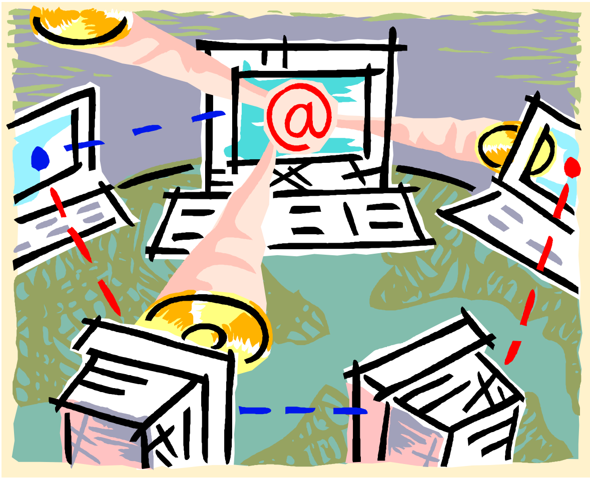
Social Choice and Social Networks

Aggregation of General Biased Signals

(DRAFT)

Elchanan Mossel
UC Berkeley
8 Sep 2010



Extending Condorect’s Jury Theorem

- We want to consider extensions of Condorcet’s Jury Theorem to situations where the signals are not binary.

- n people need to take a decision between k alternatives denoted 1,…,k.

- One of the k alternatives is correct.

- We can make various assumptions on the type of signals individuals receive.

- Such as …

Extending Condorect's Jury Theorem

• We can make various assumptions on the type of signals individuals receive.

• The most general assumption is:
• A signal space X and distributions $P_1,\ldots,P_k$ on X
• where if state of the word = i, signal ~ P_i

• A nice non-general assumption is:
• Signal= correct alternative with probability $p > 1/k$ and each other answer with probability $(1-p)/(k-1)$.

• We will discuss the least general case in class and leave the most general case as HW ☺

Extending Condorect's Jury Theorem

- Consider the setup where each signal equals the correct state of the world with probability $p>1/k$ and each of the other states with probability $q=(1-p)/(k-1)$.

- What should the aggregation function be?

Extending Condorect's Jury Theorem

- What should be the aggregation function?

- Write $n_a(x)$:= # of a's in the vector x,
- Def: A Pluraliry function $f : [k]^n \rightarrow [k]$ is defined in the following way: $f(x) = a$ if $n_a(x) > n_b(x)$ for all $b \neq a$.

- Def: A function $f : [k]^n \rightarrow [k]$ is fair if for all $\sigma \in S_k$ it holds that $f(\sigma x) := f(\sigma x_1, \ldots, \sigma x_n) = \sigma f(x)$.

- Note: fairness corresponds to treating all alternatives equally – their names do not matter.

- Claim (HW): For all k,n there exists a fair Plurality function.

Extending Condorect’s Jury Theorem

- Consider the setup where each signal equals the correct state of the world with probability $p>1/k$ and each of the other states with probability $q=(1-p)/(k-1)$. Assume further a uniform prior.

- <u>Thm:</u> Assume $p>1/k$. Write $c(n) = P[\text{Plurality is correct}]$ then:
- As $n \to \infty \quad c(n) \to 1$.
- The same is true even for $p(n) – 1/k >> n^{-1/2}$.
- If $p(n)-1/k = o(^{-1/2})$ then $c(n) \to 1/k$.
- Writing $a(n) = p(n)-q(n)$ we have for all n that
- $c(n) \geq 1 – 2 k \exp(-a^2 n(n))$

- <u>Pf:</u> ???

Extending Condorect's Jury Theorem

- Thm: Assume $p>1/k$. Write $c(n)$ = P[Plurality is correct] then:
- As $n \rightarrow \infty$ $c(n) \rightarrow 1$.
- The same is true even for $p(n) - 1/k >> n^{-1/2}$.
- If $p(n)-1/k = o(^{-1/2})$ then $c(n) \rightarrow 1/k$.
- Writing $a(n) = p(n)-q(n)$ we have for all n that
- $c(n) \geq 1 - 2\,k \exp(-a^2 n)$

- Pf: Generalize proofs of the binary case.

