Dean&Francis

ISSN 2959-6130

Revisiting the Secretary Problem under Noisy Evaluation: Robustness and Performance Analysis

Zihang Xu

Department of Statistics, University of Illinois at Urbana Champaign, Champaign, 61820, United States of America

Corresponding author: zihangx5@illinois.edu

Abstract:

This paper investigates how the existence of different degrees of noise in the real world affects the final results of the classic secretarial problem by constructing a noise-inclusive secretarial strategy model. Rather than assuming that decision-makers know the precise rankings of all candidates, this study has set up a situation where the observed rankings of each candidate by the decision-maker are subject to noise and, as a result, deviate from the optimal probability-maximizing outcome of the classical secretarial problem. To study the effect of this uncertainty, the noisy secretarial strategy model also uses Monte Carlo simulation. A set of indices for the new model environment also includes the success rate, average selection rank and regret value. Based on the above results, a relatively small amount of noise will still significantly reduce the final decision. With an increase in initial noise, the proportion of the selected best candidate decreases significantly. Based on the above study, the strength of the classic secretarial problem is highly sensitive to the initial observations. Although this strategy has almost the same success rate in a low-noise environment, it is still not feasible for real-world application.

Keywords: Secretary Problem, Optimal Stopping, Noisy Observations, Sequential Decision-Making, Monte Carlo Simulation.

1. Introduction

Sequential decision making problems are basic problems in many real-world processes in which the alternatives are offered sequentially in time, and decisions must be made irrevocably. Examples of such include online auctions, in which the bids are dynamically updated as they are made, and hiring, in which the candidates are interviewed in sequence. In such settings, decision makers have to take decisions based on incomplete and evolving information, and cannot anticipate future observations [1].

One of the classic ways of modelling such problems is "Secretary Problem" which is to maximize the chance of picking the best candidate in a random order. It is best strategy is to have approximately 1/e of the candidates eliminated and then hire the first candidate better than all those observed. This approach gives the decision-maker a $1/e$ chance of success, and its success hinges on the decision-maker's ability to compare candidates by the complete observable relative rankings [2,3]. In practical situations, however, it may be hard to maintain these assumptions.

In practice the qualities of a candidate are not directly visible, and decisions are frequently made on noisy or imperfect signals (e.g. the interview, or prediction of valuation). More important, however, is that this noise does not simply add noise to the voting process; it actually undermines the relative ranking of candidates, which is the key to the "secretary strategy". So it is still uncertain if the classic optimal stopping rule will be able to keep the same decision-making performance with noisy and real environments [4,5].

According to this theoretical difference, this study explores the robustness of the secretary strategy in a noisy evaluation environment. The author proposes an "estimated scores with noisy secretary-based strategy model", where the estimated score of each candidate is randomly perturbed.

Based on the above model, the author will systematically examine the effect of changing noise intensity on the final performance of the secretary strategy. A Monte Carlo simulation is used to determine how much the performance of the classical optimal stopping rule declines in the presence of noise and to identify the conditions under which the secretary problem can still be performed well.

The results of this study provide new references for the optimal-stopping strategies in various actual circumstances, such as corporate recruitment, online auctions and sports drafts. In these actual cases, there will be some uncertainty and information deficiencies.

This paper differs from most classic variants of the secretary problem, which focus either on changing the selection goals or introducing the possibility of multiple-choice selection, in the emphasis placed on the presence of noisy and imperfect information in the sequential decision-making setting. The principal novelty of this paper is to modify the secretary problem by adding an observation noise term, but keeping the classical secretary rule unchanged. By doing so, this paper focuses on the robustness-effects of uncertainty on the classical optimal stopping strategy. Moreover, this paper not only examines the probability of choosing the best candidate, but also includes average rank and regret as evaluation metrics to give a more complete evaluation of decision quality in realistic situations.

2. Related Works

Decision-making methods under uncertainty have been studied in the framework of the secretary problem. The classic secretary problem assumes that the candidates arrive in a random order with equal probability. Given the above conditions, the academic community has reached a general consensus and usually presents this in the form of a threshold rule for secretary strategies: reject the first n/e candidates, and then select the next candidate whose score exceeds all of them. Many optimal stopping theorems in probability theory and computer science have originated from this paper and are now in widespread use.

