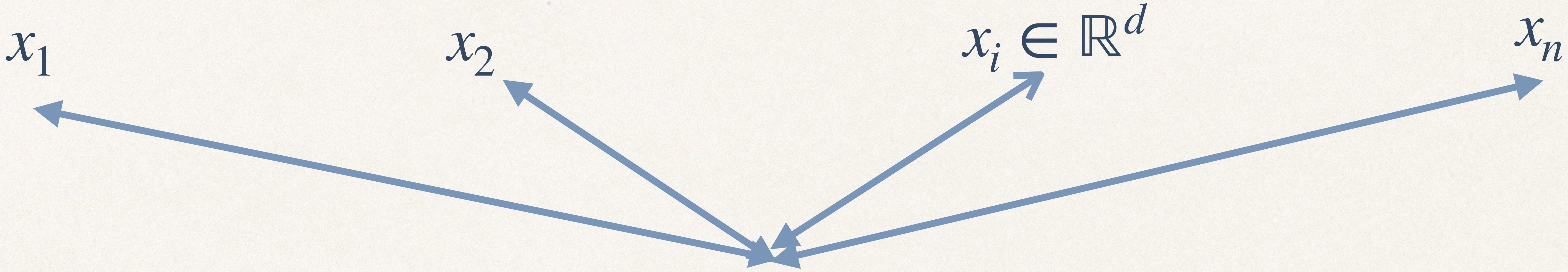
Correlated quantization for distributed mean estimation and optimization

Ananda Theertha Suresh, Ziteng Sun, Jae Hun Ro, and Felix Yu
Google Research New York

Problem statement

High dimensional mean estimation with communication constraints



Goal: estimate $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

What is the best estimate for a given communication cost?

Motivation: distributed optimization

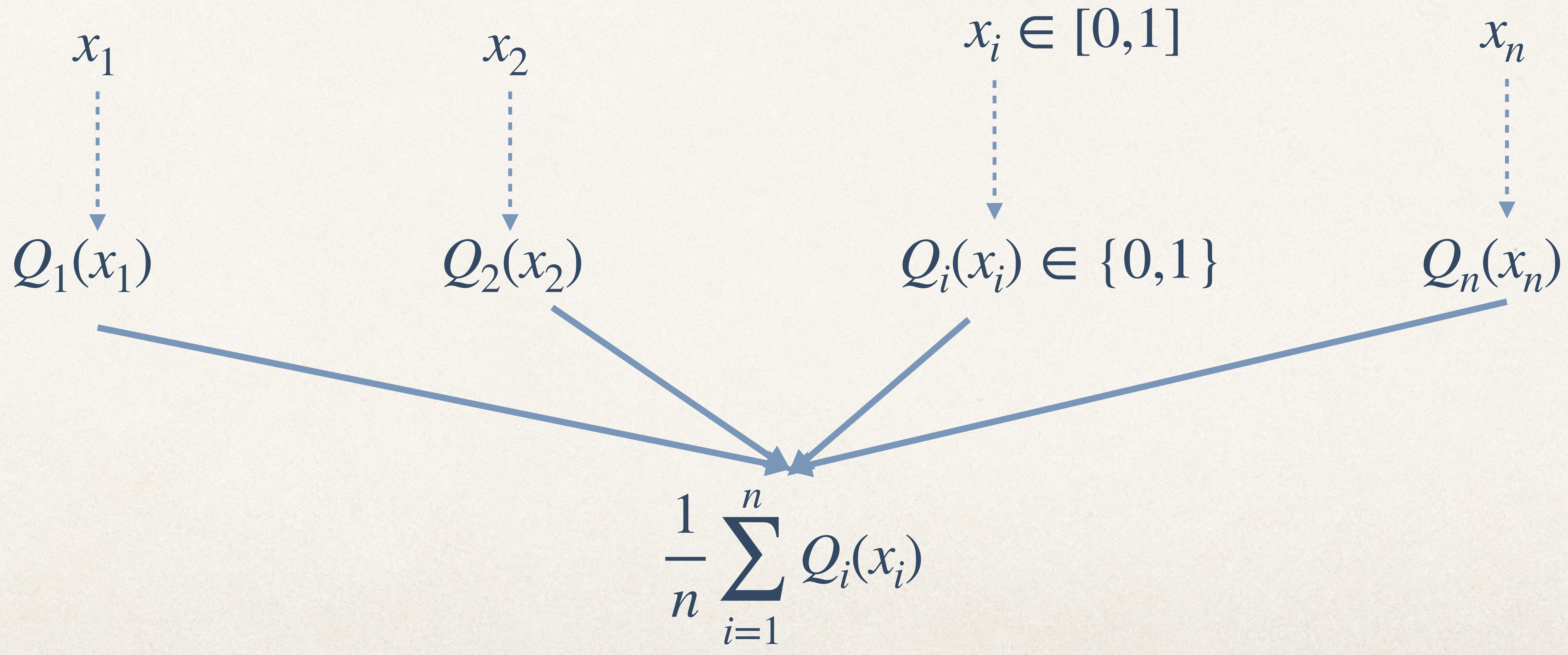
- Distributed optimization uses mean estimation as subroutine
  - At each round users send gradients / model updates to the server
  - Server aggregates gradients to compute the new model
  - Communication cost is a bottleneck

