

Article

Closed-Form Expressions for the Normalizing Constants of the Mallows Model and Weighted Mallows Model on Combinatorial Domains

Jean-Pierre van Zyl ^{1,*} and Andries Petrus Engelbrecht ^{1,2,3}

¹ Division of Computer Science, Stellenbosch University, Stellenbosch 7599, South Africa; engel@sun.ac.za

² Department of Industrial Engineering, Stellenbosch University, Stellenbosch 7599, South Africa

³ Center for Applied Mathematics and Bioinformatics, Gulf University for Science and Technology, Mubarak Al-Abdullah 7207, Kuwait

* Correspondence: 20706413@sun.ac.za

Abstract

This paper expands the Mallows model for use in combinatorial domains. The Mallows model is a popular distribution used to sample permutations around a central tendency but requires a unique normalizing constant for each distance metric used in order to be computationally efficient. In this paper, closed-form expressions for the Mallows model normalizing constant are derived for the Hamming distance, symmetric difference, and the similarity coefficient in combinatorial domains. Additionally, closed-form expressions are derived for the normalizing constant of the weighted Mallows model in combinatorial domains. The weighted Mallows model increases the versatility of the Mallows model by allowing granular control over likelihoods of individual components in the domain. The derivation of the closed-form expression results in a reduction of the order of calculations required to calculate probabilities from exponential to constant.

Keywords: combinatorial; Mallows model; normalizing constant; weighted Mallows model



Academic Editor: Maria Cristina Mariani

Received: 28 July 2025

Revised: 11 September 2025

Accepted: 24 September 2025

Published: 30 September 2025

Citation: van Zyl, J.-P.; Engelbrecht, A.P. Closed-Form Expressions for the Normalizing Constants of the Mallows Model and Weighted Mallows Model on Combinatorial Domains. *Mathematics* **2025**, *13*, 3126. <https://doi.org/10.3390/math13193126>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Permutation and combination spaces are an integral part of the foundations of mathematics and have been studied for centuries. Berge provides excellent explanations of the intricacies of combinatorial counting problems, permutation groups, and the set theory which underpins these issues [1]. Additionally, Biggs outlines the history of the study of combinations and permutations, dating back to the ancient Greek and Hindu historians, which highlights the influence that this field has had on mathematics [2].

Permutations are not simply an ancient concept confined to old texts but are still a relevant and actively studied part of the contemporary literature, e.g., see [3]. The issue of sampling random permutations famously became a contentious issue during the conscription of United States citizens by executive order of president Nixon [4]. The randomness of the sampled conscription dates was questioned by Fienberg [5], who found that conscription ranks were correlated with birthdates with a Spearman rank correlation coefficient of -0.226 . The “non-random” randomly sampled permutation of dates highlights the challenge of creating a distribution defined on a permutation domain.

Closely related to permutations are combinations, which are non-repeated unordered collections of discrete objects. Hall and Knuth emphasize the increased attention combinatorial problems have gained since the advent of computers and how computational complexity remains one of the limiting factors for combinatorial problems [6]. Papadimitriou and Steiglitz further illustrate the immense scale and ubiquity of combinatorial optimization problems (COPs) [7]. The high level of difficulty of COPs is not surprising, since COPs are often underpinned by non-deterministic polynomial time (NP)-hard decision problems. This NP-hardness, combined with the combinatorial explosion of the search space of COPs, necessitates highly specialized control of the exploration–exploitation trade-off in combinatorial optimization algorithms (COAs). Often there are high levels of exploration required, which are emphasized by new “combinatorial pure exploration” approaches in multi-armed bandit algorithms [8].

The normalizing constant, sometimes referred to as the partition function in physics, is an integral part of the definition of a statistical distribution. The normalizing constant ensures that the probability density function (PDF) of a distribution is valid and that the area under the curve of the PDF is one. For many models, a normalizing constant without a closed-form expression results in a doubly intractable problem; a popular example of a doubly intractable sampling problem is in exponential random graph models (ERGMs) for social networks [9]. As a result of the intractability of this class of problems, authors in the contemporary literature have proposed approximation approaches to sample from doubly intractable distributions, e.g., Monte Carlo Metropolis–Hastings (MCMH) [10] and Markov chain Monte Carlo (MCMC) [11].

The Mallows model (MM) [12] is a popular distribution used to model permutations; however, no research has been conducted on the adaptation of the MM for combinations. This paper derives the closed-form expression for the normalizing constant of the MM with different combinatorial discrepancy functions. The contribution of this paper allows the MM to be applied to combinatorial problems with better computational efficiency.

2. The Mallows Model

In a seminal 1957 paper, Mallows proposed the MM as a popular distribution for ranked data [12]. Mallows started with a general non-null model to calculate the probability of putting two objects U_i and U_j in the correct order. The probability of correctly ordering two objects is formulated as

$$P = \frac{1}{2} + \frac{1}{2} \tanh(k \log \theta + \log \phi), \quad (1)$$

where \tanh is the hyperbolic tangent function, $k = j - i$ is the discrepancy between the true ranks of the two objects, and θ and ϕ are model parameters.

The parameters ϕ and θ change the behavior of the model, with ϕ closely associated with the Spearman’s ρ [13], while θ is associated with Kendall’s τ [14]. These two parameters can be varied in order to change two different effects. Firstly, the parameter θ is interpreted as the weights assigned to each of the n objects in the ordering, i.e., the level of importance assigned to a correct ordering in each position. Secondly, the parameter ϕ is interpreted as the probabilities that will be assigned to a newly ranked object, i.e., if $(n - 1)$ objects have been ranked, and an additional object is introduced. Both the weights from θ and the probabilities from ϕ follow a geometric progression, decreasing away from the true position in the ordering.

The model parameters can be varied to obtain different behaviors, with the null hypothesis corresponding to $\theta = \phi = 1$. An additional behavior is obtained when $\phi = 1$, in which case the MM gives a special case of the Bradley–Terry model [15]. When $\theta = 1$,

the joint distribution of ρ and τ asymptotically tends to the bivariate normal form, which makes the two parameters indistinguishable.

Definition 1. Let $\pi_i \in \Omega$ be an arbitrary permutation of length n . Then,

$$P(\pi|\pi_0, \theta) = \frac{1}{\Psi(\theta, \pi_0)} e^{-\theta d(\pi, \pi_0)}, \quad (2)$$

where $\Psi(\theta, \pi_0)$ is the normalizing constant, π_0 is the central tendency, θ is the dispersion parameter, and $d(\cdot, \cdot)$ is the discrepancy function. When $\theta > 0$, π_0 is the mode of the distribution; when $\theta < 0$, π_0 is the antimode; and when $\theta = 0$, the distribution is uniform.

