Formalization of Basic Dynamic Programming

UNH Team

March 17, 2026

Remark 0.0.1. This document omits basic definitions, such as the comparison or arithmetic operations on random variables. Arithmetic operations on random variables are performed element-wise for each element of the sample set. Please see the Lean file for complete details.

1 Probability

The treatment of the probability follows mostly the treatment in

Deisenroth, M.P., Faisal, A.A. and Ong, C.S. (2021) Mathematics for machine learning.

Although this document states that the values are in $\mathbb{R}$, the lean file is all defined in terms of rational numbers $\mathbb{Q}$ for computability reasons.

1.1 Prelude

This section sets up some basic tools for working with probabilities. See `Probability/Prelude.lean`.

Definition 1.1.1. A $p \in \mathbb{R}$ is a *probability value* when $p \geq 0 \wedge p \leq 1$. (*LEAN*: `Prob`)

The lean file also defines several basic properties for Theorem 1.1.1 which we do not summarize here.

Definition 1.1.2. A finite set of natural numbers is defined as $\mathbb{N}_n := \{0, \ldots, n-1\}$.

Definition 1.1.3. A list $\mathbb{L}_n : \mathbb{N}_n \to \tau$ of type τ and length n is a function defined for $\mathbb{N}_n := \{0, \ldots, n-1\}$ of type τ.

The following definition generalizes the inverse of a function to non-injective function which do not have an inverse.

Definition 1.1.4 (Preimage)**.** The *preimage* $f^{-1} : \mathcal{Y} \to \mathcal{X}$ of a function $f : \mathcal{X} \to \mathcal{Y}$ is defined as

$$f^{-1}(y) := \{x \in \mathcal{X} \mid f(x) = y\}. \tag{1}$$

1.2 Defs

This section contains definitions of probability spaces, probability operators, and the expectation operator. See `Probability/Defs.lean`.

Definition 1.2.1. A *finite probability distribution* $\Delta(n)$ for $n \in \mathbb{N}$ is defined as

$$\Delta_n := \left\{ p \in \mathbb{L}_n \mid \sum_{i \in \mathbb{N}_n} p_i = 1, p_i \geq 0, \forall i \in \mathbb{N}_n \right\}.$$

We call a probability distribution *degenerate* if $p_i = 1$ for some i; otherwise the probability distribution is *supported*.

The formal model we use to represent random events and processes is the following.

Definition 1.2.2 (Complete Probability Space)**.** A *complete probability space* consists of:

1. A sample space $\Omega := \mathbb{N}_n$, which is the set of all possible outcomes.
2. An *implicit* event space $\mathcal{F} := 2^\Omega$, which we do not define explicitly in Lean.
3. A probability function $\mathbb{P} \in \Delta_N$ that satisfies that
   (a) $\mathbb{P}(\Omega) = 1$
   (b) $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$ for all $A_1, A_2 \in \mathcal{F}$ such that $A_1 \cap A_2 = \emptyset$.

The notation 2^Ω in Theorem 1.2.2 represents the power set of Ω.

The set $\mathcal{F}$ is often, but not always, the power set of Ω. One reason it may not be a power set is that it is necessary when Ω is infinite; see Borel sets. Another reason to define $\mathcal{F}$ that is not a power set is to denote the available information. We will see an example of this later. Unless specified otherwise, we assume that the event set for a finite sample space is its power set and the σ-algebra of $\mathbb{R}$ is the Borel set.

Definition 1.2.3 (Measure)**.** The measure $\mathbb{P}[\mathcal{S}]$ of the set $\mathcal{S} \subseteq \mathbb{N}$ given a probability space is defined as

$$\mathbb{P}[\mathcal{S}] := \sum_{i \in \Omega} \mathbb{I}[i \in \mathcal{S}].$$

The definition requires that the set is decidable.

Example 1.2.4. Suppose I want to model the behavior of coin flips. I have a coin that is equally likely to come up heads and tails. I flip the coin twice. One possible probability space would be

1. $\Omega := \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$.
2. $\mathcal{F} := 2^\Omega = \{\emptyset, \Omega, \{\mathrm{HH}\}, \{\mathrm{HH}, \mathrm{HT}\}, \dots\}$.
3. $P(\{\mathrm{HH}\}) := \frac{1}{4}, P(\{\mathrm{HH}, \mathrm{HT}\}) := \frac{1}{2}$. The probabilities of other sets are derived using the properties in Theorem 1.2.2.

For the sake of completeness, we include also the notion of measurability. However, since we are only concerned with complete probability spaces, all random variables we define will be measurable.

Definition 1.2.5. A function $f\colon \mathcal{X} \to \mathcal{Y}$ with σ-algebras $\mathcal{D} \subseteq 2^{\mathcal{X}}$ and $\mathcal{E} \subseteq 2^{\mathcal{Y}}$ is *measurable* when the *pre-image* of each $E \in \mathcal{E}$ is in $\mathcal{D}$:

$$f^{-1}(E) \in \mathcal{D}.$$

Here, the application to f^{-1} to a set $E \subseteq \mathcal{Y}$ is defined as:

$$f^{-1}(E) := \bigcup_{e \in E} f^{-1}(e). \tag{2}$$

The measurability in the definition of random variables is required for us to reason about their probability distributions.

Definition 1.2.6 (Random Variable)**.** A *random variable* is a measurable function $\tilde{x}\colon \mathbb{N} \to \mathcal{Y}$ where $\mathcal{Y}$ is a measurable set. Note that the way we define random variables, they are independent of a particular probability space. The measurability requirement is vacuous in probability spaces.

When manipulating random variables we denote them with a tilde, such as $\tilde{x}, \tilde{s}$. This corresponds to the common practice to use uppercase letters to denote random variables. We drop the tilde when treating the random variable as an ordinary function.

[TODO: We could get a lot of mileage from treating random variables as vectors. Expectation and probability are then linear operators on this vector space of random variables. This is only possible when using a complete vector space. Question: Do we need finiteness for this? That is, do random variables need to be defined for a finite Ω?]

Remark 1.2.7. We define some ad-hoc operations on random variables. However, it would be better to treat them as a vector space. Then probability and expectation can be seen as linear operators on this vector space.

Other notable definition is that the equality with a scalar $\tilde{x} = x_0$ is interpreted as a boolean random variable $\tilde{b} : \mathbb{B} \to \{0, 1\}$:

$$b(\omega) := (\tilde{x} = x_0)(\omega) := \mathbb{I}[x(\omega) = x_0].$$

We use the operator (function) $\mathbb{P}$ denote the probability of the set of sample space that corresponds to a given condition. We use a non-standard definition for convenience (and show that this definition equals to the standard one later).

Definition 1.2.8. For any complete probability space, $(\Omega, \mathcal{F}, P)$ and a random variable $\tilde{b}\colon \Omega \to \mathbb{B}$, the operator $\mathbb{P}$ is defined as

$$\mathbb{P}\left[\tilde{b}\right] := \sum_{i \in \Omega} \mathbb{P}(i) b(i),$$

where the boolean is interpreted as 0 for false and 1 for true.

The definition can be seen as an inner product between the probabilities and the random variable.

The following theorem shows that our definition of a probability operator is equal to the standard definition.

Theorem 1.2.9. *For any probability space, $(\Omega, \mathcal{F}, P)$ and a random variable $\tilde{b} \colon \Omega \to \mathbb{B}$:*

$$\mathbb{P}\left[\tilde{b}\right] := \mathbb{P}[b^{-1}(\text{true})].$$

Probability distributions are typically defined for random variables. The distribution of a random variable is characterized by the *cumulative distribution function* (CDF), defined as follows.

Definition 1.2.10 (CDF)**.** The *cumulative distribution function* $F_{\tilde{x}} \colon \mathbb{R} \to [0, 1]$ of a real-valued random variable $\tilde{x} \colon \Omega \to \mathbb{R}$ is defined as

$$F_{\tilde{x}}(x) := \mathbb{P}\left[\tilde{x} \leq x\right], \qquad \forall x \in \mathbb{R}.$$

Some random variables also have a *probability density function* (PDF), and discrete random variables have a *probability mass function* (PMF). In this class, we will be primarily dealing with discrete random variables.

