Application of Balance Matrices to the Interpretation of Negative Probabilities (working draft)

Orges Leka

February 12, 2025

Abstract

In this paper, we develop a comprehensive framework for interpreting negative probabilities via balance matrices. By decomposing a quasi-probability matrix into a proper joint probability matrix and a balance matrix—whose row and column sums vanish—we preserve the marginal distributions while isolating the non-classical (negative) components. Our approach leverages iterative proportional fitting to implement this decomposition and explores the algebraic properties of balance matrices, including ring isomorphisms and the behavior of Moore-Penrose inverses. Beyond establishing a rigorous mathematical foundation, we discuss the potential of these methods to address challenges in quantum mechanics, game design, and other disciplines where non-standard probabilities emerge.

Contents

1 Introduction

Negative probabilities have long presented a conceptual challenge, appearing in diverse contexts from quantum mechanics to economic modeling. In traditional probability theory, all probabilities are nonnegative; however, certain phenomena—such as interference effects in quantum systems or subtle biases in complex systems—can give rise to quantities that behave like probabilities yet take on negative values.

This paper introduces balance matrices as a powerful tool to reinterpret negative probabilities. By decomposing a given quasi-probability matrix Q into the sum

$$Q = P + B,$$

where P is a standard joint probability matrix and B is a balance matrix (with all row and column sums equal to zero), we preserve the observable marginals while isolating the contributions responsible for negativity. In doing so, we not only reconcile non-classical probabilities with conventional theory but also provide an algebraic framework complete with ring isomorphisms, pseudoinverse formulations, and effective computational techniques such as iterative proportional fitting (IPF).

Our work develops both the theoretical underpinnings and practical implementations of this approach. By connecting abstract algebraic properties with applied algorithms, we aim to offer a versatile method for researchers confronting negative probabilities in a variety of settings.

2 Definitions

In this section we summarize the key definitions introduced throughout the paper. For clarity, we list them here:

1. **Joint Probability Matrix.** Let X and Y be discrete random variables with outcome spaces

$$\{x_1, x_2, \ldots, x_m\} \quad \text{and} \quad \{y_1, y_2, \ldots, y_n\},$$

respectively. The joint probability matrix $P(X, Y)$ is defined by

$$P(X,Y) = \begin{bmatrix} P(x_1, y_1) & P(x_1, y_2) & \cdots & P(x_1, y_n) \\ P(x_2, y_1) & P(x_2, y_2) & \cdots & P(x_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ P(x_m, y_1) & P(x_m, y_2) & \cdots & P(x_m, y_n) \end{bmatrix}.$$

Each entry represents the probability that $X = x_i$ and $Y = y_j$.

2. **Normalization.** A probability matrix is normalized if the sum of all its entries equals one:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} P(x_i, y_j) = 1.$$

3. **Non-Negativity.** A probability matrix satisfies non-negativity if every entry is greater than or equal to zero:

$$P(x_i, y_j) \geq 0 \quad \text{for all } i, j.$$

4. **Marginal Distributions.** The marginal distribution of X (or Y) is obtained by summing the joint probabilities over the outcomes of Y (or X):

$$P(X = x_i) = \sum_{j=1}^{n} P(x_i, y_j) \quad \text{and} \quad P(Y = y_j) = \sum_{i=1}^{m} P(x_i, y_j).$$

5. **Conditional Distributions.** Given $P(X = x_i) > 0$, the conditional probability of Y given $X = x_i$ is defined by

$$P(Y = y_j \mid X = x_i) = \frac{P(x_i, y_j)}{P(X = x_i)}.$$

6. **Independence.** Two random variables X and Y are independent if

$$P(x_i, y_j) = P(X = x_i)\, P(Y = y_j) \quad \text{for all } i, j.$$

7. **Balance Matrix.** A balance matrix B is a matrix (not necessarily square) whose rows and columns each sum to zero:

$$\sum_j B_{ij} = 0 \quad \text{and} \quad \sum_i B_{ij} = 0.$$

8. **Joint Quasi-Probability Matrix.** Given a balance matrix B and a joint probability matrix P, the joint quasi-probability matrix Q is defined as

$$Q = P + B.$$

Unlike P, the entries of Q may fall outside the interval $[0, 1]$ while still summing to one.

9. **Iterative Proportional Fitting (IPF).** IPF is an algorithm used to adjust the entries of a matrix so that its row and column sums match given target margins. In our context, it is applied to scale an initial matrix to have the same marginals as a given quasi-probability matrix.

10. **Semantic Space.** A semantic space is defined on a finite set Ω together with a kernel

$$k : \Omega \times \Omega \to [-1, 1],$$

which is positive semidefinite, satisfies $k(\alpha, \alpha) = 1$ for all $\alpha \in \Omega$, and has the property that $k(\alpha, \beta) = 1$ if and only if $\alpha = \beta$.

11. **Balance Space.** A balance space is a triple (Ω, Σ, B) where:
    - Ω is an outcome set.
    - Σ is a σ-algebra over Ω.
    - $B : \Sigma \to \mathbb{R}$ is a σ-additive function with $B(\Omega) = 0$.

12. **Moore-Penrose Inverse.** Given a matrix (or operator) X, its Moore-Penrose inverse X^+ is the unique matrix satisfying the four Penrose conditions. In our work, it is used to define pseudoinverses for balance matrices.

13. **Quasi-Probability.** A quasi-probability is a generalization of a probability distribution that allows for negative values. The decomposition $Q = P + B$ serves to separate the classical (probability) part P from the non-classical (balance) part B.

14. **Cholesky Decomposition.** For a positive semidefinite matrix, the Cholesky decomposition factors it as a product of a lower triangular matrix and its transpose. This technique is used to obtain the mapping ϕ in the construction of a semantic space.

3 Joint Quasi-Probability Matrix

Suppose we have a *balance matrix* B, which is a matrix (not necessarily square) whose row sums and column sums are all zero, i.e.,

$$\sum_j B_{ij} = 0 \quad \text{for all } i, \qquad \sum_i B_{ij} = 0 \quad \text{for all } j.$$

Let P be a joint probability matrix, so that

$$\sum_{i,j} P_{ij} = 1 \quad \text{and} \quad P_{ij} \geq 0 \quad \text{for all } i, j.$$

We then define the *joint quasi-probability matrix* Q by

$$Q = B + P.$$

The matrix Q exhibits the following properties:

1. **Normalization:** Since B has zero sum over all its entries (because its rows and columns sum to zero), the total sum of Q is preserved:

$$\sum_{i,j} Q_{ij} = 1.$$

2. **Marginal Preservation:** The marginal distributions of Q are identical to those of P, since for any fixed row or column the contribution from B is zero.

