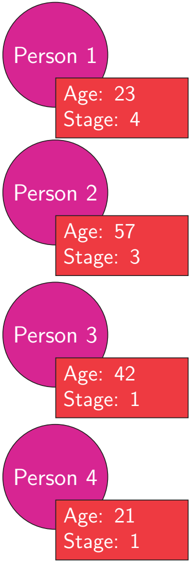
Infinite and Irregular:

Developments for Dynamic Treatment Regimes with Stochastic Decision Points

Dylan Spicker

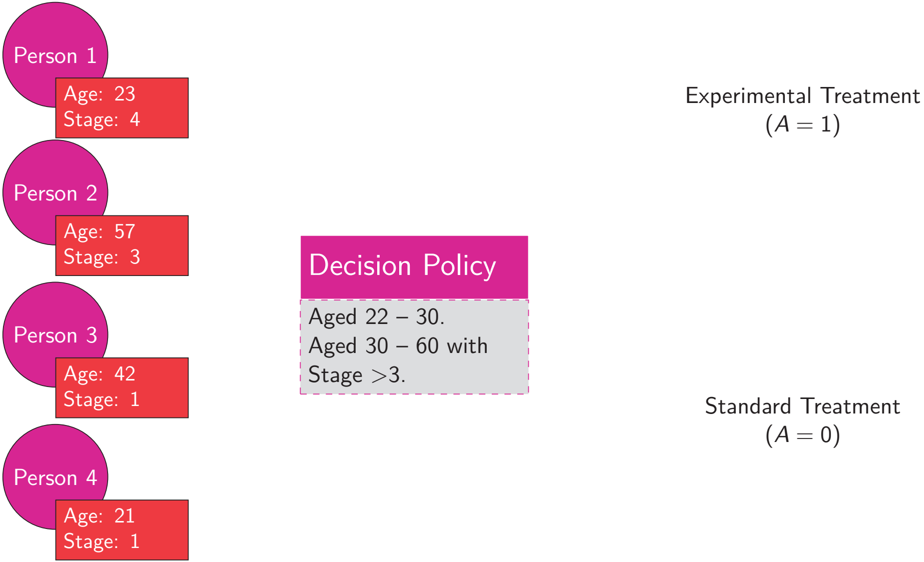
University of New Brunswick, Saint John
Department of Mathematics and Statistics

