ROBUSTNESS TO INCORRECT PRIORS IN PARTIALLY OBSERVED STOCHASTIC CONTROL[*]

ALİ DEVRAN KARA[†] AND SERDAR YÜKSEL[†]

Abstract. We study the continuity properties of optimal solutions to stochastic control problems with respect to initial probability measures and applications of these to the robustness of optimal control policies applied to systems with incomplete or incorrect priors. It is shown that for single and multistage optimal cost problems, continuity and robustness cannot be established under weak convergence in general, but that the optimal expected cost is continuous in the priors under the convergence in total variation under mild conditions. By imposing further assumptions on the measurement models, robustness and continuity also hold under weak convergence of priors. We thus obtain robustness results and bounds on the mismatch error that occurs due to the application of a control policy which is designed for an incorrectly estimated prior as the incorrect prior converges to the true one. Positive and negative practical implications of these results in empirical learning for stochastic control are presented, where almost surely weak convergence of independently and identically distributed empirical measures occurs but stronger notions of convergence, such as total variation convergence, in general, do not.

1. Introduction.

1.1. Preliminaries. We start with the probabilistic setup of the problem. Let $\mathbb{X} \subset \mathbb{R}^n$, be a Borel set in which elements of a controlled Markov process $\{X_t, t \in \mathbb{Z}_+\}$ live. Here and throughout the paper $\mathbb{Z}_+$ denotes the set of non-negative integers and $\mathbb{N}$ denotes the set of positive integers. Let $\mathbb{Y} \subset \mathbb{R}^m$ be a Borel set, and let an observation channel Q be defined as a stochastic kernel (regular conditional probability) from $\mathbb{X}$ to $\mathbb{Y}$ such that $Q(\cdot|x)$ is a probability measure on the (Borel) σ-algebra $\mathcal{B}(\mathbb{Y})$ on $\mathbb{Y}$ for every $x \in \mathbb{X}$, and $Q(A|\cdot) : \mathbb{X} \to [0,1]$ is a Borel measurable function for every $A \in \mathcal{B}(\mathbb{Y})$. Let a decision maker (DM) be located at the output of an observation channel Q, with inputs X_t and outputs Y_t. Let $\mathbb{U}$, the action space, be a Borel subset of some Euclidean space. An *admissible policy* γ is a sequence of control functions $\{\gamma_t, t \in \mathbb{Z}_+\}$ such that γ_t is measurable with respect to the σ-algebra generated by the information variables

$$I_t = \{Y_{[0,t]}, U_{[0,t-1]}\}, \quad t \in \mathbb{N}, \qquad I_0 = \{Y_0\},$$

where

$$(1.1) \qquad U_t = \gamma_t(I_t), \quad t \in \mathbb{Z}_+$$

are the $\mathbb{U}$-valued control actions and

$$Y_{[0,t]} = \{Y_s, 0 \le s \le t\}, \quad U_{[0,t-1]} = \{U_s, 0 \le s \le t-1\}.$$

[†]Mathematics and Statistics, Queen's University, Kingston, ON, Canada (yuksel@mast.queensu.ca, 16adk@queensu.ca).

We define Γ to be the set of all such admissible policies.

The joint distribution of the state, control, and observation processes is determined by (1.1) and the following relationships:

$$\Pr\big((X_0, Y_0) \in B\big) = \int_B P(dx_0)Q(dy_0|x_0), \quad B \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}),$$

where P is the (prior) distribution of the initial state X_0, and

$$\Pr\big((X_t, Y_t) \in B \mid (X, Y, U)_{[0,t-1]} = (x, y, u)_{[0,t-1]}\big)$$
$$= \int_B \mathcal{T}(dx_t|x_{t-1}, u_{t-1})Q(dy_t|x_t), B \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}), t \in \mathbb{N},$$

where $\mathcal{T}(\cdot|x, u)$ is a stochastic kernel from $\mathbb{X} \times \mathbb{U}$ to $\mathbb{X}$ and $Q(\cdot|x)$ is a stochastic kernel from $\mathbb{X}$ to $\mathbb{Y}$.

We let the objective of the agent be the minimization of the cost for the static or single stage case,

$$J(P, Q, \gamma) = E_P^{Q,\gamma}\left[c(X_0, U_0)\right],$$

over the set of admissible policies $\gamma \in \Gamma$, where $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is a Borel-measurable stagewise cost function and $E_P^{Q,\gamma}$ denotes the expectation with initial state probability measure P and measurement channel Q under policy γ. Note that $P \in \mathcal{P}(\mathbb{X})$, where we let $\mathcal{P}(\mathbb{X})$ denote the set of probability measures on $\mathbb{X}$.

For the multistage case, we will discuss the discounted cost infinite horizon setting, with the following cost criterion to be minimized:

$$J_\beta(P, Q, \gamma) = E_P^{Q,\gamma}\left[\sum_{t=0}^{\infty} \beta^t c(X_t, U_t),\right]$$

for some $\beta \in (0, 1)$.

We define the optimal cost for the single-stage and the discounted infinite horizon as a function of the priors as

$$J^*(P, Q) = \inf_{\gamma \in \Gamma} J(P, Q, \gamma),$$
$$J_\beta^*(P, Q) = \inf_{\gamma \in \Gamma} J_\beta(P, Q, \gamma),$$

respectively.

Note that for the discounted infinite horizon case, the cost function is also affected by the transition kernel $\mathcal{T}$. Thus, in the following, we may sometimes use $J_\beta(P, \mathcal{T}, \gamma)$ and $J^*(P, \mathcal{T})$ instead of $J_\beta(P, Q, \gamma)$ and $J^*(P, Q)$ depending on the context.

The focus of the paper will be to address the following problems.

Problem P1: Continuity of $J^*(P, Q)$ and $J_\beta^*(P, Q)$ on the space of prior distributions. Suppose $\{P_n, n \in \mathbb{N}\}$ is a sequence of priors converging in some sense to P. When does $P_n \to P$ imply $J^*(P_n, Q) \to J^*(P, Q)$ or $J_\beta^*(P_n, Q) \to J_\beta^*(P, Q)$?

Problem P2: Robustness to incorrect priors. A problem of major practical importance is robustness of an optimal controller to modeling errors. Suppose that an optimal policy is constructed according to a model that is incorrect: how does the application of the control to the true model affect the system performance and does the error decrease to zero as the models become closer to each other? In particular, suppose that γ_n is an optimal policy designed for P_n, an incorrect prior model for a true model P. Is it the case that if $P_n \to P$, then $J(P, Q, \gamma_n) \to J^*(P, Q)$?

Problem P3: Empirical consistency of optimal costs designed under learned priors. Let $P \in \mathcal{P}(\mathbb{X})$ be a fixed initial distribution for some state variable X which

is unknown to a decision maker. Suppose that the decision maker learns $\widetilde{P}_n$, its estimate of P, from the collection of empirical observations of realizations from independently and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$ distributed with P, and applies an optimal policy for the control problem with initial distribution $\widetilde{P}_n$ (i.e., through a *plug-in* or *separated* controller design). Defining for every (fixed) Borel $B \subset \mathbb{X}$, and $n \in \mathbb{N}$, the empirical occupation measures

$$\widetilde{P}_n(B) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \in B\}},$$

do we have that the optimal cost computed for the empirical measures converges to the true optimal cost (for the case where the true P is known) as $n \to \infty$, almost surely?

Here is a summary of the rest of the paper: in the following two subsections we present a literature review and some basic properties with regard to the convergences of probability measures. In section 2, we study the continuity properties of the optimal cost functions with respect to prior measures under different convergence notions for both the single-stage and multistage settings. In section 3, we use the results from section 2 to obtain robustness results on control policies designed under incorrect prior estimates. Finally, in section 4, an application of the results to control systems where the prior measures are estimated through empirical measurements is presented.

1.2. Literature review. The H_∞ criterion in robust control [54, 1, 55] addresses the problem of robustness of control policies with respect to unmodeled dynamics. The goal in robust control is to design control policies that work sufficiently well for systems with model or disturbance uncertainty.

Researchers have developed robust controllers through a game formulation, where the minimizer is the controller and the maximizer is the uncertainty, and have established the equivalence with a risk-sensitive cost minimization for a class of systems [31, 49, 16]. Through such a formulation, and by Legendre-type transforms, the operational use of the relative entropy methods have come to the literature; see, e.g., [37, eq. (4)] or [16, eq. (2)–(3)]. Here, one selects a nominal system which satisfies a relative entropy bound between the actual measure and the nominal measure, solves a risk sensitive optimal control problem, and this solution provides an upper bound for the original system. As such, a common approach in robust stochastic control has been to model the stochastic disturbance affecting a system and to consider perturbations which lead to finite deviations according to the Kullback–Leibler divergence (or relative entropy) between the actual measure and a reference measure, or embed the uncertainty with a penalty term imposed on the cost function under such a distance measure; see, e.g., [16, 10, 36]. Along a similar theme, [32] studies an optimization problem for the expected cost of an uncontrolled i.i.d. model under relative entropy bounds for the probability measures on the state variables; this can be considered to be similar to the setup considered in our paper where the convergence notion is in the relative entropy on the priors, as the considered process is i.i.d. We note here that the relative entropy is a very restrictive *distance* measure (note though that this does not define a metric) and in particular, through Pinsker's inequality [22, Lem. 5.2.8], it is stronger than even total variation which has also been studied in robust stochastic control: [44] has studied a min-max formulation for robust control where the one-stage transition kernel belongs to a ball under the total variation metric for each state action pair and develops a dynamic programming based solution for both finite and discounted cost infinite horizon problems. Further related work with

model uncertainty includes [34, 3], with some further work in the economics literature [25, 21].

The results are also related to the input estimation problem from finitely many samples, as reviewed in [20] and empirical risk minimization [56], although in our context, we will investigate robustness in the context of a *separated* design: an input model is estimated through empirical data, and an optimal policy is constructed with the assumption that the estimated model is correct (see, e.g., [30] for an application). Can we guarantee empirical (asymptotic) consistency? We will discuss a number of general results in section 4 and observe that in many situations empirical consistency may not hold.

On continuity properties in prior measures, [51] and [50] have studied the special case of minimum mean-square estimation, that is, with $c(x, u) = (x - u)^2$ for a single-stage problem across additive noisy channels of the form $y = x + w$, and established conditions leading to continuity or upper semicontinuity properties under weak convergence and Wasserstein metrics.

Related work also includes the recent studies [52] and [2]; [52] considers various topologies on the sets of observation channels and quantizers in partially observed stochastic control and provides some supporting results, whereas [2] presents a number of continuity properties for single-stage stochastic control problems.

The problems we consider are also related to, in the control-free context, the theory and applications of nonlinear filtering with incorrect initial specifications. Here, the problem is to identify conditions on when an incorrectly initialized nonlinear filter asymptotically gets corrected with the accumulation of additional measurements; these often require strong ergodicity properties of the Markov process [8, 9, 11] or regularity properties (such as absolute continuity) of incorrect prior with respect to the true one and conditions on the measurement processes [24].

1.3. Convergence of probability measures and some supporting results. Three important notions of convergences for sets of probability measures are weak convergence, setwise convergence, and convergence under total variation (see, e.g., [29] and [52]). For some $N \in \mathbb{N}$ a sequence $\{\mu_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{R}^N)$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^N)$ *weakly* if

$$(1.2) \qquad \int_{\mathbb{R}^N} c(x)\mu_n(dx) \to \int_{\mathbb{R}^N} c(x)\mu(dx)$$

for every continuous and bounded $c : \mathbb{R}^N \to \mathbb{R}$. $\{\mu_n\}$ is said to converge *setwise* to $\mu \in \mathcal{P}(\mathbb{R}^N)$ if (1.2) holds for all measurable and bounded $c : \mathbb{R}^N \to \mathbb{R}$. Setwise convergence can also be equivalently defined through pointwise convergence on Borel subsets of $\mathbb{R}^N$, that is, $\mu_n(A) \to \mu(A)$ for all $A \in \mathcal{B}(\mathbb{R}^N)$.

For probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$, the *total variation* metric is given by

$$\|\mu - \nu\|_{TV} = 2 \sup_{B \in \mathcal{B}(\mathbb{R}^N)} |\mu(B) - \nu(B)|$$

$$= \sup_{f : \|f\|_\infty \leq 1} \left| \int f(x)\mu(\mathrm{d}x) - \int f(x)\nu(\mathrm{d}x) \right|,$$

where the supremum is taken over all measurable real f such that $\|f\|_\infty = \sup_{x \in \mathbb{R}^N} |f(x)| \leq 1$. A sequence $\{\mu_n\}$ is said to converge in total variation to $\mu \in \mathcal{P}(\mathbb{R}^N)$ if $\|\mu_n - \mu\|_{TV} \to 0$.

We next introduce the Wasserstein metric. The *Wasserstein metric* of order 1 for two distributions $\mu, \nu \in \mathcal{P}(\mathbb{X})$ is defined as

$$W_1(\mu, \nu) = \inf_{\eta \in \mathcal{H}(\mu, \nu)} \int_{\mathbb{X} \times \mathbb{X}} \eta(\mathrm{d}x, \mathrm{d}y)|x - y|,$$

where $\mathcal{H}(\mu, \nu)$ denotes the set of probability measures on $\mathbb{X} \times \mathbb{X}$ with first marginal μ and second marginal ν.

A sequence $\{\mu_n\}$ is said to converge in W_1 to $\mu \in \mathcal{P}(\mathbb{R}^N)$ if $W_1(\mu_n, \mu) \to 0$. For compact $\mathbb{X}$, the Wasserstein distance of order 1 metrizes the weak topology on the set of probability measures on $\mathbb{X}$ (see [47, Thm. 6.9]). For noncompact $\mathbb{X}$ convergence in the W_1 metric implies weak convergence (in particular this metric bounds from above the Bounded-Lipschitz metric [47, p. 109], which metrizes the weak convergence).

The following result shows the relation between convergence of prior measures and convergence of joint measures of channel and initial distribution given a fixed channel $Q \in \mathcal{Q}$.

The joint measure PQ is induced on $(\mathbb{X} \times \mathbb{Y}, \mathcal{B}(\mathbb{X} \times \mathbb{Y}))$ for $Q \in \mathcal{Q}$ and $P \in \mathcal{P}(\mathbb{X})$ where $\mathcal{Q}$ is a set of communication channels,

$$PQ(A) = \int_A Q(\mathrm{d}y|x)P(\mathrm{d}x), \ A \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}).$$

The following is a result that will be used later in the paper.

LEMMA 1.1. *Let $Q : \mathcal{B}(\mathbb{Y}) \times \mathbb{X} \to [0, 1]$ be a stochastic kernel on $\mathbb{Y}$ given $\mathbb{X}$.*

(i) *Assume that $Q(dy|x)$ is weakly continuous in x in the sense that $\int Q(dy|x)c(y)$ is continuous in x for every continuous and bounded c. If $\{P_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{X})$ converges to $P \in \mathcal{P}(\mathbb{X})$ weakly, then $P_nQ \to PQ$ weakly.*

(ii) *If $\{P_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{X})$ converges to $P \in \mathcal{P}(\mathbb{X})$ setwise, then $P_nQ \to PQ$ setwise.*

(iii) *If $\{P_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathbb{X})$ converges to $P \in \mathcal{P}(\mathbb{X})$ in total variation, then $P_nQ \to PQ$ in total variation. In particular, $\|P_nQ - PQ\|_{TV} = \|P_n - P\|_{TV}$.*

Proof.

(i) The sequence $\{f_n\}$ is said to converge continuously to f when $\lim_{n\to\infty} f_n(x_n) = f(x)$ for any $x_n \to x$.

First we show that under the given assumptions, and with any continuous and bounded $c : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$, $\int_{\mathbb{Y}} c(x, y)Q(\mathrm{d}y|x)$ is continuous in x. Let $x_n \to x$ in $\mathbb{X}$, then $c(x_n, y) \to c(x, y)$ for all $y \in \mathbb{Y}$ by continuity of c. In particular, with $c_n(y) := c(x_n, y)$ and $c(y) := c(x, y)$, it follows that as $y_n \to y$, $c_n(y_n) \to c(y)$; that is, $c_n(\cdot)$ *continuously converges* to $c(\cdot)$ as $(x_n, y_n) \to (x, y)$ by [43, Thm. 3.3] or [33, Thm. 3.5], it follows that the continuity property

$$\lim_{n\to\infty} \int_{\mathbb{Y}} c(x_n, y)Q(\mathrm{d}y|x_n) = \int_{\mathbb{Y}} c(x, y)Q(\mathrm{d}y|x)$$

holds.

Then, we have

$$\lim_{n\to\infty} \int_{\mathbb{X} \times \mathbb{Y}} c(x, y)P_nQ(\mathrm{d}x, \mathrm{d}y)$$

$$= \lim_{n\to\infty} \int_{\mathbb{X}} \left(\int_{\mathbb{Y}} c(x, y)Q(\mathrm{d}y|x) \right) P_n(\mathrm{d}x)$$

$$= \int_{\mathbb{X}} \left(\int_{\mathbb{Y}} c(x, y)Q(\mathrm{d}y|x) \right) P(\mathrm{d}x)$$

(1.3) $$= \int_{\mathbb{X} \times \mathbb{Y}} c(x, y)PQ(\mathrm{d}x, \mathrm{d}y),$$

where we have applied Fubini's theorem and the fact that $\int_{\mathbb{Y}} c(x,y)Q(\mathrm{d}y|x)$ is a bounded and continuous function in x under the given assumptions.