Definition 1.2.11 (PMF)**.** The *probability mass function* $p_{\tilde{x}} \colon \mathcal{E} \to [0, 1]$ for a discrete random variable $\tilde{x} \colon \Omega \to \mathcal{E}$ with a finite $\mathcal{E}$ is defined as

$$p_{\tilde{x}}(e) := \mathbb{P}\left[\tilde{x} = e\right], \qquad \forall e \in \mathcal{E}.$$

Definition 1.2.12. The *joint probability* of two random variables $\tilde{x} \colon \Omega \to \mathcal{X}$ and $\tilde{y} \colon \Omega \to \mathcal{Y}$ is defined as

$$\mathbb{P}\left[\tilde{x} = x, \tilde{y} = y\right] := \mathbb{P}\left[\tilde{x} = x \wedge \tilde{y} = y\right], \forall x \in \mathcal{X}, y \in \mathcal{Y}.$$

One can also define a joint PMF for two discrete random variables $p_{\tilde{x},\tilde{y}}(x, y)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ as

$$p_{\tilde{x},\tilde{y}}(x, y) = \mathbb{P}\left[\tilde{x} = x, \tilde{y} = y\right].$$

Definition 1.2.13. The *conditional probability* of $\tilde{b} \colon \Omega \to \mathbb{B}$ on $\tilde{c} \colon \Omega \to \mathbb{B}$ is defined as

$$\mathbb{P}\left[\tilde{b} \mid \tilde{c}\right] := \frac{\mathbb{P}[\tilde{b} \wedge \tilde{c}]}{\mathbb{P}\left[\tilde{c}\right]}$$

Definition 1.2.14 (Independent RV)**.** Events $A, B \in \mathcal{F}$ are independent when

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B].$$

Random variables $\tilde{x} \colon \Omega \to \mathcal{X}$ and $\tilde{y} \colon \Omega \to \mathcal{Y}$ are *independent* when

$$F_{\tilde{x},\tilde{y}}(x, y) = F_{\tilde{x}}(x) \cdot F_{\tilde{y}}(y), \forall x \in \mathcal{X}, y \in \mathcal{Y}.$$

Discrete random variables $\tilde{x}$ and $\tilde{y}$ are independent if and only if

$$\mathbb{P}\left[\tilde{x} = x, \tilde{y} = y\right] = \mathbb{P}\left[\tilde{x} = x\right] \cdot \mathbb{P}\left[\tilde{y} = y\right], \forall x \in \mathcal{X}, y \in \mathcal{Y}.$$

The expectation of a random variable is its mean. As with probabilities, we use a non-standard definition for the sake of computational convenience and then show that it is equivalent to the common definition of probabilities.

Definition 1.2.15 (Expected Value)**.** Suppose that $\tilde{x} \colon \Omega \to \mathbb{R}$ is a random variable and that $x(\Omega) := \{x(\omega) \mid \omega \in \Omega\}$ is finite. Then

$$\mathbb{E}[\tilde{x}] := \sum_{\omega \in \Omega} x(\omega) \cdot \mathbb{P}(\omega).$$

The following theorem shows that our definition of expected value is equal to the standard definition of expected values.

Theorem 1.2.16 (Expected Value)**.** *Suppose that $\tilde{x} \colon \Omega \to \mathbb{R}$ is a random variable and that $x(\Omega) := \{x(\omega) \mid \omega \in \Omega\}$ is finite. Then*

$$\mathbb{E}[\tilde{x}] = \sum_{x \in \tilde{x}(\Omega)} x \cdot \mathbb{P}[\tilde{x} = x].$$

In general probability spaces, the expectation is defined using the Lebesgue integral:

$$\mathbb{E}[\tilde{x}] := \int_{\Omega} \tilde{x} \, dP,$$

which is equivalent to the form in the theorem above in finite probability spaces.

Some of the most important properties of expectations are as follows.

Theorem 1.2.17. *Assume $\tilde{x}, \tilde{y} \colon \Omega \to \mathbb{R}$ and $c \in \mathbb{R}$. Then:*

$$\begin{aligned} \mathbb{E}[c \cdot \tilde{x}] &= c \cdot \mathbb{E}[\tilde{x}] \\ \mathbb{E}[\tilde{x} + \tilde{y}] &= \mathbb{E}[\tilde{x}] + \mathbb{E}[\tilde{y}] \\ \mathbb{E}[\tilde{x} \cdot \tilde{y}] &= \mathbb{E}[\tilde{x}] \cdot \mathbb{E}[\tilde{y}] \quad \textit{for independent } \tilde{x}, \tilde{y}. \end{aligned}$$

Definition 1.2.18. A *random variable* defined on a finite probability space P is a mapping $\tilde{x} \colon \Omega \to \mathbb{R}$.

1.3 Induction

This section includes properties that aim to enable one to perform induction on probability spaces.

1.4 Basic

This section states the basic properties of probability spaces, and operators.

Definition 1.4.1. A *finite probability measure* $p \colon \Omega \to \mathbb{R}_+$ on a finite set Ω is any function that satisfies

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

Theorem 1.4.2 (Law of the unconscious statistician)**.** *For any discrete* $\tilde{x} \colon \Omega \to \mathbb{R}$ *and* $g \colon \mathbb{R} \to \mathbb{R}$*:*

$$\mathbb{E}\left[g(\tilde{x})\right] = \sum_{x \in \mathcal{X}} g(x) \cdot \mathbb{P}\left[\tilde{x} = x\right].$$

For the remainder of **??**, we assume that $P = (\Omega, p)$ is a *finite probability space*. All random variables are defined on the space P unless specified otherwise.

1.5 OLD CONTENT: TO BE MOVED

Here, we define probability and expectation operators.

Definition 1.5.1. A *boolean* set is $\mathbb{B} = \{\text{false}, \text{true}\}$.

Definition 1.5.2. The *expectation* of a random variable $\tilde{x} \colon \Omega \to \mathbb{R}$ is

$$\mathbb{E}\left[\tilde{x}\right] := \sum_{\omega \in \Omega} p(\omega) \cdot \tilde{x}(\omega).$$

Definition 1.5.3. An *indicator* function $\mathbb{I} \colon \mathbb{B} \to \{0, 1\}$ is defined for $b \in \mathbb{B}$ as

$$\mathbb{I}(b) := \begin{cases} 1 & \text{if } b = \text{true}, \\ 0 & \text{if } b = \text{false}. \end{cases}$$

Definition 1.5.4. The *conditional expectation* of $\tilde{x} \colon \Omega \to \mathbb{R}$ conditioned on $\tilde{b} \colon \Omega \to \mathbb{B}$ is defined as

$$\mathbb{E}\left[\tilde{x} \mid \tilde{b}\right] := \frac{1}{\mathbb{P}[\tilde{b}]} \mathbb{E}\left[\tilde{x} \cdot \mathbb{I} \circ \tilde{b}\right],$$

where we define that $x/0 = 0$ for each $x \in \mathbb{R}$.

Remark 1.5.5. It is common to prohibit conditioning on a zero probability event both for expectation and probabilities. In this document, we follow the Lean convention, where the division by 0 is 0; see `div_zero`. However, even some basic probability and expectation results may require that we assume that the conditioned event does not have probability zero for it to hold.

Definition 1.5.6. The *random conditional expectation* of a random variable $\tilde{x} \colon \Omega \to \mathbb{R}$ conditioned on $\tilde{y} \colon \Omega \to \mathcal{Y}$ for a finite set $\mathcal{Y}$ is the random variable $\mathbb{E}\left[\tilde{x} \mid \tilde{y}\right] : \Omega \to \mathbb{R}$ is defined as

$$\mathbb{E}\left[\tilde{x} \mid \tilde{y}\right](\omega) := \mathbb{E}\left[\tilde{x} \mid \tilde{y} = \tilde{y}(\omega)\right], \quad \forall \omega \in \Omega.$$

Remark 1.5.7. The Lean file defines expectations more broadly for a data type ρ which is more general than just $\mathbb{R}$. The main reason to generalize to both $\mathbb{R}$ and $\mathbb{R}_+$. However, in principle, the definitions could be used to reason with expectations that go beyond real numbers and may include other algebras, such as vectors or matrices.

Lemma 1.5.8. *Suppose that* $\tilde{b}, \tilde{c} \colon \Omega \to \mathbb{B}$. *Then:*

$$\mathbb{I}\left(\tilde{b} \wedge \tilde{c}\right) = \mathbb{I}(\tilde{b}) \cdot \mathbb{I}(\tilde{c}),$$

where the equality applies for all $\omega \in \Omega$.

