



HAL
open science

On the Rank Distribution in Discrete Choice Models

Karim Kilani

► **To cite this version:**

| Karim Kilani. On the Rank Distribution in Discrete Choice Models. 2024. hal-04670369

HAL Id: hal-04670369

<https://cnam.hal.science/hal-04670369v1>

Preprint submitted on 12 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Rank Distribution in Discrete Choice Models

Karim Kilani*

LIRSA, Conservatoire National des Arts et Métiers,
292 Rue Saint-Martin, 75003 Paris, France
karkil2205@gmail.com

August 2024

Abstract

This paper establishes a new identity within the framework of probabilistic discrete choice models, linking the ranking probabilities of alternatives in a set to best-choice probabilities in subsets through an alternating sum weighted by binomial coefficients. Inspired by results from the order statistics literature, our approach enriches the analysis of revealed and stated preferences by providing comprehensive estimates of the choice probabilities at any rank of alternatives. It is particularly useful for analyzing voting systems, such as the Borda count method, by enabling the prediction of outcomes under different voting rules.

Keywords: best-choice probabilities, borda count, discrete choice models, logit model, order statistics, rank of an alternative.

JEL Classification: C35

*Corresponding author.

1 Introduction

The pioneering work on discrete choice models, particularly that of McFadden (1974, 1981), primarily focused on the probability that an alternative is ranked first in a set of alternatives. More recently, de Palma et al. (2017) and de Palma and Kilani (2023) have shown that the probability that an alternative is the worst choice in a set can, through the principle of inclusion-exclusion, be expressed as an alternating sum of best-choice probabilities in subsets containing that alternative.

This article establishes a new identity that generalizes this approach to the probability that an alternative is ranked at any given position. We demonstrate that this probability can be expressed as an alternating sum of best-choice probabilities in subsets, weighted by binomial coefficients. Similar types of alternating weighted sums are found in the literature on order statistics.

Order statistics focus on the distribution of ordered random variables within a sample (David and Nagaraja, 2004). Our work establishes similar identities for discrete choice models. To our knowledge, the identities presented in our article are not found in any existing literature, including that of order statistics. Thus, this work contributes both to the economic theory of discrete choice and to the statistical theory of order statistics.

In Section 2, we present the notation and review the basic concepts regarding the best-choice probabilities within subsets. Section 3 will introduce the concept of the rank of an alternative within a choice set and then provide the distribution of ranks. We will show that the elements of this distribution can be expressed as a sum of Block-Marschak polynomials, according to a lemma proven in this paper. The main theorem of this paper will establish this distribution of ranks. In Section 4, we will propose an application of these results to voting theory, particularly to the Borda count rule for determining the winner of an election. Finally, Section 5 will conclude with a summary of the results and perspectives for future work.

2 Notations

Let C be the total set of alternatives, which can be candidates, projects, goods, or services, with $|C| \geq 2$, where $|C|$ denotes the cardinality of the set C . Let Ω be the sample space of all possible permutations of the alternatives in C . Each permutation $\omega \in \Omega$ represents a possible ordering of the alternatives. We assume that there exists a probability distribution P over this sample space.

For each alternative i in C , we define R_i as the random variable representing the rank of alternative i within the set C . Thus, R_i takes integer values between 1 and $|C|$.

Let A and B be two subsets of C . We denote (A, B) as a part of Ω consisting of all permutations in which the alternatives in A are ranked before those in B . Formally, we

have:

$$(A, B) = \{\omega \in \Omega \mid R_i \leq R_j, \forall i \in A \text{ and } \forall j \in B\}.$$

By convention, if one of the subsets is empty, (A, B) is defined as the entire sample space Ω .

For a singleton $\{i\}$, we simply use the label i of the alternative, without the braces. Thus, (i, B) denotes all permutations $\omega \in \Omega$ where i is ranked before every alternative in B . We are particularly interested in the probabilities of these events, denoted $P(i, B)$, which are the best-choice probabilities within subsets and are crucial values in our paper.

A well-known and widely used model in the literature is the logit model. In this model, the probability that alternative i is ranked the highest in B is given by:

$$P(i, B) = \frac{e^{v_i}}{\sum_{k \in B} e^{v_k}}, \quad i \in B \subset C. \quad (1)$$

When the model is derived as an additive random utility model, v_i refers to the deterministic component of the utility for each alternative i . The remaining random components are assumed to be independent and identically distributed according to an Extreme Value Type I distribution.

We use the logit model as an example, but there are many other interesting and richer formulations within additive random utility models. For instance, the nested logit model allows for a hierarchical structure with nests (or groups) of alternatives, which can be particularly useful in voting contexts where candidates are categorized into groups such as right-wing, centrist, and left-wing. The Generalized Extreme Value (GEV) model and the mixed logit model offer further extensions for modeling more complex behaviors. Our approach encompasses all these models.

