Achieving budget-optimality with adaptive schemes in crowdsourcing

Sewoong Oh

Department of Industrial and Enterprise Systems Engineering
University of Illinois Urbana-Champaign

joint work with Ashish Khetan
(David Karger, Devavrat Shah, Jungseul Ok, Jinwoo Shin, Yung Yi)

Adaptive task assignment for crowdsourced classification

Sewoong Oh

Department of Industrial and Enterprise Systems Engineering
University of Illinois Urbana-Champaign

joint work with Ashish Khetan
(David Karger, Devavrat Shah, Jungseul Ok, Jinwoo Shin, Yung Yi)

Crowdsourcing systems

Marketplace to get labels for training data

Quality of the labels can be very low

70/100

30/100

Add redundancy to cope with noise

Help find missing...

Help find missing...

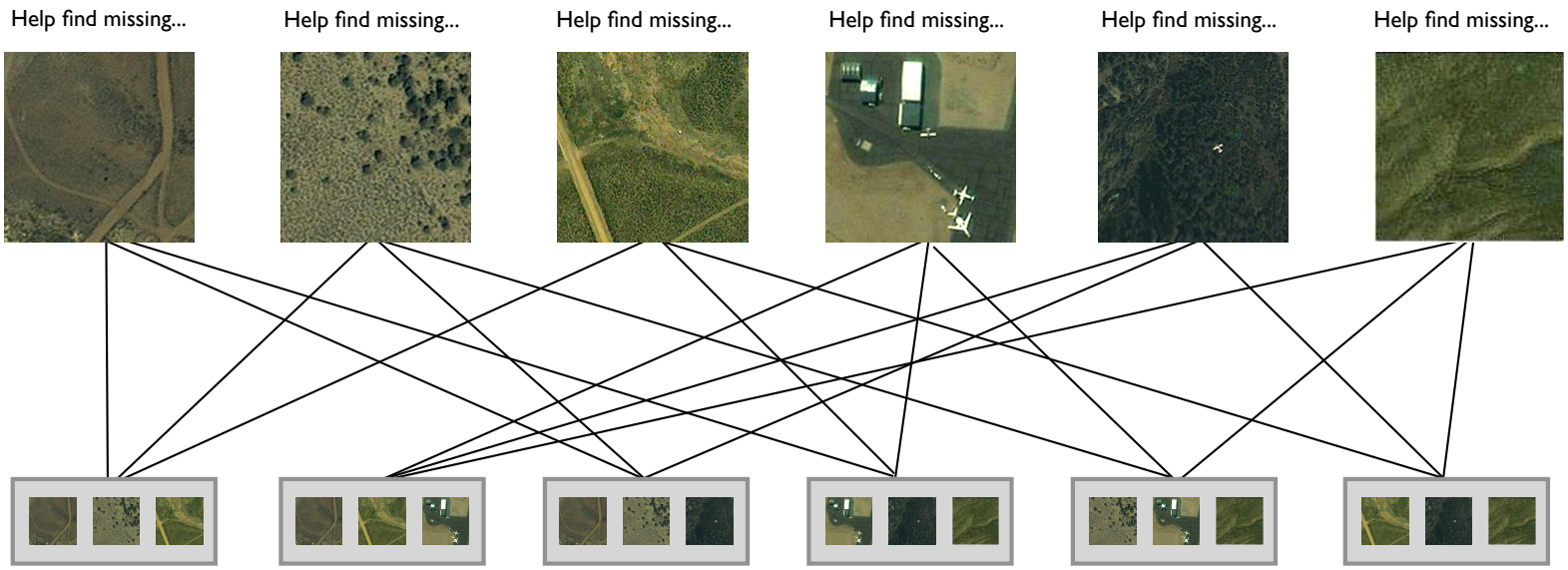
Help find missing...

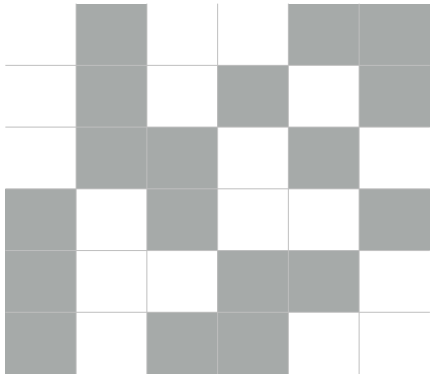
Help find missing...

Help find missing...

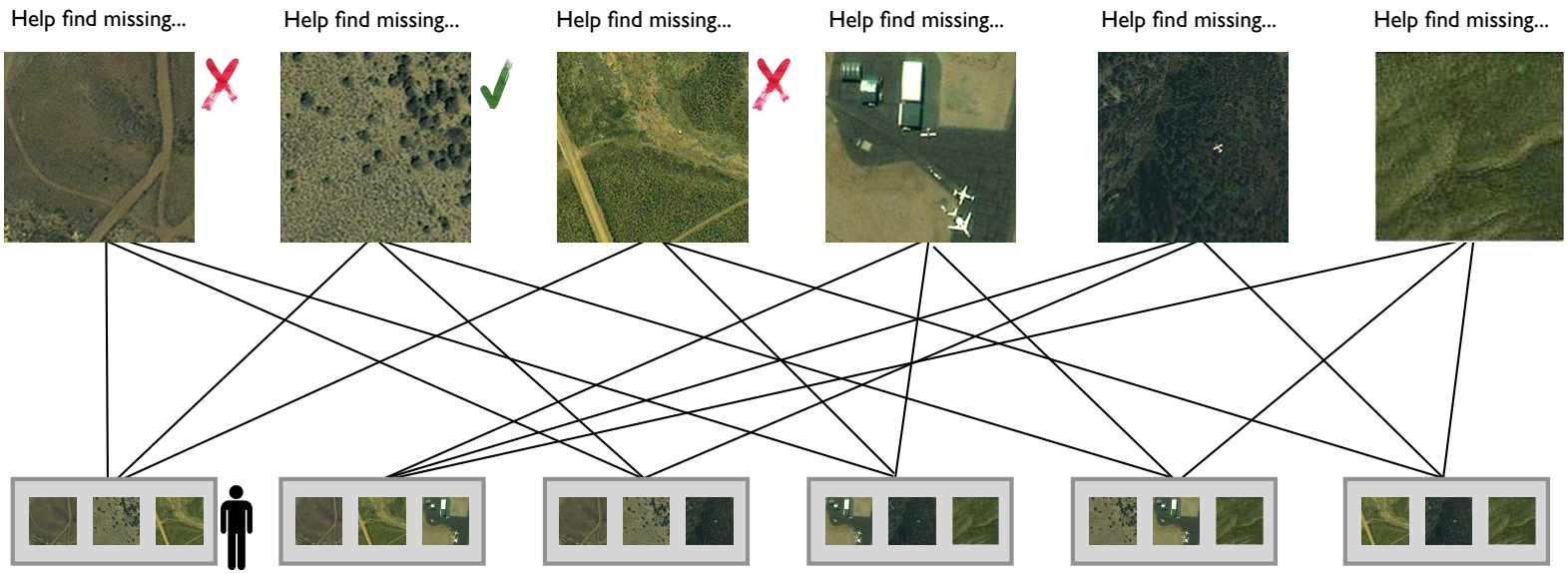
Help find missing...