The estimation point of view

- Claim: Plurality maximizes the probability of being correct among fair functions.

-

The estimation point of view

- Claim: Plurality maximizes the probability of being correct among fair functions.

- Pf:
- Same as proof for majority.
-
- In a way this is a classical estimation problem.
- There is a random variable S with uniform prior.
- Our goal is to estimate S given the signals $X_1,\ldots X_n$.
- We choose the S which maximizes $P[S \mid X_1,\ldots X_n]$.
- Since the prior is uniform this is the same as finding the S maximizing $P[X_1,\ldots X_n \mid S]$

The estimation point of view

• The estimation point of view is valid also for the general signals picture:

• optimal choice function chooses the s maximizing
• $P[S = s \mid X_1,\ldots,X_n]$

• Need to think carefully how to measure bias to obtain guarantees.

• Can apply general results from statistics to obtain similar results to the ones above.

More general signals – Example 1

- Two alternatives: +,-:
- Vote for correct alternative with probability p
- Vote for wrong alternative with probability q < p
- Do not vote with probability 1-p-q

- Q: Assuming prior correctness (1/2,1/2) what is the optimal aggregation function?

More general signals – Example 2

- Two alternatives: +,-:
- Vote for correct alternative with probability p
- Vote for wrong alternative with probability q
- Do not vote with probability 1-p-q

- Q: Assuming prior (1/2,1/2) what is the optimal aggregation function?
- Q: How large should p-q be to aggregate well?

- HW
- For the second question $p-q \gg n^{-1/2}$ always suffice thought in some cases less suffices (e.g. $q=0$, $p = \log n/n$)

Example 2 of more general signals

- There are k possible alternatives.
- Each voter receives a ranking where:
- The correct alternative is at location i with prob. p_i
- All other alternatives are placed uniformly at random.
- $p_1 > p_2 > \ldots > p_k$
-

•Q: Assuming a uniform prior what is the optimal a
function?

• Q: What is the difference needed
between the p_i to aggregate well?

• A generalization of a voting method invented
by Borda (1733 – 1799; mathematician, physicist, political scientist, and sailor)

Beyond the Plurality Function

- Further questions:

- What about other aggregation functions?
- E.G: U.S Electoral college?
- Other functions?

- We'll assume simple signals: correct outcome with probability $p>1/k$ -all other outcome equally likely.

Beyond Condorcet's Jury Theorem

- We want to consider again functions that are:

- Fair – names of alternatives do not matter.

- Monotone – stronger vote in one direction should not hurt this direction.

Fairness

- Recall: in the binary case we said f is fair if
- $f(-x) = -f(x)$.
-
- In the general case the definition is:
- <u>Def:</u> A function $f : [k]^n \rightarrow [k]$ is <u>fair</u> if for all $\sigma \in S_k$ it holds that $f(\sigma x) := f(\sigma x_1, \ldots, \sigma x_n) = \sigma f(x)$.

- How to define monotone? – stronger vote in one direction should not hurt this direction.

Monotonicity

- Def: for two vector $x,y \in [k]^n$ and $a \in [k]$ write:
- $x \leq_a y$ to indicate that:
- Whenever $x_i \neq y_i$ it holds that $y_i = a$.

- "y is more leaning towards a than x".

- Def: A function $f : [k]^n \rightarrow [k]$ is monotone if
- for all $a \in [k]$ and all $x,y \in [k]^n$ it holds that:
- $x \leq_a y \Rightarrow f(x) \leq_a f(y)$

- If a wins for vote x it also wins for vote y.

- Definition from Kalai-Mossel(????)