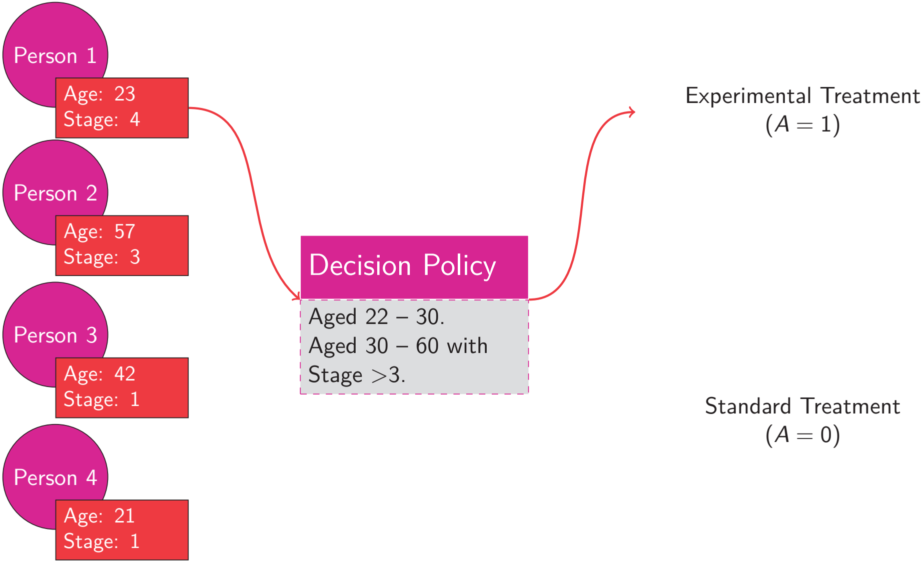
Wednesday June 5, 2024

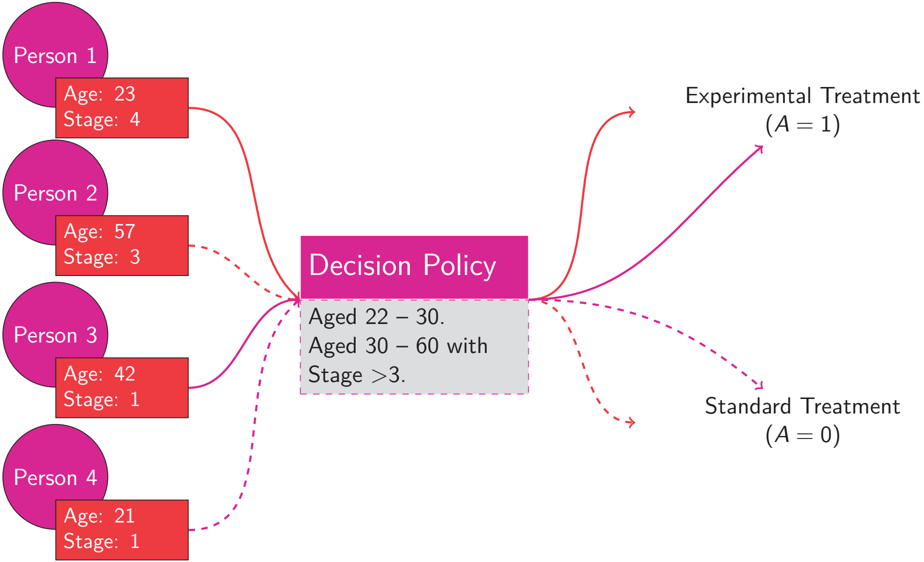


Experimental Treatment
$(A = 1)$

Standard Treatment
$(A = 0)$

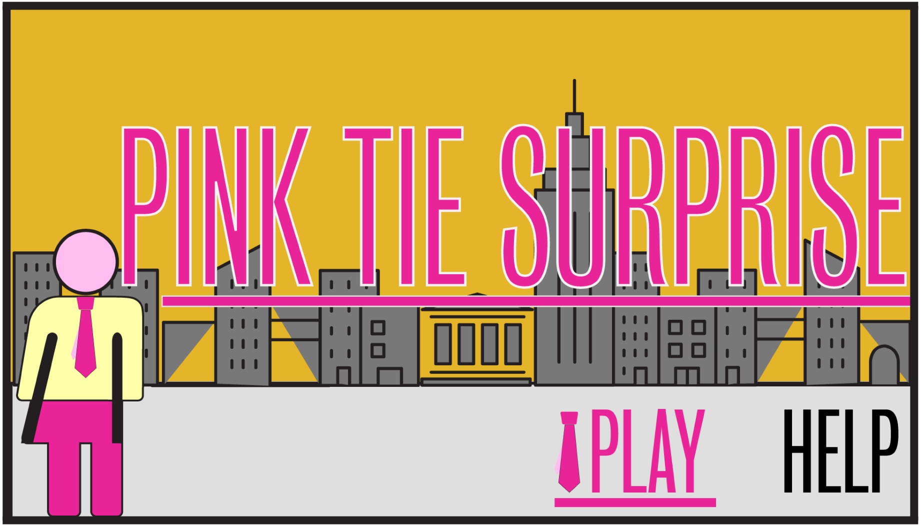

Person 1
Age: 23
Stage: 4
Person 2
Age: 57
Stage: 3
Person 3
Age: 42
Stage: 1
Person 4
Age: 21
Stage: 1
Decision Policy
Aged 22 – 30.
Aged 30 – 60 with
Stage >3.
Experimental Treatment
(A = 1)
Standard Treatment
(A = 0)


