Multivariate Spearman’s rho for aggregating ranks using copulas

Justin Bedő
IBM Research, Level 5, 204 Lygon St, Carlton, VIC 3053, Australia
The Department of Computing and Information Systems, The University of Melbourne, Australia
BEDOJ@UNIMELB.EDU.AU

Cheng Soon Ong
NICTA Canberra Research Laboratory, Australia
Department of Electrical and Electronic Engineering, the University of Melbourne, Australia
CHENGSOON.ONG@ANU.EDU.AU

Abstract

We study the problem of rank aggregation: given a set of ranked lists, we want to form a consensus ranking. Furthermore, we consider the case of extreme lists: i.e., only the rank of the best or worst elements are known. We impute missing ranks by the average value and generalise Spearman’s \( \rho \) to extreme ranks. Our main contribution is the derivation of a non-parametric estimator for rank aggregation based on multivariate extensions of Spearman’s \( \rho \), which measures correlation between a set of ranked lists. Multivariate Spearman’s \( \rho \) is defined using copulas, and we show that the geometric mean of normalised ranks maximises multivariate correlation. Motivated by this, we propose a weighted geometric mean approach for learning to rank which has a closed form least squares solution. When only the best or worst elements of a ranked list are known, we impute the missing ranks by the average value, allowing us to apply Spearman’s \( \rho \). Finally, we demonstrate good performance on the rank aggregation benchmarks MQ2007 and MQ2008.

1. Introduction

Ranking is a central task in many applications such as information retrieval, recommender systems and bioinformatics. It may also be a subtask of other learning problems such as feature selection, where features are scored according to their predictiveness, and then the most significant ones are selected. One major advantage of ranks over scores is that the resulting predicted ranks are automatically normalised and hence can be used to combine diverse sources of information. However, unlike many other supervised learning problems, the problem of learning to rank [Lebanon and Mao (2008), Liu (2011)] does not have the simple one example one label paradigm. This has led to many formulations of learning tasks, depending on what label information is available, including pairwise ranking, listwise ranking and rank aggregation.

This paper considers a novel formulation of rank aggregation based on multivariate extensions to Spearman’s \( \rho \). For a set of \( n \) objects from the domain \( \Omega \), we are given a set of \( d \) experts that rank these objects providing ranks \( R_1, \ldots, R_d \). Each rank is a permutation of the \( n \) objects, and can be represented as a vector of unique integers from 1 to \( n \). The problem of rank aggregation is to construct a new vector \( R \) that is most similar to the set of \( d \) ranks provided by the experts. In this paper we use Spearman’s correlation \( \rho \), a widely used correlation measure for ranks [Spearman (1904)]. Instead of decomposing the association into a combination of pairwise similarities, \( \rho(R, R_1), \rho(R, R_2), \ldots, \rho(R, R_d) \), we...
directly maximise the multivariate correlation

\[ R^* = \arg \max_R \rho(R, R_1, R_2, \ldots, R_d). \]

Measures of association such as Spearman’s \( \rho \) capture the concordance between random variables (Nelsen, 2006). Informally, random variables are concordant if large values of one tend to be associated with large values of the other. Let \((x_i, y_i)\) and \((x_j, y_j)\) be two observations from a vector \((X, Y)\) of continuous random variables. We say that \((x_i, y_i)\) and \((x_j, y_j)\) are concordant if \(x_i < x_j\) and \(y_i < y_j\) or if \(x_i > x_j\) and \(y_i > y_j\). If the inequalities disagree, we say that the samples are discordant. The concept of concordance captures only the order of the random variables, and is invariant to their values, and therefore is ideal for analysing ranks. As will be described in Section 3.3, the multivariate version of Spearman’s correlation is based on the difference between the concordance and discordance of the samples.

In short, Spearman’s correlation can be defined as the concordance \( Q \) between the copula \( C \) corresponding to the data and the independent copula \( \pi \)

\[ \rho \propto Q(C, \pi). \]

We review the concept of copulas in Section 2 and derive our generalisation of concordance in Section 3.

While the mathematical machinery to derive our proposed algorithm relies on constructions which may not be familiar to some machine learners, the resulting algorithm for rank aggregation is straightforward. We solve a least squares problem for \( n \) items,

\[
\min_\omega \sum_{x=1}^{n} \left( l(x) - \sum_{d=1}^{d} \omega_d r_d(x) \right)^2,
\]

where we minimise the weights \( \omega_1, \ldots, \omega_d \) corresponding to the \( d \) experts. It turns out that the appropriate transformation to learn weights between experts is to use logarithmic scaled ranks. In the above equation, \( l(x) \) and \( r(x) \) denote the logarithm of the labels and individual expert ranks respectively, with all ranks normalised uniformly to the interval \((0, 1)\). Since it is a least squares problem, there is a closed form solution for the optimal weights. This is in contrast to previous approaches to rank aggregation which involve complex optimisation methods or sampling.

1.1 Our contributions

In the rest of this paper we will theoretically justify why the above least squares problem provides a meaningful way to weight experts. We show that the geometric mean of a set of normalised ranks maximises multivariate Spearman’s \( \rho \). This motivates our method which finds a setting of weights that maximise multivariate Spearman’s \( \rho \) for a specific target (supervised rank aggregation).

As previously mentioned, in many applications of rank aggregation, only extreme ranks are available, whereas the standard definitions of Spearman’s \( \rho \) require full ranks. For practical problems, the expert may only rank the most liked (top-k) or most disliked (bottom-k) objects where \( k \) can be different for each expert. We propose a method for estimating Spearman’s \( \rho \) for extreme ranks by imputing the remaining ranks. We describe this method and show that it is an unbiased estimator in Section 4.
This results in a non-parametric approach for rank aggregation that learns the weights of experts by solving a least squares problem. In Section 5, we describe our empirical results for rank aggregation and show that our simple algorithm performs better than current state-of-the-art results.

### 1.2 Related Work

There are two related rank aggregation tasks: score-based rank aggregation and order-based rank aggregation. For score-based rank aggregation, objects are associated with scores, while for order-based rank aggregation, only the relative order of objects is available. There has been recent work on combining both scores and ranks (Sculley, 2010; Iyer and Bilmes, 2013). We consider the learning task referred to as the listwise approach in Liu (2011), where the input is a set of ranked lists of documents from multiple experts, and the learner has to predict the final ranks. Numerous proposals for solving the problem of combining multiple lists into a single list are surveyed in Liu (2011). Niu et al. (2012) has focused on learning a good ranking from given features.

Spearman’s $\rho$ is a natural measure of similarity for distributions of permutations (Mallows, 1957; Fligner and Verducci, 1986). Interestingly, there has not been much work using Spearman’s $\rho$ for dealing with ranked data, but instead the focus has been on Kendall’s $\tau$. One difficulty of inference with the Mallows model for Spearman’s $\rho$ is that it involves estimating the permanent of a matrix.

