CE 273
Markov Decision Processes

Lecture 5

Applications of Finite Horizon MDPs - Part I

Previously on Markov Decision Processes

Suppose there are N time steps $k = 0, 1, 2, \ldots, N-1$. For each k, define

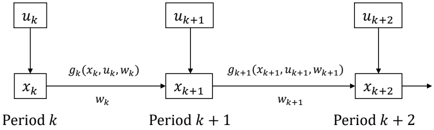
Notation	*Description*
x_k	State of the system at time k
u_k	Action/control/decision variable to be chosen at k
w_k	Disturbance, a random variable with known distribution
$f_k(x_k, u_k, w_k)$	System dynamics

The distribution of w_k may depend on x_k and u_k and is usually independent across time.

Additionally, we incur a **one-step cost** of $g_k(x_k, u_k, w_k)$ due to taking an action in a particular state. We also assume that the final state x_N results in a terminal cost of $g_N(x_N)$.

The total cost is

$$g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$$



Previously on Markov Decision Processes

The above cost is a random variable because $w_0, \ldots, w_{N-1}$ are random variables. Hence, we are typically interested in minimizing the expected total cost

$$\mathbb{E}\left\{g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)\right\}$$

Denote the set of all states at time k using S_k.Let $U_k(x_k)$ be the set of actions available at time step k and at state x_k. We say that a policy π is **admissible** if $\mu_k(x_k) \in U_k(x_k) \, \forall \, x_k \in S_k$. Let Π be the set of all admissible policies.

Think of π as the decision variable and Π as the feasible region. The objective for a given π is

$$J_\pi(x_0) = \mathbb{E}\left\{g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)\right\}$$

where expected is taken over w and states evolve according to $x_{k+1} = f_k(x_k, u_k, w_k)$. The goal is to find π^* that minimizes the above cost.

$$J^*(x_0) = J_{\pi^*}(x_0) = \min_{\pi \in \Pi} J_\pi(x_0)$$

$J^*(x_0)$ and is called the **optimal value or cost function**. Note that it is a function of the initial state just like DTMC with costs.

Previously on Markov Decision Processes

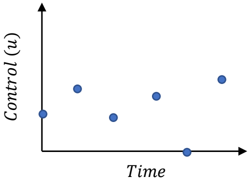


Figure: Open-loop Policy

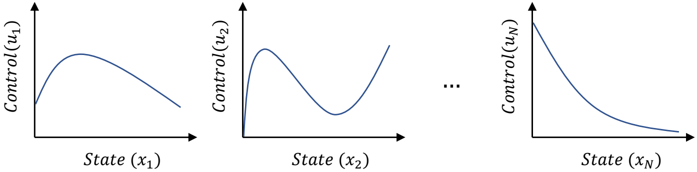


Figure: Closed-loop Polciy

Previously on Markov Decision Processes

Proposition (Principle of Optimality)

Let $\pi^* = \{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ *be an optimal policy. Consider the subproblem in which we are at* x_i *and seek the minimum cost-to-go from* i *to* N*.*

$$\mathbb{E}\left\{ g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right\}$$

The truncated policy $\{\mu_i^*, \mu_{i+1}^*, \ldots, \mu_{N-1}^*\}$ *is optimal for this subproblem.*

Theorem (DP Algorithm)

The optimal cost $J^*(x_0)$ *equals* $J_0(x_0)$ *which solves*

$$J_N(x_N) = g_N(x_N)$$

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} \mathbb{E}_{w_k}\Big\{ g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k)) \Big\}$$

$$\forall\, k = N-1, \ldots, 1, 0$$

Further, if $u_k^* = \mu_k^*(x_k)$ *minimizes the RHS of the above expression then* $\pi^* = \{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ *is optimal.*

Lecture Outline

1. DP Example
2. Inventory Control without Fixed Costs
3. Inventory Control with Fixed Costs

Lecture Outline

DP Example

DP Example

Problem Statement

Imagine an instance in which back orders are disallowed. Hence, the dynamics can be expressed as

$$x_{k+1} = f(x_k, u_k, w_k) = [x_k + u_k - w_k]^+$$

Let w_k represent the demand and assume that it can either be 1 or 2 with equal probability.

