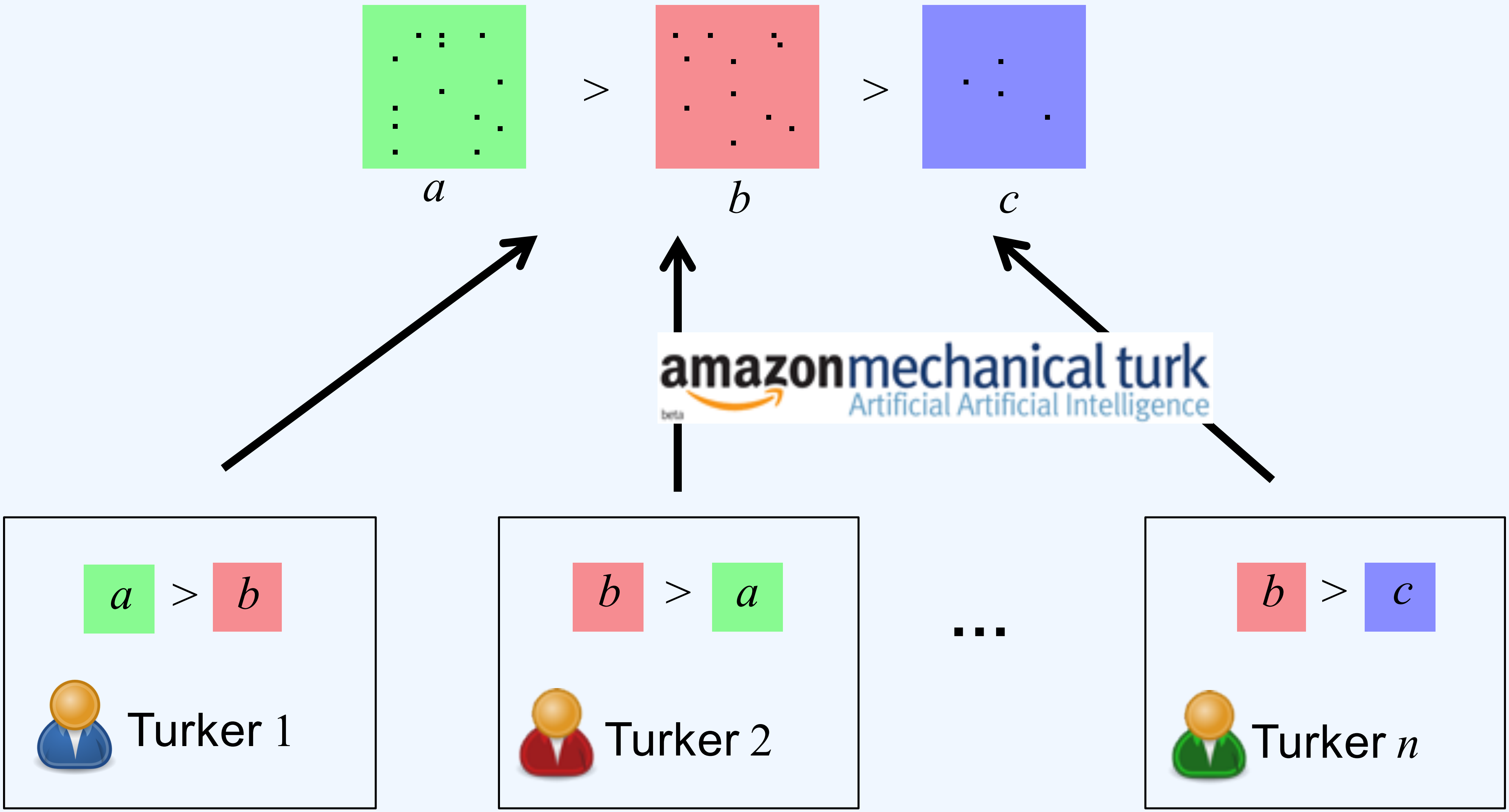
Computational social processes

Lirong Xia

Rensselaer

Fall, 2016

Example: Crowdsourcing

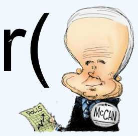
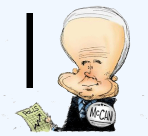


The Condorcet Jury theorem
[Condorcet 1785]

The Condorcet Jury theorem.

