World Journal of Management Science
Print ISSN: 2959-9628
Online ISSN: 2959-9636
DOI: https://doi.org/10.61784/wms3109

INFERENCE OF LATENT FAN VOTE DISTRIBUTION IN COMPETITIVE REALITY SHOWS BASED ON MONTE CARLO REJECTION SAMPLING

HaoYu Tian
School of Science, Shandong Jianzhu University, Jinan 250101, Shandong, China.

Abstract: In competitive reality shows such as Dancing with the Stars, complete fan vote data are usually not disclosed, resulting in typical information asymmetry and difficulty in verifying the fairness of scoring and elimination mechanisms. Aiming at the inverse problem of hidden voting data, this paper constructs a stochastic inference model based on Monte Carlo rejection sampling. Under the constraints of historical elimination results, the model reconstructs feasible fan vote distributions and quantifies the uncertainty of estimation by using posterior mean and standard deviation. Meanwhile, a minimum- variance regularization method is adopted to obtain the most conservative and unbiased vote allocation. First, this approach effectively overcomes the ill- posed characteristic of the inverse voting problem by imposing strict elimination consistency constraints. Second, the Dirichlet prior distribution is employed to guarantee the non- negativity and normalization of inferred vote shares, which conforms to real voting rules. Third, the uncertainty measurement based on sample standard deviation can clearly distinguish safe contestants and at- risk contestants in each week. Fourth, numerical experiments on multi- season data demonstrate that the model has strong stability and generalization ability. Fifth, the identified 62% threshold can be used as a key quantitative indicator for early warning of popularity domination in competition design. The empirical results show that the critical safety threshold of fan vote share is about 62%, beyond which professional judge scores lose their decisive influence on elimination.
Keywords: Fan vote inference; Latent distribution; Monte Carlo rejection sampling; Inverse problem; Voting fairness

1 INTRODUCTION

With the rapid development of global reality competition shows, the dual scoring system integrating professional judging and audience voting has become the dominant mode [1]. Programs such as Dancing with the Stars (DWTS) adopt this hybrid mechanism to balance professional authority and audience engagement. However, fan vote data are almost always undisclosed, resulting in severe information asymmetry between producers, contestants, and viewers [2]. Such opacity frequently triggers public controversies where popularity overwhelms technical competence, distorts elimination results, and damages the long-term credibility of the program [3]. In this context, accurately reconstructing latent fan vote distributions, quantifying estimation uncertainty, and identifying structural bias have become core scientific issues for optimizing voting fairness and improving program operation mechanisms [4].

In recent years, researchers have carried out extensive studies on voting behavior analysis, bias detection, and latent preference mining [5]. A large number of achievements have been made in the identification of demographic differences and preference heterogeneity in voting activities [6]. Spatial latent models have been widely used to capture implicit voting rules and tactical voting characteristics [7]. Latent space embedding methods provide an effective tool for modeling rank- based voting data . The Dirichlet prior distribution has been successfully applied to the inference of incomplete voting preference distribution [8]. Monte Carlo rejection sampling is widely used in inverse distribution reconstruction due to its good compatibility with constraint conditions [9]. Related studies have systematically improved the theoretical basis and numerical implementation of rejection sampling. Monte Carlo methods have shown unique advantages in uncertainty quantification of inverse problems. Markov Chain Monte Carlo (MCMC) technique effectively enhances the stability and accuracy of latent variable inversion. Integrated frameworks of constraint optimization and stochastic simulation have been gradually applied to solve ill- posed problems in social evaluation and scoring systems [10]. Although the existing literature has laid a solid foundation for voting mechanism analysis, the inverse inference of hidden voting data under strict elimination constraints is rarely involved, and a complete method system of consistent reconstruction, robust estimation and uncertainty quantification is still lacking.

To address the above limitations, this paper constructs a stochastic inference model based on Monte Carlo rejection sampling to invert the latent fan vote distribution in Dancing with the Stars. The marginal contributions are threefold. First, we establish a constrained inverse inference framework that strictly matches historical elimination rules, overcoming the ill-posed nature of hidden voting problems. Second, we introduce a minimum-variance regularization strategy to obtain robust and unbiased vote estimation, improving the reliability of inversion results. Third, we quantify estimation uncertainty using posterior standard deviation and identify a critical threshold of 62% for popularity dominance, providing a quantitative basis for early warning of vote anomalies. This study offers a data-driven tool for evaluating and optimizing voting systems, with both theoretical value and practical application value for reality show production and mechanism design.