Previous approaches (Klementiev et al., 2008; Iyer and Bilmes, 2012) to rank aggregation considers pairwise comparisons between ranked lists. We prove a result saying that the geometric mean of normalised ranks maximise Spearman’s $\rho$ (Theorem 15), which is similar in spirit to the result in Iyer and Bilmes (2012), which shows that for Lovasz-Bregman divergences the best aggregator is the arithmetic mean. Our work builds heavily on copula theory, and we use results from Nelsen (2006). Brief introductions to copulas can be found in Trivedi and Zimmer (2005); Genest and Favre (2007). Further details are available in Joe (2014).

The work on partial ranks goes back to at least Critchlow (1985), who describes the rank aggregation task in terms of distances between rankings. A good review of probability and statistics applied to permutations is Diaconis (1988).

### 2. Copulas

Copulas are functions from the unit hypercube to the unit interval (Elidan, 2013). In this section we briefly review the bivariate setting, in preparation for the multivariate setting in the next section. The expert reader may skip directly to section 3 to see the definition of multivariate Spearman’s $\rho$ in terms of the multivariate copula.

#### 2.1 Definition of Copulas

Intuitively, for continuous random variables copulas model the dependence component of a multivariate distribution after discounting for univariate marginal effects. We let $\mathbb{R}$ denote the ordinary real line $(-\infty, \infty)$, and $\overline{\mathbb{R}}$ denote the extended real line $[-\infty, \infty]$. The following algebraic definition of bivariate copulas is generalised to the multivariate setting in section 3. It essentially constrains copulas to be functions that are “monotonically increasing” along each dimension as well as towards the diagonal of the volume.

**Definition 1** Let $A_1$ and $A_2$ be nonempty subsets of $\overline{\mathbb{R}}$, and let $H(\cdot, \cdot)$ be a real function such that the domain of $H = A_1 \times A_2$. Let $B = [x_1, x_2] \times [y_1, y_2]$ be a rectangle all of whose vertices are in $\overline{\mathbb{R}}$. Then $H$ is a copula if and only if:

1. **Monotonicity:** $H(z_1, z_2) \leq H(y_1, z_2)$ and $H(z_1, z_2) \leq H(z_1, y_2)$ for all $z_1 \leq y_1$ and $z_2 \leq y_2$.
2. **Boundary Conditions:** $H(z_1, \cdot) = z_1$ and $H(\cdot, z_2) = z_2$ for all $z_1 \in A_1$ and $z_2 \in A_2$.
3. **Continuity:** $H$ is continuous on $A_1 \times A_2$.

Copulas are used to model the dependence structure of random variables and are fundamental in the study of multivariate distributions. They provide a way to separate the marginal distributions from the dependence structure, allowing for flexible modeling of complex data dependencies.
domain of \( H \). Then the \( H \)-volume of \( B \) is given by:
\[
V_H(B) = H(x_1,y_1) + H(x_2,y_2) - H(x_1,y_2) - H(x_2,y_1).
\]

**Definition 2** A real function \( H(\cdot, \cdot) \) is 2-increasing if its \( H \)-volume is non-negative, that is \( V_H(B) \geq 0 \) for all rectangles \( B \) whose vertices lie in the domain of \( H \).

**Definition 3** A copula is a function \( C : [0,1]^2 \to [0,1] \) with the following properties:

1. For every \( u, v \in [0,1] \),
   \[
   C(u,0) = 0 = C(0,v) \\
   C(u,1) = u \quad \text{and} \quad C(1,v) = v
   \]
2. \( C \) is 2-increasing.

### 2.2 Relation between bivariate CDFs and Copulas

Sklar’s theorem is central to the theory of copulas and is the foundation of many applications in statistics. Actually, Sklar’s theorem can be defined for general distribution functions, with nothing to do with probability. However, since we are interested in statistical applications we will consider cumulative distribution functions.

**Theorem 4 (Sklar’s theorem)** Let \( H(\cdot, \cdot) \) be a cumulative distribution function with marginals \( F(\cdot) \) and \( G(\cdot) \). Then there exists a copula \( C : [0,1]^2 \to [0,1] \) such that for all \( x,y \) in \( \mathbb{R} \),
\[
H(x,y) = C(F(x), G(y)).
\]

If \( F(\cdot) \) and \( G(\cdot) \) are continuous then \( C(\cdot, \cdot) \) is unique; otherwise \( C(\cdot, \cdot) \) is uniquely determined on the ranges of \( F(\cdot) \) and \( G(\cdot) \).

Conversely, if \( C(\cdot, \cdot) \) is a copula and \( F(\cdot) \) and \( G(\cdot) \) are cumulative distribution functions then the function \( H(\cdot, \cdot) \) is a bivariate cumulative distribution function with marginals \( F(\cdot) \) and \( G(\cdot) \).

### 3. Spearman’s rho

We briefly review the bivariate model to lay out the approach for estimating the copula using data, the so-called empirical copula.

#### 3.1 Empirical bivariate Spearman’s \( \rho \)

Let \( R \) and \( S \) be ranking functions, which are bijections mapping elements \( x \) in the domain \( U \) to \([1,2,\ldots,n]\). The domain \( U \) represents the space of objects that we are interested in ranking, such as documents retrieved in response to a query or the biomarkers most associated with a disease. Since we consider only the ranks of the object \( R(x) \) and \( S(x) \), the actual domain \( \Omega \) does not affect the analysis. The sums below are over the \( n \) objects \( x \). Similar to the approach of Pearson’s correlation for the measure of dependence, Spearman’s \( \rho \) is a measure of correlation between ranks, empirically given by:
\[
\rho_n = \frac{\sum_x (R(x) - \bar{R})(S(x) - \bar{S})}{\sqrt{\sum_x (R(x) - \bar{R})^2 \sum_x (S(x) - \bar{S})^2}}
\]
\[(1)\]
where \( \bar{R} := \frac{1}{n} \sum_x R(x) \) and \( \bar{S} := \frac{1}{n} \sum_x S(x) \) are the empirical means of the respective random variables. This is equivalent to applying Pearson’s correlation to the ranks instead of the values of the score function itself. There is no direct way to generalise this expression to more than two ranking functions, but as we shall see in Section 3.3 we can obtain an expression via the copula.

By substituting the definitions of the empirical means and rearranging the terms, we obtain

\[
\rho_n = \left( \frac{n+1}{n-1} \right) \left[ \frac{12}{n} \sum_x \frac{R(x) S(x)}{n+1} - 3 \right]
\]

The constants 12 and 3 seem strange, but are a natural consequence of the mean and variance of a list of ranks. As we will see later, these constants are dependent only on the dimension of the copula.

Similar to the definition of an empirical CDF, we define an empirical copula as:

\[
C_n(u, v) = \frac{1}{n} \sum_x \mathbf{1} \left( R(x) \leq u, S(x) \leq v \right),
\]

where \( \mathbf{1} \) is the indicator function. This allows us to re-express the form of \( \rho_n \) above in terms of an integral over the unit square,

\[
\rho_n = \left( \frac{n+1}{n-1} \right) \left[ \frac{12}{n} \sum_x \frac{R(x) S(x)}{n+1} - 3 \right] = \left( \frac{n+1}{n-1} \right) \left[ 12 \int_{[0,1]^2} u v C_n(u, v) - 3 \right]
\]

It can be shown (Nelsen, 2006; Genest and Favre, 2007) that \( \rho_n \) is an asymptotically unbiased estimator of

\[
\rho = 12 \int_{[0,1]^2} C(u, v) \, du \, dv - 3,
\]

where \( C \) is the population version of \( C_n \).