Theorem 1.5.9. *Suppose that* $\tilde{c} \colon \Omega \to \mathbb{B}$ *such that* $\mathbb{P}[\tilde{c}] = 0$. *Then for any* $\tilde{x} \colon \Omega \to \mathbb{R}$:

$$\mathbb{E}[\tilde{x} \mid \tilde{c}] = 0.$$

Proof. Immediate from the definition and the fact that $0 \cdot x = 0$ for $x \in \mathbb{R}$. □

Theorem 1.5.10. *Suppose that* $\tilde{c} \colon \Omega \to \mathbb{B}$ *such that* $\mathbb{P}[\tilde{c}] = 0$. *Then for any* $\tilde{b} \colon \Omega \to \mathbb{R}$:

$$\mathbb{P}\left[\tilde{b} \mid \tilde{c}\right] = 0.$$

Proof. Immediate from Theorem 1.5.9. □

Theorem 1.5.11. *Suppose that* $\tilde{b}, \tilde{c} \colon \Omega \to \mathbb{B}$, *then*

$$\mathbb{P}\left[\tilde{b} \wedge \tilde{c}\right] = \mathbb{P}\left[\tilde{b} \mid \tilde{c}\right] \cdot \mathbb{P}[\tilde{c}].$$

Proof. The property holds immediately when $\mathbb{P}[\tilde{c}] = 0$. Assume that $\mathbb{P}[\tilde{c}] > 0$. Then:

$$\begin{aligned}
\mathbb{P}\left[\tilde{b} \wedge \tilde{c}\right] &= \mathbb{E}\left[\mathbb{I}(\tilde{b} \wedge \tilde{c})\right] && [\textbf{??}] \\
&= \mathbb{E}\left[\mathbb{I}(\tilde{b}) \cdot \mathbb{I}(\tilde{c})\right] && [\text{Theorem 1.5.8}] \\
&= \frac{1}{\mathbb{P}[\tilde{c}]} \mathbb{E}\left[\mathbb{I}(\tilde{b}) \cdot \mathbb{I}(\tilde{c})\right] \cdot \mathbb{P}[\tilde{c}] && \cdot 1 \\
&= \mathbb{E}\left[\mathbb{I}(\tilde{b}) \mid \tilde{c}\right] \cdot \mathbb{P}[\tilde{c}] && [\text{Theorem 1.5.4}] \\
&= \mathbb{P}\left[\tilde{b} \mid \tilde{c}\right] \cdot \mathbb{P}[\tilde{c}] && [\text{Theorem 1.2.13}].
\end{aligned}$$

□

Lemma 1.5.12. *Let* $\tilde{y} \colon \Omega \to \mathcal{Y}$ *with a finite* $\mathcal{Y}$. *Then*

$$\mathbb{P}[\tilde{y} = y(\omega)] \geq p(\omega), \quad \omega \in \Omega.$$

Proof.

$$\begin{aligned}
\mathbb{P}[\tilde{y} = y(\omega)] &= \sum_{\omega' \in \Omega} p(\omega) \cdot \mathbb{I}(\tilde{y}(\omega') = \tilde{y}(\omega)) && [\textbf{??}] \\
&\geq p(\omega) && \omega \in \Omega[\text{ and }] p(\omega') \geq 0, \forall \omega' \in \Omega.
\end{aligned}$$

□

Remark 1.5.13. Theorem 1.5.19 shows the equivalence of expectations for surely equal random variables.

Theorem 1.5.14. *Random variables $\tilde{x}, \tilde{y} \colon \Omega \to \mathbb{R}$ satisfy that*

$$\mathbb{E}\left[\tilde{x} + \tilde{y}\right] = \mathbb{E}\left[\tilde{x}\right] + \mathbb{E}\left[\tilde{y}\right].$$

Proof. From the distributive property of sums. □

Theorem 1.5.15. *A random variable $\tilde{x} \colon \Omega \to \mathbb{R}$ and $c \in \mathbb{R}$ satisfies that*

$$\mathbb{E}\left[c\right] = c.$$

Theorem 1.5.16. *Suppose that $\tilde{x} \colon \Omega \to \mathbb{R}$ and $c \in \mathbb{R}$. Then*

$$\mathbb{E}\left[c + \tilde{x}\right] = c + \mathbb{E}\left[\tilde{x}\right].$$

Proof. From Theorems 1.5.14 and 1.5.15. □

Theorem 1.5.17. *Suppose that $\tilde{x}, \tilde{y} \colon \Omega \to \mathbb{R}$ and $\tilde{z} \colon \Omega \to \mathcal{V}$ are random variables and $c \in \mathbb{R}$, such that $\tilde{y}(\omega) = c + \tilde{x}(\omega)$. Then*

$$\mathbb{E}\left[\tilde{y} \mid \tilde{z}\right](\omega) = c + \mathbb{E}\left[\tilde{x} \mid \tilde{z}\right](\omega), \quad \forall \omega \in \Omega.$$

Proof. From Theorem 1.5.16. □

Theorem 1.5.18. *Suppose that $\tilde{x}, \tilde{y} \colon \Omega \to \mathbb{R}$ satisfy that*

$$\forall \omega \in \Omega, p(\omega) > 0 \Rightarrow \tilde{x}(\omega) \geq \tilde{y}(\omega).$$

Then

$$\mathbb{E}\left[\tilde{x}\right] \geq \mathbb{E}\left[\tilde{y}\right].$$

Theorem 1.5.19 (Congruence of Expectation)**.** *Suppose that $\tilde{x}, \tilde{z} \colon \Omega \to \mathbb{R}$ satisfy that*

$$\forall \omega \in \Omega, p(\omega) > 0 \Rightarrow \tilde{x}(\omega) = \tilde{z}(\omega).$$

Then

$$\mathbb{E}\left[\tilde{x}\right] = \mathbb{E}\left[\tilde{z}\right].$$

Proof. Immediately from the congruence of sums. □

1.6 The Laws of The Unconscious Statisticians

Theorem 1.6.1. *Let $\tilde{x} \colon \Omega \to \mathbb{R}$ be a random variable. Then:*

$$\mathbb{E}\left[\tilde{x}\right] = \sum_{x \in \tilde{x}(\Omega)} \mathbb{P}\left[\tilde{x} = x\right] \cdot x.$$

Proof. Let $\mathcal{X} := \tilde{x}(\Omega)$, which is a finite set. Then:

$$\begin{aligned}
\mathbb{E}[\tilde{x}] &= \sum_{\omega\in\Omega} p(\omega)\cdot\tilde{x}(\omega) && \text{[Theorem 1.5.2]}\\
&= \sum_{\omega\in\Omega}\sum_{x\in\mathcal{X}} p(\omega)\cdot\tilde{x}(\omega)\cdot\mathbb{I}(x=\tilde{x}(\omega)) && \text{[??]}\\
&= \sum_{\omega\in\Omega}\sum_{x\in\mathcal{X}} p(\omega)\cdot x\cdot\mathbb{I}(x=\tilde{x}(\omega)) && \text{[??]}\\
&= \sum_{x\in\mathcal{X}} x\cdot\sum_{\omega\in\Omega} p(\omega)\cdot\mathbb{I}(x=\tilde{x}(\omega)) && \text{[??]}\\
&= \sum_{x\in\mathcal{X}} x\cdot\mathbb{E}[\mathbb{I}(x=\tilde{x}(\omega))] && \text{[Theorem 1.5.2]}\\
&= \sum_{x\in\mathcal{X}} x\cdot\mathbb{P}[x=\tilde{x}(\omega)]\,. && \text{[??]}
\end{aligned}$$

□

The following theorem generalizes the theorem above.

Theorem 1.6.2. *Let $\tilde{x}\colon\Omega\to\mathbb{R}$ and $\tilde{b}\colon\Omega\to\mathcal{Y}$ be random variables. Then:*

$$\mathbb{E}\left[\tilde{x}\mid\tilde{b}\right] = \sum_{x\in\tilde{x}(\Omega)}\mathbb{P}\left[\tilde{x}=x\mid\tilde{b}\right]\cdot x.$$

Theorem 1.6.3. *Let $\tilde{x}\colon\Omega\to\mathbb{R}$ and $\tilde{y}\colon\Omega\to\mathcal{Y}$ be random variables with $\mathcal{Y}$ finite. Then:*

$$\mathbb{E}[\mathbb{E}[\tilde{x}\mid\tilde{y}]] = \sum_{y\in\mathcal{Y}}\mathbb{E}[\tilde{x}\mid\tilde{y}=y]\cdot\mathbb{P}[\tilde{y}=y]\,.$$

1.7 Total Expectation and Probability

Theorem 1.7.1 (Law of Total Probability)**.** *Let $\tilde{b}\colon\Omega\to\mathbb{B}$ and $\tilde{y}\colon\Omega\to\mathcal{Y}$ be random variables with a finite set $\mathcal{Y}$. Then:*

$$\sum_{y\in\mathcal{Y}}\mathbb{P}\left[\tilde{b}\wedge(\tilde{y}=y)\right] = \mathbb{P}\left[\tilde{b}\right].$$

Theorem 1.7.2 (Law of Total Expectation)**.** *Let $\tilde{x}\colon\Omega\to\mathcal{X}$ and $\tilde{y}\colon\Omega\to\mathcal{Y}$ be random variables with a finite set $\mathcal{Y}$. Then:*

$$\mathbb{E}[\mathbb{E}[\tilde{x}\mid\tilde{y}]] = \mathbb{E}[\tilde{x}]\,.$$

Proof. Recall that we are allowing the division by 0 and assume that $x/0 = 0$.