Based on the above classic secretary model, many other forms of models have been proposed in recent years to deal with scenarios where multiple candidates need to be selected. The multi-selection secretary problem allows the decision-maker to choose several candidates from the sequence of candidates. For instance, in the paper "A Multiple-Choice Secretary Algorithm with Applications to Online Auctions", the threshold strategy is improved to dynamic or hierarchical thresholds to support multi-round selection. Given that most organizations use multi-round selection processes in actual circumstances, such models are also widely employed in practice, such as hiring.

Another main direction of research is to extend another part of the secretary problem, that is, "value-based" and "weighted" approaches. These models no longer aim to maximize the probability of selecting the best candidate as their final objective, but rather to maximize the expected value of the candidates. For example, in "The Secretary Problem: Weights and Discounting", a new theoretical system is introduced that gives all candidates different inherent values and may be discounted over time. In the decision-making situations in real life that correspond to such models, the result is not simply a binary choice (i.e., success or failure); instead, all the candidates need to be considered in terms of trade-offs [4,6].

At the same time, the above models also address the problem of missing information in mechanisms based on the secretary problem. In such decision-making situations (e.g., auction items, bid quotes), candidates appear sequentially, and the decision-maker must make a decision in a blind environment without knowing future information. All of the above models are concerned with decision-making under uncertainty. However, they usually assume that the observed candidate signals are reliable and do not consider how noise affects each candidate separately.

Therefore, in this paper, the author will add the problem of noise in evaluations to the theory of the secretary problem and examine its effects on decision-making performance systematically. This paper provides an "add-on" view to the above extensions of secretary strategies; that is to say,

ISSN 2959-6130

although the previous extensions have focused on adjusting decision goals and structural constraints (modifying selection conditions), the method in this paper offers a more practical final evaluation based on the actual effectiveness of secretary strategies in real-life decision-making scenarios.

3. Methodology

First, the author will briefly introduce the classic secretary problem in this part and then present a noisy extension that better reflects the decisions in real life. Finally, the author will show a simulation framework to examine the performance of the above strategies.

3.1 The Secretary Problem

The classical secretary problem has the following assumptions. Among a total of *n* candidates, all candidates appear in a random order with equal probability. The decision-maker observes each candidate and immediately decides whether to accept or reject them. Furthermore, any rejection decision is irreversible. The objective of the problem is to maximize the probability of selecting the single best candidate. Since the absolute performance of the candidates is unknown, the decision-maker can only make judgments based on the relative rankings of the currently observed candidates [3]. The decisive factor in the optimal strategy lies in the "threshold rule" of the secretary strategy-let $r = n/e$. The decision-maker first observes and rejects the first *r* candidates, using this group as a reference sample. After observing the first *r* candidates, the decision-maker selects the first candidate in the subsequent phase whose performance exceeds that of all the previously selected *r* candidates. Under ideal conditions, the success probability of this strategy is approximately $1/e \approx 0.3678$.

3.2 The Noisy Secretary Problem

Although the classical model presumes that the decision-maker can make an ideal and open assessment of all candidates, such conditions are not available in reality. To address the above deficiency, the author proposes a "noisy secretary model" in which the observed performance scores of each candidate by the decision-maker are subject to random noise.

Let the true performance score of the i-th candidate be denoted by q_i. However, the decision-maker does not see the true performance score but a noisy version of it, where $\left(? \sim N\left(0, ?\sigma^2\right)\right)$ is random noise with a mean of 0 and a variance of σ^2. This form is for uncertainty in the evaluation process (e.g., interview scoring or bid valuation). Based on the above model, a decision-maker makes decisions according to a number of uncertain observations, and finally determines whether the actual outcomes of those decisions meet a set of ideal performance criteria. Noise may cause the decision-maker to incorrectly rank candidates, and thus the final result of optimal stopping theory will not be optimal [4].

3.3 Retaining the Original Threshold Rule

The author will keep the "threshold rule" of the secretary problem in this study and apply it to noisy observational data. The specific procedure is as follows: the decision-maker first observes and rejects the first $r = ?n/e?$ candidates, and records the candidate with the highest observed value among the first r candidates. If no candidate with a higher observed value appears in the following rounds, then the last candidate in the sequence is selected by default. This rule will show the effect of noise on this strategy directly, without changing the operating environment for the secretary.

To isolate and study the effect of noise more directly, the author intentionally kept the threshold rule of the secretary strategy. Thus, one can assess how robust the original stopping rule is in the presence of imperfect information [1].

