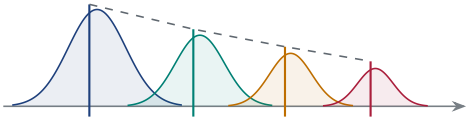
Part II

The Distributional Viewpoint and Risk

4

Beyond the Mean: Bellman-Closed Statistics



The value function propagates one number per state by an exact recursion; the return distribution propagates a whole measure, exactly too, but at a cost that explodes unless the rewards live on a lattice. Between the two lie the statistics *of the return: its variance, its higher moments, its exponential moments, its extreme values, its quantiles. Which of them admit an exact backward recursion of their own? This chapter answers the question completely: a single Bellman-closed utility $\int f \,\mathrm{d}\nu$ is, up to an affine reparametrization, the mean or an exponential moment $\mathbb{E}e^{\beta W}$ —equivalently, a β-mean of Chapter 2 (Theorem 4.8)—and the* families *that close are spanned by exponential polynomials, the moments and the exponential moments being the two examples that matter. The chapter ends with the converse problem: how much of the distribution can be recovered from finitely many exact statistics, and how this compares with propagating approximate distributions.*

4.1 Propagating moments

Throughout the chapter, $\mathcal{M} = (\mathcal{S}, \mathcal{A}, (P_t), (r_t), H)$ is a finite-horizon MDP, $\pi \in \Pi_{\mathrm{MR}}$ is a Markov policy, and $v_t^\pi(s)$, $v_t^\pi(s,a)$ are the return distributions of Definition 3.1. We write $V_t^\pi(s) = \int x\, v_t^\pi(s)(\,\mathrm{d}x)$ for the mean, as in Chapter 3, and $M_t^{(i),\pi}(s) \coloneqq \int x^i\, v_t^\pi(s)(\,\mathrm{d}x)$ for the moment of order $i \geq 0$, with $M^{(0)} \equiv 1$ and $M^{(1)} = V$. The superscript π is dropped when no confusion is possible. More generally, for a law ν we write $M_i(\nu) \coloneqq \int x^i \,\mathrm{d}\nu$ for its raw moment of order i and $\mathbb{V}\mathrm{ar}[\nu] \coloneqq M_2(\nu) - M_1(\nu)^2$ for its variance, so that $M_t^{(i),\pi}(s) = M_i(v_t^\pi(s))$: a functional of laws induces a value function.

The distributional Bellman equation (Theorem 3.2) writes $v_t^\pi(s)$ as a mixture of shifted copies of the $v_{t+1}^\pi(s')$. Both operations act simply on moments: a shift by r transforms the moments by the binomial formula, and a mixture averages them.

Proposition 4.1 (Moment recursions). For every $i \geq 0$, $t \in [H]$, $s \in \mathcal{S}$ and $a \in \mathcal{A}$,

$$\begin{aligned} M_t^{(i)}(s,a) &= \sum_{s' \in \mathcal{S}} P_t(s' \mid s,a) \sum_{j=0}^{i} \binom{i}{j} r_t(s,a,s')^{\,i-j}\, M_{t+1}^{(j)}(s'), \\ M_t^{(i)}(s) &= \sum_{a \in \mathcal{A}} \pi_t(a \mid s)\, M_t^{(i)}(s,a), \end{aligned} \tag{4.1}$$

Algorithm 4.1: Backward induction for the first K moments of the return

Input: MDP $\mathcal{M}$, Markov policy π, order K

$M_{H+1}^{(0)}(s) \leftarrow 1$, $M_{H+1}^{(i)}(s) \leftarrow 0$ for $1 \le i \le K$ and all s

for $t = H, H-1, \dots, 1$ **do**

  for $s \in \mathcal{S}$, $i = 0, \dots, K$ **do**

    $M_t^{(i)}(s) \leftarrow \sum_a \pi_t(a \mid s) \sum_{s'} P_t(s' \mid s, a) \sum_{j=0}^{i} \binom{i}{j} r_t(s,a,s')^{i-j} M_{t+1}^{(j)}(s')$

Output: $(M_t^{(i)})_{t\in[H],\, i\le K}$; in particular $V_t^\pi = M_t^{(1)}$ and $\mathbb{V}\mathrm{ar}[v_t^\pi(s)] = M_t^{(2)}(s) - V_t^\pi(s)^2$

with $M_{H+1}^{(i)} \equiv 0$ for $i \ge 1$. In particular, for the second moment,

$$M_t^{(2)}(s) \;=\; \sum_{a,s'} \pi_t(a \mid s) P_t(s' \mid s, a) \Big[r_t(s,a,s')^2 + 2\, r_t(s,a,s')\, V_{t+1}(s') + M_{t+1}^{(2)}(s') \Big].$$

Proof. If $X \sim \nu$ then $\int x^i \, (b_r)_\# \nu(\mathrm{d}x) = \mathbb{E}[(r+X)^i] = \sum_j \binom{i}{j} r^{i-j} \mathbb{E}[X^j]$, and the i-th moment of a mixture $\sum_k p_k \nu_k$ is $\sum_k p_k \int x^i \, \mathrm{d}\nu_k$. Apply both facts to (3.1)–(3.2). ∎

The recursion (4.1) is *triangular*: the moment of order i at time t involves only the moments of order $j \le i$ at time $t+1$. The first K moments can thus be propagated together, exactly, without any knowledge of the rest of the distribution (Algorithm 4.1). The cost is $O(HS^2AK^2)$ arithmetic operations, against $O(H^2S^2AR_{\max})$ for the exact distributional recursion with integer rewards (Proposition 3.7), and against an unbounded cost with general rewards (Remark 3.9).

The variance $\sigma_t^2(s) \coloneqq M_t^{(2)}(s) - V_t(s)^2$ of the return distribution deserves a formula of its own, which separates the two sources of dispersion.

Corollary 4.2 (Variance of the return). For every $t \in [H]$ and $s \in \mathcal{S}$, with $\sigma_{H+1}^2 \equiv 0$,

$$\begin{aligned} \sigma_t^2(s) \;=\; & \underbrace{\sum_{a,s'} \pi_t(a \mid s) P_t(s' \mid s, a)\, \sigma_{t+1}^2(s')}_{\text{inherited from the future}} \\ & + \underbrace{\sum_{a,s'} \pi_t(a \mid s) P_t(s' \mid s, a) \Big[r_t(s,a,s') + V_{t+1}(s') - V_t(s) \Big]^2}_{\text{created at time } t}. \end{aligned} \tag{4.2}$$

Proof. Under $\mathbb{P}_{t,s}^\pi$, write $W_{t:H} = R + W'$ with $R = r_t(s, A_t, S_{t+1})$ and $W' = W_{t+1:H}$. Given (A_t, S_{t+1}), R is fixed and W' has law $\nu_{t+1}(S_{t+1})$ by the restart property (Proposition 1.15), so $\mathbb{E}[W_{t:H} \mid A_t, S_{t+1}] = R + V_{t+1}(S_{t+1})$ and $\mathbb{V}\mathrm{ar}[W_{t:H} \mid A_t, S_{t+1}] = \sigma_{t+1}^2(S_{t+1})$. The law of total variance (Proposition A.2(ii)), $\mathbb{V}\mathrm{ar}[W] = \mathbb{E}[\mathbb{V}\mathrm{ar}[W \mid A_t, S_{t+1}]] + \mathbb{V}\mathrm{ar}[\mathbb{E}[W \mid A_t, S_{t+1}]]$, gives the two terms, since $\mathbb{E}[R + V_{t+1}(S_{t+1})] = V_t(s)$ by the Bellman equation (Corollary 3.11). ∎

The second term of (4.2) is the variance of the *one-step target* $r_t(s, A_t, S_{t+1}) + V_{t+1}(S_{t+1})$, the quantity that temporal-difference methods estimate (Chapter 12): it is the dispersion created by the randomness of the action and of the transition at time t, while the first term is the dispersion inherited from later times. Note that the recursion for σ_t^2 is *not* autonomous: it requires the value functions V_{t+1} and V_t. This is the first sign of a general phenomenon studied in the next section.

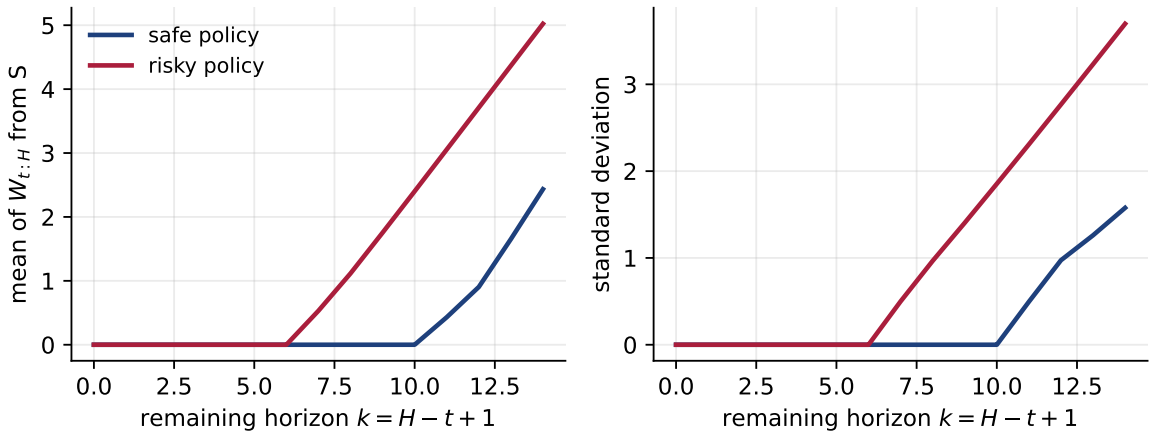


Figure 4.1. Mean (left) and standard deviation (right) of the reward-to-go from the starting cell of the windy hike, as functions of the remaining horizon $k = H - t + 1$, for the safe and risky policies of Example 3.6; computed by Algorithm 4.1.

Example 4.3 (Mean and dispersion of the hikers' returns). Figure 4.1 shows, for the two policies of Example 3.6, the mean and the standard deviation of the reward-to-go $W_{t:H}$ from the starting cell as functions of the remaining horizon $k = H - t + 1$, computed by Algorithm 4.1 with $K = 2$ (the numbers agree with the exact distributions of Figure 3.2 to machine precision). Nothing happens before the hiker can possibly reach the bar ($k \geq 7$ for the risky path, $k \geq 11$ for the safe one); then the mean grows almost linearly with the remaining horizon, with slope $1 - \mathbb{P}(\text{pit})$ for the risky hiker. For $k = H = 14$ one recovers the means 5.02 and 2.43 and the standard deviations 3.70 and 1.58 of Example 3.6. The standard deviation of the risky return grows linearly too: this return is essentially $(k - 6)$ times a Bernoulli variable of parameter $1 - \mathbb{P}(\text{pit}) \approx 0.66$—arrive and enjoy the bar, or fall—whose standard deviation 0.47 $(k - 6)$ matches the curve; in the recursion, this dispersion is the "created" term of (4.2) at the cells bordering the pit, where a gust changes the return from about $k - 6$ to 0. The safe return is less dispersed mostly because it is smaller: its coefficient of variation, 0.65, is only slightly below that of the risky return, 0.74.

4.2 Which statistics can be propagated?

The moment recursions rest on a structural property: the moments of a mixture of shifted laws are functions of the moments of the laws being shifted and mixed. We now isolate this property and ask which other statistics enjoy it. Throughout, laws are assumed to have *bounded support*—as do the return distributions of a finite-horizon MDP with finite state and action spaces (Remark 1.18)—so that every continuous function is integrable against them. A *statistic* is a map $\psi : \mathcal{M}_1(\mathbb{R}) \to \mathbb{R}^m$ defined on such laws (for instance $\psi(\nu) = (\int x \, \mathrm{d}\nu, \int x^2 \, \mathrm{d}\nu)$); m is its *dimension*. The word *functional* is used as a synonym; a real-valued statistic is called a *criterion* in Chapter 5, where it serves to compare policies, and, when it is defined on random variables rather than on their laws, a *risk measure* in Chapter 6. For weights $p = (p_1, \ldots, p_n)$ in the simplex and shifts $r = (r_1, \ldots, r_n) \in \mathbb{R}^n$, we call

$$\mathcal{T}_{p,r}(\nu_1, \ldots, \nu_n) \coloneqq \sum_{i=1}^{n} p_i \, (b_{r_i})_{\#} \nu_i$$

a *mixture of shifts*; every $(\mathcal{T}_t^\pi \nu)(s)$ is of this form (Proposition 3.4), with the *SA* pairs (a, s') as indices, weights $\pi_t(a \mid s)P_t(s' \mid s, a)$ and shifts $r_t(s, a, s')$.

Definition 4.4 (Bellman-closed statistic). A statistic $\psi : \mathcal{M}_1(\mathbb{R}) \to \mathbb{R}^m$ is *Bellman closed* if for every $n \geq 1$, every p in the simplex of $\mathbb{R}^n$ and every $r \in \mathbb{R}^n$ there exists a map $\Phi_{p,r} : (\mathbb{R}^m)^n \to \mathbb{R}^m$ such that

$$\psi\big(\mathcal{T}_{p,r}(\nu_1, \dots, \nu_n)\big) \;=\; \Phi_{p,r}\big(\psi(\nu_1), \dots, \psi(\nu_n)\big) \qquad \text{for all laws } \nu_1, \dots, \nu_n \text{ with bounded support.}$$

In words, the statistic of a mixture of shifts is determined by the statistics of the components. A Bellman-closed statistic satisfies its own Bellman equation in every MDP and for every Markov policy,

$$\begin{aligned} \psi\big(v_t^\pi(s)\big) \;&=\; \Phi_{p_t(s), r_t(s)}\Big(\big(\psi(v_{t+1}^\pi(s'))\big)_{(a,s')}\Big), \\ p_t(s) = \big(\pi_t(a \mid s)P_t(s' \mid s, a)\big)_{a,s'}, &\qquad r_t(s) = \big(r_t(s, a, s')\big)_{a,s'}, \end{aligned} \tag{4.3}$$

which can be run backward from $\psi(\delta_0)$ exactly as Algorithm 4.1 does for moments: the collection $(\psi(v_t^\pi(s)))_s$ is a "value function" with values in $\mathbb{R}^m$, and the map $\mathcal{T}_{\psi,t}^\pi : (\psi(v(s')))_{s'} \mapsto (\psi((\mathcal{T}_t^\pi v)(s)))_s$ that (4.3) defines on $(\mathbb{R}^m)^{\mathcal{S}}$ is the Bellman operator induced on the statistic (so that $M_t^{(i)}(s) = M_i(v_t^\pi(s))$, in the notation of Section 4.1). Closedness in the sense of Definition 4.4 thus implies the existence of such a recursion in every MDP, which is the original formulation of Rowland et al. (2019). Conversely, since any finitely supported law is the return distribution of some MDP (a tree whose leaves carry the atoms), a recursion valid in every MDP forces closedness on finitely supported laws—which is all that the results below use.

