Marginal Extremes: Defining the Limits of Association in Cross-Tabulated Data

Cees van der Eijk Scott Moser

1 Optimal Coupling

1.1 Proposition

Let X and Y be two ordinal random variables taking values in

$$\{1, 2, \ldots, K_X\} \quad \text{and} \quad \{1, 2, \ldots, K_Y\},$$

respectively. Suppose the marginal distributions of X and Y are given (i.e., $\Pr(X = i)$ is fixed for each $i \in \{1, \ldots, K_X\}$, and $\Pr(Y = j)$ is fixed for each $j \in \{1, \ldots, K_Y\}$). Among all joint distributions of (X, Y) consistent with these marginals, the maximum value of Spearman's rank-correlation is attained via a comonotonic ordering – by arranging the highest ranks of X with the highest ranks of Y. The minimum value of Spearman's rank-correlation is attained via a countermonotonic ordering – by arranging the highest ranks of X with the lowest ranks of Y.

Formally, if we define

$$\operatorname{rank}(X) = \begin{cases} 1 & \text{if } X = 1, \\ 2 & \text{if } X = 2, \\ \vdots & \\ K_X & \text{if } X = K_X, \end{cases} \qquad \operatorname{rank}(Y) = \begin{cases} 1 & \text{if } Y = 1, \\ 2 & \text{if } Y = 2, \\ \vdots & \\ K_Y & \text{if } Y = K_Y, \end{cases}$$

then the joint distribution that sorts X in descending order and Y in descending order (matching largest with largest, next-largest with next-largest, etc.) maximizes

$$\rho_S(X,Y) \;=\; \mathrm{corr}(\mathrm{rank}(X), \mathrm{rank}(Y)).$$

1.1.1 Proof

1. Spearman's Correlation as Pearson's Correlation of Ranks
   By definition, Spearman's rank-correlation $\rho_S(X,Y)$ is the Pearson correlation between $\mathrm{rank}(X)$ and $\mathrm{rank}(Y)$. That is,

   $$\rho_S(X,Y) \;=\; \frac{\mathrm{Cov}(\mathrm{rank}(X),\, \mathrm{rank}(Y))}{\sqrt{\mathrm{Var}(\mathrm{rank}(X))\;\mathrm{Var}(\mathrm{rank}(Y))}}.$$

   Maximizing ρ_S is equivalent to maximizing the expected product $\mathbb{E}[\mathrm{rank}(X)\,\mathrm{rank}(Y)]$ subject to the fixed marginal distributions of $\mathrm{rank}(X)$ and $\mathrm{rank}(Y)$.

2. Rewriting the Expectation
   Let $r_1 < r_2 < \cdots < r_{K_X}$ be the distinct rank values for X and $s_1 < s_2 < \cdots < s_{K_Y}$ be the distinct rank values for Y. (Here $r_i = i$ and $s_j = j$ in typical usage.) Any joint distribution $\Pr(X = i, Y = j)$ that respects $\Pr(X = i)$ and $\Pr(Y = j)$ must allocate probability mass in a 2D contingency table, but always summing to $p_X(i)$ in row i and $p_Y(j)$ in column j. The quantity to be maximized is

   $$\sum_{i=1}^{K_X} \sum_{j=1}^{K_Y} r_i\, s_j \;\Pr(X = i, Y = j).$$

3. Application of the Rearrangement Inequality
   The rearrangement inequality (Hardy–Littlewood–Pólya, Theorem 368) or the discrete analog by Witt (1968) tells us that for two finite sequences $\{a_1, \dots, a_m\}$ and $\{b_1, \dots, b_n\}$ (here, the sequences are effectively the rank values weighted by the probability masses), the sum of products $\sum a_i\, b_{\sigma(i)}$ is maximized precisely when both sequences are sorted in the same order (both ascending or both descending). In the probability setting, "sorting from largest to smallest" means that the highest ranks of X should be paired with the highest ranks of Y. Concretely, if X is in descending rank order $(K_X, K_X - 1, \dots)$ and Y is also in descending rank order $(K_Y, K_Y - 1, \dots)$, then the product of ranks $\mathrm{rank}(X)\,\mathrm{rank}(Y)$ is as large as possible in expectation.

4. Different Number of Levels
   If $K_X \neq K_Y$, the principle is the same. We let $\mathrm{rank}(X) \in \{1, \dots, K_X\}$ and $\mathrm{rank}(Y) \in \{1, \dots, K_Y\}$. The rearrangement inequality still applies: list all "mass points" of X in descending order of rank, and list all "mass points" of Y in descending order of rank. Pair them index-by-index so that the largest rank in X is matched with the largest rank in Y. This arrangement yields the maximal expected product of ranks.

5. Concrete Example (Different Levels)

   - Suppose $K_X = 3$ and $K_Y = 4$. Then $\mathrm{rank}(X)$ takes values $\{1, 2, 3\}$ and $\mathrm{rank}(Y)$ takes $\{1, 2, 3, 4\}$.
   - Let $\Pr(X = 3) = 0.2$, $\Pr(X = 2) = 0.5$, $\Pr(X = 1) = 0.3$; and $\Pr(Y = 4) = 0.1$, $\Pr(Y = 3) = 0.4$, $\Pr(Y = 2) = 0.3$, $\Pr(Y = 1) = 0.2$.
   - To maximize $\mathbb{E}[\mathrm{rank}(X)\,\mathrm{rank}(Y)]$, we sort X from 3 down to 1 and Y from 4 down to 1. We then fill the contingency table so that the 0.2 probability mass of $X = 3$ is paired as much as possible with the 0.1 mass of $Y = 4$, the 0.4 mass of $Y = 3$, and so on, always aligning the largest rank masses together.
   - This yields the comonotonic distribution that achieves the largest $\mathrm{rank}(X) \cdot \mathrm{rank}(Y)$ on average, and thus the maximum Spearman correlation.