Definition 2. The explicit formulation of the normalizing constant of the MM is

$$\Psi(\pi_0, \theta) = \sum_{\pi_i \in \Omega} e^{-\theta d(\pi_0, \pi_i)}, \quad (3)$$

where Ω is the permutation space.

The explicit form of the normalizing constant in Equation (3) is valid for any $d(\cdot, \cdot)$, but is computationally expensive to calculate. The calculation time complexity is $O(n!)$ for permutation spaces and $O(2^n)$ for combinatorial domains. Therefore, closed-form expressions for the normalizing constant, which do not require the summation over all possible permutations in Ω , are required to make the MM computationally feasible. The issue of computational complexity can be circumvented by the definition of a closed-form expression which is independent of the modal tendency. A closed-form expression of the normalizing constant, $\Psi(\theta)$, depends on only the dispersion parameter and is often calculable in $O(n)$. However, $\Psi(\theta)$ is discrepancy-function-specific, with no general form, and so has to be derived for each new $d(\cdot, \cdot)$.

3. Discrepancy Functions

3.1. Permutation Discrepancy Functions

The original MM was defined with two variations (i.e., Mallows' ϕ and Mallows' θ), which correspond to the use of Kendall's τ and Spearman's ρ , respectively. However, Diaconis proposed that any right-invariant discrepancy metric can be used with the MM and outlined six existing metrics which are compatible [16].

Kendall proposed a “measure of rank correlation”, named τ which measures the number of discordant pairs between two permutations [14]. Practically, Kendall's τ counts the number of inversions required to transform two permutations to be alike. The distance measure based on τ is

$$d_\tau(\pi, \sigma) = \frac{2\Sigma}{n(n-1)}, \quad (4)$$

where $\Sigma = |\{(i, j) \in \Omega \times \Omega : i < j \wedge \pi(i) > \sigma(j)\}|$ is the inversion number, i.e., the number of inversions needed to transform π into σ . The closed-form expression for Kendall's τ was derived by Fligner and Verducci as $\Psi(\theta) = \prod_{j=1}^{n-1} \frac{1-e^{-\theta(n-j+1)}}{1-e^{-\theta}}$ [17].

Spearman's correlation coefficient, alternatively named Spearman's ρ , was proposed by [13]. Spearman's ρ is a non-parametric measure of correlation between the rankings of two variables and is used to measure the discrepancy between permutations. Closely related to ρ is Spearman's footrule [18], which was introduced as an alternative to Spearman's correlation coefficient and can be used with the MM. The ρ -based distance measure for permutations is defined as

$$d_\rho(\pi, \sigma) = \sum_{j=1}^n (\pi(j) - \sigma(j))^2. \quad (5)$$

No known closed-form expression for $\Psi(\theta)$ with the correlation coefficient distance exists.

Hamming developed a system to detect and correct errors on digital computers, a system which relies heavily on the calculation of a distance between binary vectors [19]. The distance chosen by Hamming is equivalent to the complement of logical and (\wedge) of binary vectors, and is referred to as the Hamming distance. The Hamming distance is a count of the number of positions for which the symbols of two permutations differ:

$$d_H(\pi, \sigma) = n - \sum_{j=1}^n \delta_{\pi(j), \sigma(j)} \quad (6)$$

Fligner and Verducci derived the closed-form expression for the normalizing constant with the Hamming distance as $\Psi(\theta) = n! e^{n\theta} \sum_{j=0}^n \frac{(e^{-\theta}-1)^j}{j!}$ [17].

The Ulam distance is the minimum number of insertions and deletions required to transform one permutation to another and is the permutation edit distance [20]. The Ulam distance is a specific case of the generalized Levenshtein edit distance from [21]. Alternatively, the Ulam distance can be thought of as the complement of the longest increasing subsequence of a permutation as follows:

$$d_U(\pi, \sigma) = n - \max_k \{ \{ \pi \circ \sigma^{-1}(i_1), \dots, \pi \circ \sigma^{-1}(i_k) \} \mid \pi \circ \sigma^{-1}(i_1) < \dots < \pi \circ \sigma^{-1}(i_k) \}. \quad (7)$$

There is no additive decomposition form for the Ulam distance. Therefore, Irurozki defined an alternative formula for the normalizing constant as $\Psi(\theta) = \sum_{d=0}^n S_u(n, d) e^{-\theta d}$, where $S_u(n, d)$ is the number of permutations at distance d from the identity [22].

Cayley proposed the distance which counts the minimum number of transpositions required to change a permutation π into σ [23]. The minimum number of transpositions can be calculated as the complement of the number of cycles in the composition of the two permutations which are compared. The resulting formula for the distance is

$$d_C(\pi, \sigma) = n - |\{ \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\} \mid \pi \circ \sigma^{-1}(i_j) = i_{(j+1) \bmod k} \}|. \quad (8)$$

The initial closed-form expression for the Cayley distance was given by Fligner and Verducci [17], with errors corrected by Irurozki [22]. The resulting expression is $\Psi(\theta) = \prod_{j=1}^{n-1} ((n-j)e^{-\theta} + 1)$.

3.2. Combinatorial Discrepancy Functions

Combinatorial discrepancy functions calculate the dissimilarity between sets within the defined universal set of elements, $\kappa, \eta \subseteq U$, with $|U| = n$. Unlike discrepancy functions on permutation domains, combinatorial discrepancy functions do not account for the ordering of elements.

Although the Hamming distance was originally developed for binary vectors, and is popularly applied to permutations, it is also used for sets. The Hamming distance in set notation makes use of the intersection operation, instead of the logical and, which results in

$$d_H(\kappa, \eta) = n - |\kappa \cap \eta|. \quad (9)$$

The Jaccard distance is one of the most popular set-based distance measures proposed by botanist Jaccard [24]. The Jaccard index was originally developed to measure the amount

of diversity of plant species in different plots of land, with the Jaccard distance defined as the complement of the Jaccard index. The Jaccard distance is

$$d_J(\kappa, \eta) = 1 - \frac{|\kappa \cap \eta|}{|\kappa \cup \eta|}. \quad (10)$$

Similarity-coefficient-based distance metrics are popular in the existing literature. Similarity coefficients measure the amount of agreement (in present attributes and non-present attributes) between two sets. Matching coefficient (MC) is a term often used to describe similarity coefficients related to the Rand index [25] or the coefficients attributed to either Goodall [26] or Sokal and Michener [27]. MCs are often normalized to the range $[0, 1]$, with a corresponding distance metric defined as the complement of the coefficient. The generally used form of similarity-coefficient-based distances is

$$d_M(\kappa, \eta) = \frac{|\kappa \cap \eta| + |\kappa^c \cap \eta^c|}{n}. \quad (11)$$

A simple measure of dissimilarity which can be used for sets is the count of the number of insertions and deletions required to transform set A to set B . The count of insertions and deletions is simply the symmetric difference between two sets, denoted by Δ , which is formulated as

$$d_S(\kappa, \eta) = |\eta \setminus \kappa| + |\kappa \setminus \eta| = |\kappa \Delta \eta|. \quad (12)$$

The symmetric difference is closely related to the Jaccard distance, since the Jaccard distance is equivalent to $\frac{|A \Delta B|}{|A \cup B|}$.