An Aggregation Theorem

- Recall that $f : [k]^n \to [k]$ is invariant to a transitive action G on [n] if
- for all $\sigma \in G$ it holds that
- $f(x_\sigma) = f(x)$

- <u>Thm (Kalai-Mossel???):</u>
- $\forall$ k $\exists$ C = C(k) s.t.
- $\forall \varepsilon < 1/3$, $\forall$ monotone transitive $f : [k]^n \to [k]$ where for $p=1/k$ it holds that $P[f=a] \geq 1/(2k)$ for all a it holds that:
- for $p > 1/k + C (\log(1-\varepsilon) - \log(1/2k)) \log\log n / \log n$:
- P[f is correct] $\geq 1-\varepsilon$.
-
- Proof similar to previous proof we haven't seen …
- Examples?

An Aggregation Theorem

- Thm (Kalai-Mossel???):
- $\forall$ k $\exists$ C = C(k) s.t.
- $\forall\ \varepsilon < 1/3$, $\forall$ monotone transitive $f : [k]^n \rightarrow [k]$ where for $p=1/k$ it holds that $P[f=a] \geq 1/(2k)$ for all a it holds that:
- for $p > 1/k + C (\log(1-\varepsilon) - \log(1/2k)) \log \log n / \log n$:
- $P[f \text{ is correct}] \geq 1-\varepsilon$.

- <u>Examples:</u>
- "Electoral college" with all states of equal size.
- Plurality

- In fact in all of the examples above a bias of $Cn^{-1/2}$ suffice.

Aggregation of opinions with additional structure

• So far we assumed that the different alternatives and signals have no “additional structure”.

• We now consider two examples of such structures.

• The first example deals with signals that are real numbers.

• The second example deals with rankings.

Aggregating real #'s

- Assume that the true state of the world is s=+ or s=-

- Each voter receives a real signal N(a s, 1) where a> 0 is some constant.

- how should people vote?

Aggregating real #’s

- Assume that the true state of the world is s=+ or s=-

- Each voter receives a real signal N(a s, 1) where a> 0 is some constant.

- how should people vote?

- One option is for voters who got a positive signal to vote + and voters who got negative signal to vote -.

- Is this good?

Aggregating real #'s

- Assume that the true state of the world is s=+ or s=-

- Each voter receives a real signal N(a s, 1) where a> 0 is some constant.

- how should people vote?

- One option is for voters who got a positive signal to vote + and voters who got negative signal to vote -.

- Is this good?

- It is pretty good. For example $a \gg n^{-1/2}$ suffice to get the correct answer with probability $\rightarrow$ 1.

Aggregating real #'s

- Is this good?

- It is pretty good. For example $a >> n^{-1/2}$ suffice to get the correct answer with probability → 1.

- But it is not optimal.

- The optimal rule is: each voter declare their signal X_i.
- Winner is the sign of $\sum X_i$
- This is the best Bayesian decision rule (assuming (½, ½) prior).
- But note: this rule let's one cheater determine the outcome of election (while majority is more robust).

Aggregating Real #'s

- More generally:
- Thm (Keller, Mossel, Sen 10):
- If f is a monotone transitive function $f : R^n \rightarrow \{-,+\}$ with $E_{a=0}[f] = 0$ then $E_a[f] \rightarrow 1$ if $a >> \log n^{-1/2}$

- So any democratic function would work.
- Non democratic functions (e.g. dictator, functions of a few voters) will not aggregate even for a constant a.

Aggregating Permutations

- Next we discuss n voters who rank k alternatives.

- The outcome should be a ranking of the k alternatives.

- <u>Q: </u>Should we use a plurality vote?

Aggregating Permutations

• Next we discuss n voters who rank k alternatives.

•The outcome should be a ranking of the k alternatives.

•

• Q: Should we use a plurality vote?

• A1: May not be a good idea:

• Consider a distribution P where the true permutation is at least twice as likely as any other permutation.

• If we apply plurality rule we may need many (order k!) voters to get a good answer.

• If k is large – this is too big!

Aggregating Permutations

- Next we discuss n voters who rank k alternatives.

- The outcome should be a ranking of the k alternatives.

- A2: Suppose 99% of voters rank some alternative at the top. It is a “no brainer” that this alternative should be at the top. However plurality may very well not do it.

