Computational social choice
Statistical approaches

Lirong Xia

Rensselaer

Mar 1, 2016

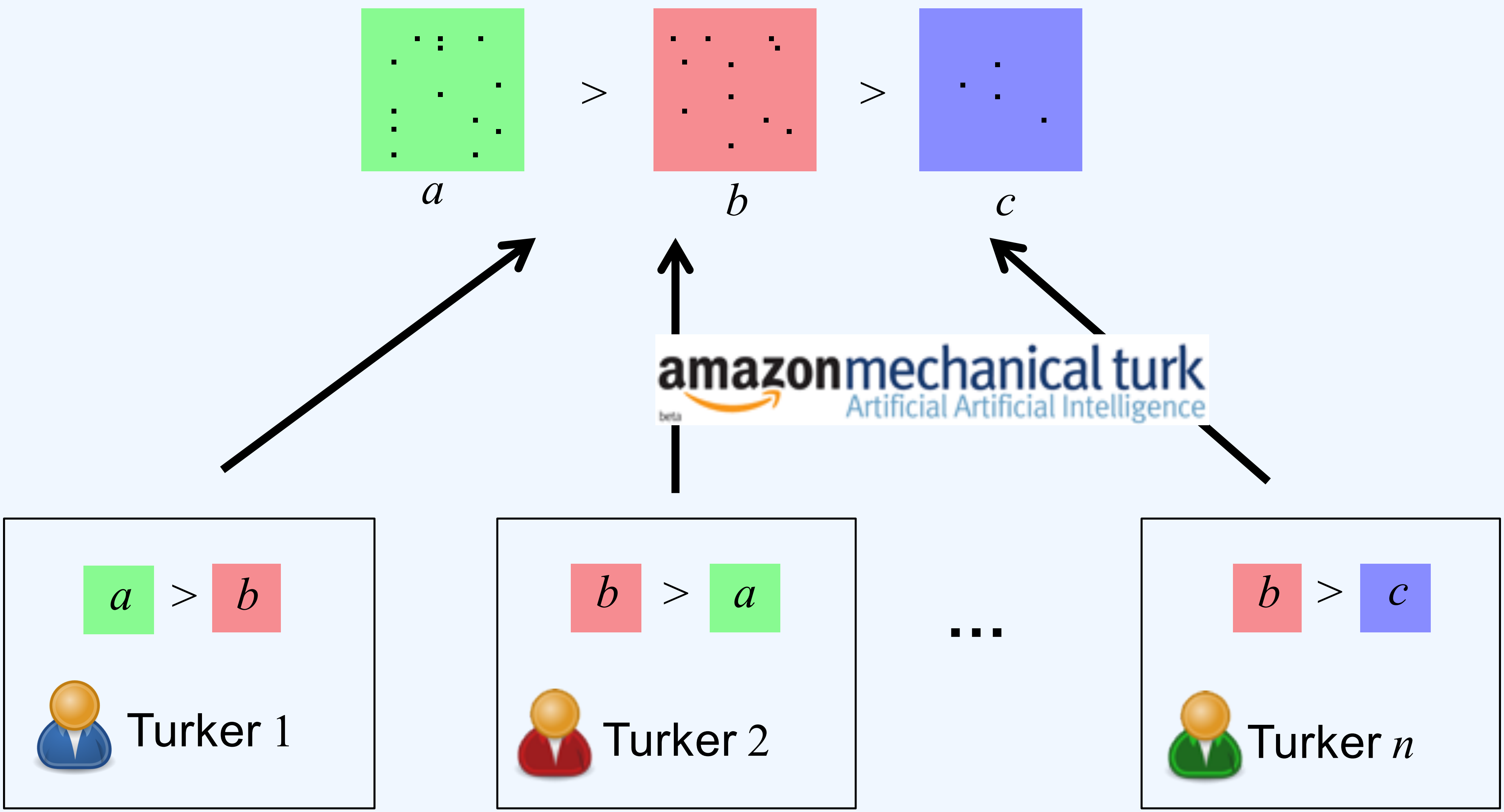
Announcement

- Start to think about the topic for project

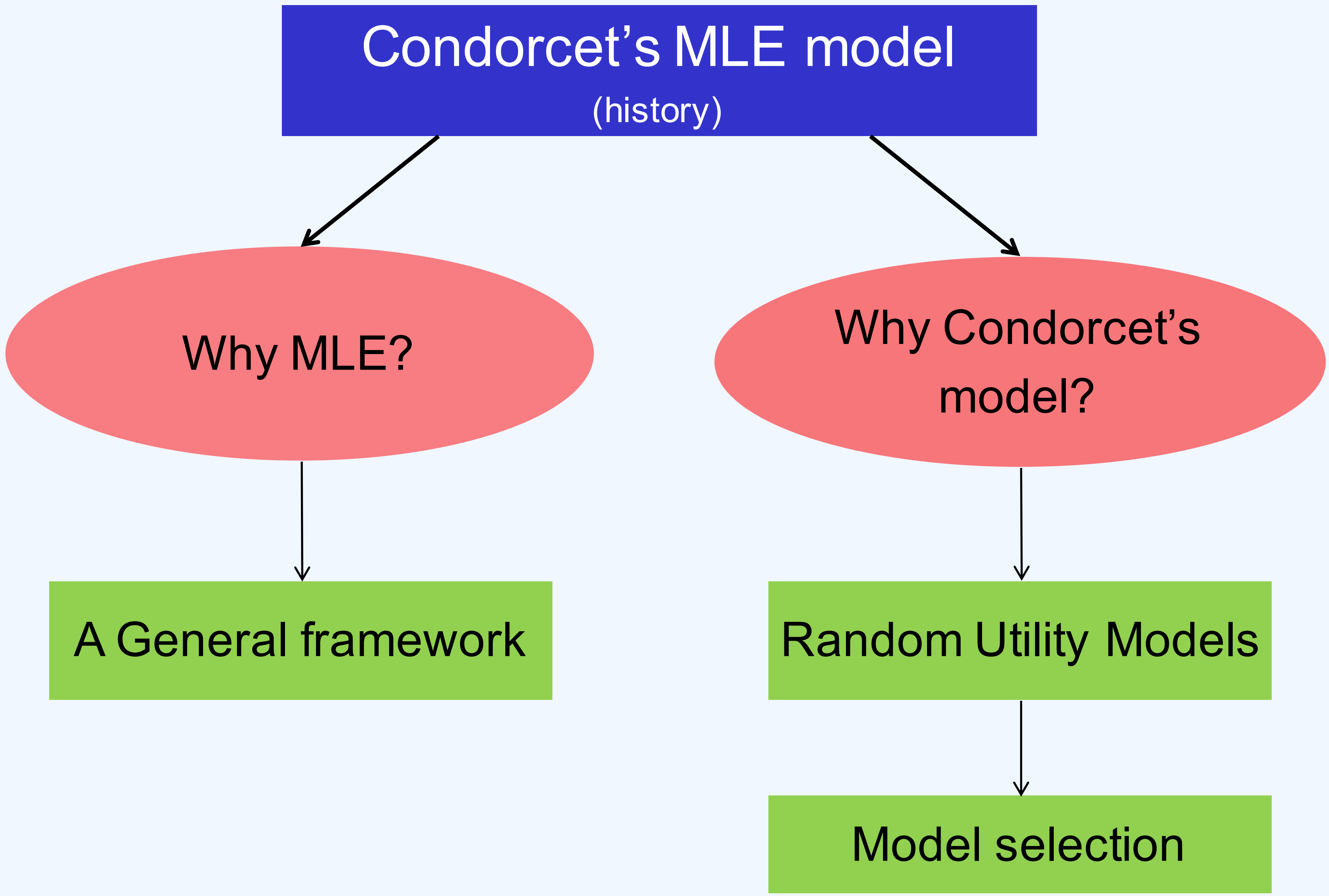
Last class: manipulation

- Various “undesirable” behavior
  - manipulation
  - bribery
  - control

Example: Crowdsourcing



Outline: statistical approaches



The Condorcet Jury theorem
[Condorcet 1785]

The Condorcet Jury theorem.

- Given
  - two alternatives $\{a,b\}$.
  - $0.5<p<1$,
- Suppose
  - each agent's preferences is generated i.i.d., such that
  - w/p p, the same as the ground truth
  - w/p $1-p$, different from the ground truth
- Then, as $n\to\infty$, the majority of agents' preferences converges in probability to the ground truth

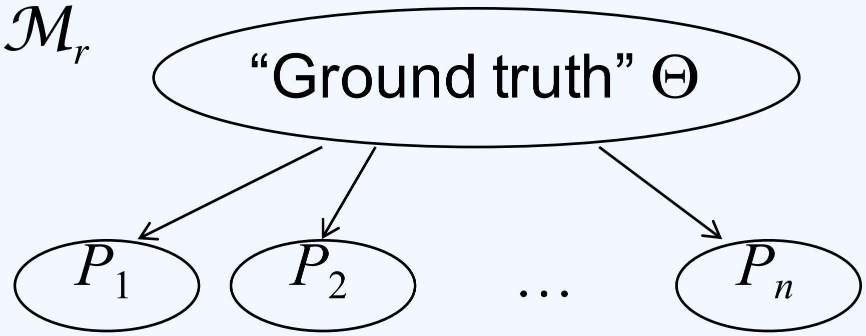
Parametric ranking models

- Composed of three parts
  - A parameter space: Θ
  - A sample space: S = Rankings$(C)^n$
    - C = the set of alternatives, n=#voters
    - assuming votes are i.i.d.
  - A set of probability distributions over S:

    {$\Pr(s|\theta)$ for each $s \in$ Rankings(C) and $\theta \in \Theta$}

Maximum likelihood estimator (MLE) mechanism

Model: $\mathcal{M}_r$



- For any profile $D=(P_1,\ldots,P_n)$,
  - The likelihood of Θ is $L(\Theta|D)=\Pr(D|\Theta)=\prod_{P\in D}\Pr(P|\Theta)$
  - The MLE mechanism

    $$\text{MLE}(D)=\text{argmax}_{\Theta}L(\Theta|D)$$

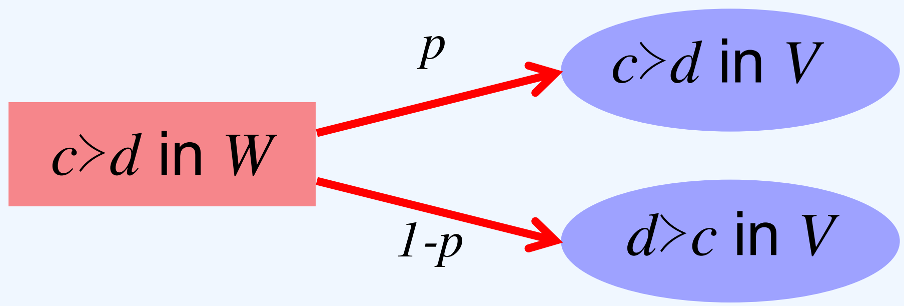
  - Decision space = Parameter space

Condorcet's MLE approach
[Condorcet 1785]

- Use a statistical model to explain the data (preference profile)
