WHAT SHOULD BE THE FOCUS OF STATISTICAL PERSONALISED MEDICINE?

Philip Dawid
University of Cambridge

Joint work with Mats Stensrud, Aaron Sarvet and Stephen Senn

OVERVIEW

- Scott Mueller and Judea Pearl have recently been promoting a new statistical approach to personalised medicine. In contrast to the traditional statistical approach based on decision theory, their idea is that medical decisions should focus on the "counterfactual contrast" between a patient's actual response to the treatment they get, and the (necessarily unobservable) "potential response" to the treatment they do not get.
- Mueller and Pearl further claim that combining observational and experimental data can lead to improved decisions.
- We deconstruct and critique their arguments, and reaffirm the appropriateness of the traditional approach.

Medical Decisions

L: background characteristics of patient
X: give treatment?
Y: response

- Data available on (L, X, Y) for many past patients

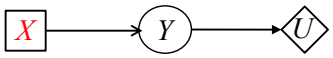
- *How should doctor use L to decide X?*
  - *or to prioritize patients for treatment?*

BIG Assumption

L: background characteristics of patient
X: give treatment?
Y: response

- Data available on (L, X, Y) for many past patients
- *We assume that new patient is exchangeable with previous cases*
  - *typically not so for clinical trials*

Decision-Theoretic (DT) Approach

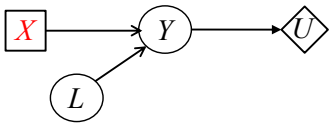


- L absent
- Value outcome $Y=y$ with utility $U = u(y)$
- From data, estimate distribution of Y for each hypothesised intervention $X \leftarrow x$
- Choose x to maximise expected utility

$$E\{U(Y) \mid X \leftarrow x\}$$

If Y is binary, maximise $P(Y = 1 \mid X \leftarrow x)$

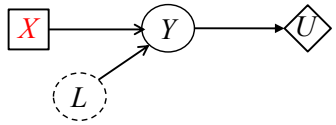
Use of Covariate Information



- **For current patient, can observe *L=l***
- From data, can estimate conditional distributions of Y, given L, $X \leftarrow x$
  - statistical/ML techniques
- Choose x to maximise "personalised" expected utility

$$E\{U(Y) \mid L=l, X \leftarrow x\}$$

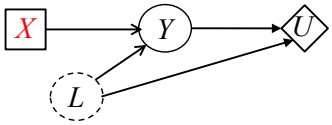
Use of Covariate Information



- **For current patient, can't observe *L***
- From data, can estimate conditional distributions of *Y*, given *L*, $X \leftarrow x$
  - *but can't use this conditional distribution*
- **Revert**: Still choose x to maximise **overall** expected utility

$$E\{U(Y) \mid X \leftarrow x\}$$

Use of Covariate Information



- **For current patient, can't observe *L***
- **But utility is now** $U(L, Y)$
- From data, estimate joint distribution of Y and L, given $X \leftarrow x$
- Choose x to maximise expected utility

$$E\{U(L, Y) \mid X \leftarrow x\}$$

Mueller and Pearl (MP) Approach

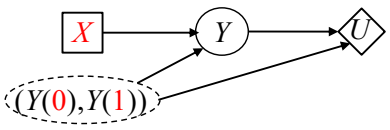
- Introduce $Y(x)$, "potential response" to intervention $X \leftarrow x$
- Consider $(Y(0), Y(1))$ as having a joint distribution, unaffected by decision $X \leftarrow x$ taken
- Compare outcome $Y(1)$ if treated with outcome $Y(0)$ if not
  - $Y(1) = Y(0)$ – no effect
  - $Y(1) > Y(0)$ – benefit
  - $Y(1) < Y(0)$ – harm
- Treat accounting for expected benefit/harm

MP Approach

Three strands to their approach:

1. Assess value of treatment in terms of pair of potential outcomes $(Y(0), Y(1))$
   e.g., P(benefit), P(harm)
2. Use data to set bounds on their probabilities
3. Combine experimental and observational data

1. Utility Function



- Assign utility U depending on both actual *and counterfactual* response (takes $L = (Y(0), Y(1))$
  - e.g., 1 if benefit, −3 if harm, 0 if untreated
    - P(benefit) − 3 × P(harm) vs. 0

