Calibration with Privacy in Peer Review

Wenxin Ding [1], Gautam Kamath [2], Weina Wang [3], Nihar B. Shah [3]

[1] University of Chicago, [2] University of Waterloo, [3] Carnegie Mellon University

Our **goal** is to design methods for the conference to accept better papers while guarding against privacy leakage due to calibration.

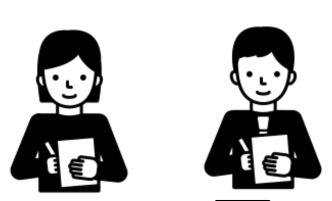
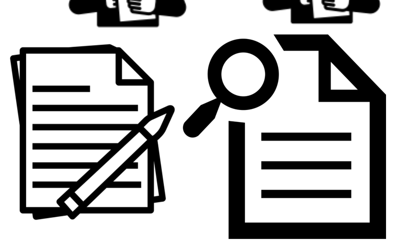
(1) We provide explicit computationally-efficient algorithms for calibration with privacy that optimally trades off the error of the conference and the error of the adversary.

(2) We establish the structure of the Pareto optimal curve between the errors.

MOTIVATION

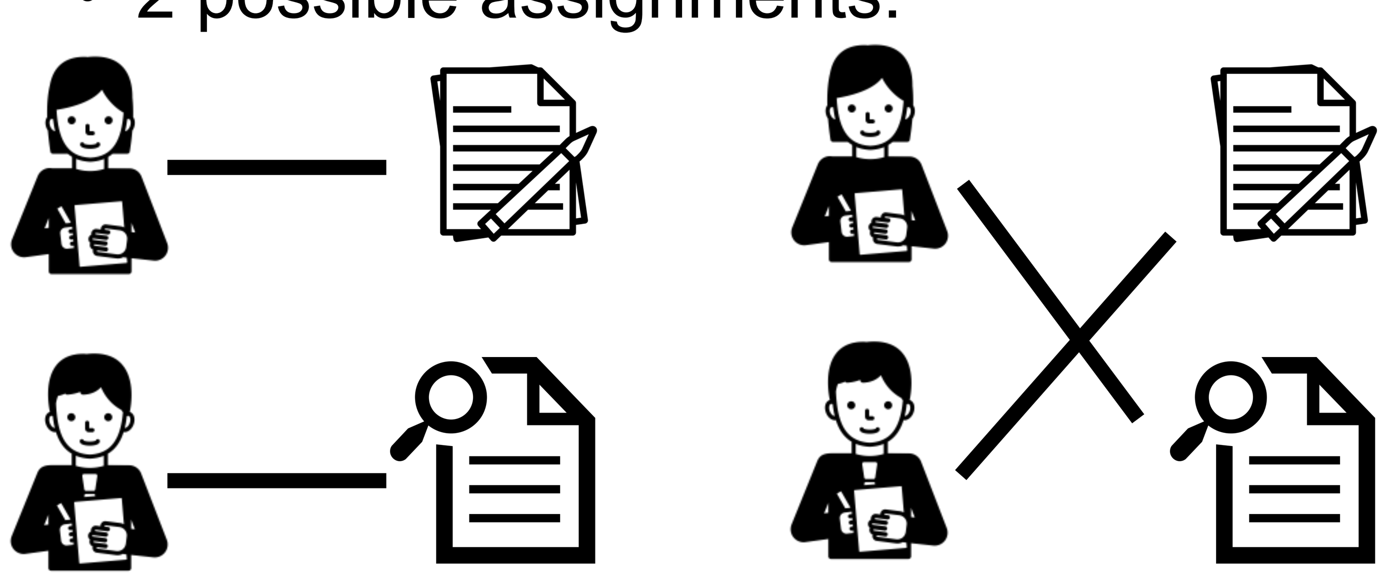
- Reviewers in peer review are often miscalibrated.
- A number of algorithms have been proposed to calibrate reviews.
- Attempts of calibration can leak sensitive information about which reviewer reviewed which paper.
- Another challenge is a small number of samples (reviews) per reviewer.

