**Explosion of ML/AI conferences**

![Number of papers over years](image)

**Trouble with peer review**

Inconsistency in Conference Peer Review: Revisiting the 2014 NeurIPS Experiment

Outsider Critic and Neil D. Lawrence

Google Research, New York

Computer Lab, University of Cambridge

September 20, 2023

Abstract

In this paper we revisit the 2014 NeurIPS experiment that measured inconsistency in conference peer review. We compare short, full papers and ICML to full papers. We observe that an earlier version of this paper contained a paradox of choice result. Further, after seven years passing since the experiment, we find that for accepted papers, there is no correlation between full paper and short paper length of the paper (or even in the selection of "inconsistency conferences") when examining the quality of individual reviewers.

**Assumptions**

1. Given (final) scores \( R_1, \ldots, R_n \) of \( n \) items in the possession of the owner, then the owner's utility takes the form

\[
U(R) = \sum_{i=1}^{n} U(R_i),
\]

where \( U \) is a non-decreasing convex function.

2. The owner has knowledge of the true ranking of her items. That is, the owner knows which permutation \( \pi^* \) that makes \( \pi^* \rightarrow R \) 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 \( z \sim \mathcal{N}(0, \sigma^2) \) for any permutation \( \pi_1, \ldots, \pi_n \).

**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 \( \pi^* \). That is, the author’s optimal strategy is to honestly report the true ranking.

- **Intuition**: inconsistent ranking leads to averaging that leads to less

**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), (5.7, 7) (rejected)

Applying the Isotonic Mechanism, the adjusted ratings are:

\[
\begin{align*}
7, & 6.75, 5.334, 5.334, 5.334, 5.334, 5.334 \\
\end{align*}
\]

**Improvised 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

\[
E \left[ \sum_{i=1}^{n} \left( \hat{R}_i - R_i \right)^2 \right] \leq E \left[ \sum_{i=1}^{n} (y_i - R_i)^2 \right]
\]

**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

\[
0.4096 + o_n(1) \leq \sup_{\pi \in \Pi} \left( \frac{E \left[ \sum_{i=1}^{n} \left( \hat{R}_i - R_i \right)^2 \right]}{n \sigma^2 V^2} \right) \leq 7.5625 + o_n(1)
\]

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**

**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 authors are to better understand expectations of the review process

**OpenReview** - openreview.net

\( \rightarrow \)

Author Acknowledgments

\( \rightarrow \)


date May 9, 2021

We are writing authors of NeurIPS submissions to fill out a very short (10 question) survey before June 1, 2021. The authors of the survey to include the best and worst aspects of the review process openly no-responding anonymous.

Please visit our Author Accreditation OpenReview to log into and follow instructions at the top of the page.

http://openreview.net/reviewping/NeurIPS2021/AuthorAccreditation

At author's discretion, we use this data to improve the author's experience. We will also do not publish any personal information, which will be kept anonymous. We will not use personal information for any commercial purpose.

Q: Besides the NeurIPS administrators, who can see your data after the survey process? Only \( \rightarrow \), if you are more specific answer.

Q: Besides the NeurIPS administrators, who can see your data after the survey process? Only \( \rightarrow \), if you are more specific answer.

Q: Besides the NeurIPS administrators, who can see your data after the survey process? Only \( \rightarrow \), if you are more specific answer.

---

**Understanding**

**Conclusion**

**Extensions**

**Ranking in a block form**

The author only knows partial information of the true ranking \( R_{11} \geq R_{12} \geq \cdots \geq R_{1n} \), but the ranking within each block is completely unknown to the owner.

The Isotonic Mechanism in a block form:

\[
\begin{align*}
\min_{\pi} & \quad \frac{1}{2} \| y - r \|^2 \\
\text{s.t.} & \quad r_{\pi(1)} \geq r_{\pi(2)} \geq \cdots \geq r_{\pi(n)} \\
\end{align*}
\]

**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 \( R \) than the latter in the sense that

\[
\pi_1 \triangleright R \geq \pi_2 \triangleright R
\]

**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 \( r \equiv (r_1, \ldots, r_n) \)

\[
(r_i - r_j) \frac{\partial f(r)}{\partial r_i} - \frac{\partial f(r)}{\partial r_j} \geq 0
\]

**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), (5.7, 7) (rejected)

Applying the Isotonic Mechanism, the adjusted ratings are:

\[
\begin{align*}
7, & 6.75, 5.334, 5.334, 5.334, 5.334, 5.334 \\
\end{align*}
\]

**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

**The Isotonic Mechanism**

- The Isotonic Mechanism improves the estimation accuracy of the true underlying scores in the sense that

\[
\begin{align*}
E \left[ \sum_{i=1}^{n} \left( \hat{R}_i - R_i \right)^2 \right] \leq E \left[ \sum_{i=1}^{n} (y_i - R_i)^2 \right]
\end{align*}
\]

**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

\[
0.4096 + o_n(1) \leq \sup_{\pi \in \Pi} \left( \frac{E \left[ \sum_{i=1}^{n} \left( \hat{R}_i - R_i \right)^2 \right]}{n \sigma^2 V^2} \right) \leq 7.5625 + o_n(1)
\]

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