A System-Level Analysis of Conference Peer Review

Yichi Zhang[1], Fang-Yi Yu[2], Grant Schoenebeck[1] and David Kempe[3]

[1]University of Michigan, [2]Harvard University, [3]University 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

Conference

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 τ on the expected quality.

Author(s)

Given the acceptance policy, in each round, we each will

- compute my paper's expected quality based on my private signal and previous reviews;
- decide whether to submit to the conference (utility V if accepted), or a sure bet (utility 1). Discount factor η.

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. $\mathbb{R}$;
- Review signal = true quality + continuous and zero-mean noise, e.g. Gaussian noise with std σ.

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

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

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

Proposition (informal)

In this setting, an author will

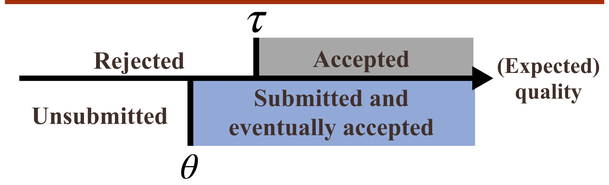
- submit and keep resubmitting a paper until acceptance if its quality is $q \geq \theta$;
- submit to the sure bet in the first round otherwise.

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: τ

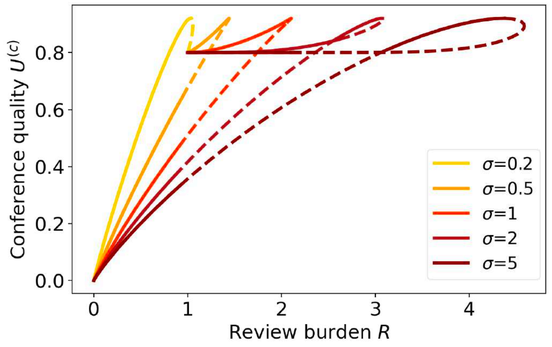
De facto threshold: θ

Resubmission gap: $\tau - \theta$

- The conference quality is maximized at $\theta = 0$;
- The resubmission gap is usually positive.

QB-tradeoff

The conference can trade off its quality and the review burden by varying its acceptance threshold:

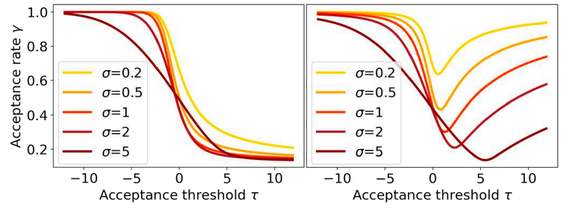


- Whether increasing or decreasing τ from $\theta = 0$ is Pareto optimal depends on the prior and the review quality.
- To improve the QB-tradeoff:
  - Improve review quality;
  - Decrease the discount factor or the conference value.

Acceptance Rate

Higher τ implies lower acceptance rate?

- It depends on the hazard rate of the prior of quality.



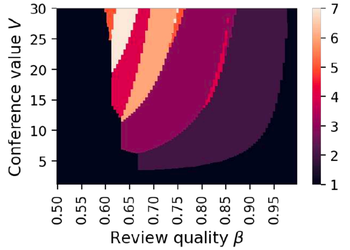
Gaussian prior Cauchy prior

Review Quality v.s. Quantity

How much does increasing m help?

- Larger $m \rightarrow$ 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 β is very low or very high;
- Larger V (and larger η)$\rightarrow$ larger m is optimal;

Generalizations

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