Two elementary observations. First, if ψ and ψ' are Bellman closed, so is the pair (ψ, ψ'), and so is $g \circ \psi$ for any *bijection* g of $\mathbb{R}^m$: the property is about the information carried by ψ, not about its parametrization ($(M^{(1)}, M^{(2)})$ and (V, σ^2) are the same statistic). Second, closedness of ψ does *not* pass to its components: σ^2 can be computed from (V, σ^2), but it is not itself Bellman closed, as we shall see. The following criterion separates the two operations of the distributional Bellman operator.

Proposition 4.5 (Closedness = translation compatibility + mixture compatibility). A statistic ψ is Bellman closed if and only if it satisfies the two following properties, for all laws ν, ν', ν_i, ν_i' with bounded support:

(i) *translation compatibility:* $\psi(\nu) = \psi(\nu')$ implies $\psi((b_r)_\# \nu) = \psi((b_r)_\# \nu')$ for every $r \in \mathbb{R}$;

(ii) *mixture compatibility:* $\psi(\nu_i) = \psi(\nu_i')$ for all $i \in [n]$ implies $\psi(\sum_i p_i \nu_i) = \psi(\sum_i p_i \nu_i')$ for all weights p.

Proof. If ψ is closed, (i) is Definition 4.4 with $n = 1$: $\psi((b_r)_\# \nu) = \Phi_{1,r}(\psi(\nu))$ depends on ν only through $\psi(\nu)$; and (ii) is the definition with all shifts equal to 0. Conversely, assume (i) and (ii) and let $y_1, \dots, y_n$ lie in the range of ψ. Define $\Phi_{p,r}(y_1, \dots, y_n) \coloneqq \psi(\sum_i p_i (b_{r_i})_\# \nu_i)$ for any laws with $\psi(\nu_i) = y_i$; this does not depend on the choice of the ν_i, since if $\psi(\nu_i') = y_i$ as well, (i) gives $\psi((b_{r_i})_\# \nu_i) = \psi((b_{r_i})_\# \nu_i')$ for every i, and then (ii) gives $\psi(\sum_i p_i (b_{r_i})_\# \nu_i) = \psi(\sum_i p_i (b_{r_i})_\# \nu_i')$. Outside the range of ψ, define $\Phi_{p,r}$ arbitrarily. ∎

Example 4.6 (Closed and non-closed statistics).

(i) The mean is closed, and so is the vector of the first K moments (Proposition 4.1); the pair (mean, variance) is closed, with $\mathbb{V}\mathrm{ar}[\sum_i p_i \nu_i] = \sum_i p_i \sigma_i^2 + \sum_i p_i (m_i - \bar{m})^2$, $\bar{m} = \sum_i p_i m_i$, and $\mathbb{V}\mathrm{ar}[(b_r)_\# \nu] = \mathbb{V}\mathrm{ar}[\nu]$.

(ii) The variance alone is translation compatible but *not* mixture compatible: δ_0 and δ_1 both have variance 0, yet $\frac{1}{2}(\delta_0 + \delta_0) = \delta_0$ has variance 0 and $\frac{1}{2}(\delta_1 + \delta_0)$ has variance $\frac{1}{4}$. Mixing creates variance out of differences between means, which the variance alone does not record.

(iii) The median $\nu \mapsto F_\nu^{-1}(\frac{1}{2})$ (lower quantile, $F_\nu^{-1}(u) = \inf\{x : F_\nu(x) \geq u\}$ as in Appendix D), and more generally any quantile, is not mixture compatible: δ_0 and $\nu' = 0.6\,\delta_0 + 0.4\,\delta_5$ both have median 0, but $\frac{1}{2}(\delta_0 + \delta_1)$ has median 0 while $\frac{1}{2}(\nu' + \delta_1) = 0.3\,\delta_0 + 0.5\,\delta_1 + 0.2\,\delta_5$ has median 1. The exact backward computation of quantiles is therefore impossible without the whole distribution (Exercise 4.3 extends this to all finite families of quantiles). This counterexample and the next are those of Bellemare et al. (2023, Examples 8.10–8.11), with different numbers.

(iv) The probability of exceeding a fixed threshold, $\nu \mapsto \nu([u, \infty))$, is mixture compatible (it is linear in ν) but not translation compatible: δ_u and δ_{u+1} both give probability 1, and shifting by -1 gives 0 and 1. Shifting moves the mass across the threshold in a way the statistic cannot anticipate.

(v) On finitely supported laws, the Shannon entropy is translation compatible but not mixture compatible: δ_0 and δ_1 have entropy 0, while $\frac{1}{2}(\delta_0 + \delta_0)$ and $\frac{1}{2}(\delta_0 + \delta_1)$ have entropies 0 and $\ln 2$.

(vi) The essential infimum and supremum are closed (Section 4.5), and so is the identity map $\nu \mapsto \nu$ (of infinite dimension): the exact distributional recursion of Algorithm 3.1 is the closed recursion of the statistic "the whole law". With integer rewards in $\{0, \dots, R_{\max}\}$, the probabilities of the $HR_{\max} + 1$ lattice points form, on the laws supported by the lattice and for integer shifts, a closed statistic of finite dimension.

4.3 Bellman-closed utilities

A *utility* is a statistic of the form $\psi_f(\nu) = \int f \, \mathrm{d}\nu$ for a continuous $f : \mathbb{R} \to \mathbb{R}$—$f$ is the *utility function* and ψ_f the (*expected*) utility of decision theory; continuity of f is assumed in this chapter only—or a vector of such, $\psi_{\mathbf{f}} = (\psi_{f_1}, \dots, \psi_{f_K})$ for $\mathbf{f} = (f_1, \dots, f_K)$. Being linear in ν, utilities are automatically mixture compatible; by Proposition 4.5, $\psi_{\mathbf{f}}$ is Bellman closed if and only if it is translation compatible. The next lemma turns this condition into a functional equation.

Lemma 4.7 (Translation compatibility of utilities). Let $f_1, \dots, f_K$ be continuous and $\mathbf{f} = (f_1, \dots, f_K) : \mathbb{R} \to \mathbb{R}^K$. The utility $\psi_{\mathbf{f}}$ is translation compatible if and only if for every $a \in \mathbb{R}$ there exist a matrix $A(a) \in \mathbb{R}^{K \times K}$ and a vector $\gamma(a) \in \mathbb{R}^K$ such that

$$\mathbf{f}(x + a) \;=\; A(a)\,\mathbf{f}(x) + \gamma(a) \qquad \text{for all } x \in \mathbb{R}, \tag{4.4}$$

that is, if and only if the linear span $V_{\mathbf{f}}$ of $\{1, f_1, \dots, f_K\}$ is invariant under translations $f \mapsto f(\cdot + a)$.

Proof. Suppose (4.4) holds. Integrating against ν, $\psi_{\mathbf{f}}((b_a)_\# \nu) = \int \mathbf{f}(x + a)\, \nu(\mathrm{d}x) = A(a)\psi_{\mathbf{f}}(\nu) + \gamma(a)$ is a function of $\psi_{\mathbf{f}}(\nu)$: translation compatibility holds. Conversely, let E be the vector space of finitely supported signed measures on $\mathbb{R}$, and consider the linear maps $\Lambda_a : E \to \mathbb{R}^K$, $\mu \mapsto \int \mathbf{f}(x + a)\, \mu(\mathrm{d}x)$, and $\widetilde{\Lambda}_0 : E \to \mathbb{R}^{K+1}$, $\mu \mapsto (\mu(\mathbb{R}), \Lambda_0 \mu)$. We claim that $\ker \widetilde{\Lambda}_0 \subset \ker \Lambda_a$. Indeed, let $\mu \in \ker \widetilde{\Lambda}_0$, $\mu \neq 0$; its positive and negative parts μ_+ and μ_- have the same total mass $c > 0$, since $\mu(\mathbb{R}) = 0$, so $\mu = c(\nu - \nu')$ with $\nu = \mu_+/c$, $\nu' = \mu_-/c$ two finitely supported probability measures; $\Lambda_0 \mu = 0$ says $\psi_{\mathbf{f}}(\nu) = \psi_{\mathbf{f}}(\nu')$,

hence $\psi_{\mathbf{f}}((b_a)_\# \nu) = \psi_{\mathbf{f}}((b_a)_\# \nu')$ by translation compatibility, which is $\Lambda_a \mu = 0$. The inclusion of kernels implies that Λ_a factors through $\widetilde{\Lambda}_0$: the map $B_a : \widetilde{\Lambda}_0 \mu \mapsto \Lambda_a \mu$ is well defined and linear on the range of $\widetilde{\Lambda}_0$, and extends to a linear map $\mathbb{R}^{K+1} \to \mathbb{R}^K$, $(c, y) \mapsto c\,\gamma(a) + A(a)y$. Applying $\Lambda_a = B_a \circ \widetilde{\Lambda}_0$ to $\mu = \delta_x$ gives (4.4). The last statement rephrases (4.4): each $f_i(\cdot + a)$ lies in $V_{\mathbf{f}}$, and constants are translation invariant. ∎

One utility

For a single function ($K = 1$), the functional equation (4.4) can be solved completely.

Theorem 4.8 (Bellman-closed utilities). Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and nonconstant. The utility ψ_f is Bellman closed if and only if either $f(x) = c_1 x + c_0$ or $f(x) = c_1 e^{\beta x} + c_0$ for some constants $c_1 \neq 0$, $c_0 \in \mathbb{R}$ and $\beta \neq 0$. Up to an affine reparametrization, the Bellman-closed utilities are thus the mean and the exponential moments $\nu \mapsto \int e^{\beta x}\, \nu(\mathrm{d}x)$, $\beta \neq 0$.

Proof. The mean is closed by Proposition 4.1, and $\int e^{\beta(x+a)}\, \mathrm{d}\nu = e^{\beta a} \int e^{\beta x}\, \mathrm{d}\nu$ shows that exponential moments are closed; an affine reparametrization does not affect closedness. Conversely, by Lemma 4.7, for every a there are $\alpha(a), \gamma(a) \in \mathbb{R}$ with

$$f(x + a) = \alpha(a) f(x) + \gamma(a) \qquad \text{for all } x. \tag{4.5}$$

Step 1: α is continuous, positive and multiplicative. Since f is not constant, choose x_1, x_2 with $f(x_1) \neq f(x_2)$; subtracting the two instances of (4.5) gives $\alpha(a) = (f(x_1 + a) - f(x_2 + a))/(f(x_1) - f(x_2))$, a continuous function of a with $\alpha(0) = 1$. If $\alpha(a) = 0$ for some a, then $f(\cdot + a)$ is constant, a contradiction; so α never vanishes. Let $g(x) \coloneqq f(x) - f(0)$; taking $x = 0$ in (4.5) gives $\gamma(a) = f(a) - \alpha(a) f(0)$, whence

$$g(x + a) = \alpha(a) g(x) + g(a). \tag{4.6}$$

Expanding $g(x + a + b)$ in two ways, $\alpha(a + b) g(x) + g(a + b) = \alpha(b)\alpha(a) g(x) + \alpha(b) g(a) + g(b)$; the difference $(\alpha(a+b) - \alpha(a)\alpha(b)) g(x)$ is independent of x, and g is not constant, so $\alpha(a+b) = \alpha(a)\alpha(b)$. In particular $\alpha(a) = \alpha(a/2)^2 > 0$, and $\ln \alpha$ is a continuous additive function on $\mathbb{R}$, hence linear: $\alpha(a) = e^{\beta a}$ for some $\beta \in \mathbb{R}$.

Step 2: $\beta = 0$. Then (4.6) says that g is additive and continuous, hence $g(x) = c_1 x$ and $f(x) = c_1 x + f(0)$, with $c_1 \neq 0$.

Step 3: $\beta \neq 0$. Exchanging the roles of x and a in (4.6), $e^{\beta a} g(x) + g(a) = g(x + a) = e^{\beta x} g(a) + g(x)$, that is, $g(x)(e^{\beta a} - 1) = g(a)(e^{\beta x} - 1)$. Fix $a \neq 0$ and set $c_1 = g(a)/(e^{\beta a} - 1)$: then $g(x) = c_1(e^{\beta x} - 1)$ for all x, so $f(x) = c_1 e^{\beta x} + (f(0) - c_1)$, and $c_1 \neq 0$ since f is not constant. ∎

Remark 4.9 (On the continuity assumption). Continuity was used twice: to make α continuous, hence $\ln \alpha$ linear, and to solve the additive equation when $\beta = 0$. Both steps go through if f is merely Borel measurable, since measurable solutions of Cauchy's equation $\phi(a + b) = \phi(a) + \phi(b)$ are linear (Aczél, 1966, Chap. 2). Without any regularity, however, the statement fails on finitely supported laws: if ϕ is a non-measurable additive function, then $f = e^{\phi}$ satisfies (4.5) with $\alpha = e^{\phi}$ and $\gamma = 0$. Such f are not even measurable, so that $\int f\, \mathrm{d}\nu$ is undefined as soon as ν is not finitely supported; they play no role in what follows.

Families of utilities ♦

When several utilities are propagated together, (4.4) allows each $f_i(\cdot + a)$ to be a combination of *all* the f_j: the moments x^j are the basic example, none of them being closed alone for $j \geq 2$ while all of them are

closed together. The structure of such families is governed by a classical result on translation-invariant spaces of functions.

Lemma 4.10 (Finite-dimensional translation-invariant spaces). Let V be a finite-dimensional linear space of continuous functions $\mathbb{R} \to \mathbb{R}$ such that $f(\cdot + a) \in V$ for every $f \in V$ and $a \in \mathbb{R}$. Then every $f \in V$ is infinitely differentiable, V is stable under differentiation, and V is spanned by functions of the form $x^j e^{ux} \cos(vx)$ and $x^j e^{ux} \sin(vx)$ ($j \geq 0$, $u, v \in \mathbb{R}$).