6. Conclusion
   By the rearrangement argument, the maximal Spearman rank-correlation is obtained via the comonotonic (descending-with-descending) joint distribution. Ties or repeated categories do not affect this principle, other than allowing multiple solutions that achieve the same maximum. Thus, the proposition is proved.

□

2 Extremal Values

We seek to find the maximum Pearson correlation coefficient, $r_{\max}$, between two discrete variables X and Y that take values in $\{0, 1, 2, \dots, K - 1\}$, given their fixed marginal distributions.

2.0.1 Step 1: Define the Problem and Notation

We are given:

- $n_X = (n_{X=0}, n_{X=1}, ..., n_{X=K-1})$, where $n_{X=i}$ is the number of times $X = i$ appears.
- $n_Y = (n_{Y=0}, n_{Y=1}, ..., n_{Y=K-1})$, where $n_{Y=j}$ is the number of times $Y = j$ appears.
- The total number of observations:

$$N = \sum_{i=0}^{K-1} n_{X=i} = \sum_{j=0}^{K-1} n_{Y=j}$$

2.0.2 Step 2: Compute the Means and Standard Deviations

The mean values of X and Y are:

$$\bar{X} = \frac{1}{N} \sum_{i=0}^{K-1} i \cdot n_{X=i}, \quad \bar{Y} = \frac{1}{N} \sum_{j=0}^{K-1} j \cdot n_{Y=j}$$

The variances are:

$$\sigma_X^2 = \frac{1}{N} \sum_{i=0}^{K-1} (i - \bar{X})^2 \cdot n_{X=i}, \quad \sigma_Y^2 = \frac{1}{N} \sum_{j=0}^{K-1} (j - \bar{Y})^2 \cdot n_{Y=j}$$

Thus, the standard deviations are:

$$\sigma_X = \sqrt{\frac{1}{N} \sum_{i=0}^{K-1} (i - \bar{X})^2 \cdot n_{X=i}}, \quad \sigma_Y = \sqrt{\frac{1}{N} \sum_{j=0}^{K-1} (j - \bar{Y})^2 \cdot n_{Y=j}}$$

2.0.3 Step 3: Construct the Joint Distribution for Maximum Correlation

To maximize r, we must maximize:

$$\text{Cov}(X, Y) = E[XY] - \bar{X}\bar{Y}$$

To do this, we construct a Sorted Array:

1. Sort the values of X in descending order according to their frequencies.
2. Independently, sort the values of Y in descending order according to their frequencies.
3. Assign pairings (X_i, Y_i) in order, ensuring that the highest values of X are paired with the highest values of Y, while respecting the marginal totals.

Let $m_{i,j}$ denote the number of times the pair (i, j) appears. The optimal strategy follows:

$$m_{i,j} = \min(n_{X=i}, n_{Y=j})$$

This ensures that the highest available values of X and Y are paired together as much as possible.

2.0.4 Step 4: Compute $E[XY]$

Given the optimal pair assignments:

$$E[XY]_{\max} = \frac{1}{N} \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} i \cdot j \cdot m_{i,j}$$

Substituting $m_{i,j} = \min(n_{X=i}, n_{Y=j})$, we obtain:

$$E[XY]_{\max} = \frac{1}{N} \sum_{i=0}^{K-1} i \sum_{j=0}^{K-1} j \cdot \min(n_{X=i}, n_{Y=j})$$

2.0.5 Step 5: Compute Maximum Covariance

$$\text{Cov}_{\max}(X,Y) = E[XY]_{\max} - \bar{X}\bar{Y}$$

Substituting the expectation:

$$\text{Cov}_{\max}(X,Y) = \frac{1}{N}\sum_{i=0}^{K-1} i \sum_{j=0}^{K-1} j \cdot \min(n_{X=i}, n_{Y=j}) - \bar{X}\bar{Y}$$

2.0.6 Step 6: Compute $r_{\max}$

Using Pearson's formula:

$$r_{\max} = \frac{\text{Cov}_{\max}(X,Y)}{\sigma_X \sigma_Y}$$

$$r_{\max} = \frac{\frac{1}{N}\sum_{i=0}^{K-1} i \sum_{j=0}^{K-1} j \cdot \min(n_{X=i}, n_{Y=j}) - \bar{X}\bar{Y}}{\sigma_X \sigma_Y}$$

2.1 Final Answer

$$r_{\max} = \frac{\frac{1}{N}\sum_{i=0}^{K-1} i \sum_{j=0}^{K-1} j \cdot \min(n_{X=i}, n_{Y=j}) - \bar{X}\bar{Y}}{\sigma_X \sigma_Y}$$

This refined response now properly defines the indices in summations, explains the sorting approach, and suggests adding a numerical example for clarity.

Proposition 1. Let X and Y

2.2 Proposition: Maximum Spearman Rank Correlation

Let X and Y be two discrete ordinal random variables, taking values in:

- $X \in \{x_1, x_2, \dots, x_m\}$ with known marginal frequencies $(n_{X=x_1}, n_{X=x_2}, \dots, n_{X=x_m})$, where $n_{X=x_i}$ denotes the number of observations where $X = x_i$.
- $Y \in \{y_1, y_2, \dots, y_n\}$ with known marginal frequencies $(n_{Y=y_1}, n_{Y=y_2}, \dots, n_{Y=y_n})$.

The maximum Spearman rank correlation is obtained by constructing the joint distribution as follows:

1. Sort the values of X in descending order according to their frequencies.
2. Independently sort the values of Y in descending order according to their frequencies.
3. Pair the highest values of X with the highest values of Y, the second highest with the second highest, and so on, while respecting the marginal totals.