4. Normalizing Constant for Combinatorial Domains

Fligner and Verducci outline a process by which the closed-form expression of the normalizing constant can be derived, representing the distance to the central tendency as a random variable v [17]. The moment-generating function (MGF) of the distribution over v , $D(v)$, is calculated under assumed conditions of uniformity over the input domain. The Taylor series expansion of the additive decomposition of the discrepancy measure is then substituted into the MGF of $D(v)$, which results in the closed-form expression for the normalizing constant.

The MGF of a discrepancy function as a random variable is given in Definition 3.

Definition 3. Let $D(v)$ be the distance of an arbitrary combination to the central tendency expressed as a random variable.

$$\text{MGF}(D(v)) := M_{D,\theta}(t) \quad (13)$$

Given the condition $\theta = 0$, the following shorthand notation is defined:

$$M_{D,0}(t) = M(t) \quad (14)$$

and

$$P(D(v) = d | \theta = 0) = P_0(d) \quad (15)$$

Under the assumption of uniformity, the normalizing constant can be expressed using the MGF of $D(v)$ as in Lemma (1).

Lemma 1. The normalizing constant, under the assumption of uniformity, is related to the MGF by

$$\Psi(\theta) = 2^n M(-\theta) \quad (16)$$

Proof.

$$\begin{aligned}
 \Psi(\theta) &= \sum_{\pi \in \Omega} e^{-\theta d(\pi, \pi_0)} \\
 &= 2^n \sum_{\substack{d_i = d(\pi, \pi_0) \\ \pi \in \Omega}} P_0(d_i) e^{-\theta d_i} \\
 &= 2^n M(-\theta)
 \end{aligned} \tag{17}$$

□

The exponential tilting, as first defined in [28], of the MM is given in Definition 4.

Definition 4. *Since MM is an exponential family, the MGF of $D(\nu)$ with the MM is*

$$M_{D,\theta} = \frac{M(t - \theta)}{M(-\theta)} \tag{18}$$

According to Fligner and Verducci, the expected value and variance of $D(\nu)$ was derived through direct argument by Lehman [29] and is given in Definition 5.

Definition 5. *The expected value of $D(\nu)$ is*

$$E_\theta[D(\nu)] = \frac{d}{dt} \log(M(t)) \Big|_{t=-\theta} \tag{19}$$

and the variance of $D(\nu)$ is

$$Var_\theta[D(\nu)] = \frac{d^2}{dt^2} \log(M(t)) \Big|_{t=-\theta} \tag{20}$$

Consider an appropriate discrepancy function, $d(\cdot, \cdot)$, and a combination, κ , (for distinguishing nomenclature, permutations are referenced by π and combinations by κ). The distance from κ to the central tendency, κ_0 , can then be defined element-wise as in Definition 6.

Definition 6. *The element-wise representation of a discrepancy function between a combination κ and a central tendency κ_0 is captured by the quantity ϵ_i as*

$$\epsilon_i(\kappa) = \begin{cases} 1 & \text{if } \kappa(i) \circ \kappa_0(i), \\ 0 & \text{otherwise,} \end{cases} \tag{21}$$

where \circ is a dyadic operator for element-wise comparison and $\kappa(i)$ is an element within the combination of κ .

The element-wise representation of the discrepancy function in Equation (21) can then be used to define the additive decomposition of $d(\kappa, \kappa_0) = d(\kappa)$, given in Definition 7.

Definition 7. *The additive decomposition of a discrepancy function is*

$$d(\kappa) = g(X(\kappa)) \tag{22}$$

with

$$X(\kappa) = \sum_{i=1}^n \epsilon_i(\kappa), \tag{23}$$

where g is a function which transforms $X(\kappa)$ if needed (e.g., linear transform, translation, or identity).

Definition 8. Assume that κ is uniformly distributed over Ω ; define a function f such that

$$f(t) = E[t^X] \quad (24)$$

where, for brevity, X is the distance of a sampled combination to the central tendency.

Definition 9. The j th derivative of $f(t)$ evaluated at $t = 1$ is

$$f^{(j)}(1) = E\left[\frac{X!}{(X-j)!}\right], \quad (25)$$

Proof. From the power rule

$$\frac{d}{dt}f(t) = E\left[X \cdot t^{(X-1)}\right], \quad (26)$$

which after repeated application becomes

$$\frac{d^j}{dt^j}f(t) = E\left[X \cdot \dots \cdot (X-j) \cdot t^{(X-j)}\right]. \quad (27)$$

When evaluated at $t = 1$, Equation (27) results in

$$\begin{aligned} \frac{d^j}{dt^j}f(1) &= E\left[X \cdot \dots \cdot (X-j) \cdot 1^{(X-j)}\right] \\ &= E\left[\frac{X!}{(X-j)!}\right]. \end{aligned} \quad (28)$$

□

Lemma 2. From the definition of f and $f^{(j)}$, the Taylor series expansion of $f(t)$ at $t = 1$ is

$$f(t) = \left(\frac{t-1}{2} + 1\right)^n. \quad (29)$$

Proof. The n th degree Taylor series expansion of $f(t)$ around $t = 1$ is

$$f(t) = \sum_{j=0}^n \frac{f^{(j)}(1)}{j!} (t-1)^j. \quad (30)$$

From Definition 9,

$$\begin{aligned} f(t) &= \sum_{j=0}^n \frac{E\left[\frac{X!}{(X-j)!}\right]}{j!} (t-1)^j \\ &= \sum_{j=0}^n E\left[\frac{X!}{j!(X-j)!}\right] (t-1)^j \\ &= \sum_{j=0}^n E\left[\binom{X}{j}\right] (t-1)^j \end{aligned} \quad (31)$$

where the binomial coefficient $\binom{X}{j}$ is zero when $X < j$. The binomial coefficient can also be redefined as

$$\binom{X}{j} = \sum_{A_j} \epsilon_{i_1} \cdot \dots \cdot \epsilon_{i_j} \quad (32)$$

where $A_j = \{S \subseteq \Omega \mid |S| = j\}$. Consequently,

$$E\left[\binom{X}{j}\right] = \sum_{A_j} E\left[\epsilon_{i_1} \cdot \dots \cdot \epsilon_{i_j}\right] = \frac{\binom{n}{j}}{2^j}. \quad (33)$$

Hence,

$$\begin{aligned} f(t) &= \sum_{j=0}^n E\left[\binom{X}{j}\right] (t-1)^j \\ &= \sum_{j=0}^n \frac{\binom{n}{j}}{2^j} (t-1)^j \\ &= \sum_{j=0}^n \binom{n}{j} \left(\frac{t-1}{2}\right)^j. \end{aligned} \quad (34)$$

Finally, from the binomial theorem of $(x+y)^n = \sum_{j=0}^n \binom{n}{j} x^j y^{n-j}$ with the substitution of $x = \frac{t-1}{2}$ and $y = 1$, the result is

$$\begin{aligned} f(t) &= \sum_{j=0}^n \binom{n}{j} \left(\frac{t-1}{2}\right)^j \\ &= \left(\frac{t-1}{2} + 1\right)^n. \end{aligned} \quad (35)$$

□

4.1. Normalizing Constant with the Hamming Distance

This section proves the existence of a closed-form expression for the normalizing constant of the MM with the Hamming distance.