Proof. Let $f_1, \dots, f_n$ be a basis of V and write $f_i(x + a) = \sum_j c_{ij}(a) f_j(x)$. Since the f_j are linearly independent, there are points $x_1, \dots, x_n$ such that the matrix $(f_j(x_l))_{j,l}$ is invertible (the evaluation map $V \to \mathbb{R}^{\{x_1,\dots,x_n\}}$ is injective for a suitable choice of points, obtained by adding one point at a time while the rank increases). Solving the linear system $f_i(x_l + a) = \sum_j c_{ij}(a) f_j(x_l)$, $l \in [n]$, shows that the c_{ij} are continuous. Two consequences follow, both resting on the fact that a finite-dimensional space of functions is closed under pointwise convergence: if $g_k \in V$ and $g_k \to g$ pointwise, then the vectors $(g_k(x_l))_l$ converge, and since every $h \in V$ is determined by $(h(x_l))_l$ through the dual basis, g_k converges pointwise to an element of V, which must be g.

First, let φ be a smooth function with compact support and unit integral. For $f = f_i$, $f_i * \varphi(x) = \int f_i(x - u)\varphi(u)\,\mathrm{d}u = \sum_j \big(\int c_{ij}(-u)\varphi(u)\,\mathrm{d}u \big) f_j(x)$ lies in V, hence so does $f * \varphi$ for every $f \in V$ by linearity, and $f * \varphi$ is infinitely differentiable. Taking $\varphi_\varepsilon(u) = \varepsilon^{-1}\varphi(u/\varepsilon)$, $f * \varphi_\varepsilon \to f$ pointwise as $\varepsilon \to 0$ by continuity of f; since $V \cap C^\infty$ is a finite-dimensional space closed under pointwise limits, $f \in C^\infty$. Second, $f'(x) = \lim_{h\to 0}(f(x+h) - f(x))/h$ is a pointwise limit of elements of V, hence $f' \in V$. The differentiation operator D is thus a linear endomorphism of V; by the Cayley–Hamilton theorem, its characteristic polynomial χ satisfies $\chi(D) = 0$ on V. Every $f \in V$ solves the linear differential equation $\chi(D)f = 0$ with constant coefficients, whose real solutions are spanned by the functions $x^j e^{ux} \cos(vx)$, $x^j e^{ux} \sin(vx)$ with $u + iv$ a root of χ of multiplicity larger than j (here $\mathrm{i}^2 = -1$). ∎

Theorem 4.11 (Bellman-closed families of utilities). Let $f_1, \dots, f_K$ be continuous. The utility $\psi_{\mathbf{f}} = (\int f_1\,\mathrm{d}\nu, \dots, \int f_K\,\mathrm{d}\nu)$ is Bellman closed if and only if the linear span of $\{1, f_1, \dots, f_K\}$ is invariant under translations. In that case, each f_i is an *exponential polynomial*: a finite linear combination of functions $x^j e^{ux} \cos(vx)$ and $x^j e^{ux} \sin(vx)$. Conversely, if V is a finite-dimensional translation-invariant space of continuous functions with basis $g_1, \dots, g_n$, then $\psi_{\mathbf{g}} = (\int g_1\,\mathrm{d}\nu, \dots, \int g_n\,\mathrm{d}\nu)$ is Bellman closed, and every utility $\psi_{\mathbf{f}}$ with $f_1, \dots, f_K \in V$ is a linear function of $\psi_{\mathbf{g}}$.

Proof. The equivalence is Lemma 4.7 combined with Proposition 4.5, and the structure of the f_i follows from Lemma 4.10 applied to $V_{\mathbf{f}}$. For the converse, write $g_l(x + a) = \sum_{l'} c_{ll'}(a) g_{l'}(x)$: this is (4.4) for $\mathbf{g}$ (with $\gamma = 0$), so $\psi_{\mathbf{g}}$ is closed by Lemma 4.7; and if $f_i = \sum_l \theta_{il} g_l$ then $\psi_{f_i} = \sum_l \theta_{il} \psi_{g_l}$. ∎

Thus a utility that is not closed on its own may still be computed exactly by dynamic programming, provided it is embedded in a closed family: the variance and the third central moment are the standard examples. The Bellman-closed families of utilities are, in this sense, exactly the finite-dimensional translation-invariant spaces of functions.

Example 4.12 (Closed families).

(i) *Polynomial moments* $\{x^j : j \leq K\}$: the span is translation invariant by the binomial formula; this is Proposition 4.1.

(ii) *Exponential polynomials at rate* β, $\{x^j e^{\beta x} : j \leq K\}$: since $(x+r)^j e^{\beta(x+r)} = e^{\beta r} \sum_l \binom{j}{l} r^{j-l} x^l e^{\beta x}$,

the statistics $N_\beta^{(j)}(\nu) = \int x^j e^{\beta x}\,\mathrm{d}\nu$ satisfy

$$N_\beta^{(j)}\big(\mathcal{T}_{p,r}(\nu_1,\dots,\nu_n)\big) = \sum_i p_i e^{\beta r_i} \sum_{l=0}^{j} \binom{j}{l} r_i^{j-l} N_\beta^{(l)}(\nu_i),$$

a triangular recursion of the same form as (4.1). For $K = 0$ this is the exponential moment $\int e^{\beta x}\,\mathrm{d}\nu$ alone, the case of Theorem 4.8; for $j \geq 1$, $N_\beta^{(j)}$ is the j-th derivative in β of the exponential moment.

(iii) *Fourier statistics* $\{\cos(\omega x), \sin(\omega x)\}$: the identities $\cos(\omega(x+r)) = \cos(\omega r)\cos(\omega x) - \sin(\omega r)\sin(\omega x)$ (and similarly for the sine) show that the pair is closed, with $\Phi_{p,r}$ a p-weighted sum of rotations by the angles ωr_i. In complex form, the characteristic function value $\int e^{\mathrm{i}\omega x}\,\mathrm{d}\nu$ is a closed statistic, multiplicative under shifts like the exponential moment. With integer rewards, the $HR_{\max}+1$ values at the frequencies $\omega_k = 2\pi k/(HR_{\max}+1)$ determine the law by the inverse discrete Fourier transform (Exercise 4.5).

The three families are the same object seen at different complex exponents $u + \mathrm{i}v$: real $u = \beta$ and $v = 0$ for exponential moments, $u = 0$ for Fourier statistics, and $u = v = 0$ with polynomial factors for moments. Theorem 4.11 says that there is nothing else.

Remark 4.13 (Discounting changes the picture). All of this is specific to the undiscounted case. With a discount factor $\gamma < 1$ the bootstrap map $b_{r,\gamma}(x) = r + \gamma x$ also scales the return, and integrating $e^{\beta x}$ against a mixture of such maps gives, with the notation $Z^\pi_{\beta,t}(s) = \int e^{\beta x}\,\nu_t^\pi(s)(\,\mathrm{d}x)$ of Definition 4.14 below, $Z^\pi_{\beta,t} = \sum_{a,s'} \pi_t P_t\, e^{\beta r} Z^\pi_{\beta\gamma,t+1}$: the exponential moment at rate β calls for the one at rate $\beta\gamma$, and no single exponential moment—no finite family of them—is Bellman closed for the discounted operator (mixtures of the maps $b_{r,\gamma}$ in place of the shifts b_r of Definition 4.4). The polynomial moments, on the other hand, survive with a factor γ^j in front of $M^{(j)}_{t+1}$, and in the discounted setting the only finite Bellman-closed families of utilities are spanned by moments (Rowland et al., 2019). Chapters 7 and 8 return to this point and show how time-dependent rates $\beta_t = \beta\gamma^{t-1}$ restore a recursion for the β-values of Definition 4.14.

4.4 Exponential moments and the exponential Bellman equation

Among the closed utilities, the exponential moments are, with the mean, the only ones that are closed *one at a time* (Theorem 4.8); and unlike the mean, their recursion is multiplicative rather than affine in the rewards. This section develops this recursion, whose logarithmic form is a Bellman equation for the β-means of Definition 2.16.

Definition 4.14 (Exponential moments and β-values). For $\beta \in \mathbb{R}$, the *exponential moment* and the *β-value* of π at (t,s) are

$$Z^\pi_{\beta,t}(s) \coloneqq \mathbb{E}^\pi_{t,s}\big[e^{\beta W_{t:H}}\big] = \int e^{\beta x}\,\nu_t^\pi(s)(\,\mathrm{d}x), \qquad U^\pi_{\beta,t}(s) \coloneqq U_\beta\big(\nu_t^\pi(s)\big) = \frac{1}{\beta}\ln Z^\pi_{\beta,t}(s),$$

with $U^\pi_{0,t} = V^\pi_t$; $Z^\pi_{\beta,t}(s,a)$ and $U^\pi_{\beta,t}(s,a)$ are defined likewise from $\nu_t^\pi(s,a)$.

Theorem 4.15 (Multiplicative Bellman equation). For every $\beta \in \mathbb{R}$, $t \in [H]$, $s \in \mathcal{S}$ and $a \in \mathcal{A}$, with $Z^\pi_{\beta,H+1} \equiv 1$,

$$Z^\pi_{\beta,t}(s,a) = \sum_{s' \in \mathcal{S}} P_t(s' \mid s,a)\, e^{\beta r_t(s,a,s')}\, Z^\pi_{\beta,t+1}(s'), \qquad Z^\pi_{\beta,t}(s) = \sum_{a \in \mathcal{A}} \pi_t(a \mid s)\, Z^\pi_{\beta,t}(s,a). \tag{4.7}$$

Equivalently, $Z^\pi_{\beta,t} = \widetilde{P}^\pi_{\beta,t} Z^\pi_{\beta,t+1}$ where $\widetilde{P}^\pi_{\beta,t}(s,s') \coloneqq \sum_a \pi_t(a \mid s) P_t(s' \mid s,a) e^{\beta r_t(s,a,s')}$ is the *tilted kernel*, a nonnegative matrix whose rows need not sum to one unless $\beta = 0$; hence $Z^\pi_{\beta,1} = \widetilde{P}^\pi_{\beta,1} \widetilde{P}^\pi_{\beta,2} \cdots \widetilde{P}^\pi_{\beta,H} \mathbf{1}$.

Proof. Integrate $e^{\beta x}$ against both sides of (3.1): since $\int e^{\beta x}\, (b_r)_\# \nu(\mathrm{d}x) = e^{\beta r} \int e^{\beta x}\, \mathrm{d}\nu$, the mixture of shifted laws has exponential moment $\sum_{s'} P_t(s' \mid s,a) e^{\beta r_t(s,a,s')} Z_{\beta,t+1}(s')$. The second identity is (3.2). The matrix form follows by combining the two. ∎

The classical Bellman equation is the derivative of (4.7) at $\beta = 0$: $\frac{\mathrm{d}}{\mathrm{d}\beta} Z_{\beta,t}(s)\big|_{\beta=0} = V_t(s)$, and differentiating the product gives $V_t(s) = \sum_{a,s'} \pi P\,[r + V_{t+1}(s')]$. Higher derivatives at 0 give the moment recursions of Proposition 4.1, and derivatives at $\beta \neq 0$ the recursions for $\int x^j e^{\beta x}\, \mathrm{d}\nu$ of Example 4.12: the exponential moment, as a function of β, is a generating function for all the closed utilities at once.

Taking logarithms in (4.7) produces a recursion for β-values in which the expectation of the classical Bellman equation is replaced by a β-mean. The key tool is the tower property of β-means, Proposition 2.17(vii), with the conditional β-mean $U_\beta(X \mid \mathfrak{G}) = \frac{1}{\beta} \ln \mathbb{E}[e^{\beta X} \mid \mathfrak{G}]$ (conditional expectation given a σ-algebra, Appendix A.3) in the role of the conditional expectation.

Corollary 4.16 (Exponential Bellman equation). For every $\beta \neq 0$, $t \in [H]$ and $s \in \mathcal{S}$, with $U^\pi_{\beta,H+1} \equiv 0$,

$$\begin{aligned} U^\pi_{\beta,t}(s) &= U_\beta\Big(r_t(s,A,S') + U^\pi_{\beta,t+1}(S')\Big) \\ &= \frac{1}{\beta} \ln \sum_{a,s'} \pi_t(a \mid s) P_t(s' \mid s,a) \exp\Big(\beta\big[r_t(s,a,s') + U^\pi_{\beta,t+1}(s')\big]\Big), \end{aligned} \tag{4.8}$$

where, in the first expression, (A,S') is distributed as (A_t, S_{t+1}) under $\mathbb{P}^\pi_{t,s}$.

Proof. Under $\mathbb{P}^\pi_{t,s}$, condition on $\mathfrak{G} = \sigma(A_t, S_{t+1})$: as in the proof of Corollary 4.2, $W_{t:H} = r_t(s,A_t,S_{t+1}) + W'$ with $W' \sim \nu_{t+1}(S_{t+1})$ given $\mathfrak{G}$, so that $U_\beta(W_{t:H} \mid \mathfrak{G}) = r_t(s,A_t,S_{t+1}) + U_{\beta,t+1}(S_{t+1})$ by translation (Proposition 2.17(iii)). Apply the tower property, Proposition 2.17(vii). The second expression spells out the β-mean of a finitely supported law; it is also (4.7) after taking logarithms. ∎

Algorithm 4.2 implements (4.8). Working with U rather than Z is not only a matter of interpretation: $Z_{\beta,t}$ ranges over $e^{\beta[\text{returns}]}$ and overflows or underflows as soon as $|\beta| H R_{\max}$ exceeds a few hundreds, whereas the β-values stay in $[\min W, \max W]$ (Proposition 2.17(ii)); the logarithm of a sum of exponentials is computed stably by factoring out the largest term (the "log-sum-exp" trick).