3. **Quasi-Probability Nature:** Although every entry of P is nonnegative and bounded between 0 and 1, the balance matrix B may include negative values or values exceeding 1. Thus, some entries of Q may lie outside the interval $[0, 1]$, justifying the designation as a quasi-probability matrix.

4. **Dimensional Flexibility:** Neither B nor P need be square; accordingly, Q may be rectangular, accommodating random variables with differing numbers of outcomes.

4 Balance Matrices and Their Properties

In this section we summarize the key theorems and properties for matrices whose row and column sums are zero (i.e., *balance matrices*), covering both the square and rectangular cases. For brevity, only the statements are given. The interested reader can find the original proofs and statements in [1], from which we have copied the theorems here:

4.1 Square Balance Matrices

Theorem 4.1 (Bijection)**.** *There exists a bijection*

$$\phi : M_n \to S_{n+1}, \quad X \mapsto J_n^* X J_n,$$

where

$$J_n = [\, I_n \mid -\mathbf{1}\,],$$

and S_{n+1} *denotes the set of* $(n+1) \times (n+1)$ *matrices with all row and column sums equal to zero.*

Corollary 4.2 (Rank Preservation)**.** *If* $\tilde{X} = J_n^* X J_n$*, then* $\tilde{X}$ *and* X *have the same rank.*

Corollary 4.3 (Self-Adjointness Preservation)**.** *We have* $\tilde{X} = \tilde{X}^*$ *if and only if* $X = X^*$*.*

Theorem 4.4 (Ring Isomorphism)**.** *Let* $\times$ *denote standard matrix multiplication and define the twisted product* $\circ$ *on* M_n *by*

$$X \circ Y = X K_n Y, \quad \text{with } K_n = J_n J_n^*.$$

Then ϕ *is an isomorphism between the rings* $(S_{n+1}, +, \times)$ *and* $(M_n, +, \circ)$*.*

Theorem 4.5 (Identity Element)**.** *The ring* $(S_{n+1}, +, \times)$ *has a unique multiplicative identity given by*

$$\phi(K_n^{-1}) = J_n^* K_n^{-1} J_n = I_{n+1} - \frac{1}{n+1} \mathbf{1}_{(n+1)\times(n+1)},$$

where $\mathbf{1}_{(n+1)\times(n+1)}$ *denotes the* $(n+1) \times (n+1)$ *matrix of all ones.*

Theorem 4.6 (Moore-Penrose Inverse for Full-Rank Square Balance Matrices)**.** *For any rank* n *matrix* $\tilde{X} \in S_{n+1}$ *with* $\tilde{X} = J_n^* X J_n$*, the unique Moore-Penrose inverse of* $\tilde{X}$ *is*

$$\tilde{X}^+ = J_n^* K_n^{-1} X^{-1} K_n^{-1} J_n.$$

Corollary 4.7. *For any rank* n *matrix* $\tilde{X} \in S_{n+1}$*,*

$$\tilde{X}\tilde{X}^+ = \tilde{X}^+\tilde{X} = J_n^* K_n^{-1} J_n = I_{n+1} - \frac{1}{n+1} \mathbf{1}_{(n+1)\times(n+1)}.$$

4.2 Rectangular Balance Matrices

For matrices with zero row and column sums that are not necessarily square, one may represent them as

$$\tilde{X} = J_m^* X J_n,$$

where X is an $m \times n$ matrix.

Theorem 4.8 (Moore-Penrose Inverse for Rectangular Balance Matrices)**.** *For* $\tilde{X} = J_m^* X J_n$*, the Moore-Penrose inverse is given by*

$$\tilde{X}^+ = J_n^* (k_n^{-1})^* (k_m^* X k_n)^+ k_m^{-1} J_m,$$

where $K_m = J_m J_m^* = k_m k_m^*$ *and* $K_n = J_n J_n^* = k_n k_n^*$*.*

Theorem 4.9 (Row-Only and Column-Only Zero Sum Cases)**.**

1. *If* $\tilde{X} = X J_n$ *with* X *left-invertible, then*

$$\tilde{X}^+ = J_n^* K_n^{-1} X^+.$$

2. *If* $\tilde{X} = J_n^* X$ *with* X *right-invertible, then*

$$\tilde{X}^+ = X^+ K_n^{-1} J_n.$$

4.3 Balance Matrices with Extra Zero Rows or Columns

Let $\mathbf{a}$ be a fixed index set (with m entries, $m < n$) specifying the positions where rows and/or columns are identically zero. Denote by $S_{n+1}^{\mathbf{a}}$ the set of $(n+1) \times (n+1)$ matrices with all row and column sums zero and with zeros in the rows and columns indexed by $\mathbf{a}$.

Theorem 4.10 (Isomorphism for Matrices with Extra Zeros)**.** *There exists an isomorphism*

$$\phi : (M_{n-m}, +, \circ) \to (S_{n+1}^{\mathbf{a}}, +, \times), \quad X \mapsto J_{m,\mathbf{a}}^{*} X J_{m,\mathbf{a}},$$

where $J_{m,\mathbf{a}}$ is defined by inserting zero columns (and rows) at the positions indicated by $\mathbf{a}$, and the product $\circ$ is defined by $X \circ Y = X K_m Y$ with $K_m = J_{m,\mathbf{a}} J_{m,\mathbf{a}}^{*}$.