Assume that the cost of ordering one unit of the item is 1 and penalty for holding and falling short is $(x_k + u_k - w_k)^2$. Thus, the one-step costs are

$$g_k(x_k, u_k, w_k) = u_k + (x_k + u_k - w_k)^2$$

Additionally, suppose that only a maximum of 2 items can be stored. Assuming, $N = 2$ and $g_N(x_N) = 0$ for all x_N find the optimum value functions and policies.

DP Example

Backward Induction

Table: Terminal Value Functions

x_2	$J_2^*(x_2)$
0	0
1	0
2	0

For $k = 1$ and 0 solve,

$$J_k(x_k) = \min_{u_k \leq 2 - x_k} \mathbb{E}\left\{ u_k + (x_k + u_k - w_k)^2 + J_{k+1}\left([x_k + u_k - w_k]^+\right) \right\}$$

DP Example

Backward Induction

For $k = 1$, the value functions can be written as

$$J_1(0) = \min_{u_1 \in \{0,1,2\}} \mathbb{E}\left\{ u_1 + (u_1 - w_1)^2 + J_2\left([u_1 - w_1]^+\right) \right\} = \min\left\{5/2, 3/2, 5/2\right\}$$

$$J_1(1) = \min_{u_1 \in \{0,1\}} \mathbb{E}\left\{ u_1 + (1 + u_1 - w_1)^2 + J_2\left([1 + u_1 - w_1]^+\right) \right\} = \min\left\{1/2, 3/2\right\}$$

$$J_1(2) = \min_{u_1 \in \{0\}} \mathbb{E}\left\{ u_1 + (2 + u_1 - w_1)^2 + J_2\left([2 + u_1 - w_1]^+\right) \right\} = 1/2$$

Table: Optimal value function and policy for $k = 1$

x_1	$J_1^*(x_1)$	$\mu_1^*(x_1)$
0	3/2	1
1	1/2	0
2	1/2	0

DP Example

Backward Induction

For $k = 0$, the value functions can be written as

$$J_0(0) = \min_{u_0 \in \{0,1,2\}} \mathbb{E}\left\{ u_0 + (u_0 - w_0)^2 + J_1\left([u_0 - w_0]^+\right) \right\} = \min\left\{4, 3, 7/2\right\}$$

$$J_0(1) = \min_{u_0 \in \{0,1\}} \mathbb{E}\left\{ u_0 + (1 + u_0 - w_0)^2 + J_1\left([1 + u_0 - w_0]^+\right) \right\} = \min\left\{2, 2\right\}$$

$$J_0(2) = \min_{u_0 \in \{0\}} \mathbb{E}\left\{ u_0 + (2 + u_0 - w_0)^2 + J_1\left([2 + u_0 - w_0]^+\right) \right\} = 3/2$$

Table: Optimal value functions and policy for $k = 0$

x_0	$J_0^*(x_0)$	$\mu_0^*(x_0)$
0	3	1
1	2	0 or 1
2	3/2	0

DP Example

Interpreting Results

The solutions obtained can be used as 'look-up' tables. For instance, if we were told that $x_0 = 2$, we do not order anything in period 0.

Further, if there was a demand of 2 units in period 0, at $k = 1$, the new future state is 0. Hence, we order 1 unit in period 1.

DP Example

Observations

The DP algorithm is the only decent approach to solve finite horizon problems. However, as you might have noticed, it can become very clumsy very quickly and

- A solution in the form of multiple look-up tables is not very insightful.
- Continuous and countably infinite state spaces cannot be handled without discretization and some approximation.
- It is intractable for large problem instances. This feature is also called **the curse of dimensionality**.

DP Example

Observations

However, for certain classes of problems, the optimal policies can be shown to satisfy certain properties. This shrinks the search space and can tackle the above three issues.

These type of properties are also called structural results. Unfortunately, there's no unified theory behind it and it is best understood from multiple examples.

But the proof techniques in establishing these results almost always involves induction or recursion! In the following lecture(s), we will study few such problems.

Lecture Outline

Inventory Control without Fixed Costs

Inventory Control without Fixed Costs

Introduction

Consider the inventory model introduced in the last class. x_k and u_k denote the inventory at the beginning of period k and order quantity in k.

Assume that the demands are bounded random variables w_k, with some known probability distributions, and are independent across time. The state/demand/control can be continuous or countably infinite.