- We are using the wrong model!
- A better model coming next …

Consensus Ranking, Rearrangements and the Mallows Model

- Given a set of rankings $\{\pi_1, \pi_2, \ldots \pi_N\}$ a natural output to consider is **consensus ranking** which is the following "average":

$$\pi_0 = \text{argmin} \sum_{i=1}^{N} d(\pi_i, \pi_0)$$

for d = distance on the set of permutations of n objects

Most natural is d_K which is the Kendall distance.

$$d_K(\pi, \text{id}) = \sum_{i<j} 1_{j \prec_\pi i}$$

$$d_K(\pi, \pi') = d_K(\pi(\pi')^{-1}, \text{id}) = \sum_{i \prec_{\pi'} j} 1_{j \prec_\pi i}$$

- Kendall – tau rank statistics uses the Kendall rank to test if two variables are statistically independent.
- Kendall (1907-1988) was an English statistician.
- One of the first to argue shares perform a "random walk"

The Mallows Model – A distribution on rearrangements

- Mallow's voting model:

- The Mallow's model is an Exponential family model in β:

  - $P(\pi \mid \pi_0) = Z(\beta)^{-1} \exp(-\beta\, d_K(\pi,\pi_0))$

- If $\beta>0$ then given rankings $\pi_1,..,\pi_k$, the consensus ranking output is the ML estimator of the original ranking assuming a uniform prior.

- This model was suggested by
  Collin Mallows in
  "Non-null ranking models" I, (1957).

-

Natural Questions Regarding the Model

- How many voters are needed in order to be able to recover the true ranking with good probability?

- Algorithmically: how can one find the average of these permutations?

- Assume β is some fixed constant.

Related work

- Cohen,Schapire,Singer 99: Greedy algorithm (**CSS)**
- Meila,Phadnis,Patterson,Bilmes 07:

  Branch and Bound algorithm – exponential running time.
- J. Bartholdi, III, C. A. Tovey, and M. A. Trick 98: Proved NP-hardness

  "Voting Scheme for which it can be difficult to tell who won the election"
- Ailon,Newman,Charikar 05 Randomized algorithm
  - guaranteed 11/7 factor approximation (**ANC**)

- Mathieu, 07: $(1+\varepsilon)$ approximation, time $O(n^6/\varepsilon^{+2\text{^}2O(1/\varepsilon)})$

Efficient Sorting of Mallow's model of rearrangements (problem 3)

- [Braverman-Mossel-10]:

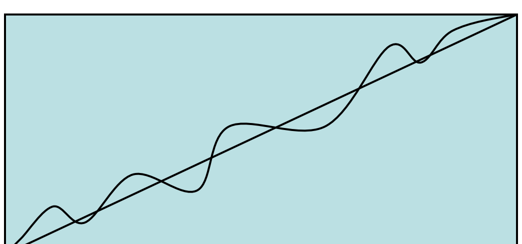
- Given r independent samples from the Mallows Model, find ML solution exactly! in time n^b,where
- $b = 1 + O((\beta\, r)^{-1})$,
- where r is the number of samples
- with high probability (say $\geq 1\text{-}n^{-100}$)

Sorting the Mallow's Model (Problem 2)

- [Braverman-M-10]: Optimal order can be found in polynomial time and $O(n \log n)$ queries.
- <u>Proof Ingredient 1:</u> "statistical properties" of generated permutations π_i in terms of the original order π_0
- Let τ rank elements according to their average location on the r generate permutations then:
- With high probability: $\sum_x |\pi_0(x) - \tau(x)| = O(n / \beta r)$,