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[\tilde{x} \mid \tilde{y}\right]\right] &= \sum_{\omega\in\Omega} p(\omega)\cdot \mathbb{E}\left[\tilde{x} \mid \tilde{y}\right](\omega) && \text{[Theorem 1.5.2]}\\
&= \sum_{\omega\in\Omega} p(\omega)\cdot \mathbb{E}\left[\tilde{x} \mid \tilde{y} = \tilde{y}(\omega)\right] && \text{[Theorem 1.5.6]}\\
&= \sum_{\omega\in\Omega} \frac{p(\omega)}{\mathbb{P}\left[\tilde{y} = \tilde{y}(\omega)\right]} \sum_{\omega'\in\Omega} p(\omega')\cdot \tilde{x}(\omega')\cdot \mathbb{I}\left(\tilde{y}(\omega') = \tilde{y}(\omega)\right) && \text{[Theorem 1.5.4]}\\
&= \sum_{\omega'\in\Omega} p(\omega')\cdot \tilde{x}(\omega')\cdot \sum_{\omega\in\Omega} \frac{p(\omega)}{\mathbb{P}\left[\tilde{y} = \tilde{y}(\omega)\right]} \mathbb{I}\left(\tilde{y}(\omega') = \tilde{y}(\omega)\right) && \text{[rearrange]}\\
&= \sum_{\omega'\in\Omega} p(\omega')\cdot \tilde{x}(\omega')\cdot \sum_{\omega\in\Omega} \frac{p(\omega)}{\mathbb{P}\left[\tilde{y} = \tilde{y}(\omega')\right]} \mathbb{I}\left(\tilde{y}(\omega') = \tilde{y}(\omega)\right) && \text{[equals when }]\tilde{y}(\omega') = \tilde{y}(\omega)\\
&= \sum_{\omega'\in\Omega} p(\omega')\cdot \tilde{x}(\omega') && \text{[see below]}\\
&= \mathbb{E}\left[\tilde{x}\right].
\end{aligned}$$

Above, we used the fact that

$$p(\omega')\cdot \sum_{\omega\in\Omega} \frac{p(\omega)}{\mathbb{P}\left[\tilde{y} = \tilde{y}(\omega')\right]} \mathbb{I}\left(\tilde{y}(\omega') = \tilde{y}(\omega)\right) = p(\omega'),$$

which follows by analyzing two cases. First, when $p(\omega') = 0$, then the equality holds immediately. If $p(\omega') > 0$ then by Theorem 1.5.12, $\mathbb{P}\left[\tilde{y} = \tilde{y}(\omega')\right] > 0$, we get from **??** that

$$\sum_{\omega\in\Omega} \frac{p(\omega)}{\mathbb{P}\left[\tilde{y} = \tilde{y}(\omega')\right]} \mathbb{I}\left(\tilde{y}(\omega') = \tilde{y}(\omega)\right) = \frac{\mathbb{P}\left[\tilde{y} = \tilde{y}(\omega')\right]}{\mathbb{P}\left[\tilde{y} = \tilde{y}(\omega')\right]} = 1,$$

which completes the step. □

The following proof is simpler but may require some more advanced properties.

Alternate proof.

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[\tilde{x} \mid \tilde{y}\right]\right] &= \sum_{y\in\mathcal{Y}} \mathbb{E}\left[\tilde{x} \mid \tilde{y} = y\right]\cdot \mathbb{P}\left[\tilde{y} = y\right]\\
&= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} x\cdot \mathbb{P}\left[\tilde{x} = x \mid \tilde{y} = y\right]\cdot \mathbb{P}\left[\tilde{y} = y\right]\\
&= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} x\cdot \mathbb{P}\left[\tilde{x} = x, \tilde{y} = y\right]\\
&= \sum_{x\in\mathcal{X}} x\cdot \sum_{y\in\mathcal{Y}} \mathbb{P}\left[\tilde{x} = x, \tilde{y} = y\right]\\
&= \sum_{x\in\mathcal{X}} x\cdot \mathbb{P}\left[\tilde{x} = x\right] = \mathbb{E}\left[\tilde{x}\right].
\end{aligned}$$

□

1.8 Non-Degeneracy

Theorem 1.8.1 (Non-degeneracy of L_1-norm)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a discrete probability space with $\Omega = \{\omega_1, \omega_2, ...\}$ countable, and let $p_i = P(\{\omega_i\}) \geq 0$ with $\sum_i p_i = 1$. Let X be a random variable with $X_i = X(\omega_i)$. Then*

$$\mathbb{E}[|X|] = 0 \iff \mathbb{P}(\{\omega \in \Omega : X(\omega) = 0\}) = 1.$$

Proof. Proof of ($\Leftarrow$): Assume $\mathbb{P}(X = 0) = 1$. If $\mathbb{P}(X = 0) = 1$, then $\mathbb{P}(X \neq 0) = 0$. In discrete terms $p_i = 0$ for all ω_i such that $X_i \neq 0$. Now compute $\mathbb{E}[|X|]$ such that

$$\mathbb{E}[|X|] = \sum_i |X_i| \cdot p_i.$$

Case 1: $X_i = 0 \Rightarrow |X_i|p_i = 0 \cdot p_i = 0$.
Case 2: $X_i \neq 0 \Rightarrow p_i = 0 \Rightarrow |X_i|p_i = |X_i| \cdot 0 = 0$. Every term is zero, so $\mathbb{E}[|X|] = 0$.

Proof of ($\Rightarrow$): Assume $\mathbb{E}[|X|] = 0$. We have

$$\mathbb{E}[|X|] = \sum_i |X_i| \cdot p_i = 0,$$

where $|X_i|p_i \geq 0$ for all i. For a sum of nonnegative terms to be zero, each term must be zero

$$|X_i|p_i = 0 \quad \text{for all } i.$$

Thus, for each i, either $p_i = 0$ or $|X_i| = 0$ (i.e., $X_i = 0$). Let $N = \{\omega_i : X_i \neq 0\}$. For $\omega_i \in N$, we have $X_i \neq 0 \Rightarrow |X_i| \neq 0 \Rightarrow$ from $|X_i|p_i = 0$ we must have $p_i = 0$. Therefore:

$$\mathbb{P}(X \neq 0) = \sum_{\omega_i \in N} p_i = \sum_{\omega_i \in N} 0 = 0.$$

Thus ¶$(X = 0) = 1 - \mathbb{P}(X \neq 0) = 1$. Hence, we conclude that

$$\mathbb{E}[|X|] = 0 \iff \mathbb{P}(X = 0) = 1.$$

□

2 Formal Decision Framework

2.1 Markov Decision Process

Definition 2.1.1. A *Markov decision process* $M := (S, A, P, r)$ consists of:

- a positive integer $S \in \mathbb{N}_{>0}$ representing the number of states, with index set $\text{Fin}(S) = \{0, 1, ..., S-1\}$
- a positive integer $A \in \mathbb{N}_{>0}$ representing the number of actions, with index set $\text{Fin}(A) = \{0, 1, ..., A-1\}$
- a transition function $P\colon \text{Fin}(S) \times \text{Fin}(A) \to \Delta(\text{Fin}(S))$, mapping each state–action pair to finite probability distribution over next states
- a reward function $r\colon \text{Fin}(S) \times \text{Fin}(A) \times \text{Fin}(S) \to \mathbb{R}$, mapping each transition (s, a, s') to a real number

2.2 Histories

We implicitly assume in the remainder of the section an MDP $M = (S, A, P, r)$.

Definition 2.2.1. A *history* h in a set of histories $\mathcal{H}$ is a sequence of states and actions defined for M recursively as

$$h := \langle s_0 \rangle, \quad \text{[or]} \quad h := \langle h', a, s \rangle,$$

where $s_0, s \in \mathrm{Fin}(S)$, $a \in \mathrm{Fin}(A)$, and $h' \in \mathcal{H}$.

Definition 2.2.2. The *length* $|h| \in \mathbb{N}$ of a history $h \in \mathcal{H}$ is defined as

$$\begin{aligned} |\langle s \rangle| &:= 0, \\ |\langle h', s, a \rangle| &:= 1 + |h'|, \qquad h' \in \mathcal{H}. \end{aligned}$$

Definition 2.2.3. The set $\mathcal{H}_{\mathrm{NE}}$ of *non-empty histories* is

$$\mathcal{H}_{\mathrm{NE}} := \{h \in \mathcal{H} \mid |h| \geq 1\}.$$

Definition 2.2.4. *Following histories* $\mathcal{H}(h, t) \subseteq \mathcal{H}$ for $h \in \mathcal{H}$ of length $t \in \mathbb{N}$ are defined recursively as

$$\mathcal{H}(h, t) := \begin{cases} \{h\} & \text{if } t = 0, \\ \{\langle h', a, s \rangle \mid h \in \mathcal{H}(h', t-1), a \in \mathcal{A}, s \in \mathcal{S}\} & \text{otherwise.} \end{cases}$$

Definition 2.2.5. The set of *histories* $\mathcal{H}_t$ of *length* $t \in \mathbb{N}$ is defined recursively as

$$\mathcal{H}_t = \begin{cases} \{\langle s \rangle \mid s \in \mathcal{S}\} & \text{if } t = 0, \\ \{\langle h, a, s \rangle \mid h \in \mathcal{H}_{t-1}, a \in \mathcal{A}, s \in \mathcal{S}\} & \textit{textotherwise.} \end{cases}$$

Theorem 2.2.6. *For $h \in \mathcal{H}$:*

$$|h'| = |h| + t, \qquad \forall h' \in \mathcal{H}(h, t).$$

Proof. The theorem follows by induction on t from the definition. □

Definition 2.2.7. We use $\tilde{s}_k \colon \mathcal{H} \to \mathcal{S}$ to denote the 0-based k-th state of each history.