Theorem 4.11 (Moore-Penrose Inverse for Matrices with Extra Zeros)**.** *For a matrix $\tilde{X} = J_{m,\mathbf{b}}^{*} X J_{m,\mathbf{a}}$ of rank m, the Moore-Penrose inverse is*

$$\tilde{X}^{+} = J_{m,\mathbf{a}}^{*} K_m^{-1} X^{-1} K_m^{-1} J_{m,\mathbf{b}}.$$

Theorem 4.12 (Projection Property)**.** *For $\tilde{X} = J_{m,\mathbf{b}}^{*} X J_{m,\mathbf{a}}$ of rank m,*

$$\tilde{X}^{+} \tilde{X} = J_{m,\mathbf{a}}^{*} K_m^{-1} J_{m,\mathbf{a}}.$$

Theorem 4.13 (Invariance under Projection)**.** *Let M be any matrix with m rows, and let $\tilde{X} = J_{m,\mathbf{b}}^{*} X J_{m,\mathbf{a}}$ be a rank m square matrix. If $\tilde{M} = J_{m,\mathbf{a}}^{*} M$, then*

$$\tilde{X}^{+} \tilde{X} \tilde{M} = \tilde{M}.$$

Theorem 4.14 (Range of the Projection Operator)**.** *The vectors of the form $\tilde{M} = J_{m,\mathbf{a}} M$ constitute an m-dimensional subspace that spans the range of the projection operator $\tilde{X}^{+} \tilde{X}$.*

5 Iterative Proportional Fitting and Quasi-Probability Matrices

In this section we describe the iterative proportional fitting (IPF) algorithm and its role in decomposing a quasi-probability matrix Q into a joint probability matrix P and a balance matrix B such that

$$Q = P + B.$$

The balance matrix B has the property that all its row and column sums are zero, ensuring that the marginal distributions of Q are identical to those of P.

5.1 Iterative Proportional Fitting (IPF)

Iterative proportional fitting is an algorithm used to adjust the entries of an $m \times n$ matrix so that its row and column sums match given target margins. Suppose we are given an initial matrix

$$A = (a_{ij}), \quad 1 \le i \le m, \ 1 \le j \le n,$$

and target row sums $R_1, \ldots, R_m$ and column sums $C_1, \ldots, C_n$. The IPF algorithm proceeds as follows:

1. **Initialization:** Choose a starting matrix $A^{(0)}$ (typically with all positive entries).
2. **Row Adjustment:** For each row i, compute the current row sum

$$S_i^{(k)} = \sum_{j=1}^{n} a_{ij}^{(k)},$$

and update every entry in that row by multiplying by $\frac{R_i}{S_i^{(k)}}$:

$$a_{ij}^{(k+1/2)} = a_{ij}^{(k)} \frac{R_i}{S_i^{(k)}}.$$

3. **Column Adjustment:** For each column j, compute the new column sum

$$T_j^{(k+1/2)} = \sum_{i=1}^{m} a_{ij}^{(k+1/2)},$$

and update every entry in that column by multiplying by $\frac{C_j}{T_j^{(k+1/2)}}$:

$$a_{ij}^{(k+1)} = a_{ij}^{(k+1/2)} \frac{C_j}{T_j^{(k+1/2)}}.$$

4. **Iteration:** Repeat the row and column adjustments until the row and column sums converge to the targets.

Example: Let

$$A^{(0)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

with desired row sums $R_1 = 4$ and $R_2 = 6$ and desired column sums $C_1 = C_2 = 5$.

Row adjustment:

- Row 1 sum $= 1 + 1 = 2$. Multiply row 1 by $\frac{4}{2} = 2$ to obtain $[2, 2]$.
- Row 2 sum $= 1 + 1 = 2$. Multiply row 2 by $\frac{6}{2} = 3$ to obtain $[3, 3]$.

Thus, the intermediate matrix is

$$A^{(1/2)} = \begin{bmatrix} 2 & 2 \\ 3 & 3 \end{bmatrix}.$$

Column adjustment: Both columns already sum to $2 + 3 = 5$, so no further scaling is required. The resulting matrix satisfies the desired margins:

$$A^{(1)} = \begin{bmatrix} 2 & 2 \\ 3 & 3 \end{bmatrix}.$$

5.2 Decomposition of Quasi-Probability Matrices

A matrix Q is called a *quasi-probability matrix* if:

1. $\sum_{i=1}^{m}\sum_{j=1}^{n} Q_{ij} = 1$,
2. $0 \le \sum_{i=1}^{m} Q_{ij} \le 1$ for every column j,
3. $0 \le \sum_{j=1}^{n} Q_{ij} \le 1$ for every row i.

Given such a Q, we wish to decompose it as

$$Q = P + B,$$

where P is a joint probability matrix (with nonnegative entries summing to 1) and B is a balance matrix (with zero row and column sums).

One approach is as follows:

1. Let
$$M = \max_{i,j} |q_{ij}|.$$
Define the matrix
$$A = (a_{ij}) \quad \text{with} \quad a_{ij} = \frac{|q_{ij}|}{M},$$
so that $a_{ij} \in [0, 1]$.
2. Apply the IPF algorithm to A using the target row and column sums taken from Q. Let the resulting matrix be $\widehat{A}$ and set
$$P := \widehat{A}.$$
3. Define the balance matrix as
$$B := Q - P.$$

Since the balance matrix B satisfies

$$\sum_j B_{ij} = \sum_j (q_{ij} - p_{ij}) = (\text{row sum of } Q) - (\text{row sum of } P) = 0,$$

(and similarly for column sums), the decomposition is valid.

Below is some SageMath code that implements the above method:

```
# Define the iterative proportional fitting (IPF) algorithm.
def ipf(A, row_targets, col_targets, tol=1e-6, max_iter=1000):
    A = matrix(RR, A)  # ensure we work with real numbers
    m, n = A.nrows(), A.ncols()
    for it in range(max_iter):
        # Row adjustment
        for i in range(m):
```

```
            current_row_sum = sum(A[i,j] for j in range(n))
            if current_row_sum != 0:
                factor = row_targets[i] / current_row_sum
            else:
                factor = 1
            for j in range(n):
                A[i,j] *= factor
        # Column adjustment
        for j in range(n):
            current_col_sum = sum(A[i,j] for i in range(m))
            if current_col_sum != 0:
                factor = col_targets[j] / current_col_sum
            else:
                factor = 1
            for i in range(m):
                A[i,j] *= factor
        # Check convergence
        row_diffs = [abs(sum(A[i,j] for j in range(n)) - row_targets[i]) for i in range(m)]
        col_diffs = [abs(sum(A[i,j] for i in range(m)) - col_targets[j]) for j in range(n)]
        if max(row_diffs + col_diffs) < tol:
            break
    return A

# Function to decompose a quasi-probability matrix Q into a joint probability matrix P and a balance matrix B.
def decompose_qpm(Q, tol=1e-6, max_iter=1000):
    Q = matrix(RR, Q)
    m, n = Q.nrows(), Q.ncols()
    # Compute target row and column sums from Q.
    row_targets = [sum(Q[i,j] for j in range(n)) for i in range(m)]
    col_targets = [sum(Q[i,j] for i in range(m)) for j in range(n)]
    # Check normalization.
    total = sum(row_targets)
    if abs(total - 1) > tol:
        print("Warning: Q is not normalized; total sum =", total)
    # Determine the maximum absolute entry in Q.
    M = max(abs(Q[i,j]) for i in range(m) for j in range(n))
    if M == 0:
        raise ValueError("Matrix Q is zero!")
    # Construct matrix A with entries a_ij = |q_ij|/M.
    A = Q.apply_map(lambda x: abs(x)/M)
    # Apply IPF to A with targets from Q.
    P = ipf(A, row_targets, col_targets, tol=tol, max_iter=max_iter)
    # Define the balance matrix B = Q - P.
    B = Q - P
    return P, B

# Example usage:
# Define a quasi-probability matrix Q.
Q = matrix(RR, [[0.2, 0.1],
                [-0.1, 0.8]])
print("Quasi-Probability Matrix Q:")
print(Q)

# Decompose Q into P (joint probability matrix) and B (balance matrix).
P, B = decompose_qpm(Q)
print("Joint Probability Matrix P:")
print(P)
```

```python
print("Balance Matrix B:")
print(B)

# Verify that Q = P + B.
print("Verification (P + B):")
print(P + B)
```