### 3.2 Multivariate copulas

We now generalise the definitions in Section 2.1 to the multivariate case. The concepts are essentially the same, constraining the copula to be “monotonically increasing” in the interval \([0, 1]\) and also towards the center of the volume.

**Definition 5** Let \( A_j \) be nonempty subsets of \( \mathbb{R} \) for \( j = 1, \ldots, d \), and let \( H_d : A_1 \times \cdots \times A_d \rightarrow \mathbb{R} \). Let \( B = [a_1, b_1] \times \cdots \times [a_d, b_d] \) be the \( d \)-box where all vertices are contained in \( \text{Dom} \, H_d \). Then the \( H_d \)-volume of \( B \) is the \( d \)-th order difference:

\[
V_{H_d}(B) = \Delta^{b_d}_{a_d} \cdots \Delta^{b_1}_{a_1} H_d(\vec{t}),
\]

where

\[
\Delta^{b_i}_{a_i} H(\vec{t}) = H_d(t_1, \ldots, t_{i-1}, b_i, t_{i+1}, \ldots, t_d) - H_d(t_1, \ldots, t_{i-1}, a_i, t_{i+1}, \ldots, t_d)
\]

**Definition 6** A real function \( H_d \) is grounded if \( H_d(\vec{t}) = 0 \) for all \( t \in \text{Dom} \, H_d \) such that \( t_j = a_j \) for at least one \( j \in \{1, \ldots, d\} \).
Definition 7 A real function $H_d$ is $d$-increasing if $V_{H_d}(B) \geq 0$ for all $n$-boxes $B$ whose vertices lie in the domain of $H$.

Definition 8 A multivariate copula has the following properties:

1. $\text{Dom}C = [0, 1]^d$
2. $C$ has margins $C_j(u) = C(1, \ldots, 1, u, 1, \ldots, 1) = u$ for all $j$ and $u \in I$
3. $C$ is grounded
4. $C$ is $d$-increasing.

There is an alternative probabilistic definition which may be more familiar to readers with a statistical background.

Definition 9 Let $U_1, \ldots, U_d$ be real uniformly distributed random variables on the unit interval $\sim U([0, 1])$. A copula function $C : [0, 1]^d \rightarrow [0, 1]$ is a joint distribution $C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d)$.

Let $X \sim F$ be a continuous random variable such that the inverse of the CDF $F^{-1}$ exists. What is the distribution of $F(x) = P(X \leq x)$?

$$P(F(X) \leq u) = P(F^{-1}(F(X)) \leq F^{-1}(u)) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u$$

The above calculation shows that the distribution is uniform, i.e. $F(x) \sim U([0, 1])$. This can be considered to be the “copula trick”, as the user has the freedom to choose the copula independently of the marginal distributions.

3.3 Multivariate extension of Spearman’s $\rho$

We generalise the concept of concordance to the multivariate setting such that we can define multivariate Spearman’s $\rho$ in an analogous way to the bivariate $\rho$ as defined in [Nelsen(2006)]. Further details of multivariate concordance can be found in [Taylor(2007)].

Recall that two random variables are concordant if they tend to be in the same order, that is $(x_i, y_i)$ and $(x_j, y_j)$ are concordant if $(x_i - x_j)(y_i - y_j) > 0$, and are discordant if $(x_i - x_j)(y_i - y_j) < 0$. The concordance function $Q$ denotes the difference between the probabilities of concordance and discordance, and as the following theorem shows, can be expressed in terms of the copulas.

Theorem 10 (Concordance function) Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be two independent vectors with joint distributions $H_1(x, y) = C_1(F(x), G(y))$ and $H_2(x, y) = C_2(F(x), G(y))$ respectively. Then the concordance function $Q$ is given by

$$Q(C_1, C_2) := P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] = 4 \int_{[0,1]^2} C_2(u, v) \, dC_1(u, v) - 1$$
Proof See [Nelsen (2006)].

We generalize the proof in [Nelsen (2006)] to obtain a similar expression for $d$ random variables. First, two lemmas are required for the proof of the multivariate concordance result.

**Lemma 11**

1. There are $2^{d-1}$ possible configurations that result in a positive product $x_1 x_2 x_3 \ldots x_d$ where $x_i \in \mathbb{R}$.

2. There are $2^{d-1}$ possible configurations that result in a negative product $x_1 x_2 x_3 \ldots x_d$ where $x_i \in \mathbb{R}$.

**Proof**

Proof is by induction. For the base case when $d = 1$, $x_1$ has only one possible option to be positive ($x_1 > 0$). Similarly, $x_1$ has only one way to be negative ($x_1 < 1$). Now for the inductive case. Assume true for $x_1 x_2 x_3 \ldots x_{d-1}$, therefore there are $2^{d-2}$ ways to form either a positive or negative product. Let us consider the case of the positive product, $x_1 x_2 x_3 \ldots x_d > 0$. If $x_d$ is positive then for the product to be positive $x_1 x_2 x_3 \ldots x_{d-1}$ must be positive. Similarly, if $x_d$ is negative then for the product to be positive $x_1 x_2 x_3 \ldots x_{d-1}$ must be negative. By the inductive assumption, there are $2^{d-2}$ ways to form a positive product and $2^{d-2}$ ways to form a negative product for $x_1 x_2 x_3 \ldots x_{d-1}$, therefore there are $2^{d-2} + 2^{d-2} = 2 \times 2^{d-2} = 2^{d-1}$ ways to form a positive product for $x_1 x_2 x_3 \ldots x_{d-1}$. The negative case follows similarly.

An alternative way to consider the above lemma is by observing that there are $2^d$ possible configurations of signs for individual $x_i$ in the product $x_1 x_2 x_3 \ldots x_d$. Half of the configurations, when there are an even number of positive $x_i$, result in a positive product. The other half result in a negative product.

The following theorem relates positive products to integrals of copulas.

**Lemma 12** Each possible positive product such as $P[X_1 > Y_1, X_2 > Y_2, \ldots, X_d > Y_d]$ can be expressed as the integral $\int_{[0,1]^d} C_X(u) \ dC_Y(u)$, where $C_X$ and $C_Y$ denote the multivariate copulas of samples $X$ and $Y$ respectively.