Remark 4.17 (The two limits). As $\beta \to 0$, $\frac{1}{\beta} \ln \sum_i p_i e^{\beta x_i} \to \sum_i p_i x_i$ and (4.8) becomes the classical Bellman equation. As $\beta \to +\infty$, $\frac{1}{\beta} \ln \sum_i p_i e^{\beta x_i} \to \max\{x_i : p_i > 0\}$, and the recursion becomes the $(\max, +)$ recursion for the largest attainable return of Exercise 3.1; as $\beta \to -\infty$, it becomes the $(\min, +)$ recursion for the smallest one (Section 4.5). The exponential Bellman

Algorithm 4.2: Backward induction for β-values

Input: MDP $\mathcal{M}$, Markov policy π, $\beta \neq 0$

$U_{H+1}(s) \leftarrow 0$ for all $s \in \mathcal{S}$

for $t = H, H-1, \dots, 1$ **do**

  for $s \in \mathcal{S}$ **do**

    $U_t(s) \leftarrow \frac{1}{\beta} \ln \sum_{a,s'} \pi_t(a \mid s) P_t(s' \mid s, a) \exp\left(\beta[r_t(s,a,s') + U_{t+1}(s')]\right)$ ▷ *log-sum-exp*

Output: $(U_t)_{t \in [H]}$, with $U_1(s) = U_\beta(v^\pi(s))$

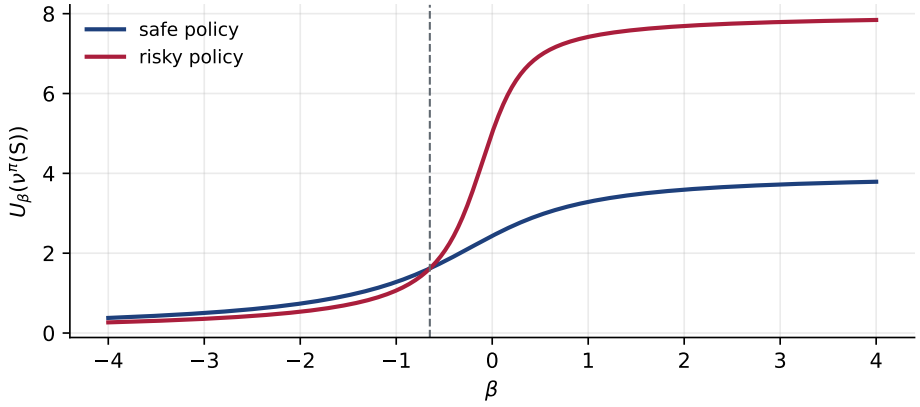


Figure 4.2. β-means of the return from the starting cell for the safe and risky policies of the windy hike (Example 3.6), computed by Algorithm 4.2; the curves cross at $\beta_\times \approx -0.65$ (dashed line).

equation interpolates between the risk-neutral, the most optimistic and the most pessimistic evaluations of a policy; the next chapter shows that it is, essentially, the only Bellman equation that can be *optimized*.

Example 4.18 (The β-means of the two hikers). Figure 4.2 shows $\beta \mapsto U_\beta(v^\pi(S))$ for the safe and risky policies of Example 3.6, computed by Algorithm 4.2; some values are

β	-2	-1	-0.5	0	0.5	1	2
safe	0.74	1.28	1.79	2.43	2.96	3.29	3.59
risky	0.53	1.07	2.04	5.02	6.95	7.42	7.69

The two curves cross at $\beta_\times \approx -0.65$. A hiker whose criterion is U_β prefers the risky path for $\beta > \beta_\times$—the child, and also a mildly prudent adult—and the safe path only when $\beta < \beta_\times$. The mean–variance approximation of Remark 2.18, $U_\beta \approx m + \beta\sigma^2/2$, predicts a crossing at $\beta \approx -2(5.02 - 2.43)/(13.7 - 2.5) \approx -0.46$, the variances being $\sigma^2 \approx 13.7$ and 2.5: with returns that are far from Gaussian, the $O(\beta^2)$ remainder matters. The Chernoff reading of Corollary 2.19 makes U_β concrete for $\beta < 0$: $\mathbb{P}(W \leq u) \leq e^{-|\beta|(U_\beta(W)-u)}$, so that U_β is a level below which the return falls only with a probability that decays at rate $|\beta|$ per unit of shortfall: for $\beta = -1$, the probability that the safe path yields fewer than $1.28 - x$ steps at the bar decays like e^{-x}, against $1.07 - x$ for the risky one. The safe path is not very rewarding either, which is why it takes a rather strong risk aversion to prefer it.

Remark 4.19 (Change of measure). The tilted kernel $\widetilde{P}^{\pi}_{\beta,t}$ of Theorem 4.15 has a probabilistic meaning. Write $Z_t = Z^{\pi}_{\beta,t}$ and fix the initial state s. The normalized kernel

$$K^{\beta}_t(a, s' \mid s) \coloneqq \frac{\pi_t(a \mid s) P_t(s' \mid s, a)\, e^{\beta r_t(s,a,s')} Z_{t+1}(s')}{Z_t(s)}$$

from $\mathcal{S}$ to $\mathcal{A} \times \mathcal{S}$ is a Markov kernel (it sums to one by (4.7)), and the product of these kernels is the exponential tilt $e^{\beta W_H}/\mathbb{E}e^{\beta W_H}$ of the law of the state–action trajectory started at s, since the factors $Z_{t+1}(s_{t+1})$ telescope: the process "conditioned on a high return" when $\beta > 0$, "on a low return" when $\beta < 0$ (compare with the tilted laws of Proposition 2.15). Its state marginal $P^{\pi,\beta}_t(s' \mid s) = \widetilde{P}^{\pi}_{\beta,t}(s, s') Z_{t+1}(s')/Z_t(s)$, the *Doob h-transform* of $\widetilde{P}^{\pi}_{\beta,t}$ by $h = Z_{t+1}$, describes the tilted law of the states alone; it does not determine the tilted trajectory law, because the tilt also reweights the actions, $\pi_t(a \mid s)$ becoming $\pi_t(a \mid s) Z^{\pi}_{\beta,t}(s, a)/Z_t(s)$. (With one decision, two equiprobable actions leading to the same state with rewards 0 and 1, the state kernel is unchanged while the action probabilities become proportional to 1 and e^{β}.) If the initial state is drawn from μ, it is tilted as well: $\mu^{\beta}(s) \propto \mu(s) Z_1(s)$. The Gibbs variational principle (Theorem C.11) then expresses $\beta U_{\beta,1}(s)$ as $\sup_Q \{\beta \mathbb{E}_Q[W_H] - \mathrm{KL}(Q \,\|\, \mathbb{P}^{\pi}_s)\}$, a formula that will be the source of the dual representations of Chapter 6.

4.5 Support statistics

Two closed statistics of a different nature are the extreme values of the return, which record where the distribution lives rather than how it is spread. For a law ν with bounded support, let $\underline{x}(\nu) = \operatorname{ess\,inf} \nu = \min \operatorname{supp} \nu$ and $\overline{x}(\nu) = \operatorname{ess\,sup} \nu = \max \operatorname{supp} \nu$.

Proposition 4.20 (Extreme values are Bellman closed). For weights p with $p_i > 0$ and shifts r, $\underline{x}(\mathcal{T}_{p,r}(\nu_1, \dots, \nu_n)) = \min_i \{r_i + \underline{x}(\nu_i)\}$ and $\overline{x}(\mathcal{T}_{p,r}(\nu_1, \dots, \nu_n)) = \max_i \{r_i + \overline{x}(\nu_i)\}$. Consequently the smallest and largest returns attainable with positive probability under π satisfy the $(\min, +)$ and $(\max, +)$ Bellman equations

$$\underline{x}_t(s) = \min \big\{ r_t(s, a, s') + \underline{x}_{t+1}(s') : \pi_t(a \mid s) P_t(s' \mid s, a) > 0 \big\},$$
$$\overline{x}_t(s) = \max \big\{ r_t(s, a, s') + \overline{x}_{t+1}(s') : \pi_t(a \mid s) P_t(s' \mid s, a) > 0 \big\},$$

with $\underline{x}_{H+1} = \overline{x}_{H+1} = 0$.

Proof. The support of $(b_r)_{\#}\nu$ is $r + \operatorname{supp} \nu$, and the support of a finite mixture with positive weights is the union of the supports; the minimum of a finite union is the minimum of the minima. The pair $(\underline{x}, \overline{x})$ is closed with $\Phi_{p,r}$ depending on p only through its support, which is what the Bellman equations express. ∎

These recursions are the $\beta \to \pm\infty$ limits of Remark 4.17. They are cheap ($O(HS^2A)$) and useful in two ways. First, they bound the return distribution: $\nu^{\pi}_t(s)$ is supported in $[\underline{x}_t(s), \overline{x}_t(s)]$, an interval usually much shorter than the a priori bound $[(H - t + 1) \min r, (H - t + 1) \max r]$, which will serve to normalize returns before reconstructing distributions from moments (Section 4.6). Second, they give the guaranteed return of a policy, the criterion of an agent who cannot afford any bad outcome; the question of *optimizing* the guaranteed return leads to a deterministic Markov policy, as does the mean (Exercise 3.1), for reasons that Chapter 5 will make precise.

4.6 From statistics back to distributions

The results so far are both reassuring and disappointing. Reassuring: moments, exponential moments and extreme values can be propagated exactly, at the cost of a value function with a few coordinates. Disappointing: among the utilities $\int f \, \mathrm{d}\nu$, this is all—and a quantile (or *value at risk*, Definition 6.1), a threshold probability, a conditional value at risk (Definition 6.9) cannot be propagated either (Example 4.6 and Exercise 4.3); nothing more is claimed here about statistics that are not utilities, beyond the extreme values of Section 4.5. Since quantiles and values at risk are the quantities that risk-sensitive agents care about (Chapter 6), one has to choose between two approximate strategies:

- propagate a closed family exactly, and *reconstruct* an approximate distribution from the exact statistics at the end (*imputation*, in the terminology of Rowland et al. 2019);
- propagate an *approximate distribution* with a fixed number of parameters, *projecting* the mixture of shifts back onto the parametric family at every step.

We analyze the first strategy for moments, and compare it with the second.

What moments determine: an agnostic bound

A law with bounded support is determined by its moments (by Weierstrass' theorem, polynomials being dense in $C([a,b])$), but a finite number of moments is not enough (Exercise 3.10). The following theorem quantifies what the first m moments do determine, in the Wasserstein distance $\mathcal{W}_1$ of Appendix D, without any assumption on the laws.

Theorem 4.21 (Moment matching in Wasserstein distance). Let $\nu, \hat{\nu} \in \mathcal{M}_1(\mathbb{R})$ be supported in a common interval of length Δ and have the same moments of orders $1, \dots, m$. Then

$$\mathcal{W}_1(\nu, \hat{\nu}) \;\le\; \frac{6\Delta}{m}.$$

The proof compares a Lipschitz test function with a polynomial of degree m, and the rate $1/m$ is the rate at which polynomials of degree m approximate Lipschitz functions uniformly. This is a classical result of approximation theory, which we state without proof; Exercise 4.9 gives a completely elementary argument, with the weaker rate $1/\sqrt{m}$, which also proves Weierstrass' theorem.

Theorem 4.22 (Jackson's theorem; Rivlin, 1969, Chap. 1, Cheney, 1966, Chap. 4). For every continuous $\varphi : [-1,1] \to \mathbb{R}$ and every $m \ge 1$ there exists a polynomial P of degree at most m such that $\sup_{|x|\le 1} |\varphi(x) - P(x)| \le 6\,\omega_\varphi(1/m)$, where $\omega_\varphi(\delta) = \sup\{|\varphi(x)-\varphi(y)| : |x-y| \le \delta\}$ is the modulus of continuity of φ. In particular, a 1-Lipschitz function is uniformly approximated within $6/m$ by a polynomial of degree m. If moreover φ is k times continuously differentiable, $k \ge 1$, with $\|\varphi^{(k)}\|_\infty \coloneqq \sup_{|x|\le 1} |\varphi^{(k)}(x)|$, then for every $m \ge k$,

$$\inf_{\deg P \le m} \sup_{|x|\le 1} |\varphi(x) - P(x)| \;\le\; \left(\frac{\pi}{2}\right)^k \frac{\|\varphi^{(k)}\|_\infty}{(m+1)\,m \cdots (m-k+2)};$$

in particular, for $m \ge 2k$, each of the k factors of the denominator is at least $m-k+2 \ge m/2$, and the right-hand side is at most $(\pi/m)^k \|\varphi^{(k)}\|_\infty$.

Proof of Theorem 4.21. Let $[a,b]$ be the interval and $\ell(x) = 2(x-a)/(b-a) - 1$ the affine map onto $[-1,1]$. The push-forwards $(\ell)_\# \nu$ and $(\ell)_\# \hat{\nu}$ are supported in $[-1,1]$, have the same moments up to order m (a polynomial of degree $\le m$ in $\ell(x)$ is a polynomial of degree $\le m$ in x), and $\mathcal{W}_1((\ell)_\#\nu, (\ell)_\#\hat{\nu}) =$

$\frac{2}{\Delta}\mathcal{W}_1(\nu,\hat\nu)$ by Lemma D.7. It suffices therefore to prove the bound $12/m$ when $\Delta = 2$ and $[a,b] = [-1,1]$. By the Kantorovich–Rubinstein duality (Theorem D.4), $\mathcal{W}_1(\nu,\hat\nu) = \sup_\varphi \int \varphi \,\mathrm{d}(\nu-\hat\nu)$ over 1-Lipschitz φ, and only the values of φ on $[-1,1]$ matter. For such φ, let P be the polynomial of Theorem 4.22 for the 1-Lipschitz function φ: since $\int P\,\mathrm{d}\nu = \int P\,\mathrm{d}\hat\nu$ by the moment condition,

$$\int \varphi \,\mathrm{d}(\nu-\hat\nu) = \int (\varphi - P)\,\mathrm{d}(\nu-\hat\nu) \le 2 \sup_{|x|\le 1} |\varphi - P| \le \frac{12}{m}. \qquad \blacksquare$$

The bound has two remarkable features. It is *agnostic*: nothing is assumed about the shape of ν or of $\hat\nu$, and any law with the right moments will do. And it involves the horizon only through the length Δ of the support, which the extreme-value recursion of Proposition 4.20 computes exactly; in the worst case $\Delta \le HR_{\max}$, so that m exact moments determine the return distribution within $6HR_{\max}/m$. The rate $1/m$ cannot be improved in general (Exercise 4.8), but the constant is pessimistic for regular laws: in the windy hike, the maximum-entropy law with the same first m moments as the risky return distribution (defined below; Figure 4.4 uses Chebyshev moments, which span the same space as the raw moments) lies at $\mathcal{W}_1$ distance 0.15, 0.022, 0.0040 and 0 for $m = 2, 4, 6, 8$—the reference measure being the counting measure on the exact support $\{0,\dots,8\}$, so that eight moments determine a law on nine atoms—where the bound gives 24, 12, 8 and 6 with the exact support length 8, and 42, 21, 14, 10.5 with the a priori $\Delta = H = 14$.