The above code first defines the IPF procedure, then uses it to adjust the absolute value matrix derived from Q so that the resulting matrix P has the same row and column sums as Q. Finally, the balance matrix $B = Q - P$ is computed. Note that the decomposition is not unique; alternative decompositions may exist.

6 Half-Coin Example of G. Székely

Let

$$q_n = (-1)^{n-1}\sqrt{2}\,\frac{C_{n-1}}{4^n}, \quad n = 0, 1, 2, \ldots,$$

where

$$C_n = \frac{\binom{2n}{n}}{n+1}, \quad n = 0, 1, 2, \ldots, \quad \text{and} \quad C_{-1} = -\frac{1}{2}.$$

Then, as has been shown in [2], we have

$$\sum_{n \geq 0} q_n = 1 \quad \text{and} \quad \sum_{n \geq 0} |q_n| = \sqrt{2}.$$

Define

$$p_n := \frac{|q_n|}{\sqrt{2}}.$$

Then, we have

$$0 \leq p_n \leq 1 \quad \text{and} \quad \sum_{n \geq 0} p_n = 1.$$

Now, set

$$b_n := q_n - p_n.$$

Then, the balance property holds:

$$\sum_{n \geq 0} b_n = \sum_{n \geq 0} q_n - \sum_{n \geq 0} p_n = 1 - 1 = 0.$$

7 Balance Spaces

In the previous example, set

$$\mathbf{B} := \big(\Omega = \{0, 1, 2, \ldots\},\ \Sigma = 2^{\Omega},\ B\big),$$

with

$$B(A) := \sum_{n \in A} b_n \quad \text{for } A \in \Sigma.$$

This gives rise to a *balance space* defined by:

- Σ is a σ-algebra over Ω.
- B is a σ-additive function, i.e., $B : \Sigma \to \mathbb{R}$.
- $B(\Omega) = 0$.

A quasi-probability space $\mathbf{Q}$ is defined as

$$\mathbf{Q} = (\Omega, \Sigma, Q)$$

such that:

- Σ is a σ-algebra over Ω.
- Q is a σ-additive function $Q : \Sigma \to \mathbb{R}$.
- $Q(\Omega) = 1$.

Hence, in the previous example we have shown how to decompose the quasi-probability function defined by q_n into a balance function and a probability function:

$$B(\{n\}) := b_n = q_n - p_n = Q(\{n\}) - P(\{n\}),$$

with all three functions Q, B, and P defined on the same (Ω, Σ) tuple. It would be interesting to see if this decomposition can be done for a general quasi-probability space, analogous to the decomposition discussed in the matrix section.

8 Applications in Slot Game Design

The decomposition $Q = B + P$ offers a useful tool for game design in slot games. A game designer, who may not be a mathematician, might devise a mechanism in which the advertised "probabilities" q_{ij} sum to 1 yet sometimes take on negative values, hence the usual theorems of probability theory can not be applied in this setting and the game rules must be changed.

For the mathematician responsible for the probabilistic analysis, the approach is to decompose Q via iterative proportional fitting into a balance matrix B and a true probability matrix P (i.e., $Q = B + P$). The resulting matrix $P = [p_{ij}]$ is then proposed as the "corrected probabilities" for the game. This method ensures that Q and P have the same row and column sums, preserving the marginal distributions while providing a mathematically consistent set of probabilities for the game design.

8.1 Example: Decomposition in Game Design

Below is an example illustrating the method.

Consider the following 2×2 quasi-probability matrix Q:

$$Q = \begin{pmatrix} 0.3000 & 0.1000 \\ -0.0500 & 0.6500 \end{pmatrix}.$$

The row sums of Q are:

$$0.3000 + 0.1000 = 0.4000, \quad \text{and} \quad -0.0500 + 0.6500 = 0.6000.$$

The column sums of Q are:

$$0.3000 + (-0.0500) = 0.2500, \quad \text{and} \quad 0.1000 + 0.6500 = 0.7500.$$

The maximum absolute value in Q is

$$M = 0.6500.$$

We then construct a matrix A by scaling the absolute values of Q by M:

$$A = \left(\frac{|q_{ij}|}{M} \right) = \begin{pmatrix} 0.4615 & 0.1538 \\ 0.0769 & 1.0000 \end{pmatrix}.$$

The target marginals (row and column sums) are the same as those of Q:

- **Row targets:** [0.4000, 0.6000]
- **Column targets:** [0.2500, 0.7500]

By applying the iterative proportional fitting (IPF) algorithm to A with these targets, we obtain the corrected probability matrix P:

$$P = \begin{pmatrix} 0.2299 & 0.1701 \\ 0.0201 & 0.5799 \end{pmatrix}.$$

The row sums of P are approximately:

$$0.2299 + 0.1701 \approx 0.4000, \quad \text{and} \quad 0.0201 + 0.5799 \approx 0.6000,$$

and the column sums of P are:

$$0.2299 + 0.0201 = 0.2500, \quad \text{and} \quad 0.1701 + 0.5799 = 0.7500.$$

Finally, the balance matrix B is computed as:

$$B = Q - P = \begin{pmatrix} 0.3000 - 0.2299 & 0.1000 - 0.1701 \\ -0.0500 - 0.0201 & 0.6500 - 0.5799 \end{pmatrix} = \begin{pmatrix} 0.0701 & -0.0701 \\ -0.0701 & 0.0701 \end{pmatrix}.$$

The row sums and column sums of B are essentially zero (up to numerical rounding), confirming that B is indeed a balance matrix.