- Given
  - two alternatives $\{O,M\}$.
  - $0.5<p<1$,
- 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 $n\rightarrow\infty$, the majority of agents' preferences converges in probability to the ground truth

Pr( | )

= Pr( | )

= $p>0.5$

Today's schedule

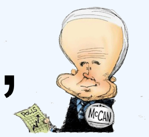
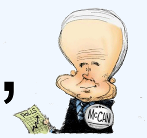
- Parametric ranking models
  - Distance-based models
    - Mallows
    - Condorcet
  - Random utility models
    - Plackett-Luce
- Decision making
  - MLE
  - Bayesian

Parametric ranking models

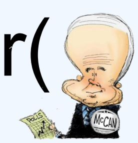
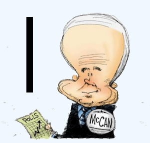
- A statistical model has three parts
  - A parameter space: Θ
  - A sample space: S = Rankings(A)^n
    - A = the set of alternatives, n=#voters
    - assuming votes are i.i.d.
  - A set of probability distributions over S:

    {$\Pr_\theta(s)$ for each $s \in$ Rankings(A) and $\theta \in \Theta$}

Example

- Condorcet’s model for two alternatives
- Parameter space Θ={  ,  }
- Sample space S = {  ,  }^{n}
- Probability distributions, i.i.d.

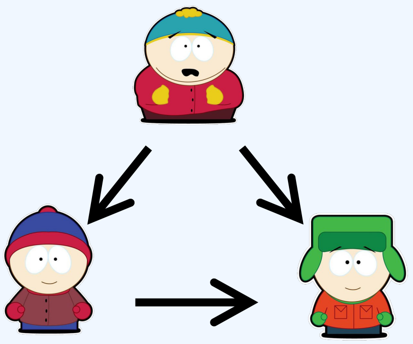
Pr( | )

= Pr( | )

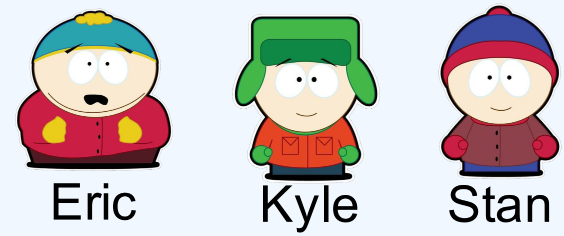
= p>0.5

Mallows' model [Mallows-1957]