Maximum-entropy reconstruction

Which law $\hat\nu$ with the prescribed moments should one pick? A principled choice, going back to Jaynes (1957), is the one of maximal entropy relative to a reference measure λ on the support—the least committal law compatible with the data. Let λ be a finite measure on $[a,b]$ (Lebesgue measure, or the counting measure on a lattice), $T = (T_1,\dots,T_m)$ be polynomials spanning, with the constants, the polynomials of degree $\le m$ (for instance $T_j(x) = x^j$), and $c \in \mathbb{R}^m$ be a target moment vector. For $\theta \in \mathbb{R}^m$ define the *exponential family*

$$p_\theta(x) \coloneqq \frac{e^{\langle \theta, T(x)\rangle}}{Z(\theta)}, \qquad Z(\theta) \coloneqq \int_a^b e^{\langle \theta, T(x)\rangle} \lambda(\,\mathrm{d}x),$$

a probability density with respect to λ, and let $\nu_\theta(\,\mathrm{d}x) \coloneqq p_\theta(x)\,\lambda(\,\mathrm{d}x)$ be the corresponding law. In this section, laws are denoted by ν (with decorations), their densities with respect to λ by p or q, and polynomials by P.

Proposition 4.23 (Maximum entropy and its dual). Let $J(\theta) \coloneqq \ln Z(\theta) - \langle \theta, c\rangle$. Then J is convex, with $\nabla J(\theta) = \mathbb{E}_{\nu_\theta}[T] - c$ and $\nabla^2 J(\theta) = \mathbb{C}\mathrm{ov}_{\nu_\theta}[T]$. If $\theta^\star$ is a critical point of J, then $\nu^\star \coloneqq \nu_{\theta^\star}$ has moment vector c, and among all laws $\nu \ll \lambda$ with $\int T\,\mathrm{d}\nu = c$ it is the unique minimizer of the relative entropy $\mathrm{KL}(\nu\,\|\,\lambda)$ (equivalently, the unique maximizer of the entropy $-\int \ln \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\,\mathrm{d}\nu$). We write $p^\star \coloneqq p_{\theta^\star}$ for its density.

Proof. Derivatives. Since T is bounded on $[a,b]$ and λ is finite, $Z(\theta)$ is finite and positive for every θ. For θ ranging in a bounded set, the first and second partial derivatives $T_j e^{\langle\theta,T\rangle}$ and $T_jT_k e^{\langle\theta,T\rangle}$ of the integrand are bounded by a constant, which is λ-integrable, so that differentiation under the integral sign is legitimate (Proposition A.1(iv)) and gives

$$\partial_j Z(\theta) = \int T_j\, e^{\langle\theta,T\rangle}\,\mathrm{d}\lambda = Z(\theta)\,\mathbb{E}_{\nu_\theta}[T_j], \qquad \partial_{jk} Z(\theta) = \int T_jT_k\, e^{\langle\theta,T\rangle}\,\mathrm{d}\lambda = Z(\theta)\,\mathbb{E}_{\nu_\theta}[T_jT_k].$$

Hence $\partial_j \ln Z = \partial_j Z/Z = \mathbb{E}_{\nu_\theta}[T_j]$ and

$$\partial_{jk}\ln Z = \frac{\partial_{jk}Z}{Z} - \frac{\partial_j Z}{Z}\frac{\partial_k Z}{Z} = \mathbb{E}_{\nu_\theta}[T_jT_k] - \mathbb{E}_{\nu_\theta}[T_j]\,\mathbb{E}_{\nu_\theta}[T_k] = \mathbb{C}\mathrm{ov}_{\nu_\theta}(T_j,T_k),$$

which are the announced formulas, the term $\langle\theta, c\rangle$ being linear in θ. (For $m = 1$ and $T_1(x) = x$ this is Proposition 2.15: $\ln Z$ is then, up to the constant $\ln\lambda([a,b])$, the cumulant generating function of $X \sim \lambda/\lambda([a,b])$, and ν_θ is its tilted law.)

Convexity and critical points. For $u \in \mathbb{R}^m$, $u^\top \nabla^2 J(\theta)\, u = \sum_{j,k} u_j u_k \mathbb{C}\mathrm{ov}_{\nu_\theta}(T_j, T_k) = \mathbb{V}\mathrm{ar}_{\nu_\theta}\big[\langle u, T\rangle\big] \geq 0$: the Hessian is positive semidefinite everywhere, so that J is convex and every critical point is a global minimizer. By the formula for the gradient, $\nabla J(\theta^\star) = 0$ means exactly that $\mathbb{E}_{\nu^\star}[T] = c$.

Maximum entropy. For a law $\nu \ll \lambda$ with density $q = \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}$, write $\mathrm{KL}(\nu \,\|\, \lambda) \coloneqq \int q \ln q \,\mathrm{d}\lambda = \int \ln q \,\mathrm{d}\nu$. Although λ is not a probability measure, this quantity is well defined in $(-\infty, +\infty]$, because the negative part of $q \ln q$ is at most $1/e$ and λ is finite; it may be negative, and only differences of such quantities will matter. Let $\nu \ll \lambda$ have moment vector c; if $\mathrm{KL}(\nu \,\|\, \lambda) = +\infty$ there is nothing to prove, so assume it finite. The key observation is that $\ln p_{\theta^\star} = \langle\theta^\star, T\rangle - \ln Z(\theta^\star)$ is an affine function of T, so that its integral against ν depends on ν only through the moments:

$$\int \ln p_{\theta^\star} \,\mathrm{d}\nu = \langle\theta^\star, c\rangle - \ln Z(\theta^\star).$$

Since $\nu^\star$ has moment vector c as well (first part), the right-hand side is also $\int \ln p_{\theta^\star} \,\mathrm{d}\nu^\star = \mathrm{KL}\big(\nu^\star \,\big\|\, \lambda\big)$. As $\frac{\mathrm{d}\nu}{\mathrm{d}\nu^\star} = q/p_{\theta^\star}$, it follows that

$$\mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) = \int \ln q \,\mathrm{d}\nu - \int \ln p_{\theta^\star} \,\mathrm{d}\nu = \mathrm{KL}(\nu \,\|\, \lambda) - \mathrm{KL}\big(\nu^\star \,\big\|\, \lambda\big)\,.$$

The left-hand side is the relative entropy between two probability measures: it is nonnegative, and zero only if $\nu = \nu^\star$ (Jensen's inequality for the strictly convex function $u \mapsto u \ln u$ applied to $\frac{\mathrm{d}\nu}{\mathrm{d}\nu^\star}$; see Appendix D.4). Hence $\mathrm{KL}(\nu \,\|\, \lambda) \geq \mathrm{KL}\big(\nu^\star \,\big\|\, \lambda\big)$, with equality only for $\nu = \nu^\star$, which is the claim; the entropy $-\int \ln \frac{\mathrm{d}\nu}{\mathrm{d}\lambda} \,\mathrm{d}\nu = -\mathrm{KL}(\nu \,\|\, \lambda)$ is maximized at the same point. ∎

Existence depends on the reference measure, not only on c. If $\operatorname{supp}\lambda$ has at least $m+1$ points, $\mathbb{C}\mathrm{ov}_{\nu_\theta}[T]$ is positive definite (a nonzero polynomial of degree at most m cannot vanish on $m+1$ points), so that J is strictly convex and a critical point is unique if it exists; and it exists if and only if c lies in the interior of the convex hull of $\{T(x) : x \in \operatorname{supp}\lambda\}$, which is the case when c is the moment vector of a law $\nu \ll \lambda$ whose support has at least $m+1$ points (an *interior* moment vector); see Csiszár (1975), Barron and Sheu (1991) and Exercise 4.10. A law charging $m+1$ points of $[a,b]$ outside $\operatorname{supp}\lambda$ does not qualify: for $m = 1$, λ the counting measure on $\{0, \frac{1}{4}\}$ and $c = \frac{3}{4}$, the mean of $\frac{1}{2}(\delta_{1/2} + \delta_1)$, no law $\nu \ll \lambda$ has mean c. The same caution applies to moments that are only approximately known—estimated, or propagated through approximations: they may fall outside, or on the boundary of, the set of moment vectors realizable on $\operatorname{supp}\lambda$, in which case J has no minimizer and the target must first be moved into that set; we return to this point below.

Computing the maximum-entropy law

In practice λ is a finite measure with finite support, $\lambda = \sum_{k=0}^{n} w_k \delta_{x_k}$ with $w_k > 0$ and $n \geq m$: the counting measure ($w_k = 1$) on the support $\{x_0, \ldots, x_n\}$ delivered by Proposition 4.20 when the rewards take finitely many values, or, for Lebesgue measure on $[a,b]$, its n-point Gauss–Legendre quadrature (Exercise 4.8) with n much larger than m, which integrates exactly all polynomials of degree $\leq 2n-1$. The law ν_θ is then the vector of masses $\nu_\theta(\{x_k\}) = w_k p_\theta(x_k) = w_k e^{\langle\theta, T(x_k)\rangle}/Z(\theta)$, and J, its gradient and its Hessian are finite sums over the atoms, computable in $O(mn)$ and $O(m^2 n)$ operations from the $m \times (n+1)$ matrix $\mathbf{T} = (T_j(x_k))_{j,k}$, which is formed once. Minimizing J is a smooth, strictly convex problem in m unknowns, with m rarely larger than ten, and the method of choice is Newton's. Its step from the current θ solves the linearization around θ of the moment-matching equation $\nabla J(\theta') = 0$,

$$\mathbb{E}_{\nu_\theta}[T] + \mathbb{C}\mathrm{ov}_{\nu_\theta}[T]\,(\theta' - \theta) = c, \qquad \text{that is,} \qquad \theta' = \theta - \mathbb{C}\mathrm{ov}_{\nu_\theta}[T]^{-1}\big(\mathbb{E}_{\nu_\theta}[T] - c\big)\,:$$

Algorithm 4.3: Damped Newton's method for the maximum-entropy law

Input: atoms $x_0, \dots, x_n$ and weights $w_k > 0$ of $\lambda = \sum_k w_k \delta_{x_k}$ ($n \geq m$); polynomials $T_1, \dots, T_m$; interior target $c \in \mathbb{R}^m$; tolerance $\varepsilon > 0$

$\mathbf{T} \leftarrow \big(T_j(x_k)\big)_{j \leq m,\, k \leq n}$ ▷ *$m \times (n+1)$ matrix, formed once*

$\theta \leftarrow 0$ ▷ *start from $\nu_0 = \lambda/\lambda([a,b])$*

repeat

  $\ell_k \leftarrow \ln w_k + \langle \theta, T(x_k) \rangle$ for $k \leq n$; $\quad L \leftarrow \max_k \ell_k + \ln \sum_k e^{\ell_k - \max_k \ell_k}$ ▷ *$L = \ln Z(\theta)$, log-sum-exp*

  $p_k \leftarrow e^{\ell_k - L}$ for $k \leq n$ ▷ *the masses $p_k = \nu_\theta(\{x_k\})$*

  $\mu \leftarrow \mathbf{T}p$; $\quad g \leftarrow \mu - c$ ▷ *moments of ν_θ, gradient $\nabla J(\theta)$*

  $\mathbf{H} \leftarrow \mathbf{T}\,\mathrm{diag}(p)\mathbf{T}^\top - \mu\mu^\top$ ▷ *Hessian $\mathbb{C}\mathrm{ov}_{\nu_\theta}[T]$*

  solve $\mathbf{H}d = -g$ for d ▷ *Newton direction, by a Cholesky factorization*

  $\alpha \leftarrow 1$

  while $J(\theta + \alpha d) > J(\theta) + \frac{1}{4}\alpha\langle g, d \rangle$ **do**

    $\alpha \leftarrow \alpha/2$ ▷ *backtracking; $J(\theta) = L - \langle \theta, c \rangle$ by the same log-sum-exp*

  $\theta \leftarrow \theta + \alpha d$

until $\|g\| \leq \varepsilon$

Output: θ and the masses $p = (p_k)_{k \leq n}$ of ν_θ, a law of the exponential family with $\|\mathbb{E}_{\nu_\theta}[T] - c\| \leq \varepsilon$

the covariance matrix measures how much each moment of ν_θ responds to a change of each parameter, and the step corrects the parameters so as to cancel the moment mismatch to first order. Far from the solution the response is nonlinear and the full step may overshoot; the standard remedy is a backtracking line search, which halves the step until J decreases by at least a quarter of what the linear model predicts (Armijo's rule). The resulting Algorithm 4.3 is the damped Newton method of Boyd and Vandenberghe (2004, §9.5), and inherits its two guarantees for a strictly convex function with bounded sublevel sets: it converges from any starting point—here coercivity is what makes the sublevel set $\{J \leq J(0)\}$ bounded, and it holds exactly when c is interior (Exercise 4.10(b))—and once θ is close enough to $\theta^\star$ the full step is always accepted and the convergence is quadratic, the number of correct digits doubling at each iteration. On the laws of this chapter, machine precision is reached in a few dozen iterations at most (five to nine in Figure 4.4), of which only the last three or four are in the quadratic regime; the count grows when ν is much more concentrated than λ, because the iterates must then travel far from $\theta = 0$.

The algorithm certifies its own output. The gradient $\nabla J(\theta) = \mathbb{E}_{\nu_\theta}[T] - c$ is the moment mismatch of the current law, so that the stopping rule $\|\nabla J(\theta)\| \leq \varepsilon$ returns a law of the exponential family whose moments are within ε of the target; and the optimality gap $J(\theta) - J(\theta^\star)$ equals the relative entropy $\mathrm{KL}\big(\nu^\star \,\big\|\, \nu_\theta\big)$ of the exact maximum-entropy law to the returned one (Lemma 4.24 below): stopping early costs exactly that much relative entropy. Each iteration costs $O(m^2 n)$ operations for the Hessian and $O(m^3)$ for the linear solve, negligible amounts: in Figure 4.4, $n + 1 = 9$ atoms and $m \leq 8$.

