# Cancer Screening Decision Making

An Introduction of Markov Decision Process and its Applications

Yu Ding

May 19, 2025

# Schedule

| | |
|---|---|
| • Optimization 1 (ADMM) | Oct 24, 2024 |
| • Statistical Modeling 1 (Mixture Models) | Nov 11, 2024 |
| • Optimization 2 (MDP) | May 27, 2025 |
| • Statistical Modeling 2 (Graphical Model) | … |
| … | |

# Decision Science

- Decision science is an interdisciplinary field devoted to understanding and improving how individuals and organizations make choices—especially under uncertainty and complexity.

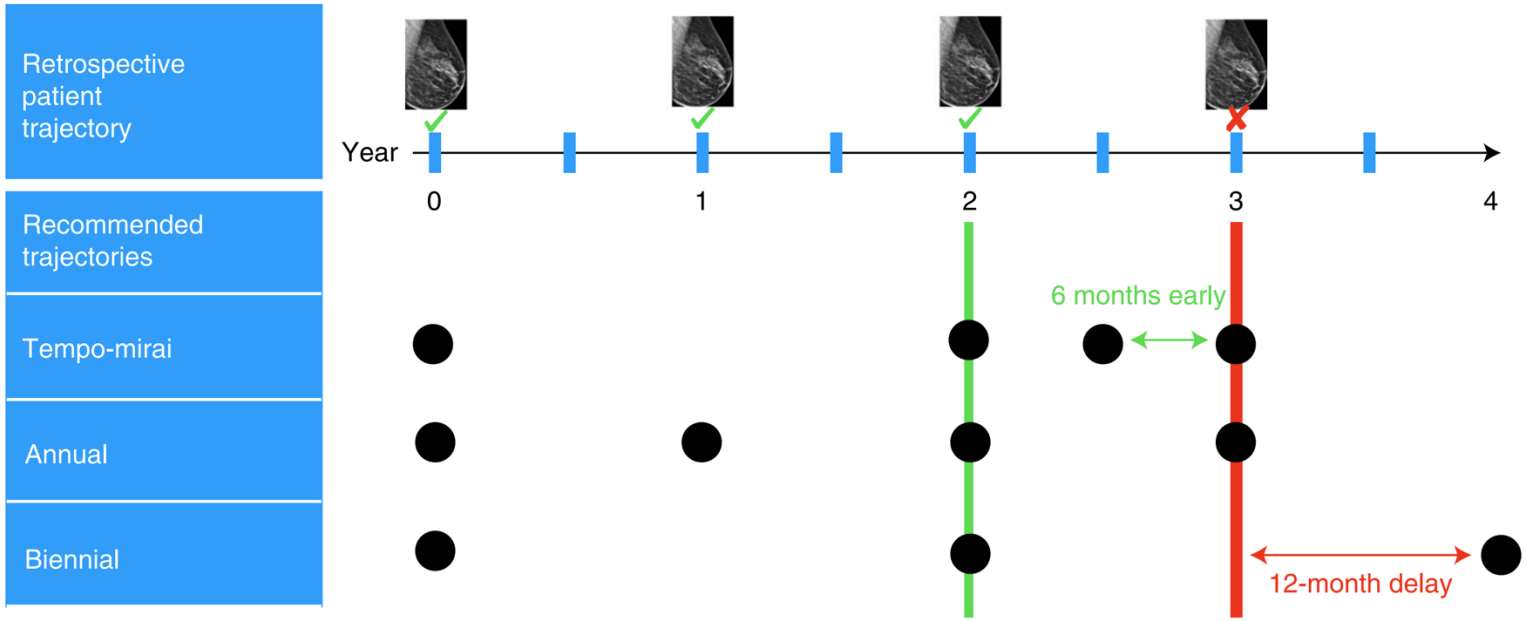
  -- ChatGPT

- Methods for decisions involving trade-offs among conflicting criteria:
  - Markov Decision Process
  - Game theory
  - Queue theory
  - …

# Content

- Literature review
- ❑ An introduction example: news vendor problem
- Markov decision process (MDP)
  - Problem formulation
  - ❑ A multi-system maintenance problem with social equity
- Partially observable Markov decision process (POMDP)
  - Problem formulation
  - ❑ Cancer screening decision making: an LFS example
  - Reinforcement learning
  - LLM-based personal healthcare advisor

# Breast cancer screening with image information


Reinforcement learning-based framework for personalized screening based on an image-based artificial intelligence risk model. Balancing early detection benefit and screening cost.

Con: MDP but not a POMDP approach, meaning the risk model itself is not personalized.

# Disease screening with chronic information

THE WALL STREET JOURNAL.

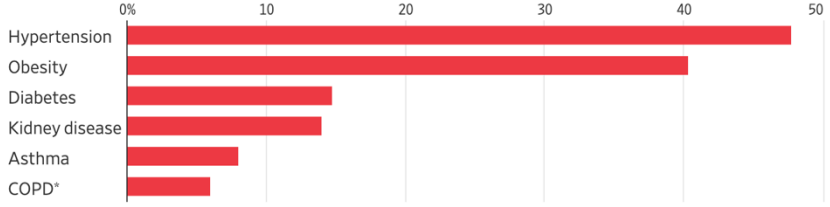
HEALTH | WELLNESS

## How Chronic Disease Became the Biggest Scourge in American Health

Americans live shorter and sicker lives than people in other high-income countries

1054 Gift unlocked article Listen (4 min)

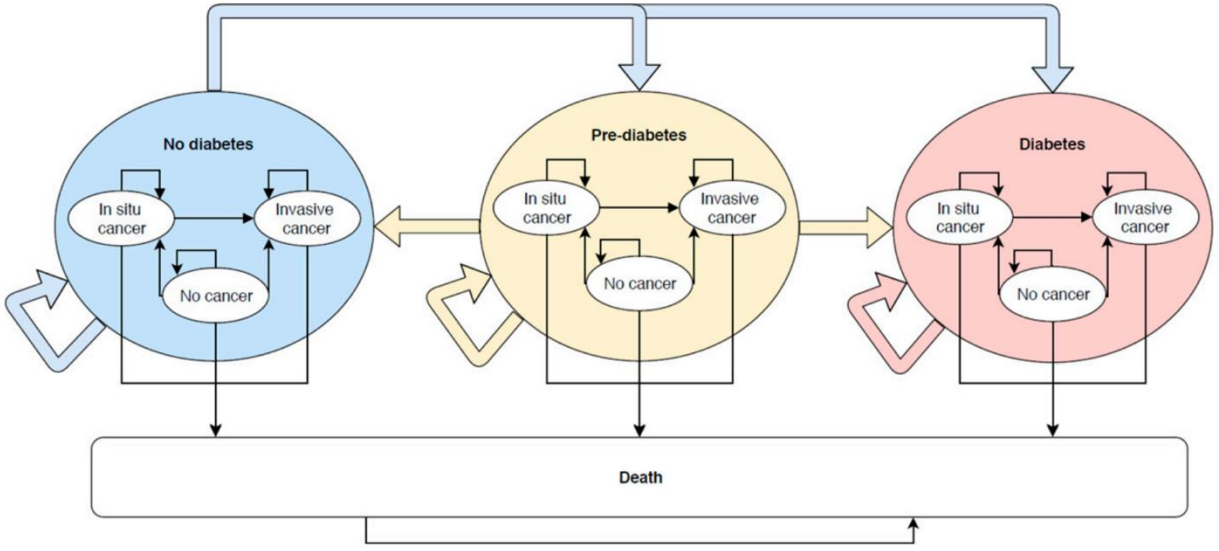
**Percentage of U.S. adults with common chronic conditions**




*Chronic obstructive pulmonary disease. Note: For most recent years available.
Source: U.S. Centers for Disease Control and Prevention

*By Brianna Abbott* Follow | *Graphics by Josh Ulick* Follow

*May 14, 2025 9:00 pm ET*




informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 69, No. 1, January 2023, pp. 260–282
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

## Personalized Disease Screening Decisions Considering a Chronic Condition

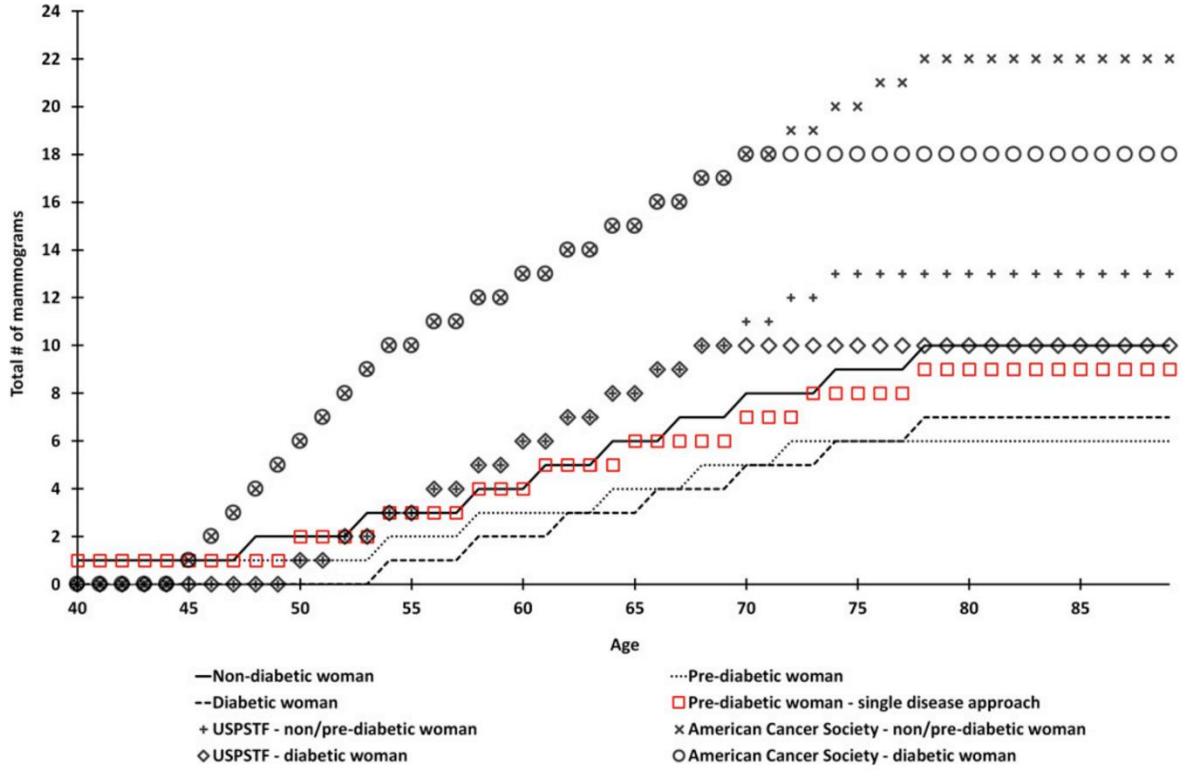
**Ali Hajjar,[a,b] Oguzhan Alagoz[c,*]**

[a] Harvard Medical School, Boston, Massachusetts 02115; [b] Institute for Technology Assessment, Massachusetts General Hospital, Boston, Massachusetts 02114; [c] Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53705
*Corresponding author
**Contact:** ahjaar@mgh.harvard.edu, https://orcid.org/0000-0003-4654-9798 (AH); alagoz@engr.wisc.edu, https://orcid.org/0000-0002-5133-1382 (OA)

Partially observed Markov decision process approach, updating belief state by Bayesian method.