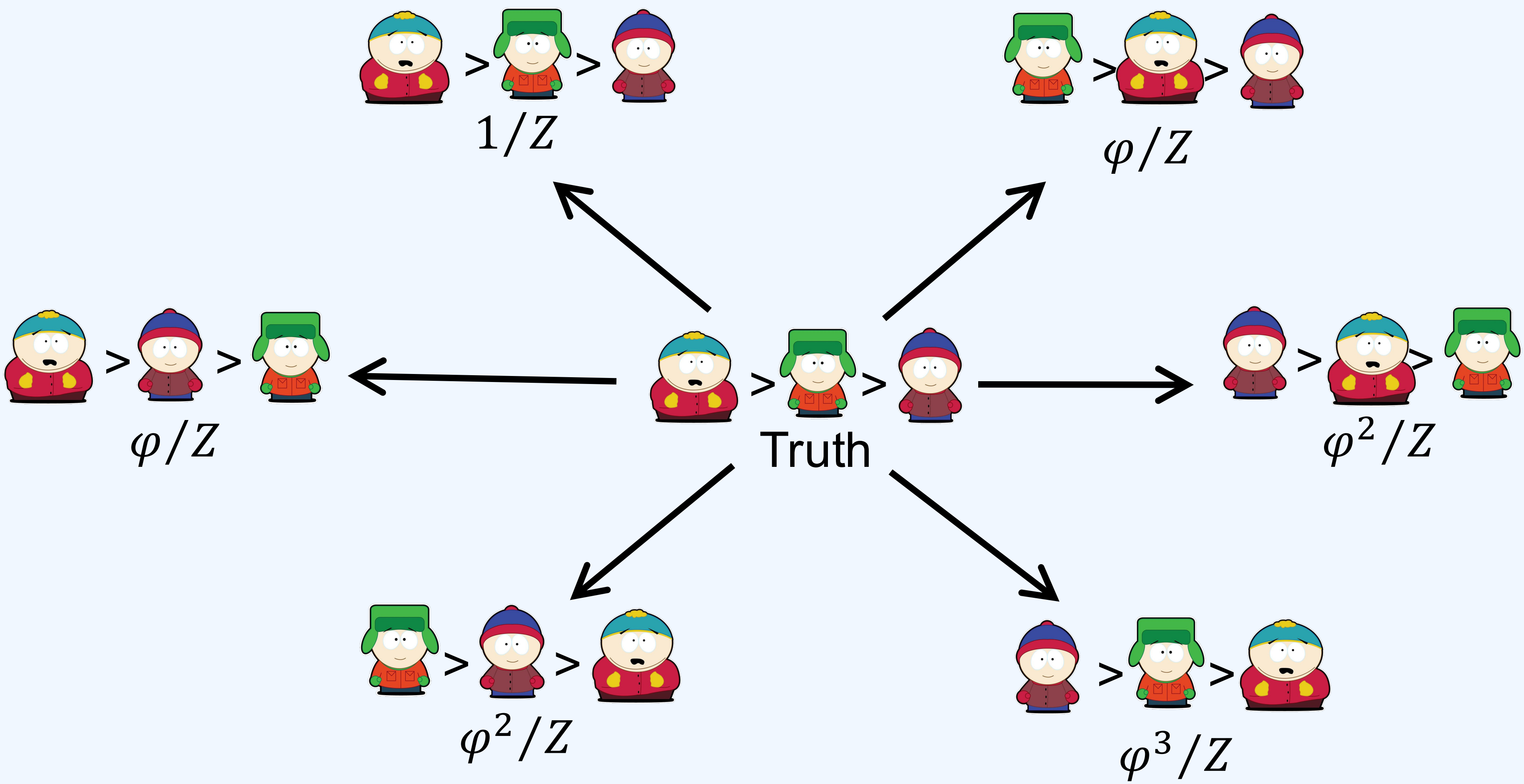
- Fixed dispersion φ <1
- Parameter space
  - – all full rankings over candidates
- Sample space
  - – i.i.d. generated full rankings
- Probabilities:



$$\Pr_W(V) \propto \varphi^{\text{Kendall}(V,W)}$$

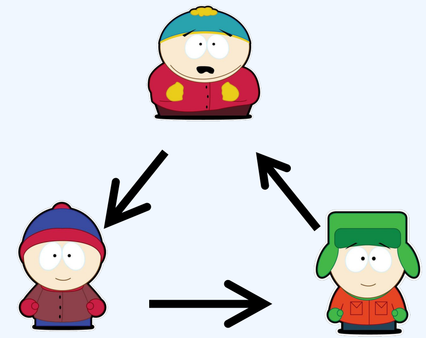
Example: Mallows for 

- Probabilities: $Z = 1 + 2\varphi + 2\varphi^2 + \varphi^3$



Condorcet's model

[Condorcet-1785, Young-1988, ES UAI-14, APX NIPS-14]

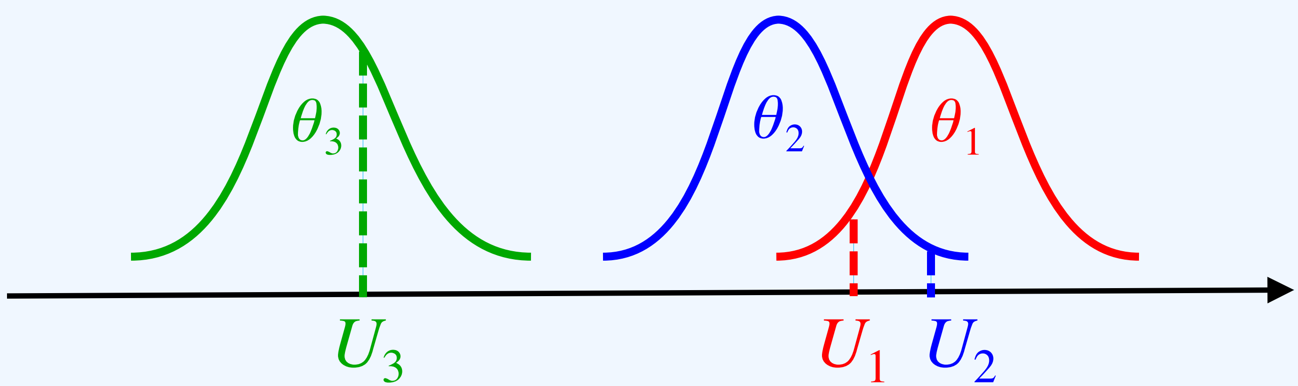


- Fixed dispersion φ <1
- Parameter space
  - all binary relations over candidates
- Sample space
  - i.i.d. generated binary relations
- Probabilities:

$$\Pr_W(V) \propto \varphi^{\text{Kendall}(V,W)}$$

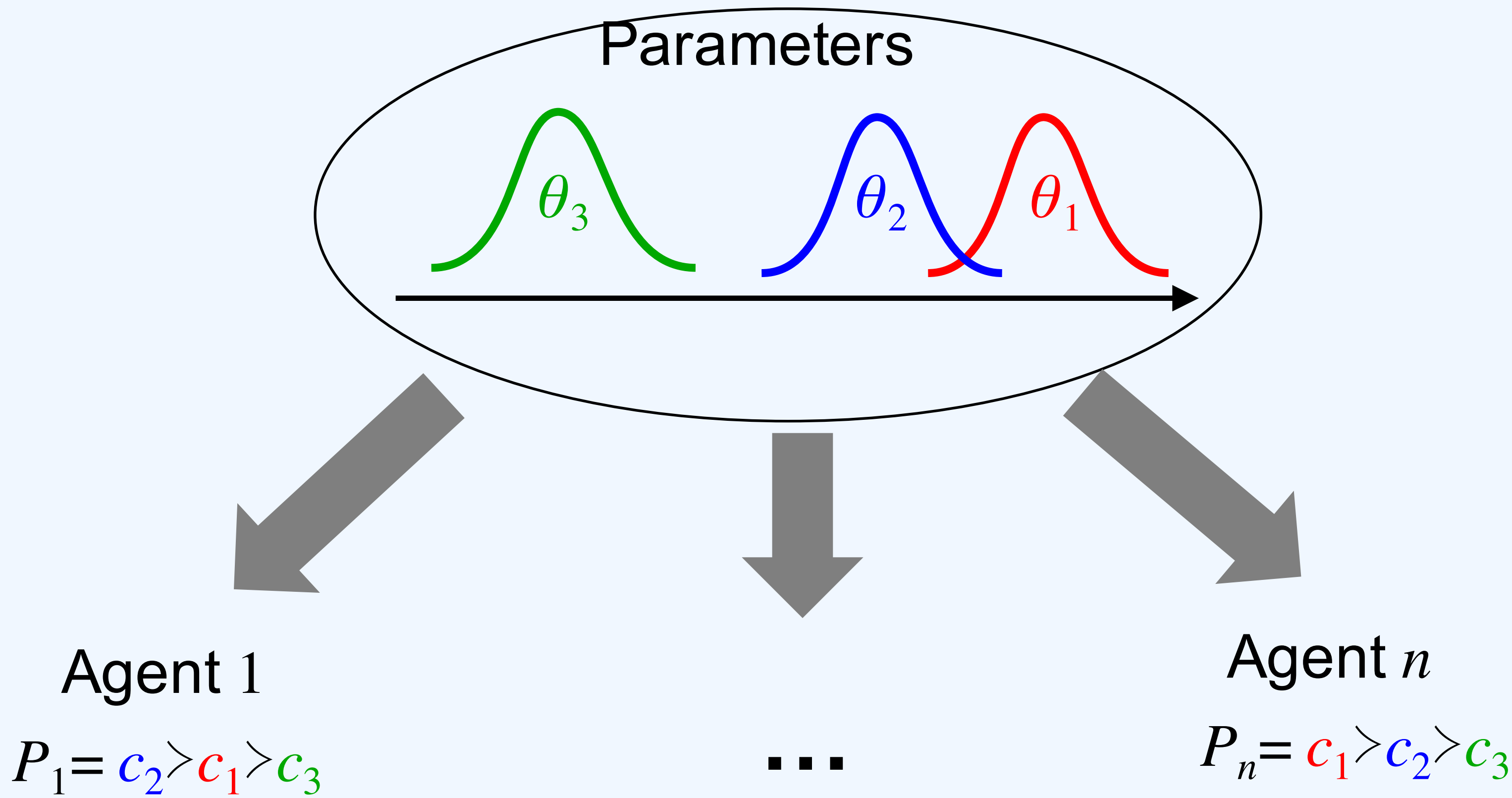
Random utility model (RUM) [Thurstone 27]