PROBLEM SETTING

- 2 reviewers 
- 2 papers 
- Miscalibration function of reviewer j: β_j
- Noise of reviewer j : ε_j
- Quality of paper i: θ_i^*
- Score of paper i reviewed by reviewer j : $s_i = \beta_j(\theta_i^*) + \varepsilon_j$

β, distributions of ε and θ^* are **known**

- Marginal p.d.f. of scores given by reviewer j : f_j
- 2 possible assignments:



Assignment 1 Assignment 2

Conference: accept higher-quality paper by estimating paper quality

Error of the conference E_C: probability of accepting the lower-quality paper

Adversary: guess true assignment by MAP

Error of the adversary: probability of guessing the wrong assignment

Per-instance error: error under specific scores

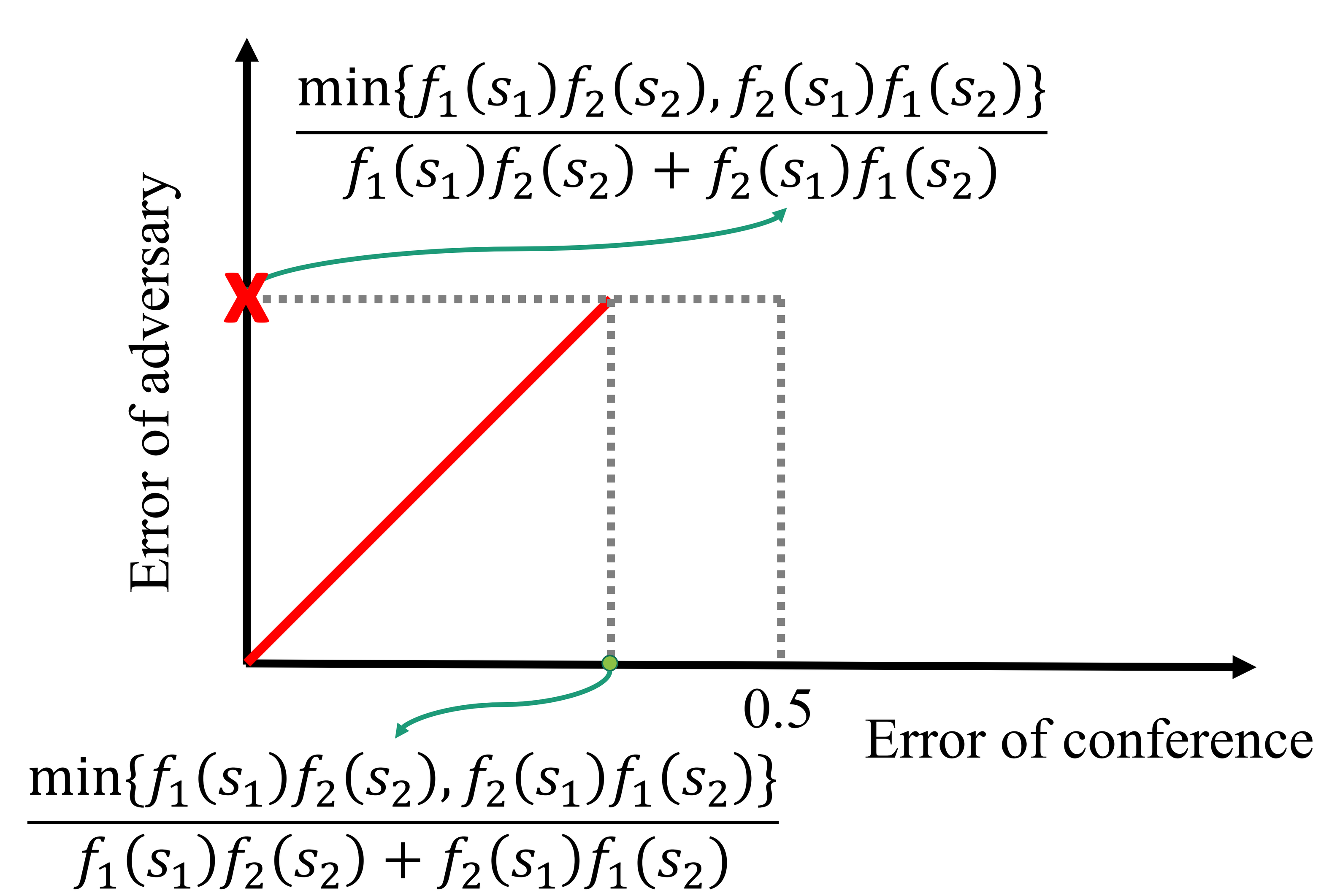
Average-case error: error over the distribution of scores

MAIN RESULTS

1. Establish the Pareto frontier of the tradeoff between privacy and utility.
2. Design explicit computationally-efficient algorithms that we prove are Pareto optimal.