  - Condorcet's model
- Use likelihood inference to make a decision
  - Decision space = Parameter space
  - not necessarily MLE

Condorcet's model
[Condorcet 1785]

- Parameterized by an opinion (simple directed graphs)
- Given a "ground truth" opinion W and p>$1/2$, generate each pairwise comparison in V independently as follows (suppose $c \succ d$ in W)



$$\Pr(b \succ c \succ a \mid a \succ b \succ c) = p(1-p)^2$$

- MLE ranking is the Kemeny rule [Young APSR-88]

Condorcet's model for more than 2 alternatives [Young 1988]

- Not very clear in Young's paper, email Lirong for a working note that proofs this according to Young's calculations
  - message 1: Condorcet's model is different from the Mallows model
  - message 2: Kemeny is not an MLE of Condorcet (but it is an MLE of Mallows)
- Fix 0.5<p<1, parameter space: all binary relations over the alternatives
  - may contain cycles
- Sample space: each vote is a all binary relations over the alternatives
- Probabilities: given a ground truth binary relation
  - comparison between a and b is generated i.i.d. and is the same as the comparison between a and b in the ground truth with probability p
- Also studied in [ES UAI-14]

Mallows model [Mallows 1957]

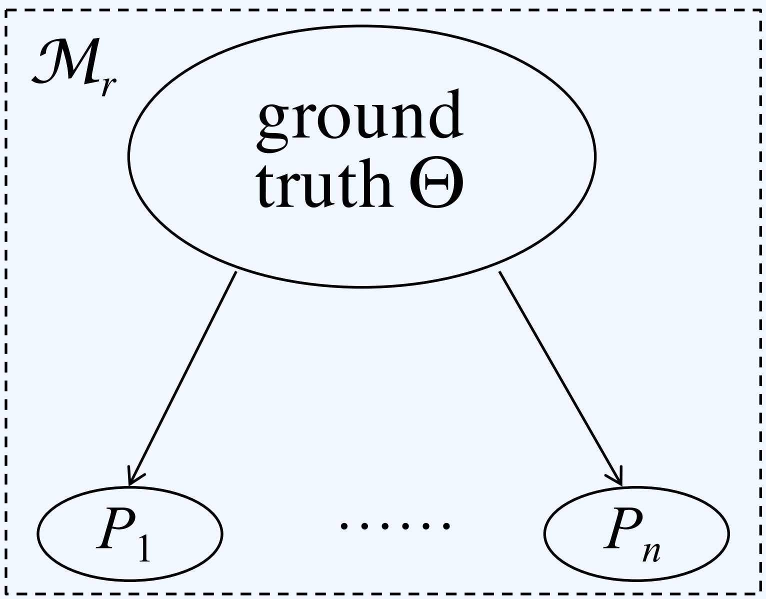
- Fix ϕ<1, parameter space
  - all full rankings over alternatives
  - different from Condorcet's model
- Sample space
  - i.i.d. generated full rankings over alternatives
  - different from Condorcet's model
- Probabilities: given a ground truth ranking W, generate a ranking V w.p.
  - $\Pr(V|W) \propto \phi^{\text{Kendall}(V,W)}$