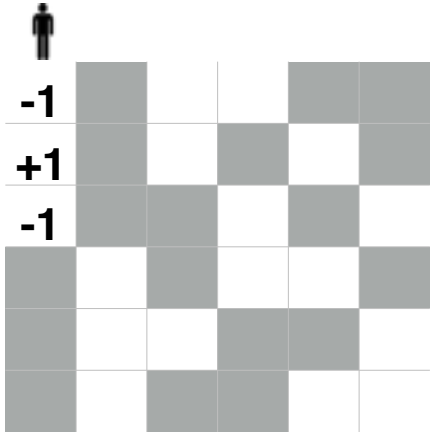
Add redundancy to cope with noise



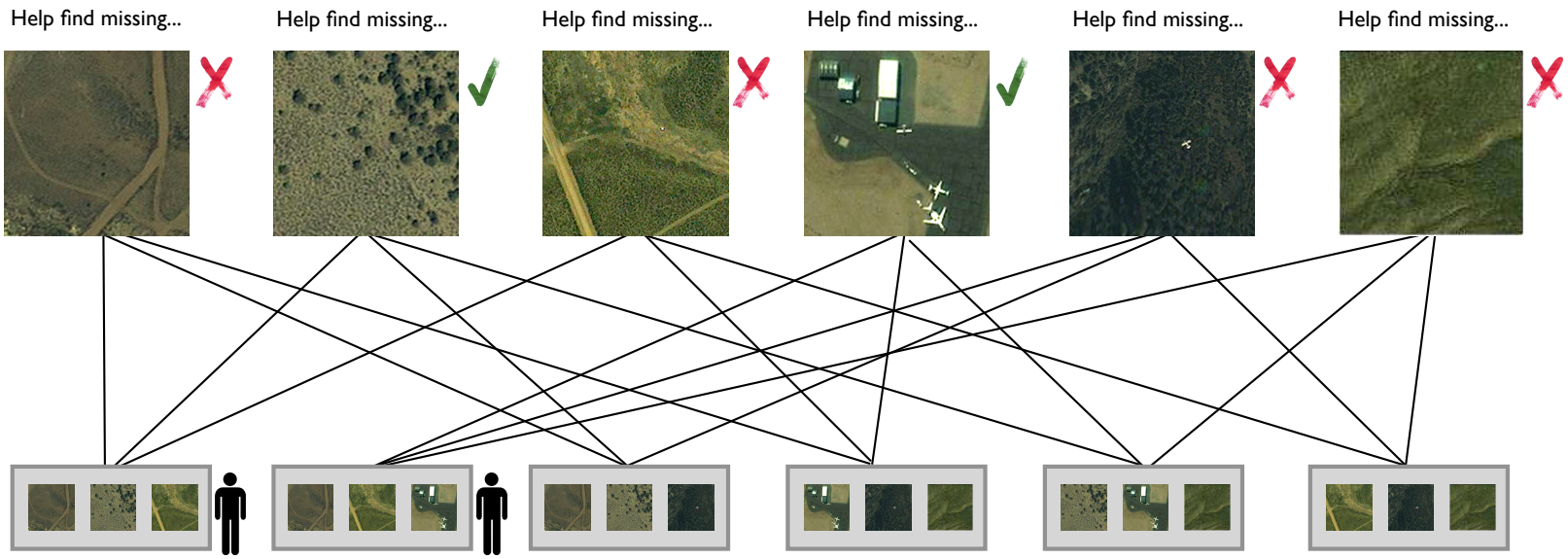


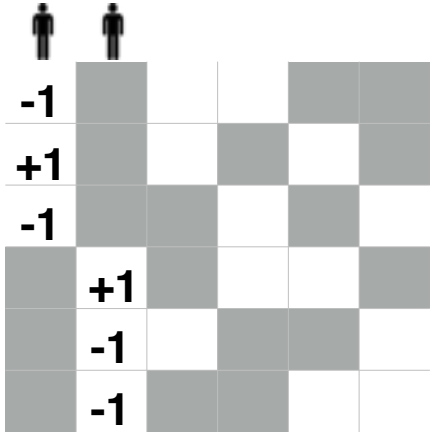
Add redundancy to cope with noise



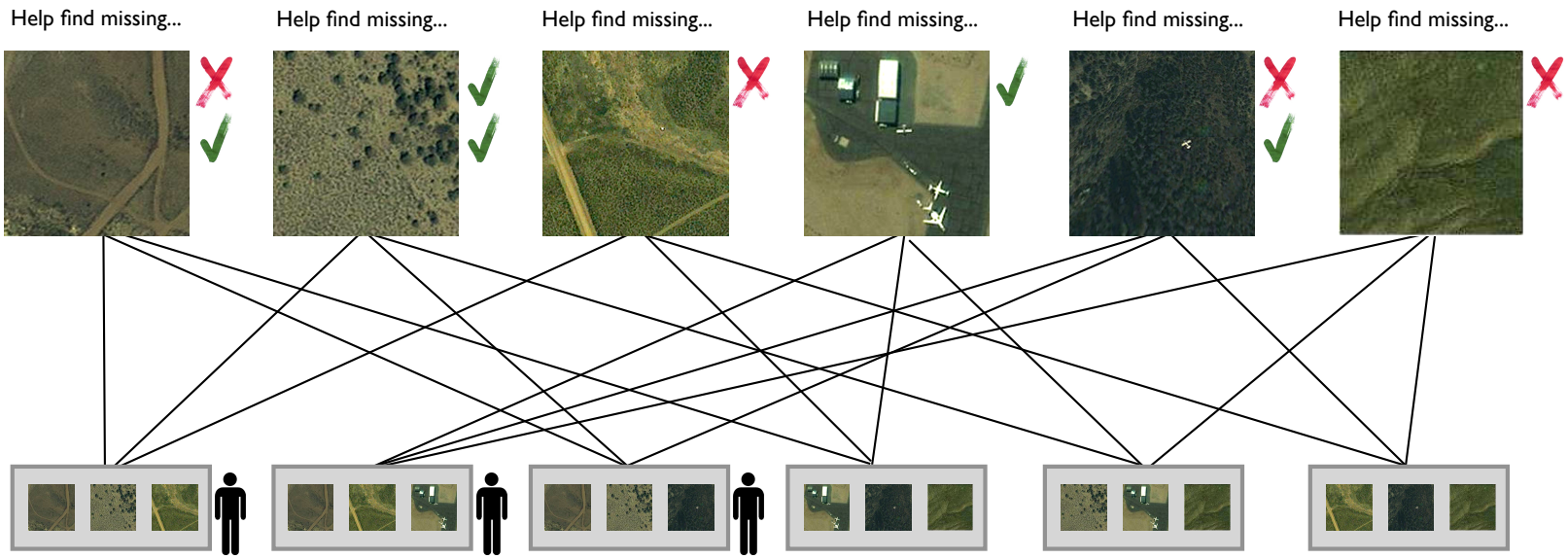


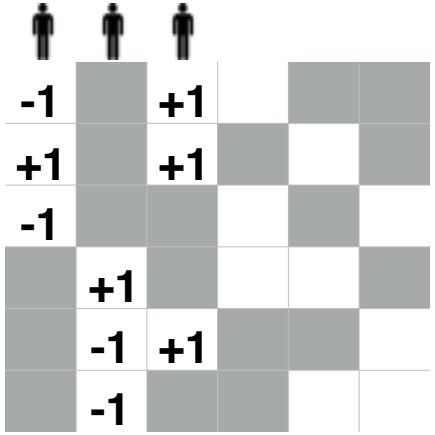
Add redundancy to cope with noise



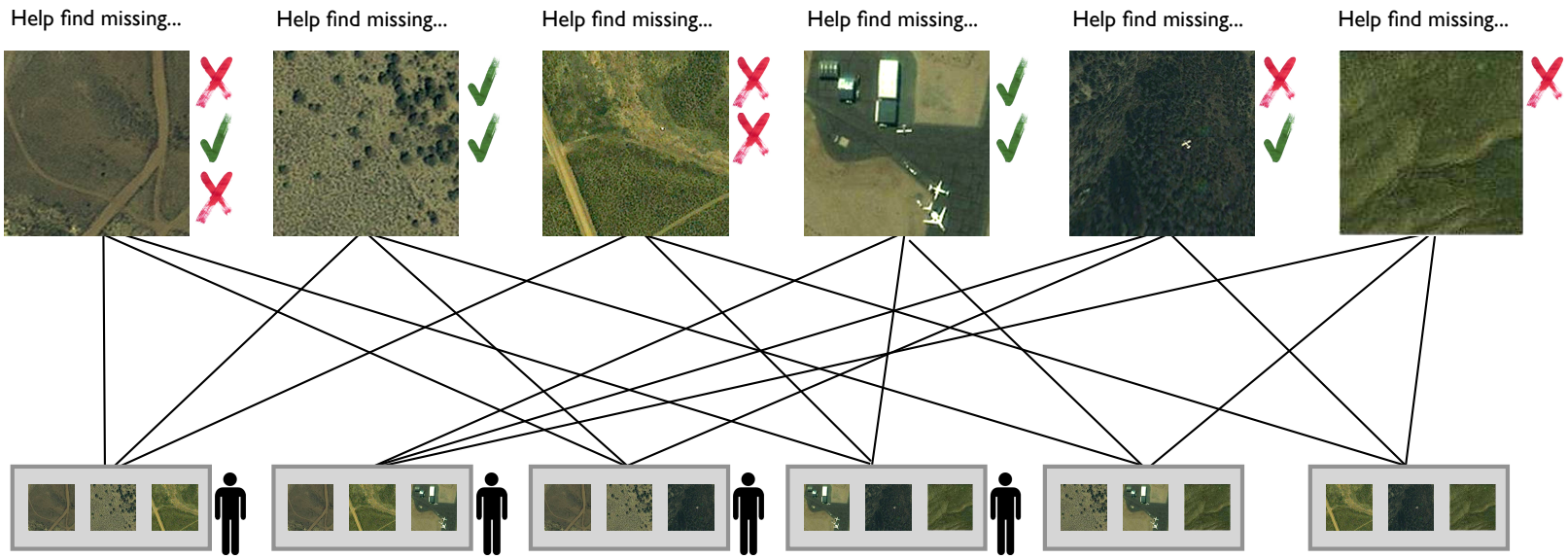


Add redundancy to cope with noise



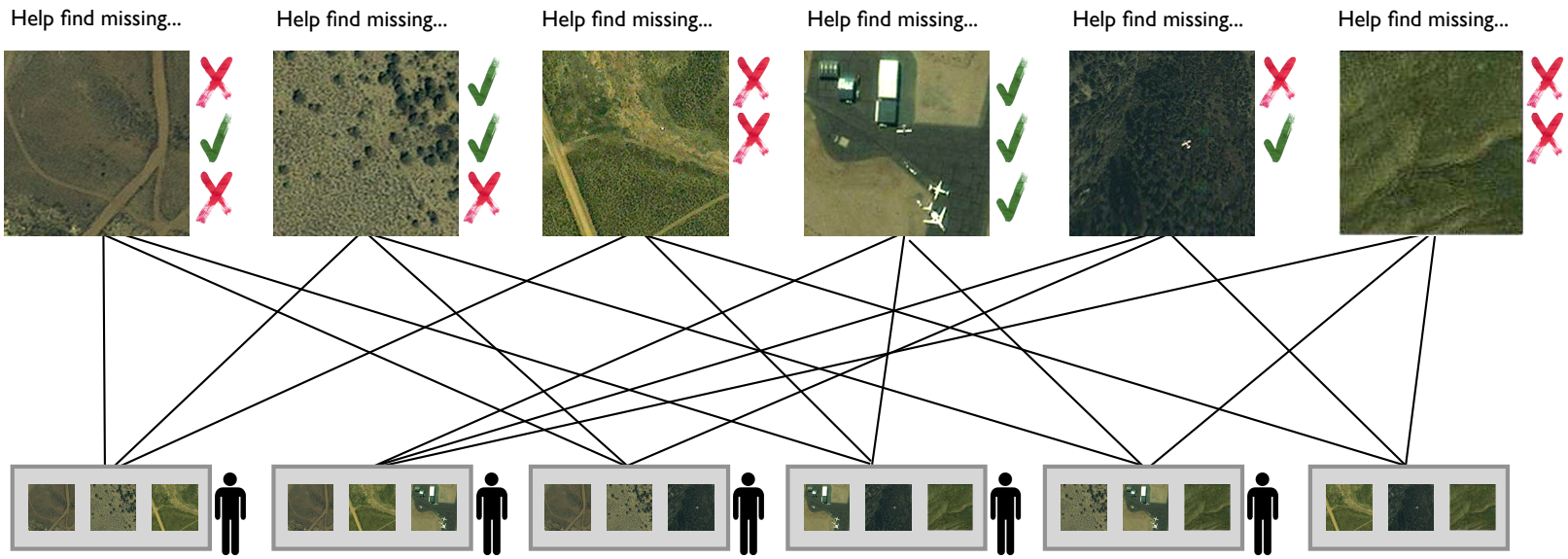


Add redundancy to cope with noise



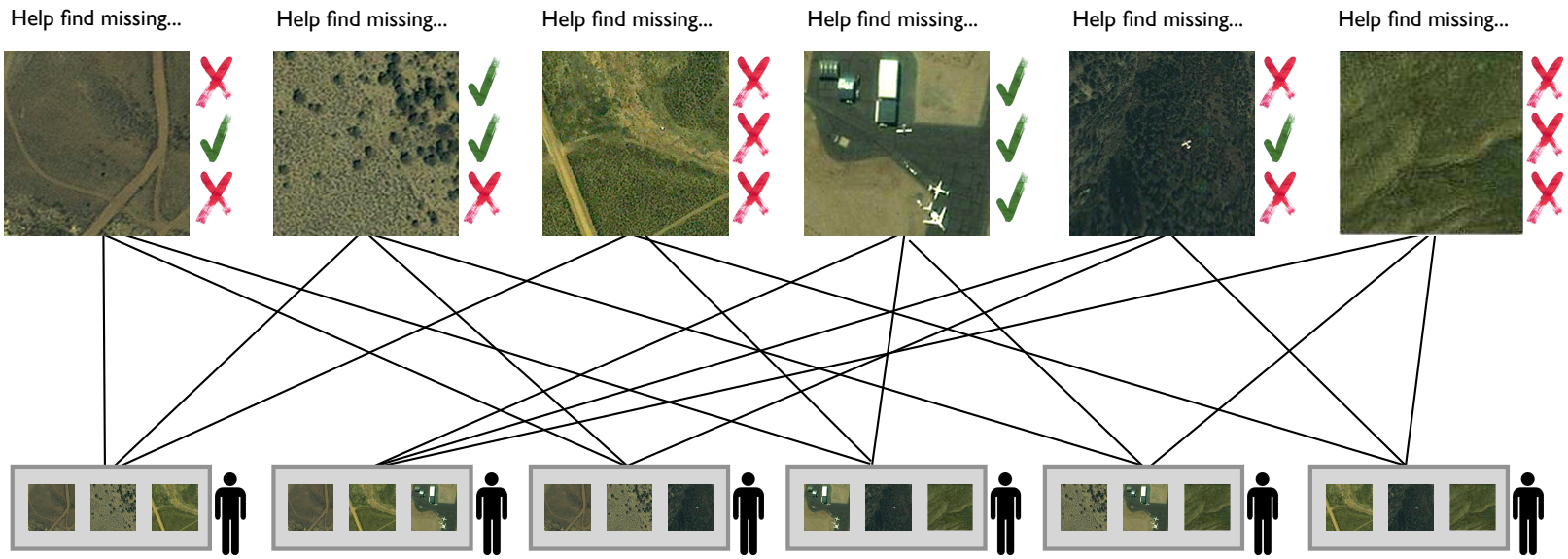
-1		+1	-1		
+1		+1			
-1			-1		
	+1		+1		
	-1	+1			
	-1				

Add redundancy to cope with noise



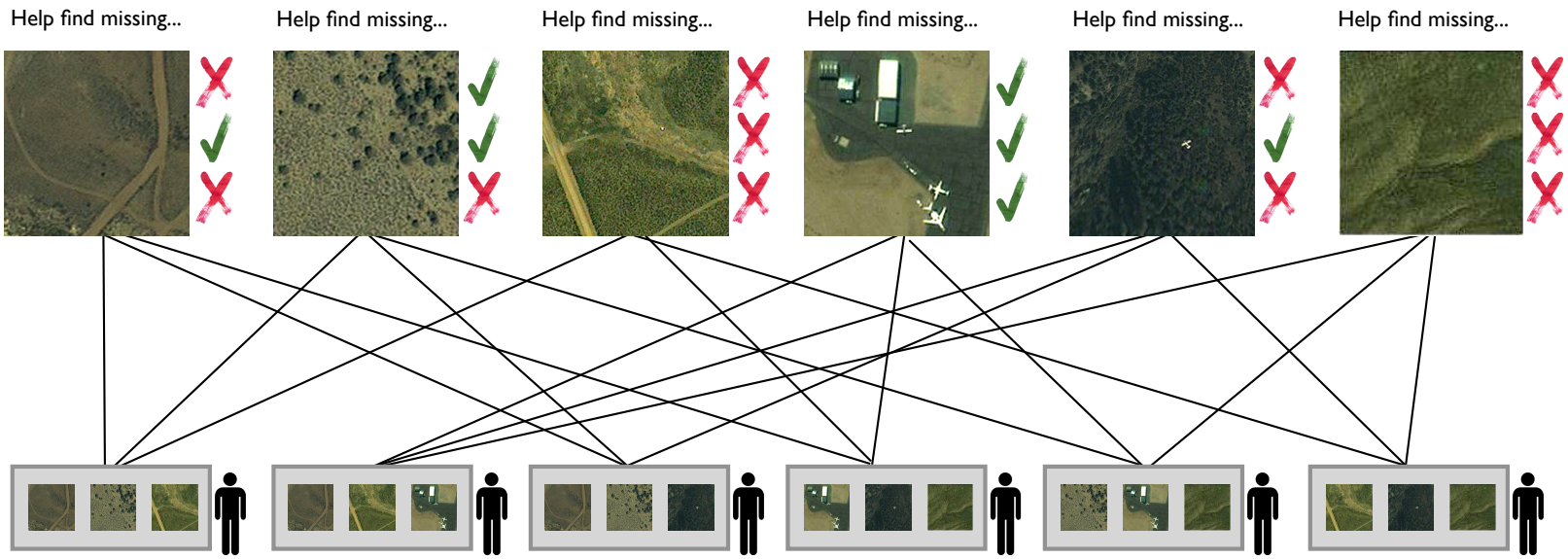
-1		+1	-1		
+1		+1		-1	
-1			-1		
	+1		+1	+1	
	-1	+1			
	-1			-1	

Add redundancy to cope with noise



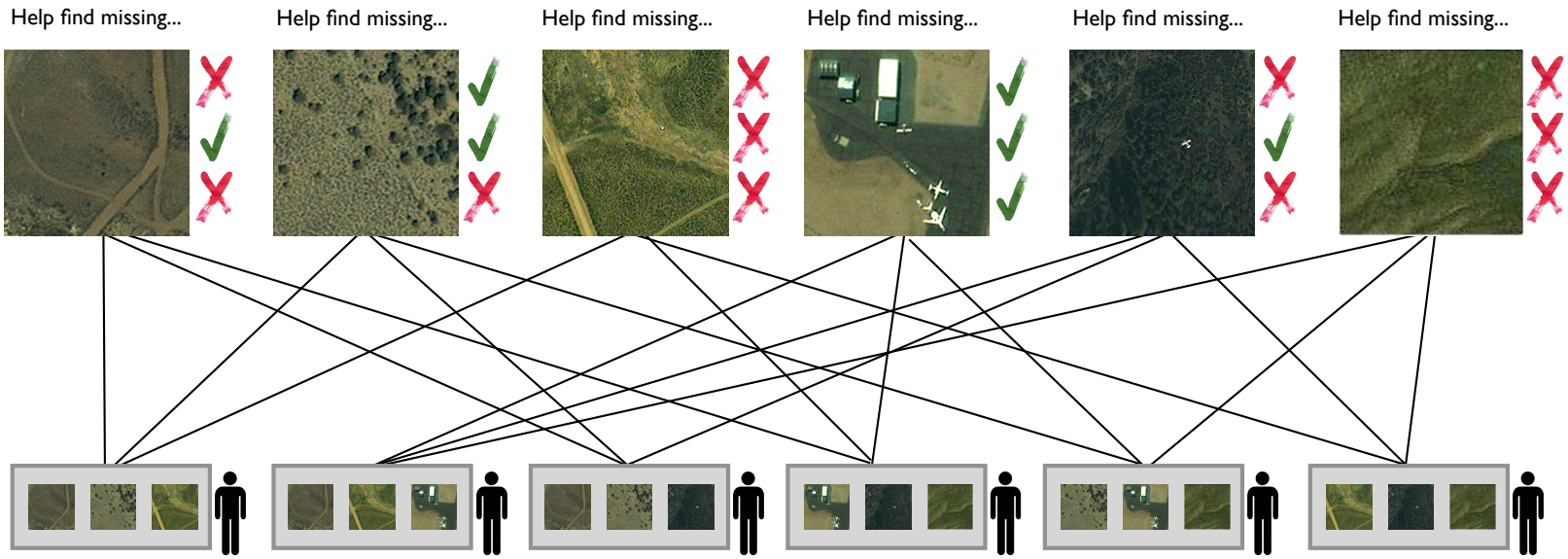
-1		+1	-1		
+1		+1		-1	
-1			-1		-1
	+1		+1	+1	
	-1	+1			-1
	-1			-1	-1

Add redundancy to cope with noise



						Majority Voting
-1		+1	-1			-1
+1		+1		-1		+1
-1			-1		-1	-1
	+1		+1	+1		+1
	-1	+1			-1	-1
	-1			-1	-1	-1

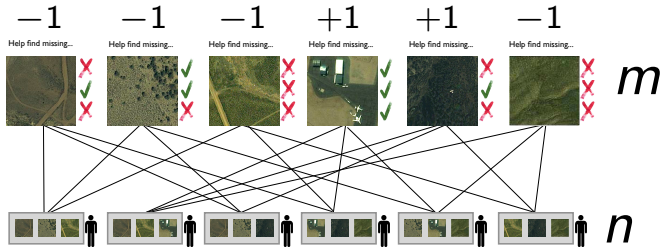
Add redundancy to cope with noise



Tradeoff: redundancy vs. accuracy

- special to crowdsourcing system:
  - we pay for each response
  - we design the graph
  - workers arrive online fashion

Notations



- m questions
- unknown true answers $t \in \{-1, +1\}^m$
- task-assignment graph $G([m], [n], E)$

  $(i, j) \in E$ indicates that question i is asked to the j-th arriving worker
- n workers arrive online fashion and submit responses
- response matrix $A \in \{-1, 0, 1\}^{m \times n}$

$$A_{ij} = \begin{cases} 0 & \text{if not assigned} \\ +1 & \text{if worker answers "}+1\text{"} \\ -1 & \text{if worker answers "}-1\text{"} \end{cases}$$

$$\begin{array}{cc} & n \text{ workers} \longrightarrow \\ m \text{ tasks} \downarrow & \begin{bmatrix} \quad A \quad \end{bmatrix} \end{array}$$

- Dawid-Skene model from 1979:
  - Each worker is parametrized by a scalar value $p_j \in [0, 1]$
  - For each assigned question, answers correctly with probability p_j

$$A_{ij} = \begin{cases} t_i & \text{with probability } p_j \\ -t_i & \text{with probability } 1 - p_j \end{cases}$$

  - Criticism: all tasks are assumed to be equally difficult

Design:

1. Task assignment graph E
2. Inference algorithm $\hat{t}(A) \in \{-1, +1\}^m$

How does error rate trade-off with cost/redundancy?

$$\mu \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1) \qquad \sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$$

	p_1	p_2	p_3	p_4	p_5	p_6
t_1	-1		+1	-1		
t_2	+1		+1		-1	
t_3	-1			-1		-1
t_3		+1		+1	+1	
t_4		-1	+1			-1
t_5		-1			-1	-1

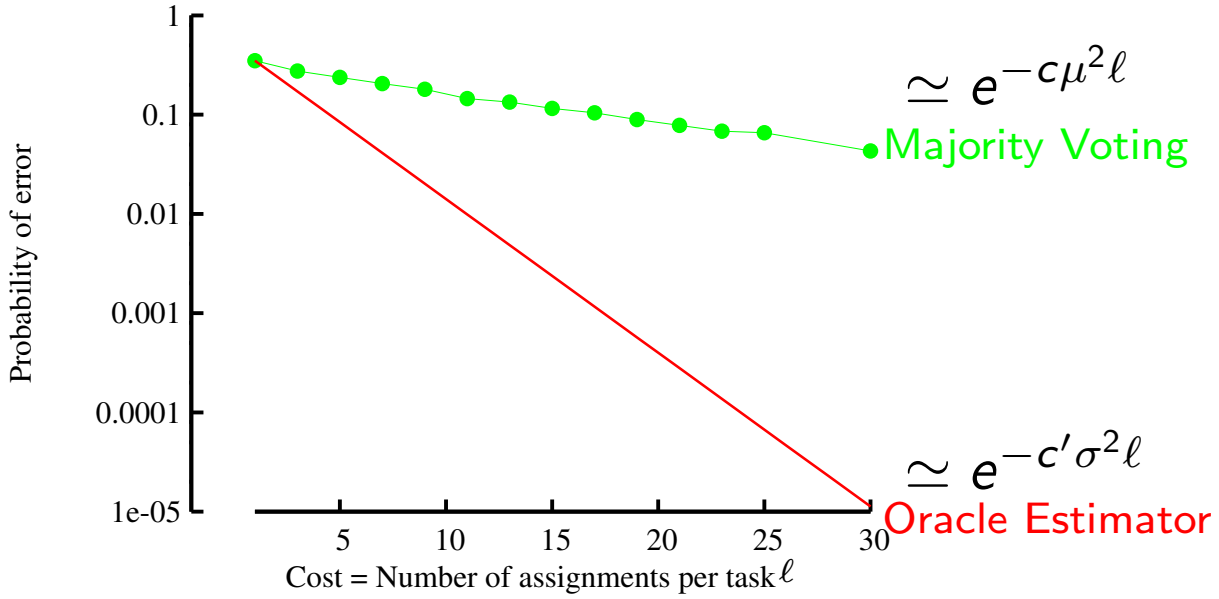
Design:

1. Task assignment graph E
2. Inference algorithm $\hat{t}(A) \in \{-1, +1\}^m$

How does error rate trade-off with cost/redundancy?

$$\mu \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1) \qquad \sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$$

	p_1	p_2	p_3	p_4	p_5	p_6
t_1	-1		+1	-1		
t_2	+1		+1		-1	
t_3	-1			-1		-1
t_3		+1		+1	+1	
t_4		-1	+1			-1
t_5		-1			-1	-1