**Proof** By theorem [11] we know that positive products can only be formed by changing the order of pairs, thus the proof reduces to showing

$$P[X_1 > Y_1, X_2 > Y_2, X_3 > Y_3, \ldots, X_d > Y_d] = P[X_1 < Y_1, X_2 < Y_2, X_3 > Y_3, \ldots, X_d > Y_d]$$

with the arbitrary choice of changing the order of the first pair of random variables. We have:

$$P[X_1 > Y_1, X_2 > Y_2, X_3 > Y_3, \ldots, X_d > Y_d] = \int_{[0,1]^d} P[X_1 \leq y_1, X_2 \leq y_2, \ldots, X_d \leq y_d \ dC_Y(u)$$

$$= \int_{[0,1]^d} C_X(u) \ dC_Y(u)$$
by definition of the $d$-copula. Similarly,

$$P[X_1 < Y_1, X_2 < Y_2, X_3 > Y_3 \ldots, X_d > Y_d] = \int_{[0,1]^d} 1 - u_1 - u_2 + P[X_1 \leq y_1, X_2 \leq y_2, \ldots, X_d \leq y_d] \, dC_Y(u)$$

$$= \int_{[0,1]^d} 1 - u_1 - u_2 + C_X(u) \, dC_Y(u).$$

However, $E(u_1) = E(u_2) = \frac{1}{2}$ as $u$ is drawn from the uniform distribution. Therefore

$$P[X_1 < Y_1, X_2 < Y_2, X_3 > Y_3 \ldots, X_d > Y_d] = \int_{[0,1]^d} 1 - u_1 - u_2 + C_X(u) \, dC_Y(u)$$

$$= \int_{[0,1]^d} C_X(u) \, dC_Y(u)$$

$$= P[X_1 > Y_1, X_2 > Y_2, X_3 > Y_3 \ldots, X_d > Y_d].$$

All other positive terms follow by induction.

We now state our generalisation of concordance to the multivariate case.

**Theorem 13 (Multivariate concordance)** Let $(X_1, \ldots, X_d)$ and $(Y_1, \ldots, Y_d)$ be two independent $d$-vectors with joint distributions $C_X(F(x))$ and $C_Y(F(y))$ where $F(x) = F_1(x_1), \ldots, F_d(x_d)$ and $F(y) = F_1(y_1), \ldots, F_d(y_d)$ are the marginal distributions, and $C_X, C_Y$ are the respective $d$ copulas. Then the concordance function $Q$ is given by

$$Q(C_X, C_Y) := P \left[ \prod_{j=1}^{d} (X_j - Y_j) > 0 \right] - P \left[ \prod_{j=1}^{d} (X_j - Y_j) < 0 \right]$$

$$= 2^d \int_{[0,1]^d} C_X(u) \, dC_Y(u) - 1.$$

**Proof** Following similar arguments to [Nelsen (2006)] for the bivariate case, as $P[\prod_{j=1}^{d} (X_j - Y_j) > 0] = 1 - P[\prod_{j=1}^{d} X_j - Y_j < 0]$ we have

$$Q(C_X, C_Y) = P \left[ \prod_{j=1}^{d} (X_j - Y_j) > 0 \right] - P \left[ \prod_{j=1}^{d} (X_j - Y_j) < 0 \right] = 2P \left[ \prod_{j=1}^{d} (X_j - Y_j) > 0 \right] - 1. \quad (2)$$

By theorem 11 there are $2^d - 1$ possible ways to form a positive product $\prod_{j=1}^{d} (X_j - Y_j) > 0$. One possibility is when all the differences are positive, that is

$$P[X_1 > Y_1, X_2 > Y_2, \ldots, X_d > Y_d].$$

By the definition of the $d$ copula, this probability can be expressed as the integral,

$$P[X_1 > Y_1, X_2 > Y_2, \ldots, X_d > Y_d] = \int_{[0,1]^d} C_X(u) \, dC_Y(u).$$

As shown in theorem 12 each of the possible $2^d - 1$ configurations that form a positive product can be expressed in the same way. Hence the probability of the product is

$$P \left[ \prod_{j=1}^{d} (X_j - Y_j) > 0 \right] = 2^{d-1} \int_{[0,1]^d} C_X(u) \, dC_Y(u).$$
By using the expression of $Q(C_X, C_Y)$ in Equation 2, we obtain the desired result.

In terms of the concordance function, Spearman’s $\rho$ is given by the concordance between the copula $C$ and the independent copula $\pi$. For example in two dimensions $\pi(u, v) = uv$. There are several possible generalisations of Spearman’s $\rho$ to the multivariate case, all of which are equivalent in the bivariate case. We consider the generalisation $\rho_2$ in Schmid and Schmidt (2007)

$$Q(C, \pi) = h(d) \left[ 2^d \int_{[0,1]^d} \pi(u) \, dC(u) - 1 \right],$$

where $h(d)$ is the normalisation factor discussed below. When $d = 2$, this is equivalent to

$$Q(C, \pi) = h(d) \left[ 2^d \int_{[0,1]^2} C(u) \, d\pi(u) - 1 \right] = h(d) \left[ 2^d \int_{[0,1]^2} C(u) \, du - 1 \right].$$

There are thus several possible generalisations of Spearman’s $\rho$ to the multivariate case, two of which are based on (3) and (4). We consider the generalisation of (3), which is equivalent to the $\rho_2$ in Schmid and Schmidt (2007). Note that in the bivariate case, both integrals are equivalent, resulting in the bivariate expression in section 2.1.

The scaling factor $h(d)$ can be derived using theorem 13. The Fréchet-Hoeffding bounds represent the maximum (minimum) correlations possible at full dependence. Thus, for Spearman’s $\rho$, this is the concordance between the maximum copula $M$ and the independent copula $\pi$. In two dimensions the maximum copula is $M(u, v) = \min(u, v)$, with $M(u)$ below defined correspondingly for $d$ dimensions.

$$h(d) = 1/Q(M, \pi) = \frac{1}{2^d \int_{[0,1]^d} \pi(u) \, dM(u) - 1}$$

$$= \frac{1}{2^d \int_0^1 u^d \, du - 1}$$

$$= \frac{d + 1}{2^d - (d + 1)}.$$  \hspace{1cm} (5)

Recall that for a set of $n$ objects from the domain $\Omega$, we are given a set of $d$ experts that rank these objects providing ranks $R_1, \ldots, R_d$, where each $R_j$ is a bijection to $(0, 1)$. Putting (3) and (6) together, we obtain the following expression for multivariate Spearman’s correlation:

$$\rho(R_1, \ldots, R_d) = Q(C, \pi) = \frac{d + 1}{2^d - (d + 1)} \left[ 2^d \int_{[0,1]^d} \pi(u) \, dC(u) - 1 \right].$$

In practice, we do not have access to the population version of the copula $C(u)$ but have the empirical copula $C_n(u)$. We discuss this further in Section 4.2.