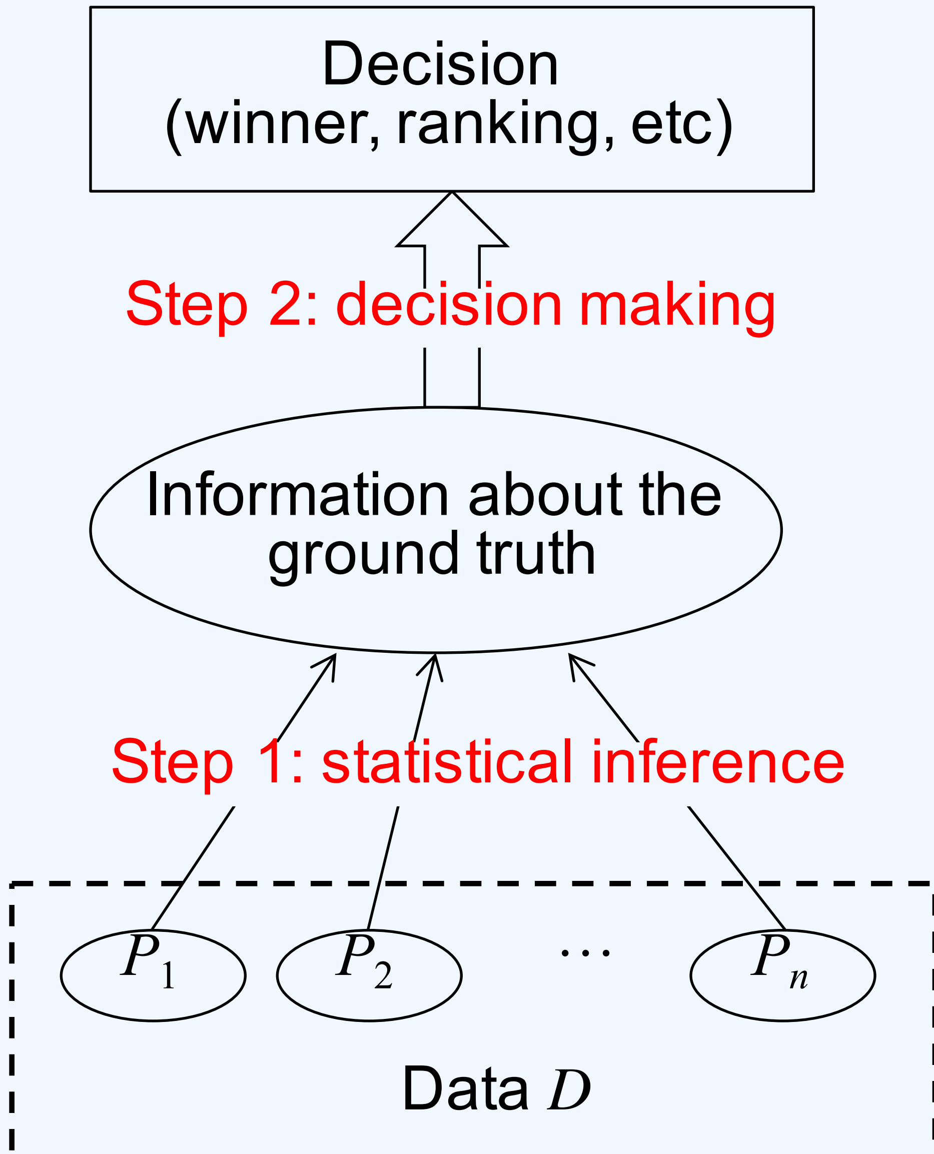
Statistical decision theory

- Given
  - statistical model: Θ, S, $\Pr(s|\theta)$
  - decision space: D
  - loss function: $L(\theta, d) \in \mathbb{R}$
- Make a good decision based on data
  - decision function f : data⟶D
  - Bayesian expected lost:
    - $EL_B(\text{data}, d) = E_{\theta|\text{data}} L(\theta, d)$
  - Frequentist expected lost:
    - $EL_F(\theta, f) = E_{\text{data}|\theta} L(\theta, f(\text{data}))$
  - Evaluated w.r.t. the objective ground truth
    - different from the approaches evaluated w.r.t. agents' subjective utilities [BCH+ EC-12]