Outline

- One-dimensional mean estimation (this talk)
  - Existing algorithms: min-max bounds
  - New algorithm: instance-specific bounds
  - Optimality
- Extensions to high dimensions
- Applications to distributed optimization

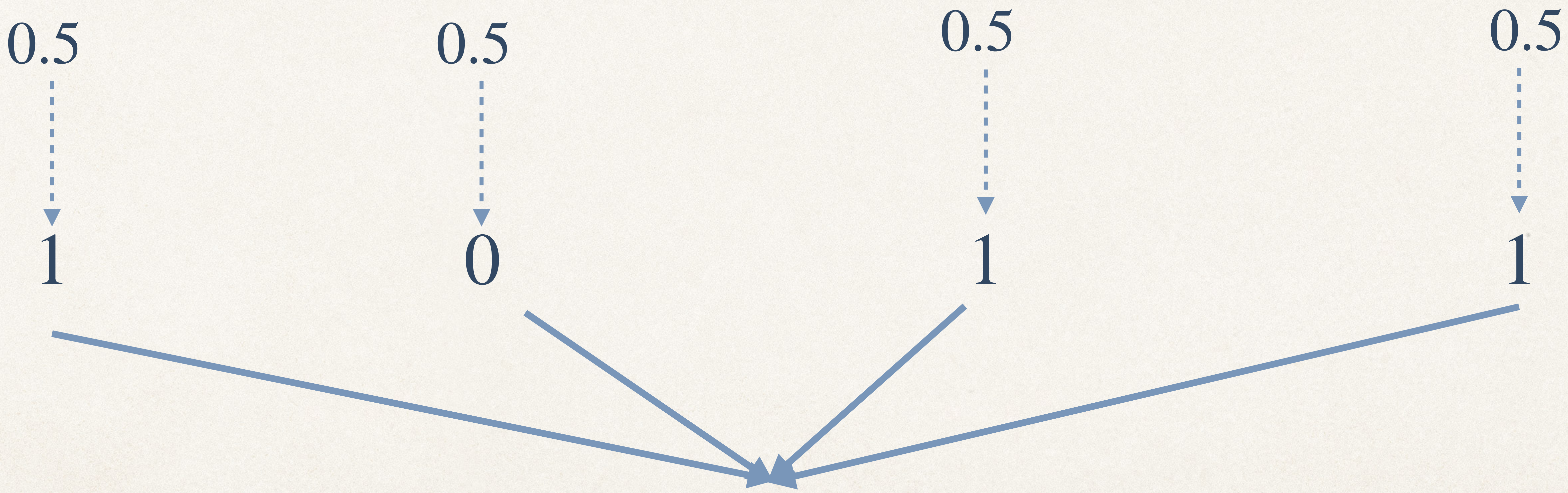
One-bit one-dimensional mean estimation

Goal: estimate $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$



Standard stochastic rounding

$$Q_i(x_i) = B(x_i) = \begin{cases} 1 \text{ with prob. } x_i \\ 0 \text{ with prob. } 1 - x_i \end{cases}$$



Users quantize independently

$$\frac{1}{n}\sum_{i=1}^{n} Q_i(x_i) = 0.75$$

What if data is more favorable?

- Standard stochastic quantization

$$E\left(\frac{1}{n}\sum_{i=1}^{n} Q_i(x_i) - \bar{x}\right)^2 \lesssim \frac{1}{n}$$

Stochastic rounding is min-max optimal up to constants

What if data is more favorable?