Con: Only one cancer type in consideration

# Disease screening with chronic information




informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 69, No. 1, January 2023, pp. 260–282
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

## Personalized Disease Screening Decisions Considering a Chronic Condition

**Ali Hajjar,[a,b] Oguzhan Alagoz[c,*]**

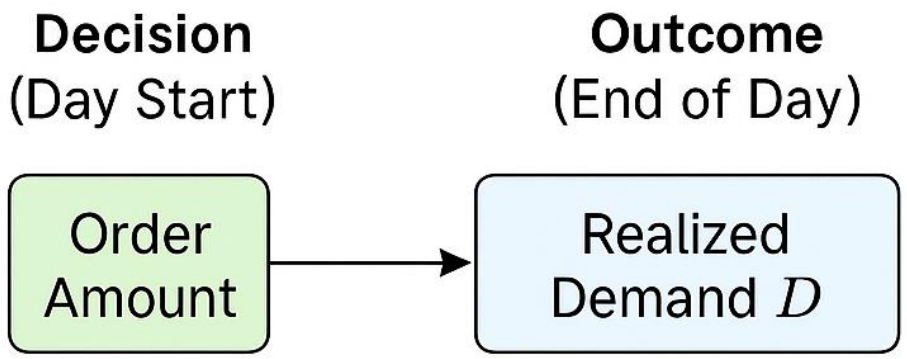
[a] Harvard Medical School, Boston, Massachusetts 02115; [b] Institute for Technology Assessment, Massachusetts General Hospital, Boston, Massachusetts 02114; [c] Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53705
*Corresponding author

**Contact:** ahjaar@mgh.harvard.edu, https://orcid.org/0000-0003-4654-9798 (AH); alagoz@engr.wisc.edu, https://orcid.org/0000-0002-5133-1382 (OA)

# Content

- Literature review
- ❑ An introduction example: news vendor problem
- Markov decision process (MDP)
  - Problem formulation
  - ❑ A multi-system maintenance problem with social equity
- Partially observable Markov decision process (POMDP)
  - Problem formulation
  - ❑ Cancer screening decision making: an LFS example
  - Reinforcement learning
  - LLM based personal healthcare advisor

# News vendor problem: one period




$a$: order quantity
$D$: random demand
$c$: purchase price per unit
$p$: selling price per unit
$s$: salvage price per unit

Reward (Profit) function:

$$R(a, D) = \underbrace{p \min(a, D)}_{\text{Sales Revenue}} \underbrace{- ca}_{\text{Cost}} + \underbrace{s \max(a - D, 0)}_{\text{Salvage revenue}}$$

**Goal**: $\max_{a \geq 0} \mathbb{E}_D [R(a, D)]$

# News vendor problem: one period

**Goal**:

$$\max_{a \geq 0} \mathbb{E}_D \left[ R(a, D) \right] = \max_{a \geq 0} \mathbb{E}_D \left[ p \min(a, D) - ca + s \max(a - D, 0) \right]$$

Assume demand $D \sim F(\cdot)$,

$$\mathbb{E}_D \left[ R(a, D) \right] = \int_{\mathcal{D}} R(a, D) f(D) dD$$

$$a^* = F^{-1}\left(\frac{p - c}{p - s}\right)$$

**Interpretation**

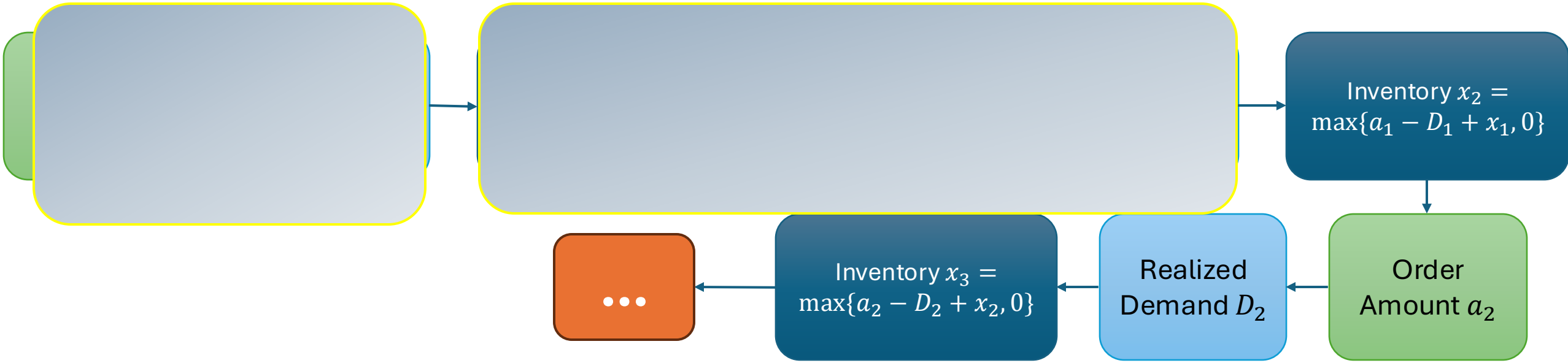
$\frac{p-c}{p-s}$ is the fraction of loss profit per stock-out

over loss profit per unsold unit

# Content

- Literature review
- ❑ An introduction example: news vendor problem
- Markov decision process (MDP)
  - Problem formulation
  - ❑ A multi-system maintenance problem with social equity
- Partially observable Markov decision process (POMDP)
  - Problem formulation
  - ❑ Cancer screening decision making: an LFS example
  - Reinforcement learning
  - LLM based personal healthcare advisor

# News Vendor Problem: infinite planning horizon




By assuming the demand distributions are identical across time, the decision can be solely made based on inventory.

Markov Property: $P(X(t+s) \in A | \{X(u) : u \leq t\}) = P(X(t+s) \in A | X(t))$

State: $x$

Action: $a$

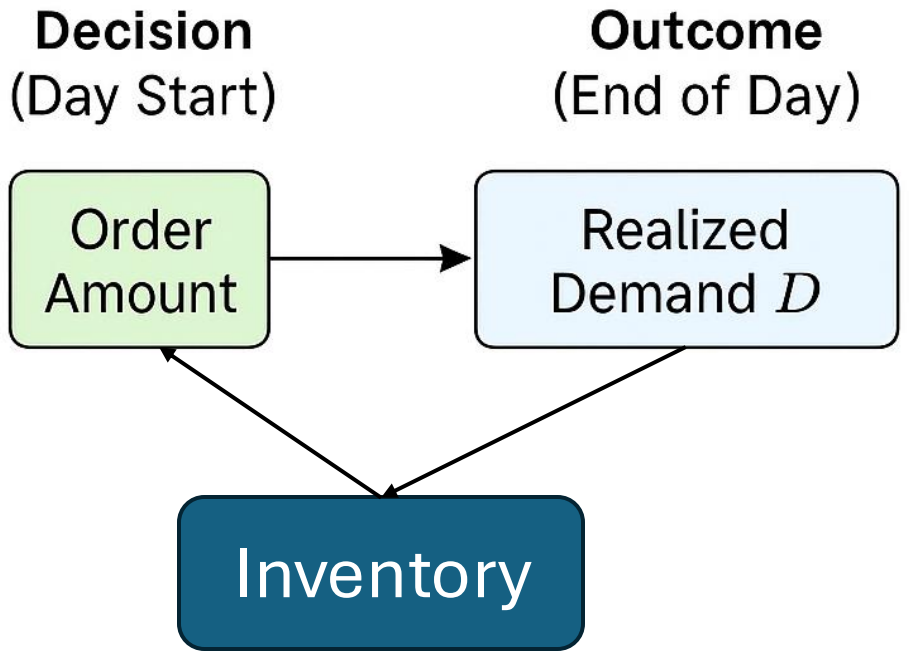
Reward: $p \min(a, D) - ca + s \max(a - D, 0)$

Transition: $\max(x + a - D, 0)$

Discount Factor: $\alpha$

# News Vendor Problem: infinite planning horizon

Markov Property:
State: $x$
Action: $a$
Reward: $p\ \min(a, D) - ca + s\ \max(a - D, 0)$
Transition: $\max(x + a - D, 0)$
Discount Factor: $\alpha$




**Goal**:

$$V^*(x) = \max_{a \geq 0} \{\mathbb{E}_D\ [R(x, a, D)] + \alpha \mathbb{E}_D [V^*(\max\{x + a - D, 0\})]\}$$

**Bellman Equation**

# News Vendor Problem: infinite planning horizon

**Bellman Equation:**

$$V^*(x) = \max_{a \geq 0}\{\mathbb{E}_D\,[R(x, a, D)] + \alpha \mathbb{E}_D[V^*(\max\{x + a - D, 0\})]\}$$

$\mathbb{E}_D\,[R(x, a, D)]$: expected immediate rewards

$\alpha \mathbb{E}_D[V^*(\max\{x + a - D, 0\})]$: discounted cumulative future rewards

