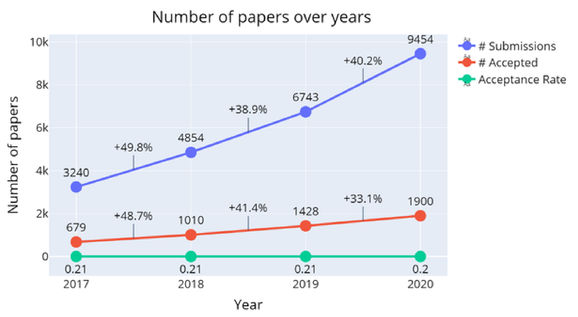
# You Are the Best Reviewer of Your Own Papers: An Owner-Assisted Scoring Mechanism

Weijie Su

University of Pennsylvania

## Explosion of ML/AI conferences




## Who are the best reviewers?

- Leading experts of your field? But there are not many of them, and they are also very busy
- How about yourself? Sounds a bit ironic, but perhaps you can:

Rank your papers according to perceived quality of the papers!

## Trouble with peer review

Inconsistency in Conference Peer Review: Revisiting the 2014 NeurIPS Experiment

Corinna Cortes* and Neil D. Lawrence†

*Google Research, New York
†Computer Lab, University of Cambridge

September 22, 2021

**Abstract**

In this paper we revisit the 2014 NeurIPS experiment that examined inconsistency in conference peer review. We determine that 50% of the variation in reviewer quality scores was subjective in origin. Further, with seven years passing since the experiment we find that for *accepted* papers, there is no correlation between quality scores and impact of the paper as measured as a function of citation count. We trace the fate of rejected papers, recovering where these papers were eventually published. For these papers we find a correlation between quality scores and impact. We conclude that the reviewing process for the 2014 conference was good for identifying poor papers, but poor for identifying good papers. We give some suggestions for improving the reviewing process but also warn against removing the subjective element. Finally, we suggest that the real conclusion of the experiment is that the community should place less onus on the notion of 'top-tier conference publications' when assessing the quality of individual researchers.

- Are you fed up with VERY noisy reviews from NeurIPS, ICML, ICLR...?
- Do you have your best papers rejected but mediocre papers accepted?

## Assumptions

1. Given (final) scores $\widehat{R}_1, \ldots, \widehat{R}_n$ of $n$ items in the possession of the owner, then the owner's utility takes the form $\mathrm{Util}(\widehat{\boldsymbol{R}}) = \sum_{i=1}^n U(\widehat{R}_i)$, where $U$ is a nondecreasing convex function
2. The owner has knowledge of the true ranking of her items. That is, the owner knows which permutation $\pi^\star$ that makes $\pi^\star \circ \boldsymbol{R}$ in nonincreasing order
3. The noise $(z_1, \ldots, z_n)$ follows an exchangeable distribution in the sense that $(z_1, \ldots, z_n)$ has the same probability distribution in $\mathbb{R}^n$ as $\rho \circ \boldsymbol{z} := (z_{\rho(1)}, \ldots, z_{\rho(n)})$ for any permutation $\rho$ of $1, \ldots, n$

## Reporting the true ranking

What's the optimal strategy of the author, if she is rational and aims to maximize her expected utility by supplying any ranking $\pi$ of her choice to the Isotonic Mechanism?

**Theorem 1** *The expected utility is maximized when the Isotonic Mechanism is provided with the true ranking $\pi^\star$. That is, the author's optimal strategy is to* ***honestly*** *report the true ranking*

- Intuition: inconsistent ranking leads to averaging that leads to loss

## Improved utility

Are the adjusted scores more accurate than the raw scores? Yes!

**Theorem 2** *The Isotonic Mechanism improves the estimation accuracy of the true underlying scores in the sense that*

$$\mathbb{E}\left[\sum_{i=1}^n \left(\widehat{R}_i(\pi^\star) - R_i\right)^2\right] \le \mathbb{E}\left[\sum_{i=1}^n (y_i - R_i)^2\right]$$

**Theorem 3** *Assume $z_1, \ldots, z_n$ are independent and identically distributed normal random variables $\mathcal{N}(0, \sigma^2)$. Letting both $\sigma > 0$ and $V > 0$ be fixed, we have*

$$0.4096 + o_n(1) \le \frac{\sup_{\boldsymbol{R}: V(\boldsymbol{R}) \le V} \mathbb{E}\left[\sum_{i=1}^n \left(\widehat{R}_i(\pi^\star) - R_i\right)^2\right]}{n^{\frac{1}{3}} \sigma^{\frac{4}{3}} V^{\frac{2}{3}}} \le 7.5625 + o_n(1),$$