(ii) Given any measurable and bounded $c : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ we have

$$\lim_{n \to \infty} \int_{\mathbb{X} \times \mathbb{Y}} c(x,y)P_nQ(\mathrm{d}x,\mathrm{d}y) = \lim_{n \to \infty} \int_{\mathbb{X}} \left(\int_{\mathbb{Y}} c(x,y)Q(\mathrm{d}y|x) \right) P_n(\mathrm{d}x)$$
$$= \int_{\mathbb{X}} \left(\int_{\mathbb{Y}} c(x,y)Q(\mathrm{d}y|x) \right) P(\mathrm{d}x)$$
$$= \int_{\mathbb{X} \times \mathbb{Y}} c(x,y)PQ(\mathrm{d}x,\mathrm{d}y),$$

where we have applied Fubini's theorem and the fact that $\int_{\mathbb{Y}} c(x,y)Q(\mathrm{d}y|x)$ is a bounded and measurable function of x under the given assumptions (see Proposition 7.29 in [4]).

(iii) Observe the following:

$$\|P_nQ - PQ\|_{TV}$$
$$= \sup_{f : \|f\|_\infty \leq 1} \left| \int_{\mathbb{X} \times \mathbb{Y}} f(x,y)P_nQ(\mathrm{d}x,\mathrm{d}y) - \int_{\mathbb{X} \times \mathbb{Y}} f(x,y)PQ(\mathrm{d}x,\mathrm{d}y) \right|,$$
$$= \sup_{f : \|f\|_\infty \leq 1} \left| \int_{\mathbb{X}} \left(\int_{\mathbb{Y}} f(x,y)Q(\mathrm{d}y|x) \right) P_n(\mathrm{d}x) \right.$$
$$\left. - \int_{\mathbb{X}} \left(\int_{\mathbb{Y}} f(x,y)Q(\mathrm{d}y|x) \right) P(\mathrm{d}x) \right|$$
$$\leq \sup_{\bar{f} : \|\bar{f}\|_\infty \leq 1} \left| \int_{\mathbb{X}} \bar{f}(x)P_n(\mathrm{d}x) - \int_{\mathbb{X}} \bar{f}(x)P(\mathrm{d}x) \right|$$
$$= \|P_n - P\|_{TV},$$

where we used the fact that $\left| \bar{f}(x) \right| := \left| \int_{\mathbb{Y}} f(x,y)Q(\mathrm{d}y|x) \right| \leq 1$ for every $x \in \mathbb{X}$.

In addition, as $P(B) = PQ(B \times \mathbb{Y})$ for any Borel B and any $P \in \mathcal{P}(\mathbb{X})$, then we have

$$\|P_n - P\|_{TV} = 2 \sup_{B \in \mathcal{B}(\mathbb{X})} |P_n(B) - P(B)|$$
$$= 2 \sup_{B \in \mathcal{B}(\mathbb{X})} |P_nQ(B \times \mathbb{Y}) - PQ(B \times \mathbb{Y})|$$
$$\leq 2 \sup_{A \in \mathcal{B}(\mathbb{X} \times \mathbb{Y})} |P_nQ(A) - PQ(A)|$$
$$= \|P_nQ - PQ\|_{TV}.$$

Therefore $\|P_nQ - PQ\|_{TV} = \|P_n - P\|_{TV} \to 0$. $\qquad\square$

2. Continuity of optimal cost with respect to convergence of prior probability measures. In this section, we study the continuity in the priors.

2.1. Single-stage setup. First, we consider a single-stage setup. The initial measure will be denoted by $P \in \mathcal{P}(\mathbb{X})$, the measurement channel will be $Q \in \mathcal{Q}$, and the control action will be from the admissible control policies, Γ. Here, Γ is just

the set of functions from $\mathbb{Y}$ to $\mathbb{U}$ that are measurable with respect to the σ-algebra generated by $I_0 = \{Y_0\}$.

The optimal single-stage cost will be defined as

$$J^*(P,Q) = \inf_{\gamma \in \Gamma} \int_{\mathbb{X} \times \mathbb{Y}} c(x, \gamma(y)) Q(\mathrm{d}y|x) P(\mathrm{d}x).$$

The following is a useful result to be used throughout the paper.

THEOREM 2.1. [52, Thm. 3.1] *Let c be continuous in u for every x, and $\mathbb{U}$ be compact. Then, for the static problem, there exists an optimal control policy for any channel Q.*

For the rest of the paper, we will use optimal policies whenever they exist otherwise we will use ϵ-optimal policies.

2.1.1. Weak convergence. We will first observe that $J^*(P,Q)$ is not always continuous in initial measures under weak convergence; then we will present an upper semicontinuity result and some sufficiency continuity conditions for weak convergence of prior measures.

THEOREM 2.2. *Let a channel $Q \in \mathcal{Q}$ be given. $J^*(P,Q)$ is not necessarily continuous in P under weak convergence. This holds even when $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{U}$ are compact and when c is bounded and continuous on $\mathbb{X} \times \mathbb{U}$.*

Proof. We prove the result with a counter example. Let $\mathbb{X} = \mathbb{U} = \mathbb{Y} = [0,1]$ and $c(x,u) = (x-u)^2$. The optimal policy for this cost function is $\gamma^*(y) = E[X|Y=y]$. The measurement channel is given in the following form:

$$Q(\cdot|x) = \frac{1}{2}\delta_x(\cdot) + \frac{1}{2}\delta_0(\cdot).$$

Let the prior distributions P and P_n be given by

$$P = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1,$$
$$P_n = \frac{1}{2}\delta_{\frac{1}{n}} + \frac{1}{2}\delta_1.$$

Note that $P_n \to P$ weakly as $n \to \infty$. Now, we calculate the optimal control actions for prior model P:

$$\gamma^*(0) = E[X|Y=0] = Pr(X=1|Y=0) = Pr(Y=0|X=1)\frac{Pr(X=1)}{Pr(Y=0)} = \frac{1}{3},$$
$$\gamma^*(1) = E[X|Y=1] = Pr(X=1|Y=1) = Pr(Y=1|X=1)\frac{Pr(X=1)}{Pr(Y=1)} = 1.$$

Thus, the optimal cost for P can be calculated as

$$J^*(P,Q) = E[(X - \gamma^*(Y))^2]$$
$$= Pr(X{=}0)Pr(Y{=}0|X{=}0)\left(0 - \frac{1}{3}\right)^2 + Pr(X{=}1)Pr(Y{=}0|X{=}1)\left(1 - \frac{1}{3}\right)^2$$
$$\quad + Pr(X{=}0)Pr(Y{=}1|X{=}0)(0{-}1)^2 + Pr(X{=}1)Pr(Y{=}1|X{=}1)(1-1)^2$$
$$= \frac{1}{18} + \frac{1}{9} = \frac{1}{6}.$$

The optimal control actions for P_n can be calculated as

$$
\begin{aligned}
\gamma_n^*(0) &= E_n[X|Y=0] = \frac{1}{n}Pr(X=1/n|Y=0) + 1Pr(X=1|Y=0) \\
&= \frac{1}{n}\frac{Pr(Y=0|X=1/n)Pr(X=1/n)}{Pr(Y=0)} + \frac{Pr(Y=0|X=1)Pr(X=1)}{Pr(Y=0)} \\
&= \frac{n+1}{2n}, \\
\gamma_n^*(1) &= E_n[X|Y=1] = \frac{1}{n}Pr(X=1/n|Y=1) + 1Pr(X=1|Y=1) \\
&= \frac{1}{n}\frac{Pr(Y=1|X=1/n)Pr(X=1/n)}{Pr(Y=1)} + \frac{Pr(Y=1|X=1)Pr(X=1)}{Pr(Y=1)} \\
&= 1, \\
\gamma_n^*(1/n) &= E_n[X|Y=1/n] = \frac{1}{n}Pr(X=1/n|Y=1/n) + 1Pr(X=1|Y=1/n) \\
&= \frac{1}{n}\frac{Pr(Y=1/n|X=1/n)Pr(X=1/n)}{Pr(Y=1/n)} + \frac{Pr(Y=1/n|X=1)Pr(X=1)}{Pr(Y=1/n)} \\
&= \frac{1}{n}.
\end{aligned}
$$

Using the optimal control actions, we can calculate the optimal cost for P_n:

$$
\begin{aligned}
J^*(P_n,Q) &= E_n[(X-\gamma^*(Y))^2] \\
&= Pr(X=0)Pr(Y=0|X=0)\left(0-\frac{n+1}{2n}\right)^2 \\
&\quad + Pr(X=1)Pr(Y=0|X=1)\left(1-\frac{n-1}{2n}\right)^2 \\
&= \frac{1}{2}\left(\frac{n+1}{2n}\right)^2 + \frac{1}{4}\left(\frac{n-1}{2n}\right)^2 = \frac{3n^2+2n+3}{16n^2}.
\end{aligned}
$$

We can see that as $n \to \infty$, $J^*(P_n,Q) \to \frac{3}{16} \neq \frac{1}{6} = J^*(P,Q)$. $\qquad\square$

The channel model in the example is a channel which either sends full information across the channel without error, or it provides no information and this is an infinite-capacity erasure channel: this channel has a practical significance as in practice we may have package drops during the transmission of the state variable which causes controller not to get any data from the system at random times.

We now provide another example which may have further practical significance. Quantizer channels are used often in practice (see [52, sect. 5]). In the following example, we show that continuity may not hold for quantizer channels either.

Before the example we define quantizers: an M-cell vector quantizer, q, is a (Borel) measurable mapping from $\mathbb{X} = \mathbb{R}^n$ to the finite set $\{1, 2, \ldots, M\}$, characterized by a measurable partition $\{B1, B2, \ldots, BM\}$ such that $B_i = \{x : q(x) = i\}$ for $i = 1, \ldots, M$. The B_i are called the cells (or bins) of q.

A quantizer q with cells $\{B_1, \ldots, B_M\}$, however, can also be characterized as a stochastic kernel Q from $\mathbb{X}$ to $\{1, \ldots, M\}$ defined by

$$Q(i|x) = 1_{x \in B_i}, \qquad i = 1, \ldots, M,$$

so that $q(x) = \sum_{i=1}^{M} Q(i|x)$.

Example 2.1. Assume the prior distributions are given by

$$P(\cdot) = \frac{1}{2}\delta_{\frac{1}{2}}(\cdot) + \frac{1}{2}\delta_1(\cdot),$$

$$P_n(\cdot) = \frac{1}{2}\delta_{\frac{1}{2}-\frac{1}{n}}(\cdot) + \frac{1}{2}\delta_1(\cdot),$$

and the channel is a quantizer with $B_1 = [0, \frac{1}{2})$ and $B_2 = [\frac{1}{2}, 1]$ where the range of the quantizer is $\{0, 1\}$ with

$$Q(0|x) = 1_{x \in [0, \frac{1}{2})}, \qquad Q(1|x) = 1_{x \in [\frac{1}{2}, 1]}.$$

If the cost function is given by $c(x, u) = (x - u)^2$, then the optimal controls are

$$\gamma^*(0) = 0, \qquad \gamma^*(1) = 3/4,$$

$$\gamma_n^*(0) = \frac{1}{2} - \frac{1}{n}, \qquad \gamma_n^*(1) = 1.$$

Under this setup the value functions can be calculated as

$$J^*(P_n, Q) = 0,$$
$$J^*(P, Q) = P(X = 1/2)(1/2 - 3/4)^2 + P(X = 1)(1 - 3/4)^2 = 1/16,$$

which shows that the optimal cost is not continuous under the weak convergence of priors when the channel is a quantization channel.

Now we show that the optimal cost is upper semicontinuous under weak convergence of priors. The next lemma, building on [52], shows that the optimal cost is unchanged when γ is restricted to the class of continuous policies. A brief proof is presented in the appendix.

LEMMA 2.3. *Let μ be an arbitrary probability measure on $(\mathbb{X} \times \mathbb{Y}, \mathcal{B}(\mathbb{X} \times \mathbb{Y}))$, and let $\mathcal{C}$ be the set of continuous functions from $\mathbb{Y}$ to $\mathbb{U}$. If $\mathbb{U}$ is convex and $c(x, u)$ is non-negative, measurable, and bounded in $\mathbb{X} \times \mathbb{U}$, then*

$$\inf_{\gamma \in \Gamma} \int c(x, \gamma(y))\mu(dx, dy) = \inf_{\gamma \in \mathcal{C}} \int c(x, \gamma(y))\mu(dx, dy).$$

The following theorem is related to Theorem 3 of Wu and Verdú [51].

THEOREM 2.4. *Let a channel $Q \in \mathcal{Q}$ be given. If $Q(dy|x)$ is weakly continuous in x in the sense that $\int Q(dy|x)c(y)$ is continuous in x for every continuous and bounded c, and if $\mathbb{U}$ is convex and $c(x, u)$ is non-negative, continuous, and bounded in $\mathbb{X} \times \mathbb{U}$, then $J^*(P, Q)$ is upper semicontinuous on $\mathcal{P}(\mathbb{X})$ under weak convergence.*

Proof. Recall from the statement of Lemma 2.3 that $\mathcal{C}$ denotes the set of all continuous functions from $\mathbb{Y}$ to $\mathbb{U}$. Let $P_n \to P$ in $\mathcal{P}(\mathbb{X})$ weakly. Then

$$\limsup_{n\to\infty} \inf_{\gamma\in\Gamma} \int_{\mathbb{X}\times\mathbb{Y}} c(x,\gamma(y)) P_n Q(\mathrm{d}x,\mathrm{d}y)$$

$$= \limsup_{n\to\infty} \inf_{\gamma\in\mathcal{C}} \int_{\mathbb{X}\times\mathbb{Y}} c(x,\gamma(y)) P_n Q(\mathrm{d}x,\mathrm{d}y)$$

$$\leq \inf_{\gamma\in\mathcal{C}} \limsup_{n\to\infty} \int_{\mathbb{X}\times\mathbb{Y}} c(x,\gamma(y)) P_n Q(\mathrm{d}x,\mathrm{d}y)$$

$$= \inf_{\gamma\in\mathcal{C}} \int_{\mathbb{X}\times\mathbb{Y}} c(x,\gamma(y)) PQ(\mathrm{d}x,\mathrm{d}y)$$

$$= \inf_{\gamma\in\Gamma} \int_{\mathbb{X}\times\mathbb{Y}} c(x,\gamma(y)) PQ(\mathrm{d}x,\mathrm{d}y),$$

where the first and last equality rely on Lemma 2.3 and the second-to-last equality holds as $c(x,\gamma(y))$ is bounded and continuous for $\gamma\in\mathcal{C}$. □

The following result shows us that if we put some continuity restrictions on the measurement channel, then we can guarantee the continuity of optimal single-stage cost function under weak convergence of prior measures.

Assumption 2.1. Q is continuous in total variation in the sense that as $x_n \to x$, $\|Q(dy|x_n) - Q(dy|x)\|_{TV} \to 0$.

THEOREM 2.5. *Let Assumption* 2.1 *hold,* $c(x,u)$ *be bounded and continuous on* $\mathbb{X}\times\mathbb{U}$, *and* $\mathbb{U}$ *be compact. If* $P_n \to P$ *weakly, then* $J^*(P_n,Q) \to J^*(P,Q)$.

Proof. We first show that under the stated assumptions, the following holds:

$$(2.1) \qquad \lim_{k\to\infty} \sup_{\gamma} \left| \int Q(\mathrm{d}y|x_k) c(x_k,\gamma(y)) - \int Q(\mathrm{d}y|x) c(x,\gamma(y)) \right| = 0$$

for $x_k \to x$. To see this, write

$$\lim_{k\to\infty} \sup_{\gamma} \left| \int Q(\mathrm{d}y|x_k) c(x_k,\gamma(y)) - \int Q(\mathrm{d}y|x) c(x,\gamma(y)) \right|$$

$$(2.2) \qquad \leq \lim_{k\to\infty} \sup_{\gamma} \left| \int Q(\mathrm{d}y|x_k) c(x_k,\gamma(y)) - \int Q(\mathrm{d}y|x) c(x_k,\gamma(y)) \right|$$

$$(2.3) \qquad + \lim_{k\to\infty} \sup_{\gamma} \left| \int Q(\mathrm{d}y|x) c(x_k,\gamma(y)) - \int Q(\mathrm{d}y|x) c(x,\gamma(y)) \right|.$$

The term in (2.2) converges to 0 by Assumption 2.1. Since we have

$$\lim_{k\to\infty} \sup_{\gamma} \left| \int Q(\mathrm{d}y|x_k) c(x_k,\gamma(y)) - \int Q(\mathrm{d}y|x) c(x_k,\gamma(y)) \right|$$

$$\leq \lim_{k\to\infty} \|Q(.|x_k) - Q(.|x)\|_{TV}$$

$$= 0.$$

By assumption the action space $\mathbb{U}$ is compact. Therefore, $c(x_k,u) \to c(x,u)$ uniformly on $\mathbb{U}$. Thus, (2.3) also goes to 0.