- 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 latent utility U_i for alternative c_i is generated independently according to $\mu_i(U_i)$
- Agents rank alternatives according to their perceived utilities
  - $\Pr(c_2 \succ c_1 \succ c_3 | \theta_1, \theta_2, \theta_3) = \Pr_{U_i \sim \mu_i}(U_2 > U_1 > U_3)$



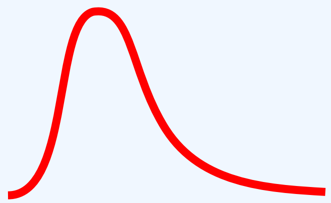
Generating a preference-profile

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



Plackett-Luce model

- μ_i's are Gumbel distributions
  - A.k.a. the Plackett-Luce (P-L) model [BM 60, Yellott 77]



- Alternative parameterization $\lambda_1,\ldots,\lambda_m$

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

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

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

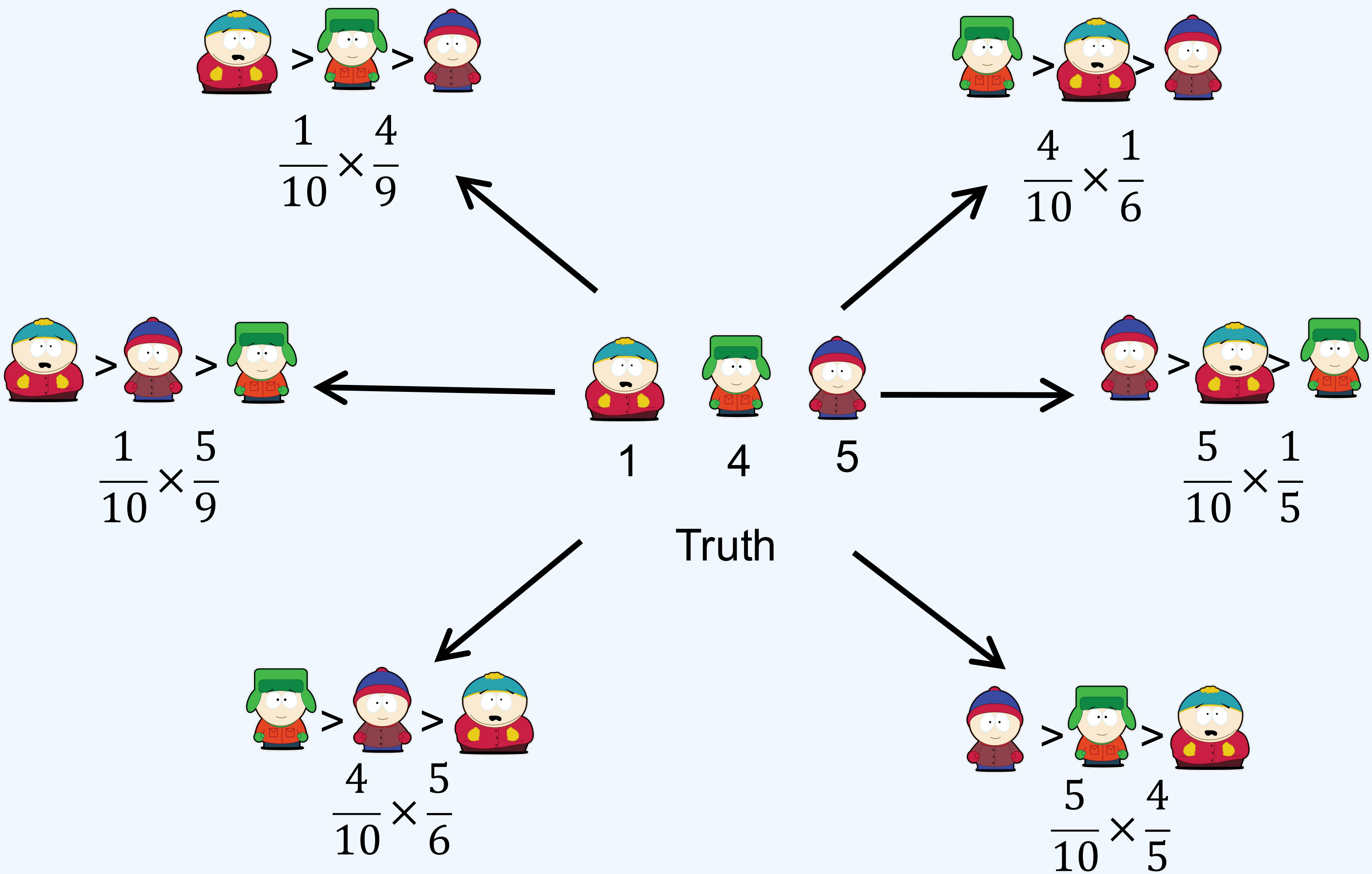
c_{m-1} is preferred to c_m

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]

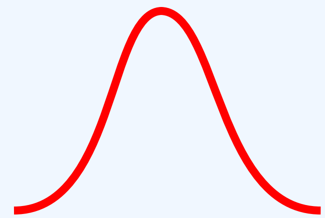
Cons: may not be the best model

Example



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

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

U_m: from -∞ to ∞ $\quad$ U_{m-1}: from U_m to ∞ $\quad \cdots \quad$ U_1: from U_2 to ∞