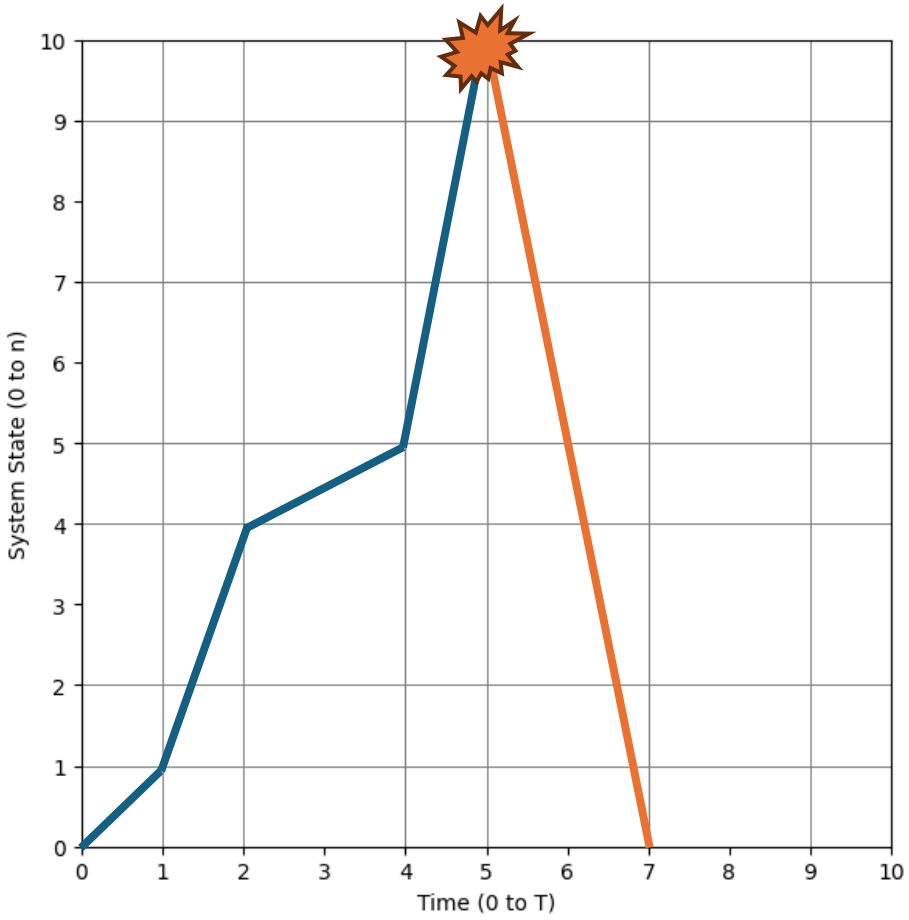
Can be solved by stochastic dynamic programming

Optimal Solution (policy function):

$$\pi^*(a|x) = \max\{S - x, 0\};$$

$S$ is a constant derived from settings represents base-stock level.

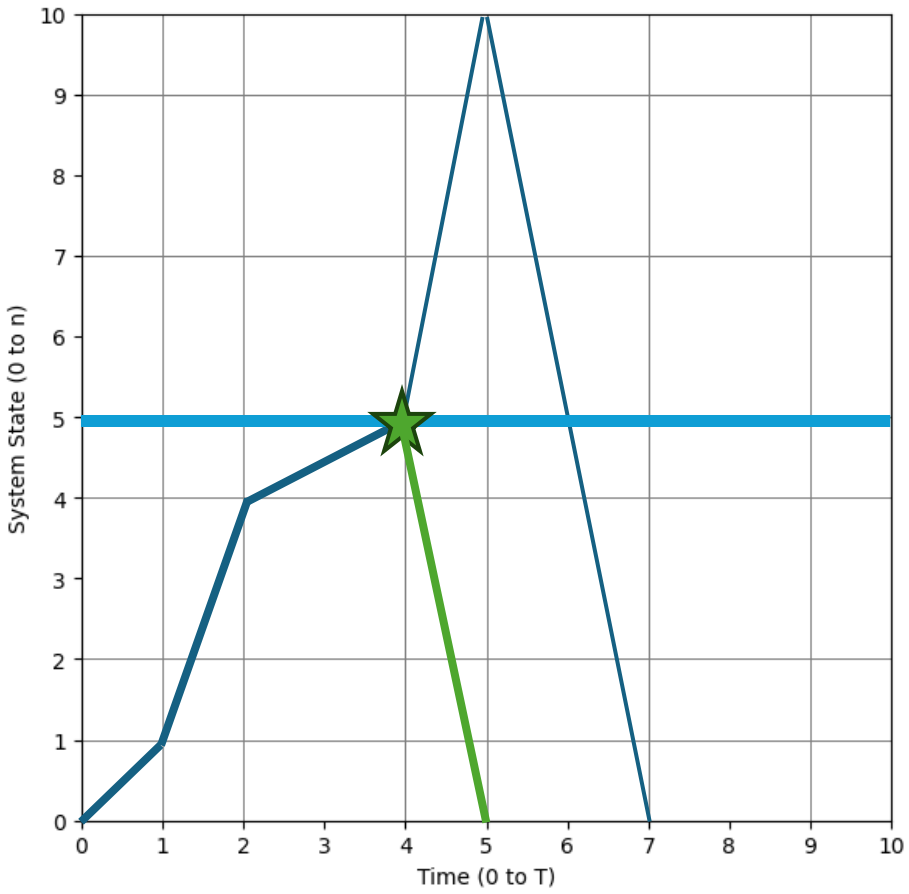
# A multi-system maintenance problem with social equity




Consider a system with state space $\{0,1,\cdots,n\}$. The system state is strictly increasing w.r.t measurement time.

If this system reach $n$, a failure occurs and generates cost $C^F$. Then the system will be reset to state 0.

# A multi-system maintenance problem with social equity




Consider a system with state space $\{0,1,\cdots,n\}$. The system state is strictly increasing w.r.t measurement time.
If this system reach $n$, a failure occurs and generates cost $C^F$. Then the system will be reset to state 0.

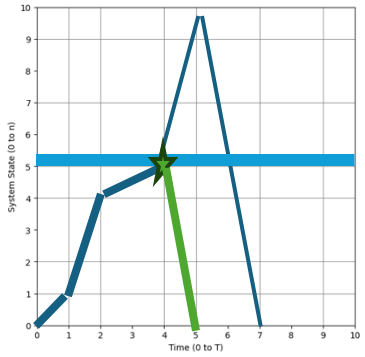
Now, we set a replacement threshold $r$. If the measurement state is larger or equal to $r$, a maintenance cost $C^R$ ($C^R \ll C^F$) is generated. Then the system will be reset to state 0.

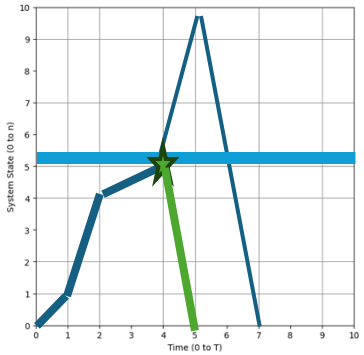
Here, the policy $r$ is constant across entire planning horizon.

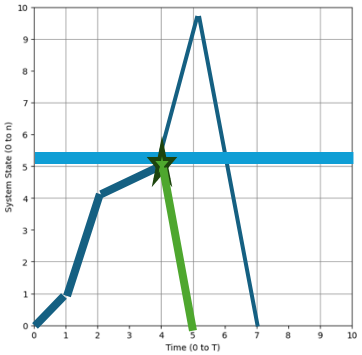
$$Q(x,r) = c(x,r) + \alpha \sum_{k=0}^{n} P_{j,k}^{r} Q(k,r)$$

$$V^* = \min_{r\in\{0,1,\cdots,n\}} \{Q(0,r)\}$$

# A multi-system maintenance problem with social equity








…… $M$ independent systems

Total cost over all systems: $F_1(\boldsymbol{r}) = \sum_{i=1}^{M} Q_i(0, r_i) = \sum_{i=1}^{M} Q_i(r_i)$

Gini index over all systems: $F_2(\boldsymbol{r}) = G(V_1(r_1), V_2(r_2), \cdots, V_M(r_M))$

$$\min_{\mathbf{r} \in \prod_{i=1}^{M} \{0,\ldots,n_i\}} \left[ F_1(\mathbf{r}),\ F_2(\mathbf{r}) \right]$$

# A multi-system maintenance problem with social equity

**Theorem 5.1** (Quasi-Convexity of $V_0(r)$)**.** *Let $C^F > C^R > 0$ and $\alpha \in (0,1)$. Suppose $P_{r,n}$ is strictly increasing in $r$. Then $V_0(r)$, viewed as a function of $r \in \{0,\dots,n\}$, can be only strictly increasing, strictly decreasing, or first decreasing and then increasing. Equivalently, $V_0(r)$ has at most one local minimum in $\{0,\dots,n\}$.*

$$\mathbf{x}(t) = (\max\{\underline{x}_1, t\}, \dots, \max\{\underline{x}_n, t\}).$$

We claim these vectors $\mathbf{x}(t)$ exactly comprise the non-dominated set *in this specific model with linear lower bounds and the given disparity measure*:

**Theorem 6.8** (All Pareto-Optimal Solutions Are Threshold Vectors *in the Sum–Disparity Model*)**.** *Under $(G,\Delta)$ on $\mathcal{F}$, the set $\{\mathbf{x}(t) : t \in [m,M]\}$ is precisely the set of non-dominated (Pareto-optimal) solutions. Specifically:*

- *No $\mathbf{y} \neq \mathbf{x}(t)$ can dominate $\mathbf{x}(t)$, meaning $\Delta(\mathbf{y}) \leq \Delta(\mathbf{x}(t))$ and $G(\mathbf{y}) \leq G(\mathbf{x}(t))$ (with at least one strict) is impossible unless $\mathbf{y} = \mathbf{x}(t)$.*
- *Any $\mathbf{x} \in \mathcal{F}$ not of that threshold form is* strictly dominated *by some $\mathbf{x}(t^\star)$.*

# Why does decision making problem pursue structural property?

1. Guide solution algorithm.
2. Ensure solutions' identification (traceability).

Most modern decision problems do not have analytical solution. Without traceability ensured by structural property, it's hard to convince audience the optimality of proposed solution (only via huge numerical test).