Theorem 1. *The closed-form expression of the MM normalizing constant with the Hamming distance is*

$$\Psi(\theta) = \left(e^{-\theta} + 1\right)^n \quad (36)$$

Proof. Let the element-wise representation of the Hamming distance between two combination be defined using the equality operator (=) as

$$\epsilon_i(\kappa) = \begin{cases} 1 & \text{if } \kappa(i) = \kappa_0(i), \\ 0 & \text{otherwise,} \end{cases} \quad (37)$$

The additive decomposition of the Hamming distance is then defined with the transform function $g_H(x) = n - x$, i.e.,

$$d_H(\kappa) = n - X(\kappa). \quad (38)$$

The MGF of $D(\nu)$ with the Hamming distance and the assumption of uniformity is

$$M(t) = E\left[e^{(n-X)t}\right] = e^{nt} f(e^{-t}) = e^{nt} \left(\frac{e^{-t} - 1}{2} + 1\right)^n. \quad (39)$$

From Lemmas (1) and (2), the closed-form expression for the normalizing constant with the Hamming discrepancy function is

$$\begin{aligned}\Psi(\theta) &= 2^n e^{-\theta n} \left(\frac{e^\theta - 1}{2} + 1 \right)^n \\ &= \left(e^{-\theta} + 1 \right)^n.\end{aligned}\tag{40}$$

□

4.2. Normalizing Constant with the Symmetric Difference

This section proves the existence of a closed-form expression for the normalizing constant of the MM with the symmetric difference.

Theorem 2. *The closed-form expression of the MM normalizing constant with the symmetric difference is*

$$\Psi(\theta) = \left(e^{-\theta} + 1 \right)^n\tag{41}$$

Proof. Let the element-wise representation of the symmetric difference between two combination be defined in terms of the logical exclusive or (\oplus) as

$$\epsilon_i(\kappa) = \begin{cases} 1 & \text{if } \kappa(i) \oplus \kappa_0(i), \\ 0 & \text{otherwise.} \end{cases}\tag{42}$$

The additive decomposition of the symmetric difference is then defined with the transform function $g_S(x) = x$, i.e.,

$$d_S(\kappa) = X(\kappa)\tag{43}$$

The MGF of $D(\nu)$ with the symmetric difference and the assumption of uniformity is

$$M(t) = E[e^{Xt}] = f(e^{-t}) = \left(\frac{e^{-t} - 1}{2} + 1 \right)^n.\tag{44}$$

From Lemmas (1) and (2), the closed-form expression for the normalizing constant with the symmetric difference discrepancy function is

$$\begin{aligned}\Psi(\theta) &= 2^n \left(\frac{e^{-\theta} - 1}{2} + 1 \right)^n \\ &= \left(e^{-\theta} + 1 \right)^n.\end{aligned}\tag{45}$$

□

The cardinality of the symmetric difference between two combinations, $|\kappa \Delta \eta|$, is equivalent to the Hamming distance between κ and η . Therefore, Equations (40) and (45) are equivalent despite the difference in Boolean operator used to define the distance metrics.

4.3. Normalizing Constant with the Similarity Coefficient Distance

This section proves the existence of a closed-form expression for the normalizing constant of the MM with the similarity coefficient.

Theorem 3. *The closed-form expression of the MM normalizing constant with the similarity coefficient is*

$$\Psi(\theta) = \left(e^{-\frac{\theta}{n}} + 1 \right)^n\tag{46}$$

Proof. Let the element-wise representation of the similarity coefficient distance be defined in terms of the negation of the logical exclusive or (\odot) as

$$\epsilon_i(\kappa) = \begin{cases} 1 & \text{if } \kappa(i) \odot \kappa_0(i), \\ 0 & \text{otherwise.} \end{cases} \quad (47)$$

The additive decomposition of the symmetric difference is then defined with the transform function $g_M(x) = \frac{x}{n}$, i.e.,

$$d_M(\kappa) = \frac{X(\kappa)}{n} \quad (48)$$

The MGF of $D(\nu)$ with the similarity coefficient and the assumption of uniformity is

$$M(t) = E[e^{Xt}] = f(e^{-\frac{t}{n}}) = \left(\frac{e^{-\frac{t}{n}} - 1}{2} + 1 \right)^n. \quad (49)$$

From Lemmas (1) and (2), the closed-form expression for the normalizing constant with the similarity coefficient discrepancy function is

$$\begin{aligned} \Psi(\theta) &= 2^n \left(\frac{e^{-\frac{\theta}{n}} - 1}{2} + 1 \right)^n \\ &= \left(e^{-\frac{\theta}{n}} + 1 \right)^n. \end{aligned} \quad (50)$$

□

5. Weighted Mallows Model for Combinatorial Domains

Fligner and Verducci proposed an extension of the MM, the generalized Mallows model (GMM) [17]. The GMM has $n - 1$ dispersion parameters (i.e., θ_j with $1 \leq j < n$), which are each used to influence the probability of a specific position within a sampled permutation. Irurozki proposed the weighted Mallows model (WMM) extension of the MM with the Hamming distance for permutation problems [22]. Since the decomposition vector of the Hamming distance has n terms instead of $n - 1$ terms, the GMM of the Hamming distance does not exist.

The WMM can be constructed for combination spaces as well. The WMM is used to control the likelihood of specific elements being included in the combinations sampled around the central tendency. In other words, the WMM can sample combinations, given a central tendency and collection of dispersion parameters, and take into account different levels of preference for the inclusion of each element $e_i \in U$.

Definition 10. Let the element-wise distance vector be

$$\mathbf{D}(\kappa) = (D_1(\kappa, \kappa_0), \dots, D_n(\kappa, \kappa_0)), \quad (51)$$

where $D_j(\kappa, \kappa_0) = 0$ if $\kappa(i) \odot \kappa_0(i)$ and 1 otherwise (i.e., $\mathbf{D}(\kappa)$ is a vector of discrepancy function components defined using the dyadic operator \odot). It follows that $d(\kappa, \kappa_0) = \sum_{j=1}^n D_j(\kappa)$.