Model selection

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

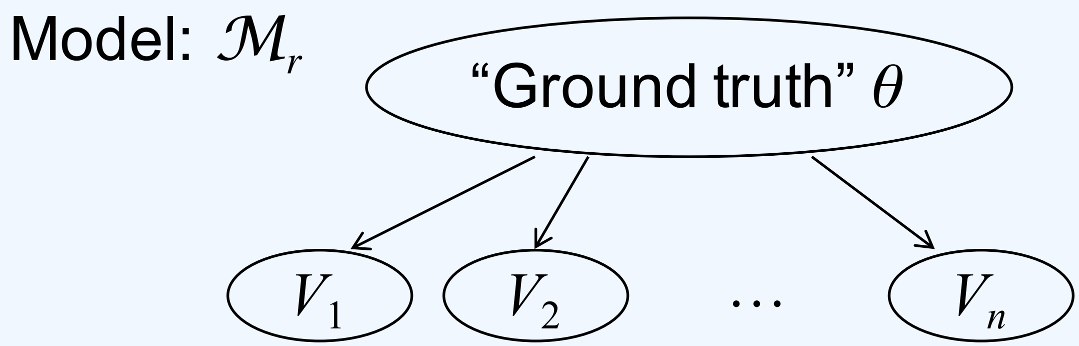
Value(Normal) - Value(PL)	LL	Pred. LL	AIC	BIC
	44.8(15.8)	**87.4(30.5)**	**-79.6(31.6)**	-50.5(31.6)

Red: statistically significant with 95% confidence

Project: model fitness for election data

Decision making

Maximum likelihood estimators (MLE)



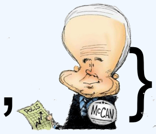
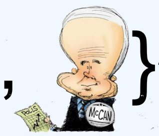
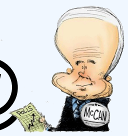
- For any profile $P=(V_1,\ldots,V_n)$,
  - The likelihood of θ is $L(\theta,P)=\Pr_\theta(P)=\prod_{V\in P}\Pr_\theta(V)$
  - The MLE mechanism

    MLE(P)=argmax$_\theta L(\theta,P)$
  - Decision space = Parameter space

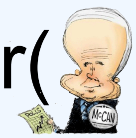
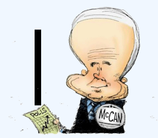
Bayesian approach

- Given a profile $P=(V_1,\ldots,V_n)$, and a prior distribution π over Θ
- Step 1: calculate the posterior probability over Θ using Bayes' rule
  - $\Pr(\theta|P) \propto \pi(\theta) \Pr_\theta(P)$
- Step 2: make a decision based on the posterior distribution
  - Maximum a posteriori (MAP) estimation
  - $\text{MAP}(P)=\text{argmax}_\theta \Pr(\theta|P)$
  - Technically equivalent to MLE when π is uniform