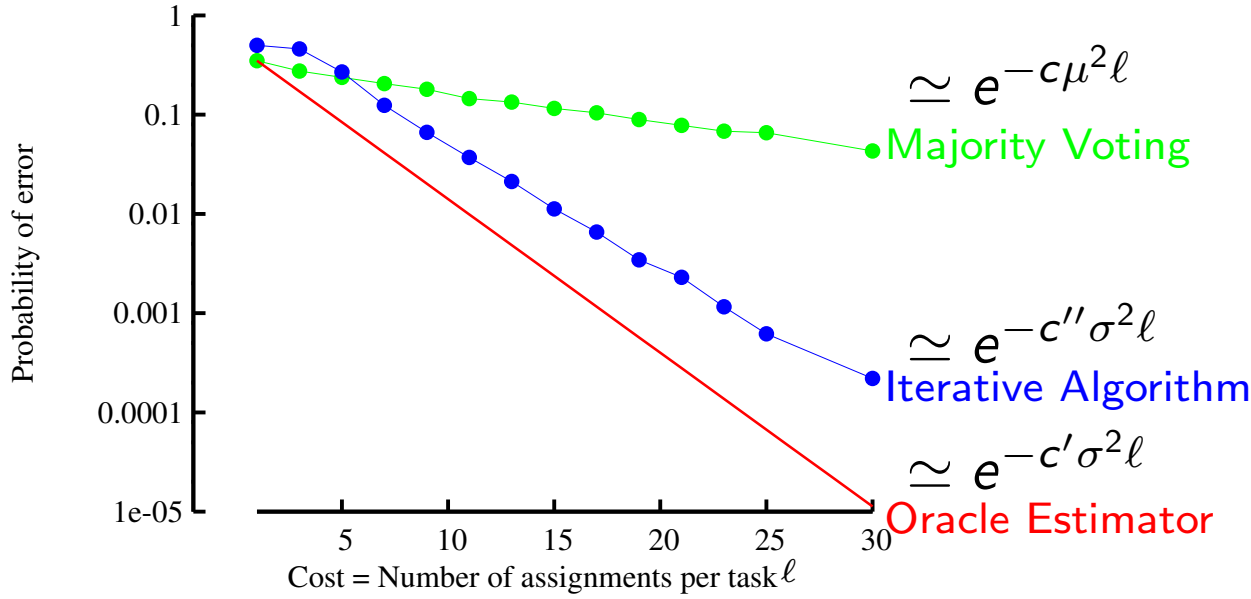
Design:

1. Task assignment graph E
2. Inference algorithm $\hat{t}(A) \in \{-1, +1\}^m$

How does error rate trade-off with cost/redundancy?

$$\mu \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1) \qquad \sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$$

	p_1	p_2	p_3	p_4	p_5	p_6
t_1	-1		+1	-1		
t_2	+1		+1		-1	
t_3	-1			-1		-1
t_3		+1		+1	+1	
t_4		-1	+1			-1
t_5		-1			-1	-1



What is known for DS model

quality of the crowd $\sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$

Achievability: random graph E and iterative inference achieves

$$P_{\mathrm{err}} \leq e^{-c\sigma^2\ell}$$

Fundamental limit: the best task-assignment and inference is limited by

$$\min_{E,\hat{t}} \max_{t\in\{\pm 1\}^m, p\in\mathcal{F}_{\sigma^2}} P_{\mathrm{err}} \geq e^{-c'\sigma^2\ell}$$

What is known for DS model

quality of the crowd $\sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$

Achievability: random graph E and iterative inference achieves

$$P_{\mathrm{err}} \leq e^{-c\sigma^2\ell}$$

Fundamental limit: the best task-assignment and inference is limited by

$$\min_{E,\hat{t}} \max_{t\in\{\pm1\}^m, p\in\mathcal{F}_{\sigma^2}} P_{\mathrm{err}} \geq e^{-c'\sigma^2\ell}$$

Fundamental limit: the best adaptive is limited by

$$\min_{\text{all adaptive schemes},\hat{t}} \max_{t\in\{\pm1\}^m, p\in\mathcal{F}_{\sigma^2}} P_{\mathrm{err}} \geq e^{-c''\sigma^2\ell}$$

Adaptive schemes are common in practice

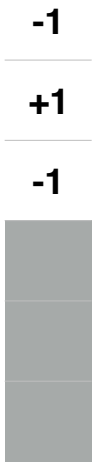
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal

Adaptive schemes are common in practice

- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal

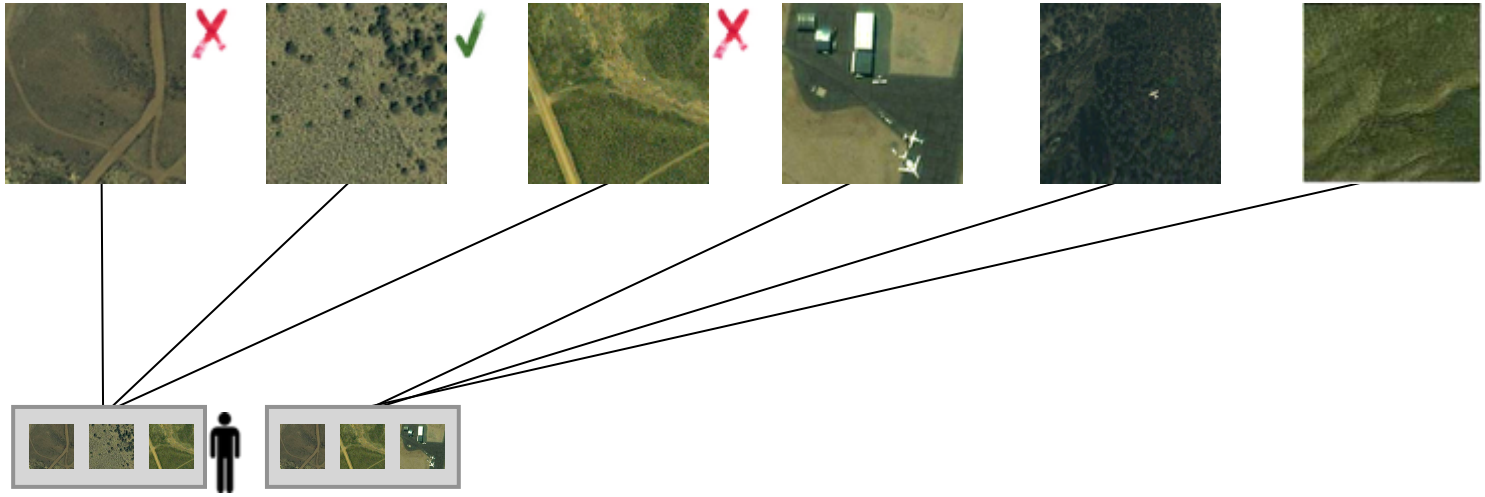
Adaptive schemes are common in practice

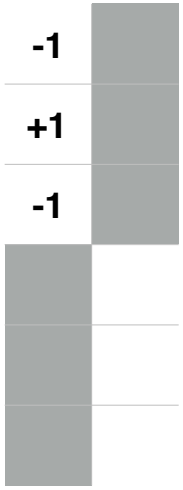
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



Adaptive schemes are common in practice

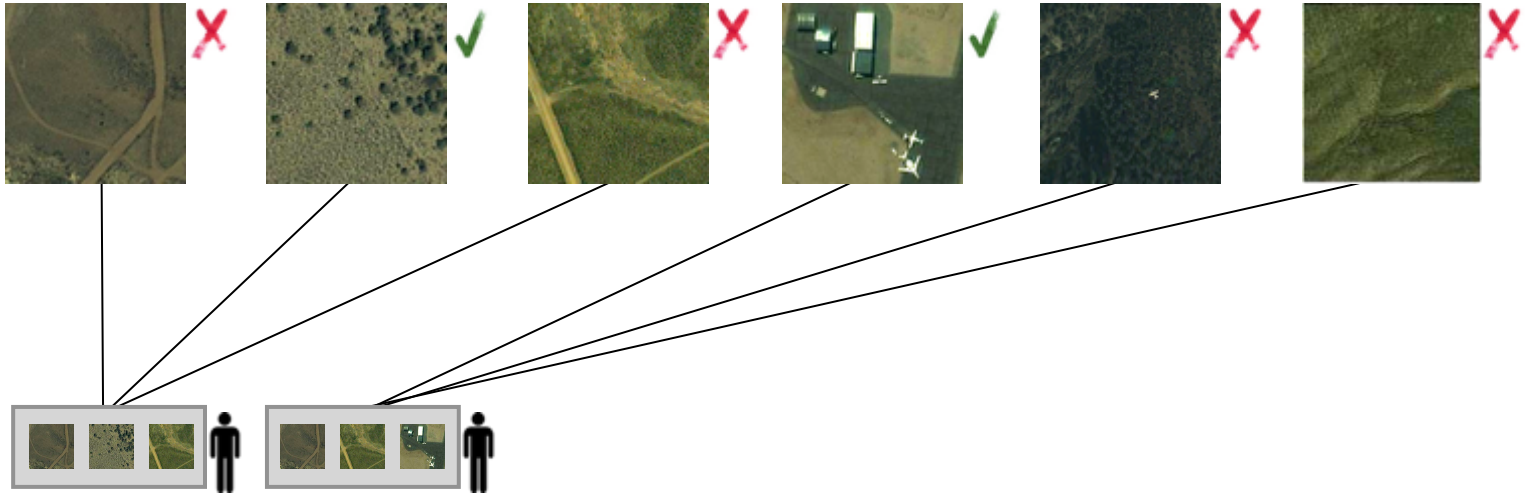
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal





Adaptive schemes are common in practice

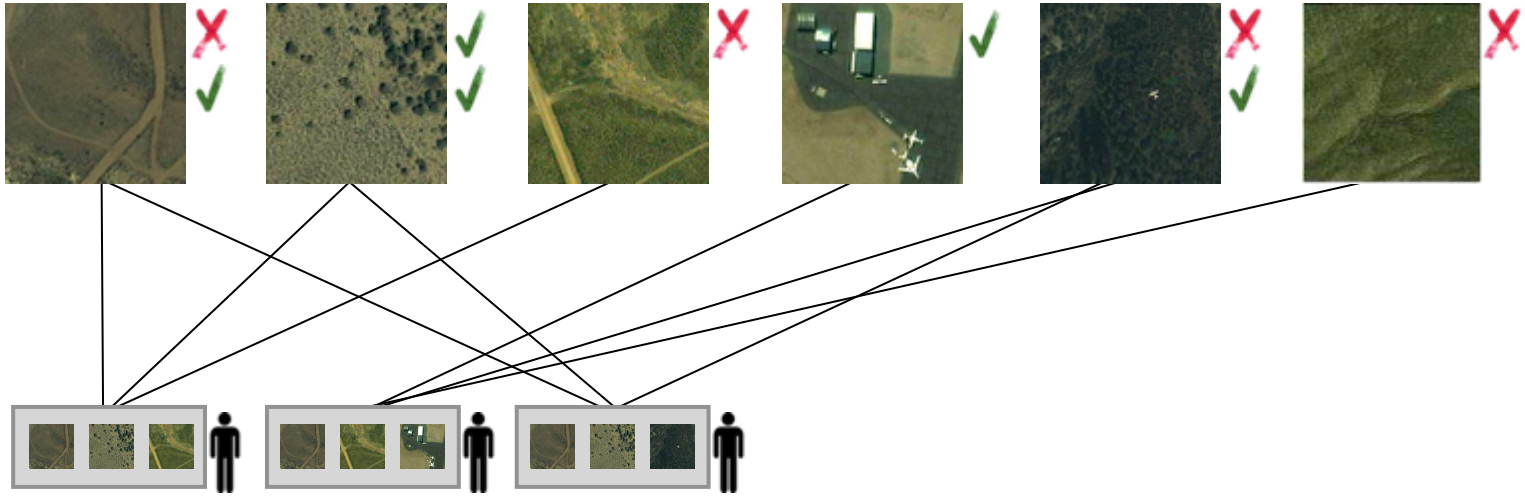
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



-1	
+1	
-1	
	+1
	-1
	-1

Adaptive schemes are common in practice

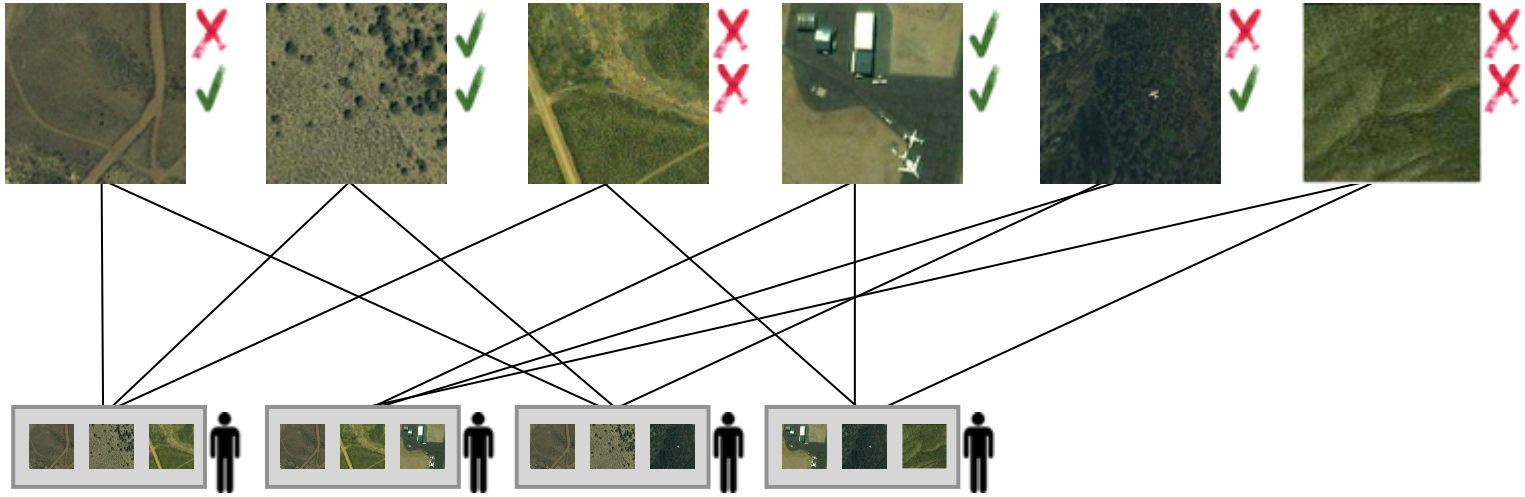
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



-1		+1
+1		+1
-1		
	+1	
	-1	+1
	-1	

Adaptive schemes are common in practice

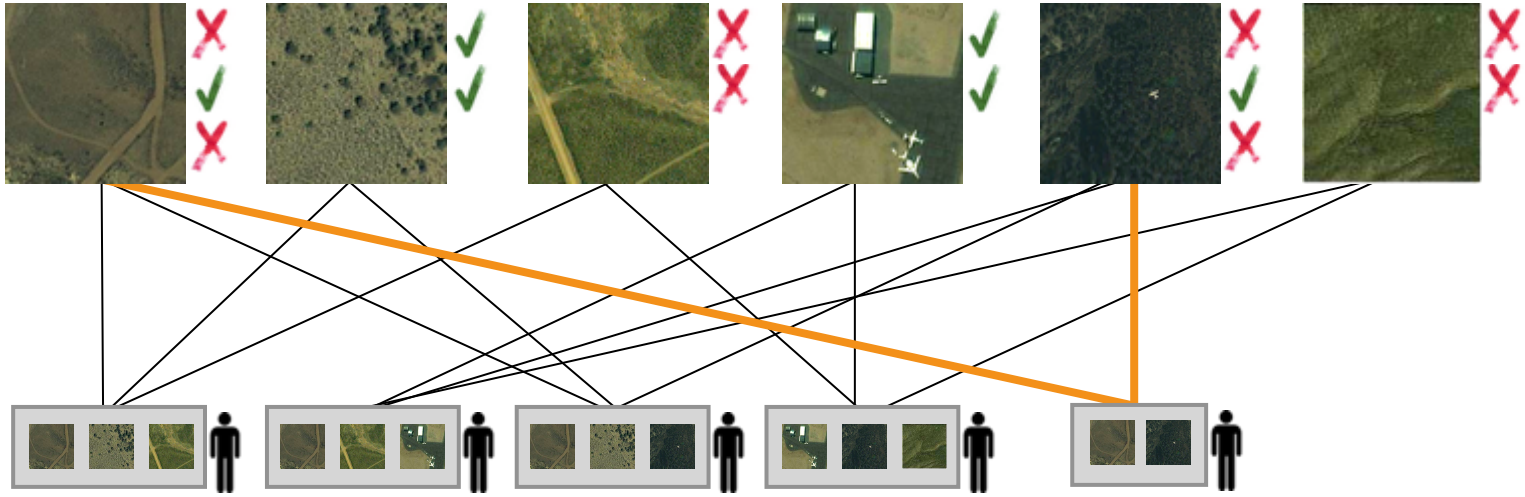
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



-1		+1	
+1		+1	
-1			-1
	+1		+1
	-1	+1	
	-1		-1

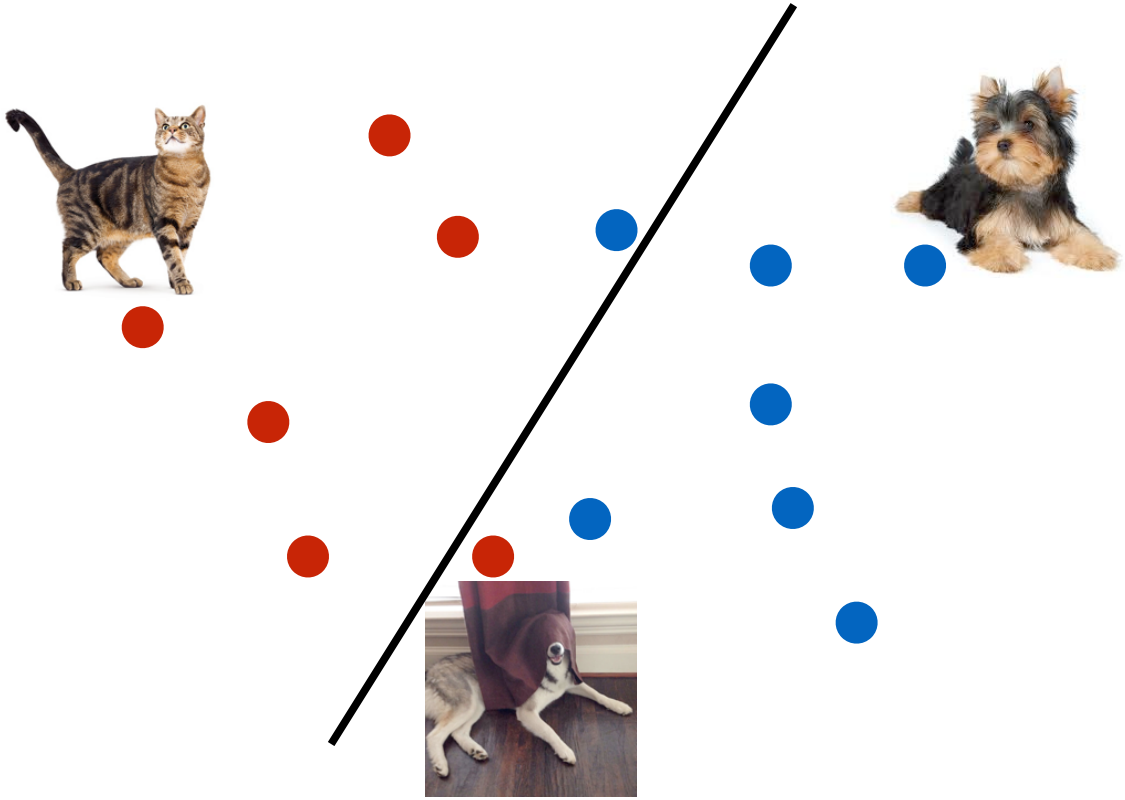
Adaptive schemes are common in practice

- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



-1		+1		-1
+1		+1		
-1			-1	
	+1		+1	
	-1	+1		-1
	-1		-1	

When does adaptivity help?