Person 1
Age: 23
Stage: 4
Person 2
Age: 57
Stage: 3
Person 3
Age: 42
Stage: 1
Person 4
Age: 21
Stage: 1
Decision Policy
Aged 22 – 30.
Aged 30 – 60 with
Stage >3.
Experimental Treatment
(A = 1)
Standard Treatment
(A = 0)



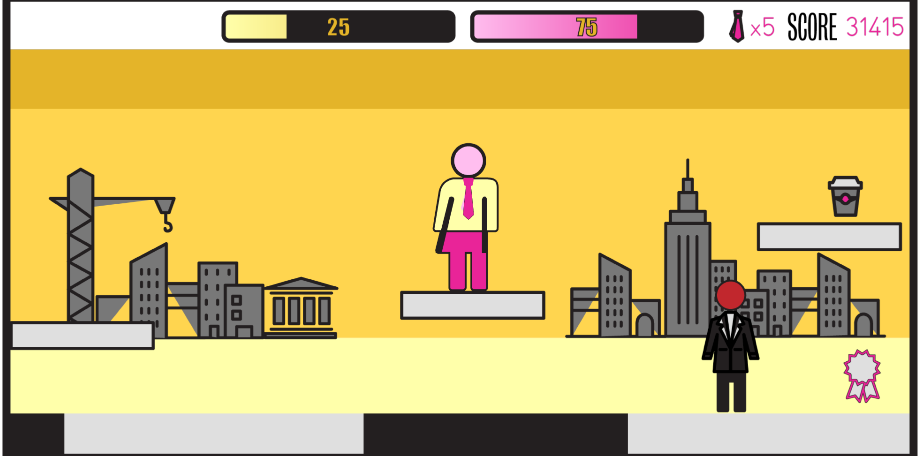
Person 1
Age: 23
Stage: 4
Person 2
Age: 57
Stage: 3
Person 3
Age: 42
Stage: 1
Person 4
Age: 21
Stage: 1
Decision Policy
Aged 22 – 30.
Aged 30 – 60 with
Stage >3.
Experimental Treatment
(A = 1)
Standard Treatment
(A = 0)

The Problem


PINK TIE SURPRISE
PLAY
HELP

The Motivation

The Setup of Our Game

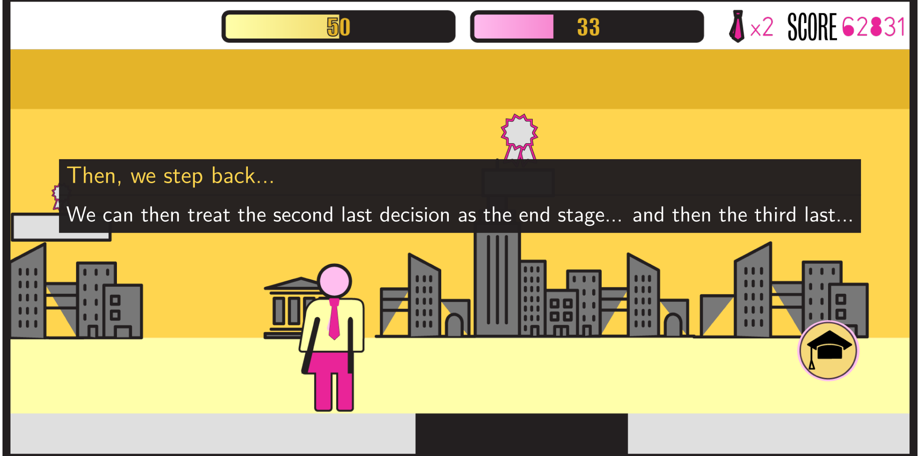


How do you know you are at the end?