Two numerical points deserve attention. The first is the evaluation of $Z(\theta)$: the exponents $\langle \theta, T(x_k) \rangle$ can reach several hundreds during the iterations, so that the sum of exponentials must be computed by factoring out its largest term, as in Algorithm 4.2; Algorithm 4.3 works throughout with the log-masses ℓ_k for this reason. The second is the choice of the polynomials $T_1, \dots, T_m$. The maximum-entropy law does not depend on it: two bases spanning, with the constants, the polynomials of degree $\leq m$ differ by an invertible affine change of coordinates on θ and a linear one on c, and the family $\{\nu_\theta\}$ is the same. The numerics, however, depend on it heavily. The Hessian $\mathbb{C}\mathrm{ov}_{\nu_\theta}[T]$ is the Gram matrix of the centered functions $T_j - \mathbb{E}_{\nu_\theta}[T_j]$ in $L^2(\nu_\theta)$. For the monomials x^j on $[0,1]$ and ν_θ close to uniform, this is essentially the Hilbert matrix $(1/(j+k+1))_{j,k}$, whose condition number grows like $e^{3.5m}$: about 10^{12} for $m = 8$, so that the linear solve loses twelve of the sixteen digits of double precision—not surprisingly, since the raw moments of orders 7 and 8 of the uniform law, $1/8$ and $1/9$, carry almost no independent information. The Chebyshev polynomials $T_j(x) = \cos(j \arccos u)$,

with $u = (2x - a - b)/(b - a) \in [-1, 1]$, computed on the atoms by the recursion $T_0 = 1$, $T_1 = u$, $T_{j+1} = 2uT_j - T_{j-1}$, are bounded by one in absolute value and orthogonal on $[-1, 1]$ for the weight $(1 - u^2)^{-1/2}$; as long as ν_θ spreads over a reasonable fraction of $[a, b]$, their Gram matrices stay well conditioned (at most $2 \cdot 10^4$ in Figure 4.4 for $m \le 8$, against 10^{14} with the monomials, for which the tolerance $\varepsilon = 10^{-12}$ then becomes unreachable), the moments $c_j = \int T_j \, \mathrm{d}\nu$ are all of order one, and the exponents $\langle \theta, T(x_k) \rangle$ stay moderate. When the law is concentrated on a small part of $[a, b]$, every basis degrades, and the remedy is to work on the exact support of Proposition 4.20 rather than on an a priori interval, which is what Figure 4.4 does. The reconstructions of that figure and the computations of Remark 4.26 were obtained in this way.

Finally, a word on failure. When c is not interior—estimated, or approximately propagated, moments—J has no minimizer, and Algorithm 4.3 does not stop with an error: the iterates θ drift to infinity, ν_θ concentrates on a subset of the atoms, the Hessian becomes nearly singular, the line search accepts ever smaller steps, and $\|\nabla J(\theta)\|$ stalls above ε or decreases very slowly (Exercise 4.10(c)). A growing $\|\theta\|$ with a stalling gradient is the signature to watch for. The cure is to move the target inside the realizable set before solving, for instance by shrinking it toward the moment vector $\nabla \ln Z(0)$ of the normalized reference measure $\lambda/\lambda([a, b])$, which is interior: $c_\eta = (1 - \eta)c + \eta \, \nabla \ln Z(0)$ is interior for every $\eta \in (0, 1]$ when c lies in the closure of the realizable set. When the moments come from the exact recursion of Algorithm 4.1 and λ charges the exact support, as in this chapter, the question does not arise.

Quality of the reconstruction

How good is the reconstruction? The answer rests on a second reading of the dual objective J: up to a constant, it is the relative entropy of the unknown law to the fitted one.

Lemma 4.24 (The dual objective is a relative entropy). For every law $\nu \ll \lambda$ with moment vector c and every $\theta \in \mathbb{R}^m$,

$$\mathrm{KL}(\nu \,\|\, \nu_\theta) \;=\; \mathrm{KL}(\nu \,\|\, \lambda) + J(\theta).$$

Consequently, if $\theta^\star$ is a critical point of J, then $\nu^\star = \nu_{\theta^\star}$ is the law of the exponential family closest to ν, and for every θ,

$$\mathrm{KL}(\nu \,\|\, \nu_\theta) \;=\; \mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) + \mathrm{KL}\big(\nu^\star \,\big\|\, \nu_\theta\big)\,.$$

Proof. Since $\ln p_\theta = \langle \theta, T \rangle - \ln Z(\theta)$ and $\int T \, \mathrm{d}\nu = c$, $\mathrm{KL}(\nu \,\|\, \nu_\theta) = \int \ln \frac{\mathrm{d}\nu}{\mathrm{d}\lambda} \, \mathrm{d}\nu - \int \ln p_\theta \, \mathrm{d}\nu = \mathrm{KL}(\nu \,\|\, \lambda) - \langle \theta, c \rangle + \ln Z(\theta)$. Thus $\theta \mapsto \mathrm{KL}(\nu \,\|\, \nu_\theta)$ differs from J by a constant and is minimized at $\theta^\star$, and $\mathrm{KL}(\nu \,\|\, \nu_\theta) - \mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) = J(\theta) - J(\theta^\star)$ does not depend on ν: computing it for $\nu = \nu^\star$, which has moments c, gives $\mathrm{KL}\big(\nu^\star \,\big\|\, \nu_\theta\big)$. ∎

The maximum-entropy law is thus, at the same time, the law with moments c closest to λ (Proposition 4.23) and the law of the exponential family closest to ν—for *every* ν with moments c, which is what makes it computable without knowing ν. Figure 4.3 shows the two orthogonality relations in the simplest case. The second one is what controls the reconstruction error: $\nu^\star$ is at least as close to ν as *any* law of the family, in particular as the one whose log-density is the best polynomial approximation of the log-density of ν. For a bounded function g on the support of λ, write

$$E_m(g) \;\coloneqq\; \inf\Big\{ \sup_{x \in \operatorname{supp}\lambda} |g(x) - P(x)| \,:\, P \text{ polynomial of degree} \le m \Big\}$$

for its best uniform approximation error by polynomials of degree at most m.

Theorem 4.25 (Quality of the maximum-entropy reconstruction ♦). Let $\nu = e^g\lambda$ with g bounded on the support of λ, and let $\nu^\star = \nu_{\theta^\star}$ be the maximum-entropy law with the moments of ν (it exists when $\operatorname{supp}\nu$, which is contained in $\operatorname{supp}\lambda$, has at least $m+1$ points: see the discussion after Proposition 4.23 and Exercise 4.10). Then

$$\mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) \;\le\; \tfrac{1}{2}\,E_m(g)^2,$$

and $\mathcal{W}_1(\nu,\nu^\star) \le \frac{\Delta}{2}E_m(g)$ when λ is carried by an interval of length Δ. In particular, if g is the restriction to $\operatorname{supp}\lambda$ of a function with k continuous derivatives on that interval, with $|g^{(k)}| \le M_k$, then for every $m \ge 2k$,

$$\mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) \;\le\; \frac{1}{2}\Big(\frac{\pi\Delta}{2m}\Big)^{2k} M_k^2 \qquad\text{and}\qquad \mathcal{W}_1(\nu,\nu^\star) \;\le\; \frac{\Delta}{2}\Big(\frac{\pi\Delta}{2m}\Big)^{k} M_k.$$

Proof. Let P be a polynomial of degree at most m. Since T spans the polynomials of degree $\le m$ together with the constants, $P = a + \langle\theta, T\rangle$ for some $a \in \mathbb{R}$ and $\theta \in \mathbb{R}^m$, and $p_\theta = e^P/\int e^P\,\mathrm{d}\lambda$. By Lemma 4.24, $\mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) \le \mathrm{KL}(\nu \,\|\, \nu_\theta)$. Let $r \coloneqq g - P$, a bounded function. As $\nu = e^g\lambda$, one has $\int e^P\,\mathrm{d}\lambda = \int e^{P-g}\,\mathrm{d}\nu = \mathbb{E}_\nu[e^{-r}]$, so that

$$\mathrm{KL}(\nu \,\|\, \nu_\theta) = \int (g-P)\,\mathrm{d}\nu + \ln\int e^P\,\mathrm{d}\lambda = \mathbb{E}_\nu[r] + \ln\mathbb{E}_\nu\big[e^{-r}\big] = \ln\mathbb{E}_\nu\big[e^{-(r-\mathbb{E}_\nu[r])}\big] \;\le\; \frac{(\sup r - \inf r)^2}{8}$$

by Hoeffding's lemma (Lemma 2.10). The oscillation $\sup r - \inf r$ of $g-P$ does not change when a constant is added to P, and equals $2\inf_a \|g-P-a\|_\infty$; taking the infimum over P gives $\mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) \le (2E_m(g))^2/8$. Pinsker's inequality and $\mathcal{W}_1 \le \Delta\|\cdot\|_{\mathrm{TV}}$ (Appendix D.4) give the bound in Wasserstein distance. Finally, let g have k continuous derivatives on an interval of length Δ and ℓ be the affine map of that interval onto $[-1,1]$, as in the proof of Theorem 4.21: the transported function $\tilde g = g\circ\ell^{-1}$ satisfies $|\tilde g^{(k)}| \le (\Delta/2)^k M_k$ on $[-1,1]$, polynomials of degree $\le m$ in x and in $\ell(x)$ are the same, and the derivative form of Jackson's theorem (Theorem 4.22) gives, for $m \ge 2k$,

$$E_m(g) = E_m(\tilde g) \;\le\; \Big(\frac{\pi}{m}\Big)^k\Big(\frac{\Delta}{2}\Big)^k M_k = \Big(\frac{\pi\Delta}{2m}\Big)^k M_k. \qquad \blacksquare$$

The gain over Theorem 4.21 comes with smoothness, and this is the point of the theorem: for a log-density with k bounded derivatives, the error decays like m^{-k} in Wasserstein distance and like m^{-2k} in relative entropy, against $1/m$ for an arbitrary law with the same moments. The quantity $\Delta^k M_k$ that multiplies m^{-k} is dimensionless (a change of unit on the returns leaves it unchanged, as it leaves the relative entropy and $\mathcal{W}_1/\Delta$): it measures how much the log-density varies over the support, in the scale of the k-th derivative. For $k=1$ the theorem is not impressive: the bound $\pi L\Delta^2/(4m)$ on $\mathcal{W}_1$ for an L-Lipschitz log-density improves on $6\Delta/m$ only when $L\Delta < 24/\pi \approx 7.6$, and $E_m(g)$ itself decays no faster than $1/m$ for some Lipschitz g (such as $|x|$). Also, the theorem does not apply to log-densities with singularities, such as a law vanishing on part of the support of λ ($g = -\infty$ there) or with a jump: Remark 4.26 returns to this restriction.

The argument is that of Barron and Sheu (1991), streamlined, and their route leads to a refinement worth stating. Hoeffding's lemma replaces the Taylor bound $e^y - 1 - y \le \frac{1}{2}y^2 e^{y_+}$, which gives instead, for a polynomial P of degree $\le m$ normalized so that $\int (g-P)\,\mathrm{d}\nu = 0$, with $r = g-P$ and $\nu_P \coloneqq e^P\lambda/\int e^P\,\mathrm{d}\lambda$ the law of the exponential family whose log-density is P up to a constant,

$$\mathrm{KL}(\nu \,\|\, \nu_P) \;=\; \ln\mathbb{E}_\nu\big[e^{-r}\big] \;\le\; \mathbb{E}_\nu\big[e^{-r} - 1 + r\big] \;\le\; \tfrac{1}{2}\int r^2\,\big(\mathrm{d}\nu + e^P\,\mathrm{d}\lambda\big), \tag{4.9}$$

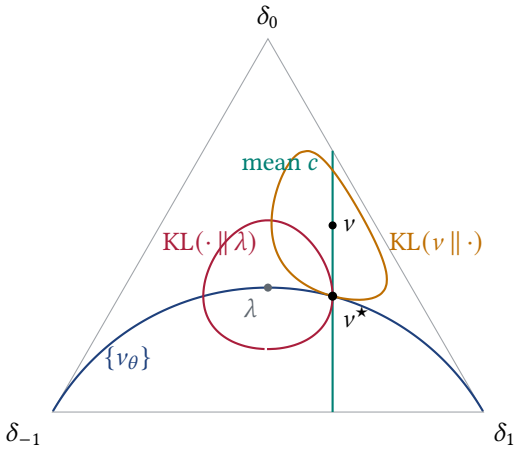


Figure 4.3. Geometry of the maximum-entropy reconstruction for laws on $\{-1, 0, 1\}$ and the mean as the only moment ($m = 1$). The simplex is the triangle with vertices δ_{-1}, δ_0, δ_1, and λ is the uniform law. The exponential family $\{\nu_\theta \propto e^{\theta x}\lambda\}$ (blue) joins δ_{-1} to δ_1 through λ; the laws with mean c form the vertical segment (teal). Their intersection $\nu^\star$ is at once the point of the segment closest to λ—the level set of $\mathrm{KL}(\cdot \,\|\, \lambda)$ through $\nu^\star$ (crimson) is tangent to the segment (Proposition 4.23)—and the point of the family closest to ν: the level set of $\mathrm{KL}(\nu \,\|\, \cdot)$ through $\nu^\star$ (amber) is tangent to the blue curve (Lemma 4.24).

since $e^{-r}\nu = e^P\lambda$. As $\mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) \le \mathrm{KL}(\nu \,\|\, \nu_P)$ for every such P (Lemma 4.24), this yields the *variance form* of the bound:

$$\mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) \;\le\; \frac{1}{2} \inf_P \int (g - P)^2 \left(\mathrm{d}\nu + e^P \,\mathrm{d}\lambda \right), \tag{4.10}$$

the infimum being over the polynomials P of degree $\le m$ with $\int (g - P)\,\mathrm{d}\nu = 0$. The measure $e^P\lambda$ has total mass $\int e^P \,\mathrm{d}\lambda = \mathbb{E}_\nu[e^{-r}] \ge 1$, close to one when the approximation is good, so that the right-hand side is essentially half the mean square of the error $g - P$ under ν and under ν_P. It is less elegant than $\frac{1}{2}E_m(g)^2$, but it involves the squared error only where ν and ν_P put their mass: the polynomial may be a poor approximation of g far in the tails without harm. This matters for laws that concentrate on a small part of their support. For $\nu = \mathrm{Bin}(400, 0.3)$ and $m = 4$, the log-density falls from about -3 at the mode to -143 at the ends of the support and no polynomial of degree 4 follows it uniformly: $E_m(g) \approx 4.4$ and the sup-norm bound $\frac{1}{2}E_m(g)^2$ is about 10, whereas $\mathrm{KL}\big(\nu \,\big\|\, \nu^\star\big) = 1.4 \cdot 10^{-8}$, and (4.10) with $P = \ln p^\star$ recentred gives $2.8 \cdot 10^{-8}$. Remark 4.26 and Exercise 4.11 rest on it.