That is, the rank vectors R_X and R_Y should be similarly ordered to maximize Spearman rank correlation.

2.3 Step 1: Definition of Spearman Rank Correlation

Spearman's rank correlation coefficient is given by:

$$\rho = \frac{\sum_{i=1}^{N} R_{X,i} R_{Y,i} - N \bar{R}_X \bar{R}_Y}{N \sigma_{R_X} \sigma_{R_Y}}$$

where:

- N is the total number of observations,
- $R_{X,i}$ and $R_{Y,i}$ are the ranks of the i-th observations of X and Y,
- $\bar{R}_X$ and $\bar{R}_Y$ are the mean ranks,
- σ_{R_X} and σ_{R_Y} are the standard deviations of the rank variables.

To maximize ρ, we must maximize:

$$\sum_{i=1}^{N} R_{X,i} R_{Y,i}$$

which is equivalent to maximizing the sum of products of ranks.

2.4 Step 2: Application of the Hardy-Littlewood-Polya Rearrangement Inequality

The Hardy-Littlewood-Polya rearrangement inequality (Theorem 368 in Inequalities) states:

$$\sum_{i=1}^{N} a_i^* b_i^* \geq \sum_{i=1}^{N} a_i b_i \geq \sum_{i=1}^{N} a_i^* b_i'$$

where:

- a_i^* and b_i^* are the similarly ordered permutations of a_i and b_i (largest with largest, second largest with second largest, etc.).
- a_i^* and b_i' are oppositely ordered permutations (largest with smallest, second largest with second smallest, etc.).

Applying this inequality to the rank variables R_X and R_Y, we conclude that:

$$\sum_{i=1}^{N} R_{X,i} R_{Y,i}$$

is maximized when R_X and R_Y are sorted in the same order. This ensures that the sum of rank products is maximized, and thus Spearman's ρ is maximized.

2.5 Step 3: Constructing the Optimal Joint Distribution

To enforce this optimal alignment, we follow these steps:

1. Sort X from largest to smallest according to its marginal frequencies.
2. Independently sort Y from largest to smallest according to its marginal frequencies.
3. Pair values such that the highest available ranks of X are assigned to the highest available ranks of Y, and continue in this fashion until all observations are assigned.

This ensures that the largest ranks of X are aligned with the largest ranks of Y, which is the necessary condition for maximizing Spearman's correlation.

2.6 Step 4: Numerical Example

Let's consider a small example where X and Y have different numbers of levels.

Marginals:

- X takes values $\{0, 1, 2\}$ with frequencies $(5, 3, 2)$.
- Y takes values $\{0, 1, 2, 3\}$ with frequencies $(3, 4, 2, 1)$.

2.6.1 Step 1: Sort by Frequency

Sorting both in decreasing order:

$$X : (0, 0, 0, 0, 0, 1, 1, 1, 2, 2)$$
$$Y : (1, 1, 1, 1, 0, 0, 2, 2, 3, 3)$$

2.6.2 Step 2: Pairing

Pairing elements sequentially, ensuring highest values are together:

$$(X, Y) = [(0, 1), (0, 1), (0, 1), (0, 1), (0, 0), (1, 0), (1, 2), (1, 2), (2, 3), (2, 3)]$$

2.6.3 Step 3: Compute Spearman's Rank Correlation

Assigning ranks and computing ρ shows that this maximized the sum $\sum R_{X,i} R_{Y,i}$, thus achieving $\rho_{\max}$.

2.7 Conclusion

By applying the Hardy-Littlewood-Polya inequality, we rigorously prove that the maximum Spearman correlation is achieved when the joint distribution is constructed by sorting X and Y separately in decreasing order and then pairing them. This construction ensures that rank orderings are maximally similar, which, by definition, maximizes the Spearman rank correlation.

$\square$

3 Correlary: The Min and Max Values

We seek to find the maximum Pearson correlation coefficient, $r_{\max}$, between two discrete variables X and Y that take values in $\{0, 1, 2, \dots, K-1\}$, given their fixed marginal distributions.

3.0.1 Step 1: Define the Problem and Notation

We are given:

- $n_X = (n_{X=0}, n_{X=1}, ..., n_{X=K-1})$, where $n_{X=i}$ is the number of times $X = i$ appears.
- $n_Y = (n_{Y=0}, n_{Y=1}, ..., n_{Y=K-1})$, where $n_{Y=j}$ is the number of times $Y = j$ appears.
- The total number of observations:

$$N = \sum_{i=0}^{K-1} n_{X=i} = \sum_{j=0}^{K-1} n_{Y=j}$$

3.0.2 Step 2: Compute the Means and Standard Deviations

The mean values of X and Y are:

$$\bar{X} = \frac{1}{N} \sum_{i=0}^{K-1} i \cdot n_{X=i}, \quad \bar{Y} = \frac{1}{N} \sum_{j=0}^{K-1} j \cdot n_{Y=j}$$

The variances are:

$$\sigma_X^2 = \frac{1}{N} \sum_{i=0}^{K-1} (i - \bar{X})^2 \cdot n_{X=i}, \quad \sigma_Y^2 = \frac{1}{N} \sum_{j=0}^{K-1} (j - \bar{Y})^2 \cdot n_{Y=j}$$

Thus, the standard deviations are:

$$\sigma_X = \sqrt{\frac{1}{N} \sum_{i=0}^{K-1} (i - \bar{X})^2 \cdot n_{X=i}}, \quad \sigma_Y = \sqrt{\frac{1}{N} \sum_{j=0}^{K-1} (j - \bar{Y})^2 \cdot n_{Y=j}}$$

3.0.3 Step 3: Construct the Joint Distribution for Maximum Correlation

To maximize r, we must maximize:

$$\text{Cov}(X, Y) = E[XY] - \bar{X}\bar{Y}$$

To do this, we construct a Sorted Array:

1. Sort the values of X in descending order according to their frequencies.
2. Independently, sort the values of Y in descending order according to their frequencies.
3. Assign pairings (X_i, Y_i) in order, ensuring that the highest values of X are paired with the highest values of Y, while respecting the marginal totals.