4. Optimal aggregation with Spearman’s $\rho$

The empirical copula requires $R$ and $S$ to comprise of ranks for the same set of elements, that is $\text{Dom} \, R = \text{Dom} \, S$. Recall from Section 3.1 that ranks map to the range $\{1, \ldots, n\}$, but the empirical
copula is expressed in terms of fractional ranks (divided by $n + 1$). In the following it is convenient to work with normalised ranks, that is to consider $R$ and $S$ as bijections to $(0, 1)$. The expression for the empirical copula then simplifies to

$$C_n(u, v) = \frac{1}{|\Omega|} \sum_{x \in \Omega} 1(R(x) \leq u, S(x) \leq v),$$

where $\Omega$ is the domain of the objects we are interested in ranking. Correspondingly, the $d$ dimensional empirical copula for $n$ objects given by

$$C_n(u) = \frac{1}{n} \sum_{x} \prod_{j=1}^{d} 1(R_j(x) \leq u_j),$$

where $R_1(x), \ldots, R_d(x)$ is the rankings of the $d$ experts. Plugging the empirical copula (9) expression into Spearman’s $\rho$ (7), and observing that integrating the product over the copula is the product of the ranks Schmid and Schmidt (2007), we obtain an empirical expression for multivariate Spearman’s correlation:

$$\rho_n(R_1, \ldots, R_d) = h(d) \left[ \frac{2d}{n} \sum_{x} \prod_{j=1}^{d} R_j(x) - 1 \right].$$

4.1 Geometric mean is optimal

We are now in a position to derive the deceptively simple result: the ranking $R$ that maximises correlation with a given set of ranks $\{R_1, \ldots, R_d\}$ is given by the geometric mean of $R_1, \ldots, R_d$. The following definition is needed to capture the notion that ranks only depend on the order.

**Definition 14 (Rank generator)** $\sigma : \mathbb{R}^{\Omega} \to [0, 1]^{\Omega}$ is a rank generator if:

- for all $x, y \in \Omega$ and $R$ with domain $\Omega$, $R(x) < R(y) \iff \sigma \circ R(x) < \sigma \circ R(y)$;
- for any ranks $R, R'$ with domain $\Omega$ there exists a permutation $\xi$ such that $\sigma \circ R' = \sigma \circ \xi \circ R$;
- for any permutation $\xi$, $\xi \circ \sigma = \sigma \circ \xi$.

A rank generator formalises the idea of generating a rank: the ranks it generates must be invariant to scale and only dependent on the ordering of elements. The standard ranking functions from statistics such as fractional ranking and dense ranking fit into this framework.

**Proposition 15** Let $\{R_1, R_2, \ldots, R_d\}$ be a set of ranks with common domain $\Omega$ and $\sigma$ be a rank generator. Then

$$\arg \max_{R \in \text{codom} \sigma} \rho_n(R, R_1, R_2, \ldots, R_d) = \sigma \left( \prod_{j=1}^{d} R_j \right).$$

**Proof** Consider the expression for Spearman’s $\rho_n$ (10):

$$\rho_n(R, R_1, R_2, \ldots, R_d) = h(d + 1) \left[ \frac{2d}{n} \sum_{x} \left( R(x) \prod_{j=1}^{d} R_j(x) \right) - 1 \right]$$
Focusing on the terms in the sum, showing that the best possible $R(x)$ is $\prod_{j=1}^{d} R_j(x)$ reduces to showing

$$\sum_{x \in U} \sigma \circ P(x) P(x)$$

is maximal, where $P := \prod_{j} R_j$. Suppose there exists an $P'$ such that

$$\sum_{x \in U} \sigma \circ P'(x) P(x) > \sum_{x \in U} \sigma \circ P(x) P(x).$$

By definition of $\sigma$, there exists a permutation $\xi$ such that

$$\sum_{x \in U} \sigma \circ P'(x) P(x) = \sum_{x \in U} \sigma \circ \xi \circ P(x) P(x)$$

$$= \sum_{x \in U} \xi \circ \sigma \circ P(x) P(x)$$

$$> \sum_{x \in U} \sigma \circ P(x) P(x).$$

This however is a contradiction for any permutation $\xi$ as $\sigma$ is order preserving.

**4.2 Empirical copulas with partial lists**

In many applications it is prohibitive to obtain complete annotations of the object ranks. For example, in the document retrieval setting, this amounts to providing ranks for all documents. The empirical copula requires the set of ranks $\{R_1, \ldots, R_d\}$ to comprise of ranks for the same set of elements, that is $\text{Dom} R_1 = \cdots = \text{Dom} R_d$. Hence a key challenge in applying Spearman’s $\rho$ to rank aggregation is to estimate the statistic on incompletely labeled lists.

Recall the definition of the empirical copula (8). We now consider the case where $\text{Dom} R \neq \text{Dom} S$, but $R$ and $S$ are generated from two top ranked lists. We define extended ranks $R', S'$ with codomain $[0, 1]$ such that $\text{Dom} R' = \text{Dom} R \cup \text{Dom} S = \text{Dom} S'$. One way to impute the missing values is to set them to a constant value for all the ranks below the top-$k$ ranks. This value is chosen to be the mid point between the start and end of the missing section. The values in the top-$k$ are retained to be the original values in the extension. The definition below formally defines this notion. Note that we have to renormalise the values.

**Definition 16 (non-informative extension)** Let $R$ be a ranking operator and $R'$ be its extension to domain $\text{Dom} R'$. Then,

$$R'(x) = \begin{cases} \frac{|\text{Dom} R| R(x)}{|\text{Dom} R'|} & x \in \text{Dom} R \\ \frac{|\text{Dom} R| + |\text{Dom} R'|}{2|\text{Dom} R'|} & \text{otherwise} \end{cases} \quad (11)$$

$\forall x \in \text{Dom} R'$.

We call this the noninformative extension since it assumes that all items that are not ranked have the same rank (the mean of the missing ranks). Note that the two experts $R_i$ and $R_j$ may have ranked different numbers of objects. An advantage of this extension is that it can easily deal with
the case of more than two experts. Consider \(d\) experts \(R_1, \ldots, R_d\), each which may have ranked a different subset of the objects. Hence the extension has to impute values on the union of items from all experts. Denote \(\text{Dom} R' := \text{Dom} R_1 \cup \ldots \cup \text{Dom} R_d\), then we can apply Definition 16 to complete each ranking operator \(R_j\). An additional advantage to the noninformative extension is that it results in a consistent ranking.

**Definition 17** An extended ranking \(R'\) of \(R\) is called consistent if the following axioms hold:

1. \(R'(x) < R'(y) \forall x, y \in \text{Dom} R\) with \(R(x) < R(y)\)
2. \(R'(x) = R'(y) \forall x, y \in \text{Dom} R\) with \(R(x) = R(y)\)
3. \(R'(y) > R'(x) \forall x \in \text{Dom} R, y \in \text{Dom} R'\)

If \(E[R] = E[R']\) also holds, then \(R'\) is called strictly consistent.

**Lemma 18** Definition 16 produces a consistent ranking. If \(E[R] = \frac{1}{2}\) then \(E[R'] = \frac{1}{2}\) produces a strictly consistent ranking.

**Proof** The notation \(|\text{Dom} R|\) can become unwieldy in following proof. We therefore adopt the shorthand notations \(r := |\text{Dom} R|\) and \(r' := |\text{Dom} R'|\) for the size of the respective sets.