One major advantage of discrete choice models, as discussed above, is the extensive existing literature, particularly on estimation methods. These methods allow us to determine the best-choice probabilities from sample data on individual preferences. Best-choice probabilities are crucial because they enable extrapolation from sample data to estimate best-choice frequencies in the population of interest. With the forthcoming identity in this paper, we will be able to extrapolate the frequencies with which individuals rank an alternative at any given position, not just the first.

3 The distribution of the rank of an alternative

We study the distribution of the rank of an alternative, i.e., the probabilities $P(R_i = r)$ for $r = 1, \dots, |C|$. We do this by partitioning the event $(R_i = r)$ into smaller events, whose probabilities are key for rationalizing a system of best-choice probabilities. These smaller events relate to central concepts in the seminal paper by Falmagne (1978), which

established the necessary and sufficient conditions for a system of best-choice probabilities to be consistent with a probability distribution P over the sample space Ω , according to the notations used in this paper.

For $R_i = r$, i must be ranked highest among all alternatives in any subset A of size $|C| - r + 1$, and ranked below all alternatives in the complement $C \setminus A$ of that subset. The event that i is ranked at position r is then given by:

$$(R_i = r) = \bigsqcup_{\substack{A:\{i\} \subseteq A \subseteq C, \\ |A|=|C|-r+1}} ((C \setminus A, i) \cap (i, A)),$$

where \bigsqcup denotes the disjoint union.

Since this is a disjoint union, the probability of the disjoint union is the sum of the individual probabilities. Thus, we have:

$$P(R_i = r) = \sum_{\substack{A:\{i\} \subseteq A \subseteq C \\ |A|=|C|-r+1}} P((C \setminus A, i) \cap (i, A)), \quad i \in C, r = 1, \dots, |C|. \quad (2)$$

For $R_i = 1$, i must be ranked before all other alternatives. This corresponds to the event where i has the highest rank in the total set C , i.e., $(R_i = 1) = (i, C)$. Therefore, the probability that i is ranked first is given by $P(R_i = 1) = P(i, C)$, which is consistent with equation (2). This represents the best-choice probability in the total set C . For lower ranks, we need to further investigate the probabilities $P((C \setminus A, i) \cap (i, A))$.

The event $(C \setminus A, i) \cap (i, A)$ means that the alternatives in $C \setminus A$ are strictly ranked before i , which, in turn, is ranked before (or equal to) all alternatives in A . We will show in the following lemma that the probabilities of these events can be expressed as an alternating sum of best-choice probabilities of the form $P(i, B)$, where B contains A .

Lemma 1. *The probability that alternative i is ranked after the alternatives in $C \setminus A$ and before those in A can be expressed as an alternating sum of best-choice probabilities:*

$$P((C \setminus A, i) \cap (i, A)) = \sum_{B:A \subseteq B \subseteq C} (-1)^{|B|-|A|} P(i, B), \quad i \in A \subseteq C. \quad (3)$$

Proof. We use the indicator function $\mathbb{1}(\cdot)$ in our proof. Since the indicator of the intersection of two events is the product of the indicators of those events, we have: $\mathbb{1}((C \setminus A, i) \cap (i, A)) = \mathbb{1}(C \setminus A, i) \times \mathbb{1}(i, A)$. The indicator $\mathbb{1}(C \setminus A, i)$ can be expressed as a product of indicators:

$$\mathbb{1}(C \setminus A, i) = \prod_{k \in C \setminus A} \mathbb{1}(R_i > R_k) = \prod_{k \in C \setminus A} (1 - \mathbb{1}(R_i < R_k)).$$

By expanding the rightmost product, we obtain:

$$\mathbb{1}(C \setminus A, i) = 1 + \sum_{\emptyset \subseteq D \subset C \setminus A} (-1)^{|D|} \prod_{k \in D} \mathbb{1}(R_i < R_k).$$

Thus, the indicator of $(C \setminus A, i) \cap (i, A)$ becomes:

$$\mathbb{1}((C \setminus A, i) \cap (i, A)) = \left(1 + \sum_{\emptyset \subsetneq D \subset C \setminus A} (-1)^{|D|} \prod_{k \in D} \mathbb{1}(R_i < R_k) \right) \times \mathbb{1}(i, A).$$