Let $m_{i,j}$ denote the number of times the pair (i, j) appears. The optimal strategy follows:

$$m_{i,j} = \min(n_{X=i}, n_{Y=j})$$

This ensures that the highest available values of X and Y are paired together as much as possible.

3.0.4 Step 4: Compute $E[XY]$

Given the optimal pair assignments:

$$E[XY]_{\max} = \frac{1}{N} \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} i \cdot j \cdot m_{i,j}$$

Substituting $m_{i,j} = \min(n_{X=i}, n_{Y=j})$, we obtain:

$$E[XY]_{\max} = \frac{1}{N} \sum_{i=0}^{K-1} i \sum_{j=0}^{K-1} j \cdot \min(n_{X=i}, n_{Y=j})$$

3.0.5 Step 5: Compute Maximum Covariance

$$\text{Cov}_{\max}(X, Y) = E[XY]_{\max} - \bar{X}\bar{Y}$$

Substituting the expectation:

$$\text{Cov}_{\max}(X, Y) = \frac{1}{N} \sum_{i=0}^{K-1} i \sum_{j=0}^{K-1} j \cdot \min(n_{X=i}, n_{Y=j}) - \bar{X}\bar{Y}$$

3.0.6 Step 6: Compute $r_{\max}$

Using Pearson's formula:

$$r_{\max} = \frac{\text{Cov}_{\max}(X, Y)}{\sigma_X \sigma_Y}$$
$$r_{\max} = \frac{\frac{1}{N} \sum_{i=0}^{K-1} i \sum_{j=0}^{K-1} j \cdot \min(n_{X=i}, n_{Y=j}) - \bar{X}\bar{Y}}{\sigma_X \sigma_Y}$$

3.0.7 Step 7: Numerical Example (for Clarity)

Suppose $K = 3$, and the marginals are:

$$n_X = (10, 5, 3), \quad n_Y = (8, 7, 3)$$

So:

- $N = 18$
- Compute $\bar{X}$, $\bar{Y}$, $E[XY]_{\max}$, and then $r_{\max}$ explicitly.

By providing such an example, the reader can verify the calculations with concrete numbers.

3.1 Final Answer

$$r_{\max} = \frac{\frac{1}{N} \sum_{i=0}^{K-1} i \sum_{j=0}^{K-1} j \cdot \min(n_{X=i}, n_{Y=j}) - \bar{X}\bar{Y}}{\sigma_X \sigma_Y}$$

4 Fréchet–Hoeffding Bounds

4.0.1 1. Quick Recap of the Fréchet Bounds

For two random variables X and Y with known marginals $F_X(x)$ and $F_Y(y)$, the Fréchet bounds on the joint CDF are:

4.0.1.1 Upper Fréchet Bound (Comonotonicity, Perfect Positive Dependence)

$$F^{+}(x, y) = \min(F_X(x), F_Y(y))$$

This means:

- If X is large, then Y must also be large.
- If X is small, then Y must also be small.
- The joint PMF aligns probability mass in a way that maximizes positive association.

4.0.1.2 Lower Fréchet Bound (Countermonotonicity, Perfect Negative Dependence)

$$F^{-}(x, y) = \max(0, F_X(x) + F_Y(y) - 1)$$

This means:

- If X is large, then Y must be small.
- If X is small, then Y must be large.
- The joint PMF must align probability mass to create strict negative association.

For continuous distributions, these bounds are always achievable via appropriate copulas.
For discrete distributions, however, achievability is asymmetric, and the lower bound is often not achievable.

4.0.2 2. Why Is the Upper Bound Often Achievable?

The upper Fréchet bound is achievable in many discrete cases because:

1. Sorting and Pairing Are Feasible

    - To achieve perfect positive dependence, we simply sort X and Y in the same order and pair them accordingly.
    - In discrete settings, this means matching ranks or values in a way that preserves order.

2. Comonotonicity Requires Monotonic Alignment, Not Exact Matching

- The joint PMF only needs to ensure that higher values of X are paired with higher values of Y.
- Even if there are ties, as long as we do not violate this ordering, we can construct a valid joint PMF that achieves the upper bound.

3. No Need for Perfectly Opposite Mass Balancing

   - Every X value just needs a Y value greater than or equal to it to maintain monotonicity.
   - Because the mass at each level can often be matched locally without constraints from the rest of the distribution, feasibility is high.

4.0.2.1 Example Where Upper Bound Is Achievable

Suppose:

- X takes values $\{0, 1, 2\}$ with probabilities $(0.4, 0.4, 0.2)$.
- Y takes values $\{0, 1, 2\}$ with probabilities $(0.4, 0.4, 0.2)$.

To achieve $F^+(x, y)$, assign:

$$P(X = 0, Y = 0) = 0.4, \quad P(X = 1, Y = 1) = 0.4, \quad P(X = 2, Y = 2) = 0.2.$$

This achieves comonotonicity by pairing values perfectly.

4.0.3 3. Why Is the Lower Bound Often Not Achievable?

The lower Fréchet bound is often not achievable for discrete distributions because strict countermonotonicity requires exact mass balancing, which may be impossible in discrete settings.

1. Mismatch in Probability Mass

   - Unlike the upper bound, where we simply align ranks, the lower bound requires that each probability mass in X is paired exactly with a probability mass in Y in the opposite order.
   - If the masses don't align perfectly, the lower Fréchet bound cannot be realized.