Axioms 1 and 2 are satisfied by definition as the map \(x \mapsto \frac{r}{r'} x\) is monotonic. For all \(x \in \text{Dom} R' \setminus \text{Dom} R\),

\[
R'(x) = \frac{r + r'}{2r'} \leq \frac{2R(y)r + r'}{2r'} \leq \frac{2R(y)r}{2r'} = \frac{r}{r'} R(y) = R'(y)
\]

for any \(y \in \text{Dom} R\), satisfying axiom 3.

Furthermore, as

\[
E[R'] = \frac{1}{r'} \left( \sum_{x \in \text{Dom} R} R'(x) + \sum_{x \in \text{Dom} R' \setminus \text{Dom} R} R'(x) \right)
\]

\[
= \frac{1}{r'} \left( \frac{r}{r'} \sum_{x \in \text{Dom} R} R(x) + (r' - r) \frac{r + r'}{2r'} \right)
\]

\[
= \frac{1}{r'} \left( \frac{r^2}{r'} E[R] + (r' - r) \frac{r + r'}{2r'} \right)
\]

\[
= \frac{r^2 (2E[R] - 1) + r'}{2r'^2},
\]

\(R'\) is strictly consistent if \(E[R] = \frac{1}{2}\).
Note also that there is a dual imputation whereby missing values are assigned to the top of the list rather than the bottom. This is equivalent to the above imputation applied to reverse rankings. The choice of top or bottom imputation is application dependent.

Though we have only presented one method of imputing missing ranks, there are other methods worth consideration, particularly when we consider more than two lists. Choice of imputation can also give further control over the model, such as introducing prior beliefs about the ordering of items. We defer the development and discussion of other imputation methods as it is out of scope of this manuscript.

5. Supervised learning to rank

We now consider the task of learning rank aggregation from extreme ranks. Theorem 15 and Definition 16 provide the core of our algorithm. Using Theorem 15 we can find an “average” rank that aggregates a set of ranks, and by extending top-$k$ and bottom-$k$ ranks to a common domain, we can apply it to partially labelled data.

5.1 Weighted mixture of experts

As a result of Theorem 15 we have a way of finding the ranking (according to some rank generator) that is closest to a set of ranks. Consider the learning problem where we have a ranking $L$ which comprise our labels, and a set of $d$ experts $\{R_j\}$. During training, we would like to find a weighting of the input rankings such that it gives the label. Given a target ranking $L$, we would like to optimise the weights $\omega$,

$$\min_\omega \rho_n(L, R_1^{\omega_1}, R_2^{\omega_2}, \ldots, R_d^{\omega_d})$$

Here we have introduced weights $\omega$ over each rank to control the influence of each rank over the final consensus rank; the intuition here is that ranks with $\omega_i > 1$ are “replicated” with more influence, which is easy to see when $\omega_i$ are natural numbers. For example, a weight of 2 would mean the ranked list has appeared twice in the calculation of the consensus rank. While it is convenient to have integer weights for interpretability, the weights $\omega$ could be any real number in general. In the following, we consider $\omega \in \mathbb{R}^n$.

Instead of performing this high-dimensional optimisation, we decompose it into a pairwise (bivariate) comparison between the label $L$ and the weighted geometric mean, where we now explicitly show the fact that the ranks are a function of the $n$ objects $x$

$$\min_\omega \sum_x \rho_n(L(x), \sigma(R_1^{\omega_1} \otimes R_2^{\omega_2} \otimes \cdots \otimes R_d^{\omega_d})(x))$$

where the notation $\otimes$ indicates the product operator. Observe that we have used Theorem 15 to convert the $d$ dimensional problem into the product of ranks $R_j$ and the Spearman’s correlation above is only two dimensional. For bivariate Spearman’s $\rho$, this can be expressed in terms of the squared difference $1$. We further assume that $\sigma$ is the identity mapping to simplify the problem, giving us:

$$\min_\omega \sum_x (L(x) - R_1^{\omega_1}(x)R_2^{\omega_2}(x)\ldots R_d^{\omega_d}(x))^2.$$  \hspace{1cm} (12)

The objective (12) minimises the distance between the label ranks and the weighted expert ranks.
5.2 Least squares method on logarithm of ranks

Recall that we consider normalised ranks (divided by $n + 1$). By using the logarithm identity, we convert the power scaling in (12) into a multiplicative scaling. Our algorithm is:

1. Extend incomplete ranks $\{R_i\}$ to $\{R'_i\}$ by imputing the average missing value such that $\text{Dom} R'_i = \text{Dom} L$;

2. Convert to log-ranks $r'_i = \log \circ R'_i$ and $l = \log \circ L$;

3. Learn weights $\omega$ by minimising

$$
\sum_x \left( l(x) - \sum_{j=1}^d \omega_j r'_j(x) \right)^2,
$$

where the outer sum is over the $n$ examples $x$.

A log transformation of the ranks is used as it naturally encodes the weights as a power scaling in the framework of theorem 15, i.e., the weighted consensus rank is given by $\prod_i r'_i(x)^{\omega_i}$. Note that this is still solving (12) as we are optimising Spearman’s $\rho$, which is sensitive only to ordering, and therefore though the final weights are different $\rho$ is maximised via (1).

In the following experiments we also included a bias/offset term in the least squares problem which can be interpreted as adding a ranking that is constant (gives all objects the same rank). It is interesting to note that the final step in this procedure is closely related to Borda Count, except our consensus rank is the geometric mean instead of the arithmetic mean. Since this is a least squares estimation problem, we directly use the closed form solution.

5.3 Benchmarking on LETOR 4.0

We tested our method on the MQ2007-agg and MQ2008-agg list aggregation benchmarks [Qin et al. (2010b)]. The goal in these challenges is to aggregate 21 and 25 different rankers respectively over a set of query-document pairs. Each dataset has 5 pre-defined cross-validation folds with each fold providing a training, testing and validation dataset (60%/20%/20%). We trained our model on the training set and tested on the testing set, leaving the validation set unused since we have no hyperparameters.

In the following we consider two types of experts: either experts $\{R_j\}$ are top-$k$ experts, that is they only rank the best $k$ samples from $\Omega$, or experts are bottom-$k$ experts, that is they identify the worst $k$ samples from $\Omega$. We call our proposed method RAGS-$\top$ and RAGS-$\bot$ respectively. We assume that the ranked documents in the benchmark datasets are either top-$k$ or bottom-$k$ respectively, with potentially different numbers of documents $k$ labeled by each expert. Ties are given the average rank of tied documents.

To evaluate the agreement, we use the standard evaluation tool from the LETOR website[1] which implements the Normalised Discounted Cumulative Gain (NDCG). In fig. 1a we see that our approach RAGS-$\top$ performs better than all other methods at any selection size on the MQ2007-agg dataset. Indeed, we also perform better than [Qin et al. (2010a)] where the best result uses a coset-permutation distance based stagewise (CPS) model with Spearman’s $\rho$ in a probabilistic model. Recall that our approach considers the multivariate Spearman’s $\rho$ whereas [Qin et al. (2010a)] uses

[1]: http://research.microsoft.com/letor

14
bivariate Spearman’s $\rho$ in a pairwise fashion. For MQ2008-agg (fig. 1b), again our approach performs better than all other methods.