3.4 Experimental Details

To evaluate the performance of the noisy secretary model, the author conducted multiple Monte Carlo simulations with varying numbers of repetitions under different noise distributions.

In each simulation, *n* candidates were generated, each with a corresponding true performance observation q_i (observations were randomly generated from $U(0,1)$). Gaussian noise is added to each candidate's score (noise is generated from a $N(0,\sigma^2)$ distribution). By adjusting the overall noise level σ, the noise distribution for each round of experiments is varied. Add this q_i to obtain the noisy observation for each candidate q_i. Randomly shuffle the order of all original candidates to simulate their arrival process.

Apply the noisy observations to the secretary strategy using the predefined threshold rule. Apply this secretary strategy to each simulation and record the candidate selected by the decision-maker in each instance. Analyze whether they were the optimal candidate and calculate the true rank of the selected candidate within the entire candidate list

This study repeats this process multiple times to estimate the following metrics. The final success rate of the secretary strategy after all experiments (i.e., the probability of selecting the truly optimal candidate). The average rank of the selected candidate's true observation. Through multiple simulation experiments, the author analyzed how uncertainty (noise) affects the secretary's decision-making performance and overall robustness. This experimental framework quantifies the extent to which the performance of the optimal stopping strategy degrades as noise intensity increases.

4. Result

4.1 Baseline Validation Under Perfect Information

The author first evaluates the decision-making performance of the secretary strategy under an ideal setting where there is no noise (i.e., when all $\sigma = 0$). As shown by the simulation results in Fig. 1, under the premise that the threshold rule satisfies $r = n / e$ in all cases, as the number of simulations runs increases, the probability of the secretary strategy selecting the best candidate in the ideal scenario peaks at the point n / e . This confirms that the threshold people selected indeed maximizes the probability of selecting the best candidate, and this result also demonstrates the fundamental accuracy of the secretary strategy.

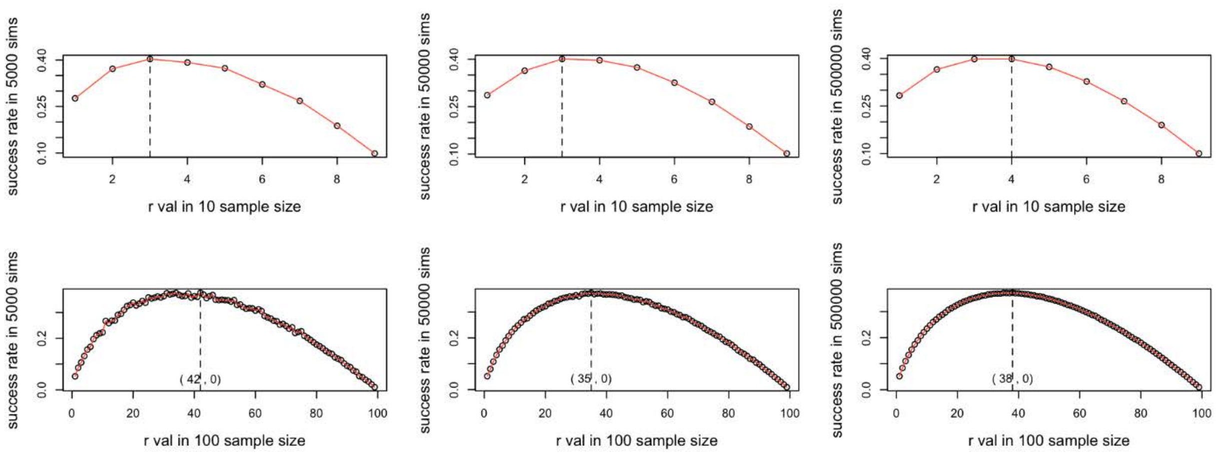

Fig. 1 Secretary success rate trending under r=n/e assumption.

4.2 Effect of Noise on the Success Probability

The most interesting finding in the noisy secretary experiment investigated here is that as the level of noise increases, the chance of eventually choosing the best candidate decreases quickly. The final success rate in all repetitions of the simulation has a downward trend even with relatively low noise levels.

The trend is clearly seen in the Fig. 2 describing Success Rate under Different Noise Levels. The three subplots show the success rate trend for candidate pool sizes of 50, 100 and 200 candidates respectively, whereas the colored curves are the simulation runs of 100, 1, 000 and 5, 000 iterations respectively. For example, if $\sigma = 0.1$, the success rate will drop from about 37% to about 13%; that is, it will fall by more than 60%. As the noise fluctuation increases further ($\sigma > 0.5$), the success rate will be very low (about 3%-5%) and perform only slightly better than random selection ($100\% / n$), with an improvement of only about 1%.