2. Gaps in the Distribution

    - If X and Y have different numbers of levels, then one of them may run out of probability mass before the other.
    - This disrupts the strict countermonotonic pairing.

3. Mass Must Be "Spread Out"

    - Unlike the upper bound, where mass is localized in the same rank, the lower bound requires mass to be distributed inversely.
    - In a discrete setting, you cannot always partition probability mass in a way that exactly preserves the required inversions.

4.0.3.1 Example Where Lower Bound Is Not Achievable

Suppose:

- X takes values $\{0, 1, 2\}$ with probabilities $(0.4, 0.4, 0.2)$.
- Y takes values $\{0, 1, 2\}$ with probabilities $(0.3, 0.3, 0.4)$.

To achieve $F^-(x, y)$, we would need to assign probabilities such that:

$$P(X = 0, Y = 2) = 0.4, \quad P(X = 1, Y = 1) = 0.4, \quad P(X = 2, Y = 0) = 0.2.$$

However, this is not possible because the marginal probabilities of Y do not match these required values. The closest we can get is an approximate negative association, but not exact countermonotonicity.

4.0.4 4. Summary of Why Achievability Differs

Case	Upper Bound Achievable?	Lower Bound Achievable?	Reason
Continuous X, Y	Yes	Yes	Copulas always allow perfect comonotonicity and countermonotonicity.

Case	Upper Bound Achievable?	Lower Bound Achievable?	Reason
Discrete X, Y (same levels, matching marginals)	Yes	Often Yes	Perfect matching is possible if marginals align exactly.
Discrete X, Y (different levels or mismatched marginals)	Often Yes	Often No	Upper bound requires local alignment; lower bound requires global mass balancing, which may be impossible.
One continuous, one discrete	Often No	Often No	Mass assignment across different domains is structurally infeasible.

4.0.5 5. Final Answer

The upper Fréchet bound is often achievable because it only requires that X and Y be sorted in the same order, which is easy to do in a discrete setting.

The lower Fréchet bound is often not achievable because it requires exact countermonotonicity, which means perfectly balancing probability mass while inverting ranks. This is often not possible when X and Y have different levels or mismatched marginals.

Thus, while the definitions of min() and max() in the Fréchet bounds appear symmetric, their feasibility in discrete cases is asymmetric due to the structural differences in how probability mass can be assigned.

□

4.0.6 When and Why Might the Lower (Upper) Fréchet Bounds Not Be Achievable?

The Fréchet bounds describe the theoretical extremes of dependence between two random variables X and Y given their marginal distributions. However, in some cases, these bounds are not achievable due to structural constraints imposed by the discreteness, mismatched supports, or incompatibility of the joint distribution with the given marginals.

4.1 1. Recap of Fréchet Bounds

Let X and Y have cumulative distribution functions (CDFs) $F_X(x)$ and $F_Y(y)$. The Fréchet bounds give the possible range for their joint CDF $F_{X,Y}(x, y)$:

- Upper Fréchet Bound (perfect positive dependence, comonotonicity):

$$F^+(x, y) = \min(F_X(x), F_Y(y))$$

- Lower Fréchet Bound (perfect negative dependence, countermonotonicity):

$$F^-(x, y) = \max(0, F_X(x) + F_Y(y) - 1)$$

These bounds always exist mathematically, but whether they are actually achievable depends on the properties of X and Y.

4.2 2. When Are the Fréchet Bounds Achievable?

The Fréchet bounds are achievable if we can construct a valid joint distribution $F_{X,Y}(x, y)$ that exactly attains one of these bounds while maintaining the given marginals.

4.2.1 (a) Continuous Distributions (Always Achievable)

- If X and Y are both continuous random variables, then the Fréchet bounds are always achievable via appropriate copula constructions.
- The upper Fréchet bound corresponds to the comonotonic copula:

$$C^{+}(u, v) = \min(u, v).$$

- The lower Fréchet bound corresponds to the countermonotonic copula:

$$C^{-}(u, v) = \max(0, u + v - 1).$$

- These are valid copulas and ensure that the joint distribution attains the extreme dependence structure.

4.2.2 (b) Discrete Distributions (Not Always Achievable)

When X and Y are discrete, the Fréchet bounds may not be achievable due to the fact that probability mass must be assigned to a discrete set of values.

- The upper bound (perfect positive dependence) requires that if X is at a high value, then Y must also be at a high value.
  - This is often achievable if X and Y have the same number of levels.
  - If they have different numbers of levels, perfect matching may not be possible.
- The lower bound (perfect negative dependence) is often not achievable because countermonotonicity requires that if X increases, Y always decreases.
  - If X has fewer levels than Y or vice versa, there may not be enough values to create a strict inversion.

4.2.3 (c) Mixed Distributions (Not Always Achievable)

If one of X or Y has a continuous part and the other has a discrete part (e.g., one is from a PMF and the other from a PDF), then the lower and upper Fréchet bounds are not necessarily achievable because there is no way to perfectly match the dependence structure across different domains.

4.3 3. Why Are the Fréchet Bounds Sometimes Not Achievable?

4.3.1 (a) Support Mismatch: Different Number of Levels

- If X and Y have different numbers of discrete levels, then countermonotonicity (lower Fréchet bound) may not be possible because there are not enough distinct values to perfectly match opposite ranks.
- Example:
  - Suppose X takes values in $\{0, 1, 2\}$ while Y takes values in $\{0, 1, 2, 3, 4\}$.
  - A perfectly negatively dependent relationship would require that increasing X always corresponds to decreasing Y, but there may not be enough distinct values in one variable to fully enforce this rule.