Example

- Θ={, }
- S = {, }^{n}
- Probability distributions:
- Data P = {10@ + 8@ }
- MLE
  - $L(O)=Pr_O(O)^6\ Pr_O(M)^4 = 0.6^{10}\ 0.4^8$
  - $L(M)=Pr_M(O)^6\ Pr_M(M)^4 = 0.4^{10}\ 0.6^8$
  - L(O)>L(M), O wins
- MAP: prior O:0.2, M:0.8
  - $Pr(O|P) \propto 0.2\ L(O) = 0.2 \times 0.6^{10}\ 0.4^8$
  - $Pr(M|P) \propto 0.8\ L(M) = 0.8 \times 0.4^{10}\ 0.6^8$
  - Pr(M|P)> Pr(O|P), M wins

Pr( | )

= Pr( | )

$= 0.6$

Decision making under uncertainty

Credit: Panos Ipeirotis & Roy Radner

- 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?

- MLE-based approach
  - 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!

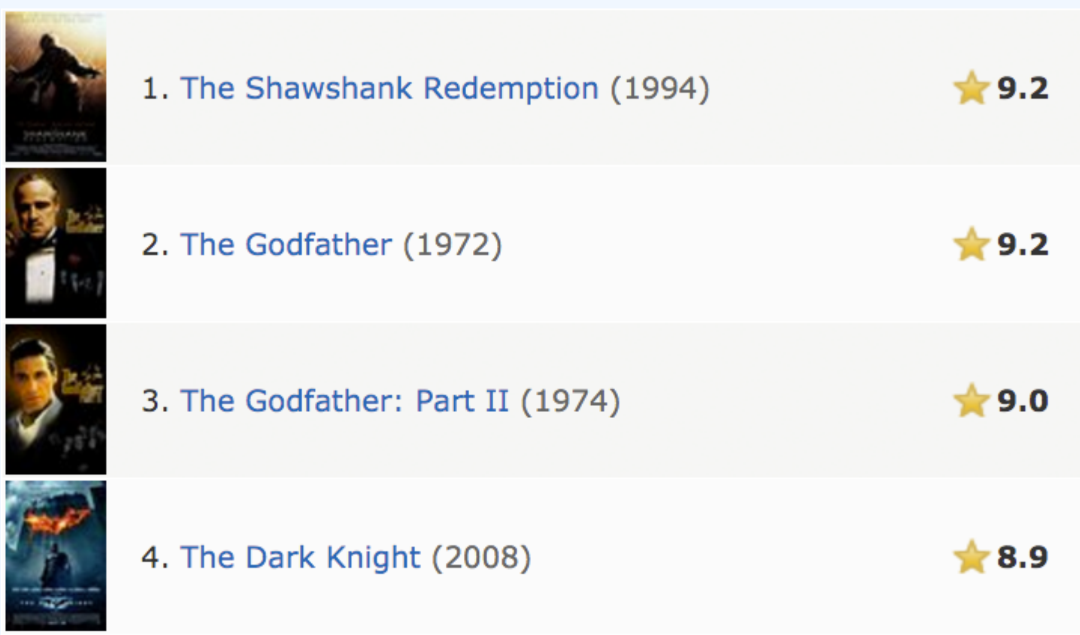
- Bayesian
  - the ground truth is captured by a belief distribution
  - Compute Pr(p|Data) assuming uniform prior
  - Compute Pr(2heads|Data)=0.485<0.5
  - No!

Statistical decision theory

- Given
  - statistical model: Θ, S, $\Pr_\theta(s)$
  - 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

Top 250 movies

- “Complex voter weighting system”
  - Claimed to be accurate
    - a “true Bayesian estimate”
  - Claimed to be fair



1.	The Shawshank Redemption (1994)	9.2
2.	The Godfather (1972)	9.2
3.	The Godfather: Part II (1974)	9.0
4.	The Dark Knight (2008)	8.9

Different Voice

- Q: “This is *unfair* ! ”
  - “That film / show has received awards, great reviews, commendations and deserves a much higher vote!”
- IMDB: “…only votes cast by IMDb users are counted. We do not delete or alter individual votes”

IMDb Votes/Ratings Top Frequently Asked Questions
http://www.imdb.com/help/show_leaf?votestopfaq

Fairness of Bayesian estimators

- Theorem: Strict Condorcet

  No Bayesian estimator satisfies strict Condorcet criterion

- Theorem: Neutrality

  Neutral Bayesian estimators

  = Bayesian estimators of "neutral" models