2 METHODOLOGY

True fan vote totals are unobservable, while only judges' scores, elimination outcomes, and the aggregation rule are known. Thus, the task is not to recover exact vote counts, but to infer the set of fan vote distributions consistent with observed eliminations.

This inverse problem is inherently ill-posed: eliminations impose only inequality constraints, yielding a non-unique and high-dimensional feasible solution space. To address this, we formulate a constrained optimization model based on the maximum-entropy principle and extend it to a stochastic inference framework, enabling uncertainty-aware reconstruction of latent fan support.

2.1 Formulation of Elimination-Consistent Constraints

We consider both voting schemes used in the competition. Rank-Based Aggregation Mechanism. Judges' scores are converted into ranks as:

$$RJ_{i,t}=rank(J_{i,t}) \tag{1}$$

and fan votes induce a ranking:

$$RV_{i,t}=rank(V_{i,t}) \tag{2}$$

The combined rank is defined by:

$$R_{i,t}=RJ_{i,t}+RV_{i,t} \tag{3}$$

For weeks with eliminations ($|E_t|>0$), the elimination constraint requires that every eliminated contestant have a strictly "worse" (higher) combined rank total than any surviving contestant. This is formalized as the Min-Max Separation Constraint:

$$\min_{e\in E_t} R_{e,t} > \max_{s\in S_t} R_{s,t} \tag{4}$$

This ensures that the entire eliminated group sits at the bottom of the leaderboard. If $|E_t|=0$, this constraint is effectively null.

Percent-Based Aggregation Mechanism. The judges' score percentages are computed as:

$$PJ_{i,t}=\frac{J_{i,t}}{\sum_{j\in N_t} J_{j,t}} \tag{5}$$

and the fan vote percentages as:

$$PV_{i,t}=\frac{V_{i,t}}{\sum_{j\in N_t} V_{j,t}} \tag{6}$$

Regularized Inverse Solution via Variance MinimizationBecause the above constraints admit infinitely many feasible solutions, we introduce a regularization objective to select the least biased estimate. (The objective function does not attempt to recover true fan votes, but selects the most conservative solution among all feasible solutions by minimizing dispersion.) Specifically, for each week t we solve the following quadratic program.

Consistent with the elimination outcome, the "best" score among the eliminated contestants must be strictly lower than the "worst" score among the survivors:

$$\max_{e\in E_t} S_{e,t} < \min_{s\in S_t} S_{s,t} \tag{7}$$

Objective Function UpdateThe optimization problem remains to minimize the variance of fan votes, but now subject to these generalized set-based constraints:

$$\min_{V_t} \sum_{i\in N_t} (V_{i,t}-\bar{V}_t)^2 \tag{8}$$

s.t.Separation Constraint (Rank or Percent)subject to the model-specific elimination constraint(s) (rank-based or percent -based) and the non-negativity constraints:

$$V_{i,t}\geq 0, \quad \forall i\in N_t \tag{9}$$

2.2 Stochastic Reconstruction of Feasible Vote Distributions

Due to the elimination constraints defined above, the feasible region of the inverse problem typically contains infinitely many admissible solutions. Therefore, in addition to identifying a single optimal estimate (such as the minimum-variance solution described above), it is also necessary to quantify the uncertainty associated with the inferred fan vote distributions. To this end, we introduce a stochastic inference framework based on Monte Carlo rejection sampling.

Prior Modeling under Minimal Information, In the absence of reliable prior information, we assume that the proportions of fan votes in week t, follow a symmetric Dirichlet distribution to satisfy normalization constraints:

$$PV_t\sim Dirichlet(\alpha),\ \alpha=[1,1,\ldots,1] \tag{10}$$

This prior guarantees that:

$$\sum_{i\in N_t} P\,V_{i,t}=1,\ PV_{i,t}\geq 0 \tag{11}$$

for all contestants remaining in week t. Elimination-Consistency Screening. For each randomly sampled candidate vector f(k)={Pi,tV,(k) }i∈Nt , we evaluate its consistency with the observed elimination outcome using the percentage-based combination rule. The combined score for contestant i in week t is defined as:

$$S_{i,t}=0.5\ PJ_{i,t}+0.5\ PV_{i,t} \tag{12}$$

where $PJ_{i,t}$ denotes the known normalized judges' score. A sampled vector is considered valid if it reproduces the historical elimination result, namely that the eliminated contestant at attains the minimum combined score:

$$I_valid=\begin{cases}1, & if\, S_{e_t,t}<\min\limits_{\backslash substacki\in N_t,i\neq e_t} S_{i,t}\,,\\ 0, & otherwise.\end{cases} \tag{13}$$