Definition 11. Define a vector

$$\boldsymbol{\epsilon}(\kappa) = (\epsilon_1(\kappa), \dots, \epsilon_n(\kappa)) \quad (52)$$

with $\boldsymbol{\epsilon}(\kappa) = g(\mathbf{D}(\kappa))$ (i.e., each $\epsilon_j(\kappa) = g(D_j(\kappa))$) where g is a function which transforms $D_j(\kappa)$ if needed (e.g., linear transform, translation, or identity). When κ is sampled uniformly,

$\epsilon(\kappa)$ is a random binary vector; therefore, the probability of sampling κ with a distance vector ϵ is $P_0(\epsilon(\kappa) = \epsilon)$.

Lemma 3. The normalizing constant, under the assumption of uniformity, is related to the MGF by

$$\Psi(\theta) = 2^n M_\epsilon(\theta) \quad (53)$$

Proof. The normalizing constant can be defined as a function of the MGF of a random vector X -associated MGF, $M_X(t) = E\left[\prod_{j=1}^n e^{t_j X_j}\right]$. Let

$$\begin{aligned} \Psi(\theta) &= 2^n \sum_{D \in \mathbb{B}^n} P_0(D(\kappa) = D) e^{-\sum_{j=1}^n \theta_j D_j(\kappa)} \\ &= 2^n \sum_{g(D) \in \mathbb{B}^n} P_0(g(D)) e^{-\sum_{j=1}^n \theta_j g(D_j)} \\ &= 2^n \sum_{\epsilon \in \mathbb{B}^n} P_0(\epsilon) e^{-\sum_{j=1}^n \theta_j \epsilon_j} \\ &= 2^n M_\epsilon(\theta). \end{aligned} \quad (54)$$

□

Definition 12. Assume that κ is uniformly distributed over Ω , define a function f similarly to the process in [17] as

$$f_\epsilon(t) = f_\epsilon(t_1, \dots, t_n) = E[t_1^{\epsilon_1} \cdots t_n^{\epsilon_n}] \quad (55)$$

Lemma 4. The Taylor series expansion of $f_\epsilon(t)$ at $t = \mathbf{1}$ is

$$f(t) = \prod_{j=1}^n \left(\frac{t_j + 1}{2} \right) \quad (56)$$

Proof. The multivariate Taylor series expansion of $f(t)$ around $t = \mathbf{1}$ is

$$f(t) = \sum_{j=0}^{\infty} \frac{1}{j!} \sum_{x_1 + \dots + x_n = j} \binom{j}{x_1 \dots x_n} \frac{\partial^j f}{\partial t_1^{x_1} \cdots \partial t_n^{x_n}} \Big|_{t=\mathbf{1}} (t_1 - 1)^{x_1} \cdots (t_n - 1)^{x_n}; \quad (57)$$

while the derivative of f_ϵ with respect to variable t_i is given as

$$\begin{aligned} \frac{\partial f_\epsilon}{\partial t_i} &= \sum_{\epsilon} P_0(\epsilon) t_1^{\epsilon_1} \cdots \epsilon_i t_i^{\epsilon_i-1} \cdots t_n^{\epsilon_n} \\ &= \left(\sum_{\epsilon | \epsilon_i = 0} P_0(\epsilon) t_1^{\epsilon_1} \cdots 0 \cdot t_i^{0-1} \cdots t_n^{\epsilon_n} \right) + \left(\sum_{\epsilon | \epsilon_i = 1} P_0(\epsilon) t_1^{\epsilon_1} \cdots 1 \cdot t_i^{1-1} \cdots t_n^{\epsilon_n} \right) \\ &= 0 + \sum_{\epsilon | \epsilon_i = 1} P_0(\epsilon) \prod_{j \neq i} t_j^{\epsilon_j}, \end{aligned} \quad (58)$$

which evaluated at $t = (1, \dots, 1)$ is

$$\begin{aligned} \frac{\partial f_\epsilon}{\partial t_i} \Big|_{t=\mathbf{1}} &= \sum_{\epsilon | \epsilon_i = 1} P_0(\epsilon) 1^{\epsilon_1} \cdots 1 \cdot 1^{1-1} \cdots 1^{\epsilon_n} \\ &= \sum_{\epsilon | \epsilon_i = 1} P_0(\epsilon). \end{aligned} \quad (59)$$

Equation (59) is simply the probability of sampling a combination uniformly given that element i is included in the combination, i.e.,

$$\sum_{\epsilon|\epsilon_i=1} P_0(\epsilon) = \frac{2^{n-1}}{2^n}. \quad (60)$$

The second-order partial derivation with respect to the same variable t_i is zero, i.e.,

$$\frac{\partial^2 f}{\partial t_i^2} = 0. \quad (61)$$

However, the second-order cross partial derivative with respect to two different variables t_{i_1} and t_{i_2} is

$$\begin{aligned} \left. \frac{\partial f_\epsilon}{\partial t_{i_1} \partial t_{i_2}} \right|_{t=1} &= \sum_{\epsilon|\epsilon_{i_1}=1, \epsilon_{i_2}=1} P_0(\epsilon) 1^{\epsilon_1} \dots 1 \cdot 1^{1-1} \dots 1 \cdot 1^{1-1} \dots 1^{\epsilon_n} \\ &= \sum_{\epsilon|\epsilon_{i_1}=1, \epsilon_{i_2}=1} P_0(\epsilon), \end{aligned} \quad (62)$$

which is the probability of sampling a combination uniformly given that elements i_1 and i_2 are included in the combination, i.e.,

$$\sum_{\epsilon|\epsilon_{i_1}=1, \epsilon_{i_2}=1} P_0(\epsilon) = \frac{2^{n-2}}{2^n}. \quad (63)$$

In general, the k -th-order cross partial derivative evaluated at $t = 1$ is

$$\left. \frac{\partial^k f}{\partial t_{i_1} \dots \partial t_{i_k}} \right|_{t=1} = \frac{2^{n-k}}{2^n} = \frac{1}{2^k} \quad (64)$$