Statistical decision framework

Given $\mathcal{M}_r$



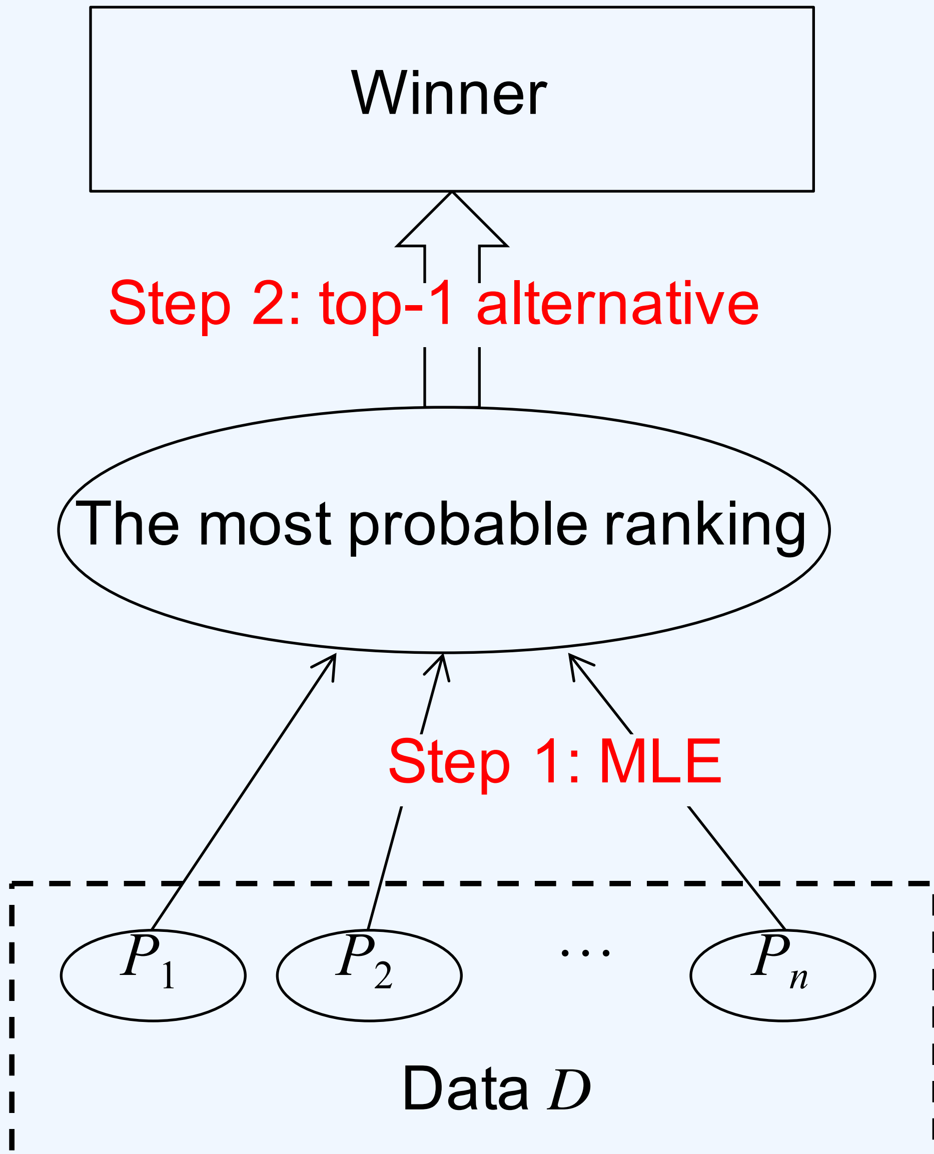


Example: Kemeny

$\mathcal{M}_r$ = Condorcet' model

Step 1: MLE

Step 2: top-alternative



Frequentist vs. Bayesian in general

- You have a biased coin: head w/p p
  - You observe 10 heads, 4 tails
  - Do you think the next two tosses will be two heads in a row?

Credit: Panos Ipeirotis & Roy Radner

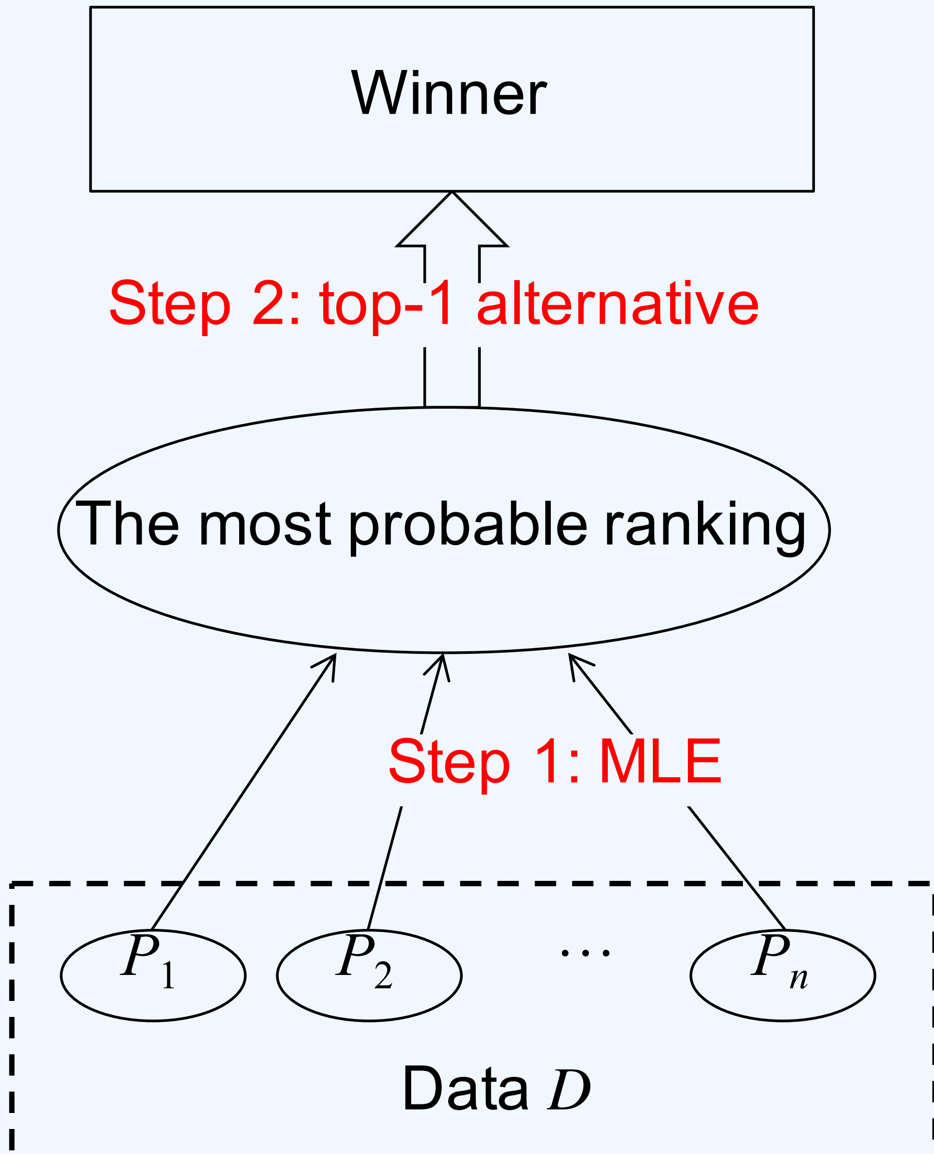
- Frequentist
  - there is an unknown but fixed ground truth
  - p = 10/14=0.714
  - Pr(2heads|p=0.714) =(0.714)^2=0.51>0.5
  - Yes!

- Bayesian
  - the ground truth is captured by a belief distribution
  - Compute Pr(p|Data) assuming uniform prior
  - Compute Pr(2heads|Data)=0.485<0.5
  - No!

Classical Kemeny [Fishburn-77]