To tease apart the effect of imputing missing ranks and the effect of weighting the experts, we compared our proposed method with and without training (uniform weights). GeoMean denotes the results for the geometric mean (uniform weights on the experts) after performing imputation assuming top-$k$ ranking by the experts. First we observe that our proposed approach outperforms the geometric mean which is a good sanity check. It is surprising that the geometric mean performs quite well in MQ2007. The major difference is that we are imputing the missing ranks, and the other methods suffer from assigning them to an arbitrary value. This demonstrates the importance of imputation.

5.4 Strictly ordered labels

One issue with the benchmark aggregation dataset is that the labels are only $\{0,1,2\}$ relevance scores, and hence it is unclear exactly what the rankings are within the relevance classes. We create a new dataset which is formed by taking the intersection between the documents retrieved by a particular query between MQ2007-agg and MQ2007-list. This new dataset contains the strictly ordered labels from MQ2007-list, but uses the aggregation data from MQ2007-agg. The same procedure is used to create the corresponding dataset for MQ2008-agg and MQ2008-list. These datasets are available for download at anonymous.website. We maintain exactly the same 5-fold cross validation splits and report our results in table 3.
Table 1: Results on MQ2007-agg: NDCG. Our method is labelled RAGS-$\top$ and RAGS-$\bot$ corresponding to top and bottom imputation respectively. The results for CPS-S was the best reported in Qin et al. (2010a). The results of $\theta$-MPM was the best among the reported results in Volkovs and Zemel (2012) from BordaCount, CPS, SVP, Bradley-Terry model, and Plackett-Luce model. The results of St.Agg was the best among the reported results in Niu et al. (2013) and was the best among MCLK, SVP, Plackett-Luce model, $\theta$-MPM, BordaCount and RRF.

| Fold | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------|---|---|---|---|---|---|---|---|---|----|
|      | 0.362 | 0.362 | 0.362 | 0.362 | 0.362 | 0.362 | 0.362 | 0.362 | 0.362 | 0.362 |
| RAGS-$\top$ | 0.44684 | 0.43342 | 0.43342 | 0.43342 | 0.43342 | 0.43342 | 0.43342 | 0.43342 | 0.43342 | 0.43342 |
| RAGS-$\bot$ | 0.32804 | 0.3448 | 0.34836 | 0.35114 | 0.3552 | 0.36132 | 0.36656 | 0.37402 | 0.37952 | 0.38458 |
| CPS-S | 0.332 | 0.341 | 0.352 | 0.362 |    |    |    |    |    |    |
| $\theta$-MPM | 0.4177 | 0.4191 | 0.4192 | 0.4234 | 0.4279 |    |    |    |    |    |
| St.Agg | 0.4195 | 0.4404 |    |    |    |    |    |    |    |    |
| GeoMean | 0.42264 | 0.42528 | 0.42570 | 0.42834 | 0.42886 | 0.43342 | 0.43664 | 0.44004 | 0.44684 | 0.45216 |
| CPS | 0.4177 | 0.4192 | 0.4234 | 0.4279 |    |    |    |    |    |    |
| RAGS-$\top$ | 0.42264 | 0.42528 | 0.42570 | 0.42834 | 0.42886 | 0.43342 | 0.43664 | 0.44004 | 0.44684 | 0.45216 |
| RAGS-$\bot$ | 0.32804 | 0.3448 | 0.34836 | 0.35114 | 0.3552 | 0.36132 | 0.36656 | 0.37402 | 0.37952 | 0.38458 |
| CPS-S | 0.4177 | 0.4192 | 0.4234 | 0.4279 |    |    |    |    |    |    |
| $\theta$-MPM | 0.4177 | 0.4192 | 0.4234 | 0.4279 |    |    |    |    |    |    |
| St.Agg | 0.4177 | 0.4192 | 0.4234 | 0.4279 |    |    |    |    |    |    |
| GeoMean | 0.42264 | 0.42528 | 0.42570 | 0.42834 | 0.42886 | 0.43342 | 0.43664 | 0.44004 | 0.44684 | 0.45216 |

The results for CPS-S was the best reported in Qin et al. (2010a). The results of $\theta$-MPM was the best among the reported results in Volkovs and Zemel (2012) from BordaCount, CPS, SVP, Bradley-Terry model, and Plackett-Luce model. The results of St.Agg was the best among the reported results in Niu et al. (2013) and was the best among MCLK, SVP, Plackett-Luce model, $\theta$-MPM, BordaCount and RRF.
|       | @1   | @2   | @3   | @4   | @5   | @6   | @7   | @8   | @9   | @10  |
|-------|------|------|------|------|------|------|------|------|------|------|
| RAGS-⊤ | 0.41158 | 0.44898 | 0.47118 | 0.4922 | 0.50696 | 0.51706 | 0.52416 | 0.48732 | 0.24498 | 0.24768 |
| RAGS-⊥ | 0.35156 | 0.40338 | 0.42624 | 0.45326 | 0.4706 | 0.48352 | 0.48994 | 0.45336 | 0.22138 | 0.22514 |
| CPS-S  | 0.314 | 0.376 | 0.419 | 0.398 | 0.419 | 0.4399 | 0.4399 | 0.419 | 0.398 | 0.398 |
| t-MPM  | 0.3817 | 0.4057 | 0.4219 | 0.4307 | 0.4399 | 0.4515 | 0.4515 | 0.4157 | 0.2157 | 0.2157 |
| St.Aggregation | 0.4052 | 0.45976 | 0.47938 | 0.49042 | 0.49986 | 0.46108 | 0.22334 | 0.22812 |
Table 3: Results on MQ2007-agglist and MQ2008-agglist. The left column shows the results for multivariate Spearman’s $\rho$ and the right column shows the result for Kendall’s $\tau$.

| Method  | MQ2007-agglist | MQ2008-agglist |
|---------|----------------|----------------|
|         | $\rho$  | $\tau$  | $\rho$  | $\tau$  |
| RAGS-$\top$ | 0.4394 | 0.6201 | 0.7235 | 0.6931 |
| RAGS-$\bot$ | 0.2992 | 0.2488 | 0.6349 | 0.5560 |
| GeoMean  | 0.2457 | 0.3011 | 0.5777 | 0.6578 |
| Borda    | 0.2217 | 0.1790 | 0.5519 | 0.5869 |

Considering the results for Spearman’s $\rho$, we observe that our learning method performs well. Note that the geometric mean outperforms Borda count on both datasets, which confirms that our theoretically justified model performs better than the heuristic model. It is interesting to observe that optimising for Spearman’s $\rho$ could result in a decrease in Kendall’s $\tau$. This demonstrates the importance of choosing the appropriate objective function for learning.