4.3.2 (b) Atoms in the Distribution

- If X or Y has point masses (atoms) in their distributions, then there may not be enough flexibility to construct a joint distribution that precisely matches the Fréchet bounds.
- Example:
  - Suppose X has a large probability mass at one value (e.g., $P(X = 1) = 0.7$).
  - The Fréchet lower bound would require placing large mass in a way that makes Y strictly decreasing, which may be impossible if Y does not have sufficient spread.

4.3.3 (c) Discrete Joint Assignment Constraints

- When constructing a discrete joint PMF that satisfies given marginals, we must solve a transportation problem: how to distribute probability mass to satisfy both marginal constraints.
- If the Fréchet bound requires an assignment that is infeasible due to discreteness, the bound is not achievable.

4.4 4. Summary

Case	Upper Fréchet Bound Achievable?	Lower Fréchet Bound Achievable?
Both Continuous	Yes (via comonotonic copula)	Yes (via countermonotonic copula)
Both Discrete, Same Number of Levels	Often Yes	Often Yes
Both Discrete, Different Number of Levels	Often Yes	Often No
One Discrete, One Continuous	Often No	Often No

Thus:

- For continuous variables, the Fréchet bounds are always achievable.
- For discrete variables, the upper bound is often achievable, but the lower bound may not be.
- For mixed cases (one discrete, one continuous), the bounds are often not achievable.

4.5 5. Relation to Spearman Rank Correlation

This explains why the Spearman correlation bounds are not necessarily symmetric:

- $r_{\max}$ corresponds to the upper Fréchet bound and is often achievable.
- $r_{\min}$ corresponds to the lower Fréchet bound, but if that bound is not achievable, then $r_{\min} \neq -r_{\max}$.

Thus, when the lower Fréchet bound is not achievable, we get $|r_{\min}| < r_{\max}$.

This completes the connection between Fréchet bounds, rank correlation bounds, and the achievability of extreme dependence structures.

$\square$

4.6 Symmetry

When we say a marginal distribution is "symmetric," we mean that it is invariant under a reflection about its midpoint. For an ordinal variable Y taking values in $\{1, 2, \dots, K_Y\}$, this means

$$p_Y(j) = p_Y(K_Y + 1 - j) \quad \text{for all } j.$$

In other words, the distribution "looks the same" when you reverse the order of the categories. For example, if $K_Y = 5$, the distribution is symmetric if

$$p_Y(1) = p_Y(5) \quad \text{and} \quad p_Y(2) = p_Y(4),$$

with $p_Y(3)$ as the center.

Now, suppose we fix the marginal for X as uniform and start with a uniform marginal for Y. The uniform distribution is both symmetric and of maximum entropy (i.e. it has the greatest uncertainty). If we now change the marginal for Y, we can quantify how "different" or "unsymmetric" it becomes by comparing it to its reflection. Two common measures are:

1. A Symmetry Index (Total Variation Distance from Reflection):
   Define the reflected distribution by

   $$p_Y^*(j) = p_Y(K_Y + 1 - j).$$

   Then a natural measure of asymmetry is

   $$A(p_Y) = \sum_{j=1}^{K_Y} \big| p_Y(j) - p_Y^*(j) \big|.$$

   Note that $A(p_Y) = 0$ if and only if p_Y is symmetric. The larger $A(p_Y)$ is, the more asymmetric (or unsymmetric) the marginal for Y is.

2. Skewness:
   The third standardized moment (skewness) is another measure of asymmetry:

   $$\gamma_Y = \frac{\mathbb{E}\big[(Y - \mu_Y)^3\big]}{\sigma_Y^3},$$

where μ_Y is the mean and σ_Y is the standard deviation of Y. For a symmetric distribution, $\gamma_Y = 0$. A nonzero skewness indicates asymmetry, with the magnitude indicating how far the distribution deviates from symmetry.

While entropy quantifies the uncertainty or "spread" of a distribution (the uniform distribution has maximum entropy), it does not directly measure symmetry. A distribution can have high entropy yet be asymmetric. Thus, if your goal is to quantify how far the marginal for Y deviates from being symmetric, the measures above—either the total variation distance from its reflection or the skewness—are more directly appropriate.

In summary:

- A marginal p_Y is symmetric if $p_Y(j) = p_Y(K_Y + 1 - j)$ for all j.
- To quantify how unsymmetric p_Y is relative to the uniform (which is symmetric), you can compute:
  - The symmetry index: $A(p_Y) = \sum_{j=1}^{K_Y} |p_Y(j) - p_Y(K_Y + 1 - j)|$.
  - The skewness: $\gamma_Y = \frac{\mathbb{E}[(Y-\mu_Y)^3]}{\sigma_Y^3}$.
- Entropy measures uncertainty rather than symmetry.

This way, as you vary the marginal distribution for Y away from uniformity, you can use these measures to quantify the degree of asymmetry.

---

5 Misc.

A variance-covariance matrix is always positive semi-definite (PSD) because of fundamental properties of variance, inner products, and expectations. Below, I'll break down why this is the case.

5.0.1 1. Definition of a Variance-Covariance Matrix

For a random vector $\mathrm{X} = (X_1, X_2, \ldots, X_n)^T$, the variance-covariance matrix (or simply covariance matrix) is:

$$= \mathbb{E}[(\mathrm{X} - \mathbb{E}[\mathrm{X}])(\mathrm{X} - \mathbb{E}[\mathrm{X}])^T]$$

Each entry in is given by:

$$\sigma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$$

where:

- Diagonal elements $\sigma_{ii} = \mathrm{Var}(X_i)$ are variances (always non-negative).
- Off-diagonal elements σ_{ij} are covariances.

5.0.2 2. Why is Always Positive Semi-Definite?