# Global optimal solution (Pareto Front)

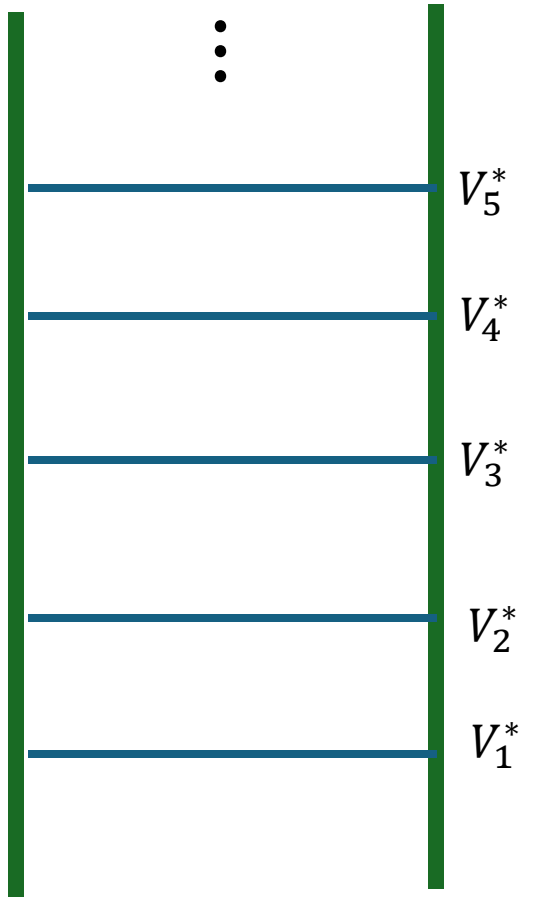
Total cost over all systems: $F_1(\boldsymbol{r}) = \sum_{i=1}^{M} Q_i(r_i)$

Gini index over all systems: $F_2(\boldsymbol{r}) = G(\sum_{i=1}^{M} Q_1(r_1), Q_2(r_2), \cdots, Q_M(r_M))$

$$\min_{\mathbf{r} \in \prod_{i=1}^{M} \{0,\ldots,n_i\}} \left[ F_1(\mathbf{r}),\ F_2(\mathbf{r}) \right]$$

$$V_i^* = \max_{r \in \{0,1,\cdots,n\}} \{Q_i(r)\}$$

Assume $V_1^* \leq V_2^* \leq \cdots V_M^*$ without loss of generality.



# Global optimal solution (Pareto Front)

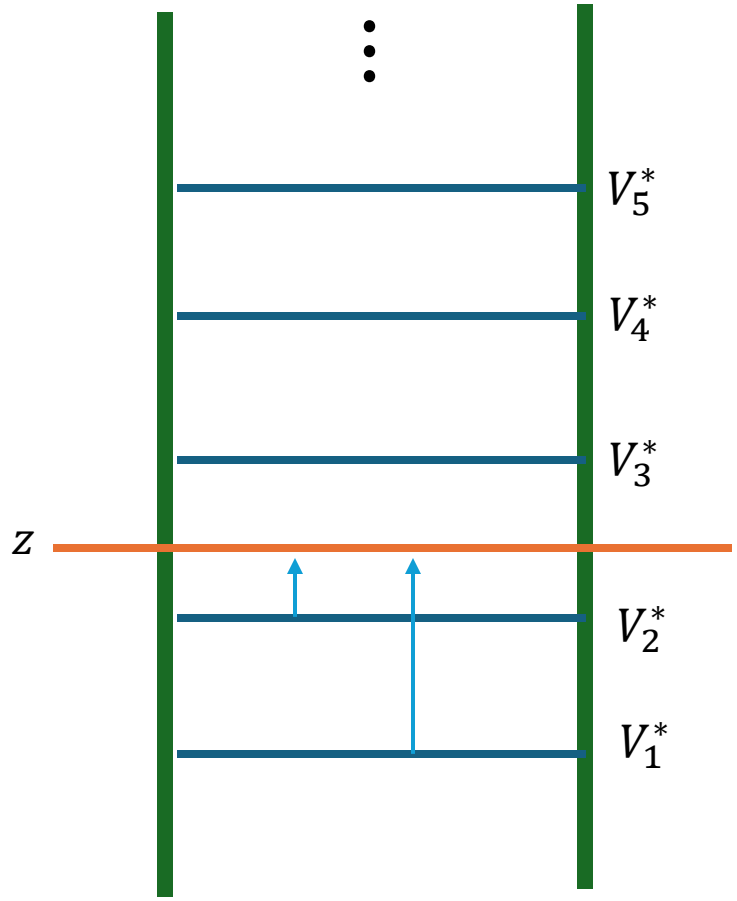
Total cost over all systems: $F_1(\boldsymbol{r}) = \sum_{i=1}^{M} Q_i(r_i)$

Gini index over all systems: $F_2(\boldsymbol{r}) = G(\sum_{i=1}^{M} Q_1(r_1), Q_2(r_2), \cdots, Q_M(r_M))$

$$\min_{\mathbf{r} \in \prod_{i=1}^{M} \{0,\ldots,n_i\}} \left[ F_1(\mathbf{r}),\ F_2(\mathbf{r}) \right]$$

$$V_i^* = \max_{r \in \{0,1,\cdots,n\}} \{Q_i(r)\}$$

Assume $V_1^* \leq V_2^* \leq \cdots V_M^*$ without loss of generality.



# Global optimal solution (Pareto Front)

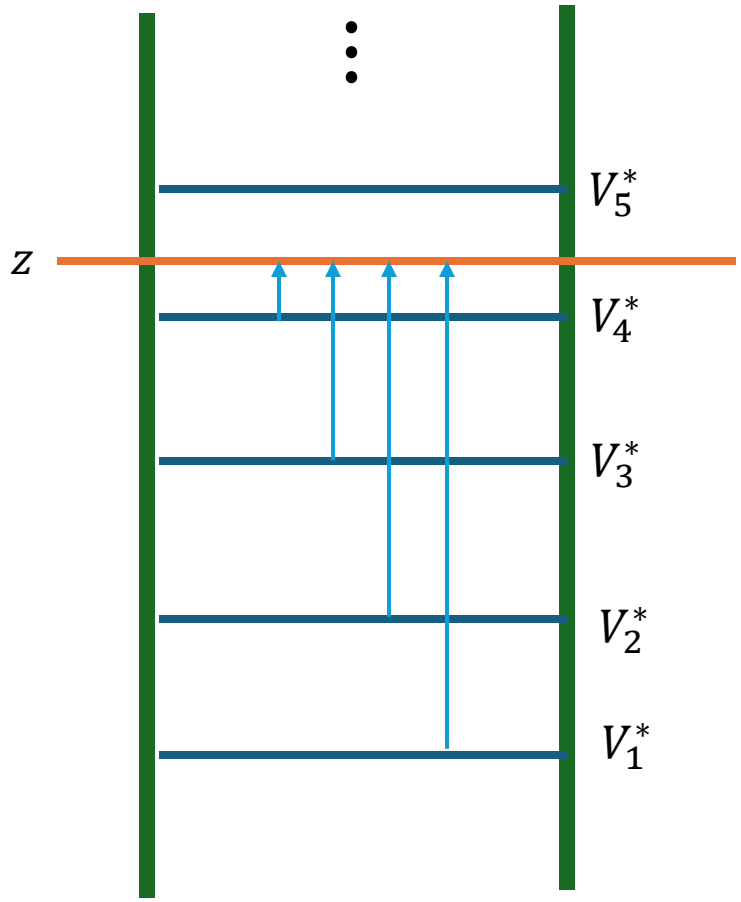
Total cost over all systems: $F_1(\boldsymbol{r}) = \sum_{i=1}^{M} Q_i(r_i)$

Gini index over all systems: $F_2(\boldsymbol{r}) = G(\sum_{i=1}^{M} Q_1(r_1), Q_2(r_2), \cdots, Q_M(r_M))$

$$\min_{\mathbf{r} \in \prod_{i=1}^{M} \{0,\ldots,n_i\}} \left[ F_1(\mathbf{r}),\ F_2(\mathbf{r}) \right]$$

$$V_i^* = \max_{r \in \{0,1,\cdots,n\}} \{Q_i(r)\}$$

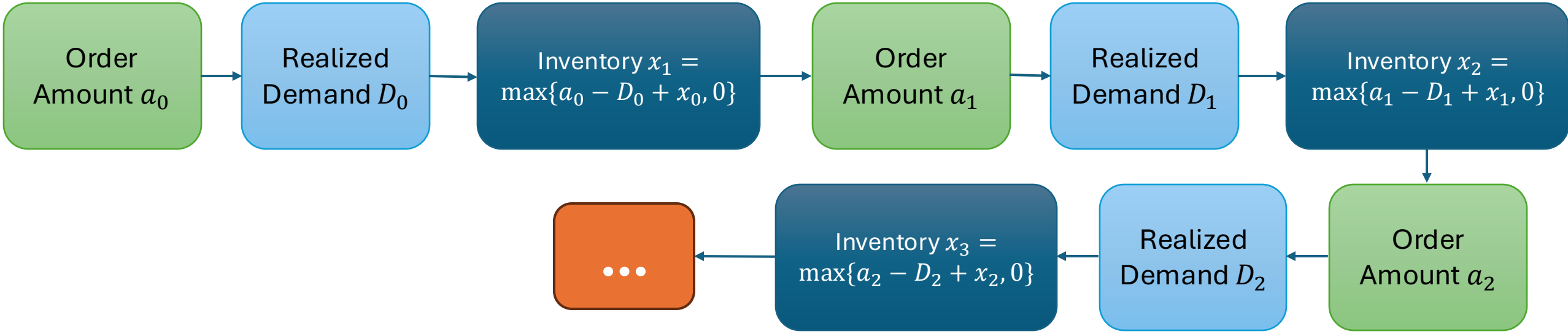
Assume $V_1^* \leq V_2^* \leq \cdots V_M^*$ without loss of generality.



