that are known to be time-consistent.

We have defined optimality-consistency and non-decreasing property of risk measures to provide sufficient conditions and necessary conditions for a process of optimization problems (POP) to be time-consistent. An optimality-consistent objective function ensures that if a policy, π , will become better than another policy, π' , with positive probability and π will never become worse than π' , then π is better than π' today. Constraints having the non-decreasing property ensure that a policy feasible today will be feasible in future.

Acknowledgements

This work was supported by “Promotion program for Reducing global Environmental load through ICT innovation (PREDICT)” of the Ministry of Internal Affairs and Communications, Japan.

References

- [1] B. Acciaio and I. Penner. Dynamic risk measures, 2010.
- [2] G. Ainslie. *Breakdown of Will*. Cambridge University Press, Cambridge, UK, first edition, 2001.
- [3] E. Altman. *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999.
- [4] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- [5] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk adjusted values and Bellman’s principle. *Annals of Operations Research*, 152(1):5–22, 2007.
- [6] M. R. Hardy and J. L. Wirch. The iterated CTE: A dynamic risk measure. *North American Actuarial Journal*, 8(4):62–75, 2004.
- [7] M. Haviv. On constrained Markov decision processes. *Operations Research Letters*, 19(1):25–28, 1996.
- [8] M. Henig. Optimal paths in graphs with stochastic or multidimensional weights. *Communications of the ACM*, 26(9):670–676, 1984.
- [9] S. C. Jaquette. A utility criterion for markov decision processes. *Management Science*, 23(1):43–49, 1976.

- [10] M. Kupper and W. Schachermayer. Representation results for law invariant time consistent functions, 2009.
- [11] T. Osogami. Iterated risk measures for risk-sensitive markov decision processes with discounted cost. Technical Report RT0922, IBM Research - Tokyo, November 2010.
- [12] T. Osogami. Overcoming limitations of expected utility with iterated risk measures. Technical Report RT0921, IBM Research - Tokyo, November 2010.
- [13] J. R. E. Lucas. Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445, 1978.
- [14] F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and their Applications*, 112:185–200, 2004.
- [15] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000.
- [16] K. W. Ross and R. Varadarajan. Markov decision processes with sample-path constraints: The communicating case. *Operations Research*, 37:780–790, 1989.
- [17] L. I. Sennott. Constrained average cost Markov decision chains. *Probability in the Engineering and Informational Sciences*, 7:69–83, 1993.
- [18] R. H. Strotz. Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23(3):165–180, 1956.
- [19] J. von Neumann, O. Morgenstern, H. W. Kuhn, and A. Rubinstein. *Theory of Games and Economic Behavior*. Princeton University Press, 60th anniversary edition, 2007.
- [20] P. Whittle. Why discount? The rationale of discounting in optimisation problems. In C. C. Heyde, Y. V. Prohorov, R. Pyke, and S. T. Rachev, editors, *Athens Conference on Applied Probability and Time Series: Volume I: Applied Probability (Lecture Notes in Statistics, Volume 114)*, pages 354–360. Springer, 1996.