Explanation of the Method:

1. **Initial Quasi-Probability Matrix Q:** The matrix Q is specified with entries that sum to 1 but may include negative values.
2. **Scaling to Form Matrix A:** We compute the maximum absolute value $M = 0.65$ in Q and form the matrix A whose entries are given by $\frac{|q_{ij}|}{M}$. This scales all entries into the interval $[0, 1]$.
3. **Setting Target Marginals:** The target row and column sums (i.e., 0.4000 and 0.6000 for rows; 0.2500 and 0.7500 for columns) are determined from Q. These marginals remain preserved throughout the decomposition.

4. **Iterative Proportional Fitting (IPF):** The IPF algorithm is applied to A to adjust its entries until the resulting matrix P matches the target marginals. This yields a corrected probability matrix P with all nonnegative entries.

5. **Computing the Balance Matrix B:** Finally, the balance matrix is obtained as $B = Q - P$. By construction, B has row and column sums equal to zero, capturing the hidden bias inherent in the original quasi-probability matrix Q.

This decomposition $Q = B + P$ is particularly useful in game design. For example, a slot game designer may initially specify a quasi-probability matrix Q that appears to offer fair odds (since the marginals match expected values), even though some entries are negative. The IPF-based decomposition then produces a corrected probability matrix P that can be used for rigorous analysis, while the balance matrix B reveals the underlying bias ensuring the game's profitability.

9 Uniqueness of the Decomposition

One natural way to achieve a unique decomposition of a quasi-probability matrix Q into a proper probability matrix P and a balance matrix B (i.e., $Q = P + B$) is to select P as the unique solution of the following optimization problem:

$$\min_{P \in \mathcal{P}} \|P - Q\|_F,$$

where $\|\cdot\|_F$ denotes the Frobenius norm and $\mathcal{P}$ is the set of all joint probability matrices having the same row and column sums as Q. In other words, $\mathcal{P}$ is the transportation polytope defined by the constraints

$$\sum_j p_{ij} = r_i \quad \text{for each row } i, \quad \text{and} \quad \sum_i p_{ij} = c_j \quad \text{for each column } j,$$

with

$$r_i = \sum_j q_{ij} \quad \text{and} \quad c_j = \sum_i q_{ij}.$$

Since the function

$$f(P) = \|P - Q\|_F^2 = \sum_{i,j} (p_{ij} - q_{ij})^2$$

is strictly convex on $\mathbb{R}^{m \times n}$ (its Hessian is $2I$), its restriction to the convex set $\mathcal{P}$ remains strictly convex. Consequently, there is a unique minimizer P^* of $f(P)$ over $\mathcal{P}$. Once this unique probability matrix P^* is obtained, the balance matrix is given by

$$B = Q - P^*.$$

This construction ensures that B automatically has zero row and column sums (since Q and P^* share the same marginals), and the decomposition $Q = B + P^*$ is unique.

Thus, minimizing the Frobenius norm $\|P - Q\|_F$ subject to P having the same row and column sums as Q and nonnegative entries indeed ensures a unique decomposition.

9.1 Theoretical Method and Python Implementation for Unique Decomposition

One natural way to achieve a unique decomposition of a quasi-probability matrix Q into a proper probability matrix P and a balance matrix B (i.e. $Q = P + B$) is to choose P as the unique solution to the following optimization problem:

$$\min_{P \in \mathcal{P}} \|P - Q\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius norm and

$$\mathcal{P} = \left\{ P \in \mathbb{R}^{m \times n} \; : \; P_{ij} \geq 0, \; \sum_j p_{ij} = r_i, \; \sum_i p_{ij} = c_j \right\}$$

is the transportation polytope defined by the constraints

$$r_i = \sum_j q_{ij} \quad \text{for each row } i, \quad \text{and} \quad c_j = \sum_i q_{ij} \quad \text{for each column } j.$$

Since the objective function

$$f(P) = \|P - Q\|_F^2 = \sum_{i,j} (p_{ij} - q_{ij})^2$$

is strictly convex, its restriction to the convex set $\mathcal{P}$ is strictly convex. Therefore, there exists a unique minimizer P^* that we denote by P^*. The balance matrix is then given by

$$B = Q - P^*.$$

This choice of P^* ensures that B automatically satisfies the zero-sum conditions on its rows and columns.

A practical way to compute P^* is by using quadratic programming. Below is an example implementation in Python using the `cvxpy` library.

Python Code Implementation:

```python
import cvxpy as cp
import numpy as np

# Define the quasi-probability matrix Q
Q = np.array([[0.3, 0.1],
              [-0.05, 0.65]])

m, n = Q.shape

# Compute the target marginals from Q
r = Q.sum(axis=1)  # row sums: [0.4, 0.6]
c = Q.sum(axis=0)  # column sums: [0.25, 0.75]

# Define the variable P (the corrected probability matrix)
P = cp.Variable((m, n))

# Define the objective: minimize the Frobenius norm squared of (P - Q)
objective = cp.Minimize(cp.sum_squares(P - Q))

```

```
# Define the constraints: nonnegativity and prescribed row/column sums
constraints = [P >= 0,
               cp.sum(P, axis=1) == r,
               cp.sum(P, axis=0) == c]

# Formulate and solve the problem
prob = cp.Problem(objective, constraints)
prob.solve()

# Extract the unique corrected probability matrix P*
P_star = P.value
B = Q - P_star  # the balance matrix

print("Corrected Probability Matrix P*:")
print(P_star)
print("Balance Matrix B:")
print(B)
```

Explanation:

1. We start with the quasi-probability matrix Q, whose entries sum to 1 but may include negative values.

2. The target row sums r_i and column sums c_j are computed directly from Q.

3. The optimization problem is set up to minimize the Frobenius norm $\|P - Q\|_F^2$ subject to P being a joint probability matrix—that is, P must be nonnegative and have the same row and column sums as Q.