  $\max |\pi_0(x) - \tau(x)| = O(\log n / \beta r)$

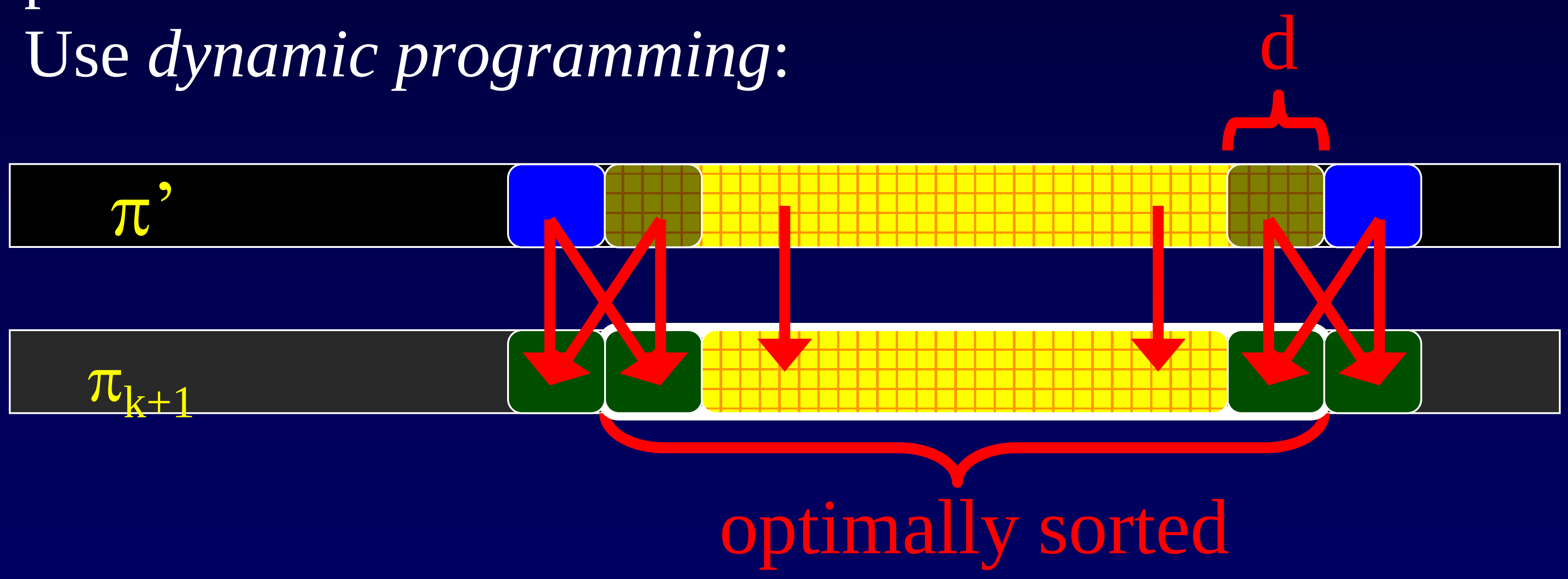
- <u>Additional ingredient:</u> A dynamic programming algorithm to find π given a starting point where each elements is at most k away with running time $O(n\, 2^{6k})$



- The proof is hard – we'll describe the algorithmic ingredient in more detail
- Following slides courtesy of M. Braverman