Suppose back orders are allowed. Hence,

$$x_{k+1} = f_k(x_k, u_k, w_k) = x_k + u_k - w_k$$

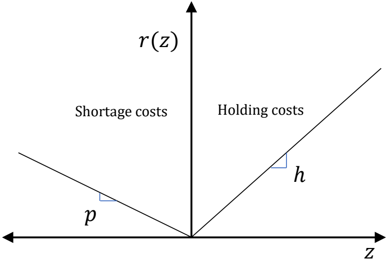
As before assume that unit cost of procuring is c but the holding/shortage costs in period k are given by

$$r(x_k + u_k - w_k) = p \max(0, -x_k - u_k + w_k) + h \max(0, x_k + u_k - w_k)$$

Inventory Control without Fixed Costs

Introduction

$$r(x_k + u_k - w_k) = p \max(0, -x_k - u_k + w_k) + h \max(0, x_k + u_k - w_k)$$



In fact, for the discussion that follows, r can be any convex function that grows larger as its arguments tend to $\pm\infty$.

Inventory Control without Fixed Costs

Optimality Conditions

The Bellman equations can be written as

$$J_N(x_N) = 0$$

$$\begin{aligned} J_k(x_k) &= \min_{u_k \geq 0} \mathbb{E}\left\{ g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k)) \right\} \\ &= \min_{u_k \geq 0} \mathbb{E}\left\{ cu_k + r(x_k + u_k - w_k) + J_{k+1}(x_k + u_k - w_k) \right\} \\ &= \min_{u_k \geq 0} \left\{ cu_k + \mathbb{E} r(x_k + u_k - w_k) + \mathbb{E} J_{k+1}(x_k + u_k - w_k) \right\} \\ &= \min_{u_k \geq 0} \left\{ cu_k + H_k(x_k + u_k) + \mathbb{E} J_{k+1}(x_k + u_k - w_k) \right\} \end{aligned}$$

where $H_k(x_k + u_k) = \mathbb{E} r(x_k + u_k - w_k)$. The function to be minimized in the above equation depends on x_k. Let $y_k = x_k + u_k$. The trick here is move x_k from the objective to the constraints.

Inventory Control without Fixed Costs

Optimality Conditions

$$J_k(x_k) = -cx_k + \min_{u_k \geq 0} \left\{ c(x_k + u_k) + H_k(x_k + u_k) + \mathbb{E}J_{k+1}(x_k + u_k - w_k) \right\}$$

$$= -cx_k + \min_{y_x \geq x_k} \left\{ cy_k + H_k(y_k) + \mathbb{E}J_{k+1}(y_k - w_k) \right\}$$

Let $G_k(y_k) = cy_k + H_k(y_k) + \mathbb{E}J_{k+1}(y_k - w_k)$. $J_k(x_k)$ can be written as

$$J_k(x_k) = -cx_k + \min_{y_k \geq x_k} G_k(y_k)$$

It turns out that $G_k(.)$ is convex for all k! (We will show this shortly.)

Definition (Convex Function)

A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if $\forall\, x, y \in X, \lambda \in [0, 1]$,

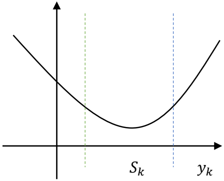
$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

Suppose the unconstrained minimum of $G_k(y_k)$ occurs at S_k. What can we say about $\min_{y_k \geq x_k} G_k(y_k)$?

$$\min_{y_k \geq x_k} G_k(y_k) = \begin{cases} G_k(S_k) & \text{if } x_k \leq S_k \\ G_k(x_k) & \text{otherwise} \end{cases}$$

Inventory Control without Fixed Costs

Optimal Solution



Thus, the optimal value function is

$$J_k(x_k) = \begin{cases} -cx_k + G_k(S_k) & \text{if } x_k \leq S_k \\ H_k(x_k) + \mathbb{E}J_{k+1}(x_k - w_k) & \text{otherwise} \end{cases}$$

Note that now we've got rid of the min operator in the above expression.