Based on the above results, the effectiveness of the secretary strategy is relatively sensitive to the amount of observable data for decision-makers. Even a small difference between the actual score and the estimated rank will significantly lower the final success rate of the secretary decision-making mechanism [4].

ISSN 2959-6130

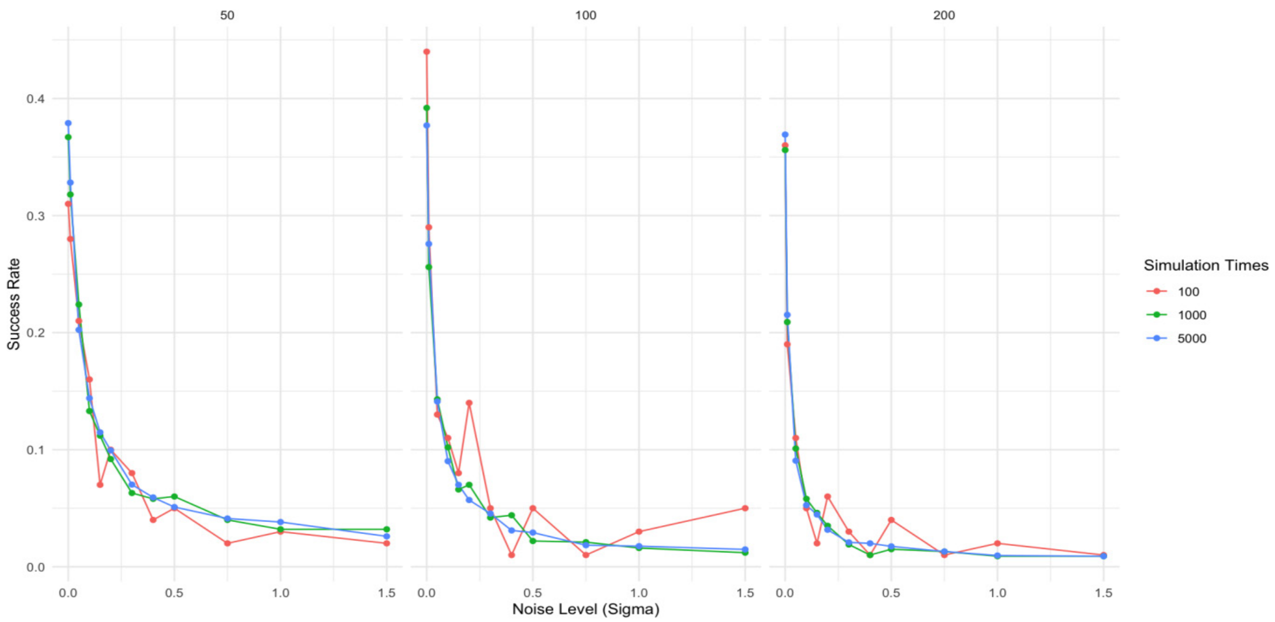

Fig. 2 Success Rate under Different Noise Levels

4.3 Degradation of Selection Quality (Average Rank Analysis)

Although the success rate shows how often the best single candidate is selected, it fails to reflect the quality of all choices in multiple experiments. To determine the general decision-making ability in all the experiments, the author added a new indicator of performance: "average rank" (the mean rank of selected results); a higher average rank value generally suggests poorer overall performance and can help observe the effect of noise.