Remark 4.26 (Where the horizon hides ♦). Theorem 4.25 does not involve the horizon, and it is worth understanding why, and to what extent. Three observations.

First, the relative entropy and the reconstruction both commute with affine changes of variable, so the bound $\frac{1}{2}E_m(g)^2$ is dimensionless: the length Δ of the support, hence H, enters only as the unit in which $\mathcal{W}_1$ is expressed—never, as in Proposition 4.27, through an accumulation of errors along the recursion. The moments are propagated exactly; the quality of the reconstruction is a property of the final law alone.

Second, this property does depend on the horizon through the shape of ν, and for laws to which a central limit theorem applies the dependence is favourable: the longer the horizon, the more Gaussian the return and the better its reconstruction from a few moments. The mechanism relies on tools (Cramér's rate function, Edgeworth-type expansions, Stirling's series) that this book does not otherwise develop; what follows is a heuristic sketch, not a theorem, and Exercise 4.11 is the only case where it is carried out in full. When W_H is a sum of H i.i.d. rewards, finitely supported on a lattice whose span is the step of the counting measure, with variance σ^2, its log-density relative

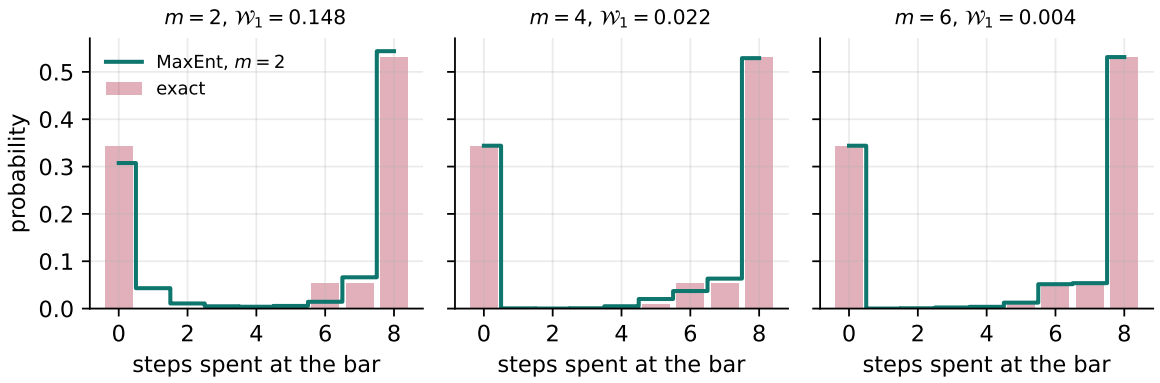


Figure 4.4. Maximum-entropy reconstruction (teal line) of the return distribution of the risky policy in the windy hike (Example 3.6; crimson bars) from its first m Chebyshev moments, with the counting measure on the exact support $\{0, \dots, 8\}$ of Proposition 4.20 as reference. Two moments already produce a U-shaped profile with the right mean and variance, but a poor balance between the two modes; four moments capture them (arrival at the bar around 8, fall into the pit at 0) up to the details of the plateau at 6–7; with $m = 6$ the two laws are indistinguishable at the scale of the figure, and $m = 8$ reproduces the nine atoms exactly.

to the counting measure, expressed in the standardized variable $x = (W_H - \mathbb{E}[W_H])/(\sigma\sqrt{H})$, is $-\frac{x^2}{2} - \frac{1}{2}\ln(2\pi H\sigma^2)$ plus a power series in x whose coefficient of x^j is of order $H^{-(j-2)/2}$: the expansion of Cramér's rate function around its minimum. This local expansion holds uniformly only away from the ends of the support (on compact subsets of the interior, in the scale W_H/H), and the tails need a separate control, as in Exercise 4.11(c); with a non-lattice law one needs instead a smooth density and the corresponding local limit theorem. A polynomial of degree m approximates it, where ν puts its mass, within $O(H^{-(m-1)/2})$, and the variance form (4.10) of the bound then gives a relative entropy of order $H^{-(m-1)}$—provided the polynomial can be kept negative on the whole support, so that the law it defines puts no mass where ν has almost none. For Bernoulli rewards, Exercise 4.11 carries this out: for even m the Taylor polynomial of $-I$ (the rate function, sign reversed) is automatically negative away from its maximum and $\mathrm{KL}\big(\nu \,\big\|\, \nu_m^\star\big) = O(H^{-(m-1)})$; for odd m it can become positive on the far side of an asymmetric support, and the rate then falls back to that of $m - 1$. Numerically—an illustration of the binomial case, not evidence for a general asymptotic statement—for $\nu = \mathrm{Bin}(H, 0.3)$ and $H = 100, 200, 400$, $\mathrm{KL}\big(\nu \,\big\|\, \nu_2^\star\big) = 6.6 \cdot 10^{-4}, 3.2 \cdot 10^{-4}, 1.6 \cdot 10^{-4}$ and $\mathrm{KL}\big(\nu \,\big\|\, \nu_4^\star\big) = 9.1 \cdot 10^{-7}, 1.1 \cdot 10^{-7}, 1.4 \cdot 10^{-8}$: a longer horizon makes the return more Gaussian and its reconstruction from a few moments better. The mean alone is the exception: $\nu_1^\star$ is a truncated geometric law, blind to the concentration of the return, and for the binomial $\mathrm{KL}\big(\nu \,\big\|\, \nu_1^\star\big)$ grows like $\frac{1}{2}\ln H$ (Exercise 4.11(a)).

Third, the theorem asks ν to charge the whole support of λ, which the lattice laws of finite MDPs need not do; this is where the extreme-value recursion of Proposition 4.20 earns its place. In Figure 4.4, λ is the counting measure on the exact support $\{0, \dots, 8\}$ of the risky return, which charges each of its nine atoms: the theorem applies, and eight moments determine the law. Had one taken the a priori support $\{0, \dots, H\} = \{0, \dots, 14\}$ instead, g would be $-\infty$ on six atoms, the bound void, and the reconstruction markedly worse: $\mathcal{W}_1 = 1.95, 0.19, 0.036$ and 0.0033 for $m = 2, 4, 6, 8$ (relative entropies 1.48, 0.17, 0.021, 0.0010), against 0.15, 0.022, 0.004 and 0 on the exact support. Laws that are not of central-limit type, like the bimodal return of the hike (a fall or an arrival), are within the theorem's reach, with an $E_m(g)$ dictated by their shape.

Comparison with projected distributional dynamic programming

The alternative strategy replaces the exact recursion on $\mathcal{M}_1(\mathbb{R})^{\mathcal{S}}$ by an approximate one on a parametric class $C \subset \mathcal{M}_1(\mathbb{R})$: given a *projection* $\Pi : \mathcal{M}_1(\mathbb{R}) \to C$, one computes $\hat{v}_{H+1} = \delta_0$ and

$$\hat{v}_t(s) \;=\; \Pi\big[(\mathcal{T}_t^\pi \hat{v}_{t+1})(s)\big], \qquad t = H, \dots, 1. \tag{4.11}$$

The class C may be the laws supported on a fixed grid of N atoms (*categorical* representation), or the laws with N equally weighted atoms at arbitrary positions (*quantile* representation); the corresponding projections are studied in Section 7.5 and used for learning in Chapter 12. Here we only need the quality of the projection, measured by $\varepsilon_\Pi \coloneqq \sup\{\mathcal{W}_1(\Pi\nu, \nu) : \nu \text{ supported in an interval of length } \Delta\}$. For the quantile representation, $\varepsilon_\Pi \le \Delta/(2N)$ (Exercise 4.12; see Dabney et al. 2018b and Bellemare et al. 2023, Chap. 5).

Proposition 4.27 (Error of projected distributional dynamic programming). Suppose that all return distributions $\nu_t^\pi(s)$ and all $(\mathcal{T}_t^\pi \hat{v}_{t+1})(s)$ are supported in an interval of length Δ. Then, for every t,

$$\overline{\mathcal{W}}_1(\hat{v}_t, \nu_t^\pi) \;\le\; (H - t + 1)\, \varepsilon_\Pi,$$

and for every statistic ψ that is L-Lipschitz on $(\mathcal{M}_1(\mathbb{R}), \mathcal{W}_1)$, $|\psi(\hat{v}_1(s)) - \psi(\nu_1^\pi(s))| \le LH\varepsilon_\Pi$.

Proof. By Lemma D.7 with $\gamma = 1$, the mixture of shifts $\mathcal{T}_t^\pi$ is non-expansive: $\overline{\mathcal{W}}_1(\mathcal{T}_t^\pi \nu, \mathcal{T}_t^\pi \nu') \le \overline{\mathcal{W}}_1(\nu, \nu')$. The triangle inequality gives

$$\overline{\mathcal{W}}_1(\hat{v}_t, \nu_t^\pi) \le \overline{\mathcal{W}}_1\big(\Pi\mathcal{T}_t^\pi \hat{v}_{t+1}, \mathcal{T}_t^\pi \hat{v}_{t+1}\big) + \overline{\mathcal{W}}_1\big(\mathcal{T}_t^\pi \hat{v}_{t+1}, \mathcal{T}_t^\pi \nu_{t+1}^\pi\big) \le \varepsilon_\Pi + \overline{\mathcal{W}}_1(\hat{v}_{t+1}, \nu_{t+1}^\pi),$$

and induction from $\hat{v}_{H+1} = \nu_{H+1}^\pi$ concludes. The second claim is the definition of Lipschitz continuity (see Corollary D.5 for utilities with Lipschitz f and for the distorted means $\nu \mapsto \int_0^1 w(u) F_\nu^{-1}(u)\, \mathrm{d}u$, of which the conditional value at risk of Chapter 6 is one). ∎

The comparison is instructive. With the quantile representation and $\Delta \le HR_{\max}$, the projected recursion with N atoms yields an error $H^2 R_{\max}/(2N)$: the projection error is paid at every step and accumulates. Propagating m moments exactly and reconstructing at the end yields, by Theorem 4.21, an error at most $6HR_{\max}/m$, one factor of H better for a comparable number of parameters, and typically far better in practice (Figure 4.4). On the other hand, the projected recursion delivers a distribution at every state and time, from which any functional can be read; the moment recursion delivers exact values of a few closed statistics, and a reconstruction where needed. Both viewpoints will be developed in the learning setting, where the recursions are driven by samples rather than by the known kernel (Chapters 12 and 13).

Distributional viewpoint

The distributional Bellman operator is a mixture of shifts. A statistic of the return can be propagated exactly if and only if it is compatible with both operations (Proposition 4.5); for utilities $\int f\, \mathrm{d}\nu$, which are automatically compatible with mixtures, compatibility with shifts forces f to be an exponential polynomial (Theorem 4.11), and a single utility is closed only if it is the mean or an exponential moment (Theorem 4.8). Exponential moments thus stand out: their Bellman equation is multiplicative, and in logarithmic form it is the classical Bellman equation with the expectation replaced by a β-mean (Corollary 4.16), interpolating between the risk-neutral value ($\beta = 0$) and the extreme values of the return ($\beta \to \pm\infty$). Everything else—quantiles, threshold probabilities, entropy—requires either the whole distribution or an

approximation; m exact moments pin the distribution down within $6\Delta/m$ in Wasserstein distance (Theorem 4.21), and the maximum-entropy law with these moments is within $\frac{1}{2}E_m(g)^2$ of a law of log-density g in relative entropy (Theorem 4.25)—in both cases without the accumulation of errors that projected recursions suffer from (Proposition 4.27), and with a dependence on the horizon that goes through the shape of the law only (Remark 4.26).

Hands-on

Session H03 (*windy hike around the pit*) is the practical counterpart of this chapter. Its first part asks for a generic backward recursion computing $\mathbb{E}[g(\sum_t h(S_t, A_t, S_{t+1}) + f(S_{H+1})) \mid S_1 = s]$ from a user-supplied one-step update $\phi(s, a, s', \cdot)$: exactly the closed recursions (4.3), to be instantiated for the mean, the second moment and the exponential moments. Its second part represents return distributions as pairs of vectors (atoms, probabilities) and implements Algorithm 3.1, from which quantiles can be read—the non-closed statistics of Example 4.6. The agent to be submitted exposes a method `setObjective` selecting a β-mean or a quantile to optimize, and is evaluated on several grids and wind strengths; the planning questions it raises are the subject of Chapter 5. The distributional value-iteration homework of session H04 (`retailStore.ipynb`) can be prepared with the moment recursions of this chapter, in their discounted form of Chapter 7.

Exercises

Exercise 4.1 (Variance: forms and non-closedness). [verify] (a) Derive (4.2) directly from the second-moment recursion of Proposition 4.1, without the law of total variance. (b) Write the state-action version $\sigma_t^2(s, a)$ of (4.2). (c) Show that for a deterministic MDP (all kernels are Dirac masses) and a deterministic policy, $\sigma_t^2 \equiv 0$, and that for a randomized policy in a deterministic MDP the "created" term of (4.2) is the variance of $a \mapsto r_t(s, a, f_t(s, a)) + V_{t+1}(f_t(s, a))$ under $\pi_t(\cdot \mid s)$.

Exercise 4.2 (Third moment and skewness). [verify] Write the recursion for the third central moment $\kappa_t(s) = \int (x - V_t(s))^3 \nu_t(s)(\mathrm{d}x)$ in terms of $(V_{t+1}, \sigma_{t+1}^2, \kappa_{t+1})$, and check that the triple (V, σ^2, κ) is Bellman closed while no proper subfamily containing κ is. Compute numerically the skewness κ/σ^3 of the two hikers' return distributions in Example 3.6 and interpret the signs.

Exercise 4.3 (Non-closed statistics). [limits] (a) Show that no finite family of quantiles $F_\nu^{-1}(\tau_1)$, $\dots$, $F_\nu^{-1}(\tau_K)$ is Bellman closed, and that the same construction shows that the conditional value at risk $\mathrm{CVaR}_\delta(\nu) = \delta^{-1} \int_0^\delta F_\nu^{-1}(u)\, \mathrm{d}u$ is not. (b) Show that no finite family of threshold probabilities $\nu([u_1, \infty)), \dots, \nu([u_K, \infty))$ is Bellman closed, although the infinite family indexed by all $u \in \mathbb{R}$ is. (c) Is the interquantile range $F_\nu^{-1}(\frac{3}{4}) - F_\nu^{-1}(\frac{1}{4})$ translation compatible? Mixture compatible? (d) Is the pair (mean, entropy) Bellman closed?