The policy is constructed based on the control values at which the minimum occurs. Recall that $y_k = x_k + u_k$. Thus, when $x_k \leq S_k$, $y_k^* = S_k$ and when $x_k > S_k$, $y_k = x_k$,

$$\mu_k^*(x_k) = \begin{cases} S_k - x_k & \text{if } x_k \leq S_k \\ 0 & \text{otherwise} \end{cases}$$

Inventory Control without Fixed Costs

Convexity Proof

We've assumed that unconstrained $G_k(.)$ has a minimum. This may not always be true for any convex function. Why?

- In such cases, we can replace the min operator with inf. In fact, when dealing with open domains, we will have to do something similar in the DP algorithm as well.
- Alternately, we can prove that $G_k(y) \to \infty$ as $|y| \to \infty$. We will ignore these details, but they are not difficult to take care of.

Let's now show that $G_k(.)$ is convex for all k using recursion.

Proposition

Let $g(y) = \mathbb{E}f(X, y)$. If f is convex in y for all realizations of X and $-\infty < \mathbb{E}f(X, y) < \infty$ for all y, $g(y)$ is convex.

Inventory Control without Fixed Costs

Optimality Conditions

Recall that $H_k(y_k) = \mathbb{E}r(y_k - w_k)$. The assumed r function is convex, and hence H_k is convex.

Theorem

The functions $G_k(.)$ and $J_k(.)$ are convex.

Proof.

Since $J_N(x_N) = 0$, it is convex. From the definition of $G_k(.)$,

$$G_{N-1}(y) = cy + H_{N-1}(y) + \mathbb{E}J_N(y - w)$$

The RHS is a sum of convex functions, hence $G_{N-1}(.)$ is convex. Now consider,

$$J_{N-1}(x_{N-1}) = \begin{cases} -cx_{N-1} + G_{N-1}(S_{N-1}) & \text{if } x_{N-1} \leq S_{N-1} \\ H_{N-1}(x_{N-1}) + \mathbb{E}J_N(x_{N-1} - w_{N-1}) & \text{otherwise} \end{cases}$$

Since $G_{N-1}(.)$ is convex, $J_{N-1}(.)$ is convex. We can proceed backwards in a similar fashion to show that the theorem is true for all k. ■

Inventory Control without Fixed Costs

Advantages

How do these structural results help?

- First, instead of finding the optimal policy functions $\mu_k(.)$ for every k, our problem reduces to searching for scalars $S_0, S_1, \ldots, S_{N-1}$ much like open-loop optimization.
- Second, we no longer need elaborate look-up tables! They are also very easy to convey to a decision maker.

The type of policies we obtained are also called **threshold policies** or **control limit policies**.

Lecture Outline

Inventory Control with Fixed Costs

Inventory Control with Fixed Costs

Introduction

Ordering new stock often involves fixed costs (for e.g., because of transportation). This can be modeled by assuming that the cost of ordering u units is $C(u)$ given by

$$C(u) = \begin{cases} K + cu & \text{if } u > 0 \\ 0 & \text{otherwise} \end{cases}$$

Inventory Control with Fixed Costs

Introduction

The Bellman equations in this case can be written as

$$J_N(x_N) = 0$$

$$\begin{aligned}
J_k(x_k) &= \min_{u_k \geq 0} \mathbb{E}\left\{ g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k)) \right\} \\
&= \min_{u_k \geq 0} \mathbb{E}\left\{ C(u_k) + r(x_k + u_k - w_k) + J_{k+1}(x_k + u_k - w_k) \right\} \\
&= \min_{u_k \geq 0} \left\{ C(u_k) + \mathbb{E}r(x_k + u_k - w_k) + \mathbb{E}J_{k+1}(x_k + u_k - w_k) \right\} \\
&= \min_{u_k \geq 0} \left\{ C(u_k) + H_k(x_k + u_k) + \mathbb{E}J_{k+1}(x_k + u_k - w_k) \right\}
\end{aligned}$$

Inventory Control with Fixed Costs

Optimality Conditions

As before, assume $y_k = x_k + u_k$, and

$$G_k(y_k) = cy_k + H_k(y_k) + \mathbb{E}J_{k+1}(y_k - w_k)$$

The value functions can be rewritten as

$$J_k(x_k) = \min\Bigg\{H_k(x_k) + \mathbb{E}J_{k+1}(x_k - w_k),$$

$$\min_{u_k>0}\Big\{K + cu_k + H_k(x_k + u_k) + \mathbb{E}J_{k+1}(x_k + u_k - w_k)\Big\}\Bigg\}$$

$$J_k(x_k) = -cx_k + \min\Bigg\{cx_k + H_k(x_k) + \mathbb{E}J_{k+1}(x_k - w_k),$$

$$\min_{u_k>0}\Big\{K + c(x_k + u_k) + H_k(x_k + u_k) + \mathbb{E}J_{k+1}(x_k + u_k - w_k)\Big\}\Bigg\}$$

