最后一题（提案）：System D 共识–分歧双通道赛制（更公平，也更刺激）

题目：

> Propose another system using fan votes and judge scores each week that you believe is more "fair" (or "better" in some other way such as making the show more exciting for the fans). Provide support for why your approach should be adopted by the show producers.

本提案给出一个**制度化、参数少、透明可解释**的新赛制：

- 用 **共识（consensus）** 决定"谁应该被淘汰"（公平）
- 用 **分歧（divergence）** 制造"可控的争议与话题"（刺激），并避免"一周定生死"的误伤

---

1. 核心动机：为什么要引入"共识"和"分歧"？

DWTS 的结构性矛盾是：

- **评委分数**代表专业/技术评价
- **观众票**代表人气/故事线/参与感

前面题目（以及我们做的 judges vs fans 对照分析）已经显示：

- judges 与 fans 的偏好**并不总一致**；
- 很多"争议淘汰"来自这种不一致。

因此，新赛制的合理目标不是"强行把两者揉成一个数字"，而是：

1. 公平：淘汰尽量发生在"双方都认为差"的选手（低共识）
2. 刺激：对"分歧很大"的选手暂时留一手，让他/她下一周用表现自证（节目叙事更好）

---

2. System D 的定义（只用每周 judge + fan）

对每个赛季的每一周 t，对仍在比赛的选手集合 $\mathcal{A}_t$：

2.1 周内百分位（消除尺度差异）

我们不直接使用原始分数/票的绝对值，而把它们转换为周内百分位（0 到 1）：

- $P_{i,t}^{J}$：选手 i 的评委信号百分位（由 `judge_percent` 排序得到）
- $P_{i,t}^{F}$：选手 i 的观众信号百分位（由 `fan_vote_normalized` 排序得到）

更具体地，若一周有 n 位选手（剔除缺失后），我们把某选手的名次 $\mathrm{rank} \in \{1, \ldots, n\}$ 映射为

$$P = \frac{\mathrm{rank} - 1}{n - 1} \in [0, 1].$$

（并对并列名次取平均名次。）

这样做的好处：

- 评委分数尺度每周可能不同，但百分位可比
- 观众票若存在系统偏差，只要"相对顺序"仍有意义，百分位就更鲁棒

原理性解释（为什么用百分位更"制度公平"）：

- 百分位只依赖"周内相对位置"，对任何严格单调变换都不敏感（例如把票数做缩放、做对数、或不同周投票总量不同）。
- 这使 System D 更像"赛制规则"，而不是"依赖某个估计模型绝对值精度"的打分器：只要当周排序信息可信，规则就稳定。

2.2 两个关键量：共识 C 与分歧 D

定义：

$$C_{i,t} = \frac{P_{i,t}^{J} + P_{i,t}^{F}}{2}$$

$$D_{i,t} = |P_{i,t}^{J} - P_{i,t}^{F}|$$

解释：

- C 高：评委和观众都支持（综合认可）
- C 低：两边都不支持（"共识底部"）
- D 高：评委与观众严重分歧（"争议选手/争议事件"）

几何/机制解释（建议写进论文，读者会更信服）：

- 把每个选手视为单位正方形 $[0, 1]^2$ 上的点 (P^J, P^F)。
- 共识 $C = (P^J + P^F)/2$ 是沿着方向向量 $(1, 1)$ 的"投影强度"：

- C 的等值线是反对角线 $P^J + P^F = \text{常数}$；越靠近左下角（两者都低）C 越小。
- 分歧 $D = |P^J - P^F|$ 是到对角线 $P^J = P^F$ 的偏离程度：
  - D 越大，点离对角线越远，表示"评委与观众判断越不一致"。
  - 在欧氏几何下，到对角线的垂直距离是 $D/\sqrt{2}$，因此 D 可以看作"分歧强度"的线性刻度。

---

3. 每周淘汰规则（可直接给制作方）

设定一个很小的参数：风险池大小 m（推荐 $m = 4$）。

每周执行：

1. **风险池（Risk Pool）**：按 C 从低到高，取 bottom-m 进入风险池。
2. **争议保护（Controversy Save）**：在风险池中，找到分歧最大的选手（D 最大者），当周保护，不被淘汰。
3. **淘汰**：在剩余的风险池选手中，淘汰 C 最低者。