# Content

- Literature review
- ❑ An introduction example: news vendor problem
- Markov decision process (MDP)
  - Problem formulation
  - ❑ A multi-system maintenance problem with social equity
- Partially observable Markov decision process (POMDP)
  - Problem formulation
  - ❑ Cancer screening decision making: an LFS example
  - Reinforcement learning
  - LLM based personal healthcare advisor

# Partially observable Markov decision process (POMDP)




1. What if the demand distributions across time are not identical?
2. **What if we have little knowledge towards demand distribution?**

We assume demand distributions are identical, but unknown at the beginning.
After each periods, we can observe the realized demand $D_t$

# Partially observable Markov decision process (POMDP)

- We assume demand distributions are **identical**, but unknown at the beginning.
- After each periods, we can observe the realized demand $d_t$.
- **Belief** $b_t$ over $\Theta$, which is the parameter space of demand function $F(\cdot)$.
- POMDP reward function for one period:

$$r(b_t, x, a) = \sum_{\theta \in \Theta} b_t(\theta) \mathbb{E}_{d \sim F_\theta(\cdot)}[p \min(a + x, d) - ca + s \max(a + x - d, 0)]$$

- Belief update: $b_{t+1}(\theta) = \frac{b_t(\theta) F_\theta(d_t)}{\sum_{\theta' \in \Theta} \theta' b_t(\theta') F_{\theta'}(d_t)}$

- POMDP Bellman Equation:

$$V^*(b_t, x) = \max_{a \geq 0} \left\{ r(b_t, x, a) + \alpha \sum_{\theta \in \Theta} b_{t+1}(\theta) \mathbb{E}_{d \sim F_\theta(\cdot)}[V^*(b_{t+1}, x')] \right\}$$

# Cancer screening decision making : an LFS example

- Equation (2) of Nam et al., 2023

  A proportional intensity model for cancer type $k$

$$\lambda_k(t|\boldsymbol{X}(t)) = \lambda_{0,k}(t)\exp(\boldsymbol{\beta}_k^T\boldsymbol{X}(t))$$

  $\boldsymbol{X}(t)$ is covariate vector as $\{G, S, D_1(t), D_2(t), \cdots, D_K(t)\}$.

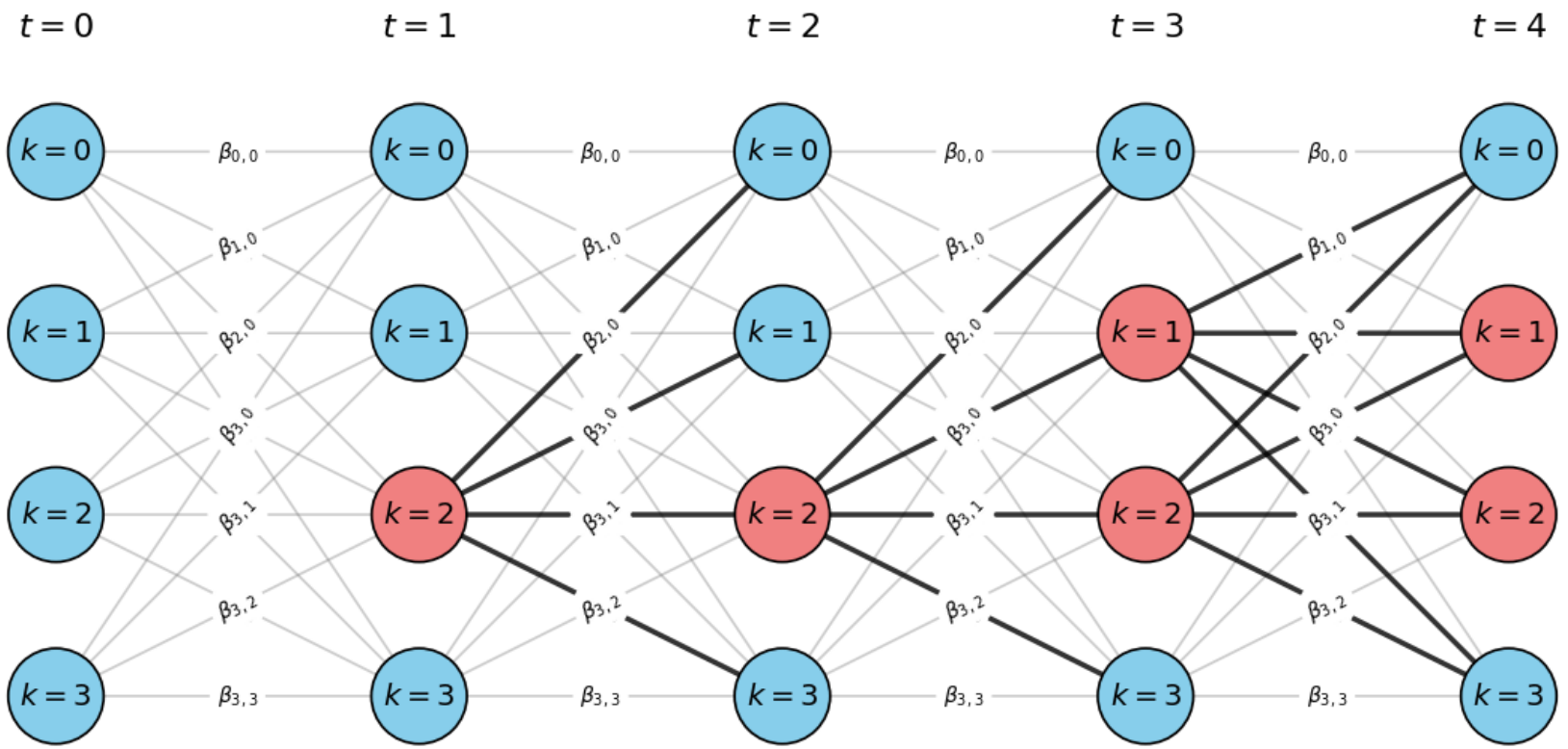
- What this work can provide:
  1. A cancer coevolution model with Markov Property.
  2. A quantitively cancer type-specific risk measurement
- What this work cannot provide:
  1. Given risk, what can people do to maximize their life gain?
  2. Does high risk necessarily mean high emergency compare with low-risk cancer types?
  3. How to transfer risk to **action**?



Nguyen, N. H., Shin, S. J., Dodd-Eaton, E. B., Ning, J., & Wang, W. (2023). Personalized risk prediction for cancer survivors: A Bayesian semi-parametric recurrent event model with competing outcomes. *bioRxiv*.

# Personalized cancer screening decision making

- A cancer coevolution model with Markov Property (Nam et al., 2023)






Nguyen, N. H., Shin, S. J., Dodd-Eaton, E. B., Ning, J., & Wang, W. (2023). Personalized risk prediction for cancer survivors: A Bayesian semi-parametric recurrent event model with competing outcomes. *bioRxiv*.

# Personalized cancer screening decision making

- Although Nam et al., 2023 is clear, dependence among cancer types will largely increase the difficulty to derive POMDP solutions (unsolvable).

- A Hierarchical Model for Multi-Tumor Screening (Hidden Markov Model)

$$\underbrace{\mathbf{z}(t)}_{\text{(observed covariates)}} \longrightarrow \underbrace{L(t)}_{\text{(latent factor, partially covariate-dependent)}} \longrightarrow \underbrace{X_k(t)}_{\text{(tumor stages, dependent on } L(t) \text{ alone)}} .$$

- $\boldsymbol{z}(t)$: education, occupation, genotype, gender, smoking, exercise habits…
- $\boldsymbol{l}(t)$: latent factors, so that $x_k(t) \perp x_{k'}(t) | \boldsymbol{l}(t)$.

# Personalized cancer screening decision making

$$\underbrace{\mathbf{z}(t)}_{\text{(observed covariates)}} \longrightarrow \underbrace{L(t)}_{\text{(latent factor, partially covariate-dependent)}} \longrightarrow \underbrace{X_k(t)}_{\text{(tumor stages, dependent on } L(t) \text{ alone)}} .$$

A Hierarchical Model for Multi-Tumor Screening

Tumor stage:

$$x_k(t) \in \{0, \cdots, M-1\}$$

0 represents no tumor;

$M-1$ represents the stage that tumor is discovered without prescheduled screening;

Action:

$$a(t) \in \{\text{NoTest}\} \cup \{\text{TestTumor}(1), \ldots, \text{TestTumor}(K)\}.$$

The screening budget (screening time per period) is fixed as a constant (can be 1).

Reward function if testing is taken for type k:

$$r\big(\text{TestTumor}(k),\, X_k(t) = m\big) = \Big[R_0\, \mathbf{1}_{\{m=0\}} + \big(1 - p_{\text{neg}}\big)\big((M-1) - m\big)\, \mathbf{1}_{\{m>0\}}\Big] - c.$$

$p_{\text{neg}}$: false negative rate; We assume there is no false positive cases.

# Personalized cancer screening decision making

$$\underbrace{\mathbf{z}(t)}_{\text{(observed covariates)}} \longrightarrow \underbrace{L(t)}_{\text{(latent factor, partially covariate-dependent)}} \longrightarrow \underbrace{X_k(t)}_{\text{(tumor stages, dependent on } L(t) \text{ alone)}} .$$

## A Hierarchical Model for Multi-Tumor Screening

### Belief: Set a prior based on a trained HMM (DL) model

For each $\ell$, define the $M \times 1$ vector

$$\mathbf{b}_k(\ell; t) = \big(b_{0,k}(\ell; t),\, b_{1,k}(\ell; t), \ldots, b_{M-1,k}(\ell; t)\big)^\top,$$

where

$$b_{m,k}(\ell; t) = \Pr\big[X_k(t) = m \mid L(t) = \ell\big].$$

Thus, each column $\mathbf{b}_k(\ell; t)$ has dimension $M$ with entries summing to 1. Our entire belief state at time $t$ is $\big(\mathbf{p}(t),\, \{\mathbf{b}_k(\ell; t)\}_{k=1,\ldots,K,\, \ell \in \mathcal{L}}\big)$.

- $\mathbf{p}(t)$ tracks the probability distribution of the latent factor $L(t)$.
- For each possible latent state $\ell$, we keep a separate distribution over $X_k(t)$ for each tumor $k$.