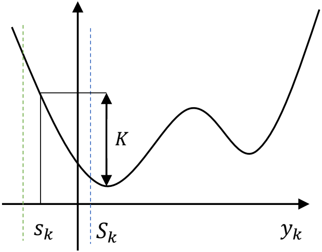
Again, we use y_k to move x_k from the objective to the constraints,

$$J_k(x_k) = -cx_k + \min\Bigg\{G_k(x_k), \min_{y_k>x_k}\Big\{K + G_k(y_k)\Big\}\Bigg\}$$

Inventory Control with Fixed Costs

Optimality Conditions

Unfortunately, $G_k(.)$ is no longer convex because of the structure of the above equation. One can however identify two distinct regimes in the optimal value functions.



Let s_k be the smallest y_k for which $G_k(y) = K + G_k(S_k)$.

Case I: When $x_k \leq s_k$,

$$\min\left\{G_k(x_k), \min_{y_k > x_k}\left\{K + G_k(y_k)\right\}\right\} = \min_{y_k > x_k}\left\{K + G_k(y_k)\right\} = K + G_k(S_k)$$

Case II: When $x_k > s_k$,

$$\min\left\{G_k(x_k), \min_{y_k > x_k}\left\{K + G_k(y_k)\right\}\right\} = G_k(x_k)$$

Inventory Control with Fixed Costs

Optimality Conditions

The optimal value functions can thus be written as

$$J_k(x_k) = \begin{cases} -cx_k + K + G_k(S_k) & \text{if } x_k \leq s_k \\ -cx_k + G_k(x_k) & \text{if } x_k > s_k \end{cases}$$

Now, let's look at the optimal policy. In Case I, minimum of the RHS occurs when $y_k = S_k$ and hence $u_k = S_k - x_k$. In Case II, the minimum occurs when $u_k = 0$. Thus, we have

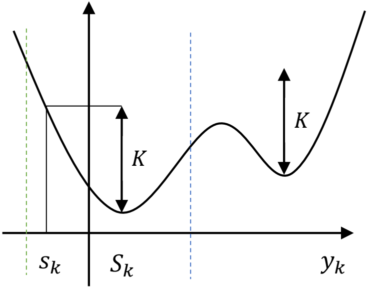
$$\mu_k^*(x_k) = \begin{cases} S_k - x_k & \text{if } x_k \leq s_k \\ 0 & \text{if } x_k > s_k \end{cases}$$

Such policies are also called (s, S) policies.

Inventory Control with Fixed Costs

Optimality Conditions

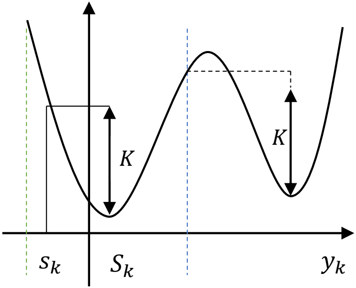
What happens to Case II, when x_k moves to the right?



Inventory Control with Fixed Costs

K-Convexity

What if $G_k(.)$ looks like this?



Case I can be left untouched, but Case II needs to be updated with more regimes.

Luckily, it turns out that $G_k(.)$ can never look like this!

Inventory Control with Fixed Costs

K-Convexity

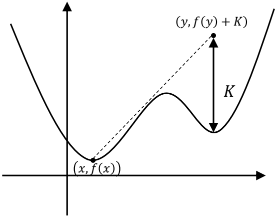
Using a recursive argument as before, it can be shown that the functions $G_k(.)$ are K-convex.

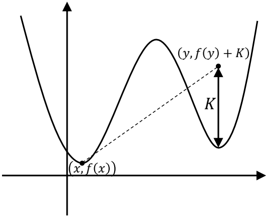
Definition (K-Convexity)

A function $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ is K-convex if $\forall\, x, y \in X, \lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda)(f(y) + K)$$

In other words, the secant between $(x, f(x))$ and $(y, f(y) + K)$ must lie above the function between $[x, y]$.