Let the optimal policies be given by γ_n^* and γ^* for initial distributions P_n and P, respectively. We now consider two cases.
First, if $J^*(P_n,Q) < J^*(P,Q)$, then

$$(2.4) \qquad J^*(P,Q) - J^*(P_n,Q) \leq J(P,Q,\gamma_n^*) - J(P_n,Q,\gamma_n^*).$$

Second, if $J^*(P,Q) < J^*(P_n,Q)$, then

$$(2.5) \qquad J^*(P_n,Q) - J^*(P,Q) \leq J(P_n,Q,\gamma^*) - J(P,Q,\gamma^*).$$

Inequalities (2.4) and (2.5) are combined to give

$$(2.6) \quad \begin{aligned} &|J^*(P,Q) - J^*(P_n,Q)| \\ &\leq \max(J(P,Q,\gamma_n^*) - J(P_n,Q,\gamma_n^*), J(P_n,Q,\gamma^*) - J(P,Q,\gamma^*)). \end{aligned}$$

Observe the following:

$$\begin{aligned} &|J^*(P_n,Q) - J^*(P,Q)| \\ &\leq \max\left(\left| \int (P_n - P)(\mathrm{d}x) \int Q(\mathrm{d}y|x) c(x,\gamma_n^*(y)) \right|, \right. \\ &\qquad\qquad \left. \left| \int (P_n - P)(\mathrm{d}x) \int Q(\mathrm{d}y|x) c(x,\gamma^*(y)) \right| \right). \end{aligned}$$

$$(2.7)$$

For continuity, we need (2.7) to tend to 0 as $n \to \infty$.

Let F be a family of functions from a complete separable metric space $\mathbb{S}$ to $\mathbb{R}$. The family F is said to be *equicontinuous at a point* $x_0 \in S$ if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that $|f(x) - f(x_0)| \leq \epsilon$ for all $f \in F$ and for every x such that $|x - x_0| \leq \delta$. The family F is said to be *equicontinuous* if it is equicontinuous at each $x \in S$. A consequence of weak convergence is that if f_n is a uniformly bounded and equicontinuous family, then, $P_n \to P$ weakly, [14, Corollary 11.3.4] implies that $\int (P_n - P)(dx) f_n(x) \to 0$. Condition (2.1) ensures that the sequence of functions $\int Q(\mathrm{d}y|x_k) c(x_k,\gamma_n(y))$ is equicontinuous as it gives us a uniform continuity over family of all admissible policies, and the result follows. $\qquad\square$

Example 2.2. Consider the following additive noisy channel:

$$y = x + w,$$

where $w \sim \mu$ with μ admitting a density, η, which is continuous. An example is the Gaussian density. Suppose that $\mathbb{U}$ is compact and $c(x,u)$ is continuous and bounded in $\mathbb{X} \times \mathbb{U}$. For $x_k \to x$ we have that

$$\begin{aligned} &\lim_{k\to\infty} \sup_\gamma \left| \int \eta(y - x_k) c(x_k,\gamma(y)) \mathrm{d}y - \int \eta(y - x) c(x,\gamma(y)) \mathrm{d}y \right| \\ &\leq \lim_{k\to\infty} \sup_\gamma \left| \int (\eta(y - x_k) - \eta(y - x)) c(x_k,\gamma(y)) \mathrm{d}y + \cdots \right. \\ &\qquad\qquad \left. + \int \eta(y - x) \left(c(x_k,\gamma(y)) - c(x,\gamma(y)) \right) \mathrm{d}y \right| \end{aligned}$$

$$(2.8) \qquad \leq \|\eta(\cdot - x_k) - \eta(\cdot - x)\|_{TV} \cdot \|c\|_\infty + \cdots$$

$$(2.9) \qquad + \lim_{k\to\infty} \sup_\gamma \left| \int \eta(y - x) \left(c(x_k,\gamma(y)) - c(x,\gamma(y)) \right) \mathrm{d}y \right|$$

$$= 0,$$

where $\|c\|_\infty$ denotes the supremum norm of c. We note that the term in (2.8) tends to zero since $\eta(\cdot - x_k)$ converges to $\eta(\cdot - x)$ pointwise and therefore by Scheffé's theorem it converges in L_1 and thus in total variation. Additionally, the term in (2.9) tends

to zero since c is uniformly continuous by assumptions. Therefore Assumption 2.1 is satisfied, and Theorem 2.5 holds. Thus, for a special but practically important class of channels weak convergence of priors is sufficient for continuity if further $c(x, u)$ is bounded and continuous on $\mathbb{X} \times \mathbb{U}$ and $\mathbb{U}$ is compact.

We note here that a related result due to Wu and Verdú [51] establishes continuity of the minimum-mean-square-error (MMSE), that is, with $c(x, u) = \|x - u\|^2$, under weak convergence when the channel is additive, the additive noise has a finite variance and it admits a continuous and bounded density function. In general, however, the following example shows that the boundedness condition cannot be relaxed even when the channel is noninformative, which can be viewed as an extreme form of regularity.

Example 2.3. Let $\mathbb{X} = \mathbb{U} = \mathbb{R}$, $\mathbb{Y} = [0, 1]$, and $c(x, u) = (x - u)^2$. With the given cost function this is a mean-square error problem; therefore, the optimal policy is $\gamma^*(y) = E[x|y]$. We let the channel be distributed uniformly on $[0, 1]$, that is, $Q \sim U([0, 1])$. Note that this channel is noninformative. Let P_n be the following discrete distribution:

$$P_n = \left(\frac{1}{2} - \frac{1}{n}\right) \cdot \delta_{\frac{1}{n}} + \left(\frac{1}{2} - \frac{1}{n}\right) \cdot \delta_{-\frac{1}{n}} + \frac{1}{2n} \cdot \delta_{a_n} + \frac{1}{2n} \cdot \delta_{-a_n},$$

a_n is a sequence of numbers defined by

$$a_n = \sqrt{n - \left(\frac{1}{n} + \frac{2}{n^2}\right)}.$$

Clearly $P_n \to \delta_0$ weakly: as for any bounded and continuous function f we have

$$\int_{\mathbb{R}} f(x) P_n(\mathrm{d}x) =$$

$$\left(\frac{1}{2} - \frac{1}{n}\right) \cdot f\left(\frac{1}{n}\right) + \left(\frac{1}{2} - \frac{1}{n}\right) \cdot f\left(-\frac{1}{n}\right) + \frac{1}{2n} \cdot f(a_n) + \frac{1}{2n} \cdot f(-a_n)$$

$$\to f(0) = \int_{\mathbb{R}} f(x) \delta_0(\mathrm{d}x)$$

by boundedness and continuity. By symmetry and the noninformative nature of Q, the optimal policy is $\gamma^*(y) = E[X|Y] = 0$ for all P_n and for $P = \delta_0$. With initial distribution P, we have

$$J^*(P, Q) = E_P^{Q, \gamma^*}[(X - U)^2]$$
$$= E_P^{Q, \gamma^*}[(X - \gamma^*(Y))^2]$$
$$= E_P^{Q, \gamma^*}[(X)^2] = 0.$$

Whereas for all $n \in \mathbb{N}$ we have

$$J^*(P_n, Q) = E_{P_n}^{Q, \gamma^*}[(X)^2]$$
$$= \left(\frac{1}{2} - \frac{1}{n}\right) \cdot \frac{1}{n^2} + \left(\frac{1}{2} - \frac{1}{n}\right) \cdot \frac{1}{n^2} + \frac{1}{2n} \cdot a_n^2 + \frac{1}{2n} \cdot a_n^2$$
$$= \left(1 - \frac{2}{n}\right) \cdot \frac{1}{n^2} + \frac{1}{n} \cdot \left(n - \left(\frac{1}{n} + \frac{2}{n^2}\right)\right)$$
$$= 1,$$

so $J^*(P_n, Q) \nrightarrow J^*(P, Q)$ as $n \to \infty$.

Now we present a result for continuity under the Wasserstein metric.

Assumption 2.2. There exists a measurable non-negative function f so that for some probability measure P_Q, the following (absolute continuity condition) holds:

$$Q(Y \in A | x) = \int_A f(x, y) P_Q(dy).$$

Furthermore, $\tilde{c}(x, y, u) := c(x, u) f(x, y)$ is such that

$$|\tilde{c}(x', y, u) - \tilde{c}(x, y, u)| \le \alpha |x' - x|$$

for all $y \in \mathbb{Y}, u \in \mathbb{U}$ and for some $\alpha \in \mathbb{R}_+$.

THEOREM 2.6. *Under Assumption 2.2,*

$$|J^*(P_n, Q) - J^*(P, Q)| \le \alpha W_1(P_n, P),$$

and thus, as $W_1(P_n, P) \to 0$, $J^*(P_n, Q) \to J^*(P, Q)$.

Proof. We first use the bound in (2.6) such that

$$\begin{aligned}
&|J^*(P, Q) - J^*(P_n, Q)| \\
&\qquad \le \max(J(P, Q, \gamma_n^*) - J(P_n, Q, \gamma_n^*), J(P_n, Q, \gamma^*) - J(P, Q, \gamma^*)).
\end{aligned}$$

Let $\mathbb{P}_n$ denote a product measure on the space $\mathbb{X} \times \mathbb{X}$ such that its first marginal is P and the second marginal is P_n. Then, for any $\gamma \in \Gamma$ we have

$$\begin{aligned}
&\int \mathbb{P}_n(dx, dx') \left(\int Q(dy|x) c(x, \gamma(y)) - \int Q(dy|x') c(x', \gamma(y)) \right) \\
&\quad = \int \mathbb{P}_n(dx, dx') \left(\int P_Q(dy) f(x, y) c(x, \gamma(y)) - \int P_Q(dy) f(x', y) c(x', \gamma(y)) \right) \\
&\quad = \int \mathbb{P}_n(dx, dx') \int P_Q(dy) \left(f(x, y) c(x, \gamma(y)) - f(x', y) c(x', \gamma(y)) \right) \\
&\quad \le \int \mathbb{P}_n(dx, dx') \int P_Q(dy) \alpha |x' - x| \\
&\quad \le \int \mathbb{P}_n(dx, dx') \alpha |x' - x|.
\end{aligned}$$

Optimizing over all such couplings $\mathbb{P}_n$ completes the proof. $\square$

2.1.2. Setwise convergence.

THEOREM 2.7. *Let a channel $Q \in \mathcal{Q}$ be given. $J^*(P, Q)$ is not necessarily continuous in P under setwise convergence. This holds even when $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{U}$ are compact and when c is bounded and continuous in both x and u.*

Proof. We present the following counterexample, building on [52]: let $\mathbb{X} = \mathbb{Y} = \mathbb{U} = [0, 1]$, and let $c(x, u) = (x - u)^2$. For $n \in \mathbb{N}$ and $k = 1, \ldots, n$, we define

$$L_{n,k} = \left[\frac{2k-2}{2n}, \frac{2k-1}{2n} \right), \; R_{n,k} = \left[\frac{2k-1}{2n}, \frac{k}{n} \right).$$

For ease of notation, we shall let $L = \{y \in \cup_{k=1}^n L_{n,k}\}$ and $R = \{y \in \cup_{k=1}^n R_{n,k}\}$. Next, we define the square-wave function by

$$h_n(t) = 1_{\{t \in L\}} - 1_{\{t \in R\}}.$$

As $\int_0^1 h_n(t)\mathrm{d}t = 0$ and $|h_n(t)| \leq 1$, the function

$$f_n(t) = (1 + h_n(t))1_{\{t\in[0,1]\}}$$

is a probability density function.

By the proof of the Riemann–Lebesgue lemma (for example, see Theorem 12.21 in [48]), we have

$$\lim_{n\to\infty} \int_0^1 h_n(t)g(t)\mathrm{d}t = 0 \text{ for all } g \in L_1\left([0,1],\mathbb{R}\right),$$

therefore

$$\lim_{n\to\infty} \int_0^1 f_n(t)g(t)\mathrm{d}t = \int_0^1 g(t) \text{ for all } g \in L_1\left([0,1],\mathbb{R}\right).$$

So if we let $P_n \sim f_n$ for $n \in \mathbb{N}$, we have that $P_n \to P \sim U([0,1])$ setwise. Next we let the channel be

$$Q(\cdot|x) \sim \frac{1}{2} \cdot \delta_x + \frac{1}{2} \cdot U([0,1]).$$

For initial distribution P, the optimal policy is

$$\gamma_P^*(y) = E[X|Y] = \frac{1}{2}\left(\frac{1}{2} + y\right).$$

This gives

$$J^*(P,Q) = \frac{1}{16}.$$

By tedious calculations (see section B.1), the optimal policy for initial distribution P_n is

$$\gamma_{P_n}^*(y) = \begin{cases} \frac{1}{2} - \frac{1}{4n} & \text{if } y \in \cup_{k=1}^n R_{n,k}, \\ \frac{1}{3} \cdot \left(\frac{1}{2} - \frac{1}{4n}\right) + \frac{2}{3}y & \text{if } y \in \cup_{k=1}^n L_{n,k}. \end{cases}$$

This gives

$$J^*(P_n,Q) = \frac{1}{18} - \frac{1}{24n^2}.$$

So we have $J^*(P_n,Q) \to \frac{1}{18} \neq \frac{1}{16}$ as $n \to \infty$, and we see that the optimal cost is clearly not continuous on the space of initial distributions under setwise convergence. □

The next result shows that the optimal cost is upper semicontinuous under setwise convergence too.

THEOREM 2.8. *Let a channel Q be given. If $c(x,u)$ is non-negative, measurable, and bounded in $\mathbb{X} \times \mathbb{U}$, then $J^*(P,Q)$ is upper semicontinuous on $\mathcal{P}(\mathbb{X})$ under setwise convergence.*

Proof. Given a fixed channel $Q \in \mathcal{Q}$, let $P_n \to P$ in $\mathcal{P}(\mathbb{X})$ setwise. Then

$$\limsup_{n\to\infty} \inf_{\gamma\in\Gamma} \int_{\mathbb{X}\times\mathbb{Y}} c(x,\gamma(y))Q(\mathrm{d}y|x)P_n(\mathrm{d}x)$$

$$\leq \inf_{\gamma\in\Gamma} \limsup_{n\to\infty} \int_{\mathbb{X}\times\mathbb{Y}} c(x,\gamma(y))Q(\mathrm{d}y|x)P_n(\mathrm{d}x)$$

$$= \inf_{\gamma\in\Gamma} \int_{\mathbb{X}\times\mathbb{Y}} c(x,\gamma(y))Q(\mathrm{d}y|x)P(\mathrm{d}x),$$

where for the last equality we used Lemma 1.1 and the fact that c is bounded and measurable. □

2.1.3. Continuity under total variation. The proof of the following result builds on [52, Thm. 3.4].

THEOREM 2.9. *The optimal cost $J^*(P, Q)$ is continuous on the set of input distributions, $\mathcal{P}(\mathbb{X})$, under the topology of total variation. In other words, if $\|P_n - P\|_{TV} \to 0$, then $|J^*(P_n, Q) - J^*(P, Q)| \to 0$.*

Proof. Let $P_n \to P$ in total variation with a fixed channel Q. Recall from section 1 that $J(P', Q, \gamma') = E_{P'}^{Q, \gamma'}[c(x, u)]$, that is, the expected cost with initial distribution $P' \in \mathcal{P}(\mathbb{X})$ and control policy $\gamma' \in \Gamma$. Let the optimal (ε-optimal) policies are given by γ_n^* and γ^* for initial distributions P_n and P, respectively. Using the bound in 2.6 we write

$$|J^*(P_n, Q) - J^*(P, Q)|$$
$$\leq \max(J(P, Q, \gamma_n^*) - J(P_n, Q, \gamma_n^*), J(P_n, Q, \gamma^*) - J(P, Q, \gamma^*)).$$

As c is bounded it follows that for any $\gamma' \in \Gamma$,

$$|J(P_n, Q, \gamma') - J(P, Q, \gamma')|$$
(2.10)
$$= \left| \int c(x, \gamma'(y)) P_n Q(\mathrm{d}x, \mathrm{d}y) - \int c(x, \gamma'(y)) PQ(\mathrm{d}x, \mathrm{d}y) \right|$$
$$\leq \|c\|_\infty \|P_n Q - PQ\|_{TV} = \|c\|_\infty \|P_n - P\|_{TV},$$

where we have used Lemma 1.1 for the last equality. Inequalities (2.6) and (2.10) together imply that $|J^*(P, Q) - J^*(P_n, Q)| \leq \|c\|_\infty \|P_n - P\|_{TV}$. Since $\|P_n - P\|_{TV} \to 0$ we have that $J^*(P_n, Q) \to J^*(P, Q)$. □

Remark 1. In this paper we only focus on the case where the channel is known by the controller; that is, the true channel model Q is available to the controller. For the case where this is no longer true, some further analysis is required. If the prior model P and the channel Q are not known, the controller can have an estimating sequence $P_n Q_n \in \mathcal{P}(\mathbb{X} \times \mathbb{Y})$ for the true joint measure $PQ \in \mathcal{P}(\mathbb{X} \times \mathbb{Y})$. The question then becomes analyzing the convergence of $P_n Q_n \to PQ$. This joint convergence might require a different set of assumptions on P_n and $Q_n(\cdot | x)$ which we do not discuss in this paper. Lemma 1.1 might give an idea on this joint convergence where we consider the convergence of the joint measure $P_n Q$ to PQ. In [52], a similar joint convergence is studied for convergence of measurement channels and fixed prior distributions. The reader can also refer to [52, Lem. 2.2] for an analysis on the convergence of $PQ_n \to PQ$.