# Belief Update

## A Hierarchical Model for Multi-Tumor Screening

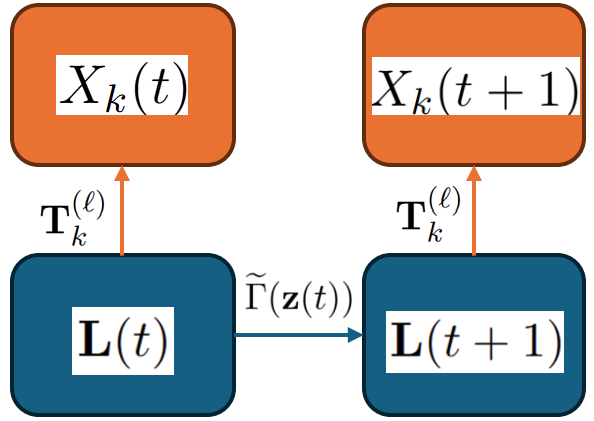
$\widetilde{\Gamma}(\mathbf{z}(t)) \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{L}|}$: Transition matrix for the joint latent vector $\mathbf{L}(t)$. Each entry

$$\widetilde{\Gamma}_{\ell,\ell'}(\mathbf{z}(t)) = \Pr\big[\mathbf{L}(t+1) = \ell' \;\big|\; \mathbf{L}(t) = \ell,\ \mathbf{z}(t)\big].$$

$\alpha_{k,m}(\ell) \in [0,1]$: Probability that tumor $k$ transitions from stage $m$ to $m+1$ given $\mathbf{L}(t) = \ell$.

$\mathbf{T}_k^{(\ell)} \in \mathbb{R}^{M\times M}$: Stage-transition matrix for tumor $k$, given $\mathbf{L}(t) = \ell$. Specifically,

$$(\mathbf{T}_k^{(\ell)})_{m,m'} = \begin{cases} \alpha_{k,m}(\ell), & \text{if } m' = m+1 \le M-1, \\ 1-\alpha_{k,m}(\ell), & \text{if } m' = m \le M-2, \\ 1, & \text{if } m = M-1 \text{ and } m' = m, \\ 0, & \text{otherwise.} \end{cases}$$




$\mathbf{O}_{k,\text{pos}}^{(\ell)}$ and $\mathbf{O}_{k,\text{neg}}^{(\ell)} \in \mathbb{R}^{M\times M}$: Diagonal observation matrices for tumor $k$. Under no false positives and uniform $p_{\text{neg}}$,

$$(\mathbf{O}_{k,\text{pos}}^{(\ell)})_{m,m} = \begin{cases} 0, & m = 0, \\ 1-p_{\text{neg}}, & m > 0, \end{cases} \qquad (\mathbf{O}_{k,\text{neg}}^{(\ell)})_{m,m} = 1 - (\mathbf{O}_{k,\text{pos}}^{(\ell)})_{m,m}.$$

# Belief Update

## A Hierarchical Model for Multi-Tumor Screening

At time $t$, we take an action

$$a(t) \in \{\text{NoTest}\} \cup \{\text{TestTumor}(1), \ldots, \text{TestTumor}(K)\}.$$

If $a(t) = \text{TestTumor}(k)$, we observe pos or neg with probabilities:

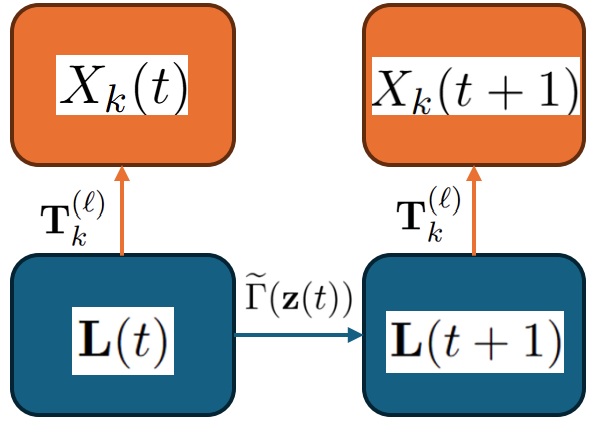
$$\Pr(\text{pos}) = \sum_{\ell=1}^{|\mathcal{L}|} p_\ell(t)\, \mathbf{1}^\top (\mathbf{O}_{k,\text{pos}}^{(\ell)}\, \mathbf{b}_k(\ell;t)),$$
$$\Pr(\text{neg}) = \sum_{\ell=1}^{|\mathcal{L}|} p_\ell(t)\, \mathbf{1}^\top (\mathbf{O}_{k,\text{neg}}^{(\ell)}\, \mathbf{b}_k(\ell;t)).$$

No false positives means if $X_k(t) = 0$, pos is impossible.

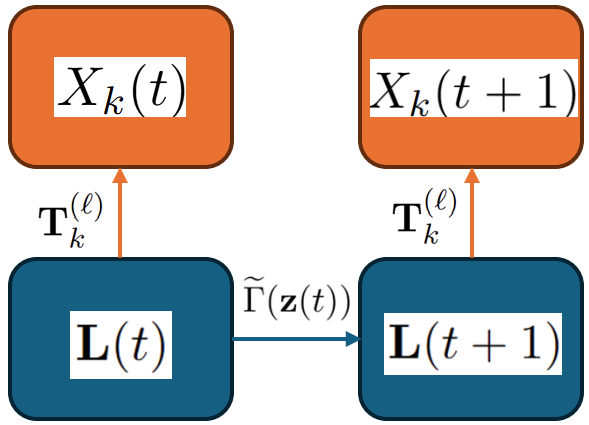
**Observation Matrices.**

$$\mathbf{O}_{k,\text{pos}}^{(\ell)} = \text{diag}\big(0,\ 1-p_{\text{neg}}, \ldots, 1-p_{\text{neg}}\big), \quad \mathbf{O}_{k,\text{neg}}^{(\ell)} = \text{diag}\big(1,\ p_{\text{neg}}, \ldots, p_{\text{neg}}\big).$$

If NoTest is chosen, no new observation is obtained at time $t$.



# Belief Update




## A Hierarchical Model for Multi-Tumor Screening

### Belief Update: Note we do not update transition probability (optional).

Suppose $a(t) = \text{TestTumor}(k)$ and we observe pos:

1. *Unnormalized posterior for* $X_k(t)$ *given* $\mathbf{L}(t) = \ell$:

$$\mathbf{O}_{k,\text{pos}}^{(\ell)}\, \mathbf{b}_k(\ell; t).$$

2. *Probability of* pos *given* $\mathbf{L}(t) = \ell$:

$$p_{\text{pos},\ell} = \mathbf{1}^\top \left( \mathbf{O}_{k,\text{pos}}^{(\ell)}\, \mathbf{b}_k(\ell; t) \right).$$

3. *Overall probability of* pos:

$$p_{\text{pos}} = \sum_{\ell=1}^{|\mathcal{L}|} p_\ell(t)\, p_{\text{pos},\ell}.$$

4. *Updated distribution over* $\mathbf{L}(t)$:

$$p_\ell^*(t) = \frac{p_\ell(t)\, p_{\text{pos},\ell}}{p_{\text{pos}}}.$$

5. *Updated distribution over* $X_k(t)$ *given* $\mathbf{L}(t) = \ell$:

$$\mathbf{b}_k^*(\ell; t) = \frac{\mathbf{O}_{k,\text{pos}}^{(\ell)}\, \mathbf{b}_k(\ell; t)}{p_{\text{pos},\ell}}.$$

An analogous update applies if neg is observed. If $a(t) = \text{NoTest}$, we have $\mathbf{p}^*(t) = \mathbf{p}(t)$ and $\mathbf{b}_k^*(\ell; t) = \mathbf{b}_k(\ell; t)$.

### 5.2 Latent Factor Transition

The next step is $\mathbf{L}(t) \to \mathbf{L}(t+1)$. We define

$$\mathbf{p}(t+1) = \widetilde{\Gamma}(\mathbf{z}(t))^\top\, \mathbf{p}^*(t).$$

### 5.3 Tumor Stage Transition

Finally, each tumor $k$ evolves conditioned on $\mathbf{L}(t+1)$. Since $\mathbf{L}(t+1) = \ell'$ occurs with probability

$$p_{\ell'}(t+1) = \sum_{\ell=1}^{|\mathcal{L}|} \left[ p_\ell^*(t)\, \widetilde{\Gamma}_{\ell,\ell'}(\mathbf{z}(t)) \right],$$

the new distribution for $X_k(t+1)$ given $\mathbf{L}(t+1) = \ell'$ is

$$\mathbf{b}_k(\ell'; t+1) = \frac{\sum_{\ell=1}^{|\mathcal{L}|} p_\ell^*(t)\, \widetilde{\Gamma}_{\ell,\ell'}(\mathbf{z}(t))\, \mathbf{T}_k^{(\ell)}\, \mathbf{b}_k^*(\ell; t)}{p_{\ell'}(t+1)}.$$

# Personalized cancer screening decision making

Expected immediate reward with belief $\big(\mathbf{p}(t),\ \{\mathbf{b}_k(\ell;t)\}_{k=1,\dots,K,\ \ell\in\mathcal{L}}\big)$.

$$\mathrm{E}[\,r_k \mid \text{belief}\,] \;=\; \sum_{\ell=1}^{|\mathcal{L}|} p_\ell(t)\,\Big[\,R_0\,b_{0,k}(\ell;t) \;+\; \sum_{m=1}^{M-1}(1-p_{\text{neg}})\,\big((M-1)-m\big)\,b_{m,k}(\ell;t)\Big] \;-\; c.$$

Bellman Equation:

Let

$$V_t\big(\mathbf{p},\ \{\mathbf{b}_k(\ell)\},\ \mathbf{z}\big)$$

denote the *optimal* value function at time $t$, with horizon $T$ or discount factor $\beta \in (0,1]$. Then:

$$V_t(\dots) = \max_{a\in\{\text{NoTest}\}\cup\{\text{TestTumor}(1),\dots,K\}} \Big\{ r(\dots,a) + \beta\,\mathbb{E}\big[V_{t+1}\big(\mathbf{p}(t{+}1),\ \{\mathbf{b}_k(\ell;t{+}1)\},\ \mathbf{z}(t{+}1)\big)\ \big|\ (\dots),\ a\big]\Big\}.$$

The expectation is taken over observation outcomes (pos or neg) if a tumor is tested.