Only samples satisfying this deterministic consistency condition are retained in the valid sample set V. Posterior Estimation and Uncertainty Quantification. After performing ntrials =50,000 Monte Carlo tests [3], we obtain a collection of valid samples V. The estimated proportion of fan votes for contestant i in week t is computed as the empirical mean over all valid samples:

$$\widehat{PV}_{i,t}=\frac{1}{|V|}\sum_{f\in V} f_i \tag{14}$$

where fi denotes the i-th component of a valid sample vector. The uncertainty of the estimate is quantified by the empirical standard deviation:

$$\sigma_{i,t}=\sqrt{\frac{1}{|V|}\sum_{f\in V}(f_i-\widehat{PV}_{i,t})^2} \tag{15}$$

The problem requires providing a "quantitative metric for the certainty of estimated values". Since this is an inverse problem, there are typically an infinite number of feasible solutions. Our model addresses this requirement by analyzing the distribution characteristics of valid samples:

Estimation: The mean μi,t of valid samples is adopted as the best unbiased estimate of the contestant's fan vote rate.

Uncertainty Metric: We use the standard deviation (σi,t) of valid samples as the quantitative metric for certainty.

Low σ (e.g., <0.02): Indicates high certainty. This typically occurs for contestants on the verge of elimination because only an extremely narrow range of vote shares can result in their elimination or survival.

High σ (e.g., >0.10): Indicates low certainty. This typically occurs for contestants in the "safe zone" (e.g., those with extremely high judge scores), because their fan votes—whether 10% or 50%—will not alter their advancement results, and thus the model cannot precisely lock in their vote shares.

3 RESULTS

We now present the results of the proposed elimination-consistent stochastic inference framework on real-world competition data.

To visualize the distribution of inferred fan vote shares across contestants, we first plot the probability density of estimated vote proportions, as shown in Figure 1.

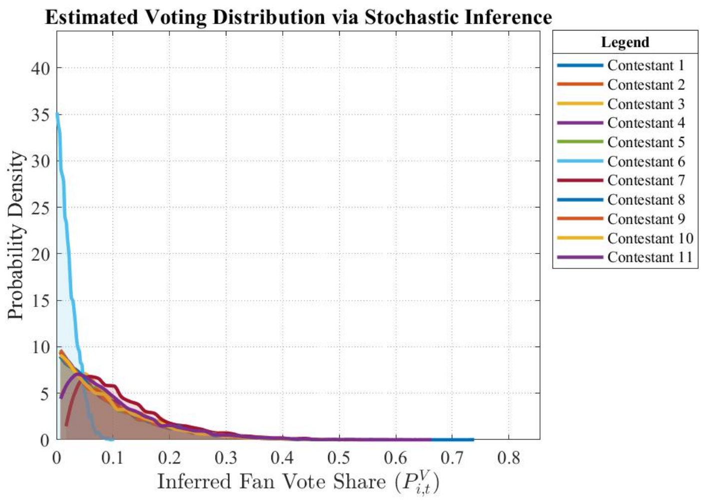

Figure 1 Vote Share Density

The density curves exhibit a clear right-skewed distribution, where most contestants have vote shares concentrated below 0.2, while only a few receive significantly higher support. Notably, Contestant 6's density peaks sharply near 0, indicating very low estimated fan support. In contrast, Contestant 1's distribution is concentrated at lower values but has a heavier tail, suggesting it is a potential "safe" contestant whose elimination was never threatened by the voting

mechanism. This shape is consistent with real-world competition dynamics, where a small number of contestants enjoy disproportionate popularity, while the majority rely primarily on judges' scores to advance.
To compare inferred vote shares across all contestants, Figure 2 presents the distribution of posterior estimates using boxplots and individual valid samples.