2.2. Multistage and infinite horizon discounted setup. We now consider continuity problems for the multistage case. For this case, our focus will be on the infinite stage discounted cost setting. Clearly, the lack of continuity for single-stage problems implies the lack of continuity of multistage problems. In the following, the emphasis will be on developing setups where continuity can be established.

In particular, if we put further restrictions on the system model and the measurement channel, we will establish sufficient conditions for continuity under weak convergence of the priors. This will be studied in the following.

2.2.1. Weak convergence. Consider a partially observed Markov decision process (POMDP), with state space $\mathbb{X}$, action space $\mathbb{A}$, and observation space $\mathbb{Y}$, all Borel spaces. Define the history spaces $\mathbb{H}_t = (\mathbb{Y} \times \mathbb{U})^t \times \mathbb{Y}$, $t = 0, 1, 2, \ldots$ endowed with their product Borel σ-algebras generated by $\mathcal{B}(\mathbb{Y})$ and $\mathcal{B}(\mathbb{U})$. A *policy* $\pi = \{\pi_t\}$ is a sequence of stochastic kernels on $\mathbb{U}$ given $\mathbb{H}_t$. We denote by Π the set of all policies. For any initial distribution μ and policy π we can think of the POMDP as

a stochastic process $\{X_t, Y_t, U_t\}_{t \geq 0}$ defined on the probability space $(\Omega, \mathcal{B}(\Omega), P_\mu^\pi)$, where $\Omega = \mathbb{H}_\infty \times \mathbb{X}^\infty$, the X_t are $\mathbb{X}$-valued random variables, the Y_t are $\mathbb{Y}$-valued random variables, the U_t are $\mathbb{U}$-valued random variables.

It is known that any POMDP can be reduced to a (completely observable) Markov decision process (MDP) [53, 39], whose states are the posterior state distributions or "beliefs" of the observer; that is, the state at time t is

$$Z_t(\cdot) := \Pr\{X_t \in \cdot | Y_0, \ldots, Y_t, U_0, \ldots, U_{t-1}\} \in \mathcal{P}(\mathbb{X}).$$

We call this equivalent MDP the belief-MDP. The belief-MDP has state space $\mathbb{Z} = \mathcal{P}(\mathbb{X})$ and action space $\mathbb{U}$. Recall that $\mathbb{Z}$ is equipped with the Borel σ-algebra generated by the topology of weak convergence [5]. Since $\mathbb{X}$ is a Borel space, $\mathbb{Z}$ is metrizable with the Prokhorov metric which makes $\mathbb{Z}$ into a Borel space [35]. The transition probability η of the belief-MDP can be constructed as follows (see also [26]). If we define the measurable function $F(z, a, y) := \Pr\{X_{t+1} \in \cdot | Z_t = z, U_t = u, Y_{t+1} = y\}$ from $\mathbb{Z} \times \mathbb{A} \times \mathbb{Y}$ to $\mathbb{Z}$ and the stochastic kernel $H(\cdot | z, u) := \Pr\{Y_{t+1} \in \cdot | Z_t = z, U_t = u\}$ on $\mathbb{Y}$ given $\mathbb{Z} \times \mathbb{U}$, then η can be written as

$$\eta(\cdot | z, u) = \int_{\mathbb{Y}} 1_{\{F(z, u, y) \in \cdot\}} H(dy | z, u).$$

The one-stage cost function c of the belief-MDP is given by

$$(2.11) \qquad \tilde{c}(z, u) := \int_{\mathbb{X}} c(x, u) z(dx).$$

Hence, the belief-MDP is a MDP with the components $(\mathbb{Z}, \mathbb{U}, \eta, \tilde{c})$.

It is a standard result that an optimal control policy will use the belief z_t as a sufficient statistic for optimal policies (see [53, 39]).

Assumption 2.3. (a) The stochastic kernel $\mathcal{T}(dx_1 | x_0 = x, u_0 = u)$ is weakly continuous in (x, u).
(b) Assumption 2.1 holds; that is, the observation channel $Q(dy | x)$ is continuous in total variation.
(c) The stagewise cost function $c(x, u)$ is non-negative, bounded, and continuous on $\mathbb{X} \times \mathbb{U}$.
(d) $\mathbb{U}$ is compact.

By [4, Prop. 7.30], the one-stage cost function $\tilde{c}$ of the belief-MDP, which is defined in (2.11), is continuous and bounded, that is, in $C_b(\mathbb{Z} \times \mathbb{U})$, under Assumption 2.3(a),(b). The following theorem is from [19, Ex. 3.7, 4.1] and [41, Ex. 2.1].

THEOREM 2.10. (i) *Under Assumption* 2.3, *the stochastic kernel η for belief-MDP is weakly continuous in (z, u).*
(ii) *If we relax the continuity in total variation of the observation channel to weak continuity, then η may not be weakly continuous even if the transition probability p of POMDP is continuous in total variation.*
(iii) *η may not be setwise continuous in u even if the observation channel is continuous in total variation.*

For an infinite horizon discounted cost problem with bounded costs, an optimal policy can be computed through the iterated use of the *discounted cost optimality operator* to be introduced below. Note that under Assumption 2.3(c), by an application of the dominated convergence theorem $\tilde{c}$ is continuous and bounded. Now, for a

MDP with weakly continuous transition probabilities and compact action spaces, it follows that an optimal control policy exists and that the optimal cost is continuous in the initial state (or probability measure in the context here): this follows because the *discounted cost optimality operator* $T : C_b(\mathbb{Z}) \to C_b(\mathbb{Z})$ (see, e.g., [28, Chap. 8.5]):

$$(T(v))(z) = \min_u(\tilde{c}(z, u) + \beta E[v(z_1)|z_0 = z, u_0 = u])$$

is a contraction from $C_b(\mathbb{Z})$ to itself under the supremum norm. As a result, there exists a fixed point, which is *continuous*. This fixed point is the value function.

This argument shows that the value function is continuous in the belief state, $z_0(Y)(x \in \cdot) = P(x \in \cdot|Y)$, which is the *posterior* distribution of the state variable given the observations. However, convergence of the distribution of the priors may not always imply the convergence of the posteriors: for the single-stage case, with the cost function of the belief process $\tilde{c}$ defined as in (2.11), the value function is given by $J^*(z) = \inf_{u \in \mathbb{U}} \tilde{c}(z, u)$. It can be seen that the value function is again continuous in the belief state z if c is continuous in x and $\mathbb{U}$ is compact (by an application of the dominated convergence theorem). However, as we have seen in the counterexample used to prove Theorem 2.2; even though $P_n \to P$ weakly, value functions do not converge. The next theorem shows that with further conditions, convergence of the posteriors (belief states) can also be guaranteed. Before the main result we first present a key lemma.

LEMMA 2.11. *Under Assumption* 2.3*, as* $P_n \to P$ *weakly,*

$$\sup_{\gamma \in \Gamma} \left| E_P\big[c(X_k, \gamma(Y_{[0,k]}))\big] - E_{P_n}\big[c(X_k, \gamma(Y_{[0,k]}))\big] \right| \to 0$$

for any time stage $k < \infty$.

Proof. The proof can be found in the Appendix A.1. □

THEOREM 2.12. *Suppose that Assumption* 2.3 *holds. Then, as* $P_n \to P$ *weakly* $|J^*_\beta(P_n, \mathcal{T}) - J^*_\beta(P, \mathcal{T})| \to 0$.

Proof. We start with the following bound;

$$\begin{aligned}(2.12) \quad &|J^*_\beta(P, \mathcal{T}) - J^*_\beta(P_n, \mathcal{T})| \\ &\leq \max(J_\beta(P, \mathcal{T}, \gamma^*_n) - J(P_n, \mathcal{T}, \gamma^*_n), J_\beta(P_n, \mathcal{T}, \gamma^*) - J(P, \mathcal{T}, \gamma^*)).\end{aligned}$$

Now, we try to show that under Assumption 2.3, as $P_n \to P$ weakly,

$$(2.13) \qquad \sup_{\gamma \in \Gamma} |J_\beta(P_n, \mathcal{T}, \gamma) - J_\beta(P, \mathcal{T}, \gamma)| \to 0.$$

To prove (2.13) we start with the following inequality:

$$\begin{aligned}\sup_{\gamma \in \Gamma} &|J_\beta(P_n, \mathcal{T}, \gamma) - J_\beta(P, \mathcal{T}, \gamma)| \\ &\leq \sup_{\gamma \in \Gamma} \sum_{k=0}^T \beta^k \left| E_P\big[c(X_k, \gamma(Y_{[0,k]}))\big] - E_{P_n}\big[c(X_k, \gamma(Y_{[0,k]}))\big] \right| + \sum_{k=T}^\infty \beta^k 2\|c\|_\infty.\end{aligned}$$

First we fix $\epsilon > 0$ and find a T_ϵ such that $\sum_{k=T_\epsilon}^\infty \beta^k 2\|c\|_\infty \leq \epsilon/2$. Now, we claim that we can find an N such that for every $n > N$

$$\sup_{\gamma \in \Gamma} \sum_{k=0}^{T_\epsilon - 1} \left| E_{P_n} \left[c(X_k, \gamma(Y_{[0,k]})) \right] - E_P \left[c(X_k, \gamma(Y_{[0,k]})) \right] \right|$$

$$\leq \sum_{k=0}^{T_\epsilon - 1} \sup_{\gamma \in \Gamma} \left| E_{P_n} \left[c(X_k, \gamma(Y_{[0,k]})) \right] - E_P \left[c(X_k, \gamma(Y_{[0,k]})) \right] \right| \leq \epsilon/2.$$

Lemma 2.11 implies that for every time stage $k < T_\epsilon$, $\sup_{\gamma \in \Gamma} |E_P[c(X_k, \gamma(Y_{[0,k]}))] - E_{P_n}[c(X_k, \gamma(Y_{[0,k]}))]|$ can be made less than $\epsilon/2T_\epsilon$ for all $n > N$ for some $N < \infty$. Since $T_\epsilon < \infty$, we can find a common N for all time stages $k < T_\epsilon$. So we can write

$$\sup_{\gamma \in \Gamma} \sum_{k=0}^{T_\epsilon - 1} \beta^k \left| E_P \left[c(X_k, \gamma(Y_{[0,k]})) \right] - E_{P_n} \left[c(X_k, \gamma(Y_{[0,k]})) \right] \right| < \epsilon/2.$$

Combining the results, we have

$$\sup_{\gamma \in \Gamma} |J_\beta(P_n, \mathcal{T}, \gamma) - J_\beta(P, \mathcal{T}, \gamma)| \leq \sum_{k=0}^{T_\epsilon - 1} \beta^k \left| E_P \left[c(X_k, \gamma(Y_{[0,k]})) \right] - E_{P_n} \left[c(X_k, \gamma(Y_{[0,k]})) \right] \right|$$

$$+ \sum_{k=T_\epsilon}^{\infty} \beta^k 2 \|c\|_\infty$$

$$< \epsilon$$

for all $n > N$ for some $N < \infty$ for every given $\epsilon > 0$, which proves (2.13).

Now looking at the term in (2.12):

$$\max(J_\beta(P, \mathcal{T}, \gamma_n^*) - J(P_n, \mathcal{T}, \gamma_n^*), J_\beta(P_n, \mathcal{T}, \gamma^*) - J(P, \mathcal{T}, \gamma^*)),$$

we can see that both terms go to zero using (2.13). $\quad\square$

In the following we give a weaker result, which holds with no restrictions, however.

2.2.2. Continuity under total variation and strategic measures. For stochastic control problems, *strategic measures* are defined (see Schäl [42], also [17, 18]) as the set of probability measures induced on the product spaces of the state and action pairs by measurable control policies: given an initial distribution on the state, and a policy, one can uniquely define a probability measure on the infinite product space consistent with finite dimensional distributions, by the Ionescu–Tulcea theorem [27]. Now, define a *strategic measure* under a policy $\gamma^n = \{\gamma_0^n, \gamma_1^n, \ldots, \gamma_k^n, \ldots\}$ as a probability measure defined on $\mathcal{B}(\mathbb{X} \times \mathbb{Y} \times \mathbb{U})^{\mathbb{Z}_+}$ by:

$$P_{P_n}^{\gamma^n}(\mathrm{d}(x_0, y_0, u_0), \mathrm{d}(x_1, y_1, u_1), \ldots)$$
$$= P_n(\mathrm{d}x_0) Q(\mathrm{d}y_0|x_0) 1_{\{\gamma^n(y_0) \in \mathrm{d}u_0\}} \mathcal{T}(\mathrm{d}x_1|x_0, u_0) Q(\mathrm{d}y_1|x_1) 1_{\{\gamma^n(y_0, y_1) \in \mathrm{d}u_1\}} \cdots$$

Under a strategic measure $P_{P_n}^{\gamma^n}$ we define,

$$J_\beta^*(P_n, \mathcal{T}) = \inf_{\gamma^n} E_{P_n}^{\gamma^n} \left[\sum_k \beta^k c(x_k, \gamma_k^n(y_{[0,k]})) \right].$$

THEOREM 2.13. *If $c(x, u)$ is a non-negative, measurable, and bounded function in $\mathbb{X} \times \mathbb{U}$, then*

$$|J_\beta^*(P_n, \mathcal{T}) - J_\beta^*(P, \mathcal{T})| \leq \|P_n(x_0 \in \cdot) - P(x_0 \in \cdot)\|_{TV} \frac{1}{1 - \beta} \|c\|_\infty.$$

Proof. From inequalities (2.4), (2.5), and (2.6) we have that $|J_\beta^*(P_n, \mathcal{T}) - J_\beta^*(P, \mathcal{T})|$ is upper bounded as follows:

$$|J_\beta^*(P_n, \mathcal{T}) - J_\beta^*(P, \mathcal{T})|$$

$$\leq \max\bigg(\sum_k \beta^k \|P_P^{\gamma^n}(x_k \in \cdot, y_{[0,k]} \in \cdot) - P_{P_n}^{\gamma^n}(x_k \in \cdot, y_{[0,k]} \in \cdot)\|_{TV} \sup c(x, \gamma^n(y_{[0,k]})),$$

$$\sum_k \beta^k \|P_{P_n}^{\gamma}(x_k \in \cdot, y_{[0,k]} \in \cdot) - P_P^{\gamma}(x_k \in \cdot, y_{[0,k]} \in \cdot)\|_{TV} \sup c(x, \gamma(y_{[0,k]}))\bigg).$$

For any $\gamma \in \Gamma$, we have

$$\sum_k \beta^k \|P_{P_n}^{\gamma}(x_k \in \cdot, y_{[0,k]} \in \cdot) - P_P^{\gamma}(x_k \in \cdot, y_{[0,k]} \in \cdot)\|_{TV} \sup c(x, \gamma(y_{[0,k]}))$$

$$= \sum_k \beta^k \|P_P^{\gamma}(x_k \in \cdot, y_{[0,k]} \in \cdot) - P_{P_n}^{\gamma}(x_k \in \cdot, y_{[0,k]} \in \cdot)\|_{TV} \|c\|_\infty$$

$$(2.14) \qquad \leq \frac{1}{1-\beta} \|c\|_\infty \|P_n(x_0 \in \cdot) - P(x_0 \in \cdot)\|_{TV}.$$

Here, we use the property that

$$\|P_{P_n}^{\gamma}(x_k \in \cdot, y_{[0,k]} \in \cdot) - P_P^{\gamma}(x_k \in \cdot, y_{[0,k]} \in \cdot)\|_{TV}$$

$$\leq \|P_{P_n}^{\gamma}((x,y,u)_{[0,\infty)} \in \cdot) - P_P^{\gamma}((x,y,u)_{[0,\infty)} \in \cdot)\|_{TV} = \|P_n(x_0 \in \cdot) - P(x_0 \in \cdot)\|_{TV},$$

similar to the derivation in Lemma 1.1. $\qquad\square$

2.3. Some remarks on the infinite horizon average cost setup. In this section, we show that the conditions presented earlier may not lead to continuity for the infinite horizon setup but alternative conditions will likely be useful. Consider an infinite horizon average cost setup with the objective function given by

$$J_\infty(P, \mathcal{T}, \gamma) = \limsup_{T \to \infty} \frac{1}{T} E_P^{Q,\gamma}\left[\sum_{t=0}^{T-1} c(X_t, U_t)\right].$$

The optimal cost is given by

$$J_\infty^*(P, \mathcal{T}) = \inf_{\gamma \in \Gamma} J_\infty(P, \mathcal{T}, \gamma).$$

Given an optimal stationary policy, whose existence follows from the conditions given in Assumption 2.3, say, through the convex analytic method (see Borkar [7]), the process process π_k becomes Markovian. It is known that (see [29, Thm. 2.3.4–2.3.5]) if this belief Markov process admits a unique invariant measure, then

$$\frac{1}{N} E_{\pi_0}\left[\sum_{k=0}^{N-1} c(\pi_k)\right] \to c$$

almost surely for all initial conditions (that is, "priors") for some constant c. Thus, if the belief process has a unique invariant measure, the continuity in priors holds immediately as the average cost does not depend on the prior measure provided that the initial prior belongs to the support set of the invariant measure. However, checking ergodicity is a challenging problem for the controlled setup; this is a subject of current research.