With the underlying assumptions $\binom{j}{\epsilon_1 \dots \epsilon_n} = j! \because \epsilon \in \mathbb{B}^n$, and $j \leq n \because \frac{\partial^j f}{\partial t_i^j} = 0 \forall j > 1$, the Taylor series expansion becomes

$$\begin{aligned} f_\epsilon(t) &= \sum_{j=0}^n \frac{1}{j!} \sum_{A_j} \binom{j}{\epsilon_{i_1} \dots \epsilon_{i_n}} \left. \frac{\partial^j f}{\partial t_{i_1} \dots \partial t_{i_j}} \right|_{t=1} \prod_{s=1}^j (t_{i_s} - 1) \\ &= \sum_{j=0}^n \frac{1}{j!} \sum_{A_j} \frac{j!}{2^j} \prod_{s=1}^j (t_{i_s} - 1) \\ &= \sum_{j=0}^n \frac{1}{2^j} \sum_{A_j} \prod_{s=1}^j (t_{i_s} - 1) \\ &= \sum_{j=0}^n \frac{1}{2^j} e_j((t_1 - 1), \dots, (t_n - 1)) \end{aligned} \quad (65)$$

where $e_j((t_1 - 1), \dots, (t_n - 1))$ is the elementary symmetric polynomial (ESP) of degree j and $A_j = \{S \subseteq \Omega \mid |S| = j\}$. The sequence of the ESP has a generating function of the form

$$\prod_{j=1}^n (1 + x_j \cdot a) = \sum_{j=0}^n a^j \cdot e_j(x_1, \dots, x_n) \quad (66)$$

which can be used to represent the weighted sum of the ESPs as

$$\sum_{j=0}^n \left(\frac{1}{2}\right)^j \mathbf{e}_j(x_1, \dots, x_n) = \prod_{j=1}^n \left(1 + x_j \cdot \frac{1}{2}\right) \quad (67)$$

Hence,

$$f(\mathbf{t}) = \prod_{j=1}^n \left(\frac{t_j + 1}{2}\right) \quad (68)$$

□

5.1. Weighted Mallows Model Normalizing Constant with the Hamming Distance

This section proves the existence of a closed-form expression for the normalizing constant of the WMM with the Hamming distance.

Theorem 4. *The closed-form expression of the WMM normalizing constant with the Hamming distance is*

$$\Psi(\boldsymbol{\theta}) = \prod_{j=1}^n \left(e^{-\theta_j} + 1\right). \quad (69)$$

Proof. Let the element-wise distance vector be $\mathbf{H}(\kappa) = (H_1(\kappa, \kappa_0), \dots, H_n(\kappa, \kappa_0))$ where $H_j(\kappa, \kappa_0) = 1$ if $\kappa(i) = \kappa_0(i)$ and 0 otherwise (i.e., a vector of Hamming distance components). It follows that

$$d_H(\kappa, \kappa_0) = \sum_{j=1}^n H_j(\kappa) \quad (70)$$

Define the vector $\boldsymbol{\epsilon}(\kappa)$ using the function $g_H(x) = 1 - x$,

$$\boldsymbol{\epsilon}(\kappa) = g_H(\mathbf{H}) = (g_H(H_1(\kappa)), \dots, g_H(H_n(\kappa))) = (1 - H_1(\kappa), \dots, 1 - H_n(\kappa)) \quad (71)$$

The MGF parameterized by $\boldsymbol{\epsilon}$ is

$$\begin{aligned} M_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}) &= M_{g_H(\mathbf{H})}(\boldsymbol{\theta}) \\ &= M_{(\mathbf{1} - \mathbf{H})}(\boldsymbol{\theta}) \\ &= e^{-\sum_{j=1}^n \theta_j} f_{\boldsymbol{\epsilon}}(e^{\boldsymbol{\theta}}) \\ &= e^{-\sum_{j=1}^n \theta_j} \prod_{j=1}^n \left(\frac{1 + e^{-\theta_j}}{2}\right) \end{aligned} \quad (72)$$

which, from Lemmas (3) and (4), results in the closed-form expression for the normalizing constant of the WMM with the Hamming distance as

$$\begin{aligned} \Psi(\boldsymbol{\theta}) &= 2^n e^{-\sum_{j=1}^n \theta_j} \prod_{j=1}^n \left(\frac{1 + e^{\theta_j}}{2}\right) \\ &= \prod_{j=1}^n \left(e^{-\theta_j} + 1\right). \end{aligned} \quad (73)$$

□

5.2. Weighted Mallows Model Normalizing Constant with the Symmetric Difference

This section proves the existence of a closed-form expression for the normalizing constant of the WMM with the symmetric difference.

Theorem 5. *The closed-form expression of the WMM normalizing constant with the symmetric difference is*

$$\Psi(\boldsymbol{\theta}) = \prod_{j=1}^n (e^{-\theta_j} + 1). \quad (74)$$

Proof. Let the element-wise distance vector be $\mathbf{S}(\kappa) = (S_1(\kappa, \kappa_0), \dots, S_n(\kappa, \kappa_0))$ where $S_j(\kappa, \kappa_0) = 1$ if $\kappa(i) \oplus \kappa_0(i)$ and 0 otherwise (i.e., a vector of symmetric difference components). It follows that

$$d_S(\kappa, \kappa_0) = \sum_{j=1}^n S_j(\kappa) \quad (75)$$

Define the vector $\boldsymbol{\epsilon}(\kappa)$ using the function $g_S(x) = x$,

$$\boldsymbol{\epsilon}(\kappa) = g_S(\mathbf{S}) = (g_S(S_1(\kappa)), \dots, g_S(S_n(\kappa))) = (S_1(\kappa), \dots, S_n(\kappa)) \quad (76)$$

The MGF parameterized by $\boldsymbol{\epsilon}$ is

$$\begin{aligned} M_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}) &= M_{g_S(\mathbf{S})}(\boldsymbol{\theta}) \\ &= M_{\mathbf{S}}(\boldsymbol{\theta}) \\ &= f_{\boldsymbol{\epsilon}}(e^{\boldsymbol{\theta}}) \\ &= \prod_{j=1}^n \left(\frac{1 + e^{-\theta_j}}{2} \right) \end{aligned} \quad (77)$$

which, from Lemmas (3) and (4), results in the closed-form expression for the normalizing constant of the WMM with the symmetric difference as

$$\begin{aligned} \Psi(\boldsymbol{\theta}) &= 2^n \prod_{j=1}^n \left(\frac{1 + e^{-\theta_j}}{2} \right) \\ &= \prod_{j=1}^n (e^{-\theta_j} + 1). \end{aligned} \quad (78)$$

□

5.3. Weighted Mallows Model Normalizing Constant with the Similarity Coefficient Distance

This section proves the existence of a closed-form expression for the normalizing constant of the WMM with the similarity coefficient.