4. Since the problem is strictly convex, the solver finds the unique minimizer P^* (denoted here as `P_star`).

5. Finally, the balance matrix B is calculated as the difference $Q - P^*$.

This approach guarantees a unique decomposition $Q = B + P^*$ and can be a valuable tool in applications such as game design, where one might need to correct quasi-probabilities into a proper probability distribution.

10 Semantic spaces on the same sample set

The goal of this section is given a finite probability space to construct a semantic space on the same set.

Let (Ω, Σ, P) be a probability space. For $\gamma \in \Omega$ with $p_\gamma := P(\{\gamma\}) \neq 0$, we define the random variable

$$X^*_\gamma(\omega) = \begin{cases} 1, & \text{if } \omega = \gamma, \\ 0, & \text{otherwise.} \end{cases}$$

Then, since these random variables are Bernoulli distributed, we have

$$E(X^*_\gamma) = p_\gamma \quad \text{and} \quad \mathrm{Var}(X^*_\gamma) = p_\gamma(1 - p_\gamma).$$

We standardize this random variable by setting

$$X_\gamma = \frac{X_\gamma^* - E(X_\gamma^*)}{\sqrt{\mathrm{Var}(X_\gamma^*)}} = \frac{X_\gamma^* - p_\gamma}{\sqrt{p_\gamma(1-p_\gamma)}}.$$

For $\alpha, \beta \in \Omega$, we define the function

$$k(\alpha, \beta) := E\big(X_\alpha X_\beta\big).$$

10.1 Calculation of $E(X_\alpha X_\beta)$

First, we write

$$X_\alpha(\omega) = \frac{1_{\{\alpha\}}(\omega) - p_\alpha}{\sqrt{p_\alpha(1-p_\alpha)}}, \quad X_\beta(\omega) = \frac{1_{\{\beta\}}(\omega) - p_\beta}{\sqrt{p_\beta(1-p_\beta)}},$$

thus obtaining:

$$E(X_\alpha X_\beta) = E\left(\frac{\big(1_{\{\alpha\}} - p_\alpha\big)\big(1_{\{\beta\}} - p_\beta\big)}{\sqrt{p_\alpha(1-p_\alpha)p_\beta(1-p_\beta)}}\right).$$

By expanding the numerator, we have:

$$E(X_\alpha X_\beta) = \frac{E\big(1_{\{\alpha\}}1_{\{\beta\}}\big) - p_\alpha E\big(1_{\{\beta\}}\big) - p_\beta E\big(1_{\{\alpha\}}\big) + p_\alpha p_\beta}{\sqrt{p_\alpha(1-p_\alpha)p_\beta(1-p_\beta)}}.$$

Since

$$E\big(1_{\{\alpha\}}1_{\{\beta\}}\big) = P\big(\{\alpha\} \cap \{\beta\}\big) = \begin{cases} p_\alpha, & \text{if } \alpha = \beta, \\ 0, & \text{if } \alpha \neq \beta, \end{cases}$$

and $E\big(1_{\{\alpha\}}\big) = p_\alpha$ as well as $E\big(1_{\{\beta\}}\big) = p_\beta$, the numerator simplifies to

$$\delta_{\alpha,\beta}\, p_\alpha - p_\alpha p_\beta - p_\alpha p_\beta + p_\alpha p_\beta = \delta_{\alpha,\beta}\, p_\alpha - p_\alpha p_\beta,$$

where $\delta_{\alpha,\beta}$ denotes the Kronecker delta (i.e., $\delta_{\alpha,\beta} = 1$ when $\alpha = \beta$ and 0 otherwise).

Thus, we obtain

$$E(X_\alpha X_\beta) = \frac{\delta_{\alpha,\beta}\, p_\alpha - p_\alpha p_\beta}{\sqrt{p_\alpha(1-p_\alpha)p_\beta(1-p_\beta)}}.$$

10.2 Piecewise Representation

This corresponds to the following case distinction:

$$E(X_\alpha X_\beta) = \begin{cases} \dfrac{p_\alpha - p_\alpha^2}{\sqrt{p_\alpha(1-p_\alpha)p_\alpha(1-p_\alpha)}} = \dfrac{p_\alpha(1-p_\alpha)}{p_\alpha(1-p_\alpha)} = 1, & \text{if } \alpha = \beta, \\ \dfrac{-p_\alpha p_\beta}{\sqrt{p_\alpha(1-p_\alpha)p_\beta(1-p_\beta)}}, & \text{if } \alpha \neq \beta. \end{cases}$$

10.3 Summary

For $\alpha, \beta \in \Omega$, we have:

$$\boxed{E(X_\alpha X_\beta) = \frac{\delta_{\alpha,\beta}\, p_\alpha - p_\alpha p_\beta}{\sqrt{p_\alpha(1-p_\alpha)p_\beta(1-p_\beta)}}}.$$

This is a formula in which only p_α and p_β appear on the right-hand side.

10.4 Proof that (Ω, k) is a Semantic Space

We define for $\alpha, \beta \in \Omega$

$$k(\alpha, \beta) = E(X_\alpha X_\beta),$$

where for each $\gamma \in \Omega$ the random variable

$$X_\gamma = \frac{1_{\{\gamma\}} - p_\gamma}{\sqrt{p_\gamma(1-p_\gamma)}}$$

is defined, with $p_\gamma = P(\{\gamma\}) \neq 0$. Note that X_γ has been standardized, i.e., we have

$$E(X_\gamma) = 0 \quad \text{and} \quad \|X_\gamma\| = \sqrt{E(X_\gamma^2)} = 1.$$

We now show the three required properties.

10.4.1 k is Positive Semidefinite

Let $\alpha_1, \ldots, \alpha_n$ be a finite subset of Ω and let $c_1, \ldots, c_n \in \mathbb{R}$ be arbitrary coefficients. Then,

$$\sum_{i,j=1}^{n} c_i c_j\, k(\alpha_i, \alpha_j) = \sum_{i,j=1}^{n} c_i c_j\, \langle X_{\alpha_i}, X_{\alpha_j} \rangle.$$

Since the inner product is linear and symmetric, we can write:

$$\sum_{i,j=1}^{n} c_i c_j\, \langle X_{\alpha_i}, X_{\alpha_j} \rangle = \left\langle \sum_{i=1}^{n} c_i X_{\alpha_i}, \sum_{j=1}^{n} c_j X_{\alpha_j} \right\rangle = \left\| \sum_{i=1}^{n} c_i X_{\alpha_i} \right\|^2 \geq 0.$$

Thus, k is positive semidefinite.