6. Conclusion

We propose an approach for learning weights between experts for the task of rank aggregation. By generalising the derivation of concordance functions, we obtain an expression for multivariate Spearman’s $\rho$. Furthermore we show that the geometric mean of the expert ranks is the optimal aggregator under Spearman’s correlation. Motivated by this, our method solves a least squares estimation problem for logarithmic normalised ranks to find optimal weights. We propose an imputation method for completing top-$k$ ranked lists that allows us to apply Spearman’s $\rho$ to aggregate ranks from partial lists. Surprisingly, our weighted geometric mean shows state of the art results on benchmark datasets, without the need for tuning hyperparameters or expensive computation. The simplicity of our model makes it easier to interpret, and the weights give a direct estimate of the influence of each expert. This problem has wide applications to ensemble learning, voting, text mining, recommender systems and bioinformatics.

7. Acknowledgements

NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

Appendix A. Bivariate Spearman’s $\rho$ and Squared Distance

This well known result shows that Spearman’s $\rho$ can be expressed in terms of the squared distance between ranks.

2. http://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient accessed on 20 May 2014
In the following derivation, we use the expressions for the sum of integers and the sum of squares of integers:

\[ \sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}. \]

Recall that Spearman is defined as:

\[ \rho_n = \frac{\sum_x (R(x) - \bar{R})(S(x) - \bar{S})}{\sqrt{\sum_x (R(x) - \bar{R})^2 \sum_x (S(x) - \bar{S})^2}}. \]

Since there are no ties, both \( R(x) \) and \( S(x) \) consist of integers from 1 to \( n \) inclusive, and the two squared sums in the denominator are the same. Recall that the mean rank is:

\[ \bar{R} = \bar{S} = \frac{n+1}{2}, \quad \text{and} \quad \sum_x R(x) = \frac{n(n+1)}{2} = n\bar{R}. \]

Therefore the denominator can be expressed as a function of \( n \).

\[ \sqrt{\sum_x (R(x) - \bar{R})^2 \sum_x (S(x) - \bar{S})^2} = \sum_x (R(x) - \bar{R})^2 \]

\[ = \sum_x (R(x)^2 - 2R(x)\bar{R} + \bar{R}^2) \]

\[ = \sum_x R(x)^2 - 2\bar{R} \sum_x R(x) + n\bar{R}^2 \]

\[ = \sum_x R(x)^2 - n\bar{R}^2 \]

\[ = \frac{n(n+1)(2n+1)}{6} - n \left( \frac{n+1}{2} \right)^2 \]

\[ = n(n+1) \left( \frac{2n+1}{6} - \frac{n+1}{4} \right) \]

\[ = n(n+1) \left( \frac{n-1}{12} \right) \]

\[ = \frac{n(n^2-1)}{12}. \]

Since both \( R(x) \) and \( S(x) \) consists of the same integers, we can express the squared difference in terms of the product.

\[ \sum_x \frac{1}{2} (R(x) - S(x))^2 = \sum_x \frac{1}{2} (R(x)^2 - 2R(x)S(x) + S(x)^2) \]

\[ = \sum_x \frac{1}{2} (R(x)^2 + S(x)^2) - \sum_x R(x)S(x) \]

\[ = \sum_x R(x)^2 - \sum_x R(x)S(x) \]

where the first term is a function of \( n \).
We express the product of the means, which appears in the numerator later, to match the denominator:

\[
n \left( \frac{n + 1}{2} \right)^2 = \frac{n(n + 1)}{12} 3(n + 1) = -\frac{n(n + 1)}{12} [(n - 1) - (4n + 2)] = \frac{n(n + 1)(n - 1)}{12} + \frac{n(n + 1)(2n + 1)}{6} = -\frac{n(n^2 - 1)}{12} + \sum_x R(x)^2
\]

where the last term in the sum is the expression for the sum of squares. We can now derive the expression for the numerator:

\[
\sum_x (R(x) - \bar{R})(S(x) - \bar{S}) = \sum_x R(x)S(x) - \bar{R} \sum_x S(x) - \bar{S} \sum_x R(x) + n\bar{R}\bar{S} = \sum_x R(x)S(x) - n\bar{R}\bar{S} = \sum_x R(x)S(x) - n \left( \frac{n + 1}{2} \right)^2 = \sum_x R(x)S(x) + \frac{n(n^2 - 1)}{12} - \sum_x R(x)^2 = \frac{n(n^2 - 1)}{12} - \sum_x \frac{1}{2}(R(x) - S(x))^2
\]

where the last line uses the expression of the sum of squared differences above. Putting together the expressions for the numerator and denominator together gives the desired result:

\[
\rho_n = 1 - \frac{6 \sum_x (R(x) - S(x))^2}{n(n^2 - 1)}
\]
References

Douglas E. Critchlow. *Metric Methods for Analyzing Partially Ranked Data*. Springer-Verlag, 1985.

Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Lecture Notes – Monograph Series*. Institute of Mathematical Statistics, 1988.

Gal Elidan. Copulas in machine learning. In *Copulae in Mathematical and Quantitative Finance*, 2013.

M.A. Fligner and J.S. Verducci. Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 359–369, 1986.

Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.

Rishabh Iyer and Jeff Bilmes. The submodular Bregman and Lovász-Bregman divergences with applications. In *NIPS*, 2012.

Rishabh Iyer and Jeff Bilmes. The Lovász-Bregman divergence and connections to rank aggregation, clustering and web ranking. In *UAI*, 2013.

Harry Joe. *Dependence Modeling with Copulas*. CRC Press, 2014.

Alexandre Klementiev, Dan Roth, and Kevin Small. Unsupervised rank aggregation with distance-based models. In *International Conference on Machine Learning*, 2008.

Guy Lebanon and Yi Mao. Non-parametric modeling of partially ranked data. *Journal of Machine Learning Research*, 9:2401–2429, 2008.

Tie-Yan Liu. *Learning to Rank for Information Retrieval*. Springer-Verlag, 2011.

C. L. Mallows. Non-null ranking models. *Biometrika*, 44(1):114–130, 1957.

Roger B. Nelsen. *An Introduction to Copulas*. Springer, second edition, 2006.

Shuzi Niu, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. Top-k learning to rank: Labeling, ranking and evaluation. In *SIGIR*, pages 751–760, 2012. ISBN 9781450314725.

Shuzi Niu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. Stochastic rank aggregation. In *UAI*, 2013.

Tao Qin, Xiubo Geng, and Tie-Yan Liu. A new probabilistic model for rank aggregation. In *NIPS*, 2010a.

Tao Qin, Tie-Yan Liu, Jun Xu, and Hang Li. LETOR: A benchmark collection for research on learning to rank for information retrieval. *Information Retrieval Journal*, 2010b.

Friedrich Schmid and Rafael Schmidt. Multivariate extensions of spearman’s rho and related statistics. *Statistics & Probability Letters*, 77:407–416, 2007.

D. Sculley. Combined regression and ranking. In *KDD*, 2010.
C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904.

Michael. D. Taylor. Multivariate measures of concordance. *Annals of the Institute of Statistical Mathematics*, 59(4):789–806, 2007.

Pravin K. Trivedi and David M. Zimmer. Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, 1(1):1–111, 2005.

Maksims N. Volkovs and Richard S. Zemel. A flexible generative model for preference aggregation. In *WWW*, 2012.