Definition 2.2.8. We use $\tilde{a}_k \colon \mathcal{H} \to \mathcal{A}$ to denote the 0-based k-th action of each history.

Definition 2.2.9. The *history-reward* random variable $\tilde{r}^{\mathrm{h}} \colon \mathcal{H} \to \mathbb{R}$ for $h = \langle h', a, s \rangle \in \mathcal{H}$ for $h' \in \mathcal{H}$, $a \in \mathcal{A}$, and $s \in \mathcal{S}$ is defined recursively as

$$\tilde{r}^{\mathrm{h}}(h) := r(s_{|h|}(h'), a, s) + r_{\mathrm{h}}(h').$$

Definition 2.2.10. The *history-reward* random variable $\tilde{r}_k^{\mathrm{h}} \colon \mathcal{H} \to \mathbb{R}$ for $h = \langle h', a, s \rangle \in \mathcal{H}$ for $h' \in \mathcal{H}$, $a \in \mathcal{A}$, and $s \in \mathcal{S}$ is defined as the k-th reward (0-based) of a history.

Definition 2.2.11. The *history-reward* random variable $\tilde{r}_{\leq k}^{\mathrm{h}} \colon \mathcal{H} \to \mathbb{R}$ for $h = \langle h', a, s \rangle \in \mathcal{H}$ for $h' \in \mathcal{H}$, $a \in \mathcal{A}$, and $s \in \mathcal{S}$ is defined as the sum of all k-th or earlier rewards (0-based) of a history.

Definition 2.2.12. The *history-reward* random variable $\tilde{r}_{\geq k}^{\mathrm{h}} \colon \mathcal{H} \to \mathbb{R}$ for $h = \langle h', a, s \rangle \in \mathcal{H}$ for $h' \in \mathcal{H}$, $a \in \mathcal{A}$, and $s \in \mathcal{S}$ is defined as the sum of k-th or later reward (0-based) of a history.

2.3 Policies

Definition 2.3.1. The set of *decision rules* $\mathcal{D}$ is defined as $\mathcal{D} := \mathcal{A}^{\mathcal{S}}$. A single action $a \in \mathcal{A}$ can also be interpreted as a decision rule $d := s \mapsto a$.

Definition 2.3.2. The set of *history-dependent policies* is $\Pi_{\mathrm{HR}} := \Delta(\mathcal{A})^{\mathcal{H}}$.

Definition 2.3.3. The set of *Markov deterministic policies* Π_{MD} is $\Pi_{\mathrm{MD}} := \mathcal{D}^{\mathbb{N}}$. A Markov deterministic policy $\pi \in \Pi_{\mathrm{MD}}$ can also be interpreted as $\bar{\pi} \in \Pi_{\mathrm{HR}}$:

$$\bar{\pi}(h) := \delta\left[\pi(|h|, s_{|h|}(h))\right],$$

where δ is the Dirac distribution, and $s_{|h|}$ is the history's last state.

Definition 2.3.4. The set of *stationary deterministic policies* Π_{SD} is defined as $\Pi_{\mathrm{SD}} := \mathcal{D}$. A stationary policy $\pi \in \Pi_{\mathrm{SD}}$ can be interpreted as $\bar{\pi} \in \Pi_{\mathrm{HR}}$:

$$\bar{\pi}(h) := \delta\left[\pi(s_{|h|}(h))\right],$$

where δ is the Dirac distribution and $s_{|h|}$ is the history's last state.

2.4 Distribution

Definition 2.4.1. The *history probability distribution* $p_T^{\mathrm{h}} \colon \Pi_{\mathrm{HR}} \to \Delta(\mathcal{H}(h,t))$ and $\pi \in \Pi_{\mathrm{HR}}$ is defined for each $T \in \mathbb{N}$ and $h \in \mathcal{H}(\widehat{h}, t)$ as

$$(p_T^{\mathrm{h}}(\pi))(h) := \begin{cases} \mathbb{I}(h = \widehat{h}) & \text{if } T = 0, \\ p_{T-1}^{\mathrm{h}}(h', \pi) \cdot \pi(h', a) \cdot p(s_{|h|}(h'), a, s) & \text{if } T > 1 \wedge h = \langle h', a, s \rangle. \end{cases}$$

Moreover, the function p^{h} maps policies to correct probability distribution.

TODO: This definition needs to be updated. A probability space $(\Omega_{h,t}, 2^{\Omega_{h,t}}, \hat{p}_{h,\pi})$ which is defined as

$$\Omega_{h,t} := \left\{ h' \in \mathcal{H}_{|h|+t} \mid s_k(h) = s_k(h') \wedge a_k(h) = a_k(h'), \forall k \leq |h| \right\}, \tag{3}$$

$$\hat{p}_{h,\pi}\left(\langle h', a, s\rangle\right) := \begin{cases} 1 & \text{if } \langle h', a, s\rangle = h, \\ \hat{p}_{h,\pi}(h') \cdot \pi(h', a) \cdot p(s_{|h'|}(h'), a, s) & \text{otherwise,} \end{cases} \tag{4}$$

for each $\langle h', a, s\rangle \in \Omega_{h,t}$. The random variables are defined as $\tilde{s}_k(h') := s_k(h')$, $\tilde{a}_k(h') := a_k(h')$, $\forall h' \in \Omega_{h,t}$. We interpret the subscripts analogously on all operators, including other risk measures, and $\mathbb{E}$, and $\mathbb{P}$.

Definition 2.4.2. The *history-dependent expectation* is defined for each $t \in \mathbb{N}$, $\pi \in \Pi_{\text{HR}}$, $\hat{h} \in \mathcal{H}$ and a $\tilde{x} \colon \mathcal{H} \to \mathbb{R}$ as

$$\mathbb{E}^{\hat{h},\pi,t}[\tilde{x}] := \mathbb{E}\left[\tilde{x}\right] = \sum_{h \in \mathcal{H}(\hat{h},t)} p^{\text{h}}(h,\pi) \cdot \tilde{x}(h).$$

In the $\mathbb{E}$ operator above, the random variable $\tilde{x}$ lives in a probability space (Ω, p) where $\Omega = \mathcal{H}(\hat{h}, t)$ and $p(h) = p^{\text{h}}(h,\pi), \forall h \in \Omega$. Moreover, if $\hat{h}$ is a state, then it is interpreted as a history with the single initial state.

Definition 2.4.3. The *history-dependent expectation* is defined for each $t \in \mathbb{N}$, $\pi \in \Pi_{\text{HR}}$, $\hat{h} \in \mathcal{H}$, $\tilde{x} \colon \mathcal{H} \to \mathbb{R}$, $\tilde{b} \colon \mathcal{H} \to \mathbb{B}$ as

$$\mathbb{E}^{\hat{h},\pi,t}[\tilde{x} \mid \tilde{b}] := \mathbb{E}\left[\tilde{x} \mid \tilde{b}\right].$$

In the $\mathbb{E}$ operator above, the random variables $\tilde{x}$ and $\tilde{b}$ live in a probability space (Ω, p) where $\Omega = \mathcal{H}(\hat{h}, t)$ and $p(h) = p^{\text{h}}(h,\pi), \forall h \in \Omega$. Moreover, if $\hat{h}$ is a state, then it is interpreted as a history with the single initial state.

Definition 2.4.4. The *history-dependent expectation* is defined for each $t \in \mathbb{N}$, $\pi \in \Pi_{\text{HR}}$, $\hat{h} \in \mathcal{H}$, $\tilde{x} \colon \mathcal{H} \to \mathbb{R}$, $\tilde{y} \colon \mathcal{H} \to \mathcal{V}$ as

$$\mathbb{E}^{\hat{h},\pi,t}[\tilde{x} \mid \tilde{y}](h) := \mathbb{E}\left[\tilde{x} \mid \tilde{y} = \tilde{y}(h)\right](h), \quad \forall h \in \mathcal{H}(\hat{h}, t).$$

In the $\mathbb{E}$ operator above, the random variables $\tilde{x}$ and $\tilde{h}$ live in a probability space (Ω, p) where $\Omega = \mathcal{H}(\hat{h}, t)$ and $p(h) = p^{\text{h}}(h,\pi), \forall h \in \Omega$. Moreover, if $\hat{h}$ is a state, then it is interpreted as a history with the single initial state.