The algorithm assuming small deviation

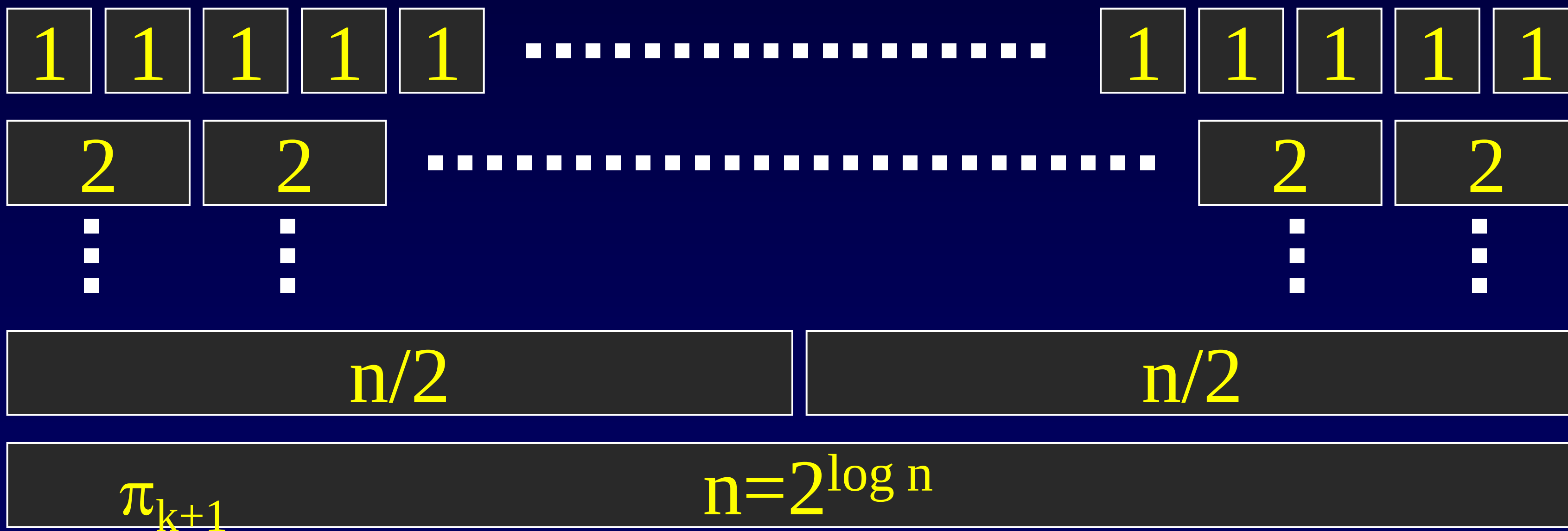
- <u>The problem now:</u> Find the optimal ordering π_{k+1} such that each element is at most $d = O(\log n)$ away from its position in π'.
- Use *dynamic programming*:



- For each interval there are $<2^{4d}$ "variations".
- A total of poly(n) variations, can store all of them.