Start at the End!
50
33
x2 SCORE 62831
Should we try to collect this?
Starting at the End...
If we only have one decision left to make, and we know all of the information, it is easy!
Collect the award **IF** the amount of energy required is less than what we have.
Do not otherwise.
Collect this to win →

Start at the End ... and Work Backwards

Start at the End!



1. Denote all decision points available $j \in \{1, 2, \ldots, K\}$.

Finite Horizon Dynamic Treatment Regimes

1. Denote all decision points available $j \in \{1, 2, \ldots, K\}$.
2. Denote the treatment at time j, $A_j \in \{0, 1\}$.

Finite Horizon Dynamic Treatment Regimes

1. Denote all decision points available $j \in \{1, 2, \ldots, K\}$.
2. Denote the treatment at time j, $A_j \in \{0, 1\}$.
3. Denote all current individual information at time j $X_j \in \mathbb{R}^{\ell_j}$.

Finite Horizon Dynamic Treatment Regimes

1. Denote all decision points available $j \in \{1, 2, \dots, K\}$.
2. Denote the treatment at time j, $A_j \in \{0, 1\}$.
3. Denote all current individual information at time j $X_j \in \mathbb{R}^{\ell_j}$.
4. Denote the outcome, observed at time K, $Y \in \mathbb{R}$.

Our goal is to determine

$$d^{\text{opt}} = \{d_1^{\text{opt}}, d_2^{\text{opt}}, \ldots, d_K^{\text{opt}}\}, \quad d_j : \mathbb{R}^{\ell_j^*} \longrightarrow \{0, 1\},$$

such that $E[Y|X_1]$ is maximized if d^{opt} is followed.

Key Assumption of DTR Estimation: Deterministic Treatment Times

We assume that there are a

known quantity of treatment decisions

to be made, occurring at known times.

Backwards Induction

1. Estimate d_K^{opt} using Y and $\{X_1, A_1, X_2, \ldots, A_{K-1}, X_K\}$.

1. Estimate d_K^{opt} using Y and $\{X_1, A_1, X_2, \ldots, A_{K-1}, X_K\}$.
2. Compute $\widetilde{Y}_K$ based on d_K^{opt}.

Backwards Induction

1. Estimate d_K^{opt} using Y and $\{X_1, A_1, X_2, \ldots, A_{K-1}, X_K\}$.
2. Compute $\widetilde{Y}_K$ based on d_K^{opt}.
3. Estimate d_{K-1}^{opt} using $\widetilde{Y}_K$ and $\{X_1, A_1, X_2, \ldots, X_{K-1}\}$

Backwards Induction

1. Estimate d_K^{opt} using Y and $\{X_1, A_1, X_2, \ldots, A_{K-1}, X_K\}$.
2. Compute $\widetilde{Y}_K$ based on d_K^{opt}.
3. Estimate d_{K-1}^{opt} using $\widetilde{Y}_K$ and $\{X_1, A_1, X_2, \ldots, X_{K-1}\}$
4. Repeat.

How do you know you are at the end?

Possible Solutions

Time-to-Event Data

Suppose the outcome is T, the time of occurence for some event of interest.

The goal is to find d^{opt} to maximize $E[T|X_1]$.

DTRs with Survival Outcomes

- Shu Yang, Anastasios A Tsiatis, and Michael Blazing (2018). "Modeling survival distribution as a function of time to treatment discontinuation: A dynamic treatment regime approach". In: Biometrics 74.3, pp. 900–909
- Rebecca Hager, Anastasios A Tsiatis, and Marie Davidian (2018). "Optimal two-stage dynamic treatment regimes from a classification perspective with censored survival data". In: Biometrics 74.4, pp. 1180–1192
- Gabrielle Simoneau et al. (2020). "Estimating optimal dynamic treatment regimes with survival outcomes". In: Journal of the American Statistical Association 115.531, pp. 1531–1539
- Hunyong Cho et al. (2023). "Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring". In: Biometrika 110.2, pp. 395–410

How do you know you are at the end, if we are trying to delay the remission of a particular disease or the onset of a symptom?