10.4.2 $-1 \leq k(\alpha, \beta) \leq 1$

Since X_α and X_β are normalized ($\|X_\alpha\| = \|X_\beta\| = 1$), it follows from the Cauchy–Schwarz inequality that

$$|k(\alpha, \beta)| = |\langle X_\alpha, X_\beta \rangle| \leq \|X_\alpha\| \|X_\beta\| = 1.$$

Therefore, for all $\alpha, \beta \in \Omega$

$$-1 \leq k(\alpha, \beta) \leq 1.$$

10.4.3 Characterization: $k(\alpha,\beta)=1 \iff \alpha=\beta$

First, note that for each $\alpha \in \Omega$

$$k(\alpha,\alpha) = E(X_\alpha^2) = \|X_\alpha\|^2 = 1.$$

Now, let α and β be two distinct elements of Ω, i.e., $\alpha \neq \beta$. Since X_α and X_β are defined as

$$X_\alpha = \frac{1_{\{\alpha\}} - p_\alpha}{\sqrt{p_\alpha(1-p_\alpha)}}, \quad X_\beta = \frac{1_{\{\beta\}} - p_\beta}{\sqrt{p_\beta(1-p_\beta)}},$$

we observe that the indicator functions $1_{\{\alpha\}}(\omega)$ and $1_{\{\beta\}}(\omega)$ never simultaneously take the value 1 when $\alpha \neq \beta$; that is, $1_{\{\alpha\}}1_{\{\beta\}} = 0$ almost surely. Consequently,

$$E\big(1_{\{\alpha\}}\,1_{\{\beta\}}\big) = P(\{\alpha\}\cap\{\beta\}) = 0.$$

Furthermore, $E\big(1_{\{\alpha\}}\big) = p_\alpha$ and $E\big(1_{\{\beta\}}\big) = p_\beta$. Thus, by expanding the numerator we obtain:

$$E\Big[(1_{\{\alpha\}} - p_\alpha)(1_{\{\beta\}} - p_\beta)\Big] = 0 - p_\alpha p_\beta - p_\alpha p_\beta + p_\alpha p_\beta = -p_\alpha p_\beta.$$

Hence,

$$k(\alpha,\beta) = \frac{-p_\alpha p_\beta}{\sqrt{p_\alpha(1-p_\alpha)p_\beta(1-p_\beta)}}.$$

Since $p_\alpha, p_\beta > 0$ and $p_\alpha, p_\beta < 1$, the fraction is strictly less than 1 (indeed, it is negative). Therefore,

$$k(\alpha,\beta) = 1 \quad \iff \quad \alpha = \beta.$$

This shows that $k(\alpha,\beta)=1$ occurs if and only if $\alpha=\beta$.

10.4.4 Conclusion

We have shown:

- k is positive semidefinite,
- $-1 \leq k(\alpha,\beta) \leq 1$ for all $\alpha,\beta \in \Omega$,
- $k(\alpha,\beta) = 1$ if and only if $\alpha=\beta$.

Hence, (Ω, k) is a semantic space.

10.5 Example: Binomial Distribution with $n = 5$ and $p = \frac{1}{2}$

We consider the probability space given by the binomial distribution with parameters $n = 5$ and $p = \frac{1}{2}$. The probabilities for the individual outcomes $k = 0,1,2,3,4,5$ are

$$p\_list = \left[\frac{1}{32}, \frac{5}{32}, \frac{5}{16}, \frac{5}{16}, \frac{5}{32}, \frac{1}{32}\right].$$

For this space, we construct the semantic space by considering, for each $\gamma \in \{0, 1, 2, 3, 4, 5\}$, the standardized Bernoulli random variable

$$X_\gamma = \frac{1_{\{\gamma\}} - p_\gamma}{\sqrt{p_\gamma(1 - p_\gamma)}},$$

and by defining the function

$$k(\alpha, \beta) = E\left(X_\alpha X_\beta\right) = \frac{\delta_{\alpha,\beta}\, p_\alpha - p_\alpha p_\beta}{\sqrt{p_\alpha(1 - p_\alpha)p_\beta(1 - p_\beta)}},$$

where $\delta_{\alpha,\beta}$ denotes the Kronecker delta.

The corresponding Gram matrix K is then given by:

$$K = \begin{pmatrix} 1 & -\frac{1}{279}\sqrt{465} & -\frac{1}{341}\sqrt{1705} & -\frac{1}{341}\sqrt{1705} & -\frac{1}{279}\sqrt{465} & -\frac{1}{31} \\ -\frac{1}{279}\sqrt{465} & 1 & -\frac{5}{99}\sqrt{33} & -\frac{5}{99}\sqrt{33} & -\frac{5}{27} & -\frac{1}{279}\sqrt{465} \\ -\frac{1}{341}\sqrt{1705} & -\frac{5}{99}\sqrt{33} & 1 & -\frac{5}{11} & -\frac{5}{99}\sqrt{33} & -\frac{1}{341}\sqrt{1705} \\ -\frac{1}{341}\sqrt{1705} & -\frac{5}{99}\sqrt{33} & -\frac{5}{11} & 1 & -\frac{5}{99}\sqrt{33} & -\frac{1}{341}\sqrt{1705} \\ -\frac{1}{279}\sqrt{465} & -\frac{5}{27} & -\frac{5}{99}\sqrt{33} & -\frac{5}{99}\sqrt{33} & 1 & -\frac{1}{279}\sqrt{465} \\ -\frac{1}{31} & -\frac{1}{279}\sqrt{465} & -\frac{1}{341}\sqrt{1705} & -\frac{1}{341}\sqrt{1705} & -\frac{1}{279}\sqrt{465} & 1 \end{pmatrix}.$$

This matrix gives the inner products $k(\alpha, \beta) = E(X_\alpha X_\beta)$ in the semantic space.