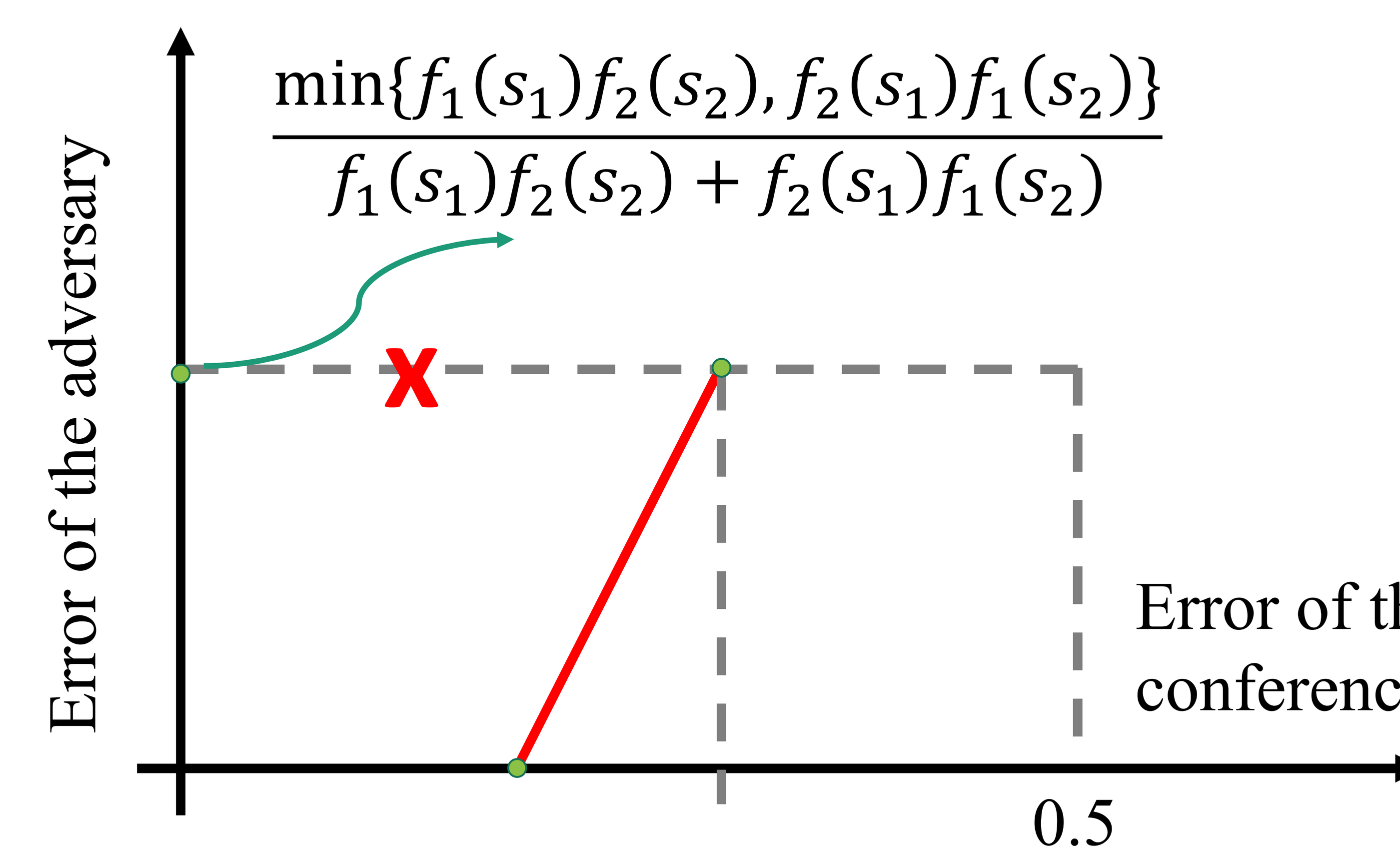
- Noiseless: $\varepsilon = 0$

Pareto frontier:



- Noisy: $\varepsilon \sim N(0, \sigma^2)$

Pareto frontier:



ALGORITHMS

Per-instance error in the noiseless case:

Input: s_1, s_2, maximum allowable $E_C(s_1, s_2)$

If one paper has higher estimated quality under both assignments: **accept the paper**

Otherwise, the conference selects probability p:

- with probability p the **conference calibrates under the true assignment**;
- with probability $1-p$ the conference **calibrates under the wrong assignment**

Average-case error in the noiseless case:

Input: maximum allowable E_C

If E_C is large:

run per-instance algorithm with $E_C = 1$

Otherwise, the conference flips a coin:

- if coin outcome is head: **run per-instance algorithm with $E_C = 1$** ;
- Otherwise, the conference **calibrates under the true assignment**

*algorithm for noisy case is available in the paper