$\mathcal{M}_r$ = Condorcet' model

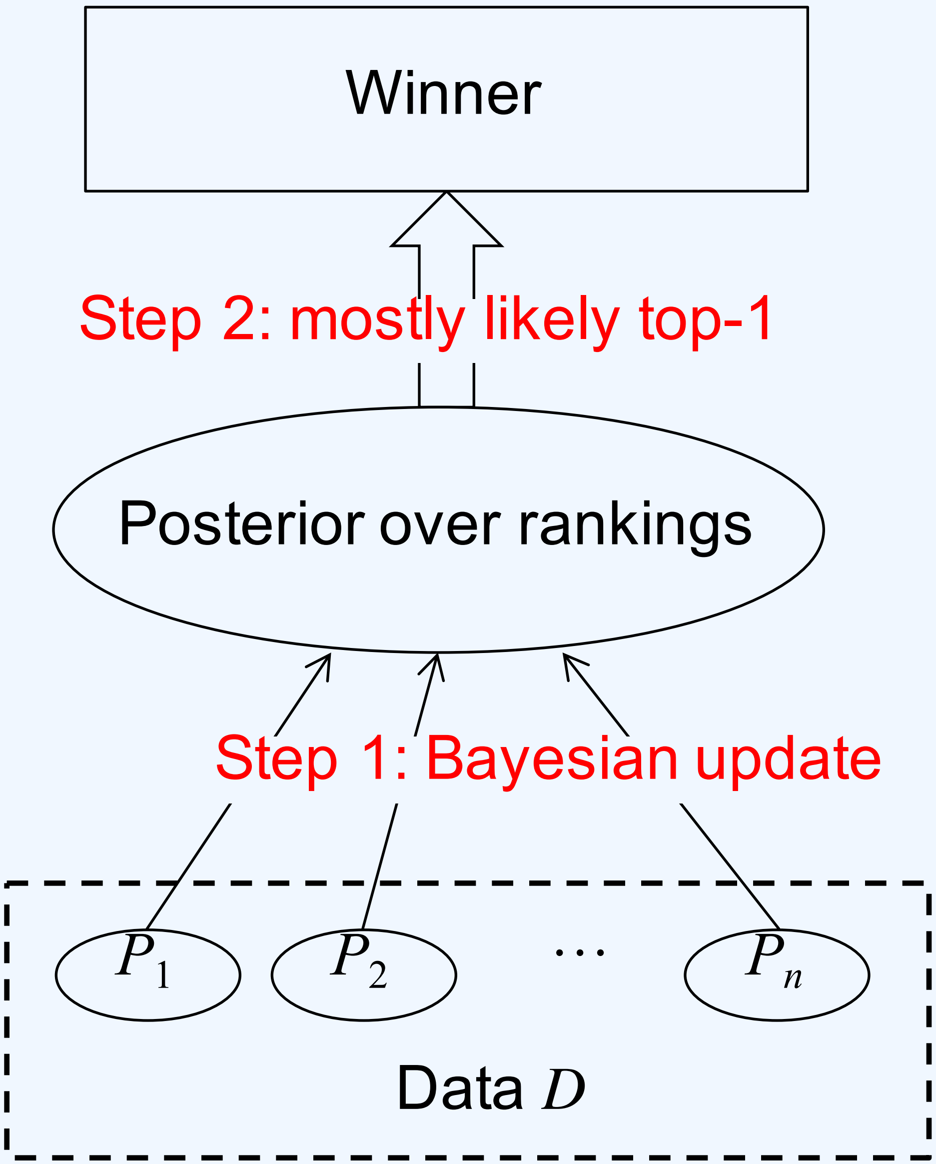
This is the Kemeny rule (for single winner)!



Example: Bayesian

$\mathcal{M}_r$ = Condorcet' model

This is a new rule!

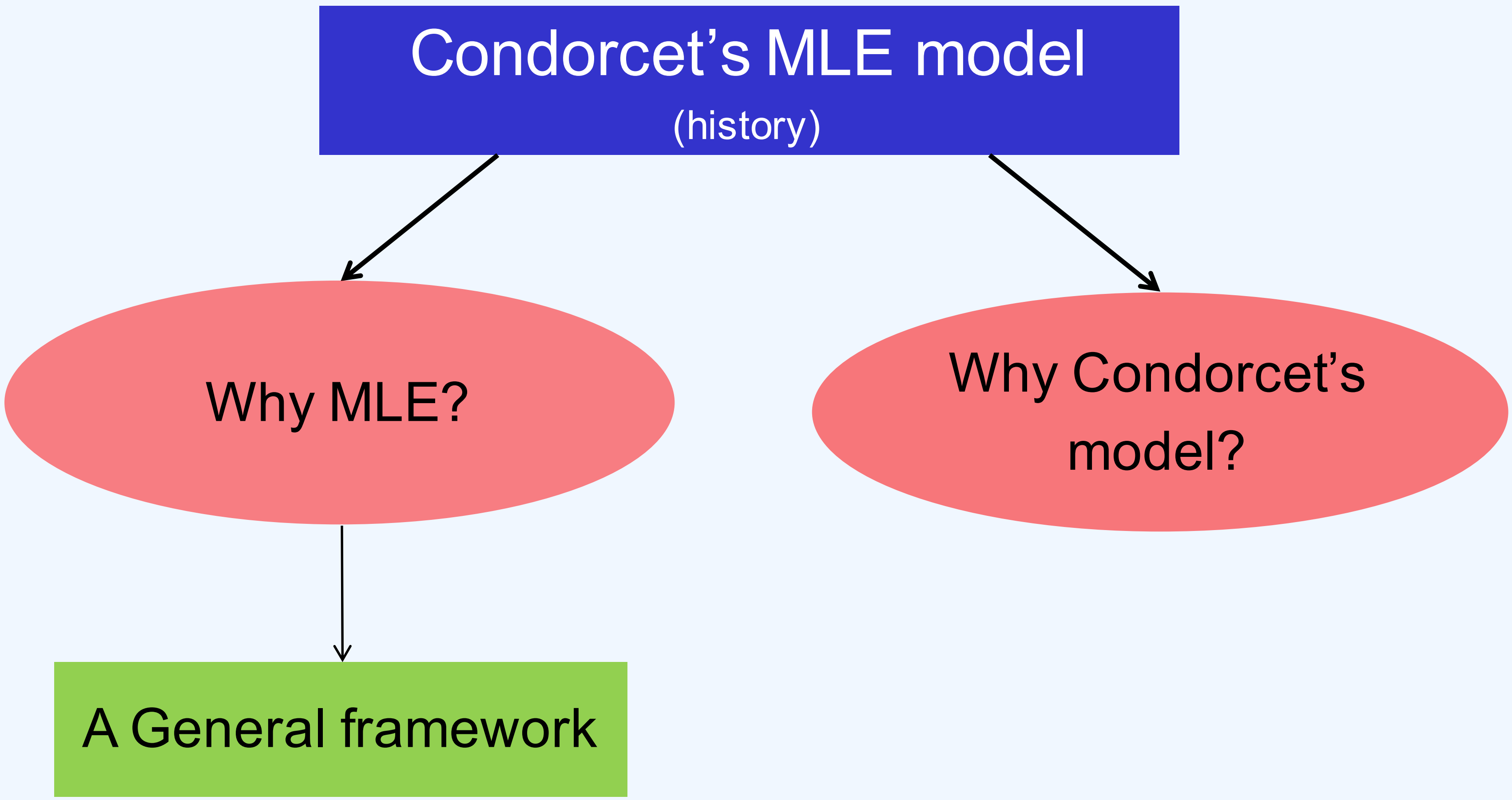


Classical Kemeny vs. Bayesian

<table>
<tr><th></th><th>Anonymity, neutrality, monotonicity</th><th>Consistency</th><th>Condorcet</th><th>Easy to compute</th></tr>
<tr><td>Kemeny (Fishburn version)</td><td rowspan="2">Y</td><td rowspan="2">N</td><td>Y</td><td>N</td></tr>
<tr><td>Bayesian</td><td>N</td><td>Y</td></tr>
</table>

Lots of open questions!