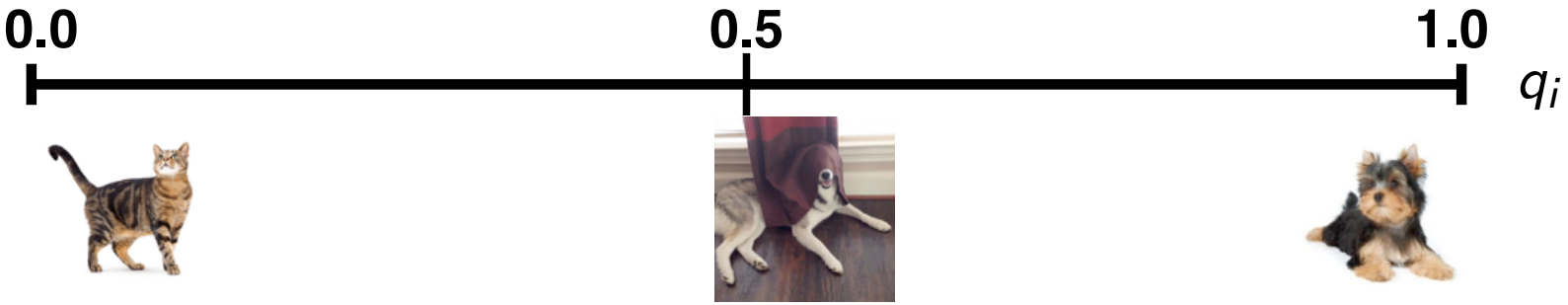


Generalized DS model [Zhou et al. '15]

- Definition
  - Each task is parametrized by a scalar value $q_i \in [0, 1]$
  - When a task is presented to a worker, it is perceived as a positive task with probability q_i

$$A_{ij} = \begin{cases} +1 & \text{with probability } q_i p_j + (1 - q_i)(1 - p_j) \\ -1 & \text{with probability } q_i(1 - p_j) + (1 - q_i)p_j \end{cases}$$

  - ground truth is $t_i = \mathbb{I}(q_i > 0.5) - \mathbb{I}(q_i < 0.5)$
  - difficulty level of a task i measured by $(2q_i - 1)^2$

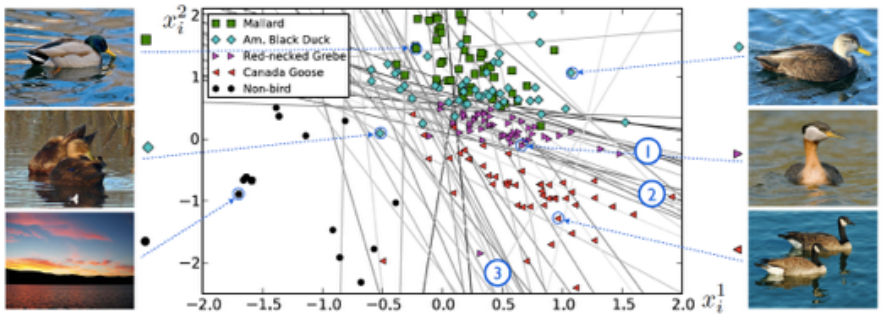


Generalized DS model [Zhou et al. '15]

- Definition
  - Each task is parametrized by a scalar value $q_i \in [0, 1]$
  - When a task is presented to a worker, it is perceived as a positive task with probability q_i

$$A_{ij} = \begin{cases} +1 & \text{with probability } q_i p_j + (1 - q_i)(1 - p_j) \\ -1 & \text{with probability } q_i(1 - p_j) + (1 - q_i)p_j \end{cases}$$

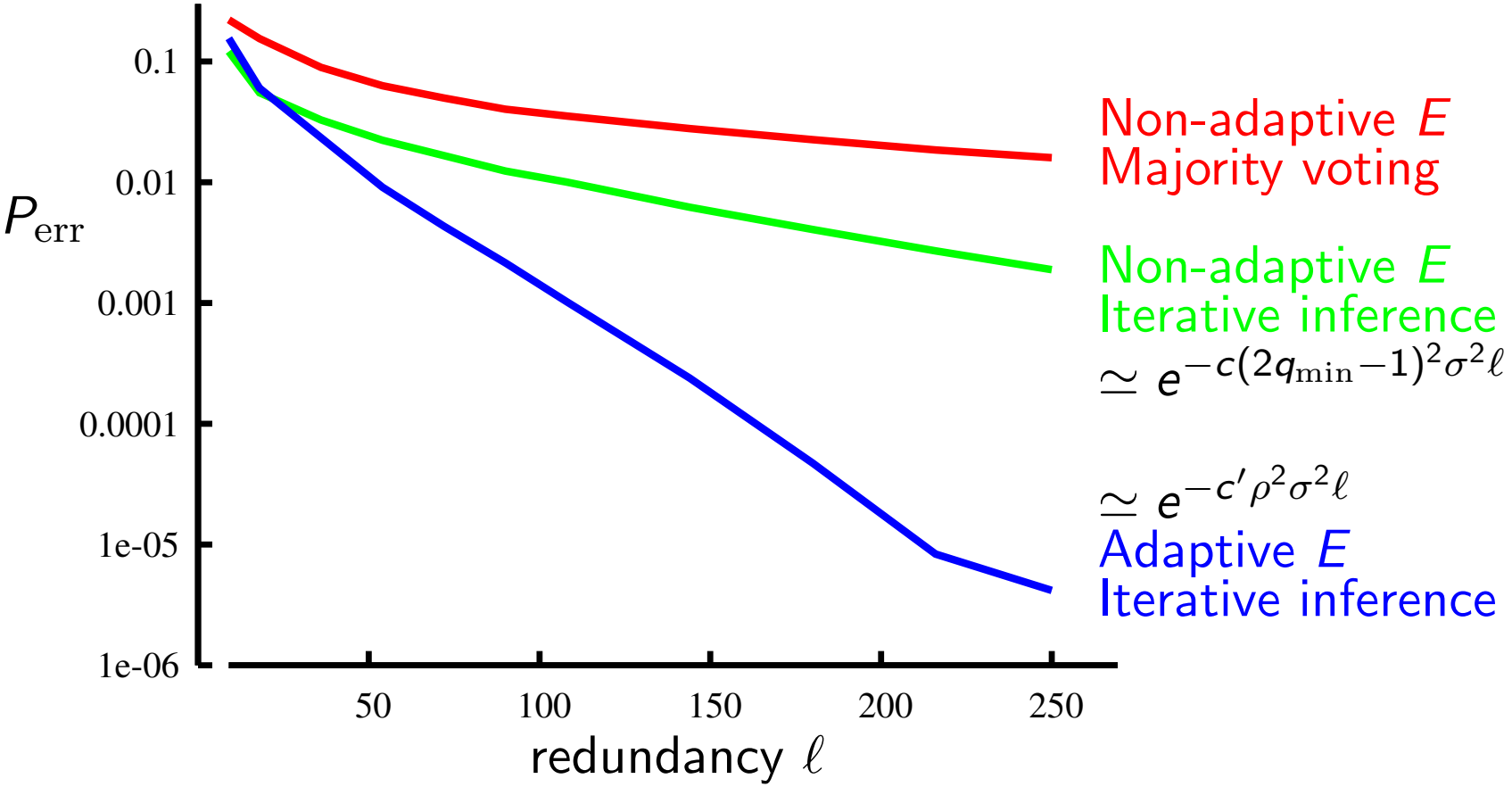
  - ground truth is $t_i = \mathbb{I}(q_i > 0.5) - \mathbb{I}(q_i < 0.5)$
  - difficulty level of a task i measured by $(2q_i - 1)^2$



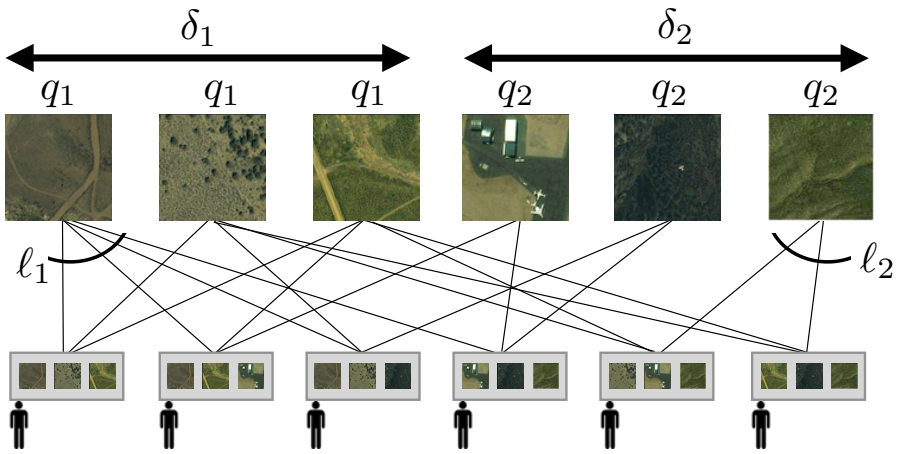
Generalized DS model with adaptive scheme

Collective task difficulty:

$$\rho^2 = \frac{1}{\frac{1}{m}\sum_{i=1}^{m}\frac{1}{(2q_i-1)^2}}$$



Minimax rate for adaptive scenario



- task difficulty $\rho^2 = \frac{1}{\frac{1}{m}\sum_{i=1}^{m}\frac{1}{(2q_i-1)^2}}$

Fundamental limit: the best adaptive scheme and inference is limited by

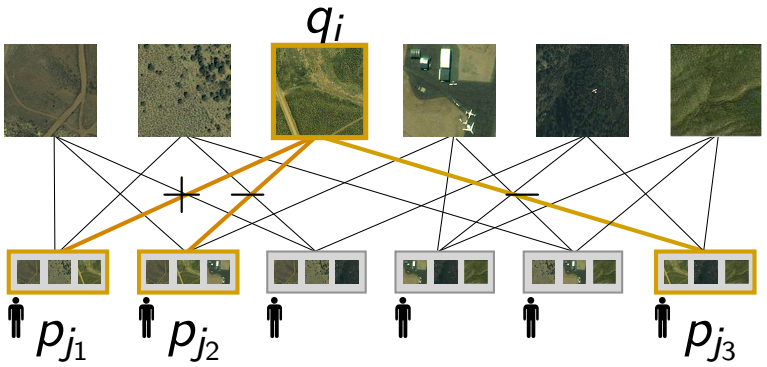
$$\min_{\text{adaptive}E,\hat{t}} \max_{q\in\mathcal{G}_{\rho^2},p\in\mathcal{F}_{\sigma^2}} P_{\text{err}} \geq e^{-C'\rho^2\sigma^2\ell}$$

Achievability: efficient adaptive scheme and inference achieves

$$P_{\text{err}} \leq e^{-C\rho^2\sigma^2\ell}$$

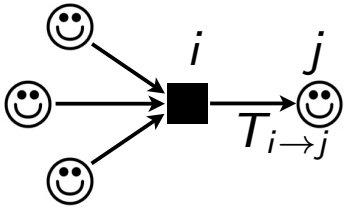
Fundamental limit: the best non-adaptive scheme is limited by

$$\min_{E,\hat{t}} \max_{q\in\mathcal{G}_{\rho^2},p\in\mathcal{F}_{\sigma^2}} P_{\text{err}} \geq e^{-C''(2q_{\min}-1)^2\sigma^2\ell}$$



- Message-passing algorithm
  - Two sets of messages:
    - ★ Task messages $\{T_{i\to j}\}$, and worker messages $\{W_{j\to i}\}$
  - Initialize worker messages as random Gaussian: $W_{j\to i} \sim \mathcal{N}(1,1)$
  - Iteratively update messages

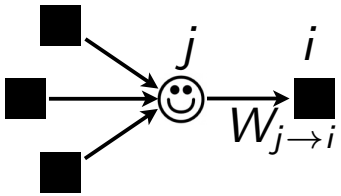
Task-likelihood update



$$\underbrace{T_{i\to j}}_{\textit{likelihood}} = \sum_{j'\neq j} \underbrace{W_{j'\to i}}_{\textit{reliability}} A_{ij'}$$

A task is likely to be '+' if reliable workers agree that it is '+'