2.5 Basic Properties

Theorem 2.5.1. *Assume $\tilde{x} \colon \mathcal{H} \to \mathbb{R}$ and $c \in \mathbb{R}$. Then $\forall h \in \mathcal{H}, \pi \in \Pi_{\text{HR}}, t \in \mathbb{N}$:*

$$\mathbb{E}^{\hat{h},\pi,t}\left[c + \tilde{x}\right] = c + \mathbb{E}^{\hat{h},\pi,t}\left[\tilde{x}\right].$$

Proof. Directly from Theorem 1.5.19. □

Theorem 2.5.2. *Suppose that $\tilde{x}, \tilde{y} \colon \mathcal{H} \to \mathbb{R}$. Then $\forall h \in \mathcal{H}, \pi \in \Pi_{\text{HR}}, t \in \mathbb{N}$:*

$$\mathbb{E}^{\hat{h},\pi,t}\left[\tilde{x} + \tilde{y}\right] = \mathbb{E}^{\hat{h},\pi,t}\left[\tilde{x}\right] + \mathbb{E}^{\hat{h},\pi,t}\left[\tilde{y}\right].$$

Proof. From Theorem 1.5.14. □

Theorem 2.5.3. *Suppose that $c \in \mathbb{R}$. Then $\forall h \in \mathcal{H}, \pi \in \Pi_{\text{HR}}, t \in \mathbb{N}$:*

$$\mathbb{E}^{\hat{h},\pi,t}\left[c\right] = c.$$

Proof. From Theorem 1.5.15. □

Theorem 2.5.4. *Suppose that $\tilde{x}, \tilde{y} \colon \mathcal{H} \to \mathbb{R}$ satisfy that $\tilde{x}(h) = \tilde{y}(h), \forall h \in \mathcal{H}$. Then $\forall h \in \mathcal{H}, \pi \in \Pi_{\text{HR}}, t \in \mathbb{N}$:*

$$\mathbb{E}^{\hat{h},\pi,t}\left[\tilde{x}\right] = c + \mathbb{E}^{\hat{h},\pi,t}\left[\tilde{y}\right].$$

Proof. From Theorem 1.5.16. □

Theorem 2.5.5. *For each $\hat{h} \in \mathcal{H}$, $\pi \in \Pi_{\text{HR}}$, and $t \in \mathbb{N}$:*

$$\mathbb{E}^{\hat{h},\pi,t}\left[\tilde{r}^{\text{h}}\right] = \mathbb{E}^{\hat{h},\pi,t}\left[\sum_{k=0}^{|\tilde{\text{id}}|-1} r(\tilde{s}_k, \tilde{a}_k, \tilde{s}_{k+1})\right],$$

where $\tilde{\text{id}}(h)$ is the identity function, $|\cdot|$ is the length of a history (0-based), $\tilde{s}_k \colon \mathcal{H} \to \mathcal{S}$ and $\tilde{a}_k \colon \mathcal{H} \to \mathcal{A}$ are the 0-based k-th state and action, respectively of each history.

Proof. Follows from Theorem 2.5.1 and the equality of the reward function $\tilde{r}^{\text{h}}$ and the sum in the expectation. □

Theorem 2.5.6. *For each $h \in \mathcal{H}$, $\pi \in \Pi_{\text{HR}}$, and $t \in \mathbb{N}$:*

$$\mathbb{E}^{h,\pi,t}\left[\tilde{r}^{\text{h}}\right] = \tilde{r}^{\text{h}}(h) + \mathbb{E}^{h,\pi,t}\left[\tilde{r}^{\text{h}}_{\geq k_0}\right],$$

where $k_0 := |h|$.

Proof. Follows from Theorem 2.5.4. □

Theorem 2.5.7. *For each $\hat{h} \in \mathcal{H}$, $\pi \in \Pi_{\text{HR}}$, $t \in \mathbb{N}$, $h \in \mathcal{H}$:*

$$\mathbb{P}^{\hat{h},\pi,t}[\tilde{s}_{k_0} = \tilde{s}_{k_0}(\omega) \wedge \tilde{a}_{k_0} = \tilde{a}_{k_0}(\omega)] > 0 \quad \Rightarrow \quad \mathbb{E}^{\hat{h},\pi,t}\left[\tilde{r}^{\text{h}}_{k_0} \mid \tilde{s}_{k_0}, \tilde{a}_{k_0}\right](h) = \tilde{r}^{\text{h}}_{k_0}(h), \forall h \in \mathcal{H}.$$

where $k_0 := |\hat{h}|$.

Proof. From Theorem 1.5.15. □

Theorem 2.5.8. *Assume $h \in \mathcal{H}$ and $f \colon \mathcal{H} \to \mathbb{R}$ such that $s_0 := s_{|h|}(h)$*

$$f(\langle h, a, s\rangle) = f(\langle s_0, a, a\rangle), \forall a \in \mathcal{A}, s \in \mathcal{S}.$$

Then

$$\mathbb{E}^{h,\pi,1}\left[\tilde{f}\right] = \mathbb{E}^{s_0,\pi,1}\left[\tilde{f}\right].$$

Proof. Directly from the definition of the expectation. □

2.6 Total Expectation

Theorem 2.6.1 (Total Expectation)**.** *For each $h \in \mathcal{H}$, $\pi \in \Pi_{\text{HR}}$, $t \in \mathbb{N}$, $\tilde{x} \colon \mathcal{H} \to \mathbb{R}$ and $\tilde{y} \colon \mathcal{H} \to \mathcal{V}$:*

$$\mathbb{E}^{h,\pi,t}\left[\mathbb{E}^{h,\pi,t}\left[\tilde{x} \mid \tilde{y}\right]\right] = \mathbb{E}^{h,\pi,t}\left[\tilde{x}\right].$$

Proof. From Theorem 1.7.2. □

Theorem 2.6.2. *Suppose that the random variable $\tilde{x} \colon \mathcal{H} \to \mathbb{R}$ satisfies for some $k, t \in \mathbb{N}$, with $k \leq t$, that*

$$\tilde{x}(h) = \tilde{x}(h_{\leq k}), \forall h \in \mathcal{H},$$

where $h_{\leq k}$ is the prefix of h of length k. Then for each $h \in \mathcal{H}, \pi \in \Pi_{\text{HR}}$:

$$\mathbb{E}^{h,\pi,t}\left[\tilde{x}\right] = \mathbb{E}^{h,\pi,k}\left[\tilde{x}\right].$$

2.7 Conditional Properties

Theorem 2.7.1. *For each $\beta > 0$, $h \in \mathcal{H}$, $\pi \in \Pi_{\text{HR}}$, $t \in \mathbb{N}$, $\tilde{x} \colon \mathcal{H} \to \mathbb{R}$, $s \in \mathcal{S}$, $a \in \mathcal{A}$:*

$$\mathbb{E}^{h,\pi,t+1}[\tilde{x} \mid \tilde{a}_{|h|} = a, \tilde{s}_{|h|+1} = s] = \mathbb{E}^{\langle h,a,s\rangle,\pi,t}[\tilde{x}],$$

Proof. Let

$$\begin{aligned} b &:= \mathbb{P}^{h,\pi,t+1}\left[\tilde{a}_{|h|} = a, \tilde{s}_{|h|+1} = s\right] \\ &= \hat{p}_{h,\pi}(h') \cdot \pi(h', a) \cdot p(s_{|h'|}(h'), a, s) \\ &> 0 \end{aligned}$$

where the inequality holds from the hypothesis. Also, let

$$\mathbb{B} := \left\{h' \in \Omega_{h,t+1} \mid a_{|h|}(h') = a \wedge s_{|h|+1}(h') = s\right\}.$$

Note that

$$\mathbb{B} = \Omega_{\langle h',a,s\rangle,t}, \tag{5}$$

which can be seen by algebraic manipulation from (3).

Using the notation above, we can show the result as

$$\begin{aligned}
\mathbb{E}^{h,\pi,t+1}[\widetilde{x} \mid \tilde{a}_{|h|} = a, \tilde{s}_{|h|+1} = s] &= \frac{1}{b} \sum_{h' \in \mathbb{B}} \widehat{p}_{h,\pi}(h') \cdot x(h') && \text{[definition]} \\
&= \frac{1}{b} \sum_{h' \in \Omega_{\langle h', a, s \rangle, t}} \widehat{p}_{h,\pi}(h') \cdot x(h') && \text{[Eq. (5)]} \\
&= \sum_{h' \in \Omega_{\langle h', a, s \rangle, t}} \widehat{p}_{\langle h,a,s \rangle,\pi}(h') \cdot x(h') && \text{[Eq. (4)]} \\
&= \mathbb{E}^{\langle h,a,s \rangle,\pi,t}[\widetilde{x}]. && \text{[definition]}
\end{aligned}$$

□

3 Dynamic Program: History-Dependent Finite Horizon

In this section, we derive dynamic programming equations for histories. We assume an MDP $M = (\mathcal{S}, \mathcal{A}, p, r)$ throughout this section.