On the other hand, we show in the following that Assumption 2.3 does not alone guarantee continuity.

Example 2.4. Let $\mathbb{X} = \mathbb{R}$, $\mathbb{U} = [-1, 1]$, $\mathbb{Y} = [-1, 1]$. Suppose we are given two initial distributions, the transition kernel and the measurement channel as

$$P(\cdot) = \delta_0(\cdot),$$
$$P_n(\cdot) = \frac{1}{2}\delta_{1/n}(\cdot) + \frac{1}{2}\delta_{-1/n}(\cdot),$$
$$\mathcal{T}(\cdot|x, u) = \delta_{2x}(\cdot),$$
$$Q(\cdot|x) = U[-1, 1].$$

Notice that the transition kernel is weakly continuous in (x, u); the measurement channel is continuous in total variation and $P_n \to P$ weakly.

The stagewise cost function is defined by as follows.

$$c(x, u) = \begin{cases} (x + u)^2 & \text{if } |x| \leq 1, \\ (1 + u)^2 & \text{if } |x| > 1. \end{cases}$$

So the cost is always bounded. The optimal control actions for both initial distributions at any time $k \geq 0$ are $\gamma(y_{[0,k]}) = 0$ and $\gamma_n(y_{[0,k]}) = 0$.

It is easy to see that the optimal cost for P is 0. The optimal cost for P_n can be calculated as follows:

$$J_\infty^*(P_n, \mathcal{T}) = \lim_{N \to \infty} \frac{1}{N} \left(\sum_{k=0}^{\log_2 n} \left(\frac{2^k}{n}\right)^2 + \sum_{k=\log_2 n + 1}^{N} (1)^2 \right)$$
$$= \lim_{N \to \infty} \frac{1}{N} \left(\frac{4(n^2)}{3} - \frac{1}{3} + N - \log_2 n + 1 \right) = 1 \neq 0.$$

A more complete treatment for the average cost case will be reported in future work.

3. Robustness.

3.1. Robustness to incorrect priors and mismatch bounds for single-stage problems. First, for this subsection, we will consider the single-stage stochastic control problem with cost function c, initial distribution P. We shall denote this problem by $\Xi = (c, P, Q)$. Consider the following problem: let $\widetilde{P}$ be another initial distribution. Decision maker DM computes an optimal policy, $\widetilde{\gamma}^*$, for the problem $\widetilde{\Xi} = (c, \widetilde{P}, Q)$ and applies it to Ξ. Can we approximate the loss in performance, that is, can we find a bound on $J(P, Q, \widetilde{\gamma}^*) - J^*(P, Q)$? This situation naturally arises when the initial distribution, P, is uncertain and DM has a prior belief, $\widetilde{P}$, which is perhaps based on an incorrect initial model.

PROPOSITION 3.1. *Assume that $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is nonnegative, measurable, and bounded. Let $\widetilde{\gamma}^*$ be an optimal (or ε-optimal) control policy for the single-stage stochastic control problem $\widetilde{\Xi} = (c, \widetilde{P}, Q)$, where c is a cost function, $\widetilde{P}$ is an initial distribution, and Q is a measurement channel. Let P be another probability distribution on $\mathbb{X}$. Then $|J(P, Q, \widetilde{\gamma}) - J^*(P, Q)| \leq 2\|c\|_\infty \|P - \widetilde{P}\|_{TV}$.*

Proof. Let γ^* denote an optimal (or ε-optimal) policy for the problem $\Xi = (c, P, Q)$. We have

$$|J(P, Q, \widetilde{\gamma}^*) - J^*(P, Q)| = |J(P, Q, \widetilde{\gamma}^*) - J(\widetilde{P}, Q, \widetilde{\gamma}^*) + J(\widetilde{P}, Q, \widetilde{\gamma}^*) - J(P, Q, \gamma^*)|$$
$$\leq |J(P, Q, \widetilde{\gamma}^*) - J(\widetilde{P}, Q, \widetilde{\gamma}^*)| + |J^*(\widetilde{P}, Q) - J^*(P, Q)| \ (+2\varepsilon)$$
$$\leq 2\|c\|_\infty \|P - \widetilde{P}\|_{TV},$$

where we have used (2.10) for the final inequality. $\square$

Similar to the continuity section, for the weak convergence of priors we have a negative result. The following result implies that the mismatch error may not diminish even if our belief model converges weakly to the true model.

PROPOSITION 3.2. *Let an initial distribution $P \in \mathcal{P}(\mathbb{X})$ and a cost function $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ be given. Assume that $P_n \to P$ weakly, and let $\gamma_{P_n}^*$ be an optimal (or ε-optimal) policy for the control problem $\Xi_n = (c, P_n)$. It does not follow that $J(P, Q, \gamma_{P_n}^*) \to J^*(P, Q)$ as $n \to \infty$. This result holds even if c is bounded and continuous in both x and u.*

Proof. We use an extension of the proof from Theorem 2.7. Let $\mathbb{X} = \mathbb{Y} = \mathbb{U} = [0, 1]$, and let $c(x, u) = (x - u)^2$. We let $P_n \sim f_n$ for $n \in \mathbb{N}$ as in the proof of Theorem 2.7. We have that $P_n \to P \sim U([0, 1])$ setwise. We let the channel be

$$Q(\cdot|x) \sim \frac{1}{2} \cdot \delta_x + \frac{1}{2} \cdot U([0, 1]).$$

As presented in the proof of Theorem 2.7, the optimal policy for the control problem $\Xi_n = (c, P_n, Q)$ is given by

$$\gamma_{P_n}^*(y) = \begin{cases} \frac{1}{2} - \frac{1}{4n} & \text{if } y \in \cup_{k=1}^n R_{n,k}, \\ \frac{1}{3} \cdot \left(\frac{1}{2} - \frac{1}{4n} \right) + \frac{2}{3} y & \text{if } y \in \cup_{k=1}^n L_{n,k}. \end{cases}$$

If the decision maker applies $\gamma_{P_n}^*$ to the control problem $\Xi = (c, P, Q)$, this results in the following cost (the calculations can be found in B.2):

$$J(P, Q, \gamma_{P_n}^*) = \frac{2}{27} + \frac{5}{72n^2}.$$

We notice that as $n \to \infty$, $|J(P, Q, \widetilde{\gamma}^*) - J^*(P, Q)| \to \frac{2}{27} - \frac{1}{16} = \frac{5}{432} \neq 0$.   □

We now present a positive result for weak convergence.

THEOREM 3.1. *Suppose that $c(x, u)$ is bounded, $\mathbb{U}$ is compact, and Assumption 2.1 holds. Then, as $P_n \to P$ weakly $\lim_{n \to \infty} |J(P, Q, \gamma_{P_n}^*) - J^*(P, Q)| \to 0$; that is, the system is robust to errors in the priors under weak convergence.*

Proof.

$$\lim_{n \to \infty} |J(P, Q, \gamma_{P_n}^*) - J^*(P, Q)|$$

$$(3.1) \qquad \leq \lim_{n \to \infty} |J(P, Q, \gamma_{P_n}^*) - J(P_n, Q, \gamma_{P_n}^*)| + \lim_{n \to \infty} |J(P_n, Q, \gamma_{P_n}^*) - J(P, Q)|$$

$$= 0.$$

We note that the first term goes to 0 with the same argument used in 2.7 and the second terms goes to zero by Theorem 2.5. We conclude that under the given assumptions, the control problem is robust under weak convergence.   □

3.2. Robustness to incorrect priors for multistage problems. The following result holds in generality.

THEOREM 3.2. *If $c(x, u)$ is a non-negative, measurable, and bounded function in $\mathbb{X} \times \mathbb{U}$, then*

$$|J_\beta(P, Q, \gamma_{P_n}^*) - J_\beta^*(P, Q)| \leq 2\|P_n(\mathrm{d}x_0) - P(\mathrm{d}x_0)\|_{TV} \frac{1}{1 - \beta} \|c\|_\infty.$$

Proof. We use that

$$J_\beta(P, Q, \gamma_{P_n}^*) - J_\beta^*(P, Q) = J_\beta(P, Q, \gamma_{P_n}^*) - J_\beta(P_n, Q, \gamma_{P_n}^*) + J_\beta(P_n, Q, \gamma_{P_n}^*) - J_\beta^*(P, Q).$$

From inequalities (2.4), (2.5), and (2.6) we have that $|J_\beta(P, Q, \gamma_{P_n}^*) - J_\beta(P_n, Q, \gamma_{P_n}^*)|$ is upper bounded as $\|P_n(dx_0) - P(dx_0)\|_{TV} \frac{1}{1-\beta} \|c\|_\infty$. The analysis is then complete by considering Theorem 2.13. □

We now develop a robustness result under weak convergence of priors for multistage case. First, we give a lemma showing that for any multistage setting with a controlled Markov chain satisfying Assumption 2.3, the cost at any time stage is continuous in priors under weak convergence.

THEOREM 3.3. *Under Assumption* 2.3, *as* $P_n \to P$ *weakly, we have,*

$$|J_\beta(P, \mathcal{T}, \gamma_{P_n}^*) - J_\beta^*(P, \mathcal{T})| \to 0.$$

Proof. We use the following bound again:

$$|J_\beta(P, \mathcal{T}, \gamma_{P_n}^*) - J_\beta^*(P, \mathcal{T})| \le |J_\beta(P, \mathcal{T}, \gamma_{P_n}^*) - J_\beta(P_n, \mathcal{T}, \gamma_{P_n}^*)| + |J_\beta(P_n, \mathcal{T}, \gamma_{P_n}^*) - J_\beta^*(P, \mathcal{T})|.$$

Here, $\gamma_{P_n}^*$ is optimal for prior P_n, the existence of $\gamma_{P_n}^*$, and an optimal policy γ for prior P follows from Theorem 2.12. The second term goes to zero by Theorem 2.12. The first term goes to zero by (2.13) which states that

$$\sup_{\gamma \in \Gamma} |J_\beta(P_n, \mathcal{T}, \gamma) - J_\beta(P, \mathcal{T}, \gamma)| \to 0. \qquad \Box$$

4. Implications for empirical learning methods in stochastic control. In engineering practice, when one does not know the probability measure for a random variable one typically attempts to learn it via test inputs or empirical observations. Let $\{(X_i), i \in \mathbb{N}\}$ be an $\mathbb{X}$-valued i.i.d. random variable sequence generated according to some distribution μ.

Defining for every (fixed) Borel $B \subset \mathbb{X}$, and $n \in \mathbb{N}$, the empirical occupation measures

$$\mu_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \in B\}};$$

one has $\mu_n(B) \to \mu(B)$ almost surely (a.s.) by the strong law of large numbers. Also, $\mu_n \to \mu$ weakly with probability one ([14], Thm. 11.4.1).

However, μ_n can not converge to μ in total variation, in general. On the other hand, if we know that μ admits a density, we can find estimators to estimate μ under total variation [12].

As discussed above, the empirical averages converge almost surely. By a similar reasoning, for a given bounded measurable function f, $\int \mu_n(dx)f(x)$ converges to $\int \mu(dx)f(x)$. This then also holds for any *countable* collection of functions, $\{f_1, f_2, f_3, ...\}$. A relevant question is the following: can one ensure uniform convergence (over a family of functions) with arbitrary precision by only guaranteeing convergence for a countable collection of functions? This entails the problem of covering a family of functions with arbitrarily small neighborhoods of finitely many functions under an appropriate distance metric. The answer to this question is studied by the theory of *empirical risk minimization*: in the learning theoretic context when one tries to estimate the source distribution, the convergence of optimal costs under μ_n to the cost optimal for μ is called the *consistency of empirical risk minimization* [46].

In particular, if the following uniform convergence holds,

$$(4.1) \qquad \lim_{n \to \infty} \sup_{f \in \mathcal{F}} \left| \int f(x) \mu_n(dx) - \int f(x) \mu(dx) \right| = 0,$$

for a class of measurable functions $\mathcal{F}$, then $\mathcal{F}$ is called a μ-*Glivenko–Cantelli class* [15]. If the class $\mathcal{F}$ is μ-Glivenko–Cantelli for every μ, it is called a *universal Glivenko–Cantelli* class. One example of a universal Glivenko–Cantelli family of real functions on $\mathbb{R}^N$ is the family $\{f : \|f\|_{BL} \le M\}$ for some $0 < M < \infty$, where $\|f\|_{BL} = \|f\|_\infty + \sup_{x_1 \ne x_2} \frac{|f(x_1)-f(x_2)|}{|x_1-x_2|}$ ([15]). For related characterizations and further examples, see [38, 45, 13].

In another direction, if (4.1) holds for any sequence of measures $\{\mu_n\}$ converging weakly to μ (rather than only empirical models), then $\mathcal{F}$ is called a μ-*uniform class*. For a subset, B, of $\mathbb{X}$, the oscillation of $\mathcal{F}$ on B is defined as

$$w_{\mathcal{F}}(B) = \sup\{|f(x) - f(y)|, f \in \mathcal{F}, x, y \in B\};$$

in the case where $\mathcal{F}$ consists of a single function f we use the notation $w_f(B)$ or $w_f B$. Then a characterization for uniformity classes is given by the following [6]: a necessary and sufficient condition for $\mathcal{F}$ to be a P-uniformity class is that

$$w_{\mathcal{F}}(\mathbb{X}) < \infty$$

and

$$\lim_{\delta \to 0} \sup_{f \in \mathcal{F}} P\{x : w_f S(x, \delta) > \epsilon\} = 0$$

for any $\epsilon > 0$, where $S(x, \delta)$ is the ball around x with radius δ. It can be seen that one example of a uniformity class on $\mathbb{R}^N$ is again the family $\{f : \|f\|_{BL} \le M\}$ for some $0 < M < \infty$. For a detailed discussion and characterization of these classes see [6].

4.1. Application to robustness to incorrect priors.

THEOREM 4.1. *Suppose that the prior model of the system is estimated with the i.i.d. measurements such that for every (fixed) Borel $B \subset \mathbb{X}$, and $n \in \mathbb{N}$, the empirical occupation measures*

$$(4.2) \qquad P_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \in B\}}.$$

Then under one of the following conditions,

(i) *if the stagewise cost function $c(x, u)$ is non-negative, bounded, and continuous on $\mathbb{X} \times \mathbb{U}$, the action space $\mathbb{U}$ is compact and the measurement channel $Q(\cdot|x)$ is continuous in total variation;*

(ii) *if we restrict the class of policies to $\mathcal{G} = \{\gamma : \int Q(dy|x)c(x, \gamma(y)) \in \mathcal{E}\}$, where $\mathcal{E}$ is a class of P-Glivenko–Cantelli family of functions;*

(iii) *if we restrict the class of policies to $\mathcal{G} = \{\gamma : c(x, \gamma(y)) \in \mathcal{E}\}$, where $\mathcal{E}$ is a class of PQ-uniformity family of functions and if the measurement channel $Q(\cdot|x)$ is weakly continuous in x,*

as $n \to \infty$, with probability 1, we have, $J^*(P_n, Q) \to J^*(P, Q)$ and $J(P, Q, \gamma_n^*) \to J^*(P, Q)$, where γ_n^* is the optimal policy designed for the estimated model P_n. That is the optimal cost function and the optimal policies are consistent under empirical estimation.

Proof.
 (i) The result follows from Theorem 2.5 and Theorem 3.1.
(ii) We write

$$\lim_{n \to \infty} \sup_{\gamma \in \Gamma} \Big| \int P_n(dx) \int Q(dy|x)c(x,\gamma(y)) - \int P(dx) \int Q(dy|x)c(x,\gamma(y)) \Big|.$$

The results follows from the discussion made for (4.1).
(iii) We make the following argument. Similar to the above discussion, if optimal policies are assumed to be from the restricted class of policies $\mathcal{G}$, a sufficient condition for the convergence of optimal costs is the following form of uniform weak convergence:

$$\lim_{n \to \infty} \sup_{\gamma \in \mathcal{G}} \Big| \int_{\mathbb{X} \times \mathbb{Y}} c(x,\gamma(y)) P_n Q(dx,dy)$$
$$- \int_{\mathbb{X} \times \mathbb{Y}} c(x,\gamma(y)) PQ(dx,dy) \Big| = 0.$$

The above argument holds when $\mathcal{G} = \{\gamma : c(x,\gamma(y)) \in \mathcal{E}\}$, where $\mathcal{E}$ is a class of PQ-uniformity family of functions and when the joint measure $P_n Q$ converges weakly to PQ. We have that the empirical measures P_n converges weakly to P with probability 1. Thus, the result follows from Lemma 1.1(i). □

Conditions (ii) and (iii) might be difficult to check in general. Therefore, we analyze these conditions from a practical point of view. Condition (ii) requires $\int Q(dy|x)c(x,\gamma(y))$ to be a P-Glivenko–Cantelli class of functions. There are various characterizations of such class of functions (see [46]). One example of Glivenko–Cantelli class of functions is bounded Lipschitz functions. Thus, if $\int Q(dy|x)c(x,\gamma(y))$ is bounded and Lipschitz in x, the condition (ii) is satisfied. This requirement is met under the following set of restrictions:

- $\|Q(\cdot|x) - Q(\cdot|x')\|_{TV} \le \alpha|x - x'|$ for some $\alpha < \infty$ and $|c(x,u) - c(x',u)| \le \beta|x - x'|$ for all $u \in \mathbb{U}$ for some $\beta < \infty$.