Outline: statistical approaches



Classical voting rules as MLEs
[Conitzer&Sandholm UAI-05]

- When the outcomes are winning alternatives
  - MLE rules must satisfy consistency: if $r(D_1) \cap r(D_2) \neq \phi$, then $r(D_1 \cup D_2) = r(D_1) \cap r(D_2)$
  - All classical voting rules except positional scoring rules are NOT MLEs
- Positional scoring rules are MLEs
- This is NOT a coincidence!
  - All MLE rules that outputs winners satisfy anonymity and consistency
  - Positional scoring rules are the only voting rules that satisfy anonymity, neutrality, and consistency! [Young SIAMAM-75]

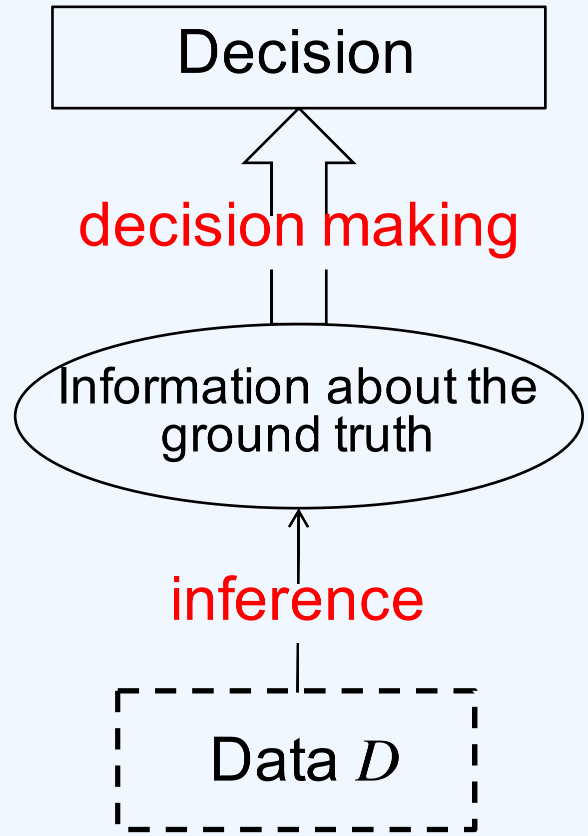
Classical voting rules as MLEs [Conitzer&Sandholm UAI-05]

- When the outcomes are winning rankings
  - MLE rules must satisfy reinforcement (the counterpart of consistency for rankings)
  - All classical voting rules except positional scoring rules and Kemeny are NOT MLEs
- This is not (completely) a coincidence!
  - Kemeny is the only preference function (that outputs rankings) that satisfies neutrality, reinforcement, and Condorcet consistency [Young&Levenglick SIAMAM-78]

Are we happy?

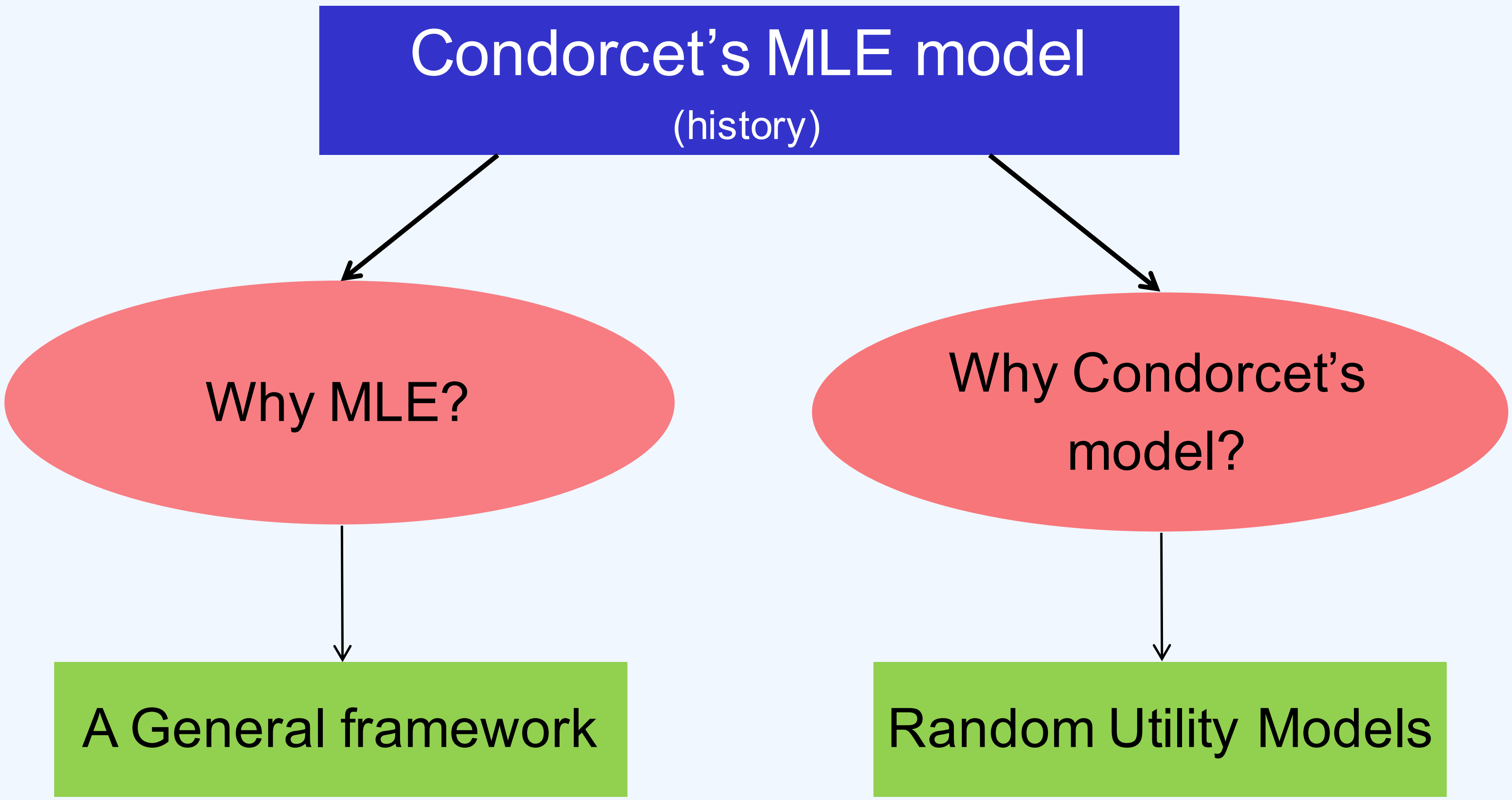
- Condorcet's model
  – not very natural
  – computationally hard
- Other classic voting rules
  – most are not MLEs
  – models are not very natural either
  – approximately compute the MLE