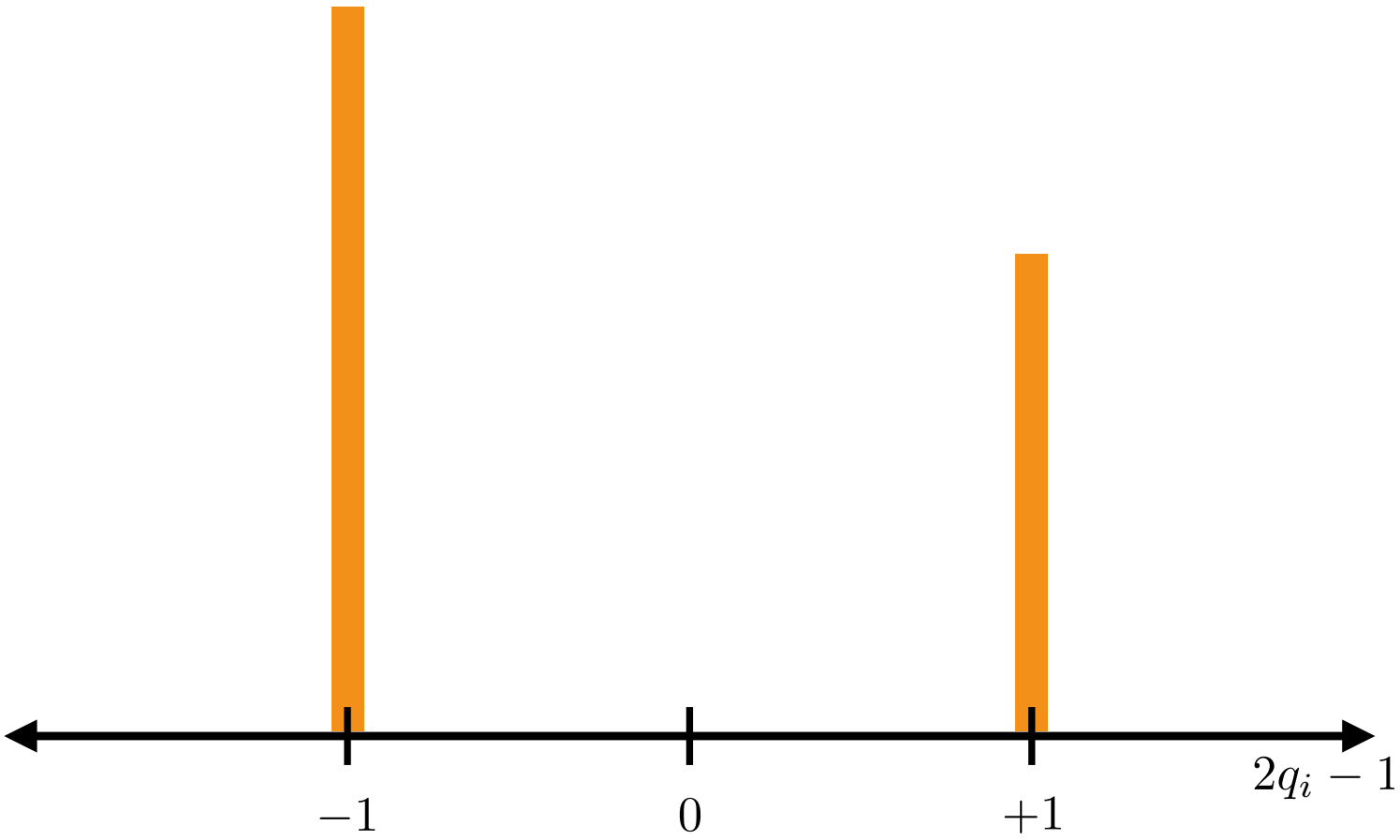
Worker-reliability update



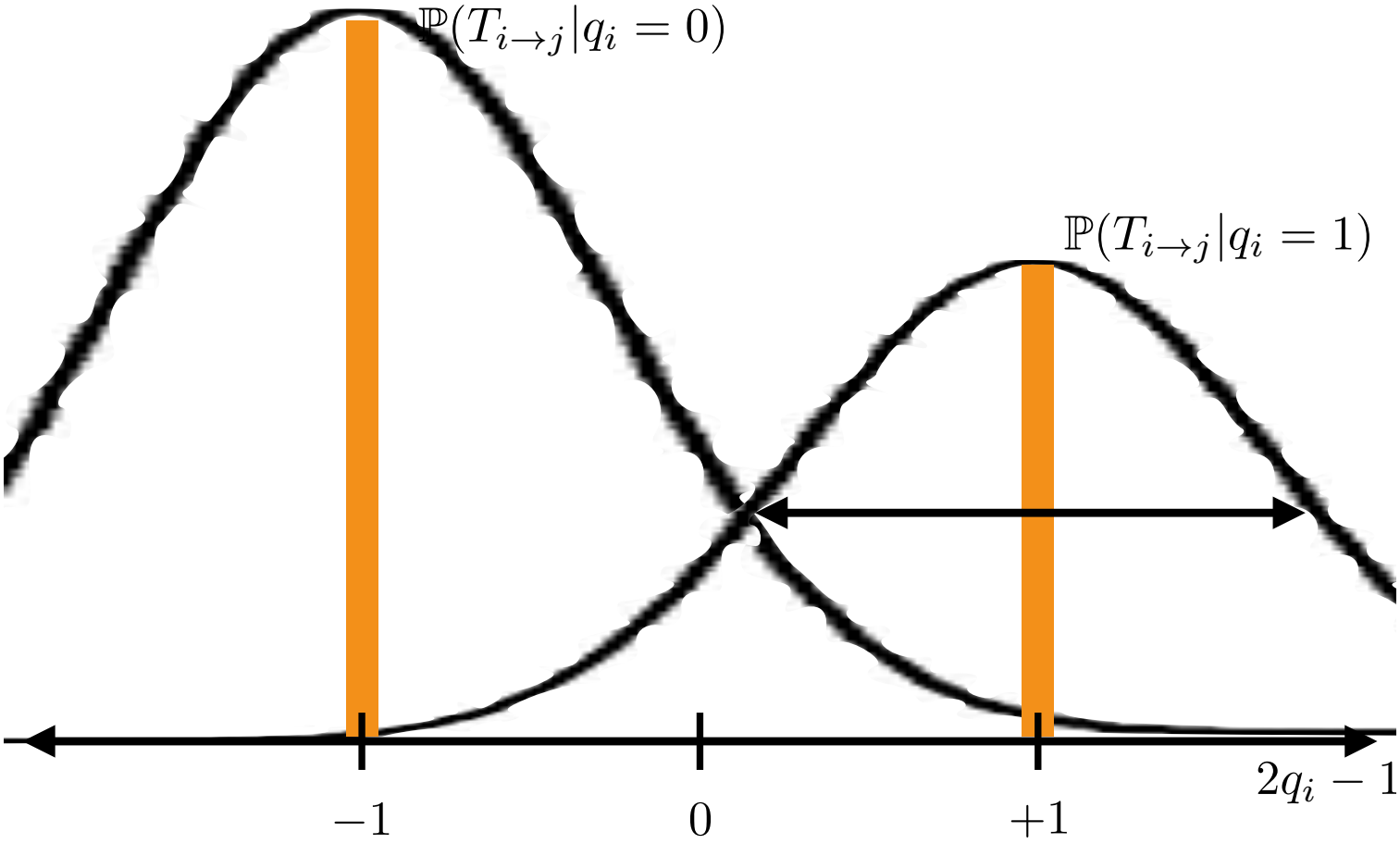
$$\underbrace{W_{j\to i}}_{\textit{reliability}} = \sum_{i'\neq i} \underbrace{T_{i'\to j}}_{\textit{likelihood}} A_{i'j}$$

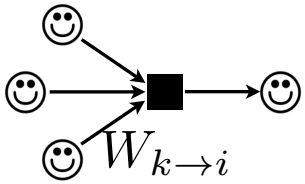
A worker is reliable if the worker agreed with our belief on other tasks