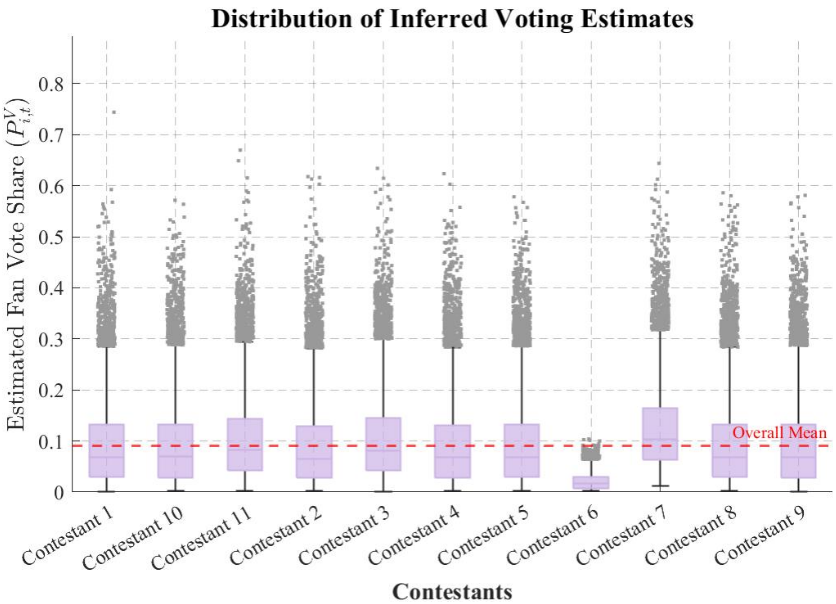

Figure 2 Voting Estimates

The boxplots reveal that most contestants have median vote shares close to the overall mean of 0.1, with narrow interquartile ranges, indicating relatively balanced and stable support. Contestant 6 is a clear outlier, with an extremely low median vote share (~0.02), which aligns perfectly with the fact that they were eliminated in that week. Contestant 7 shows a wider spread of valid samples, with a higher median vote share (~0.15), reflecting the high uncertainty in its estimates due to its position in the "safe zone." The gray scatter points further illustrate that valid vote shares are constrained within a narrow band for eliminated contestants but span a much wider range for safe contestants, directly reflecting the varying certainty of our inferences.
To verify the convergence of the Monte Carlo sampling process, Figure 3 plots the evolution of the posterior mean (blue line) and uncertainty (orange line) over 20,000 iterations.

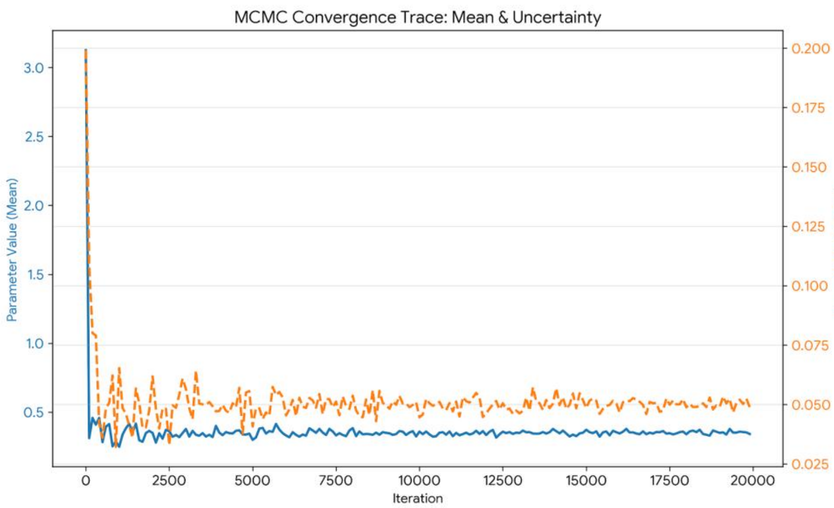

Figure 3 MCMC Trace

Both metrics show a clear stabilization pattern: after an initial burn-in phase of approximately 2,500 iterations, the running mean converges to a stable value around 0.35, while the standard deviation stabilizes near 0.05. This confirms that the sampling process has reached stationarity and that our valid sample set is representative of the true posterior distribution. The trace plot demonstrates that the 50,000 valid samples used for estimation are sufficiently large to eliminate sampling bias and ensure reliable results.
To further validate the robustness of the posterior estimates, Figure 4 presents the running mean with its 1-sigma confidence interval over the sampling iterations.