Distributing $\mathbb{1}(i, A)$ in the terms of the sum, we get:

$$\mathbb{1}((C \setminus A, i) \cap (i, A)) = \mathbb{1}(i, A) + \sum_{\emptyset \subsetneq D \subset C \setminus A} (-1)^{|D|} \mathbb{1}(i, A \cup D),$$

which simplifies to:

$$\mathbb{1}((C \setminus A, i) \cap (i, A)) = \sum_{D \subset C \setminus A} (-1)^{|D|} \mathbb{1}(i, A \cup D).$$

Applying the expectation operator to both sides of this equation, we obtain:

$$P((C \setminus A, i) \cap (i, A)) = \sum_{D \subset C \setminus A} (-1)^{|D|} P(i, A \cup D).$$

By setting $B = A \cup D$ and noting that $|D| = |B| - |A|$, we arrive at the desired expression. \square

Thus, the alternating sums on the right-hand side of equation (3) correspond to the Block-Marschak polynomials. Their non-negativity has been established by Falmagne (1978) as both a necessary and sufficient condition for the rationalizability of a system of best-choice probabilities. This interpretation, provided by our Lemma 1, has also been recognized by Fiorini (2004), who offers a shorter proof than the pioneering one by Falmagne.

Using the previous lemma, we are now able to establish the distribution of the rank R_i for any alternative $i \in C$, thereby constituting the fundamental result of this paper. Before stating our main theorem, note that we use binomial coefficients $\binom{n}{k}$, with the convention that $\binom{n}{k} = 0$ if $k > n$. This convention simplifies the notation of the alternating sums that appear in our result.

Theorem 1. *The distribution of the rank R_i , which expresses the probabilities that alternative i is at different ranks r , is given by the following identity, which relates these*

probabilities to best-choice probabilities in subsets of C :

$$P(R_i = r) = \sum_{B:\{i\}\subseteq B\subseteq C} (-1)^{|B|-|C|+r-1} \binom{|B|-1}{|C|-r} P(i, B), \quad i \in C, \quad r = 1, \dots, |C|. \quad (4)$$

Proof. Starting from Equation (2) and applying Lemma 1, we obtain:

$$P(R_i = r) = \sum_{\substack{A:\{i\}\subseteq A\subseteq C \\ |A|=|C|-r+1}} \sum_{B:A\subseteq B\subseteq C} (-1)^{|B|-|A|} P(i, B).$$

We can interchange the two sums to obtain:

$$P(R_i = r) = \sum_{\substack{A:\{i\}\subseteq A\subseteq C \\ |A|=|C|-r+1}} \left(\sum_{B:A\subseteq B\subseteq C} (-1)^{|B|-|A|} \right) P(i, B).$$

The number of subsets A of B of size $|C| - r + 1$ is given by the binomial coefficient $\binom{|B|-1}{|C|-r}$, since alternative i is already included in both A and B and should not be taken into account in the counting. Thus, we obtain:

$$P(R_i = r) = \sum_{\substack{B:\{i\}\subseteq B\subseteq C \\ |B|\geq|C|-r+1}} (-1)^{|B|-1-|C|+r} \binom{|B|-1}{|C|-r} P(i, B).$$

At this point, the convention that a binomial coefficient is zero when the lower index exceeds the upper index comes into play, eliminating the constraint on the size of the subsets and resulting in the form given in Equation (4). \square

For the case $r = 2$, corresponding to the probability that alternative i is ranked in the second-best position within the total set C , equation (4) simplifies by considering the subsets B with a cardinality of $(|C| - 1)$, that is, C minus one alternative j other than i , as well as the entire set C :

$$P(R_i = 2) = \sum_{j \in C \setminus \{i\}} P(i, C \setminus \{j\}) - (|C| - 1)P(i, C), \quad i \in C. \quad (5)$$

For $r = |C|$, corresponding to the worst-choice probability of i in the total set C , equation (4) becomes:

$$P(R_i = |C|) = P(C, i) = \sum_{B:\{i\}\subseteq B\subseteq C} (-1)^{|B|-1} P(i, B), \quad i \in C. \quad (6)$$

This is precisely the alternating sum found in de Palma et al. (2017), a special case of our more general result, which covers all ranks, not just the last one.

In the case of the logit model, by substituting the logit best-choice probabilities $P(i, B)$ given by Equation (1) into our identity given by Equation (4), we obtain:

$$P(R_i = r) = \sum_{B:\{i\}\subseteq B\subseteq C} (-1)^{|B|-1-|C|+r} \binom{|B|-1}{|C|-r} \frac{e^{v_i}}{\sum_{k\in B} e^{v_k}}, \quad i \in C, \quad r = 1, \dots, |C|. \quad (7)$$

To our knowledge, this result has not yet been reported in the literature. While it is possible to obtain full ranking probabilities compatible with logit best-choice probabilities using the exploded logit, as first demonstrated by Beggs et al. (1981), it is challenging to recover the specific form given in Equation (7) directly from the exploded logit without redoing the combinatorial work we have used here. It is worth noting that the exploded logit was derived within the framework of additive random utility models, while our identity is valid for any logit model, without requiring such a formulation.