Many outcomes of interest are **not** survival outcomes.

Infinite Time Horizons

- Denote all current state information at time t, S_t.

Infinite Time Horizons

- Denote all current state information at time t, S_t.
- Denote the action at time t, A_t.

- Denote all current state information at time t, S_t.
- Denote the action at time t, A_t.
- Denote the reward at time t, A_t.

- Denote all current **state information** at time t, S_t.
- Denote the **action** at time t, A_t.
- Denote the **reward** at time t, A_t.

The **cumulative discounted reward** at time t is

$$G_t = \sum_{k=1}^{\infty} \gamma^{k-1} R_{t+k}.$$

The goal is to **maximize** $E[G_t | S_t, A_t]$.

A Markov Decision Process is a stochastic process describing the transformations between states based on the actions that were taken, and the considering the earned rewards.

Infinite Horizon Dynamic Treatment Regimes

- Ashkan Ertefaie and Robert L Strawderman (Sept. 2018). "Constructing dynamic treatment regimes over indefinite time horizons". en. In: Biometrika 105.4, pp. 963–977
- Daniel J Luckett et al. (2020). "Estimating Dynamic Treatment Regimes in Mobile Health Using V-learning". en. In: J. Am. Stat. Assoc. 115.530, pp. 692–706
- Wenzhuo Zhou, Ruoqing Zhu, and Annie Qu (Jan. 2024). "Estimating Optimal Infinite Horizon Dynamic Treatment Regimes via pT-Learning". In: J. Am. Stat. Assoc. 119.545, pp. 625–638

How do you know you are at the end, if the process is Markovian?

For all $t \geq 1$, the **Markov assumption** assumes

$$S_{t+1} \perp \{S_1, A_1, S_2, \ldots, S_{t-1}, A_{t-1}\} \mid \{S_t, A_t\}.$$

This is commonly expressed as

$$Pr(S_{t+1} | S_t, A_t, S_{t-1}, A_{t-1}, \ldots, A_1, S_1) = Pr(S_{t+1} | S_t, A_t).$$

The Time-Homogenous Assumption

The time-homogenous assumption assumes, for all $t \geq 1$,

$$Pr(S_{t+1} | S_t, A_t) = Pr(S_1 | S_0, A_0).$$

The Problem, Re-Framed

Why make Markovian Assumptions?

"Although not imposed by other methods for estimating optimal dynamic treatment regimes, this Markov assumption is advantageous because the resulting Q-function and corresponding optimal dynamic treatment regime are both independent of time. In addition to avoiding the need for backward induction, estimation and inference become possible at decision points that lie beyond the observed time horizon."

– Ertefaie and Strawderman 2018

~~How do you know you are at the end?~~

How are you able to extrapolate?

The Benefit of the Markovian Assumption

- Extrapolation is Possible
- Backwards Induction Free
- Interpretable Regimes

Markovian assumptions are **not** always appropriate.

Other Unaddressed Considerations

- Irregular Treatment Times
- Time-homogeneity
- Covariate-drive Treatment Times

In practice, the time of a given treatment, t_j, is informed by the patient history preceding t_j.

ITRs with Covariate Driven Treatment Times

- Janie Coulombe et al. (May 2023). "Estimating individualized treatment rules in longitudinal studies with covariate-driven observation times". In: Stat. Methods Med. Res. 32.5, pp. 868–884

Another Way of Thinking About This

We typically derive the optimal DTR by implicitly conditioning on the number and timing of future treatments.

Some Areas of Pursuit

Online DTR Estimation

What if we frame the problem as one of online learning rather than offline learning?

Joint or Hierarchical Modelling

What if we explore these as separate stochastic processes that depend on one another? Renewal reward process or functional data.

Exploring Repeated ITRs

Can we explore the impact of finding ITRs, perhaps which take as predictors past treatments (if they exist) in a way to optimize single treatments, not in sequence?