*where both $o_n(1) \to 0$ as $n \to \infty$*

- The Isotonic Mechanism is especially preferable when the noise level $\sigma$ is large and $n$ is large

## The Isotonic Mechanism

Reviewer ratings $y_1, \ldots, y_n$ governed by $y_i = R_i + z_i$

- $R_1, R_2, \ldots, R_n$ are the true values of the $n$ papers; $z_i$ are the noise terms

The author provides a ranking $\pi$ (a permutation of $1, \ldots, n$) of the papers. The Isotonic Mechanism reports $\widehat{\boldsymbol{R}}(\pi)$ that is the optimal solution of the following *convex* program:

$$\begin{aligned} \min_{\boldsymbol{r}} \quad & \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{r}\|^2 \\ \text{s.t.} \quad & r_{\pi(1)} \ge r_{\pi(2)} \ge \cdots \ge r_{\pi(n)} \end{aligned}$$

## A (real) example

I submitted 7 papers to NeurIPS 2021, with (average ratings, my ranking): (7, 1) (accepted), (6.75, 2) (accepted, the present paper), (5, 3) (rejected), (5.5, 4) (rejected), (4.67, 5) (rejected), (6, 6) (accepted), (5.5, 7) (rejected)

Applying the Isotonic Mechanism, the adjusted ratings are

$$7,\ 6.75,\ 5.334,\ 5.334,\ 5.334,\ 5.334,\ 5.334$$

## Future work

- Non-convex utility function
- Authors value papers differently
- Authors may submit low-quality papers as a "stepping stone"
- How to combine ranking from multiple authors?
- How to incorporate reviewer quality?
- Investigate what would be the outcome if the Isotonic Mechanism were used for NeurIPS 2021

## A surprising coincidence

**NeurIPS 2021 author survey to better understand expectations of the review process**

OpenReview <noreply@openreview.net>
Tue 6/1/2021 3:37 PM
To: Su, Weijie <suw@wharton.upenn.edu>

Dear Weijie,

We are inviting all authors of NeurIPS submissions to fill out a very short (<5 minutes) survey between now and June 11, 2021. The purpose of this survey is to evaluate how well authors' expectations and perceptions of the review process agree with reviewing outcomes.

Please visit your Author Console in OpenReview (after logging in) and follow instructions at the top of the page:
https://openreview.net/group?id=NeurIPS.cc/2021/Conference/Authors

All authors are asked to estimate the probability that each of their papers will be accepted in the NeurIPS 2021 review process. Authors who submitted more than one paper will additionally be asked a second question: to rank their papers in terms of their own perception of the papers' scientific contributions to the NeurIPS community.

Privacy and confidentiality are of utmost importance to us, and hence we have taken strict precautions to preserve them.

Q: Besides the OpenReview administrators, who can see your data during the review process? A: Only you, so your responses cannot affect the review process in any way.

Q: Besides the OpenReview administrators, who can see your data after the review process? A: Only you, the Workflow Chair, and the NeurIPS 2021 Program Chairs (to perform statistical analysis). Your responses will never be visible to your co-authors, reviewers, area chairs, or senior area chairs.

## Extensions

**Ranking in a block form.** The author only knows partial information of the true ranking: $\boldsymbol{R}_{I_1^\star} \ge \boldsymbol{R}_{I_2^\star} \ge \cdots \ge \boldsymbol{R}_{I_m^\star}$, but the ranking within each block is completely unknown to the owner. The Isotonic Mechanism in a block form:

$$\begin{aligned} \min_{\boldsymbol{r}} \quad & \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{r}\|^2 \\ \text{s.t.} \quad & \boldsymbol{r}_{I_1} \ge \boldsymbol{r}_{I_2} \ge \cdots \ge \boldsymbol{r}_{I_m} \end{aligned}$$

**Robustness to inconsistencies.** The owner might give a ranking that is not consistent with the true values. Imagine that the owner is choosing between two rankings $\pi_1, \pi_2$. Former is more faithful with respect to $\boldsymbol{R}$ than the latter in the sense that

$$\pi_1 \circ \boldsymbol{R} \succeq \pi_2 \circ \boldsymbol{R}$$

**Non-separable utility functions.** The main theorem continues to hold for Schur-convex utility functions, which include the class of convex functions. When $f$ is differentiable and symmetric, then $f$ is Schur-convex if and only if for all $\boldsymbol{r} \equiv (r_1, \ldots, r_n)$

$$(r_i - r_j)\left(\frac{\partial f(\boldsymbol{r})}{\partial r_i} - \frac{\partial f(\boldsymbol{r})}{\partial r_j}\right) \ge 0$$