Theorem 6. *The closed-form expression of the WMM normalizing constant with the similarity coefficient is*

$$\Psi(\boldsymbol{\theta}) = \prod_{j=1}^n \left(e^{-\frac{\theta_j}{n}} + 1 \right). \quad (79)$$

Proof. Let the element-wise distance vector be $\mathbf{M}(\kappa) = (M_1(\kappa, \kappa_0), \dots, M_n(\kappa, \kappa_0))$ where $M_j(\kappa, \kappa_0) = \frac{1}{n}$ if $\kappa(i) \odot \kappa_0(i)$ and 0 otherwise (i.e., a vector of similarity coefficient components). It follows that

$$d_M(\kappa, \kappa_0) = \sum_{j=1}^n M_j(\kappa) \quad (80)$$

Define the vector $\epsilon(\kappa)$ using the function $g_M(x) = \frac{x}{n}$,

$$\epsilon(\kappa) = g_M(\mathbf{M}) = (g_M(M_1(\kappa)), \dots, g_M(M_n(\kappa))) = \left(\frac{M_1(\kappa)}{n}, \dots, \frac{M_n(\kappa)}{n} \right) \quad (81)$$

The MGF parameterized by ϵ is

$$\begin{aligned} M_\epsilon(\theta) &= M_{g_M(\mathbf{M})}(\theta) \\ &= M_{\frac{\mathbf{M}}{n}}(\theta) \\ &= f_\epsilon(e^\theta) \\ &= \prod_{j=1}^n \left(\frac{1 + e^{-\frac{\theta_j}{n}}}{2} \right) \end{aligned} \quad (82)$$

which, from Lemmas (3) and (4), results in the closed-form expression for the normalizing constant of the WMM with the similarity coefficient as

$$\begin{aligned} \Psi(\theta) &= 2^n \prod_{j=1}^n \left(\frac{1 + e^{-\frac{\theta_j}{n}}}{2} \right) \\ &= \prod_{j=1}^n \left(e^{-\frac{\theta_j}{n}} + 1 \right). \end{aligned} \quad (83)$$

□

6. Discussion

The preceding sections derived the closed-form expressions for the MM and WMM. The derived expressions for the normalizing constant of the MM and WMM with different discrepancy functions are summarized in Table 1.

Table 1. Summary of derived normalizing constants.

Discrepancy Function	Normalizing Constant
Hamming distance	$\Psi(\theta) = (e^{-\theta} + 1)^n$
Symmetric difference	$\Psi(\theta) = (e^{-\theta} + 1)^n$
Similarity index	$\Psi(\theta) = \left(e^{-\frac{\theta}{n}} + 1 \right)^n$
Weighted Hamming distance	$\Psi(\theta) = \prod_{j=1}^n \left(e^{-\theta_j} + 1 \right)$
Weighted Symmetric difference	$\Psi(\theta) = \prod_{j=1}^n \left(e^{-\theta_j} + 1 \right)$
Weighted Similarity index	$\Psi(\theta) = \prod_{j=1}^n \left(e^{-\frac{\theta_j}{n}} + 1 \right)$

Consider a hypothetical combinatorial domain defined on the universal set $\mathcal{U} = \{a, b, c\}$. Given that the cardinality of the universal set is $|\mathcal{U}| = n$, the set \mathcal{U} spans 2^n possible combinations, i.e.,

$$\mathcal{P}(\mathcal{U}) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}.$$

Consider two sets $\kappa = \{a, b\}$ and $\kappa_0 = \{b\}$; the symmetric difference is

$$d_S(\kappa, \kappa_0) = |\kappa \Delta \kappa_0| = |\kappa_0 \setminus \kappa| + |\kappa \setminus \kappa_0| = 0 + 1 = 1.$$

The probability of sampling a set κ given the central tendency κ_0 is expressed as

$$P(\kappa|\kappa_0, \theta) = \frac{1}{\Psi(\theta)} e^{-\theta d(\kappa, \kappa_0)};$$

with an arbitrary value of $\theta = 1.5$,

$$\begin{aligned} P(\{a, b\}|\{b\}, 1.5) &= \frac{1}{(e^{-1.5} + 1)^3} e^{-1.5 \cdot 1} \\ &= 0.1219. \end{aligned}$$

Alternatively, if the similarity coefficient is used, i.e.,

$$d_M(\kappa, \kappa_0) = \frac{|\kappa \cap \kappa_0| + |\kappa^c \cap \kappa_0^c|}{n} = \frac{1+1}{3} = \frac{2}{3}.$$

Then the resultant probability is

$$\begin{aligned} P(\{a, b\}|\{b\}, 1.5) &= \frac{1}{\left(e^{-\frac{1.5}{3}} + 1\right)^3} e^{-1.5 \cdot \frac{2}{3}} \\ &= 0.0887. \end{aligned}$$

If the value of the dispersion parameter is arbitrary changed to $\theta = 5$ for the similarity coefficient calculation, the result is

$$\begin{aligned} P(\{a, b\}|\{b\}, 5) &= \frac{1}{\left(e^{-\frac{5}{3}} + 1\right)^3} e^{-5 \cdot \frac{1}{3}} \\ &= 0.0212. \end{aligned}$$

The effect of the dispersion parameter is further illustrated by the following two examples with the symmetric difference: given $\kappa = \{a, b\}$ and $\kappa_0 = \{a, b\}$, the probability of sampling the central tendency for a dispersion of $\theta = 0$ is

$$\begin{aligned} P(\{a, b\}|\{a, b\}, 0) &= \frac{1}{(e^{-0} + 1)^3} e^{-0 \cdot 0} \\ &= 0.125 = \frac{1}{2^3}. \end{aligned}$$

In contrast, for a dispersion of $\theta = 10$, the probability is

$$\begin{aligned} P(\{a, b\}|\{a, b\}, 10) &= \frac{1}{(e^{-10} + 1)^3} e^{-10 \cdot 0} \\ &= 0.9998 \approx 1.0. \end{aligned}$$

When the dispersion parameter is zero, the model behaves like a uniform distribution, with all subsets of $\mathcal{P}(\mathcal{U})$ equally likely to be sampled with a probability of $\frac{1}{2^n}$. As the dispersion parameter increases, there is a point at which it becomes certain that the central tendency is the only combination which can be sampled.

The combinatorial MM holds promise for application to a wide range of problems in the combinatorial domain. Many combinatorial decision problems are in the class of NP problems [30]. Combinatorial decision problems in NP often result in COPs, which do not have tractable exact solutions. For example, the graph coloring problem, Boolean satisfiability problem, job-shop scheduling problem, and the traveling salesman problem

are all examples of COPs which are in NP. As a result, approximation approaches such as metaheuristics are employed to find potentially optimal solutions in a reasonable amount of time.

The combinatorial MM is applicable to any problem for which there is an underlying combinatorial decision problem and which can be represented as a finite set. It is a popular approach in the literature to enhance the exploration and exploitation capabilities of metaheuristics through the inclusion of a probabilistic local search mechanism. For example, the popular particle swarm optimisation (PSO) metaheuristic has a variant which uses the Gaussian density function to sample new positions [31]. Similarly, the logistic chaotic function has been employed to improve the performance of differential evolution (DE) [32], and the Wigner semicircle distribution has been used to stochastically apply different local search techniques to evolutionary strategies (ESs) [33].