The above condition is stronger than condition (i).

Condition (iii) requires $c(x,\gamma(y))$ to be a PQ-uniformity class of functions and $Q(\cdot|x)$ to be weakly continuous. One example of uniformity class of functions is again bounded Lipchitz functions. Thus, if $c(x,\gamma(y))$ is bounded and Lipschitz in x and y and if Q is weakly continuous, the condition (iii) is satisfied. The following is sufficient:

- $|c(x,u) - c(x',u')| \le \alpha|x - x'| + \beta|u - u'|$ for some $\alpha < \infty$ and $\beta < \infty$.
- Γ is restricted to be the space of Lipschitz functions such that for any $\gamma \in \Gamma$ we have $|\gamma(y) - \gamma(y')| \le \eta|y - y'|$.
- $Q(\cdot|x)$ is weakly continuous.

This set of assumptions weakens the restrictions on the channel. However, putting assumptions on the space of policies is artificial as there is usually no guarantee that the optimal policy is Lipschitz.

We next consider the multistage case.

THEOREM 4.2. *Suppose that the prior model of the system is estimated with the i.i.d. measurements as in* (4.2). *Under Assumption* 2.3, *as $n \to \infty$ we have a.s., $J_\beta^*(P_n, \mathcal{T}) \to J_\beta^*(P, \mathcal{T})$ and $J_\beta(P, \mathcal{T}, \gamma_n^*) \to J_\beta^*(P, \mathcal{T})$, where γ_n^* is the optimal policy designed for the estimated model P_n. That is, the optimal cost function and the optimal policies are consistent under empirical estimation for the multistage problem.*

Proof. Given that $P_n \to P$ weakly (a.s.), the result follows from Theorem 2.12 and Theorem 3.3. □

5. Conclusion. We studied the topological properties of single and multistage optimization problems in stochastic control on the space of initial probability measures and applications of these to robustness of the control policies applied to systems with incomplete models. We made the observation that while weak convergence is in general too *weak* for continuity and robustness, channel and transition kernel regularities often allow for continuity and robustness under weak convergence. This is a practically very important result since often in engineering applications, system models are learned through training data which only guarantees weak convergence to the true model in general.

6. Acknowledgments. The authors are grateful to Graeme Baker who as an NSERC USRA summer student worked on the initial stages for some of the results reported here and the numerical analysis utilized in the proof of Theorem 2.7 presented in section B.1. The authors are also grateful to Prof. Tamás Linder for the joint work in [52], and Prof. Naci Saldi for many technical discussions with regard to Theorem 2.10.

Appendix A. Technical Proofs.

Here, we include some additional, mainly numerical, derivations utilized in the paper. The numerical derivations may or may not be included in the final version of the paper.

A.1. Proof of Lemma 2.11.

Proof. Our goal is to show that for a given ϵ the term can be bounded by ϵ for $n > N$ for a sufficiently large N. For the ease of notation we will first study the case where $k = 2$, then we will look at the general case.

In the following, to economize the notation, we write $\gamma(Y_{[0,t]})$ to denote $\{\gamma_k(Y_0,\ldots,Y_k;U_0,\ldots,U_{k-1}), k \leq t\}$.

$$\sup_{\gamma \in \Gamma} \left| E_P\left[c(X_2,\gamma(Y_{[0,2]}))\right] - E_{P_n}\left[c(X_2,\gamma(Y_{[0,2]}))\right] \right|$$

$$= \sup_{\gamma \in \Gamma} \left| \int P(dx_0)Q(dy_0|x_0)\mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1) \times \left(E[c(X_2,\gamma(Y_{[0,2]}))|x_0,x_1,y_0,y_1] \right) \right.$$

$$\left. - \int P_n(dx_0)Q(dy_0|x_0))\mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1) \times \left(E[c(X_2,\gamma(Y_{[0,2]}))|x_0,x_1,y_0,y_1] \right) \right|$$

$$= \sup_{\gamma \in \Gamma} \left| \int P(dx_0)Q(dy_0|x_0)\mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1) \times \left(E[c(X_2,\gamma(Y_{[0,2]}))|x_1,y_0,y_1] \right) \right.$$

$$\left. - \int P_n(dx_0)Q(dy_0|x_0))\mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1) \times \left(E[c(X_2,\gamma(Y_{[0,2]}))|x_1,y_0,y_1] \right) \right|.$$

In the last equality, we used the fact that conditioned on all observations and most recent state variable, we can take out the conditioning on the earlier state variables using the Markov properties of the system. This follows from

$$P^\gamma(dx_2,dy_0,dy_1,dy_2|x_0,x_1,y_0,y_1)$$
$$= P(dy_2|x_2)P^\gamma(dx_2,dy_0,dy_1|x_0,x_1,y_0,y_1)$$
$$= P(dy_2|x_2)P(dx_2|x_1,\gamma(y_0,y_1))P(dy_0,dy_1|y_0,y_1)$$
$$= P(dy_2|x_2)P^\gamma(dx_2|x_1,y_0,y_1)P(dy_0,dy_1|y_0,y_1)$$
$$= P^\gamma(dx_2,dy_0,dy_1,dy_2|x_1,y_0,y_1),$$

where P^γ denote the induced probability measure on the state and the measurement variables given a policy $\gamma \in \Gamma$.

Now, if we can show that the term $\int Q(dy_0|x_0)\mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1) \times (E[c(X_2,\gamma(Y_{[0,2]}))|x_1,y_0,y_1])$ is a continuous and bounded function of x_0 uniformly for all $\gamma \in \Gamma$, we can make the difference less than ϵ since $P_n \to P$ weakly. To show the continuity observe the following:

$$\lim_{x_0' \to x_0} \sup_{\gamma \in \Gamma} \left| \int Q(dy_0|x_0)\mathcal{T}(dx_1|x_0,\gamma(y_0))Q(dy_1|x_1)E[c(X_2,\gamma(Y_{[0,2]}))|y_0,x_1,y_1] \right.$$

$$\left. - \int Q(dy_0|x_0')\mathcal{T}(dx_1|x_0',\gamma(y_0))Q(dy_1|x_1)E[c(X_2,\gamma(Y_{[0,2]}))|y_0,x_1,y_1] \right|$$

$$\leq \lim_{x_0' \to x_0} \sup_{\gamma \in \Gamma} \left| \int Q(dy_0|x_0) \int \mathcal{T}(dx_1|x_0',\gamma(y_0))Q(dy_1|x_1) \times E[c(X_2,\gamma(Y_{[0,2]}))|x_1,y_0,y_1] \right.$$

$$\left. - \int Q(dy_0|x_0') \int \mathcal{T}(dx_1|x_0',\gamma(y_0))Q(dy_1|x_1) \times E[c(X_2,\gamma(Y_{[0,2]}))|x_1,y_0,y_1] \right|$$

$$+ \lim_{x_0' \to x_0} \sup_{\gamma \in \Gamma} \int Q(dy_0|x_0) \left| \int \mathcal{T}(dx_1|x_0,\gamma(y_0)) \int Q(dy_1|x_1) \times E[c(X_2,\gamma(Y_{[0,2]}))|x_1,y_0,y_1] \right.$$

$$\left. - \int \mathcal{T}(dx_1|x_0',\gamma(y_0)) \int Q(dy_1|x_1) \times E[c(X_2,\gamma(Y_{[0,2]}))|x_1,y_0,y_1] \right|.$$

(A.1)

In (A.1), the first term goes to 0 by Assumption 2.3(b) with the following argument:

$$\lim_{x_0' \to x_0} \sup_{\gamma \in \Gamma} \left| \int Q(dy_0|x_0)\mathcal{T}(dx_1|x_0',\gamma(y_0))Q(dy_1|x_1)E[c(X_2,\gamma(Y_{[0,2]}))|y_0,x_1,y_1] \right.$$

$$\left. - \int Q(dy_0|x_0')\mathcal{T}(dx_1|x_0',\gamma(y_0))Q(dy_1|x_1)E[c(X_2,\gamma(Y_{[0,2]}))|y_0,x_1,y_1] \right|$$

$$\leq \lim_{x_0' \to x_0} \|c\|_\infty \|Q(y_0 \in \cdot|x_0) - Q(y_0 \in \cdot|x_0')\|_{TV} = 0.$$

Before analyzing the second term, we claim the following: for a family of functions, $\{f(\gamma,x_k)\}$ which is uniformly bounded, and equicontinuous over $\gamma \in \Gamma$,

$$\lim_{x_{k-1}' \to x_{k-1}} \sup_{\gamma \in \Gamma} \left| \int \mathcal{T}(dx_k|x_{k-1},\gamma(y_{[0,k-1]}))f(\gamma,x_k) \right.$$

(A.2)
$$\left. - \int \mathcal{T}(dx_k|x_{k-1}',\gamma(y_{[0,k-1]}))f(\gamma,x_k) \right| = 0.$$

To see this, observe the following: by the Arzela–Ascoli theorem, for any given compact set K, and $\eta > 0$ there is a finite family of continuous functions $\mathbb{F} := \{f_1, \ldots, f_N\}$ such that for any γ, we can find an $f_i \in \mathbb{F}$ with $\sup_{x_k \in K} |f(\gamma,x_k) - f_i(x_k)| \leq \eta$. Furthermore, since $\mathcal{T}(dx_k|x_{k-1},u_{k-1})$ is continuous, for any $r > 0$, the set $\{\mathcal{T}(dx_k|x_{k-1}',u_{k-1}), |x_{k-1}' - x_{k-1}| \leq r, u_{k-1} \in \mathbb{U}\}$ as a continuous image of a compact set is itself weakly compact, and thus tight, and hence, for every $\epsilon > 0$, there exists a compact set K_ϵ such that $\int_{K_\epsilon} \mathcal{T}(dx_k|x_{k-1}',\gamma(y_{[0,k-1]})) \geq 1-\epsilon$ for all $|x_{k-1}' - x_{k-1}| \leq r$ and $\gamma \in \Gamma$.

Now, we fix an $\epsilon > 0$, choose an $r > 0$ and a compact set K_ϵ according to above discussion and fix a finite family of continuous functions $\mathbb{F} := \{f_1, \ldots, f_N\}$ such that for any γ, we can find an $f_i \in \mathbb{F}$ with $\sup_{x_k \in K} |f(\gamma,x_k) - f_i(x_k)| \leq \epsilon$. We also choose a $0 < \delta \leq r$ such that $\sup_{\gamma \in \Gamma} | \int_{K_\epsilon} \mathcal{T}(dx_k|x_{k-1},\gamma(y_{[0,k-1]}))f_i(x_k) -$

$\int_{K_\epsilon} \mathcal{T}(dx_k|x'_{k-1}, \gamma(y_{[0,k-1]}))f_i(x_k)| \leq \epsilon$ for $|x'_{k-1} - x_{k-1}| \leq \delta$ and for all $f_i \in \mathbb{F}$, which we can do since $\mathcal{T}(\cdot|x, u)$ is weakly continuous in x uniformly for $u \in \mathbb{U}$ and there are finitely many f_i. With this setup, let us look at (A.2) again for $\{x'_{k-1} : |x'_{k-1} - x_{k-1}| \leq \delta\}$:

$$\sup_{\gamma \in \Gamma} \left| \int \mathcal{T}(dx_k|x_{k-1}, \gamma(y_{[0,k-1]})f(\gamma, x_k) - \int \mathcal{T}(dx_k|x'_{k-1}, \gamma(y_{[0,k-1]}))f(\gamma, x_k) \right|$$

$$\leq \sup_{\gamma \in \Gamma} \left| \int_{\mathbb{X} \setminus K_\epsilon} \mathcal{T}(dx_k|x_{k-1}, \gamma(y_{[0,k-1]}))f(\gamma, x_k) - \int_{\mathbb{X} \setminus K_\epsilon} \mathcal{T}(dx_k|x'_{k-1}, \gamma(y_{[0,k-1]}))f(\gamma, x_k) \right|$$

$$+ \sup_{\gamma \in \Gamma} \left| \int_{K_\epsilon} \mathcal{T}(dx_k|x_{k-1}, \gamma(y_{[0,k-1]}))f(\gamma, x_k) - \int_{K_\epsilon} \mathcal{T}(dx_k|x'_{k-1}, \gamma(y_{[0,k-1]})f(\gamma, x_k) \right|$$

$$\leq 2\epsilon\|c\|_\infty + \sup_{\gamma \in \Gamma} | \int_{K_\epsilon} \mathcal{T}(dx_k|x_{k-1}, \gamma(y_{[0,k-1]}))\big(f(\gamma, x_k) - f_i(x_k)\big)$$

$$+ \int_{K_\epsilon} \mathcal{T}(dx_k|x_{k-1}, \gamma(y_{[0,k-1]}))f_i(x_k) - \int_{K_\epsilon} \mathcal{T}(dx_k|x'_{k-1}, \gamma(y_{[0,k-1]}))f_i(x_k)$$

$$+ \int_{K_\epsilon} \mathcal{T}(dx_k|x'_{k-1}, \gamma(y_{[0,k-1]}))\big(f_i(x_k) - f(\gamma, x_k)\big) |$$

$$\leq 2\epsilon\|c\|_\infty + \sup_{\gamma \in \Gamma} \int_{K_\epsilon} \mathcal{T}(dx_k|x_{k-1}, \gamma(y_{[0,k-1]}))f_i(x_k)$$

$$- \int_{K_\epsilon} \mathcal{T}(dx_k|x'_{k-1}, \gamma(y_{[0,k-1]}))f_i(x_k) | + 2\epsilon$$

$$\leq 2\epsilon\|c\|_\infty + 3\epsilon,$$

where $f_i(x_k)$ is chosen according to the discussion above such that f_i is ϵ close to $f(\gamma, x_k)$. As ϵ is arbitrary, (A.2) holds true.

Now we return to (A.1) again. To show that the second term also goes to 0, we use (A.2) with the claim that

$$f(\gamma, x_1) := \int Q(dy_1|x_1) \times E[c(X_2, \gamma(Y_{[0,2]}))|x_1, y_0, y_1]$$

is a continuous and bounded function of x_1 uniformly for all $\gamma \in \Gamma$. To see this, we write

$$\lim_{x'_1 \to x_1} \sup_{\gamma \in \Gamma} \left(\int Q(dy_1|x_1) \times E[c(X_2, \gamma(Y_{[0,2]}))|x_1, y_0, y_1] \right.$$

$$\left. - \int Q(dy_1|x'_1) \times E[c(X_2, \gamma(Y_{[0,2]}))|x'_1, y_0, y_1] \right)$$

$$\leq \lim_{x'_1 \to x_1} \sup_{\gamma \in \Gamma} \left| \int Q(dy_1|x_1) \times E[c(X_2, \gamma(Y_{[0,2]}))|x_1, y_0, y_1] \right.$$

$$\left. - \int Q(dy_1|x'_1) \times E[c(X_2, \gamma(Y_{[0,2]}))|x_1, y_0, y_1] \right|$$

$$+ \lim_{x'_1 \to x_1} \sup_{\gamma \in \Gamma} \int Q(dy_1|x_1) \times \left| E[c(X_2, \gamma(Y_{[0,2]}))|x_1, y_0, y_1] - E[c(X_2, \gamma(Y_{[0,2]}))|x'_1, y_0, y_1] \right|$$

$$= 0.$$

Above, for the first term we used Assumption 2.3(b), and for the second term we have the following:

$$\lim_{x_1' \to x_1} \sup_{\gamma \in \Gamma} \left| E[c(X_2, \gamma(Y_{[0,2]}))|x_1, y_0, y_1] - E[c(X_2, \gamma(Y_{[0,2]}))|x_1', y_0, y_1] \right|$$

$$= \lim_{x_1' \to x_1} \sup_{\gamma \in \Gamma} \left| \int \mathcal{T}(dx_2|x_1, \gamma(y_{[0,1]})) Q(dy_2|x_2) c(x_2, \gamma(y_{[0,2]})) \right.$$

(A.3)
$$\left. - \int \mathcal{T}(dx_2|x_1', \gamma(y_{[0,1]})) Q(dy_2|x_2) c(x_2, \gamma(y_{[0,2]})) \right|.$$

For the last term, we used (A.2) with $f(\gamma, x_2) = \int Q(dy_2|x_2) c(x_2, \gamma(y_{[0,2]}))$. To see that $f(\gamma, x_2)$ is a uniformly bounded, equicontinuous family of functions over $\gamma \in \Gamma$, observe the following:

$$\lim_{x_2' \to x_2} \sup_{\gamma \in \Gamma} \left| \int Q(dy_2|x_2') c(x_2', \gamma(y_{[0,2]})) - \int Q(dy_2|x_2) c(x_2, \gamma(y_{[0,2]})) \right|$$

$$\leq \lim_{x_2' \to x_2} \sup_{\gamma \in \Gamma} \left| \int Q(dy_2|x_2') c(x_2', \gamma(y_{[0,2]})) - \int Q(dy_2|x_2) c(x_2', \gamma(y_{[0,2]})) \right|$$

$$+ \lim_{x_2' \to x_2} \sup_{\gamma \in \Gamma} \left| \int Q(dy_2|x_2) \big(c(x_2', \gamma(y_{[0,2]})) - c(x_2, \gamma(y_{[0,2]})) \big) \right|$$

$$= 0.$$

For the first term we used that the channel Q is continuous in total variation, and for the second term we used that $c(x_2, \gamma(y_{[0,2]}))$ is continuous in x_2 uniformly for all $\gamma \in \Gamma$, because for every γ, $\gamma(y_{[0,2]})$ corresponds to a control action $u \in \mathbb{U}$ and $\mathbb{U}$ is compact.