4 Application in Voting Theory

The Borda count is a method for assigning value to alternatives based on their ranking in a voting context. The higher an alternative is ranked, the more points it receives. For a detailed description of this rule and other voting methods, we refer the reader to Levin and Nalebuff (1995).

The Borda rule has several variations. In the version described here, m represents the number of top-ranked alternatives that will receive points, with $1 \leq m \leq |C|$, where $|C|$ is the total number of alternatives. Each individual assigns m points to their top-ranked choice, $m - 1$ points to their second-ranked choice, and so on, down to 1 point for their m -th ranked choice. The remaining $|C| - m$ alternatives receive 0 points. The alternative with the highest total points wins the vote. If $m = 1$, this rule reduces to a simple majority vote.

Mathematically, and within the context of our probabilistic choice framework, the Borda points assigned to an alternative i , based on the parameter m , are represented by the random variable $S_i^{[m]}$ and are related to the rank variable R_i as follows:

$$S_i^{[m]} = \max(0, m - R_i + 1), \quad i \in C, \quad 1 \leq m \leq |C|.$$

Let s be the number of points assigned, with $1 \leq s \leq m$. We first consider the case where the alternative $i \in C$ has received points. The case $s = 0$ will be handled immediately after using the expression above. We have:

$$P(S_i^{[m]} = s) = P(m - R_i + 1 = s) = P(R_i = m - s + 1), \quad i \in C, \quad 1 \leq s \leq m.$$

By applying the rank distribution described in Theorem 1 (cf. Equation 4), we obtain:

$$P(S_i^{[m]} = s) = \sum_{B:\{i\}\subseteq B\subseteq C} (-1)^{|B|-|C|+m-s} \binom{|B|-1}{|C|-m+s-1} P(i, B), \quad (8)$$

$$i \in C, \quad 1 \leq s \leq m.$$

For the case $s = 0$, we can simply use complementarity to obtain:

$$P(S_i^{[m]} = 0) = 1 - \sum_{s=1}^m P(S_i^{[m]} = s), \quad (9)$$

where the probabilities on the right-hand side should be replaced by the expressions given above.

5 Conclusion

The main identity presented in this paper, within the framework of probabilistic discrete choice models, expresses the rank probabilities of alternatives within a given set as alternating sums of products of binomial coefficients and best-choice probabilities in subsets of alternatives. This identity is inspired by results from identities appearing in the order statistics literature.

Our identity leverages the well-established properties of best-choice probabilities, commonly used in econometrics, to estimate the probabilities of the highest rank as well as those of lower ranks. By utilizing these properties, it enriches analyses based on revealed or stated preferences by providing more comprehensive estimates of the various ranks of alternatives.

This identity is particularly useful for analyzing voting systems, such as the Borda count method briefly presented in this paper. It allows for predicting the outcomes of different voting systems, providing a valuable tool for testing and evaluating various voting rules, which is crucial for the effective functioning of democratic processes. Additionally, it is relevant for addressing rationing issues, enabling the examination of impacts when individuals do not have access to the best alternative. By enriching the analysis of preferences in discrete choice models, this identity opens new perspectives for evaluating voting systems and the effects of access limitations on collective choices.

Acknowledgements

I would like to thank Jean Lainé and Paolo Delle Site for their encouragement throughout all stages of my research.

References

- Beggs, S., Cardell, S., and Hausman, J. (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, 17(1):1–19.
- David, H. A. and Nagaraja, H. N. (2004). *Order statistics*. John Wiley & Sons.
- de Palma, A. and Kilani, K. (2023). Best, worst, and bestworst choice probabilities for logit and reverse logit models. *Journal of Choice Modelling*, 49:100449.
- de Palma, A., Kilani, K., and Laffond, G. (2017). Relations between best, worst, and best–worst choices for random utility models. *Journal of Mathematical Psychology*, 76:51–58.
- Falmagne, J. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18(1):52–72.
- Fiorini, S. (2004). A short proof of a theorem of falmagne. *Journal of Mathematical Psychology*, 48(1):80–82.
- Levin, J. and Nalebuff, B. (1995). An introduction to vote-counting schemes. *Journal of Economic Perspectives*, 9(1):3–26.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.
- McFadden, D. L. (1981). Econometric models of probabilistic choice. In Manski, C. F. and McFadden, D. L., editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 198–272. MIT Press, Cambridge, MA, USA.