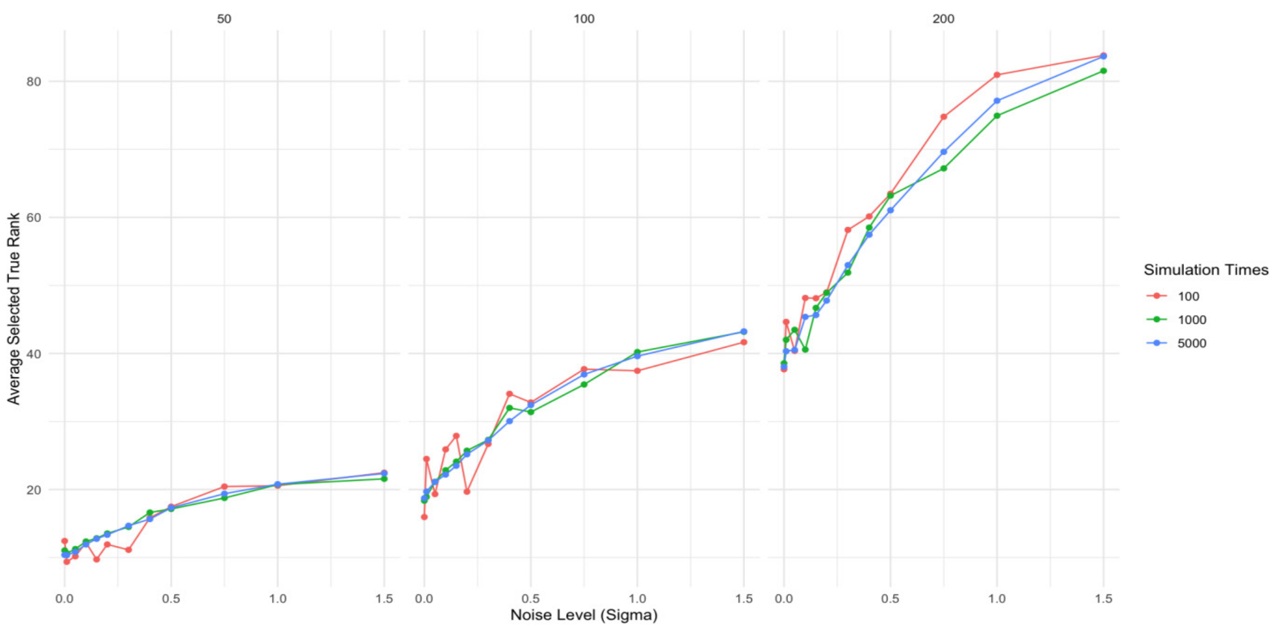

Fig. 3 Average Selected Rank under Different Noise Levels

As shown in the Fig. 3 about "Average Selected Rank under Different Noise Levels", the mean of the rank values is gradually increasing with an increase in the overall noise fluctuation. Under ideal conditions with complete information by the decision-maker (i.e., no noise), the average rank values for a total of 50, 100, and 200 candidates are 11.29, 17.68 and 38.09 respectively (corresponding to the top 22.58%, 17.68% and 19.05% of all rankings). When the noise is relatively large (σ=1.5), the mean ranks are as follows: 14.85, 28.45 and 55.73, corresponding to the top 29.69%, 28.45% and 27.86% of all scores, respectively. Although the model has some noise variations and thus a lower average rank, it can still select candidates from the top 30% of the entire score pool for the secretary.

A monotonic increase is therefore a relatively steady deterioration of decision quality. Although it is not as steep a decline in success rate, the average rank is also relatively high, and therefore the algorithm will still be able to identify some relatively good candidates [7].

4.4 Regret Analysis and Decision Loss

To further quantify the impact of noise on secretary strategy, the author created and analyzed the "average regret" metric (the higher the average regret value, the lower the quality of the selected candidates).