- Standard stochastic quantization

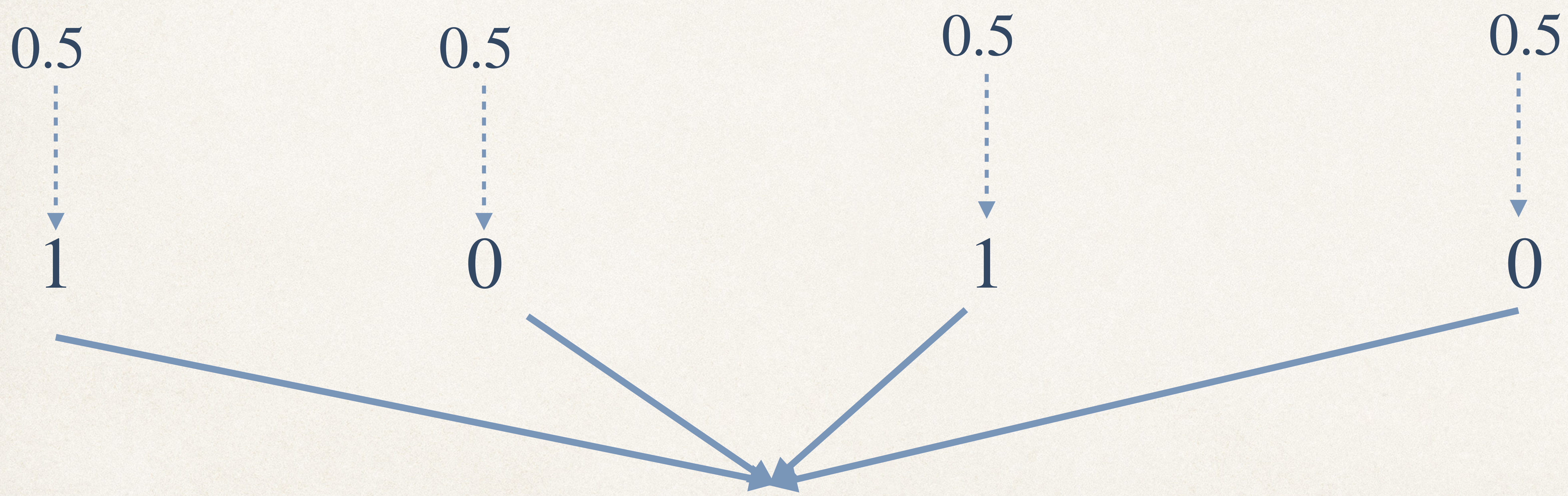
$$E\left(\frac{1}{n}\sum_{i=1}^{n} Q_i(x_i) - \bar{x}\right)^2 \lesssim \frac{1}{n}$$

Stochastic rounding is min-max optimal up to constants

- Suppose points are close-by e.g., $x_i \sim N(0.5, 0.0001)$
  - Can we improve the estimate?
  - Can we provide instance-specific bounds?

Correlated quantizer

$$Q_i(x_i) = B(x_i) = \begin{cases} 1 \text{ with prob. } x_i \\ 0 \text{ with prob. } 1 - x_i \end{cases}$$



$$\frac{1}{n}\sum_{i=1}^{n} Q_i(x_i) = 0.5$$

Correlated quantization: If one user rounds up, other users tend to round down

Results on correlated stochastic quantizer

Theorem: For any $x_1, x_2, \ldots, x_n$ such that each $x_i \in [0,1]$,

$$E\left(\frac{1}{n}\sum_{i=1}^{n} Q_i(x_i) - \bar{x}\right)^2 \lesssim \frac{\sigma_{\text{md}}}{n} + \frac{1}{n^2}$$

where $\sigma_{\text{md}} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$.

Results on correlated stochastic quantizer

Theorem: For any $x_1, x_2, \ldots, x_n$ such that each $x_i \in [0,1]$,

$$E\left(\frac{1}{n}\sum_{i=1}^{n} Q_i(x_i) - \bar{x}\right)^2 \lesssim \frac{\sigma_{\text{md}}}{n} + \frac{1}{n^2}$$

where

$$\sigma_{\text{md}} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|\,.$$

- A simple modification of the stochastic rounding algorithm
- Does not require prior knowledge of σ_{md}
- Data dependent bound

Optimality of the correlated quantizer

Theorem: For any one bit quantizer and any σ_{md}, there exists $x_1, x_2, \ldots, x_n$ such that $\sigma_{\text{md}} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$ and

$$E\left(\frac{1}{n} \sum_{i=1}^{n} Q_i(x_i) - \bar{x} \right)^2 \gtrsim \frac{\sigma_{\text{md}}}{n}$$

Thank you! Poster tonight @ **Hall E #1111**

- One-bit one-dimensional mean estimation
    - Existing algorithms: min-max bounds
    - New algorithm: instance-specific bounds
    - Optimality
- Extensions to multiple bits
- Extensions to high dimensions
- Applications to distributed optimization