1. DS model + non-adaptive scheme



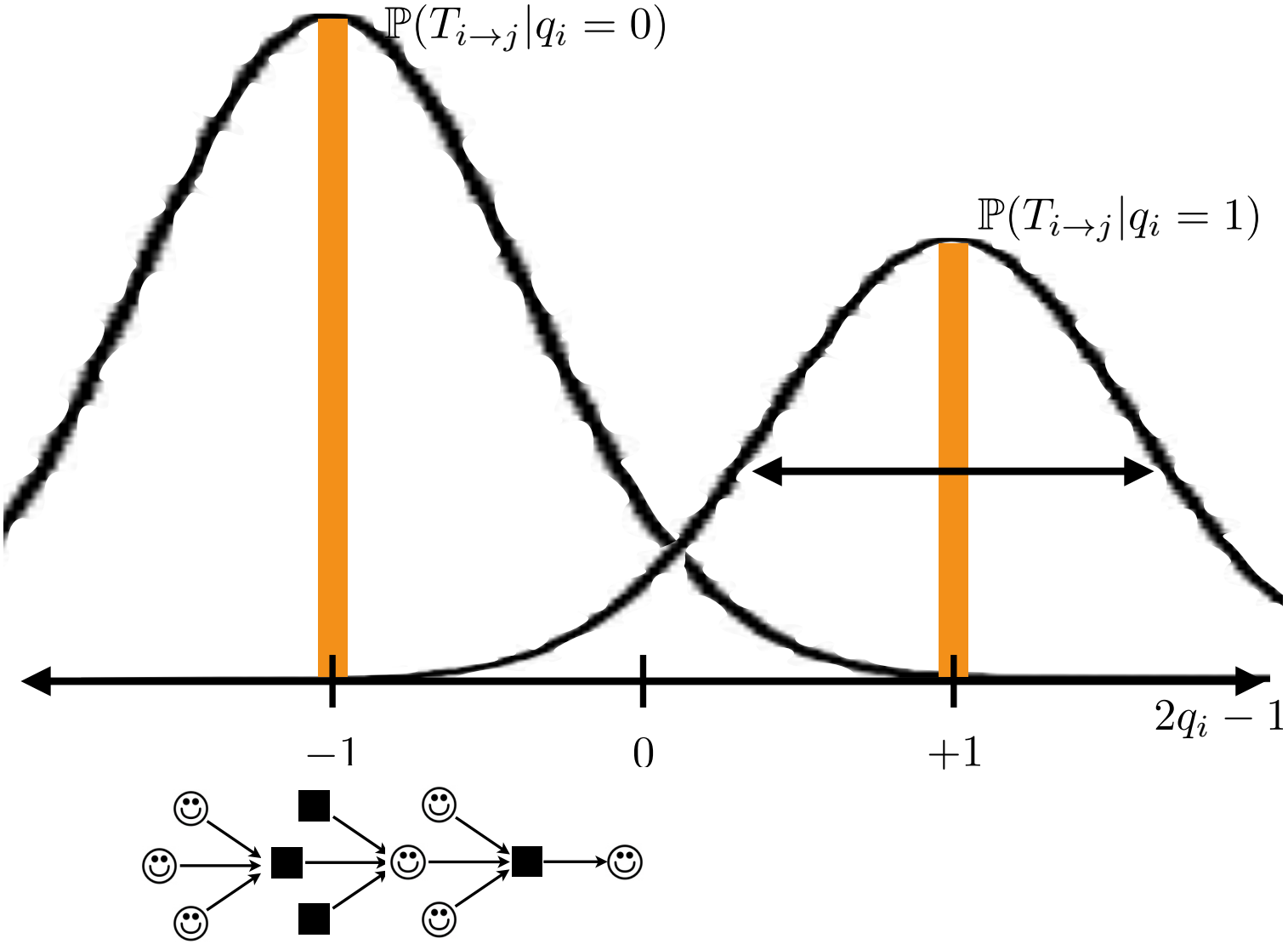
1. DS model + non-adaptive scheme



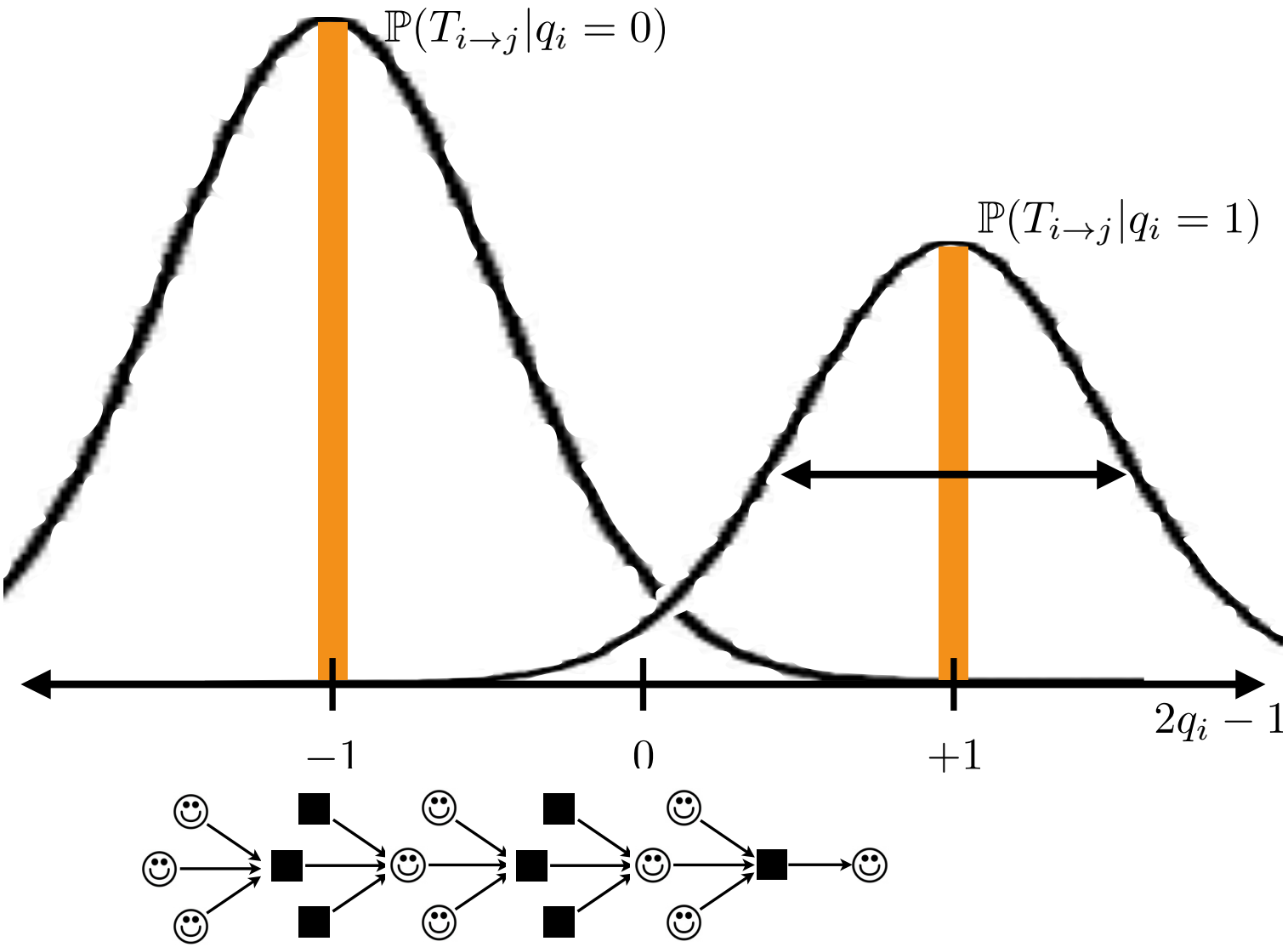


$$T_{i\to j} = \sum A_{ik} W_{k\to i}$$

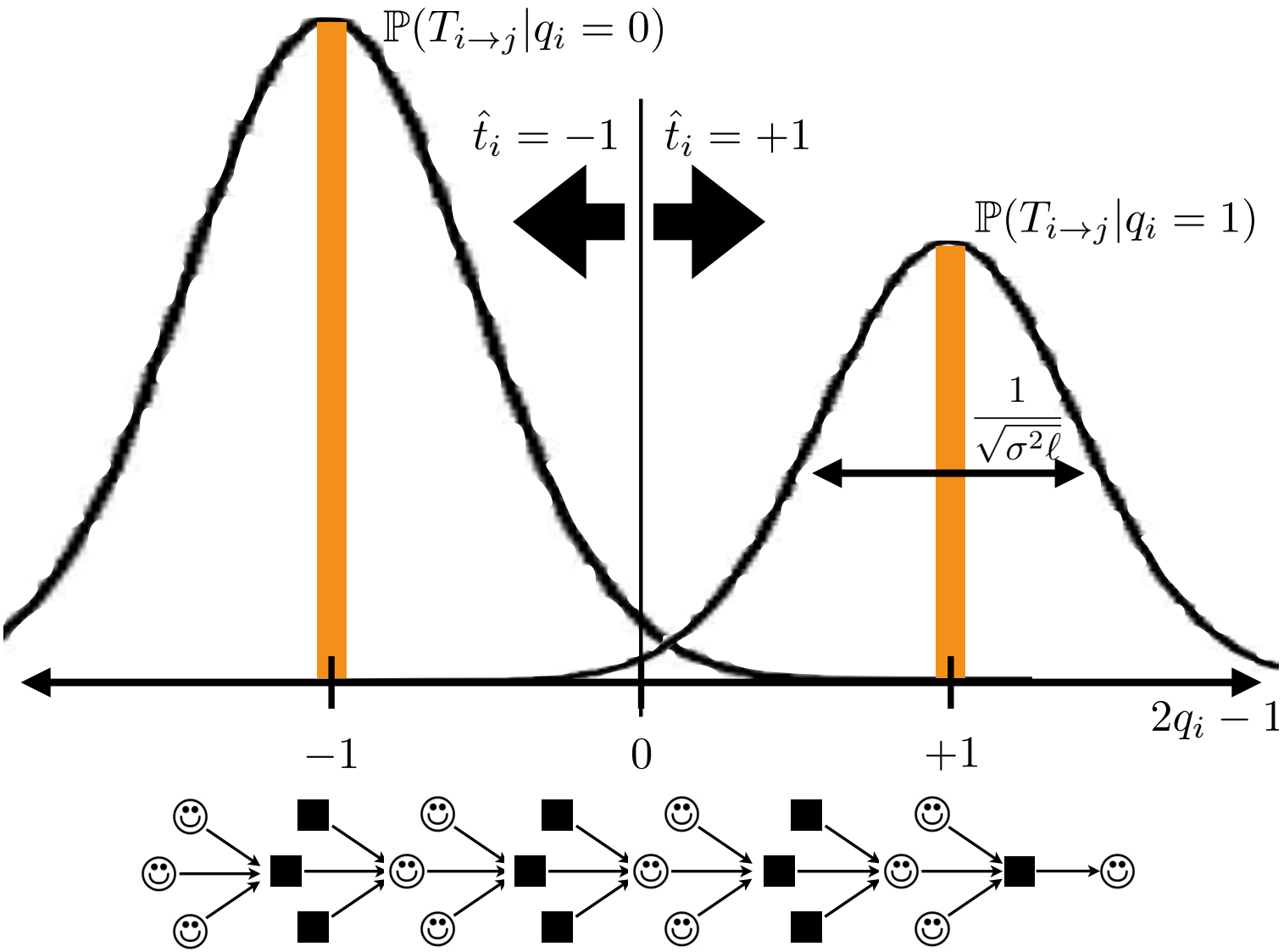
1. DS model + non-adaptive scheme



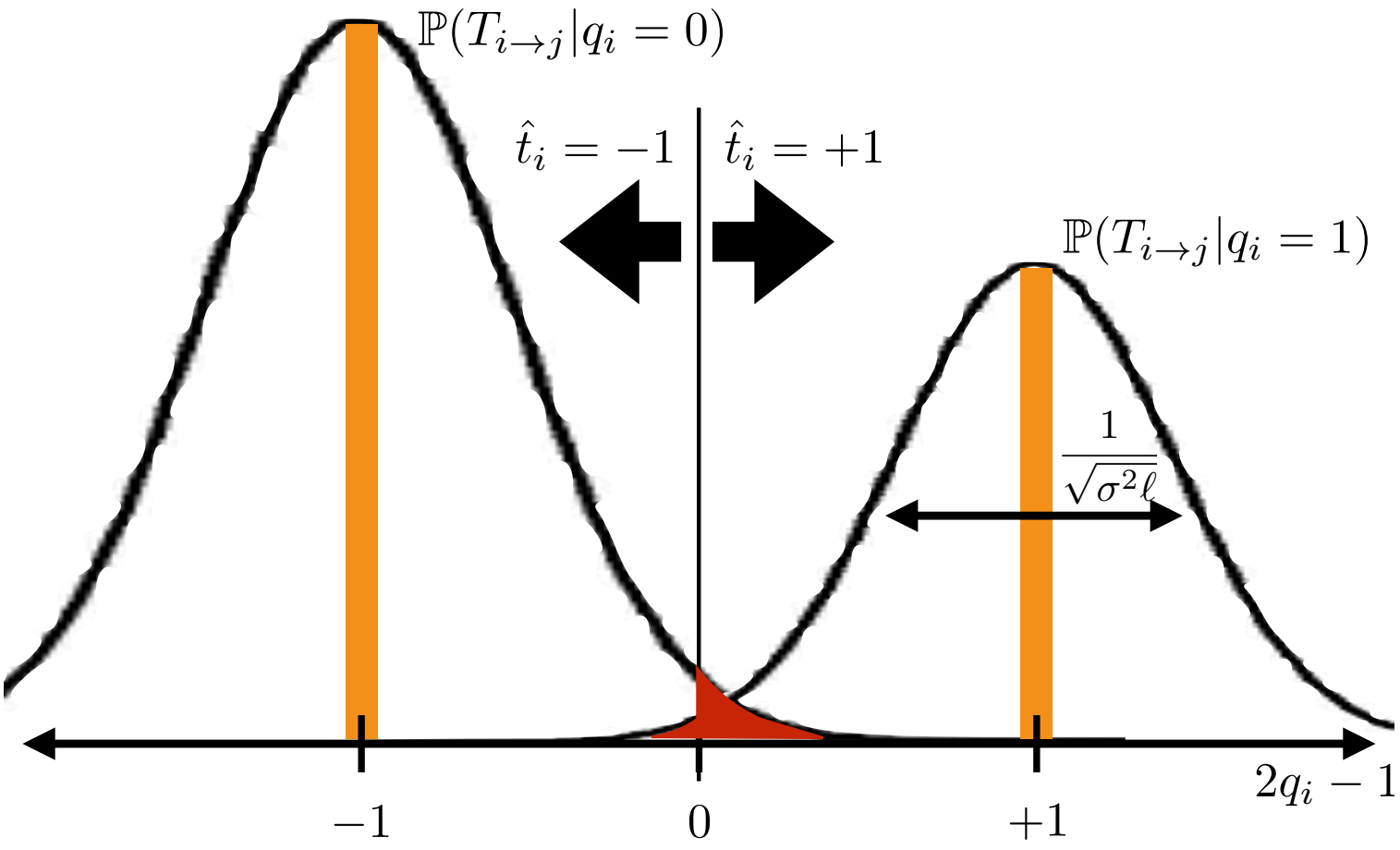
1. DS model + non-adaptive scheme



1. DS model + non-adaptive scheme

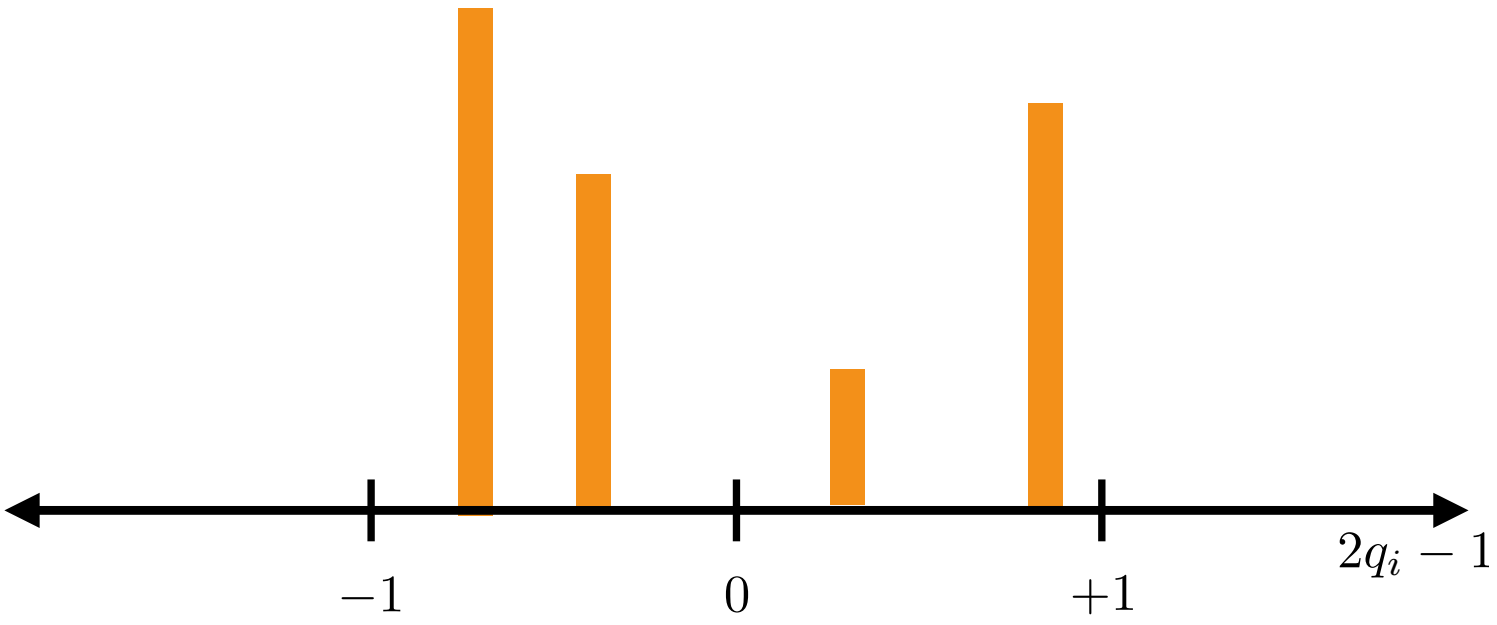


1. DS model + non-adaptive scheme

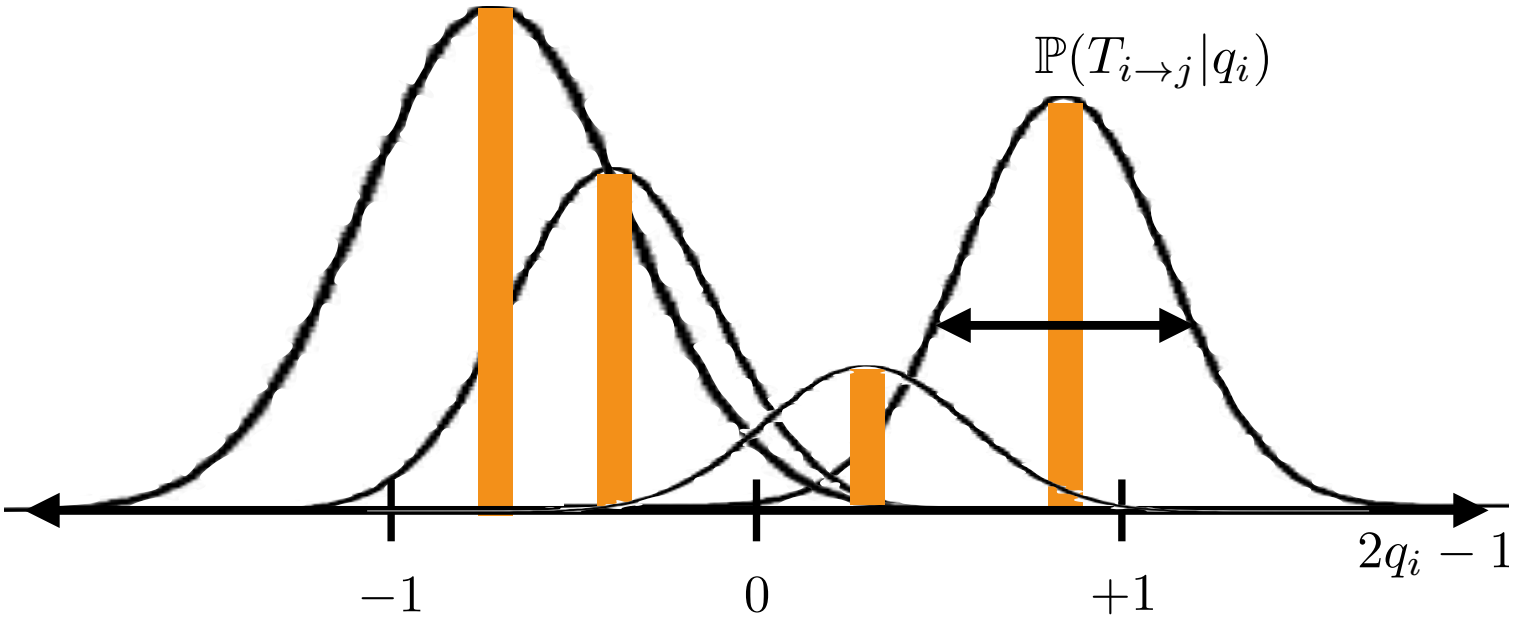


$$P_{\text{err}} \leq e^{-C\sigma^2 \ell}$$

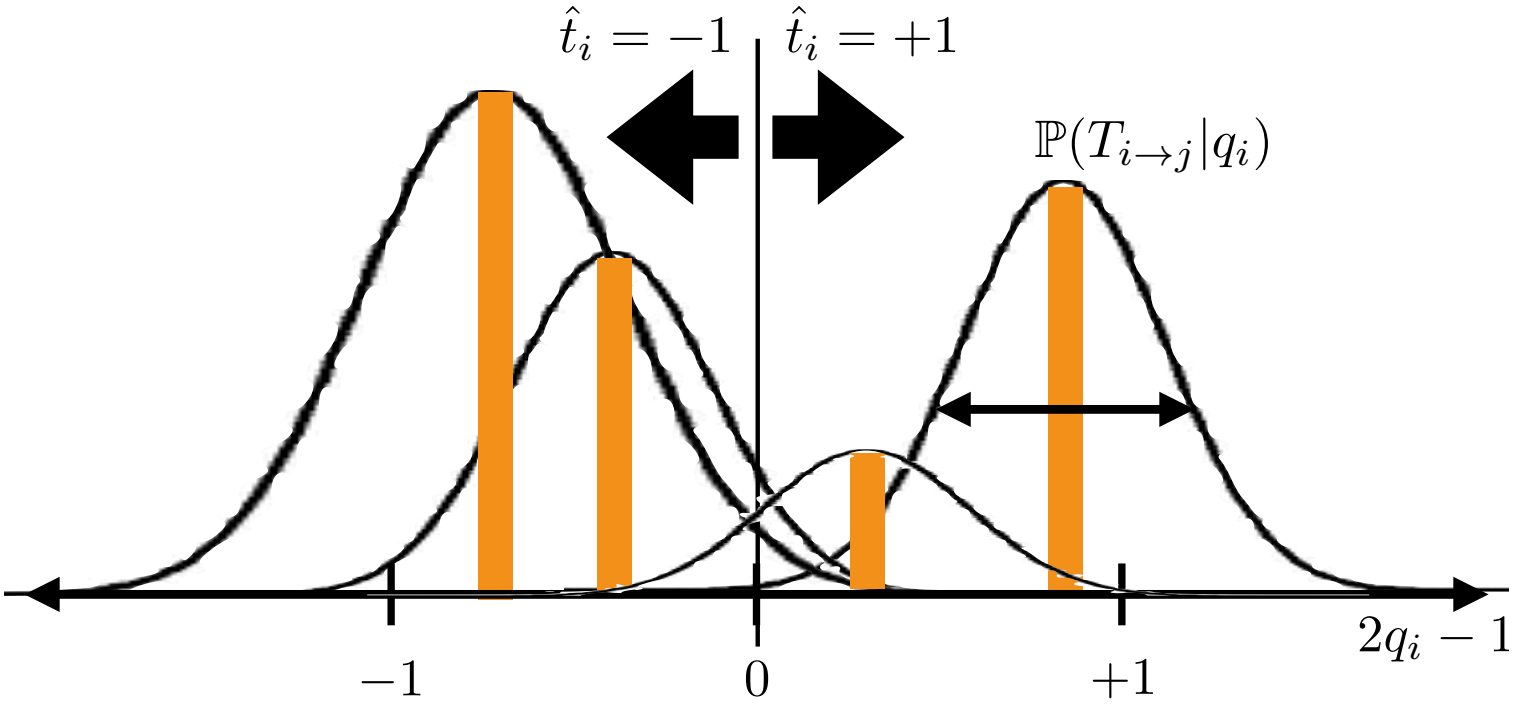
2. Generalized DS model + non-adaptive scheme



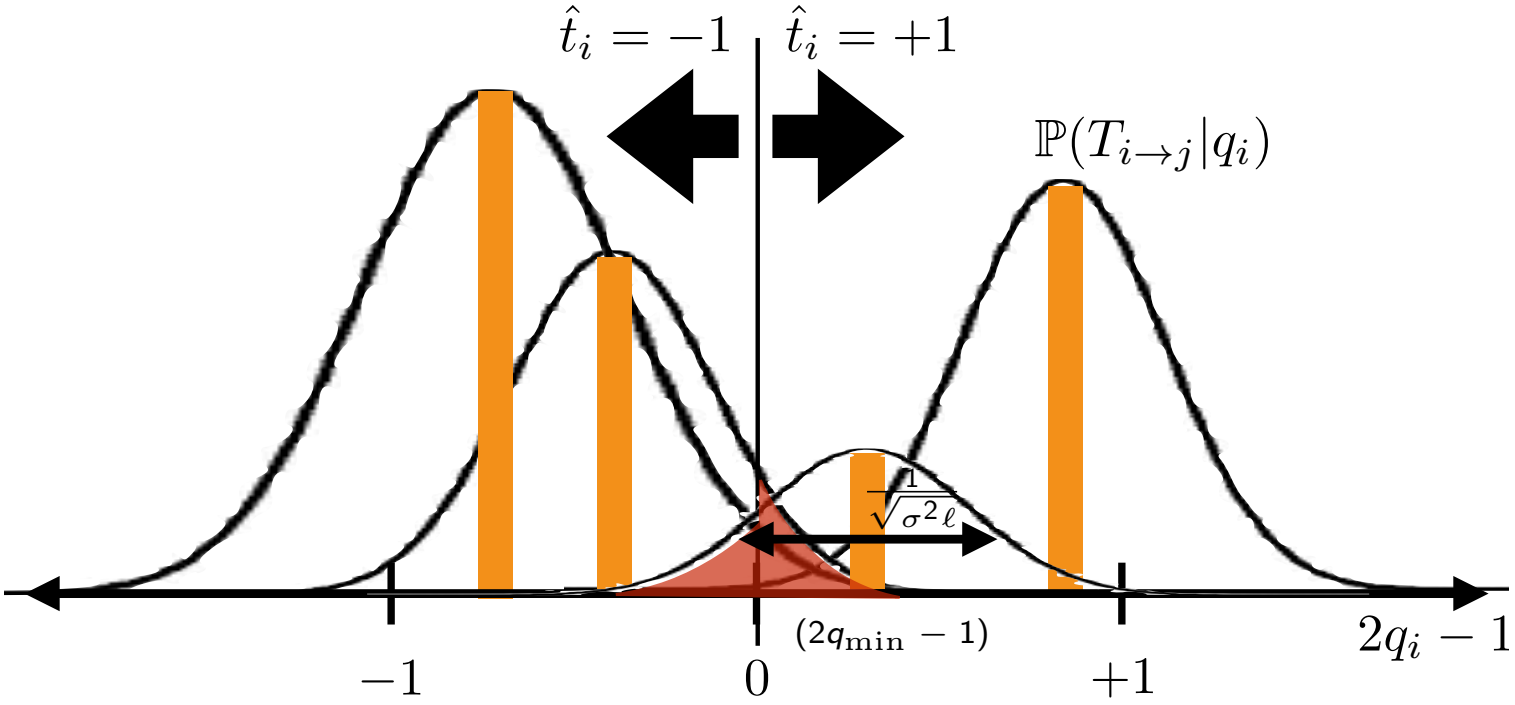
2. Generalized DS model + non-adaptive scheme