> 节目解释话术也非常自然：
> "你们两边都不支持的人更应该走；
> 但如果分歧巨大，我们给一次'争议保护'，下周再用表现说话。"

规则的"数学等价说法"（帮助解释公平性与可视化）：

- 风险池 bottom-m 本质上对应一个周内阈值 C_{cut}（即风险池里最大的 C）。
- 因为 $C = (P^J + P^F)/2$，所以风险池大致落在半平面

$$P^J + P^F \le 2C_{\text{cut}},$$

  也就是图上靠近左下角的一块区域。
- 争议保护选的是该区域内 D 最大者，即离对角线 $P^J = P^F$ 最远者（"最有争议"）。

---

4. 为什么它更公平？（可写进论文的论证点）

4.1 "淘汰发生在低共识"更可辩护

传统的固定加权/固定排名融合，有时会出现：

- 评委很高、观众很低（或相反）的人被直接淘汰 → 容易引发"robbed"争议。

System D 把淘汰候选限制在 bottom-m 的低共识（低综合支持）区间：

- 这意味着淘汰更像"整体支持不足"的结果，而不是"一方极端影响"的结果。

4.2 对数据不确定性更鲁棒（不依赖 fan vote 的绝对精度）

System D 的输入是周内百分位 P^J, P^F，而不是票的绝对值。

- 即使 fan vote 的估计存在偏差，只要"相对高低"在周内仍有意义，该规则仍能工作。
- 与依赖模型方差/置信区间的规则相比，它更像"制度公平"，对估计模型更不敏感。

需要强调的边界：

- **鲁棒的是尺度/总量变化**（例如投票总量波动、分数尺度变化），而不是"排序完全错误"。
- 若 fan vote 的周内排序误差很大，则任何基于排序/百分位的规则都会受影响；System D 的优势在于不要求票的绝对值精确。

---

5. 为什么它更刺激？（制作方会喜欢的点）

5.1 把争议量化成节目资产

$D_{i,t}$ 直接是"争议强度"。节目可以公开：

- "本周分歧最大选手"
- "本周争议保护触发"

这比"暗箱调整权重"更透明，也更容易引发讨论。

5.2 争议保护不是'免死金牌'，而是'延迟审判'

保护发生在风险池内，且保护对象本身也处于 bottom-m（共识偏低），因此：

- 它不会让强者无故受罚
- 也不会让弱者无限苟活
- 反而形成更强叙事：下周必须自证，否则更容易被淘汰

---

6. 数据支持（我们用历史数据做了可复现对照）

可复现脚本：

- `proposed_system_outputs/proposed_system_D.py`

输出：

- `tables/systemD_weekly.csv` ：每周风险池、保护对象、预测淘汰
- `tables/systemD_summary.csv` ：汇总统计
- `figures/fig_systemD_example_week.png` ：示例周的 P^J vs P^F 散点图（标出风险池/保护/淘汰）
- `figures/fig_systemD_divergence_distribution.png` ：分歧 D 的分布对比（基线 vs System D）
- `figures/fig_systemD_geometry_example_week.png` ：示例周的"决策几何图"（画出风险边界线、对角线、争议保护的分歧带）
- `figures/fig_systemD_divergence_ecdf.png` ：分歧 D 的 ECDF 对照图（比直方图更便于比较整体偏移）

6.1 可引用的关键统计（来自 systemD_summary.csv）

- 分析到的淘汰周数：264
- System D 与 50/50 基线在 42.0% 的周上给出不同淘汰人选（disagree rate \approx 0.420）
- System D 的"争议保护"在 34.8% 的周上**恰好保护了**基线规则会淘汰的人（protected == baseline eliminated rate \approx 0.348）

这两点在论文里可解释为：

- 我们的规则并非"换汤不换药"，它系统性改变了争议周的命运；
- 它把"潜在争议淘汰"更频繁地转化为"下一周对决"的节目结构。

> 注意：这里不强调"命中历史淘汰的准确率"，避免循环论证。
> 我们强调的是：规则满足的机制性性质（低共识淘汰、分歧保护）以及可视化证据。

---

7. 图的作用与论文写法（可直接粘贴）

图 1：System D 示例周（如何工作）

文件： `fig_systemD_example_week.png`

- 横轴：评委百分位 P^J
- 纵轴：观众百分位 P^F
- 橙色：风险池；绿色：争议保护；红色：当周淘汰

论文图注模板：

> Figure X. Example week under System D. Contestants are plotted by judges percentile P^J and fans percentile P^F. The risk pool is defined by the bottom-m consensus score $C = (P^J + P^F)/2$. The most divergent contestant within the risk pool (max $D = |P^J - P^F|$) receives a controversy save, and the remaining lowest-consensus contestant is eliminated.

