A System-Level Analysis of Conference Peer Review

Yichi Zhang¹, Fang-Yi Yu², Grant Schoenebeck¹ and David Kempe³
¹University of Michigan, ²Harvard University, ³University of Southern California

Motivation

Three constituencies have diverse objectives:
- Authors want their papers to be accepted;
- Conferences want to accept more high-quality papers and fewer low-quality papers;
- Reviewers want to avoid being overburdened with reviewing tasks.

Several attempts to navigate the tradeoffs:
- Increasing the bar of acceptance;
- Soliciting more and more reviews per submission;
- Requiring historical reviews to be included with each resubmission.

Question: How well do various policies work, and why do they work or not?

Model

1. Stackelberg game

We’ll solicit \( m \) i.i.d. reviews per paper in each round of (re)submission;
Compute the posterior expected quality of each paper;
Decide whether to accept or not based on a threshold \( \tau \) on the expected quality.

2. Quality and signal models

Binary model:
- Paper quality: \([-1, 1]\);
- Review signal: flip the true quality with \( p = 1 - \beta \).

Continuous model:
- Paper quality: convex domain, e.g. \( \mathbb{H} \);
- Review signal = true quality + continuous and zero-mean noise, e.g. Gaussian noise with std \( \sigma \).

3. Noiseless authors with unlimited resubmissions

Noiseless: authors perfectly know papers’ true qualities.
Unlimited resubmission: each paper can be submitted an unlimited number of times.

An author will submit if

\[
P_{acc}(\phi, q) > \frac{1 - \eta}{V - \eta} \]

- \( P_{acc}(\phi, q) \): probability of acceptance under policy \( \phi \) and quality \( q \).

Proposition (informal)

In this setting, an author will
- submit and keep resubmitting a paper until acceptance if its quality is \( q \geq \theta \);
- submit to the sure bet in the first round otherwise.

4. Conference quality (Q) and review burden (B)

Quality: the summation of accepted papers’ quality;
Burden: the expected number of reviews per paper.

The QB-tradeoff: Quality and Burden cannot be optimized at the same time.

De Facto Threshold and Resubmission Gap

Acceptance threshold: \( \tau \)
De facto threshold: \( \theta \)
Resubmission gap: \( \tau - \theta \)
- The conference quality is maximized at \( \theta = 0 \);
- The resubmission gap is usually positive.

Acceptance Rate

Higher \( \tau \) implies lower acceptance rate?
- It depends on the hazard rate of the prior of quality.

Review Quality v.s. Quantity

How much does increasing \( m \) help?
- Larger \( m \) → fewer rounds of review but a heavier burden in each round;
In the binary model, what is the optimal \( m \)?

- \( m = 1 \) is optimal when \( \beta \) is very low or very high;
- Larger \( V \) (and larger \( \eta \)) → larger \( m \) is optimal;

Generalizations

- Authors have noisy signals;
- Categorical model: finite paper qualities + ICLR data estimated review noise.