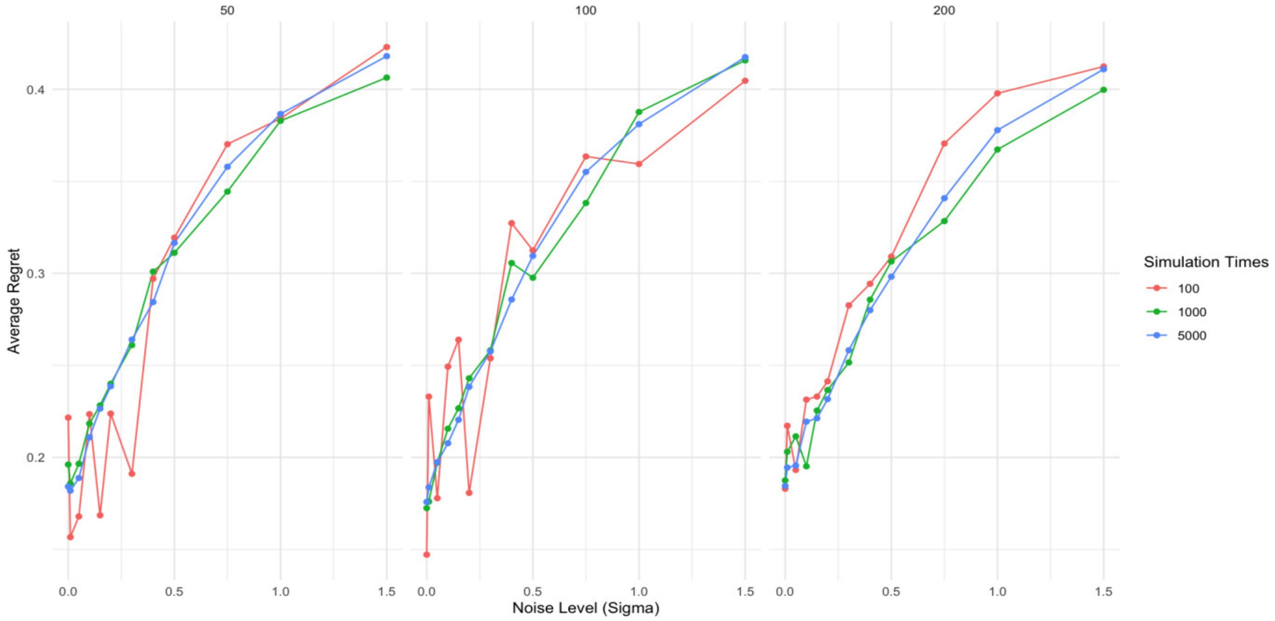

Fig. 4 Average Regret under Different Noise Levels

According to the trend presented in Fig. 4 about "Average Regret under Different Noise Levels" graph, all curves show a consistent upward trajectory overall. Compared to success rate and average rank, average regret more directly reflects the cost corresponding to the value of a chosen objective.

In reality, as noise fluctuations increase, the overall trend of average regret begins to rise. In a noise-free environment with $\sigma = 0$, the average regret value is approximately 0.18–0.20. However, when the noise variance increases to $\sigma = 1.0$, the average regret rises to around 0.38–0.4, meaning the decision loss nearly doubles [8].

Another important observation is that the results remain highly consistent across different simulation scales. While there are significant fluctuations in the results obtained from 100 repeated simulations, the stability of the estimation results improves significantly with a substantial increase in the number of simulations (1000 or 5000). This summary of results indicates that the decrease in the effectiveness of the secretary strategy due to increased noise (e.g., decreased success rate, increased average rank, increased average regret) has been confirmed by multiple simulation experiments and is not due to randomness.

4.5 Real-World Applications: Recruitment and Online Auctions

The observed noise effects described above can be directly applied to real-world decision-making environments. In recruitment scenarios, a low noise level (σ= 0) corresponds to a decision-making environment where all candidate information is highly reliable, such as structured technical interviews or standardized testing environments. In this case, the secretary strategy is a highly effective stopping strategy because the observed experimental results accurately reflect the performance of the selected candidates in that ranking.

However, in more realistic environments with decision-maker subjectivity, such as behavioral interviews, cultural fit assessments of interviewers, or incomplete resume information, noise fluctuations significantly increase and become unpredictable. These conditions represent the moderate noise level in this study (ranging from σ ≈ *0.2 to 0.5*), and the results show that the effectiveness of the secretary strategy declines significantly in such decision-making situations. This extended study of the sec-

ISSN 2959-6130

retary problem emphasizes the criticality of "incomplete information" and "candidate value heterogeneity" in decision-making environments. In this model, decision quality depends not only on the order in which candidates appear but also on the reliability of candidate information [9].

In the decision-making scenario of online auctions, noise fluctuations mostly stem from incomplete information about the value of the auctioned item or the behavior of bidders. Bidders may have incomplete information about the auctioned item and the auction environment, be influenced by strategic incentives, or have behavioral biases among different bidders, resulting in bidders' bids failing to reflect the true value of the auctioned item. This situation is more likely to occur in high-noise decision-making environments (i.e., $\sigma \geqslant 0.5$) in this study. According to the experimental results of this study, in such scenarios, the performance of the secretary strategy is only slightly better than random selection. It is worth noting that previous studies on the "multiple-choice secretary algorithm" in auction environments have pointed out that allowing multiple candidates to be selected simultaneously can alleviate the limitations caused by this noise to some extent [10].