1. Utility Function

- Assign utility U depending on both actual *and counterfactual* response (takes $L = (Y(0), Y(1))$
  - e.g., 1 if benefit, −3 if harm, 0 if untreated
    - P(benefit) − 3 × P(harm) vs. 0

BUT—

- Impossible ever to observe both $Y(0)$ and $Y(1)$ together
- So typically *can not identify joint distribution* of $(Y(0), Y(1))$ from data
  – – likewise, P(benefit), P(harm)

2. Probability Bounds

Can not identify joint distribution of ($Y(0)$, $Y(1)$) from data

HOWEVER:

From *experimental data*, yielding the marginal distributions of each of $Y(0)$, $Y(1)$ separately, we can *set bounds* on their joint distribution (and so on P(benefit), P(harm),..)

2. Probability Bounds

- From *experimental data*, yielding the marginal distributions of each of $Y(0)$, $Y(1)$ separately, we can *set bounds* on their joint distribution (and so on P(benefit), P(harm), …)

- $P(Y{=}1 \mid X{\leftarrow}1) = 0.49$
- $P(Y{=}1 \mid X{\leftarrow}0) = 0.21$

$0 \le x \le .21$

		$Y(1)$ = 1	$Y(1)$ = 0	
$Y(0)$ = 1		x	$.21 - x$	.21
$Y(0)$ = 0		$.49 - x$	$.3 + x$	.79
		.49	.51	1

2. Probability Bounds

- From *experimental data*, yielding the *marginal* distributions of each of $Y(0)$, $Y(1)$ separately, we can *set bounds* on their *joint* distribution (and so on P(benefit), P(harm), …)

- This joint distribution is *fully identified* just when *either* $Y(0)$ *or* $Y(1)$ is degenerate (at 0 or 1)
  - In this case either P(benefit) = 0, or P(harm) = 0
- Then clear what to do
- DT and MP agree

Use of Covariate

- From experimental data, yielding the marginal distributions of each of $Y(0)$, $Y(1)$ separately, we can set bounds on their joint distribution (and so on P(benefit), P(harm), …)
- We can *improve* these bounds if there is a pre-treatment covariate L such that we know the joint distribution of (L, Y) under either intervention
  - *even if we don't observe L for current patient*
- Now we have full identification just when, *given* any value of L, either $Y(0)$ or $Y(1)$ is degenerate
- DM and MP agree if L observed

3. Combining experimental and observational data

- One possible covariate is the *intention to treat* (ITT) variable X^* , which determines the *actual* treatment X in the observational case, but is overridden by the *applied* X in an experiment
- X^* sometimes observable in an experiment
- *But even if not, by combining experimental and observational data we can still compute the experimental joint distributions of* (X^*, Y)
- In either case this gives improved MP bounds

Another BIG Assumption

- From *experimental data*, yielding the marginal distributions of each of $Y(0)$, $Y(1)$ separately, we can *set bounds* on their joint distribution (and so on P(benefit), P(harm),..)
- If we also have *observational data*, we can improve these bounds

- We must assume observational cases are exchangeable with experimental cases
  - *typically not so for clinical trials*

EXAMPLE 1

- *Experimental data*:
  P(survive | drug) = 0.49, P(survive | no drug) = 0.21

DT: Better to give treatment

MP: We have

$$0.28 \le \text{P(benefit)} \le 0.49$$
$$0 \le \text{P(harm)} \le 0.21$$

What to do?

- Adding *Observational data*:

P(drug) = 0.7, P(survive | drug) = 0.27, P(survive | no drug) = 0.7

we get

P(benefit) = 0.28
P(harm) = 0
—so give treatment

19

EXAMPLE 2

- *Experimental data*:
  P(survive | drug) = 0.49, P(survive | no drug) = 0.21

DT: Better to give treatment

MP: We have

$$0.28 \le \text{P(benefit)} \le 0.49$$
$$0 \le \text{P(harm)} \le 0.21$$

What to do?

- Adding *Observational data*:

P(drug) = 0.7, P(survive | drug) = 0.7, P(survive | no drug) = 0.7

we get

P(benefit) = 0.49
P(harm) = 0.21
$0.49 - 3 \times 0.21 < 0$. So don't treat.
different from DT because of utility function

20

EXAMPLE 2

- *Experimental data*:
  P(survive | drug) = 0.49, P(survive | no drug) = 0.21

DT: Better to give treatment

MP:

P(benefit) = 0.49
P(harm)= 0.21
$0.49 - 3 \times 0.21 < 0$. So don't treat.