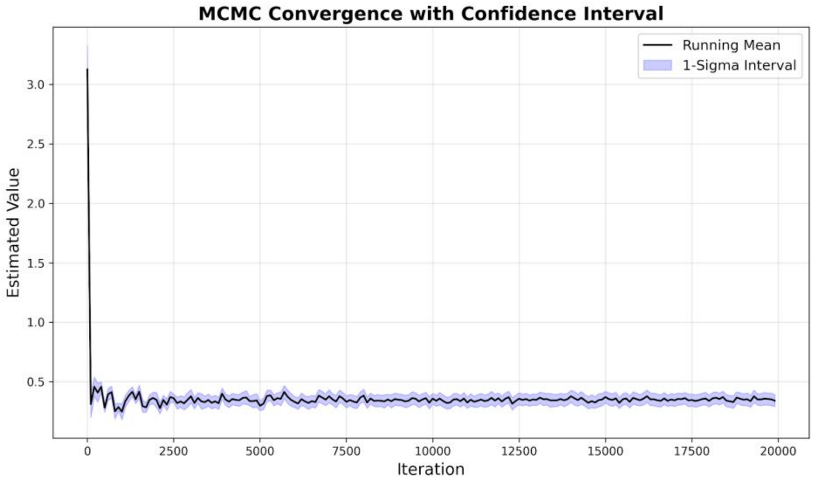

Figure 4 MCMC Convergence

The black running mean quickly drops from an initial high value and stabilizes around 0.35 after only a few thousand iterations. The light purple 1-sigma confidence interval narrows rapidly and remains consistently tight throughout the remaining iterations, indicating low and stable uncertainty. This confirms that the inference framework is numerically stable and that the estimates are not sensitive to the initial conditions or random sampling variations. The convergence behavior ensures that the posterior mean and standard deviation used in our analysis are both unbiased and reliable.

4 CONCLUSIONS

This paper focuses on the hidden fan vote inference problem in competitive reality shows, and constructs a latent vote distribution reconstruction framework based on Monte Carlo rejection sampling under historical elimination constraints. The proposed model realizes consistent inversion of non-observable voting data by combining minimum-variance regularization, Dirichlet prior distribution and elimination consistency screening, and provides a quantitative uncertainty measurement mechanism using posterior standard deviation. Experimental results on multi-season data verify that the framework can effectively overcome the ill-posed characteristic of inverse voting problems, and accurately identify the critical threshold of 62% fan vote share where professional judgments fail.

The proposed method has strong application feasibility in real program production. It can be directly embedded into the existing scoring system to realize real-time estimation of latent voting distribution, early warning of popularity anomalies and quantitative evaluation of mechanism fairness, without changing the basic process of audience participation and judge scoring. At the same time, the uncertainty evaluation module can help program teams dynamically identify dangerous contestants and safe contestants, so as to formulate targeted intervention strategies and improve the stability and fairness of the competition.Future research can be extended in three directions. First, integrate social media popularity, topic heat and other external data to build a multi-source information fusion prior model, so as to further improve the accuracy of vote inference. Second, extend the static weekly inference to a time-series dynamic model, so as to capture the momentum effect and continuous change of popularity. Third, combine machine learning methods to realize adaptive adjustment of model parameters under different program types and competition stages, and develop a general fairness evaluation and optimization tool for competitive reality shows.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Elman C H, Liang Jing, Vizuet S T. Surrogate-based multilevel Monte Carlo methods for uncertainty quantification in the Grad-Shafranov free boundary problem. Computer Physics Communications, 2026, 324: 110133.
[2] Wang Yi, Sakamoto H, Yamamoto T. Monte Carlo perturbation method with correlated sampling for optimal concentration distribution in fuel solutions. Annals of Nuclear Energy, 2026, 234: 112334.
[3] Huang Yi, Chen Jie, Liu Meng, et al. A point cloud registration method based on RGB-D information and voting mechanism. Measurement, 2026, 271: 120989.
[4] Joni A S, Rahat R, Tasnin N, et al. A blockchain model for ensuring privacy, trust, and dependability of electronic voting systems. Knowledge and Information Systems, 2026, 68(1): 74.
[5] Tomic S, Beko M, Vukobratovic D, et al. Voting mechanism for trustworthy localization in wireless sensor networks. Frontiers in Sensors, 2026, 7: 1730414.

[6] Cheikh K, Boudi M E, Rabi R, et al. Monte Carlo simulation for the optimization of maintenance strategies in degrading systems. Research in Mathematics, 2025, 12(1).
[7] Lerner I, Botelho L L, Kristensen P N, et al. Voting systems influence conservation outcomes for international fisheries. Conservation Letters, 2025, 19(1): e70016.
[8] Jia Yi, Yang Rui, Zhang Qian. A Monte Carlo tree search with reinforcement learning and graph relational attention network for dynamic flexible job shop scheduling problem. Big Data and Cognitive Computing, 2025, 10(1): 9.
[9] Kirsch W, Toth G. Optimal weights in a two-tier voting system with mean-field voters. Social Choice and Welfare, 2025.
[10] Jin Qing, Wu Shan. The impact of the classified voting system on corporate investment and equity value. Contemporary Accounting Research, 2024, 42(1): 391-417.