Thus, we can find an N such that for all $n > N$,

$$\sup_{\gamma \in \Gamma} \left| E_P\big[c(X_2, \gamma(Y_{[0,2]}))\big] - E_{P_n}\big[c(X_2, \gamma(Y_{[0,2]}))\big] \right| < \epsilon.$$

Now, we generalize the argument for any time stage $k \geq 0$,

$$\sup_{\gamma \in \Gamma} \left| E_P\big[c(X_k, \gamma(Y_{[0,k]}))\big] - E_{P_n}\big[c(X_k, \gamma(Y_{[0,k]}))\big] \right|$$

$$= \sup_{\gamma \in \Gamma} \left| \int P(dx_0) Q(dy_0|x_0) \mathcal{T}(dx_1|x_0, \gamma(y_0)) Q(dy_1|x_1) \ldots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) \right.$$

$$Q(dy_{k-1}|x_{k-1}) \times \left(E[c(X_k, \gamma(Y_{[0,k]}))|x_0, x_1, \ldots x_{k-1}, y_0, \ldots, y_{k-1}] \right)$$

$$- \int P_n(dx_0) Q(dy_0|x_0)) \mathcal{T}(dx_1|x_0, \gamma(y_0)) \ldots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) Q(dy_{k-1}|x_{k-1})$$

$$\left. \times \left(E[c(X_k, \gamma(Y_{[0,k]}))|x_0, x_1, \ldots x_{k-1}, y_0, \ldots, y_{k-1}] \right) \right|$$

$$= \sup_{\gamma \in \Gamma} \left| \int P(dx_0) Q(dy_0|x_0) \mathcal{T}(dx_1|x_0, \gamma(y_0)) Q(dy_1|x_1) \ldots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) \right.$$

$$Q(dy_{k-1}|x_{k-1}) \times \left(E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \ldots, y_{k-1}] \right)$$

$$- \int P_n(dx_0) Q(dy_0|x_0)) \mathcal{T}(dx_1|x_0, \gamma(y_0)) \ldots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) Q(dy_{k-1}|x_{k-1})$$

$$\left. \times \left(E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \ldots, y_{k-1}] \right) \right|.$$

The last equality follows from the Markov property of the system as we do not depend on the past state variables given all the past observations and most recent state variable. We again need to show the continuity in x_0.

$$\lim_{x_0' \to x_0} \sup_{\gamma \in \Gamma} \left[\int Q(dy_0|x_0) \dots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0 \dots y_{k-1}] \right.$$

$$\left. - \int Q(dy_0|x_0') \dots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]}) E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \dots, y_{k-1}] \right]$$

$$\leq \lim_{x_0' \to x_0} \sup_{\gamma \in \Gamma} \left| \int Q(dy_0|x_0) \int \mathcal{T}(dx_1|x_0', \gamma(y_0)) \dots T(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) Q(dy_{k-1}|x_{k-1}) \right.$$

$$\times E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \dots, y_{k-1}]$$

$$- \int Q(dy_0|x_0') \int \mathcal{T}(dx_1|x_0', \gamma(y_0)) \dots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) Q(dy_{k-1}|x_{k-1})$$

$$\left. \times E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \dots, y_{k-1}] \right|$$

$$+ \lim_{x_0' \to x_0} \sup_{\gamma \in \Gamma} \int Q(dy_0|x_0) \left| \int \mathcal{T}(dx_1|x_0, \gamma(y_0)) \dots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) \right.$$

$$Q(dy_{k-1}|x_{k-1}) \times E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \dots, y_{k-1}]$$

$$- \int \mathcal{T}(dx_1|x_0', \gamma(y_0)) \dots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]}))$$

$$\left. Q(dy_{k-1}|x_{k-1}) \times E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \dots, y_{k-1}] \right|$$

$$= 0.$$

(A.4)

Similar to the special case where $k = 2$, at (A.4), for the first term we used Assumption 2.3(b).

For the second term, we used the fact that the term

$$\int Q(dy_1|x_1) \dots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) Q(dy_{k-1}|x_{k-1})$$

$$E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \dots, y_{k-1}]$$

is a continuous and bounded function of x_1 uniformly for all $\gamma \in \Gamma$. For this, it is necessary to show that

$$\int Q(dy_2|x_2) \dots \mathcal{T}(dx_{k-1}|x_{k-2}, \gamma(y_{[0,k-2]})) Q(dy_{k-1}|x_{k-1})$$

$$E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \dots, y_{k-1}]$$

is a continuous function of x_2 uniformly over Γ. Then, inductively, the last step is to show that

$$E[c(X_k, \gamma(Y_{[0,k]}))|x_{k-1}, y_0, \dots, y_{k-1}]$$

$$= \int c(x_k, \gamma(y_{[0,k]})) \mathcal{T}(dx_k|x_{k-1}, \gamma(y_{[0,k-1]})) Q(dy_k|x_k)$$

is a continuous function of x_{k-1} uniformly over Γ, which holds since $\mathcal{T}(dx_k|x_{k-1}, \gamma(y_{[0,k-1]}))$ is weakly continuous in x_{k-1} and the continuity is uniform on Γ since for every $\gamma \in \Gamma$, $\gamma(y_{[0,k-1]})$ corresponds to a control action from $\mathbb{U}$ and $\mathbb{U}$ is compact. The fact that $Q(dy_k|x_k)$ is continuous in total variation in x_k and $c(x_k, \gamma(y_{[0,k]}))$ is continuous in x_k uniformly over Γ completes the proof. $\square$

A.2. Proof of Lemma 2.3. The proof is adapted from the proof of Theorem 3.2 in [52].

Proof. Let $\mu_{\mathbb{Y}}$ satisfy $\mu_{\mathbb{Y}}(A) = \mu(\mathbb{X} \times A)$ for $A \in \mathcal{B}(\mathbb{Y})$. By Lusin's theorem ([40, Thm. 2.24]), for all $\gamma \in \Gamma$ and $\varepsilon > 0$, there is a continuous function $f : \mathbb{U} \to \mathbb{Y}$ such that

$$\mu_{\mathbb{Y}}\{y : f(y) \neq \gamma(y)\} < \varepsilon.$$

For convenience of notation we let $B = \{y : f(y) \neq \gamma(y)\}$. Observe the following:

$$\int |c(x, \gamma(y)) - c(x, f(y))| \, \mu(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{X} \times B} |c(x, \gamma(y)) - c(x, f(y))| \, \mu(\mathrm{d}x, \mathrm{d}y)$$
$$< \varepsilon \cdot \|c\|_\infty,$$

where $\|c\|_\infty$ denotes the supremum norm of c, which is finite by assumption. We have that

$$(A.5) \qquad \int c(x, f(y)) \mu(\mathrm{d}x, \mathrm{d}y) < \int c(x, g(y)) \mu(\mathrm{d}x, \mathrm{d}y) + \varepsilon \cdot \|c\|_\infty.$$

Define

$$j(\mu, \mathcal{C}) = \inf_{\gamma \in \mathcal{C}} \int c(x, \gamma(y)) \mu(\mathrm{d}x, \mathrm{d}y), \text{ and } j(\mu, \Gamma) = \inf_{\gamma \in \Gamma} \int c(x, \gamma(y)) \mu(\mathrm{d}x, \mathrm{d}y).$$

Note that $j(\mu, \mathcal{C}) \geq j(\mu, \Gamma)$ as $\mathcal{C} \subset \Gamma$. From (A.5) we have $j(\mu, \mathcal{C}) \leq j(\mu, \Gamma) + \varepsilon \cdot \|c\|_\infty$, which gives $j(\mu, \mathcal{C}) \leq j(\mu, \Gamma)$ as ε was arbitrary. Hence $j(\mu, \mathcal{C}) = j(\mu, \Gamma)$. □

Appendix B. Additional proofs.

B.1. Calculations for proof of Theorem 2.7. In the following, p will denote the density of the corresponding (possibly conditional) probability measure.

B.1.1. Computing γ_P^*.

$$\gamma_P^* = E[X|Y = y] = \int_0^1 x p(x|y) \mathrm{d}x = \int_0^1 x \frac{p(y|x)}{\int_0^1 p(y|x) p(x) \mathrm{d}x} \mathrm{d}x$$

by Bayes' rule. The denominator reads as

$$\int_0^1 p(y|x) p(x) \mathrm{d}x = \int_0^1 \left(\frac{1}{2} + \frac{1}{2} \delta_x(y) \right) \mathrm{d}x = 1.$$

This gives

$$\gamma_P^* = \int_0^1 x p(y|x) \mathrm{d}x = \int_0^1 x \left(\frac{1}{2} + \frac{1}{2} \delta_x(y) \right) \mathrm{d}x = \int_0^1 \frac{1}{2} x \mathrm{d}x + \frac{1}{2} y$$
$$= \frac{1}{4} + \frac{1}{2} y = \frac{1}{2} \left(\frac{1}{2} + y \right)$$

as desired.

B.1.2. Computing $J^*(P, Q)$.

$$J^*(P, Q) = E[c(x, \gamma_P^*(y))] = E[(x - \gamma_P^*(y))^2]$$
$$= \int_0^1 \int_0^1 \left(x - \frac{1}{4} - \frac{1}{2}y \right)^2 Q(\mathrm{d}y|x) P(\mathrm{d}x).$$

We now compute the inner integral for some fixed x:

$$\int_0^1 \left(x - \frac{1}{4} - \frac{1}{2}y \right)^2 Q(\mathrm{d}y|x)$$
$$= \frac{1}{2} \int_0^1 \left(x - \frac{1}{4} - \frac{1}{2}y \right)^2 \mathrm{d}y + \frac{1}{2} \int_0^1 \left(x - \frac{1}{4} - \frac{1}{2}y \right)^2 \delta_x(y) \mathrm{d}y$$
$$= \frac{1}{2} \left(\frac{13}{48} - x + x^2 \right) + \frac{1}{2} \left(\frac{1}{16} - \frac{x}{4} + \frac{x^2}{4} \right) = \frac{1}{6} - \frac{5x}{8} + \frac{5x^2}{8}.$$

This gives

$$J^*(P, Q) = \int_0^1 \left(\frac{1}{6} - \frac{5x}{8} + \frac{5x^2}{8} \right) \mathrm{d}x = \frac{1}{16}.$$

B.1.3. Computing $\gamma_{P_n}^*$. Let $n \in \mathbb{N}$. If $y \in R$, then we have

$$p(y) = \int_0^1 p(y|x) f_n(x) \mathrm{d}x = \int_L \frac{1}{2} (2) \mathrm{d}x = \int_L \mathrm{d}x = \lambda(L) = \frac{1}{2},$$

where λ is the Lebesgue measure. We now compute $E[X|Y = y]$,

$$E[X|Y = y] = \int_0^1 x p(x|y) \mathrm{d}x = \int_0^1 x \frac{p(y|x) f_n(x)}{\int_0^1 p(y|x) P(x) \mathrm{d}x} \mathrm{d}x = \int_L x \frac{1/2 \cdot 2}{1/2} \mathrm{d}x$$
$$= \sum_{k=1}^n x^2 \big|_{\frac{2k-2}{2n}}^{\frac{2k-1}{2n}} = \sum_{k=1}^n \left(\left(\frac{2k-1}{2n} \right)^2 - \left(\frac{2k-2}{2n} \right)^2 \right)$$
(B.1) $$= \sum_{k=1}^n \frac{4k-3}{4n^2} = \frac{1}{n^2} \sum_{k=1}^n k - \frac{1}{n^2} \sum_{k=1}^n \frac{3}{4} = \frac{1}{n^2} \cdot \frac{n(n+1)}{2} - \frac{1}{n^2} \cdot \frac{3n}{4} = \frac{1}{2} - \frac{1}{4n},$$

where we have used Bayes' rule to establish the second equality. Now we consider the case where $y \in L$. We calculate $p(y)$:

$$p(y) = \int_0^1 p(y|x) f_n(x) \mathrm{d}x = \int_L \frac{1}{2} (2) \mathrm{d}x + \frac{1}{2} \cdot 2 = \int_L \mathrm{d}x + 1 = \lambda(L) + 1 = \frac{3}{2}.$$

This gives

$$E[X|Y = y] = \int_0^1 x p(x|y) \mathrm{d}x = \int_0^1 x \frac{p(y|x) f_n(x)}{\int_0^1 p(y|x) p(x) \mathrm{d}x} \mathrm{d}x$$
$$= \int_L x \frac{1/2 \cdot 2}{3/2} \mathrm{d}x + \int_L x \frac{1/2 \cdot 2}{3/2} \delta_x(y) \mathrm{d}x = \frac{1}{3} \cdot \left(\frac{1}{2} - \frac{1}{4n} \right) + \frac{2}{3} y,$$

where we use (B.1) to give the last equality. We see that the optimal policy is

$$\gamma_{P_n}^*(y) = E[X|Y = y] = \begin{cases} \frac{1}{2} - \frac{1}{4n} & \text{if } y \in R, \\ \frac{1}{3} \cdot \left(\frac{1}{2} - \frac{1}{4n} \right) + \frac{2}{3} y & \text{if } y \in L. \end{cases}$$

B.1.4. Computing $J^*(P_n, Q)$. Let $n \in \mathbb{N}$.

$$J^*(P_n, Q) = E[c(x, \gamma_{P_n}^*(y))] = E[(x - \gamma_{P_n}^*(y))^2]$$

$$= \int_0^1 \left(\int_0^1 (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x) \right) P_n(\mathrm{d}x)$$

$$= 2 \int_L \left(\underbrace{\int_L (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x)}_{(i)} + \underbrace{\int_R (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x)}_{(ii)} \right) \mathrm{d}x.$$

We now compute term (i) for $x \in L$:

$$\int_L (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x) = \int_L \left(x - \frac{1}{3}\left(\frac{1}{2} - \frac{1}{4n} \right) - \frac{2}{3}y \right)^2 Q(\mathrm{d}y|x)$$

$$= \frac{1}{2}\left(x - \frac{1}{3}\left(\frac{1}{2} - \frac{1}{4n} \right) - \frac{2}{3}x \right)^2 + \cdots$$

$$+ \int_L \frac{1}{2}\left(x - \frac{1}{3}\left(\frac{1}{2} - \frac{1}{4n} \right) - \frac{2}{3}y \right)^2 \mathrm{d}y$$

$$= \frac{1}{72} + \frac{1}{288n^2} - \frac{1}{72n} + \left(-\frac{1}{18} + \frac{1}{36n} \right)x + \frac{x^2}{18} + \cdots$$

$$+ \int_L \left(\frac{1}{72} + \frac{1}{288n^2} - \frac{1}{72n} + \left(-\frac{1}{6} + \frac{1}{12n} \right)x + \cdots \right.$$

$$\left. + \frac{x^2}{2} + \left(\frac{1}{9} - \frac{1}{18n} - \frac{2}{3}x \right)y + \frac{2y^2}{9} \right) \mathrm{d}y$$

$$= \frac{1}{72} + \frac{1}{288n^2} - \frac{1}{72n} + \left(-\frac{1}{18} + \frac{1}{36n} \right)x + \frac{x^2}{18} + \cdots$$

$$+ \left(\frac{1}{72} + \frac{1}{288n^2} - \frac{1}{72n} + \left(-\frac{1}{6} + \frac{1}{12n} \right)x + \frac{x^2}{2} \right) \int_L \mathrm{d}y + \cdots$$

$$+ \left(\frac{1}{9} - \frac{1}{18n} - \frac{2}{3}x \right) \int_L y\mathrm{d}y + \frac{2}{9} \int_L y^2 \mathrm{d}y.$$

For ease we compute the integrals in the above step separately. Some are familiar from section B.1.3.

$$\int_L \mathrm{d}y = \lambda(L) = \frac{1}{2},$$

$$\int_L y\mathrm{d}y = \frac{1}{2} \sum_{k=1}^n y^2 \Big|_{\frac{2k-2}{2n}}^{\frac{2k-1}{2n}} = \frac{1}{2}\left(\frac{1}{2} - \frac{1}{4n} \right) \text{ (by (B.1))},$$

$$\int_L y^2 \mathrm{d}y = \frac{1}{3} \sum_{k=1}^n y^3 \Big|_{\frac{2k-2}{2n}}^{\frac{2k-1}{2n}} = \frac{1}{3}\left(\sum_{k=1}^n \left(\left(\frac{2k-1}{2n} \right)^3 - \left(\frac{2k-2}{2n} \right)^3 \right) \right)$$

$$= \frac{1}{3}\left(\sum_{k=1}^n \frac{7 - 18k + 12k^2}{8n^3} \right) = \frac{1}{3}\left(\frac{7}{8n^3} \sum_{k=1}^n 1 - \frac{9}{4n^3} \sum_{k=1}^n k + \frac{3}{2n^3} \sum_{k=1}^n k^2 \right)$$

$$= \frac{7}{24n^3}(n) - \frac{3}{4n^3}\left(\frac{n^2 + n}{2} \right) + \frac{1}{2n^3}\left(\frac{2n^3 + 3n^2 + n}{6} \right) = \frac{1}{6} - \frac{1}{8n}.$$