# Personalized cancer screening decision making

Let

$$V_t\big(\mathbf{p},\,\{\mathbf{b}_k(\ell)\},\,\mathbf{z}\big)$$

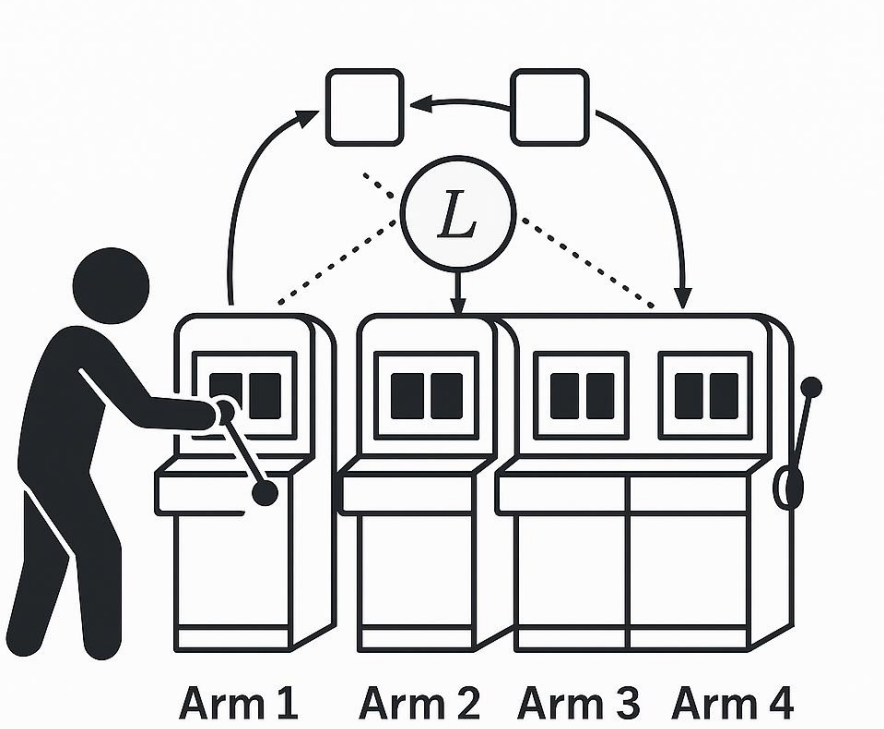
denote the *optimal* value function at time $t$, with horizon $T$ or discount factor $\beta \in (0,1]$. Then:

$$V_t(\dots) = \max_{a\in\{\text{NoTest}\}\cup\{\text{TestTumor}(1),\dots,K\}} \Big\{ r(\dots,a) + \beta\,\mathbb{E}\big[V_{t+1}\big(\mathbf{p}(t{+}1),\,\{\mathbf{b}_k(\ell;t{+}1)\},\,\mathbf{z}(t{+}1)\big)\;\big|\;(\dots),\,a\big]\Big\}.$$

- Termination:
  - Once type k is detected, then it will be dropped from decision space.
  - Once people die, the process terminated.
  - The normal planning horizon is 50 years.

# Restless Multi-Armed Bandit (MAB) Problem

- There is no analytical solution (of course).
- Solving this problem is PSPACE-hard (at least as hard as any problems in NP-hard)
- Consume more than 100GB memory for 10 cancer types, 3 latent factors, and 50 years planning horizon per person with brute force.
- This problem is a typical example of **partially independent restless multi-armed bandit (RMAB) problem** (easy to prove).



# A Whittle-Index Approach for Multi-Tumor Screening

**Definition 1** (Indexability)**.** *Consider the single-arm (single-tumor) POMDP with actions* $\{\textsc{Test}, \textsc{NoTest}\}$*. For each subsidy (or Lagrange multiplier)* $\lambda \geq 0$*, let* $\pi^\lambda$ *be an optimal policy that maximizes*

$$\mathcal{R}^\lambda(\pi) \;=\; \mathbb{E}\big[\, \textit{discounted total reward} \;+\; \lambda \times (\textit{discounted count of NoTest actions})\big].$$

*Define the* passive set $P(\lambda) \subseteq \mathcal{S}$ *of system states* $\mathcal{S}$ *as*

$$P(\lambda) \;=\; \big\{ s \in \mathcal{S} \;\big|\; \textit{the action } \textsc{NoTest} \textit{ is optimal (not worse) under } \lambda \big\}.$$

*We say the system is* indexable *if*

$$P(\lambda) \;\subseteq\; P(\lambda'), \quad \textit{for all } 0 \leq \lambda < \lambda'.$$

*In other words, the set of states for which* $\textsc{NoTest}$ *is optimal grows monotonically in* $\lambda$*.*

# A Whittle-Index Approach for Multi-Tumor Screening

Once a Whittle index $\mathrm{WI}(b_k)$ can be assigned to each tumor $k$ (where $b_k$ is the *current* belief over $(\mathbf{L}(t), X_k(t))$), the **Whittle-index scheduling policy** is:

1. At each time $t$, compute $\mathrm{WI}(b_k^t)$ for all tumors $k$.
2. *Test* the top $M$ tumors with highest indices. (If the budget is $M$.)
3. Observe the test outcomes and update each tumor's belief according to Section 5. Untested tumors also update via the "no observation" transition.
4. Repeat for $t+1$.

# A Whittle-Index Approach for Multi-Tumor Screening

- In typical RMAB theory, if each tumor is indexable (easy to prove), then the Whittle policy is Lagrangian-optimal for the relaxed average-testing constraint and generally performs near-optimally under the original per-period constraint. **This claim should be proved based on settings** (not very hard).

- Whittle-Index Approach can only provide near optimal solution (with performance guarantee).

- No time to work on the structural property so far. 

# Personalized cancer screening decision making

- Considering the data limitation, $X_k(t) \in [0,1]$: caner type $k$ risk at time $t$.

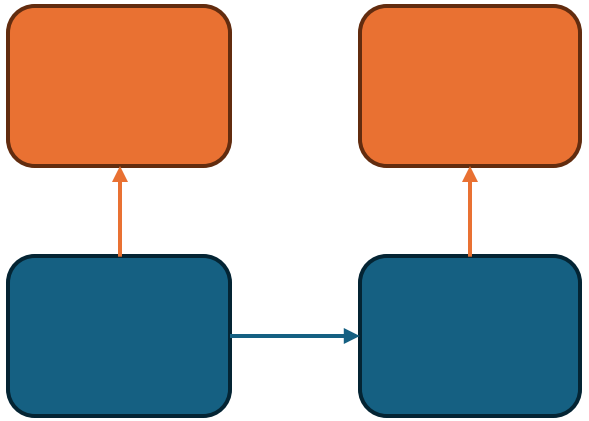
  - If we $\textsc{TestTumor}(k)$:

$$r\big(\textsc{TestTumor}(k), \mathbf{p}(t)\big) \;=\; \sum_{\ell=1}^{|\mathcal{L}|} p_\ell(t)\Big[(1 - x_{k,\ell})\, R_0 \;+\; x_{k,\ell}(1 - p_{\text{neg}})\, R_{\text{det}}\Big] \;-\; c,$$

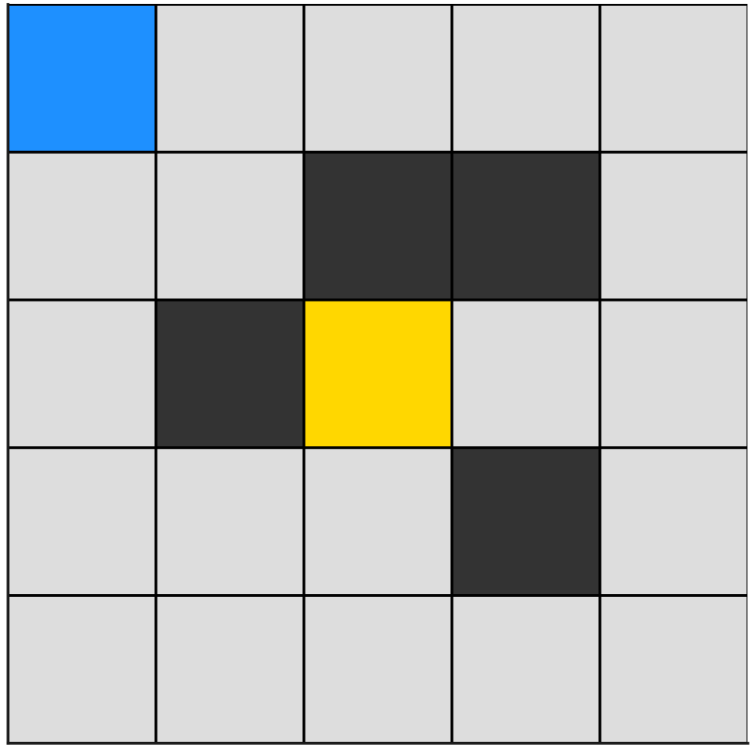
    where $R_0$ is the baseline reward for having no tumor, $R_{\text{det}}$ is the reward for detection (when the tumor is indeed present and the test is positive), and $c$ is the per-test cost.