Exercise 4.4 (The examination route to Theorem 4.8). [verify] Let f be continuous and nonconstant, and suppose that ψ_f is translation compatible. For $a \in \mathbb{R}$ define $\Psi_a(\nu) = \int f(x + a)\, \nu(\mathrm{d}x)$ on finitely supported laws, and let I be the range of f (an interval). (a) Show that Ψ_a preserves convex combinations of laws, and that there is a map $H_a : I \to \mathbb{R}$ with $\Psi_a = H_a \circ \Psi_0$.

(b) Show that $H_a(\kappa y_1 + (1-\kappa)y_2) = \kappa H_a(y_1) + (1-\kappa)H_a(y_2)$ for $y_1, y_2 \in I$, $\kappa \in [0,1]$, and deduce that H_a is affine on I: $H_a(y) = \alpha(a)y + \gamma(a)$. (c) Recover (4.5) and conclude as in the text. (d) Where does the argument use that f is nonconstant?

Exercise 4.5 (Fourier statistics and exact recovery). [verify] (a) Prove directly that the pair

$$\Big(\int \cos(\omega x)\,\mathrm{d}\nu,\ \int \sin(\omega x)\,\mathrm{d}\nu\Big)$$

is Bellman closed and write $\Phi_{p,r}$ explicitly. (b) Suppose all rewards are integers in $\{0,\dots,R_{\max}\}$ and let $N = HR_{\max}+1$. Show that the N complex statistics $\hat\nu(k) = \int e^{2i\pi kx/N}\,\mathrm{d}\nu$, $k = 0,\dots,N-1$, determine $v_t^\pi(s)$ exactly, and that propagating them costs the same as Algorithm 3.1. (c) Explain why, in contrast, a small number of exponential moments $\int e^{\beta_k x}\,\mathrm{d}\nu$ at real rates β_k determines the distribution only approximately, and badly so (*hint:* the functions $e^{\beta_k x}$ are far from orthogonal).

Exercise 4.6 (Around the exponential Bellman equation). [verify] (a) Show that (4.8) reduces, as $\beta \to 0$, to $U_{\beta,t}(s) = V_t(s) + \frac{\beta}{2}\sigma_t^2(s) + o(\beta)$ with σ_t^2 given by (4.2), by expanding the β-mean of the one-step target. (b) Show that the map $u \mapsto \big(s \mapsto \frac{1}{\beta}\ln\sum_{a,s'}\pi_t(a\mid s)P_t(s'\mid s,a)e^{\beta[r_t(s,a,s')+u(s')]}\big)$ on $\mathbb{R}^{\mathcal{S}}$ is monotone, translation equivariant ($u+c \mapsto \cdot + c$) and non-expansive for the supremum norm, but not affine. (c) Prove that $Z_{\beta,t}^\pi(s)$ is, for fixed (t,s), a log-convex function of β, and that $\beta \mapsto U_{\beta,t}^\pi(s)$ is nondecreasing, with limits $\underline{x}_t(s)$ and $\overline{x}_t(s)$ of Proposition 4.20.

Exercise 4.7 (Two-point laws and the β-mean crossing). [verify] Let $\nu = (1-q)\delta_0 + q\delta_b$ and $\nu' = \delta_c$ with $0 < c < qb$. Compute $U_\beta(\nu)$ and find the crossing $\beta_\times < 0$ at which $U_{\beta_\times}(\nu) = c$. Compare with the mean–variance prediction $\beta \approx -2(qb-c)/(q(1-q)b^2)$ for $q = \frac{2}{3}$, $b = 8$, $c = 2.4$ (a caricature of Example 4.18), and explain the sign of the discrepancy by comparing the behaviors of $\beta \mapsto U_\beta(\nu)$ and of the parabola $\beta \mapsto m + \beta\sigma^2/2$ as $\beta \to -\infty$.

Exercise 4.8 (Sharpness of the moment-matching bound). [limits] Let λ be the uniform law on $[-1,1]$ and ν_n the n-point Gauss–Legendre quadrature law (atoms at the roots of the n-th Legendre polynomial, with the quadrature weights). Recall that $\int P\,\mathrm{d}\nu_n = \int P\,\mathrm{d}\lambda$ for all polynomials P of degree $\le 2n-1$. (a) Show that $\mathcal{W}_1(\lambda,\nu) \ge 1/(2n)$ for every law ν with at most n atoms. (b) Deduce that two laws on an interval of length Δ with the same first m moments can be at $\mathcal{W}_1$ distance at least $\Delta/(2(m+1))$ for every odd m (and hence, up to a factor 2, for every m), so that the rate $1/m$ of Theorem 4.21 cannot be improved.

Exercise 4.9 (Moment matching the elementary way: Bernstein polynomials). [verify] Let $\varphi : [0,1] \to \mathbb{R}$ be L-Lipschitz and, for $m \ge 1$, let

$$B_m\varphi(x) \;\coloneqq\; \sum_{k=0}^{m} \varphi\Big(\frac{k}{m}\Big)\binom{m}{k}x^k(1-x)^{m-k} \;=\; \mathbb{E}\Big[\varphi\Big(\frac{S_x}{m}\Big)\Big], \qquad S_x \sim \mathrm{Bin}(m,x),$$

be its *Bernstein polynomial*, of degree at most m. (a) Show that $|\varphi(x) - B_m\varphi(x)| \le L\sqrt{x(1-x)/m} \le L/(2\sqrt{m})$ for every $x \in [0,1]$. (b) Deduce, by the argument of Theorem 4.21 but without Jackson's theorem, that two laws supported in an interval of length Δ with the same moments of

orders $1,\dots,m$ satisfy $\mathcal{W}_1(\nu,\hat\nu)\le\Delta/\sqrt{m}$. (c) Show that the rate $1/\sqrt{m}$ is intrinsic to Bernstein polynomials: for $\varphi(x)=|x-\tfrac12|$, $B_m\varphi(\tfrac12)\ge 1/(2\sqrt{3m})$ whereas $\varphi(\tfrac12)=0$ (*hint:* for a centered variable X with finite fourth moment, $\mathbb{E}|X|\ge(\mathbb{E}X^2)^{3/2}/(\mathbb{E}X^4)^{1/2}$ by two applications of the Cauchy–Schwarz inequality; compute $\mathbb{E}X^4$ for $X=S_{1/2}-m/2$). (d) Show that for a merely continuous φ with modulus of continuity ω, $\sup_{[0,1]}|\varphi-B_m\varphi|\le\frac32\,\omega(1/\sqrt{m})$: this is Bernstein's proof of Weierstrass' theorem, and the analogue of Theorem 4.22 with the rate $1/\sqrt{m}$ in place of $1/m$.

Exercise 4.10 (The maximum-entropy dual). [verify] In the setting of Proposition 4.23 with λ the counting measure on a finite set $\{x_0,\dots,x_n\}$, $n\ge m$, and $T_j(x)=x^j$: (a) show that J is strictly convex; (b) show that J is coercive ($J(\theta)\to\infty$ as $|\theta|\to\infty$) if and only if c lies in the interior of the convex hull of $\{T(x_0),\dots,T(x_n)\}$, and deduce the existence and uniqueness of the critical point in that case; (c) show that the moment vector of any law charging at least $m+1$ points of the support is interior, and describe the minimizing sequences of J when c lies on the boundary.

Exercise 4.11 (Reconstruction of a binomial law ♦). [verify] Let $\nu=\mathrm{Bin}(H,p)$, $0<p<1$, be the law of the return of H i.i.d. Bernoulli rewards, λ the counting measure on $\{0,\dots,H\}$ (laws on $\{0,\dots,H\}$ being identified with their mass functions, $\nu(k)=\nu(\{k\})$), $\nu_m^\star$ the maximum-entropy law with the first m moments of ν, $\sigma^2=p(1-p)$, and $I(u)=u\ln\frac{u}{p}+(1-u)\ln\frac{1-u}{1-p}$ the rate function of Cramér's theorem, so that $\nu(k)\approx e^{-HI(k/H)}$. (a) Show that $\nu_1^\star(k)\propto\rho^k$ for some $\rho\in(0,1)$ when $p<\frac12$, and, using the entropy $\frac12\ln(2\pi eH\sigma^2)+o(1)$ of the binomial law, that $\mathrm{KL}\big(\nu\,\big\|\,\nu_1^\star\big)=\frac12\ln H+O(1)$. (b) Let $f(t)=(1+t)\ln(1+t)-t$ and f_m be its Taylor polynomial of degree m at 0. Check that $I(p+d)=p\,f(d/p)+(1-p)\,f\big(-\frac{d}{1-p}\big)$, that $f_m''(t)=\big(1-(-t)^{m-1}\big)/(1+t)$, and that $f_m(t)\ge t^2/4$ for all $t\ge-1$ when m is even. Deduce that the Taylor polynomial I_m of degree m of I at p satisfies $I_m(p+d)\ge d^2/(4\sigma^2)$ whenever $p+d\in[0,1]$ and m is even, and show that this fails for $p=0.1$ and $m=3$. (c) Using Stirling's series, show that for every J, uniformly for $u=k/H$ in a compact subset of $(0,1)$, $\ln\nu(k)=-H\,I(u)-\frac12\ln\big(2\pi Hu(1-u)\big)+\sum_{j=1}^{J}H^{-(2j-1)}G_j(u)+O(H^{-2J-1})$ with G_j smooth. Deduce that for even m there is a polynomial P of degree m with $|\ln\nu(k)-P(k)|\le CH^{-(m-1)/2}(1+|x|^{m+1})$ for $|u-p|\le\frac12\min(p,1-p)$, where $x=(k-Hp)/(\sigma\sqrt{H})$, and $e^{P(k)}\le CH^{-1/2}e^{-x^2/4}$ for all k. (d) Conclude from the variance form (4.9) of the bound that $\mathrm{KL}\big(\nu\,\big\|\,\nu_m^\star\big)=O(H^{-(m-1)})$ for even m, hence $O(H^{-(m-2)})$ for odd m. Check numerically, for $p=0.3$ and $p=0.1$, whether odd m do better than this, and explain what happens when $p=\frac12$.

Exercise 4.12 (Projected recursions). [limits] (a) Let Π_N map a law ν to the law with equal weights $1/N$ at its quantiles $F_\nu^{-1}((2i-1)/(2N))$, $i\in[N]$. Show that $\mathcal{W}_1(\Pi_N\nu,\nu)\le\Delta/(2N)$ when ν is supported in an interval of length Δ (*hint:* quantile coupling, Theorem D.3), with equality for $\nu=\frac{1}{2N}\sum_{i=1}^{N}(\delta_{z_i}+\delta_{z_{i+1}})$, $z_i=(i-1)\Delta/N$. (b) Consider a single state, two actions with rewards 0 and $\Delta'>0$, the uniform policy, and the projection Π_2 ($N=2$ atoms, at the quantiles of orders $\frac14$ and $\frac34$). Show that the projected recursion (4.11) is stuck at $\hat\nu_t=\frac12(\delta_0+\delta_{\Delta'})$ for every $t\le H$, whereas ν_t^π is the binomial law of $H-t+1$ trials scaled by Δ', and deduce that $\mathcal{W}_1(\hat\nu_1,\nu_1^\pi)\ge(H-1)\Delta'/2$: projection errors accumulate linearly in H, and even the mean of the projected recursion is biased.

Exercise 4.13 (Numerical study). [implement] Implement Algorithms 4.1 to 4.3 and Proposition 4.20 for the windy hike, and reproduce the numbers of Examples 4.3 and 4.18 and the reconstructions of Figure 4.4 (the recursions are exact, so the reference is exact up to rounding). Then vary the wind probability $p \in \{0.05, 0.1, 0.2\}$ and report how the crossing point $\beta_\times$ of the two β-mean curves moves. Is there a wind probability $p \in (0, 1)$ for which the safe path becomes preferable to a risk-neutral hiker ($\beta = 0$)? Determine whether this is possible, and justify the answer.

Bibliographical notes

Recursions for the variance of the return go back at least to Sobel (1982) in the discounted setting; White (1988) reviews the mean, variance and probabilistic criteria for finite MDPs, and Filar et al. (1989) study variance-penalized MDPs. Tamar et al. (2016) learn the variance of the reward-to-go by temporal-difference methods, exploiting the pair recursion of Corollary 4.2. Moment criteria for MDPs were introduced by Jaquette (1973).

The notion of Bellman closedness is due to Rowland et al. (2019), who introduce the imputation strategy and prove, in the *discounted* setting, that the only Bellman-closed finite families of utilities are spanned by the moments x^j (their Theorem 4.3, which rests on a result of Engert 1970 on translation-invariant spaces of measurable functions); Bellemare et al. (2023, Chap. 8) give a textbook account. In the undiscounted finite-horizon setting treated here the class is larger and consists of the exponential polynomials, the sufficiency of which is proved in Marthe et al. (2023). The proof of Theorem 4.8 through the functional equation (4.5) was an examination problem of the course; the linear-algebraic reduction of Lemma 4.7 shortens it, and the structure of finite-dimensional translation-invariant spaces (Lemma 4.10) is classical (Anselone and Korevaar, 1964; Engert, 1970). The error bound for projected distributional dynamic programming in the finite-horizon setting is also in Marthe et al. (2023).

The multiplicative Bellman equation for exponential utilities is the starting point of risk-sensitive control, from Howard and Matheson (1972) to Chung and Sobel (1987) and Denardo and Rothblum (1979); the change of measure of Remark 4.19 is the finite-horizon face of the Perron–Frobenius theory that governs the average-reward case (Chapter 9).

Theorem 4.21 and the maximum-entropy reconstruction from Chebyshev moments are from Bringas Miranda et al. (2026); Jackson's theorem, with the constant 6 and in its derivative form with the constant $(\pi/2)^k$, can be found in Rivlin (1969, Chap. 1) and Cheney (1966, Chap. 4), and in the general form of DeVore and Lorentz (1993). The maximum-entropy principle is due to Jaynes (1957); the geometry of the problem, including the Pythagorean identity of Lemma 4.24, and existence results are in Csiszár (1975); the approximation rates of exponential families are due to Barron and Sheu (1991), whose argument Theorem 4.25 follows, with Hoeffding's lemma in place of their second-order bound. The categorical and quantile representations are those of Bellemare et al. (2017) and Dabney et al. (2018b), discussed in Chapter 12; the quantile projection error $\Delta/(2N)$ is elementary (Exercise 4.12) and appears in Bellemare et al. (2023, Chap. 5).