The main idea of the proof is to:

1. Derive (exponential-size) dynamic programming equations for the history-dependent value function of history-dependent policies
   (a) Define the value function
   (b) Define an optimal value function
2. Show that value functions decompose to equivalence classes
3. Show that the value function for the equivalence classes can be computed efficiently

3.1 Definitions

Definition 3.1.1. A finite horizon objective definition is given by $O := (s_0, T)$ where $s_0 \in \mathcal{S}$ is the initial state and $T \in \mathbb{N}$ is the horizon.

In the reminder of the section, we assume an objective $O = (s_0, T)$.

Definition 3.1.2. The *finite horizon objective function* for and objective O is $\pi \in \Pi_{\text{HR}}$ is defined as

$$\rho(\pi, O) := \mathbb{E}^{s_0,\pi,T}\left[\widetilde{r}^{\text{h}}\right].$$

Definition 3.1.3. A policy $\pi^\star \in \Pi_{\text{HR}}$ is *return optimal* for an objective O if

$$\rho(\pi^\star, O) \geq \rho(\pi, O), \quad \forall \pi \in \Pi_{\text{HR}}.$$

Definition 3.1.4. The *set of history-dependent value functions* $\mathcal{U}$ is defined as

$$\mathcal{U} := \mathbb{R}^{\mathcal{H}}.$$

Definition 3.1.5. A *history-dependent policy value function* $\hat{u}_t^\pi \colon \mathcal{H} \to \mathbb{R}$ for each $h \in \mathcal{H}$, $\pi \in \Pi_{\mathrm{HR}}$, and $t \in \mathbb{N}$ is defined as

$$\hat{u}_t^\pi(h) := \mathbb{E}^{h,\pi,t}\left[\tilde{r}^{\mathrm{h}}_{\geq |h|}\right],$$

Definition 3.1.6. The optimal history-dependent value function $\hat{u}_t^\star \colon \mathcal{H} \to \mathbb{R}$ is defined for a horizon $t \in \mathbb{N}$ as

$$\hat{u}_t^\star(h) := \sup_{\pi \in \Pi_{\mathrm{HR}}} \hat{u}_t^\pi(h).$$

The following definition is another way of defining an optimal policy.

Definition 3.1.7. For each $t \in \mathbb{N}$, a policy $\pi^\star \in \Pi_{\mathrm{HR}}$ is optimal if

$$\hat{u}_t^{\pi^\star}(h) \geq \hat{u}_t^\pi(h), \quad \forall \pi \in \Pi_{\mathrm{HR}}, h \in \mathcal{H}.$$

Theorem 3.1.8. *A policy $\pi^\star \in \Pi_{\mathrm{HR}}$ optimal in Theorem 3.1.7 is also optimal in Theorem 3.1.7 for any initial state s_0 and horizon T.*

3.2 History-dependent Dynamic Program

The following definitions of history-dependent value functions use a dynamic program formulation.

Definition 3.2.1. The *history-dependent policy Bellman operator* $L_{\mathrm{h}}^\pi \colon \mathcal{U} \to \mathcal{U}$ is defined for each $\pi \in \Pi_{\mathrm{HR}}$ as

$$(L_{\mathrm{h}}^\pi \tilde{u})(h) := \mathbb{E}^{h,\pi,1}\left[\tilde{r}^{\mathrm{h}}_{|h|} + \tilde{u}\right], \quad \forall h \in \mathcal{H}, \tilde{u} \in \mathcal{U},$$

where the value function $\tilde{u}$ is interpreted as a random variable on defined on the sample space $\Omega = \mathcal{H}$.

Definition 3.2.2. The *history-dependent optimal Bellman operator* $L_{\mathrm{h}}^\star \colon \mathcal{U} \to \mathcal{U}$ is defined as

$$(L_{\mathrm{h}}^\star \tilde{u})(h) := \max_{a \in \mathcal{A}} \mathbb{E}^{h,a,1}\left[\tilde{r}^{\mathrm{h}}_{|h|} + \tilde{u}\right], \quad \forall h \in \mathcal{H}, \tilde{u} \in \mathcal{U},$$

where the value function $\tilde{u}$ is interpreted as a random variable on defined on the sample space $\Omega = \mathcal{H}$.

Definition 3.2.3. The history-dependent *DP value function* $u_t^\pi \in \mathcal{U}$ for a policy $\pi \in \Pi_{\mathrm{HR}}$ and $t \in \mathbb{N}$ is defined as

$$u_t^\pi := \begin{cases} 0 & \text{if } t = 0, \\ L_{\mathrm{h}}^\pi u_{t-1}^\pi & \text{otherwise.} \end{cases}$$

Definition 3.2.4. The history-dependent *DP value function* $u_t^\star \in \mathcal{U}$ for $t \in \mathbb{N}$ is defined as

$$u_t^\star := \begin{cases} 0 & \text{if } t = 0, \\ L_{\mathrm{h}}^\star u_{t-1}^\star & \text{otherwise.} \end{cases}$$

Lemma 3.2.5. *Suppose that* $u^1, u^2 \in \mathcal{U}$ *satisfy that* $u^1(h) \geq u^2(h), \forall h \in \mathcal{H}$. *Then*

$$(L_{\mathrm{h}}^\star u^1)(h) \geq (L_{\mathrm{h}}^\pi u^2)(h), \quad \forall \pi \in \Pi_{\mathrm{HR}}, h \in \mathcal{H}.$$

Proof. From Theorem 1.5.18. □

The following theorem shows the history-dependent value function can be computed by the dynamic program. The following theorem is akin to [**?**, theorem 4.2.1].

Theorem 3.2.6. *For each* $\pi \in \Pi_{\mathrm{HR}}$ *and* $t \in \mathbb{N}$*:*

$$\hat{u}_t^\pi(h) = u_t^\pi(h), \quad \forall h \in \mathcal{H}.$$

Proof. By induction on t. The base case for $t = 0$ follows from the definition. The inductive case for $t + 1$ follows for each $h \in \mathcal{H}$ when $|h| = k_0$ as

$$\begin{aligned}
\hat{u}_{t+1}^\pi(h) &= \mathbb{E}^{h,\pi,t+1}\left[\tilde{r}_{\geq k_0}^{\mathrm{h}}\right] && \text{[Theorem 3.1.5]} \\
&= \mathbb{E}^{h,\pi,t+1}\left[\mathbb{E}^{h,\pi,t+1}\left[\tilde{r}_{\geq k_0}^{\mathrm{h}} \mid \tilde{a}_{k_0}, \tilde{s}_{k_0+1}\right]\right] && \text{[Theorem 2.6.1]} \\
&= \mathbb{E}^{h,\pi,t+1}\left[\tilde{r}_{k_0}^{\mathrm{h}} + \mathbb{E}^{h,\pi,t+1}\left[\tilde{r}_{\geq k_0+1}^{\mathrm{h}} \mid \tilde{a}_{k_0}, \tilde{s}_{k_0+1}\right]\right] && \text{[Theorem 2.5.7]} \\
&= \mathbb{E}^{h,\pi,t+1}\left[\tilde{r}_{k_0}^{\mathrm{h}} + \mathbb{E}^{\langle h, \tilde{a}_{k_0}, \tilde{s}_{k_0+1}\rangle,\pi,t}\left[\tilde{r}_{\geq k_0+1}^{\mathrm{h}}\right]\right] && \text{[Theorem 2.7.1]} \\
&= \mathbb{E}^{h,\pi,t+1}\left[\tilde{r}_{k_0}^{\mathrm{h}} + \hat{u}_t(\langle h, \tilde{a}_{k_0}, \tilde{s}_{k_0+1}\rangle; \pi)\right] && \text{[Theorem 3.1.5]} \\
&= \mathbb{E}^{h,\pi,t+1}\left[\tilde{r}_{k_0}^{\mathrm{h}} + u_t^\pi(\langle h, \tilde{a}_{k_0}, \tilde{s}_{k_0+1}\rangle)\right] && \text{[inductive assm]} \\
&= \mathbb{E}^{h,\pi,1}\left[\tilde{r}^{\mathrm{h}} + \tilde{u}_t^\pi\right] && \text{[Theorem 2.6.2]} \\
&= L_{\mathrm{h}}^\pi u_t^\pi && \text{[Theorem 3.2.1]} \\
&= u_t^\pi(h). && \text{[Theorem 3.2.3]}
\end{aligned}$$

Also, we use $\tilde{u}_t^\pi$ to emphasize when we treat u_t^π as a random variable. □

The following theorem is akin to [**?**, theorem 4.3.2].