7. Conclusions and Future Work

This paper expanded the Mallows model (MM) and weighted Mallows model (WMM) for use in combinatorial domains. Previously, the computational bottleneck for probability calculation and combination sampling has been the normalizing constant. The requirement to sum over all possible subsets of elements in the universal set made it computationally infeasible to use the MM to sample combinations in relation to a central tendency.

The closed-form expressions for the normalizing constant with the Hamming distance, symmetric difference, and similarity coefficient (Rand index) makes the MM and WMM computationally feasible for use on combinatorial problems and provides a platform for solving problems defined using set-valued representations. Undoubtedly, the most natural measurement of discrepancy in combinatorial domains remains the Jaccard distance. Unfortunately, the closed-form expression for the Jaccard distance remains unobtainable due to the lack of an additive decomposition form; however, researchers are invited to attempt a more efficient expression for the normalizing constant with the Jaccard distance.

Additionally, future work should concentrate on efficient sampling methods for the MM and determine whether existing approaches like Gibbs sampling are appropriate. Further, the parameter learning of the combinatorial MM should be explored, for example, a thorough investigation into the maximum likelihood estimation (MLE) of the parameters.

Author Contributions: Conceptualization, J.-P.v.Z. and A.P.E.; methodology, J.-P.v.Z.; formal analysis, J.-P.v.Z.; writing—original draft preparation, J.-P.v.Z.; writing—review and editing, A.P.E.; supervision, A.P.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Berge, C. *Principles of Combinatorics*; Academic Press: New York, NY, USA, 1971.
2. Biggs, N.L. The roots of combinatorics. *Hist. Math.* **1979**, *6*, 109–136. [\[CrossRef\]](#)
3. Cameron, P.J. *Permutation Groups*; Number 45; Cambridge University Press: Cambridge, UK, 1999.
4. Nixon, R. *Amending the Selective Service Regulations to Prescribe Random Selection*; Executive Order 11497, Number 229; Federal Register: Washington, DC, USA, 1969; Volume 34.
5. Fienberg, S.E. Randomization and Social Affairs: The 1970 Draft Lottery. *Science* **1971**, *171*, 255–261. [\[CrossRef\]](#)
6. Hall, M.; Knuth, D.E. Combinatorial Analysis and Computers. *Am. Math. Mon.* **1965**, *72*, 21–28. [\[CrossRef\]](#)
7. Papadimitriou, C.H.; Steiglitz, K. *Combinatorial Optimization: Algorithms and Complexity*; Courier Corporation: New York, NY, USA, 2013.

8. Chen, S.; Lin, T.; King, I.; Lyu, M.R.; Chen, W. Combinatorial pure exploration of multi-armed bandits. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 379–387.
9. Robins, G.; Pattison, P.; Kalish, Y.; Lusher, D. An introduction to exponential random graph (p^*) models for social networks. *Soc. Netw.* **2007**, *29*, 173–191. [\[CrossRef\]](#)
10. Liang, F. A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants. *J. Stat. Comput. Simul.* **2010**, *80*, 1007–1022. [\[CrossRef\]](#)
11. Park, J.; Haran, M. Bayesian inference in the presence of intractable normalizing functions. *J. Am. Stat. Assoc.* **2018**, *113*, 1372–1390. [\[CrossRef\]](#)
12. Mallows, C.L. Non-null ranking models. I. *Biometrika* **1957**, *44*, 114–130. [\[CrossRef\]](#)
13. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [\[CrossRef\]](#)
14. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93. [\[CrossRef\]](#)
15. Bradley, R.A.; Terry, M.E. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **1952**, *39*, 324–345. [\[CrossRef\]](#)
16. Diaconis, P. *Group Representations in Probability and Statistics*; Lecture Notes-Monograph Series; Institute of Mathematical Statistics: Hayward, CA, USA, 1988; Volume 11, pp. i–vi+1–192.
17. Fligner, M.A.; Verducci, J.S. Distance based ranking models. *J. R. Stat. Soc. Ser. B (Methodol.)* **1986**, *48*, 359–369. [\[CrossRef\]](#)
18. Spearman, C. Footrule for measuring correlation. *Br. J. Psychol.* **1906**, *2*, 89. [\[CrossRef\]](#)
19. Hamming, R.W. Error detecting and error correcting codes. *Bell Syst. Tech. J.* **1950**, *29*, 147–160. [\[CrossRef\]](#)
20. Ulam, S.M. Monte Carlo calculations in problems of mathematical physics. *Mod. Math. Eng.* **1961**, *261*, 281.
21. Levenshtein, V.I. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet Physics-Doklady*; American Institute of Physics: New York, NY, USA, 1966; Volume 10, pp. 707–710.
22. Irurozki, E. Sampling and Learning Distance-Based Probability Models for Permutation Spaces. Ph.D. Thesis, Universidad del País Vasco-Euskal Herriko Unibertsitatea, Biscay, Spain, 2014.
23. Cayley, A. Note on the theory of permutations. *Lond. Edinburgh Dublin Philos. Mag. J. Sci.* **1849**, *34*, 527–529. [\[CrossRef\]](#)
24. Jaccard, P. Distribution de la Flore Alpine dans le Bassin des Dranses et dans quelques régions voisines. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 241–272.
25. Rand, W.M. Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.* **1971**, *66*, 846–850. [\[CrossRef\]](#)
26. Goodall, D.W. The distribution of the matching coefficient. *Biometrics* **1967**, *23*, 647–656. [\[CrossRef\]](#)
27. Sokal, R.R.; Michener, C.D. *A Statistical Method for Evaluating Systematic Relationships*; University of Kansas Scientific Bulletin: Lawrence, KS, USA, 1958.
28. Esscher, F. On the probability function in the collective theory of risk. *Scand. Actuar. J.* **1932**, *1932*, 175–195. [\[CrossRef\]](#)
29. Lehmann, L. *Theory of Point Estimation*; A Wiley Publication in Mathematical Statistics; Wiley: Hoboken, NJ, USA, 1983.
30. Karp, R.M. On the computational complexity of combinatorial problems. *Networks* **1975**, *5*, 45–68. [\[CrossRef\]](#)
31. Krohling, R.A. Gaussian swarm: A novel particle swarm optimization algorithm. In Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems, Singapore, 1–3 December 2004; IEEE: New York, NY, USA, 2004; Volume 1, pp. 372–376.
32. Jia, D.; Zheng, G.; Khan, M.K. An effective memetic differential evolution algorithm based on chaotic local search. *Inf. Sci.* **2011**, *181*, 3175–3187. [\[CrossRef\]](#)
33. Lara, A.; Sanchez, G.; Coello, C.A.C.; Schutze, O. HCS: A new local search strategy for memetic multiobjective evolutionary algorithms. *IEEE Trans. Evol. Comput.* **2009**, *14*, 112–132. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.