A matrix is positive semi-definite (PSD) if, for any nonzero vector z,

$$\mathrm{z}^T \ \mathrm{z} \geq 0$$

To see why this holds for :

1. Quadratic Form Interpretation
   Consider the quadratic form:

   $$\mathrm{z}^T \ \mathrm{z}$$

   Expanding using the definition of :

   $$\mathrm{z}^T \mathbb{E}[(\mathrm{X} - \mathbb{E}[\mathrm{X}])(\mathrm{X} - \mathbb{E}[\mathrm{X}])^T]\mathrm{z}$$

   Using the linearity of expectation:

   $$\mathbb{E}[\mathrm{z}^T(\mathrm{X} - \mathbb{E}[\mathrm{X}])(\mathrm{X} - \mathbb{E}[\mathrm{X}])^T\mathrm{z}]$$

This simplifies to:

$$\mathbb{E}[(\mathrm{z}^T(\mathrm{X} - \mathbb{E}[\mathrm{X}]))^2]$$

Since this is an expectation of a squared quantity, it is always non-negative, ensuring:

$$\mathrm{z}^T \ \mathrm{z} \geq 0$$

Thus, is positive semi-definite.

2. Variance is Always Non-Negative
   Since variance cannot be negative, the diagonal elements (variances) satisfy:

   $$\mathrm{Var}(X_i) = \sigma_{ii} \geq 0$$

   This ensures that the matrix does not have negative eigenvalues, a key requirement for PSD matrices.

3. Covariance as an Inner Product
   The covariance function behaves like an inner product in a Hilbert space, meaning it satisfies properties like:

   $$\mathrm{Cov}(X, X) \geq 0, \quad \mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$$

   Since inner product matrices are always PSD, inherits this property.

4. Eigenvalue Interpretation
   The eigenvalues of a covariance matrix represent the variances along principal component directions. Since variance is always non-negative, all eigenvalues are non-negative, ensuring that is PSD.

5.0.3 3. When is Positive Definite (PD)?

A matrix is positive definite (PD) if z^T z $>$ 0 for all nonzero z.
For the covariance matrix:

- is positive definite if and only if the random variables $X_1, \ldots, X_n$ are linearly independent.
- If at least one variable can be written as a linear combination of others, becomes singular (rank-deficient) and is only positive semi-definite.

5.0.4 4. Summary

Always PSD: A covariance matrix is always positive semi-definite because it is the expectation of a squared quantity.
Not always PD: If some variables are perfectly correlated (or one is a linear function of others), the matrix becomes singular and is not invertible.
Eigenvalue Rule: All eigenvalues of are non-negative, ensuring PSD.

5.1 The Role of Ties

5.2 1. Visualizing the Fréchet Bounds for Discrete Variables

Suppose we have two discrete variables X and Y with the following marginal distributions:

$$\begin{aligned} X: &\quad P(X=0)=0.4, \quad P(X=1)=0.4, \quad P(X=2)=0.2 \\ Y: &\quad P(Y=0)=0.3, \quad P(Y=1)=0.3, \quad P(Y=2)=0.4 \end{aligned}$$

We will try to construct the upper and lower Fréchet bounds and see when they are achievable.

5.3 2. Upper Fréchet Bound (Perfect Positive Dependence) (Achievable)

The upper Fréchet bound requires perfect positive dependence, meaning that high values of X should match high values of Y as closely as possible.

5.3.1 Step 1: Assign Probability Mass in Increasing Order

We sort X and Y in increasing order:

X	Probability $P(X)$	Y	Probability $P(Y)$
0	0.4	0	0.3
1	0.4	1	0.3
2	0.2	2	0.4

5.3.2 Step 2: Match Probabilities

We now assign probabilities in a way that ensures X and Y are as aligned as possible:

$$P(X = 0, Y = 0) = 0.3, \quad P(X = 0, Y = 1) = 0.1$$
$$P(X = 1, Y = 1) = 0.3, \quad P(X = 1, Y = 2) = 0.1$$
$$P(X = 2, Y = 2) = 0.2$$

This results in the following joint probability table:

$X \backslash Y$	0	1	2	Marginal $P(X)$
0	0.3	0.1	0	0.4
1	0	0.3	0.1	0.4
2	0	0	0.2	0.2
Marginal $P(Y)$	0.3	0.3	0.4	

5.3.3 Step 3: Verify that it Achieves the Upper Bound

- The probability mass aligns perfectly with the upper Fréchet bound formula:

$$P(X \leq x, Y \leq y) = \min(F_X(x), F_Y(y)).$$

- Every time X increases, Y also increases or stays the same.
- The upper bound is achievable!

5.4 3. Lower Fréchet Bound (Perfect Negative Dependence) (Not Achievable)

The lower Fréchet bound requires perfect negative dependence, meaning that high values of X should match low values of Y and vice versa.

5.4.0.1 Step 1: Assign Probability Mass in Opposite Order

We sort X in increasing order but Y in decreasing order:

X	Probability $P(X)$	Y	Probability $P(Y)$
0	0.4	2	0.4
1	0.4	1	0.3
2	0.2	0	0.3

5.4.0.2 Step 2: Try to Assign Probabilities

We now try to pair the lowest values of X with the highest values of Y to enforce negative dependence.

$$P(X=0, Y=2) = 0.4, \quad P(X=1, Y=1) = 0.3, \quad P(X=2, Y=0) = 0.2.$$

This results in the following joint probability table:

$X \backslash Y$	0	1	2	Marginal $P(X)$
0	0	0	0.4	0.4

$X \backslash Y$	0	1	2	Marginal $P(X)$
1	0	0.3	0.1	0.4
2	0.2	0	0	0.2
Marginal $P(Y)$	0.3	0.3	0.4	

5.4.0.3 Step 3: Check for Validity

- We see a problem:
  - $P(Y = 2) = 0.4$, but our assignment forces all of it to $X = 0$.
  - $P(Y = 0) = 0.3$, but we only assigned 0.2 to $X = 2$.