Equally important, when the noise fluctuation is extremely high (i.e., $\sigma \geqslant 1$), this setting means that the observed information in the decision-making environment is almost completely without reference value. In practice, this reflects the extreme unreliability of the decision-maker's evaluation process or the serious information asymmetry, resulting in almost no meaningful signal from the candidates being obtained during the decision-making process. Although such conditions are extremely rare in well-informed decision-making environments, this situation represents a crucial critical point: once the noise variance exceeds $\sigma \geqslant 1$, the secretary strategy and more broadly, any ranking-based decision rule loses its effectiveness and approaches the effect of random selection.

4.6 Limitations

While this study reveals the limitations of the secretary strategy in noisy decision-making environments, caution is still needed when interpreting its findings.

First, the noise model used in this study assumes that all noise pre-assumptions are independent and Gaussian distributed. This assumption cannot fully simulate the errors in some real-world cases, and in most real-world cases, the randomness and unpredictability of errors are extremely high. Second, even in high-noise environments, the secretary strategy retains some structural characteristics (it is still based on relative comparison, and the results are still better than completely random selection), indicating that the strategy has not completely failed, but has merely become unreliable.

Moreover, the original secretary problem assumes that the decision-maker is able to provide true information about the candidates' rankings, and that is not the case in real-world decisions, where imperfect proxy variables are used. Even in the most typical situation of the "secretary problem" with a "company recruitment", noise is still present. This gap between theoretical pre-assumptions and the real world is what this study is revealing as degradation of decision-making performance.

Conclusion

In this paper the author presents some random noise in the system and further elaborate on the classic secretary problem, all in an effort to model the real world decision-making environment more faithfully. While for optimal decisions under ideal conditions (all constraints and information satisfied), the secretary strategy is the most effective, the results have proved that the performance of this strategy decreases sharply when real world factors (noise) become more and more prevalent.

Indeed, if the noise level is not high, the noise it adds to the selection mechanism itself results in a significant reduction of the success rates. At the same time, the average ranking and regret values do monotonically increase as the number of steps increases, showing the secretary strategy to be very sensitive to the accuracy of the information on which it is based; the assumption of reliable rankings is critical to its optimal behaviour.

In a noisy environment, the robustness of the secretary strategy is relatively low, and it is reliability is not consistent. In the practical aspect, the results gained after adding noise show that there are some discrepancies between these ideal preconditions of the optimal stopping model and the real world conditions. When the information about candidates is incomplete, as in the case of recruitment or online auctions, applying the classic secretary directly leads to unstable outcomes in the decision. The conclusion is that in order to make decisions more stable under real-world conditions, one needs to enrich the classic secretary problem with some mechanisms, like the multiple-choice rules, or weighted decision rules.

To sum up, this study underlines that in practice the accuracy of the pre-assumptions of the classic secretary problem is of crucial importance in determining the optimal selection criteria. Future work may involve secretary strategy models which are able to ignore noise and thus better model real-world disruptions due to noise instability.

References

[1] KAPLAN H, MANSOUR Y, NASSIE S. Secretary problem and independent sampling with limited memory. Management

Science, 2024, 70(3): 1582-1598.
[2] PEREZ-SALAZAR S, SINGH M, TORIELLO A. Robust Online Selection with Uncertain Offer Acceptance. Mathematics of Operations Research, 2025, 50(3): 2226-2260.
[3] Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, et al. Game-theoretic statistics and safe anytime-valid inference. Statistical Science, 2023, 38(4): 576-601.
[4] Salem, J., & Gupta, S. Secretary Problems with Biased Evaluations Using Partial Ordinal Information. Management Science, 2024, 70(8), 5337–5366.
[5] FUJII K, YOSHIDA Y. Secretary problem of predictions and an order of choice. International Journal of Geometric Methods in Modern Physics 2024, 21: 05.
[6] Babaioff, M., Immorlica, N., & Kleinberg, R. Matroids, secretary problems, and online mechanisms. Microsoft Research, 2007, 434–443.
[7] HARB E. Fishing For Better Constants: The Prophet Secretary Via Poissonization[EB/OL]. (2023-03-01). https://arxiv.org/abs/2303.00605.
[8] CORREA J, CRISTI A, DÜTTING P, et al. Fairness and Bias in Online Selection. Operations Research, 2025. DOI:10.1287/opre.2021.0662.
[9] Gnedin A. Best choice from the planar Poisson process. Stochastic Processes and their Applications, 2024, 111, 317–354.
[10] SALEM J, GUPTA S. Secretary Problems with Biased Evaluations Using Partial Ordinal Information[J]. Management Science, 2024, 70(8): 5337-5366.