These combine with the above to give (after some simplification)

$$\int_L (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x) = \frac{37}{432} + \frac{7}{576n^2} - \frac{11}{144n} + \left(-\frac{11}{36} + \frac{11}{72n}\right)x + \frac{11}{36}x^2.$$

Next we compute term (ii) for $x \in L$:

$$\int_R (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x) = \frac{1}{2}\int_R \left(x - \frac{1}{2} + \frac{1}{4n}\right)^2 \mathrm{d}y = \frac{1}{4}\left(x - \frac{1}{2} + \frac{1}{4n}\right)^2$$
$$= \frac{1}{16} + \frac{1}{64n^2} - \frac{1}{16n} + \left(-\frac{1}{4} + \frac{1}{8n}\right)x + \frac{x^2}{4}.$$

For $x \in L$ we compute (i)+(ii):

$$\text{(i)} + \text{(ii)} = \frac{4}{27} + \frac{1}{36n^2} - \frac{5}{36n} + \left(-\frac{5}{9} + \frac{5}{18n}\right)x + \frac{5}{9}x^2.$$

This gives

$$J(P_n, Q) = \int_0^1 \left(c(x, \gamma^*_{P_n}(y))\right)^2 P_n(\mathrm{d}x)$$
$$= 2\int_L \left(\frac{4}{27} + \frac{1}{36n^2} - \frac{5}{36n} + \left(-\frac{5}{9} + \frac{5}{18n}\right)x + \frac{5}{9}x^2\right)\mathrm{d}x$$
$$= 2\left(\frac{4}{27} + \frac{1}{36n^2} - \frac{5}{36n}\right)\int_L \mathrm{d}x + 2\left(-\frac{5}{9} + \frac{5}{18n}\right)\int_L x\mathrm{d}x + 2\left(\frac{5}{9}\right)\int_L x^2\mathrm{d}x$$
$$= \frac{1}{18} - \frac{1}{24n^2}.$$

B.2. Calculations for proof of Proposition 3.2. Let $n \in \mathbb{N}$. From section B.1.3 we know that

$$\gamma^*_{P_n}(y) = E[X|Y=y] = \begin{cases} \frac{1}{2} - \frac{1}{4n} & \text{if } y \in R, \\ \frac{1}{3}\cdot\left(\frac{1}{2} - \frac{1}{4n}\right) + \frac{2}{3}y & \text{if } y \in L. \end{cases}$$

With $P \sim U([0,1])$ we have

$$J(P, Q, \gamma^*_{P_n}) = E[c(x, \gamma^*_{P_n}(y))] = E[(x - \gamma^*_{P_n}(y))^2]$$
$$= \int_0^1 \left(\int_0^1 (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x)\right) P(\mathrm{d}x)$$
$$= \int_L \left(\int_0^1 (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x)\right)\mathrm{d}x + \int_R \left(\int_0^1 (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x)\right)\mathrm{d}x$$
$$= \frac{1}{2}J^*(P_n, Q) + \int_R \left(\int_L (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x) + \int_R (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x)\right)\mathrm{d}x$$
$$= \frac{1}{36} - \frac{1}{48n^2} + \int_R \left(\underbrace{\int_L (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x)}_{\text{(iii)}} + \underbrace{\int_R (x - \gamma^*_{P_n}(y))^2 Q(\mathrm{d}y|x)}_{\text{(iv)}}\right)\mathrm{d}x,$$

where we have used the computation from section B.1.4 for $J^*(P_n, Q)$. For $x \in R$ we compute (iii):

$$\int_L (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x) = \int_L \left(x - \frac{1}{3}\left(\frac{1}{2} - \frac{1}{4n}\right) - \frac{2}{3}y \right)^2 Q(\mathrm{d}y|x)$$

$$= \int_L \frac{1}{2}\left(x - \frac{1}{3}\left(\frac{1}{2} - \frac{1}{4n}\right) - \frac{2}{3}y \right)^2 \mathrm{d}y$$

$$= \left(\frac{1}{72} + \frac{1}{288n^2} - \frac{1}{72n} + \left(-\frac{1}{6} + \frac{1}{12n}\right)x + \frac{x^2}{2} \right)\int_L \mathrm{d}y + \cdots$$

$$+ \left(\frac{1}{9} - \frac{1}{18n} - \frac{2}{3}x \right)\int_L y\mathrm{d}y + \frac{2}{9}\int_L y^2\mathrm{d}y,$$

where the last line follows from the computation of (i) in section B.1.4. After some simplification we arrive at

$$\int_L (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x) = \frac{31}{432} + \frac{5}{576n^2} - \frac{1}{16n} + \left(-\frac{1}{4} + \frac{1}{8n}\right)x + \frac{x^2}{4}.$$

For $x \in R$ we compute term (iv):

$$\int_R (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x) = \int_R \left(x - \frac{1}{2} + \frac{1}{4n} \right)^2 Q(\mathrm{d}y|x)$$

$$= \frac{1}{2}\left(x - \frac{1}{2} + \frac{1}{4n} \right)^2 + \frac{1}{2}\int_R \left(x - \frac{1}{2} + \frac{1}{4n} \right)^2 \mathrm{d}y$$

$$= \frac{3}{4}\left(x - \frac{1}{2} + \frac{1}{4n} \right)^2$$

$$= \frac{3}{16} + \frac{3}{64n^2} - \frac{3}{16n} + \left(-\frac{3}{4} + \frac{3}{8n}\right)x + \frac{3}{4}x^2.$$

For $x \in R$ we have

$$(\mathrm{iii}) + (\mathrm{iv}) = \frac{7}{27} + \frac{1}{18n^2} - \frac{1}{4n} + \left(-1 + \frac{1}{2n}\right)x + x^2.$$

This gives

$$\int_R \left(\int_0^1 (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x) \right)\mathrm{d}x = \int_R \left(\frac{7}{27} + \frac{1}{18n^2} - \frac{1}{4n} + \left(-1 + \frac{1}{2n}\right)x + x^2 \right)\mathrm{d}x$$

$$= \left(\frac{7}{27} + \frac{1}{18n^2} - \frac{1}{4n} \right)\int_R \mathrm{d}x + \left(-1 + \frac{1}{2n}\right)\int_R x\mathrm{d}x + \int_R x^2\mathrm{d}x.$$

We compute the integrals in the previous step separately:

$$\int_R \mathrm{d}x = \lambda(R) = \frac{1}{2},$$

$$\int_R x\mathrm{d}x = \int_0^1 x\mathrm{d}x - \int_L x\mathrm{d}x = \frac{1}{2} - \left(\frac{1}{4} - \frac{1}{8n}\right) = \frac{1}{4} + \frac{1}{8n},$$

$$\int_R x^2\mathrm{d}x = \int_0^1 x^2\mathrm{d}x - \int_L x^2\mathrm{d}x = \frac{1}{3} - \left(\frac{1}{6} - \frac{1}{8n}\right) = \frac{1}{6} + \frac{1}{8n}.$$

With the above, this gives (after some simplification):

$$\int_R \left(\int_0^1 (x - \gamma_{P_n}^*(y))^2 Q(\mathrm{d}y|x) \right)\mathrm{d}x = \frac{5}{108} + \frac{13}{144n^2}.$$

Therefore, $J(P, Q, \gamma_{P_n}^*) = \frac{2}{27} + \frac{5}{72n^2}.$

REFERENCES

[1] T. BAŞAR AND P. BERNHARD, *H-infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, Birkhäuser, Boston, MA, 1995.

[2] G. BAKER AND S. YÜKSEL, *Continuity and robustness to incorrect priors in estimation and control*, in Proceedings of the IEEE Int. Symp. on Information Theory (ISIT), 2016, pp. 1999–2013.

[3] A. BENAVOLI AND L. CHISC, *Robust stochastic control based on imprecise probabilities*, IFAC Proceedings Volumes, 44 (2011), pp. 4606–4613.

[4] D. P. BERTSEKAS AND S. SHREVE, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.

[5] P. BILLINGSLEY, *Convergence of Probability Measures*, Wiley, New York, 1968.

[6] P. BILLINGSLEY AND F. TOPSØE, *Uniformity in weak convergence*, Probab. Theory Related Fields, 7 (1967), pp. 1–16.

[7] V. S. BORKAR, *Convex Analytic Methods in Markov Decision Processes*, in Handbook of Markov Decision Processes, E. A. Feinberg, A. Shwartz, eds., Kluwer, Boston, MA, 2001, pp. 347–375.

[8] A. BUDHIRAJA AND D. OCONE, *Exponential stability of discrete-time filters for bounded observation noise*, Systems Control Lett., 30 (1997), pp. 185–193.

[9] A. BUDHIRAJA AND D. OCONE, *Exponential stability in discrete-time filtering for non-ergodic signals*, Stochastic Process. Appl., 82 (1999), pp. 245–257.

[10] C. D. CHARALAMBOUS AND F. REZAEI, *Stochastic uncertain systems subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games*, IEEE Trans. Automat. Control, 52 (2007), pp. 647–663.

[11] P. CHIGANSKY, AND R. LIPTSER, AND R. VAN HANDEL, *Intrinsic methods in filter stability*, in Oxford Handbook of Nonlinear Filtering, D. Crisan and B. Rozovsky, eds., Oxford University Press, Oxford, 2011.

[12] L. DEVROYE AND L. GYÖRFI, *Non-parametric Density Estimation: The L_1 View*, Wiley, New York, 1985.

[13] R. M. DUDLEY, *Uniform Central Limit Theorems*, Cambridge University Press, Cambridge, 1999.

[14] R. M. DUDLEY, *Real Analysis and Probability*, 2nd ed., Cambridge University Press, Cambridge, 2002.

[15] R. M. DUDLEY AND E. GINE AND J. ZINN, *Uniform and universal Glivenko-Cantelli classes*, J. Theoret. Probab., 4 (1991), pp. 485–510.

[16] P. DUPUIS AND M. R. JAMES AND I. PETERSEN, *Robust properties of risk-sensitive control*, Math. Control Signals Systems, 13 (2000), pp. 318–332.

[17] E. B. DYNKIN AND A. A. YUSHKEVICH, *Controlled Markov Processes*, Springer, New York, 1979.

[18] E. A. FEINBERG, *On measurability and representation of strategic measures in Markov decision processes*, in Statistics Probability and Game Theory, Lecture Notes-Monograph Series 30, Institute of Mathematical Statistics, Hayward, CA, 1996, pp. 29–43.

[19] E.A. FEINBERG AND P.O. KASYANOV AND M.Z. ZGUROVSKY, *Partially observable total-cost Markov decision process with weakly continuous transition probabilities*, Math. Oper. Res., 41 (2016), pp. 656–681.

[20] A. GOEVA AND H. LAM AND H. QIAN AND B. ZHANG, *Reconstructing input models in stochastic simulation*, in Proceedings of the 2014 Winter Simulation Conference, IEEE Press, 2014, pp. 698–709.

[21] O. GOSSNER AND T. TOMALA, *Entropy bounds on Bayesian learning*, J. Math. Econom., 44 (2008), pp. 24–32.

[22] R. M. GRAY, *Entropy and Information Theory*, Springer, New York, 1990.

[23] A. GUPTA, AND S. YÜKSEL, AND T. BAŞAR, AND C. LANGBORT, *On the existence of optimal policies for a class of static and sequential dynamic teams*, SIAM J. Control Optim., 53 (2015), pp. 1681–1712.

[24] R. VAN HANDEL, *Discrete time nonlinear filters with informative observations are stable*, Electron. Commun. Probab., 13 (2008), pp. 562–575.

[25] L. P. HANSEN, AND T. J. SARGENT, *Robust control and model uncertainty*, Amer. Econ. Rev., 91 (2001), pp. 60–66.

[26] O. HERNÁNDEZ-LERMA, *Adaptive Markov Control Processes*, Springer, New York, 1989.

[27] O. HERNANDEZ-LERMA AND J. B. LASSERRE, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer, New York, 1996.

[28] O. HERNÁNDEZ-LERMA, AND J. B. LASSERRE, *Further Topics on Discrete-Time Markov Control Processes*, Springer, New York, 1999.

[29] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Markov Chains and Invariant Probabilities*, Birkhäuser, Basel, 2003.

[30] C-H. HSIEH AND B. R. BARMISH AND J. A. GUBNER, *Kelly Betting Can Be Too Conservative*, in Proceedings of the IEEE Conference on Decision and Control, 2016.

[31] D. JACOBSON, *Optimal Stochastic Linear systems with exponential performance criteria and their relation to deterministic differential games*, IEEE Trans. Automat. Control, 18 (1973), pp. 124–131.

[32] H. LAM, *Robust sensitivity analysis for stochastic systems*, Math. Oper. Res., 41 (2016), pp. 1248–1275.

[33] H.J. LANGEN, *Convergence of dynamic programming models*, Math. Oper. Res., 6 (1981), pp. 493–512.

[34] B. ØKSENDAL AND A. SULEM, *Forward–Backward stochastic differential games and stochastic control under model uncertainty*, J. Optim. Theory Appl., 161 (2014), pp. 22–55.

[35] K.R. PARTHASARATHY, *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.

[36] I. PETERSEN, M. R. JAMES AND P. DUPUIS, *Minimax optimal control of stochastic uncertain systems with relative entropy constraints*, IEEE Trans. Automat. Control, 45 (2000), pp. 398–412.

[37] P. DAI PRA, L. MENEGHINI AND W. J. RUNGGALDIER, *Connections between stochastic control and dynamic games*, Math. Control Signals Systems, 9 (1996), pp. 303–326.

[38] M. RAGINSKY, *Empirical processes, typical sequences and coordinated actions in standard Borel spaces*, IEEE Trans. Inform. Theory, 59 (2013), pp. 1288–1301.

[39] D. RHENIUS, *Incomplete information in Markovian decision models*, Ann. Statist., 2 (1974), pp. 1327–1334.

[40] W. RUDIN, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1987.

[41] N. SALDI, AND S. YÜKSEL, AND T. LINDER, *Near optimality of quantized policies in stochastic control under weak continuity conditions*, J. Math. Anal. Appl., 435 (2016), pp. 321–337.

[42] M. SCHÄL, *Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal*, Z. Wahrscheinlichkeitsth, 32 (1975), pp. 179–196

[43] R. SERFOZO, *Convergence of Lebesgue integrals with varying measures*, Sankhyā J. Stat. Series A, 44 (1982), pp. 380–402.

[44] I. TZORTZIS AND C.D. CHARALAMBOUS AND T. CHARALAMBOUS, *Dynamic programming subject to total variation distance ambiguity*, SIAM J. Control Optim., 53 (2015), pp. 2040–2075.

[45] R. VAN HANDEL, *The universal Glivenko–Cantelli property*, Probab. Theory Related Fields, 155 (2012), pp. 911–934.

[46] V. N. VAPNIK, *The Nature of Statistical Learning Theory*, 2nd ed., Springer, New York, 2000.

[47] C. VILLANI, *Optimal Transport: Old and New*, Springer, New York, 2008.

[48] R. L. WHEEDEN AND A. ZYGMUND, *Measure and Integral*, Marcel Dekker, New York, 1977.

[49] P. WHITTLE, *A risk-sensitive maximum principle: The case of imperfect state observation*, IEEE Trans. Automat. Control, 36 (1991), pp. 793–801.

[50] Y. WU AND S. VERDÚ, *Witsenhausen's Counterexample: A View from Optimal Transport Theory*, in Proceedings of the IEEE Conference on Decision and Control, 2016, pp. 5732–5737.

[51] Y. WU AND S. VERDÚ, *Functional properties of minimum mean-square error and mutual information*, IEEE Trans. Inform. Theory, 58 (2012), pp. 1289–1301.

[52] S. YÜKSEL AND T. LINDER, *Optimization and convergence of observation channels in stochastic control*, SIAM J. Control Optim., 50 (2012), pp. 864–887.

[53] A.A. YUSHKEVICH, *Reduction of a controlled Markov model with incomplete data to a problem with complete information in the case of Borel state and control spaces*, Theory Prob. Appl., 21 (1976), pp. 153–158.

[54] G. ZAMES, *Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms and approximate inverses*, IEEE Trans. Automat. Control, 26 (1981), pp. 301–320.

[55] K. ZHOU, J. C. DOYLE AND K. GLOVER, *Robust and Optimal Control*, vol. 40, Prentice-Hall, Upper Saddle River, New Jersey, 1996.

[56] H. ZHU AND E. ZHOU, *Risk Quantification in Stochastic Simulation under Input Uncertainty*, preprint, arXiv:1507.06015, 2015.