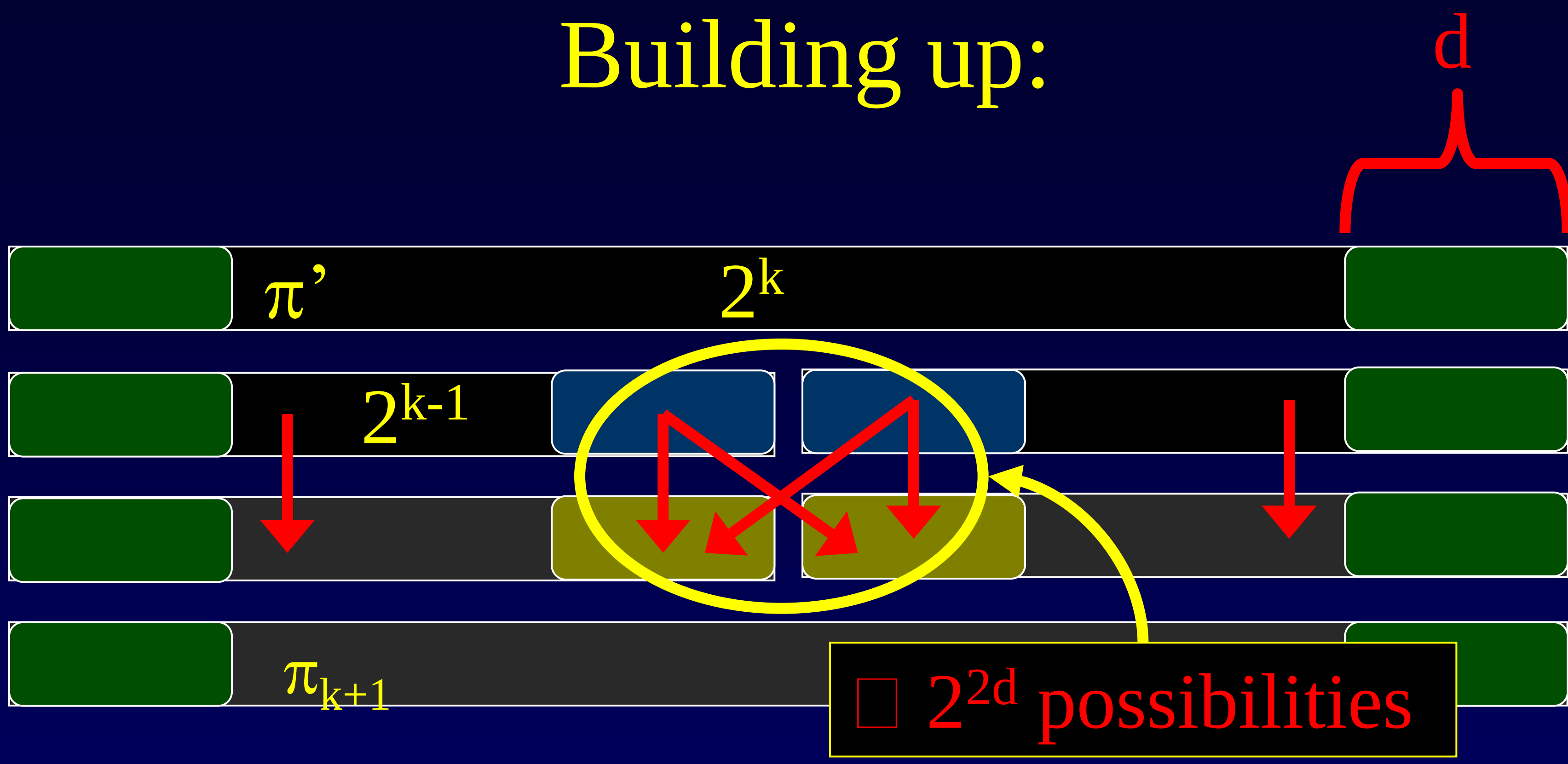
The algorithm assuming small deviation

- Store optimal permutations of all variations on the following intervals:



- A total of $\tilde{O}(2^{4d}\, n)$ storage.
- Work from shorter intervals to longer.

Building up:



- Each of the shorter intervals has been pre-sorted.
- Thus the cost of doing all intervals on level k is

#intervals × #checks × #cost/check = $(n/2^k)\ 2^{4d} \times 2^{2d} \times 2^{2k}$.

- Thus, total running time is bounded by $O(2^{6d}\ n^2)$.

Some Notes on Related Problems

- Except for it’s social context the problem above is an example of
- Sorting from noisy information.
- Here are a couple more examples of the same form.

Example Consensus Ranking, Rearrangements and the Mallows Model

- **Problem** 1: Consider a sorting problem where for each query comparing two elements x and y:
- Return correct answer with probability ½+ ε
- Independently for each query.
- Can query each pair as many times as we want.
- How many queries are needed to find correct order with probability 0.9999?
- Feige, Raghavan, Peled and Upfal.

- **Problem** 2: Consider a sorting problem where for each query comparing two elements x and y.
- Return correct answer with probability ½+ ε
- Independently for each query.
- For each element can query only once.
- What can we do?
- Again ML solution found in Braverman-Mossel-09.

HW due in 3 weeks

- Q1: The Plurality function:

Prove that for all k,n there exists a fair Plurality function.

- Prove that if voters receive independent signals with the correct alternative with probability $p>1/k$ and all other alternatives with equal probability, then Plurality maximizes the probability of correct vote among all functions (assuming a uniform prior).

- Assume $p(n)$ depends on n and write $q(n)$ for the probability of receiving incorrect signal and $a(n) = p(n)-q(n)$.
- Show have for all n that
- $c(n) := P[\text{Plurality is correct}] \geq 1 - 2\, k \exp(-a^2(n)\, n)$
- (k = # alternatives).

HW due in 3 weeks

•

- Q2 – Bonus : The weakest aggregation functions.

- Consider independent signals N(a s, 1).
- Find a transitive monotone function with $E_{a=0}[f] = 0$
- where $E_a[f] \rightarrow 0$ for all a(n) << $(\log n)^{-1/2}$

- Consider binary signals which are correct with probability 1/2 + a(n).
- Find a fair transitive monotone function that does not aggregate well for all a << $(\log n)^{-1}$