Thus, the marginals are not satisfied. The lower bound cannot be realized because we ran out of probability mass to assign!

5.4.1 4. Why Does This Happen?

- The upper Fréchet bound only requires that we pair increasing values in a monotonic fashion, which can usually be done locally at each probability level.
- The lower Fréchet bound requires strict inversion, which demands exact mass balancing across different probability levels.

If the probability masses do not align perfectly, the countermonotonic structure cannot be realized.

5.4.2 5. Summary: Why the Upper Bound is Easier to Achieve

Bound	Definition	Achievability in Discrete Case	Why?
Upper Fréchet Bound (Perfect Positive Dependence)	$P(X \leq x, Y \leq y) = \min(F_X(x), F_Y(y))$	Often Achievable	Sorting in the same order always works, even with ties.

Bound	Definition	Achievability in Discrete Case	Why?
Lower Fréchet Bound (Perfect Negative Dependence)	$P(X \leq x, Y \leq y) = \max(0, F_X(x) + F_Y(y) - 1)$	Often Not Achievable	Requires exact probability mass balancing, which is often impossible.

5.4.3 Final Conclusion

Even though the upper and lower Fréchet bounds look symmetric in their formulas, they are not symmetric in achievability for discrete distributions. The upper bound only requires local matching of probabilities, while the lower bound requires global balancing, which is much harder to satisfy.

Thus, for discrete variables:

- The upper Fréchet bound is often achievable.
- The lower Fréchet bound is often not achievable because exact countermonotonicity is difficult to construct.

This explains why $r_{\min} \neq -r_{\max}$ in general when X and Y have different numbers of levels.

$\square$

You:

Ah, i think i see. does it have to do with 'ties'?

ChatGPT:

Yes, ties play a crucial role in why the upper Fréchet bound is often achievable, while the lower Fréchet bound may not be achievable in discrete settings! Let's take a deeper look at how ties in ranks and probability masses affect achievability.

5.5 1. How Do Ties Affect Fréchet Bounds?

Ties in probability mass occur when:

- Discrete variables have repeated values with the same probability.
- The CDF jumps at certain points instead of increasing smoothly (which happens in continuous distributions).
- The number of unique levels in X and Y are different or their probabilities do not match perfectly.

Ties influence achievability differently for the upper and lower Fréchet bounds.

5.6 2. Upper Fréchet Bound: Why Ties Are Not a Problem

The upper Fréchet bound enforces perfect positive dependence:

$$F^+(x, y) = \min(F_X(x), F_Y(y))$$

5.7 Why Ties Do Not Break Achievability

The upper bound only requires sorting X and Y in the same order.

- If multiple values have the same probability, they can simply be paired together without conflict.
- The local structure of the PMF allows mass to be matched easily.

Example: Consider:

$$X : (0, 0, 1, 1, 2), \quad Y : (0, 0, 1, 1, 2)$$

We can pair them identically, and the structure remains valid. No probability mass is "left over."

5.7.1 Key Insight:

- Since we are always pairing ranks in the same order, ties do not create infeasibility.
- We can still respect the marginal distributions because increasing values in X correspond to increasing values in Y, even with ties.

Conclusion: The upper bound is often achievable because ties do not interfere with its construction.

5.8 3. Lower Fréchet Bound: Why Ties Make It Infeasible

The lower Fréchet bound enforces perfect negative dependence:

$$F^{-}(x,y) = \max(0, F_X(x) + F_Y(y) - 1)$$

5.8.1 Why Ties Break Achievability

Ties create a mass-balancing problem when trying to enforce countermonotonicity.

- To achieve perfect negative dependence, we must invert the ranks of X and Y.
- But if multiple values have the same probability, we cannot "spread out" the mass properly across the ranks.

Example: Consider:

$$X : (0,0,1,1,2), \quad Y : (2,2,1,1,0)$$

We try to pair them as:

$$(X,Y) = [(0,2),(0,2),(1,1),(1,1),(2,0)]$$

Problem:

- We need $P(X=0, Y=2) = P(X=0) = 0.4$, but $P(Y=2) = 0.3$.
- There is too much probability mass at one level of X that cannot be matched with the corresponding Y.
- This violates the marginals, making the lower bound infeasible.

5.8.2 Key Insight:

- Unlike the upper bound (where mass is aligned locally), the lower bound requires global balancing.
- If ties concentrate too much mass at one value, we cannot distribute it correctly while maintaining countermonotonicity.

Conclusion: The lower bound is often NOT achievable because ties prevent exact probability mass balancing.

5.9 4. Summary: Why Ties Matter

Bound	Definition	Effect of Ties	Achievability
Upper Fréchet Bound (Comonotonicity)	$F^+(x,y) = \min(F_X(x), F_Y(y))$	Ties do not cause problems because ranks are sorted identically.	Often Achievable
Lower Fréchet Bound (Countermonotonicity)	$F^-(x,y) = \max(0, F_X(x) + F_Y(y) - 1)$	Ties break countermonotonicity because mass cannot be spread properly.	Often Not Achievable

Thus, the presence of ties makes the lower bound infeasible in many discrete settings, explaining why:

$$r_{\min} \neq -r_{\max}$$

when X and Y have different numbers of levels or mismatched probabilities.

5.9.1 Final Conclusion

Ties do not affect the upper Fréchet bound because we can sort and pair ranks without constraint.

Ties break the lower Fréchet bound because they prevent proper balancing of probability mass across opposite ranks, making perfect negative dependence impossible in many discrete cases.

This is the fundamental reason why the lower Fréchet bound is often not achievable for discrete variables, but the upper bound is!

$\square$