2. Generalized DS model + non-adaptive scheme

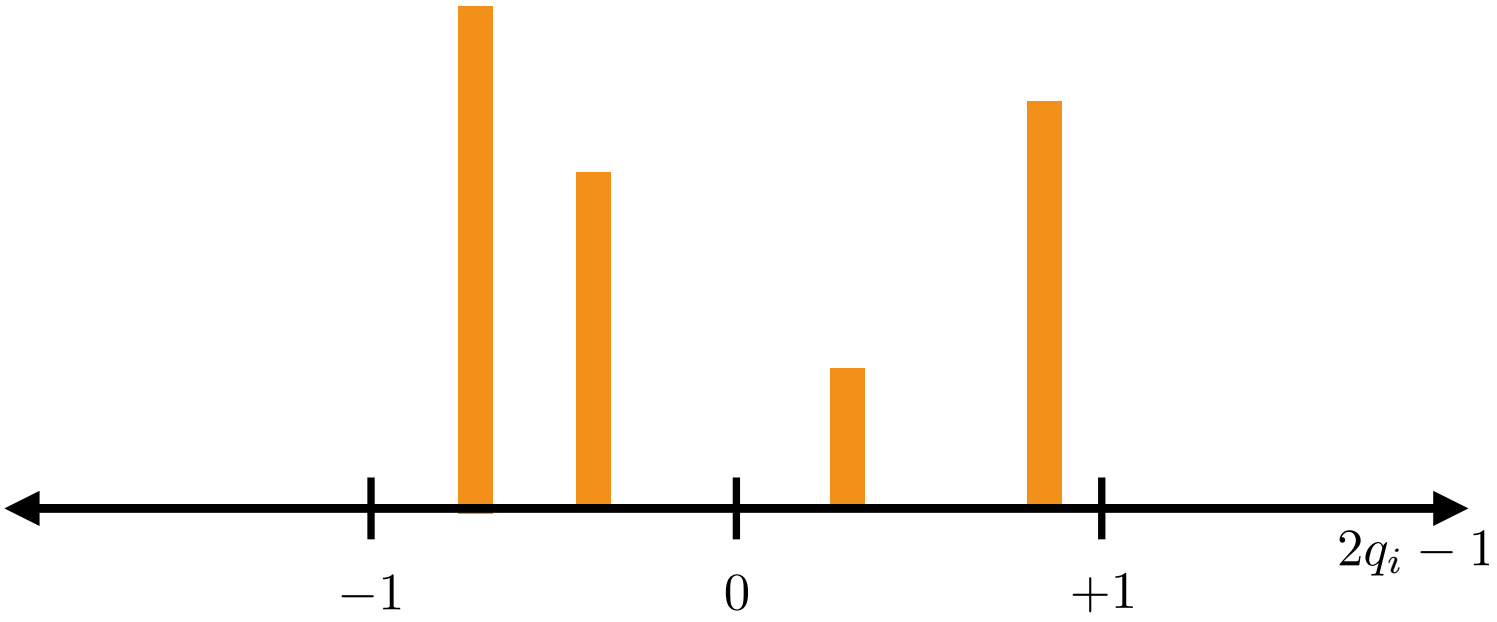


2. Generalized DS model + non-adaptive scheme



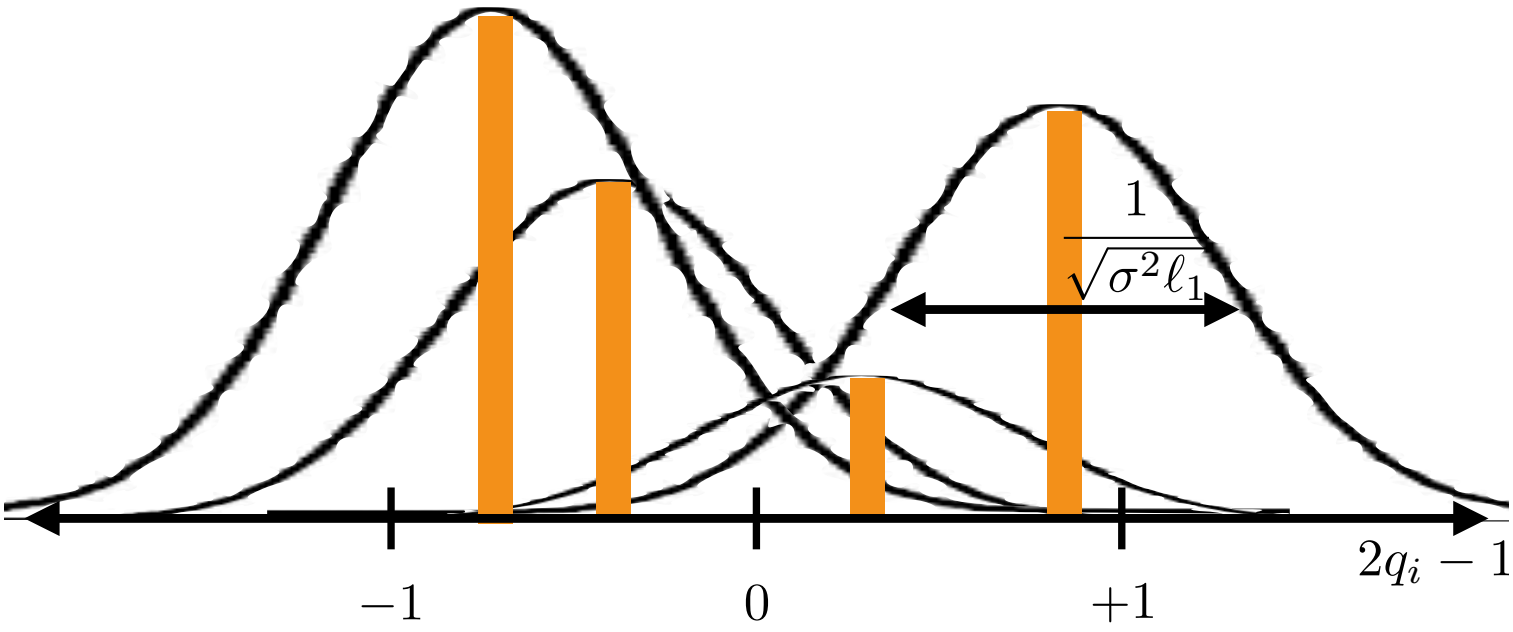
$$P_{\text{err}} \leq e^{-C(2q_{\min}-1)^2 \sigma^2 \ell}$$

3. Generalized DS model + adaptive scheme



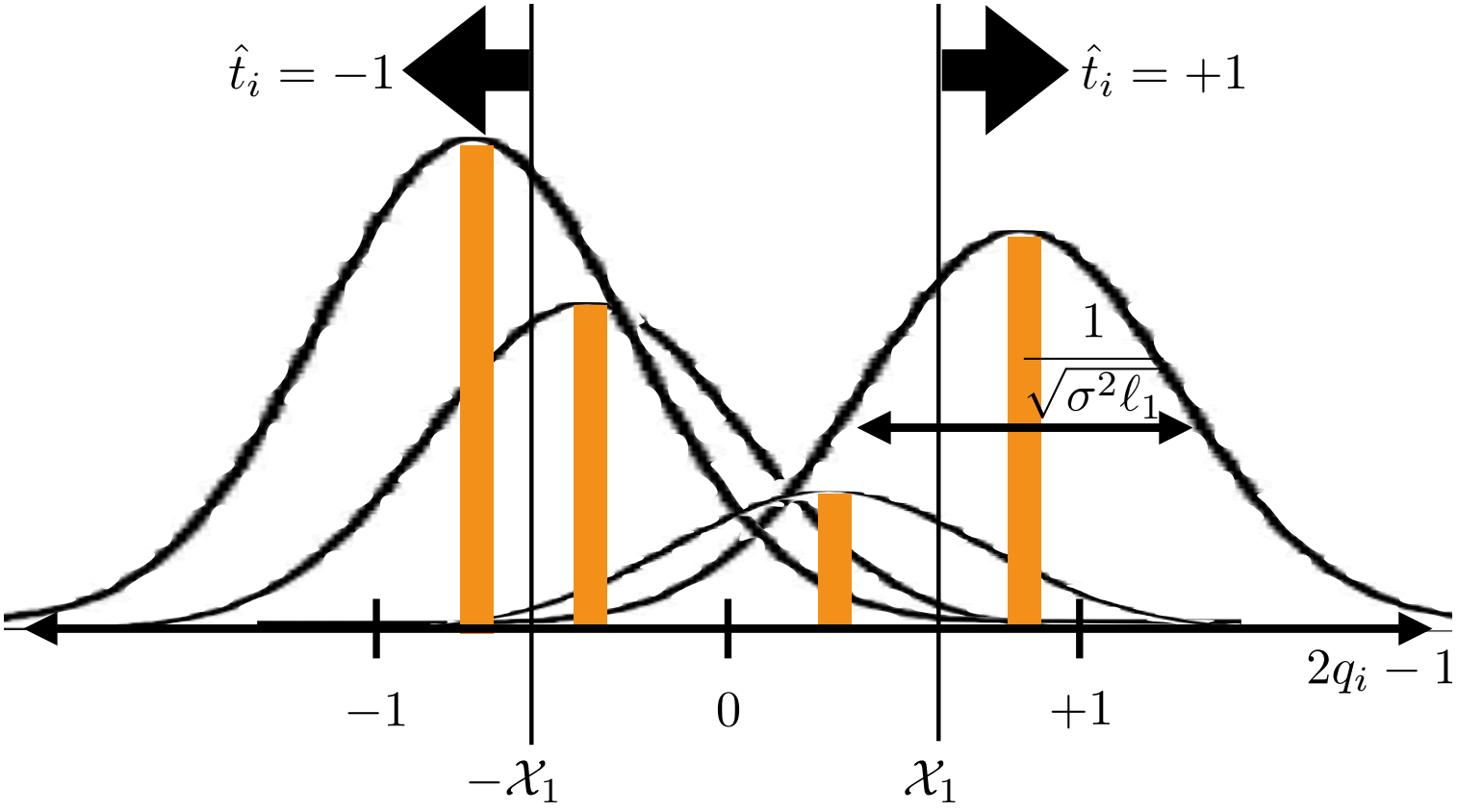
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run iterative algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

3. Generalized DS model + adaptive scheme



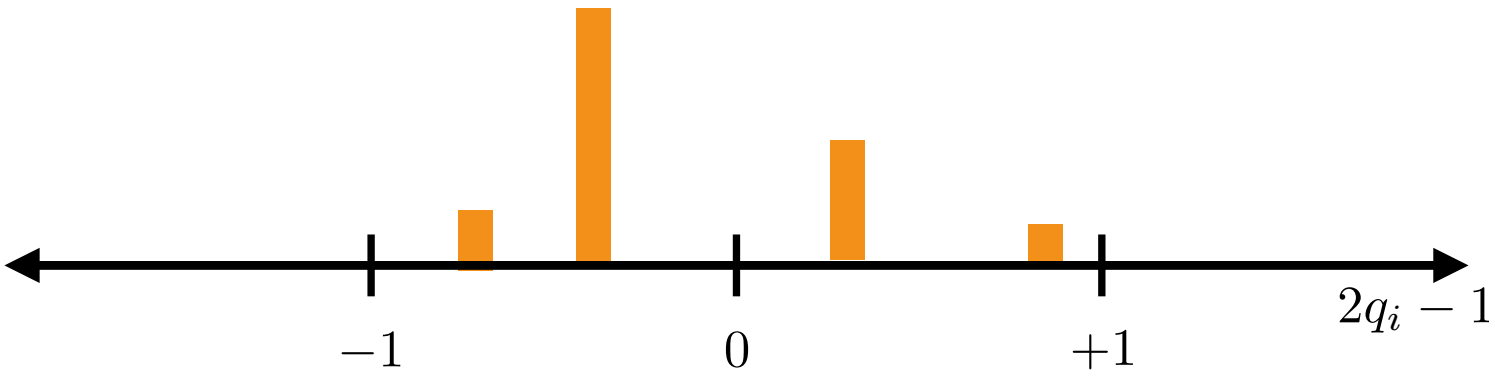
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run iterative algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

3. Generalized DS model + adaptive scheme



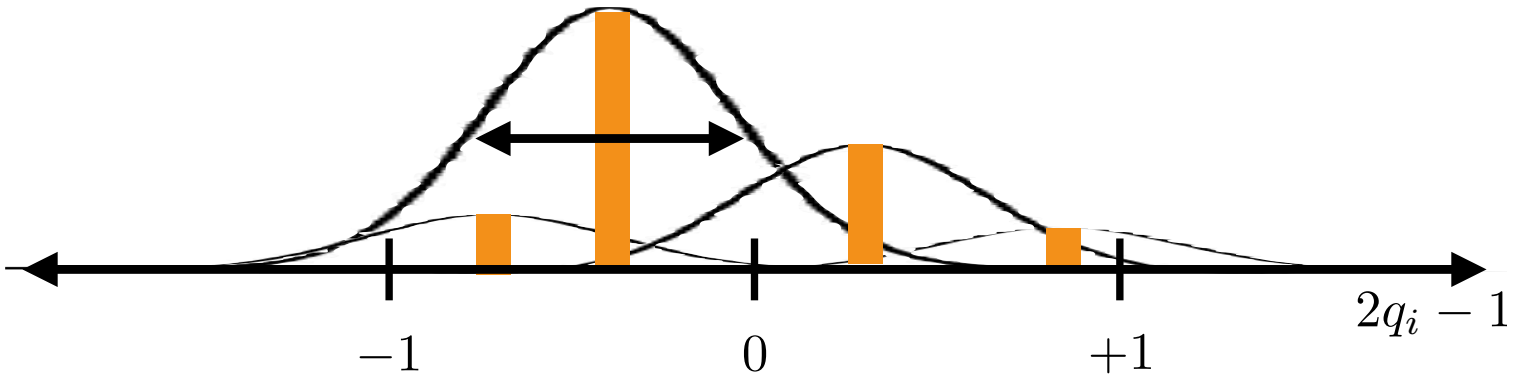
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run iterative algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

3. Generalized DS model + adaptive scheme



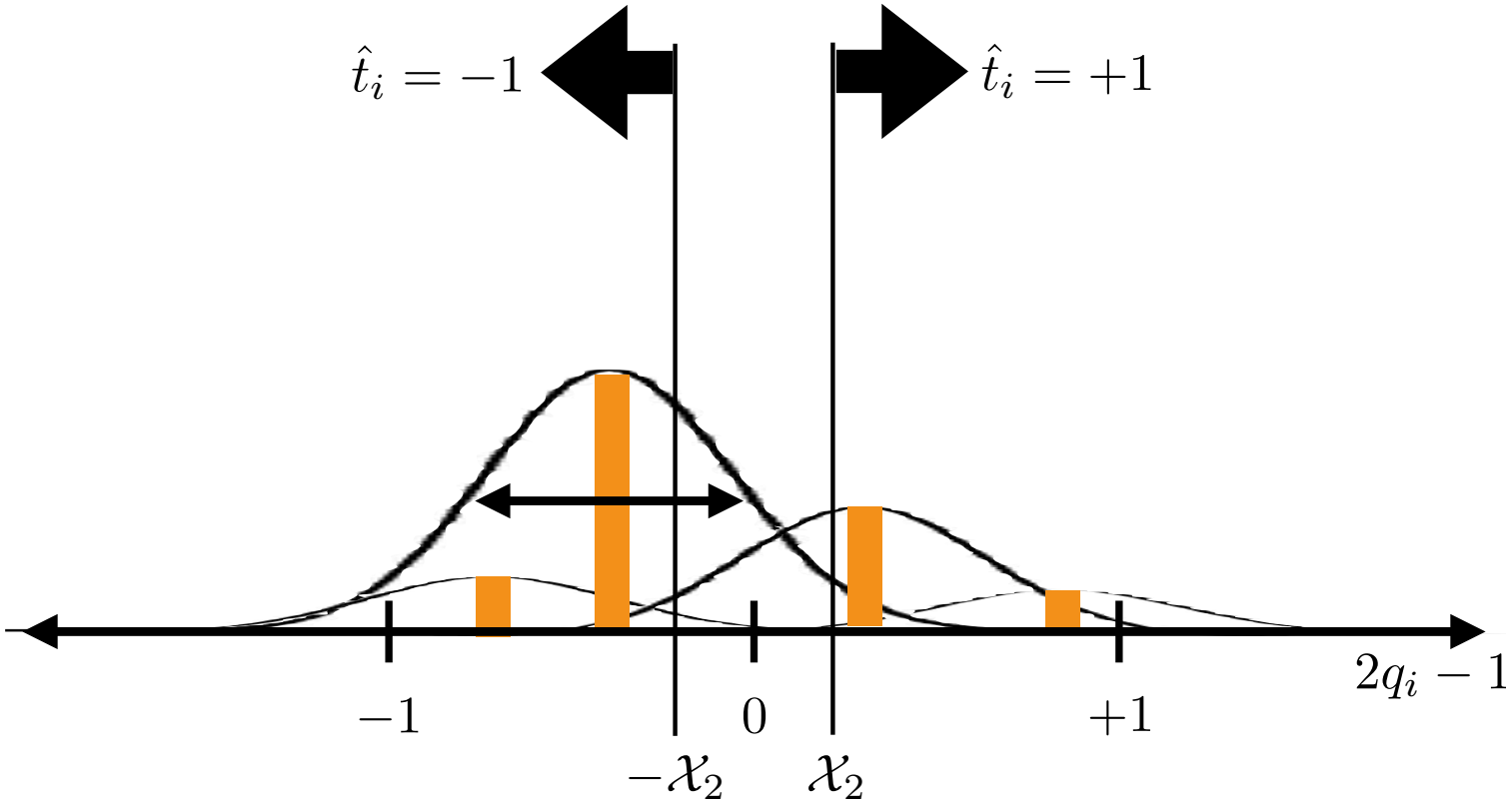
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run iterative algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