Theorem 3.2.7. *For each* $t \in \mathbb{N}$*:*

$$u_t^\star(h) \geq \hat{u}_t(h; \pi), \quad \forall h \in \mathcal{H}, \pi \in \Pi_{\mathrm{HR}}.$$

Proof. By induction on t. The base case is immediate. The inductive case follows for $t + 1$ as follows. For

each $\pi \in \Pi_{\mathrm{HR}}$:

$$\begin{aligned}
u^\star_{t+1}(h) &= (L^\star_{\mathrm{h}} u^\star_t)(h) && \text{[Theorem 3.2.2]}\\
&\geq (L^\pi_{\mathrm{h}} \widehat{u}_t(\cdot;\pi))(h) && \text{[ind asm, Theorem 3.2.5]}\\
&= (L^\pi_{\mathrm{h}} u^\pi_t)(h) && \text{[Theorem 3.2.6]}\\
&= u^\pi_t(h) && \text{[Theorem 3.1.5]}\\
&= \widehat{u}_t(h;\pi). && \text{[Theorem 3.2.6]}
\end{aligned}$$

□

4 Expected Dynamic Program: Markov Policy

4.1 Optimality

We discuss results needed to prove the optimality of Markov policies.

Definition 4.1.1. The set of *independent value functions* is defined as $\mathcal{V} := \mathbb{R}^{\mathcal{S}}$.

Definition 4.1.2. A *Markov Bellman operator* $L^\star \colon \mathcal{V} \to \mathcal{V}$ is defined as

$$(L^\star v)(h) := \max_{a \in \mathcal{A}} \mathbb{E}^{h,a,1}\left[\tilde{r}^{\mathrm{h}} + v(\tilde{s}_{|h|})\right], \quad \forall \tilde{u} \in \mathcal{U},$$

Definition 4.1.3. The *optimal value function* $v^\star_t \in \mathcal{V}, t \in \mathbb{N}$ is defined as

$$v^\star_t := \begin{cases} 0 & \text{if } t = 0 \\ (L^\star v^\star_{t-1}) & \text{otherwise.} \end{cases}$$

Theorem 4.1.4. *Suppose that $t \in \mathbb{N}$. Then:*

$$v^\star_t(s_{|h|}(h)) = u^\star_t(h), \quad \forall h \in \mathcal{H}.$$

Proof. By induction on t. The base case follows immediately from the definition. The inductive step for $t+1$ follows as:

$$\begin{aligned}
u^\star_{t+1}(h) &= \max_{a \in \mathcal{A}} \mathbb{E}^{h,a,1}\left[\tilde{r}^{\mathrm{h}}_{|h|} + \tilde{u}^\star_t\right] && \text{[Theorem 3.2.4]}\\
&= \max_{a \in \mathcal{A}} \mathbb{E}^{h,a,1}\left[\tilde{r}^{\mathrm{h}}_{|h|} + v^\star_t(\tilde{s}_l)\right] && \text{[inductive asm.]}\\
&= \max_{a \in \mathcal{A}} \mathbb{E}^{s_0,a,1}\left[\tilde{r}^{\mathrm{h}}_{|h|} + v^\star_t(\tilde{s}_l)\right] && \text{[Theorem 2.5.8]}\\
&= \max_{a \in \mathcal{A}} \mathbb{E}^{s_0,a,1}\left[\tilde{r}^{\mathrm{h}} + v^\star_t(\tilde{s}_l)\right] && \text{[Theorem 2.5.1]}\\
&= v^\star_{t+1}(s_{|h|}(h)) && \text{[Theorem 4.1.3].}
\end{aligned}$$

□

Definition 4.1.5. The *optimal finite-horizon policy* $\pi_t^\star, t \in \mathbb{N}$ is defined as

$$\pi_t^\star(k,s) := \begin{cases} \arg\max_{a\in\mathcal{A}} \mathbb{E}^{s,a,1}\left[\tilde{r}^{\mathrm{h}} + v_{t-k}^\star(\tilde{s}_{|h|})\right] & \text{if } k \le t, \\ a_0 & \text{otherwise,} \end{cases}$$

where a_0 is an arbitrary action.

Theorem 4.1.6. *Assume a horizon $T \in \mathbb{N}$. Then:*

$$v_{T-|h|}^\star(s_{|h|}(h)) = u_{T-|h|}^{\pi^\star_{T-h}}(h), \quad \forall h \in \{h \in \mathcal{H} \mid |h| \le T\}.$$

Proof. Fix some $T \in \mathbb{N}$. By induction on k from $k = T$ to $k = 0$. The base case is immediate from the definition. We prove the inductive case for $k-1$ from k as

$$\begin{aligned}
u_{T-k+1}^{\pi^\star_T}(h) &= \mathbb{E}^{h,\pi^\star_T,1}\left[\tilde{r}_k^{\mathrm{h}} + \tilde{u}_{T-k}^{\pi^\star_T}\right] && \text{[Theorem 3.2.1]} \\
&= \mathbb{E}^{h,a^\star,1}\left[\tilde{r}_k^{\mathrm{h}} + \tilde{u}_{T-k}^{\pi^\star_T}\right] && \text{[???]} \\
&= \mathbb{E}^{h,a^\star,1}\left[\tilde{r}_k^{\mathrm{h}} + v_{T-k}^\star(\tilde{s}_1)\right] && \text{[ind asm]} \\
&= \mathbb{E}^{s_0,a^\star,1}\left[\tilde{r}^{\mathrm{h}} + v_{T-k}^\star(\tilde{s}_1)\right] && \text{[Theorem 2.5.8]} \\
&= \max_{a\in\mathcal{A}} \mathbb{E}^{s_0,a,1}\left[\tilde{r}^{\mathrm{h}} + v_{T-k}^\star(\tilde{s}_1)\right] && \text{[???]} \\
&= v_{T-k+1}^\star(s_0). && \text{[Theorem 4.1.3]}
\end{aligned}$$

Here, $k := |h|$, $a^\star := \pi_T^\star(k, s_0)$, and $s_0 := s_{|h|}(h)$ □

4.2 Evaluation

We discuss results pertinent to the evaluation of Markov policies.

Markov value functions depend on the length of the history.

Definition 4.2.1. The set of *independent value functions* is defined as $\mathcal{V}_{\mathrm{M}} := \mathbb{R}^{\mathbb{N}\times\mathcal{S}}$.

Definition 4.2.2. A *Markov policy Bellman operator* $L_k^\pi \colon \mathcal{V} \to \mathcal{V}$ for $\pi \in \Pi$ is defined as

$$(L^\pi v)(k,s) := \max_{a\in\mathcal{A}} \mathbb{E}^{s,a,1}\left[\tilde{r}^{\mathrm{h}} + v(k+1, \tilde{s}_{|h|})\right], \quad \forall v \in \mathcal{V}_{\mathrm{M}}, k \in \mathbb{N}, s \in \mathcal{S}.$$

Definition 4.2.3. The *Markov policy value function* $v_t^\pi \in \mathcal{V}_{\mathrm{M}}, t \in \mathbb{N}$ for $\pi \in \Pi_{\mathrm{MD}}$ is defined as

$$v_t^\pi := \begin{cases} 0 & \text{if } t = 0, \\ (L^\pi v_{t-1}^\pi) & \text{otherwise.} \end{cases}$$

5 Probability Maticies

Definition 5.0.1. A probability matrix $\mathbf{P}$ is defined as an $n \times n$ matrix with the properties:

$$\mathbf{P}_{ij} \geq 0 \quad \forall i, j \in \{1, 2, \dots, n\}$$
$$\mathbf{P1} = \mathbf{1}$$

Theorem 5.0.2. *Let* $\mathbf{d} \in \Delta^n$. *Then*

$$\mathbf{d}^T\mathbf{P} \in \Delta^n$$

Proof. Let $i \in \{1, 2, \dots, n\}$ and $\mathbf{p} = \mathbf{d}^T\mathbf{P}$. Then $p_i = \mathbf{d}^T\mathbf{P}_i$ where $\mathbf{P}_i$ is the i^{th} column of $\mathbf{P}$. Since all the elements of both $\mathbf{P}_i$ and $\mathbf{d}$ are non-negative, their inner product must also be non-negative. Since

$$\begin{aligned}\mathbf{d}^T\mathbf{P1} &= \mathbf{d}^T\mathbf{1} \\ &= 1 \ ,\end{aligned}$$

$\mathbf{p}$ satisfies that its elements are non-negative and sum to 1 making it a member of Δ^n □