In a population of similar individuals, treating everyone (DT) would lead to 49% surviving, whereas treating no one (MP) would lead to 21% surviving!

21

EXAMPLE 2

- *Experimental data*:
  P(survive | drug) = 0.49, P(survive | no drug) = 0.21

DT: Better to give treatment

MP:

P(benefit) = 0.49
P(harm)= 0.21
$0.49 - 3 \times 0.21 < 0$. So don't treat.

Alternatively the utilities for benefit and harm imply the following "marginal utilities"—

$U(\text{treat}, Y{=}1) = 1$. $U(\text{treat}, Y{=}0) = -1.24$. $E\{U(\text{treat}, Y) = -0.14$
$U(\text{not treat}, Y{=}1) = U(\text{not treat}, Y{=}0) = 0$. $E\{U(\text{not treat}, Y) = 0$
So don't treat

But why should the utility of the outcome depend on whether or not treated?

22

EXAMPLE 3

- *Experimental data*:

P(survive | drug) = 0.49, P(survive | no drug) = 0.21

DT: Better to give treatment

MP: We have

$$0.28 \le \text{P(benefit)} \le 0.49$$
$$0 \le \text{P(harm)} \le 0.21$$

What to do?

- Adding *Observational data*:

P(drug) = 0.7, P(survive | drug) = 0.29, P(survive | no drug) = 0.33

we get

$$0.28 \le \text{P(benefit)} \le 0.40$$
$$0 \le \text{P(harm)} \le 0.12$$

What to do?

23

3. Combining experimental and observational data

- In Examples 1 & 2, after combining data, the joint distribution of $(Y(0), Y(1))$ was fully identified
- This happens just when, for either *intended treatment* $X^*{=}x^*$, there is a setting $X \leftarrow x^\dagger$ under which Y is determined (= 0 or 1)
- Ideally, observe x^* for new patient, and choose $X \leftarrow x^\dagger$ if that yields $Y = 1$; else, other setting
  – DT and MP agree
- More generally, observe X^* to improve DT

24

3. Combining experimental and observational data

- In Example 1, after combining data, the joint distribution of ($Y(0)$, $Y(1)$) was fully identified
  - P(benefit) = 0.29, P(harm)= 0
- We find:
  - For $X^*=0$, $X \leftarrow 1 \Rightarrow Y=1$. So treat
  - For $X^*=1$, $X \leftarrow 0 \Rightarrow Y=0$. So treat
- So treat anyway! DT and MP agree

25

3. Combining experimental and observational data

- In Example 2, after combining data, the joint distribution of ($Y(0)$, $Y(1)$) was fully identified
  - P(benefit) = 0.49, P(harm)= 0.21
- We find:
  - For $X^*=0$, $X \leftarrow 1 \Rightarrow Y=0$. So don't treat
  - For $X^*=1$, $X \leftarrow 0 \Rightarrow Y=0$. So treat
- Helps if we can observe X^*

26

3. Combining experimental and observational data

In Example 3, after combining data, the joint distribution of ($Y(0)$, $Y(1)$) was *not* fully identified:

$0.28 \leq$ P(benefit) ≤ 0.40, $0 \leq$ P(harm) ≤ 0.12

- In this case we find:
  - $P(Y=1 \mid X^*=0, X \leftarrow 1) = 0.97$
  - $P(Y=1 \mid X^*=0, X \leftarrow 0) = 0.33$
    - So treat if $X^*=0$
  - $P(Y=1 \mid X^*=1, X \leftarrow 1) = 0.29$
  - $P(Y=1 \mid X^*=1, X \leftarrow 0) = 0.16$
    - So treat if $X^*=1$
  - DT: Treat anyway (X^* wasn't helpful)

27

Messages?

- Reason for combining observational and experimental data is to give us a handle on unobserved ITT variable X^*
- If we can observe X^* in a new patient, this can improve decision-making
- If we can't, too bad
  - *unless* utility depends on unobserved features
  - as it does for MP
- Is such a utility reasonable?

28

REFERENCES

Ang Li and Judea Pearl (2019).
Unit selection based on counterfactual logic.
Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, 1793–1799.

Scott Mueller and Judea Pearl (2022).
Personalized decision making – a conceptual introduction.
Technical Report R-513, UCLA Cognitive Systems Laboratory.

29