New mechanisms via the statistical decision framework



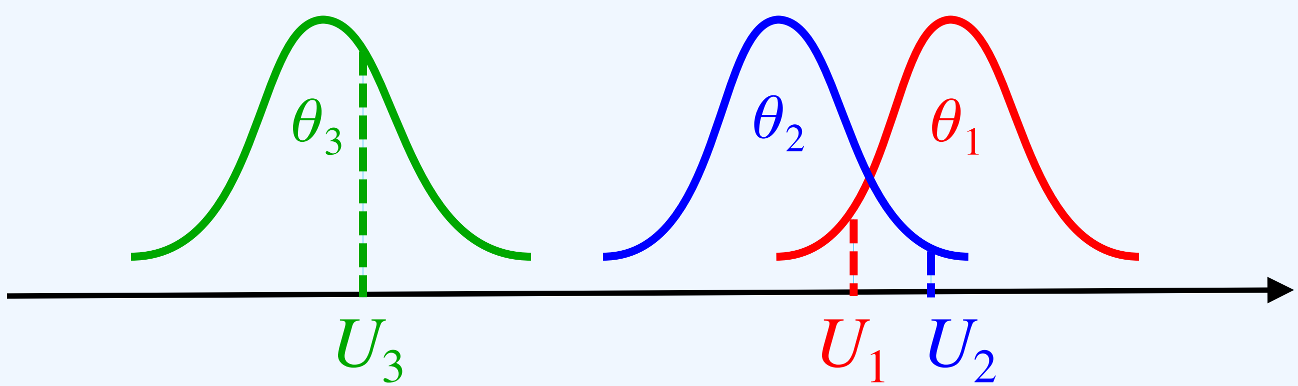
- Model selection
  - How can we evaluate fitness?
- Frequentist or Bayesian?
- Computation
  - How can we compute MLE efficiently?

Outline: statistical approaches



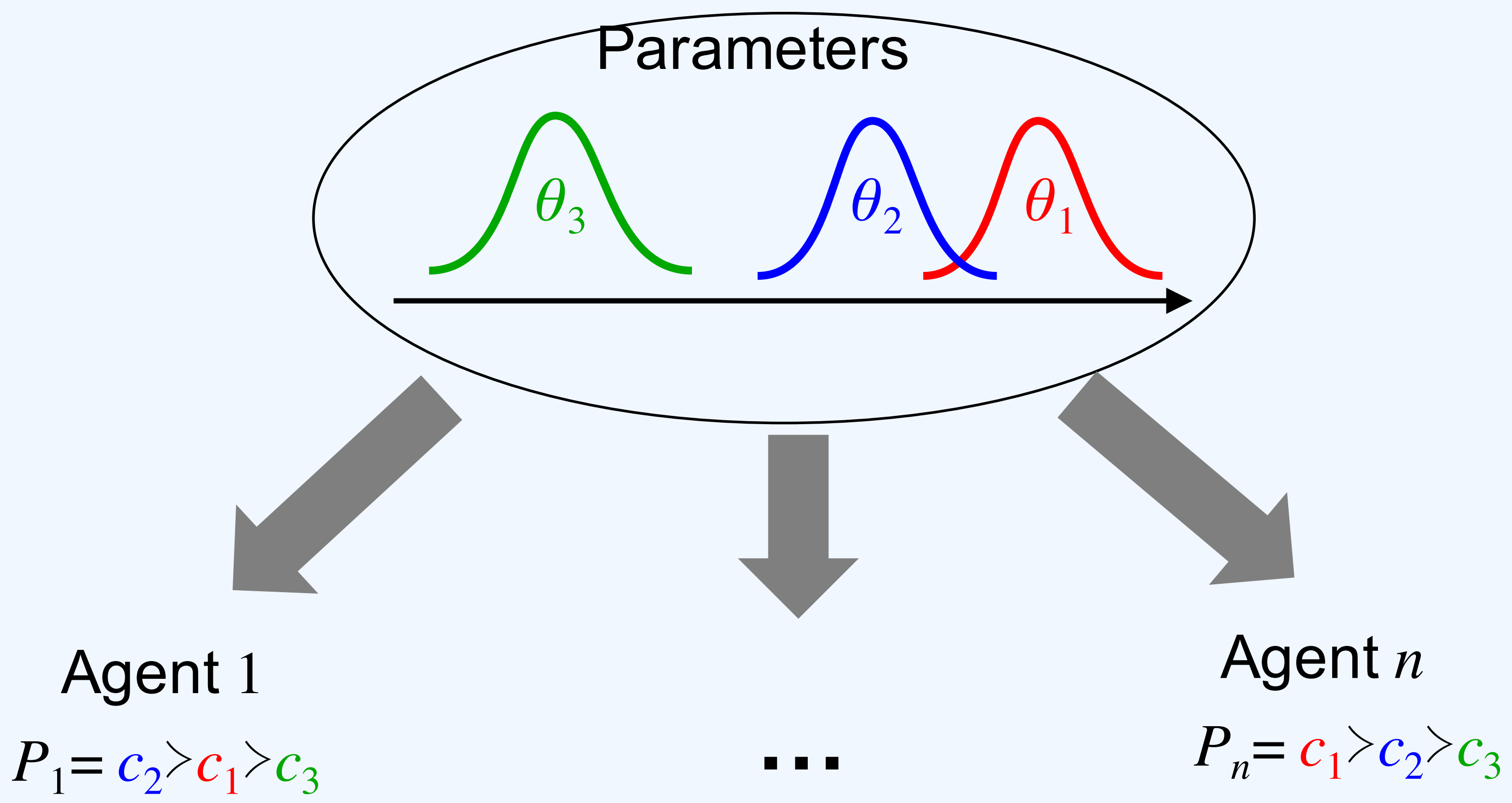
Random utility model (RUM)
[Thurstone 27]

- Continuous parameters: $\Theta=(\theta_1,\ldots, \theta_m)$
  - m: number of alternatives
  - Each alternative is modeled by a utility distribution μ_i
  - θ_i: a vector that parameterizes μ_i
- An agent's perceived utility U_i for alternative c_i is generated independently according to $\mu_i(U_i)$
- Agents rank alternatives according to their perceived utilities
  - $\Pr(c_2 \succ c_1 \succ c_3 | \theta_1, \theta_2, \theta_3) = \Pr_{U_i \sim \mu_i}(U_2 > U_1 > U_3)$

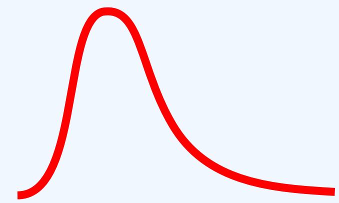


Generating a preference-profile

- $\Pr(\text{Data} \mid \theta_1, \theta_2, \theta_3) = \prod_{R \in \text{Data}} \Pr(R \mid \theta_1, \theta_2, \theta_3)$



RUMs with Gumbel distributions



- μ_i's are Gumbel distributions
  - A.k.a. the Plackett-Luce (P-L) model [BM 60, Yellott 77]
- Equivalently, there exist positive numbers $\lambda_1,\ldots,\lambda_m$

$$\Pr(c_1 \succ c_2 \succ \cdots \succ c_m \mid \lambda_1 \cdots \lambda_m) = \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_m} \times \frac{\lambda_2}{\lambda_2 + \cdots + \lambda_m} \times \cdots \times \frac{\lambda_{m-1}}{\lambda_{m-1} + \lambda_m}$$

c_1 is the top choice in $\{c_1,\ldots,c_m\}$

c_2 is the top choice in $\{c_2,\ldots,c_m\}$

c_{m-1} is preferred to c_m

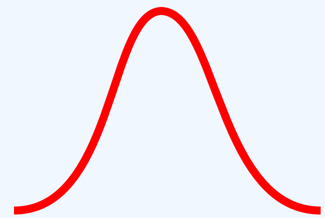
Pros:

- Computationally tractable
  - Analytical solution to the likelihood function
    - The only RUM that was known to be tractable
  - Widely applied in Economics [McFadden 74], learning to rank [Liu 11], and analyzing elections [GM 06,07,08,09]

Cons: does not seem to fit very well

RUM with normal distributions

- μ_i's are normal distributions
  - Thurstone's Case V [Thurstone 27]



Pros:

- Intuitive
- Flexible

Cons: believed to be computationally intractable

- No analytical solution for the likelihood function $\Pr(P \mid \Theta)$ is known

$$\Pr(c_1 \succ \cdots \succ c_m \mid \Theta) = \int_{-\infty}^{\infty}\int_{U_m}^{\infty}\cdots\int_{U_2}^{\infty} \boxed{\mu_m(U_m)}\boxed{\mu_{m-1}(U_{m-1})}\cdots\boxed{\mu_1(U_1)}\, dU_1 \cdots dU_{m-1}\, dU_m$$

$\boxed{U_m\text{: from } -\infty \text{ to } \infty}$ $\boxed{U_{m-1}\text{: from } U_m \text{ to } \infty}$ $\cdots$ $\boxed{U_1\text{: from } U_2 \text{ to } \infty}$

MC-EM algorithm for RUMs
[APX NIPS-12]

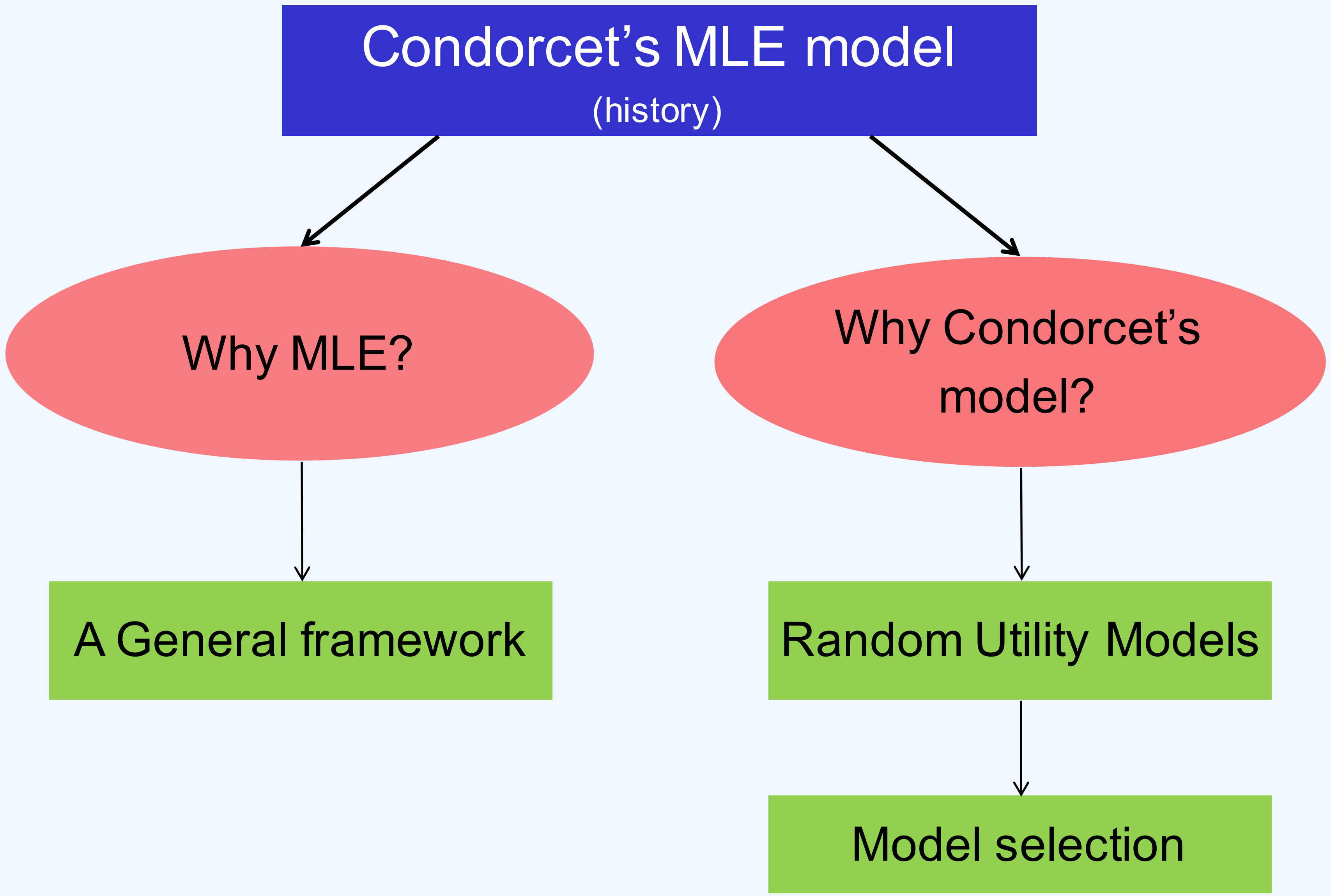
- Utility distributions μ_l's belong to the exponential family (EF)
  - Includes normal, Gamma, exponential, Binomial, Gumbel, etc.
- In each iteration t
- E-step, for any set of parameters Θ
  - Computes the expected log likelihood (ELL)

$$ELL(\Theta | \text{Data}, \Theta^t) = f(\Theta, \boxed{g(\text{Data}, \Theta^t))}$$

Approximately computed by Gibbs sampling

- M-step
  - Choose $\Theta^{t+1} = \text{argmax}_\Theta\, ELL(\Theta | \text{Data}, \Theta^t)$
- Until $|\Pr(D|\Theta^t)\text{-}\Pr(D|\Theta^{t+1})| < \varepsilon$

Outline: statistical approaches



Model selection

- Compare RUMs with Normal distributions and PL for
  - log-likelihood: log Pr(D|Θ)
  - predictive log-likelihood: E log Pr(D_{test}|Θ)
  - Akaike information criterion (AIC): $2k$-2log Pr(D|Θ)
  - Bayesian information criterion (BIC): $k$log n-2log Pr(D|Θ)
- Tested on an election dataset
  - 9 alternatives, randomly chosen 50 voters

Value(Normal) - Value(PL)	LL	Pred. LL	AIC	BIC
	44.8(15.8)	**87.4(30.5)**	**-79.6(31.6)**	-50.5(31.6)

Red: statistically significant with 95% confidence

Project: model fitness for election data