图 2：分歧 D 的分布（为什么能减少"误伤型淘汰"）

文件： `fig_systemD_divergence_distribution.png`

- 绿色：被争议保护的选手的分歧 D 往往更大
- 对比基线淘汰对象与 System D 淘汰对象的 D 分布，可展示：System D 更倾向于"留下争议更大者"，将争议转化为下一周对决

论文图注模板：

> Figure Y. Divergence distribution comparison. System D systematically protects contestants with large judge–fan disagreement (high D) within the low-consensus risk pool, shifting controversy cases from immediate elimination to next-week resolution.

图 3（高级）：决策几何（为什么风险池 = 一条边界线以下）

文件： `fig_systemD_geometry_example_week.png`

- 风险边界线：$P^J + P^F = 2C_{\text{cut}}$（底部共识阈值）
- 灰色虚线：$P^J = P^F$（无分歧对角线）
- 绿色分歧带：$|P^J - P^F| = D_{\text{protected}}$（当周被保护者的分歧强度）

论文图注模板：

> Figure Z. Decision geometry of System D in an example week. The bottom-m risk pool corresponds to the low-consensus half-plane $P^J + P^F \leq 2C_{cut}$. The controversy save selects the contestant with the largest deviation from the diagonal $P^J = P^F$ within this region.

图 4（高级）：ECDF 对照（整体比较"谁更倾向于淘汰高分歧者"）

文件： `fig_systemD_divergence_ecdf.png`

- ECDF（经验分布函数）能直观看到一组样本整体是否"更偏向小/大"
- 若 System D 的淘汰 ECDF 曲线整体更靠左，说明它更少淘汰高分歧者
- 保护者（绿色）往往更靠右，体现"争议保护"确实把高分歧选手留了下来

论文图注模板：

> Figure W. ECDF of divergence D. Compared to the 50/50 baseline, System D shifts eliminations toward smaller disagreement levels while systematically protecting contestants with higher D in the low-consensus region.

---

8. 给制作方的采纳理由（Executive Pitch）

- **更公平**：淘汰更集中在低共识选手，减少"一方极端信号导致的误淘汰"。
- **更刺激**：争议保护把分歧变成节目资产，形成可控的"下一周对决"叙事。
- **更透明**：只需要每周 judge 与 fan 两个信号，规则清晰、参数少（仅 m），容易对外解释。
- **更鲁棒**：使用周内百分位，不依赖票的绝对精度；即使投票体系/估计模型变化，规则仍稳定。

---

9. 实施细节与边界情况（让规则"可落地"）

9.1 参数 m 怎么选？

- 建议默认 $m = 4$：能形成一个足够小的"风险叙事池"，又能稳定触发争议保护。
- 也可按当周剩余人数 n 自适应：$m = \min(4, n)$（脚本实现即如此），避免人数很少时规则失效。

9.2 并列与缺失

- 并列（ties）：百分位按平均名次处理，使规则在并列情况下仍连续、可解释。
- 缺失：若某周仅剩 1 人有有效信号，则百分位固定为 0.5（脚本中做了防御性处理）。

9.3 规则口径（制作方台词）

可以把 C 公布为"综合支持度排名"，把 D 公布为"分歧指数"，并在风险池宣布：

- "综合支持度最低的人最危险（公平）"
- "分歧指数最高的人得到一次争议保护（刺激且透明）"

---

10. 题目要求对照（确认 Z 题已完成）

- **提出一个每周制度**：第 3 节给出了逐周执行的淘汰规则（风险池→争议保护→淘汰）。
- **只使用 fan votes + judge scores**：第 2 节定义仅依赖 P^J, P^F（由 `judge_percent` 与 `fan_vote_normalized` 得到）。

- **更公平/更刺激的理由**：第 4–5 节给出机制性论证；第 8–9 节给出制作方可落地话术与采纳理由。
- **提供支持（support）**：第 6–7 节给出可复现实证输出与四张图（含两张高级图）用于论文支撑。