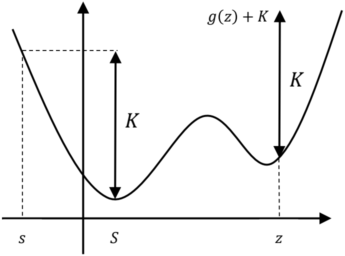




Inventory Control with Fixed Costs

K-Convexity

A key consequence of K-convexity is shown below. It proves the optimality of the (s, S) policy. (How?)



Proposition

If g is continuous K-convex function and $g(y) \to \infty$ as $|y| \to \infty$, then there exists scalars s and S with $s \leq S$ and

1. $g(S) \leq g(y)$ *for all scalars* y
2. $g(S) + K = g(s) < g(y)$, *for all* $y < s$
3. $g(y)$ *is a decreasing function on* $(-\infty, s)$
4. $g(y) \leq g(z) + K$ *for all* y, z *with* $s \leq y \leq z$

Inventory Control with Fixed Costs

K-Convexity

If the $G_k(.)$ functions are K-convex, we have concluded that (s, S) policy is optimal.

To formally show that the $G_k(.)$ functions are K-convex, we once again start with N, use the fact that $J_N(.)$ is convex and K- convex, and proceed backwards to show that $J_k(.)$ and $G_k(.)$ are K-convex for all $k = N-1, N-2, \ldots, 0$.

These structural results help reduce the problem to finding two optimal scalars for every time step instead of optimal functions.

Inventory Control with Fixed Costs

Historical Notes

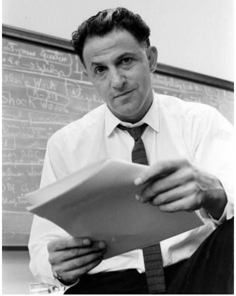
Inventory control was one of the earliest applications of MDP. The results we saw today are due to Herbert Scarf.

- Scarf, Herbert (1959). The Optimality of (S, s) Policies for the Dynamic Inventory Proble. Technical Report No. 11 (NR-047-019). Prepared for Office of Naval Research.
- Arrow, K. J., Harris, T., & Marschak, J. (1951). Optimal inventory policy. Econometrica: Journal of the Econometric Society, 250-272.

Inventory Control with Fixed Costs

Historical Notes

In fact, Markov (1856-1922) had no role to play in the development of MDPs! Richard Bellman is credited for the developed dynamic programming during his time at RAND.



- Bellman, R. (1957). A Markovian decision process. Journal of Mathematics and Mechanics, 679-684.

Inventory Control with Fixed Costs

Historical Notes

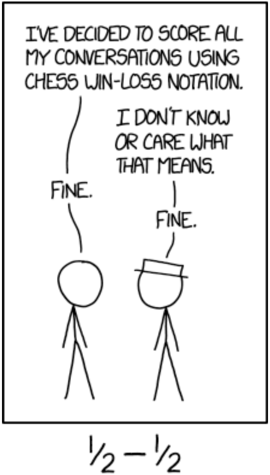
However, the core ideas were already in use in many fields. For instance, Pierre Masse, a French engineer had developed similar mathematical models for water resource management and Lloyd Shapley had some ideas in his seminal paper on Stochastic Games.

- Massé, P. (1944). Application des probabilités en chaâne á l'hydrologie statistique et au jeu des réservoirs. Journal de la société francaise de statistique, 85, 204-219.
- Shapley, L. S. (1953). Stochastic games. Proceedings of the national academy of sciences, 39(10), 1095-1100.

Ronald Howard also made fundamental contributions (especially on policy iteration methods) in his famous book "Dynamic Programming and Markov Processes" independently around the same time as Bellman.

- Howard, R. A. (2002). Comments on the origin and application of Markov decision processes. Operations Research, 50(1), 100-102.

Your Moment of Zen


I'VE DECIDED TO SCORE ALL MY CONVERSATIONS USING CHESS WIN-LOSS NOTATION.
I DON'T KNOW OR CARE WHAT THAT MEANS.
FINE.
FINE.
½ – ½