3. Generalized DS model + adaptive scheme



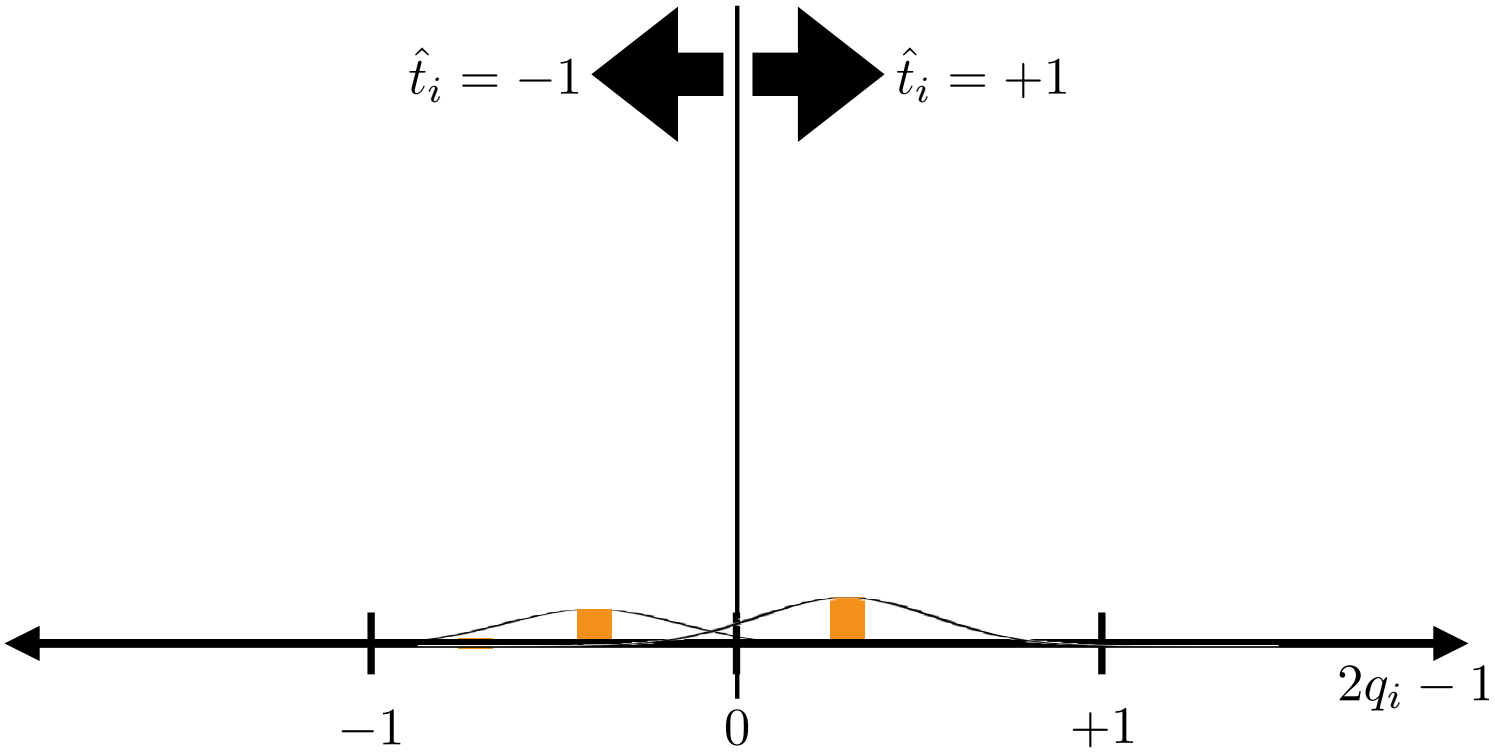
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run iterative algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

3. Generalized DS model + adaptive scheme



- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run iterative algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

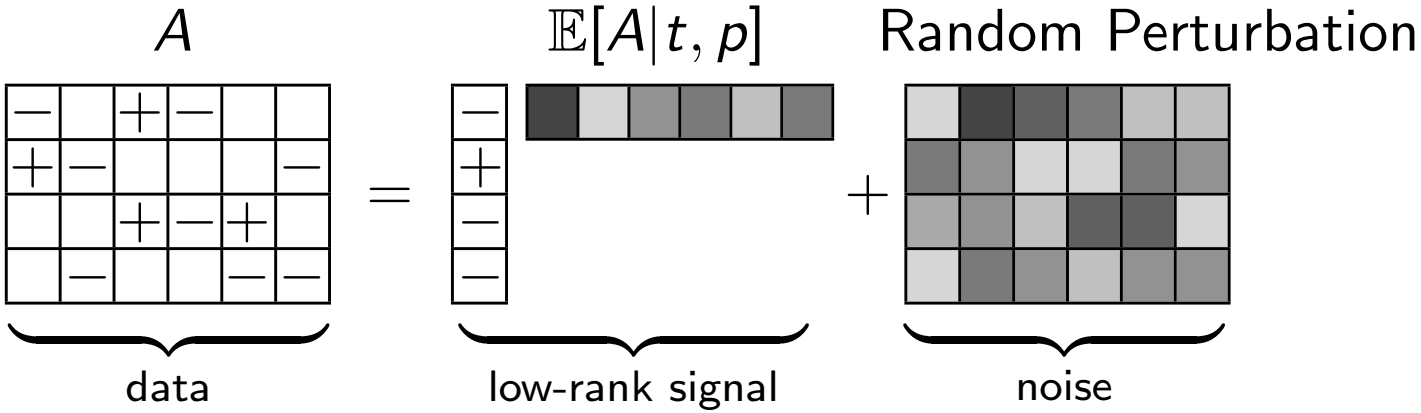
3. Generalized DS model + adaptive scheme



- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run iterative algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

Where did this algorithm come from?

- Spectral method



$$\mathbb{E}[A_{ij}|t_i, p_j] = \frac{\ell}{n} \cdot t_i \cdot (2p_j - 1)$$

$$\mathbb{E}[A|t,p] = \frac{\ell}{n} \cdot t \cdot (2p - \mathbb{1})^T$$

  - Singular vector of a non-backtracking matrix (widely used in community detection [Mossel et al. '13, Krzkala et al. '13, Bordenave et al. '14, Saade et al. '15, etc.])
- Belief propagation for approximating [Peng et al. 12]

$$\mathbb{P}(t_i|A)$$

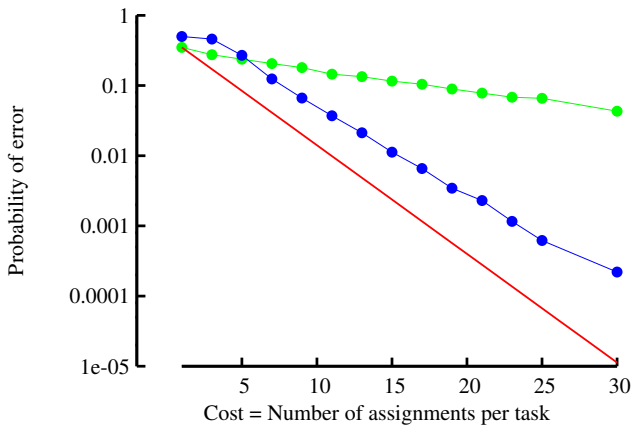
Recap

Dawid-Skene model | generalized Dawid-Skene

- Each worker with quality p_j

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$$

$$P_{\text{err}} \asymp e^{-C\sigma^2\ell}$$



- Criticism: homogeneous tasks
- No gain in adaptivity

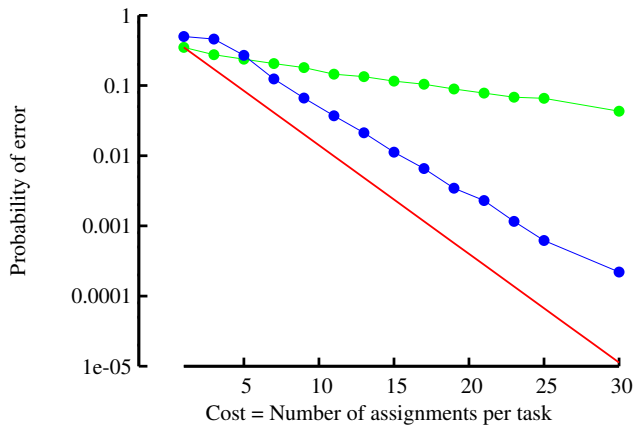
Recap

Dawid-Skene model

- Each worker with quality p_j

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$$

$$P_{\text{err}} \asymp e^{-C\sigma^2\ell}$$



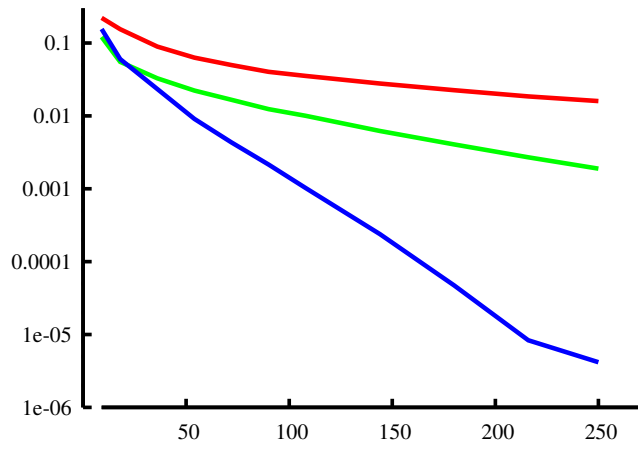
- Criticism: homogeneous tasks
- No gain in adaptivity

generalized Dawid-Skene

- Each task with difficulty q_i

$$\rho^2 = \frac{1}{\frac{1}{m}\sum_{i=1}^{m}\frac{1}{(2q_i-1)^2}}$$

$$P_{\text{err}} \asymp e^{-C\rho^2\sigma^2\ell}$$

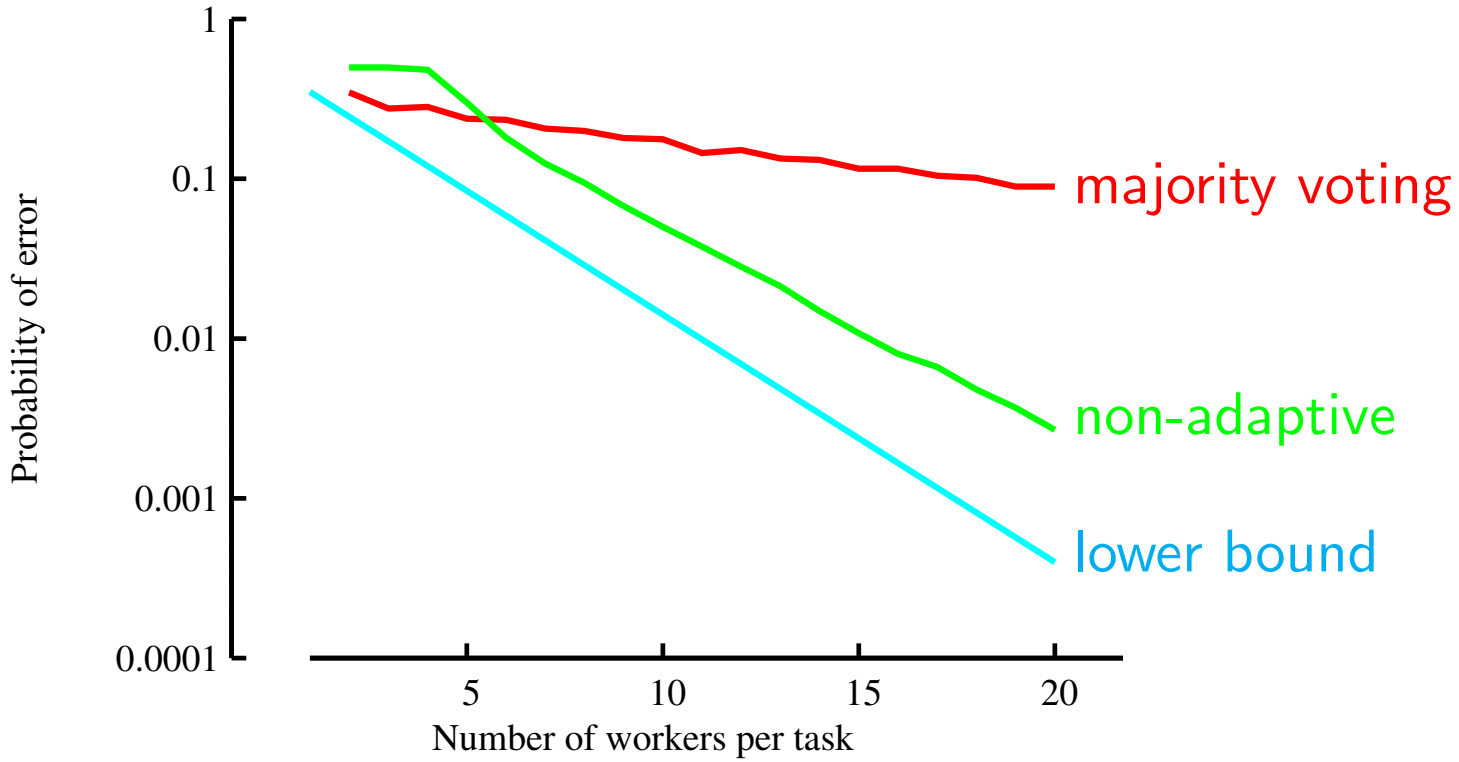


- non-adaptive limit:

$$P_{\text{err}} \asymp e^{-C(2q_{\min}-1)^2\sigma^2\ell}$$

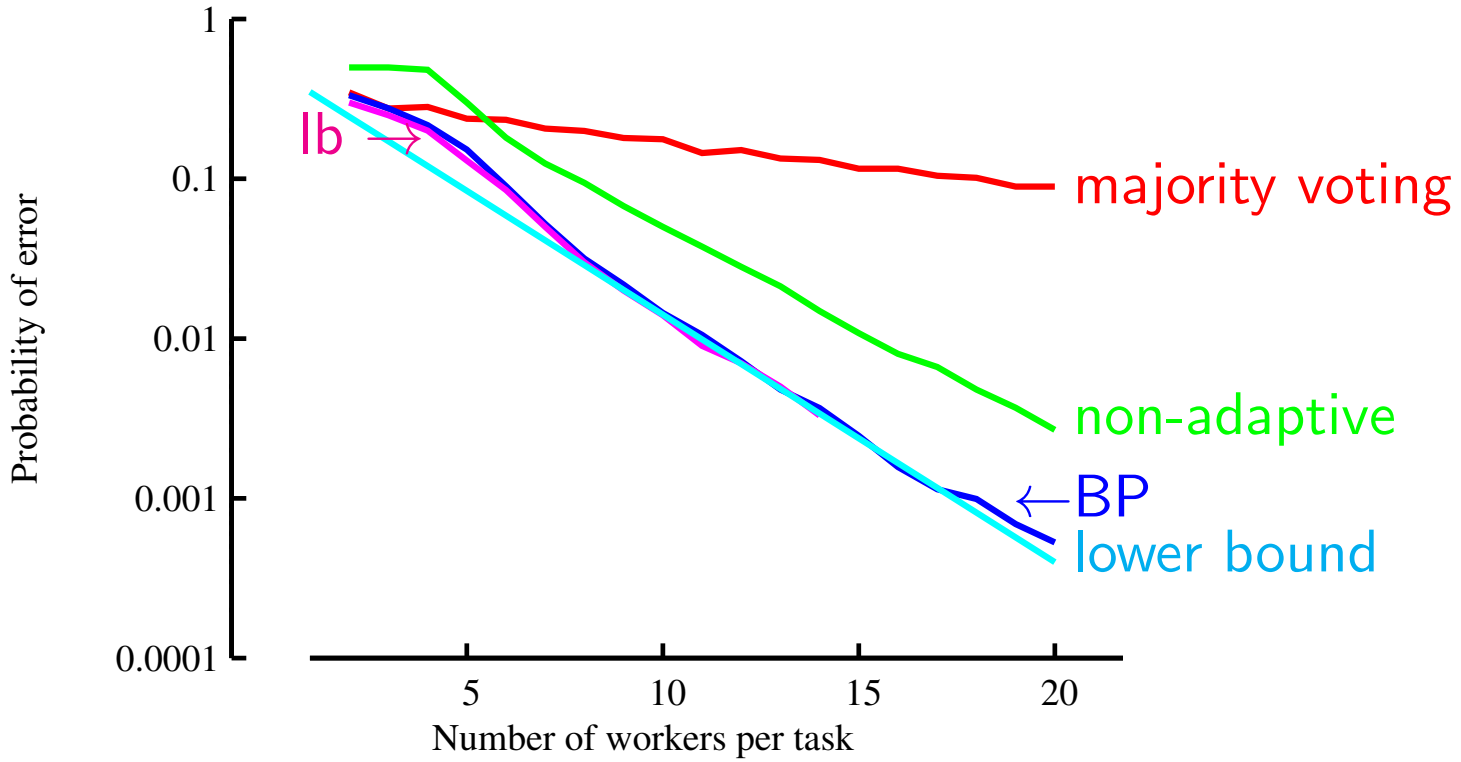
Ongoing/future work

What about Belief Propagation and Expectation Maximization?



Ongoing/future work

What about Belief Propagation and Expectation Maximization?



[Ok, Oh, Shin, Yi '16]

$$P_{\text{err}}^{\text{BP}} - P_{\text{err}}^{\text{LB}} \leq m^{-\gamma} \qquad \text{for some } \gamma > 0$$

Related work

- Crowdsourcing in machine learning

  [ShengProvostIpeirotis2008] - first modern application of DS model
  [WhitehillWuBergsmaMovellanRuvolo2009] - NIPS, release datasets
  [WelinderBransonPeronaBelongie2010] - NIPS, release datasets
  [GhoshKaleMcAfee2011] - first analysis of DS model $P_{\mathrm{err}} \leq \frac{C}{\sigma^2 \ell}$
  [HoJabbariVaughan2013] - online arrival of workers, mixture of DS model

- Lots of recent papers...
- In this talk...
  - *Optimality of Belief Propagation for Crowdsourced Classification*, Ok, O. Shin, Yi, ICML 2016
  - *Achieving Budget-optimality with Adaptive Schemes in Crowdsourcing*, Khetan, O. NIPS 2016