  - If we $\textsc{NoTest}$:

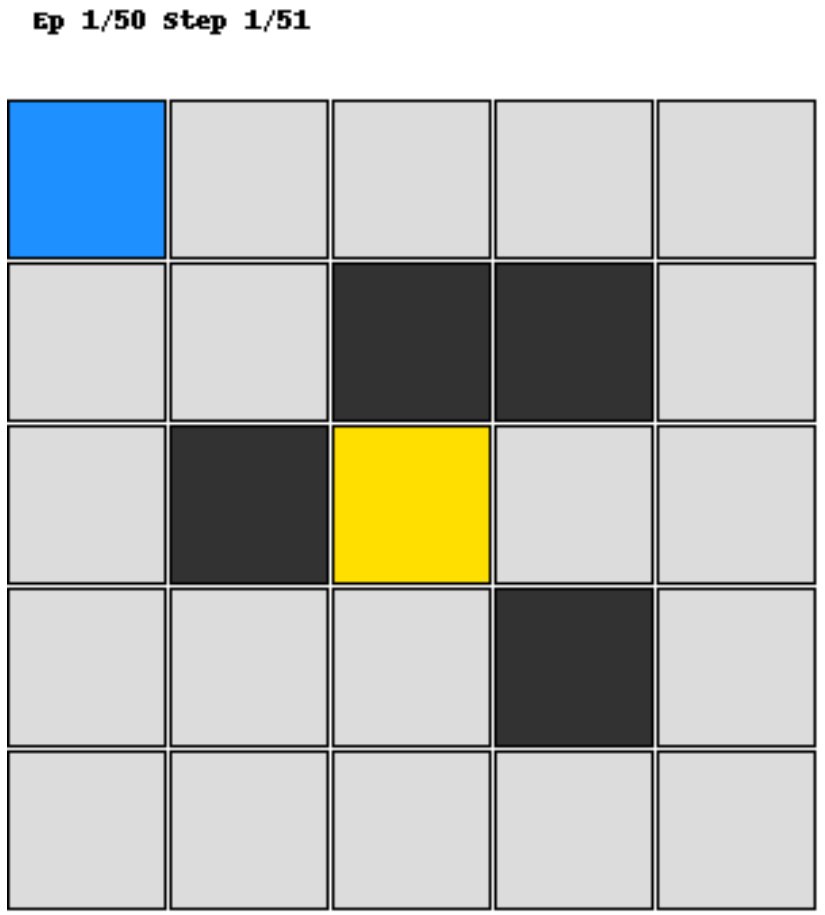
$$r\big(\textsc{NoTest}, \mathbf{p}(t)\big) \;=\; \sum_{\ell=1}^{|\mathcal{L}|} p_\ell(t)\Big[p_{k,\ell}^{(\text{passive})}\, R_{\text{pass}}\Big],$$



# Reinforcement Learning Approach



# Reinforcement Learning Approach



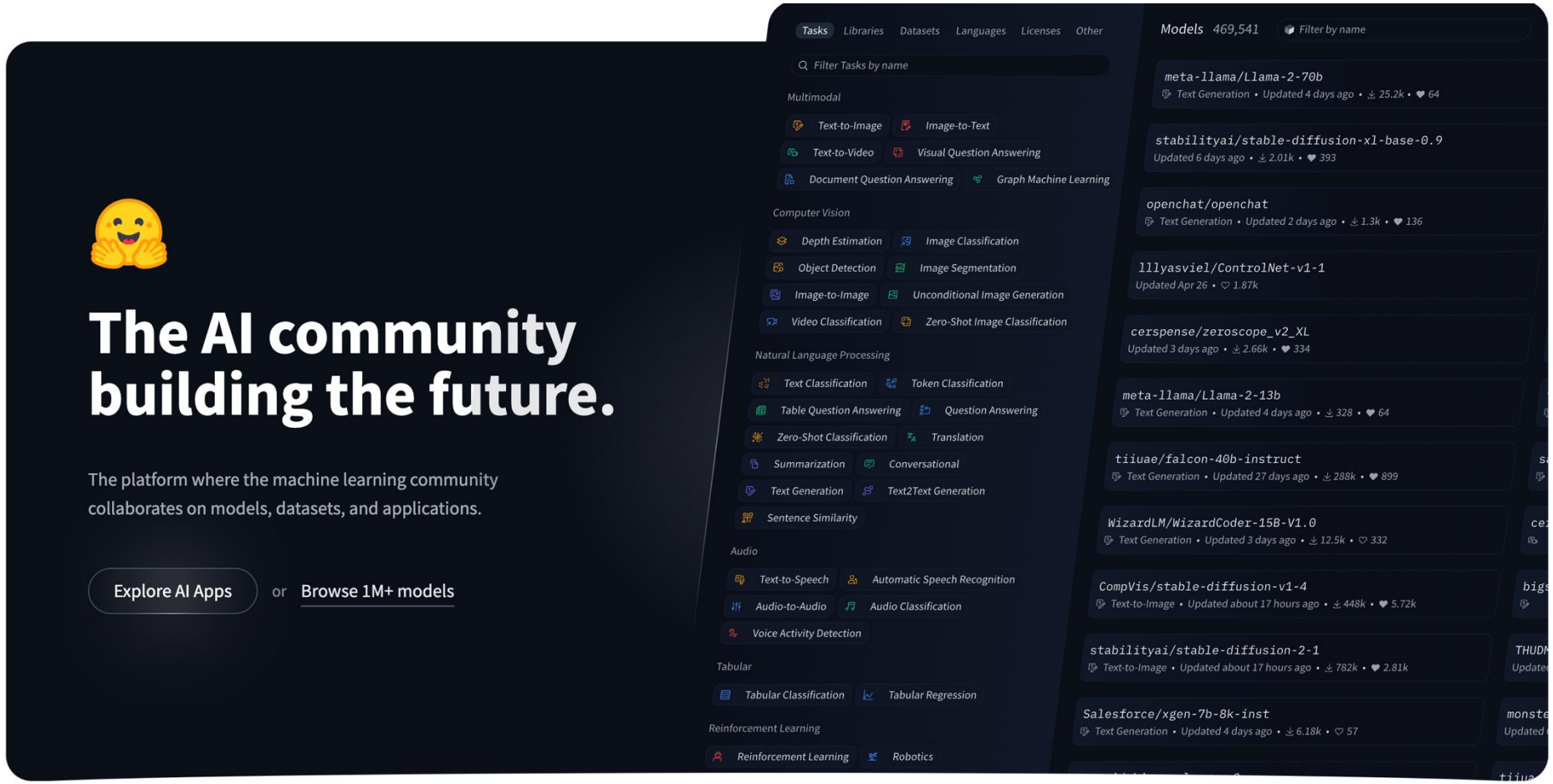

Build based on OpenAI GYM

# LLM-based personal healthcare advisor

- LLM fine tune by LoRA (Low-rank Adaptation) + Huggingface
- Weight matrix for LLM: $\mathbf{W} \in \mathbb{R}^{d \times k}$
- LoRA parameterizes the update as $\Delta\mathbf{W} = \mathbf{AB}$, where $\mathbf{A} \in \mathbb{R}^{d \times r}$, and $\mathbf{B} \in \mathbb{R}^{r \times k}$, with rank $\mathrm{r} \ll \min(d, k)$.
- Instead of learning $d \times k$ updates, LoRA learn only $(d + k) \times r$.

# LLM-based personal healthcare advisor

- LLM fine tune by LoRA (Low-rank Adaptation) + Huggingface

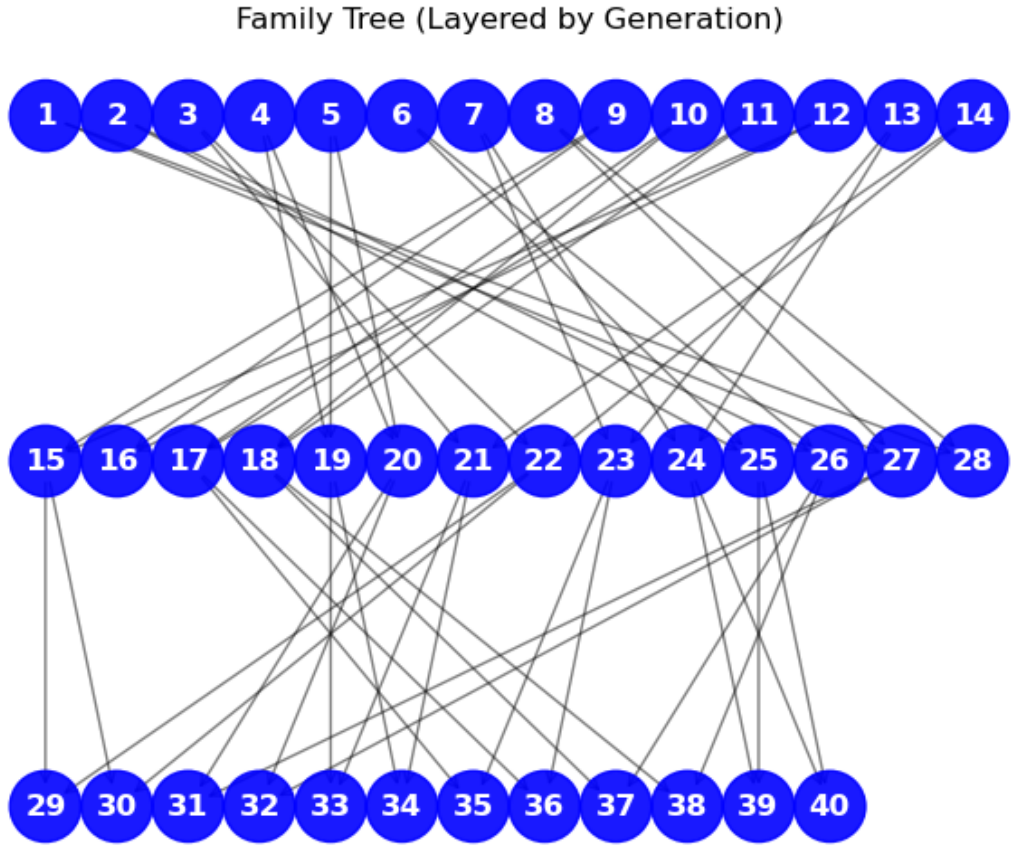


# LLM-based personal healthcare advisor


```
{
  "instruction": "Given this patient screening state, recommend the next test and justify your choice.",
  "input": {
    "covariates": {
      "age": 62,
      "smoking_pack_years": 30,
      "quit_years_ago": 5,
      "family_history_colon": false
    },
    "belief": {
      "latent_probs": { "l1": 0.60, "l2": 0.40 },
      "tumor_beliefs": {
        "1": { "stage0": 0.70, "stage1": 0.20, "stage2": 0.10 },
        "2": { "stage0": 0.50, "stage1": 0.30, "stage2": 0.20 }
      }
    },
    "cost":  100,
    "budget": 1
  },
  "output": {
    "action": "TestTumor(2)",
    "justification": "Tumor 2 has the highest expected net benefit (Whittle index 3.2) given its
    0.5 probability of being present and cost 100."
  }
}
```

# Follow-up directions

- Incorporate family history for personal decision making
- Family-wise decision making
- Genotype unknown (?)



- Thanks