11 Balanced extension of a finite semantic space

A finite semantic space (Ω, k) is called *balanced* if the corresponding Gram matrix is a balance matrix, that is, if one of the following equivalent conditions holds:

$$\textbf{(1)} \quad \sum_{a,b\in\Omega} k(a, b) = 0,$$

$$\textbf{(2)} \quad \forall\, a \in \Omega : \quad \sum_{b\in\Omega} k(a, b) = 0.$$

Let (Ω, k) be a finite semantic space, i.e., $k\colon \Omega \times \Omega \to [-1, 1]$ is a positive semidefinite kernel with

$$k(a, a) = 1 \quad \text{and} \quad k(a, b) = 1 \iff a = b, \quad \forall\, a, b \in \Omega.$$

Since the Gram matrix is positive semidefinite, one may apply the Cholesky decomposition to obtain an injective mapping

$$\phi\colon \Omega \to \mathbb{R}^m$$

such that

$$k(a, b) = \langle \phi(a), \phi(b) \rangle, \quad \forall\, a, b \in \Omega.$$

We now distinguish two cases.

Case A: $\sum_{a,b\in\Omega} k(a,b) = 0$

Define

$$w := \sum_{a\in\Omega} \phi(a).$$

Then it follows that

$$0 = \sum_{a,b\in\Omega} k(a,b) = \Big\langle \sum_{a\in\Omega} \phi(a), \sum_{b\in\Omega} \phi(b) \Big\rangle = \langle w, w\rangle = \|w\|^2.$$

Hence, $w = 0$ (the zero vector). In particular, for every $a \in \Omega$ we have

$$\sum_{b\in\Omega} k(a,b) = \langle \phi(a), w\rangle = 0.$$

Thus, the matrix

$$B := (k(a,b))_{a,b\in\Omega}$$

is a *balance matrix.*

We now define, according to Born's rule, the joint probabilities by

$$P(x,y) := \frac{k(x,y)^2}{\sum_{a,b\in\Omega} k(a,b)^2}.$$

Then, P is a joint-probability matrix and it follows that

$$Q := B + P$$

is a quasi-probability matrix.

Case B: $w \neq 0$

Suppose that

$$w := \sum_{a\in\Omega} \phi(a) \neq 0.$$

In order to enforce balance, we extend the original space as follows. First, define

$$\Delta := \{\, a \in \Omega \mid \exists \text{exactly one } b \in \Omega \text{ such that } -\phi(a) = \phi(b) \,\}.$$

Clearly, Δ must be a proper subset of Ω, since otherwise for every $a \in \Omega$ we would have $-\phi(a) = \phi(b)$ for some $b \in \Omega$, which would imply

$$w = \sum_{a\in\Omega} \phi(a) = 0,$$

contradicting the assumption $w \neq 0$.

Now set

$$M := \Omega \setminus \Delta.$$

For each $a \in M$, define a new element a^* and set

$$M^* := \{\, a^* \mid a \in M \,\}.$$

We then define the extended space as the disjoint union

$$\Omega^* := \Omega \,\dot{\cup}\, M^*.$$

On Ω^* we define the mapping $\phi^* : \Omega^* \to \mathbb{R}^m$ by

$$\phi^*(x) := \begin{cases} \phi(x), & x \in \Delta, \\ \phi(x), & x \in M, \\ -\phi(a), & x = a^* \in M^*, \text{ where } a \in M. \end{cases}$$

Since for every $a \in M$ both $\phi(a)$ and $-\phi(a)$ occur in Ω^*, it immediately follows that

$$\sum_{x \in M \cup M^*} \phi^*(x) = \sum_{a \in M} \big[\phi(a) - \phi(a)\big] = 0.$$

Furthermore, one sees that

$$\sum_{x \in \Delta} \phi^*(x) = 0.$$

It then follows that

$$\sum_{x \in \Omega^*} \phi^*(x) = \sum_{x \in \Delta} \phi^*(x) + \sum_{x \in M \cup M^*} \phi^*(x) = 0.$$

We define the extended kernel k^* on Ω^* by

$$k^*(x, y) := \langle \phi^*(x), \phi^*(y) \rangle, \quad \forall\, x, y \in \Omega^*.$$

Then, for all $x, y \in \Omega^* = \Delta \cup M \cup M^*$, the extended kernel $k^*(x, y)$ is given by

$$k^*(x, y) = \begin{cases} k(x, y), & \text{if } x, y \in \Delta \cup M, \\ -k(x, y), & \text{if } x \in \Delta \cup M \text{ and } y \in M^*, \\ k(x, y), & \text{if } x, y \in M^*. \end{cases}$$

Here, $k(x, y)$ is the original kernel on Ω.

This formulation allows one to compute $k^*(x, y)$ directly in terms of $k(x, y)$ without the need for performing a Cholesky decomposition—a practical advantage in applications.

It is clear that (Ω^*, k^*) is a semantic space, and for $x, y \in \Omega \subset \Omega^*$ we already have

$$k^*(x, y)^2 = k(x, y)^2.$$

We can now, analogously to Case A, apply Born's rule by defining, for $x, y \in \Omega^*$, the probabilities

$$P(x, y) := \frac{k^*(x, y)^2}{\sum\limits_{a, b \in \Omega^*} k^*(a, b)^2},$$

and setting

$$B := K^*, \quad Q := B + P.$$

11.1 Summary

We have shown that for every finite semantic space (Ω, k) there exists an balanced semantic space (Ω^*, k^*) with the properties:

1. $\Omega \subseteq \Omega^*$,
2. $k^*(x,y)^2 = k(x,y)^2$ for all $x, y \in \Omega$,
3. $\sum_{x,y\in\Omega^*} k^*(x,y) = 0.$

12 Extension of a finite semantic space via a new element

Let (Ω, P) be a finite probability space with $P(\omega) > 0$ for all $\omega \in \Omega$. For subsets $A, B \subseteq \Omega$, define

$$k(A,B) := \frac{P(A\cap B)}{P(A\cup B)},$$

i.e. the Jaccard kernel, which is known to be positive semidefinite. Hence,

$$k : 2^\Omega \times 2^\Omega \to [0,1].$$

Moreover, we have

$$k(A,B) = 1 \quad \text{if and only if} \quad A = B,$$

since $P(\omega) > 0$ for all $\omega \in \Omega$. Thus, $(2^\Omega, k)$ forms a semantic space (with the subsets of Ω as the objects) that is not balanced.

We now extend this space by adding a new element $\{t\}$ to Ω, i.e.,

$$\Omega^* = \Omega \cup \{t\}.$$

Define the mapping ϕ^* on subsets $A^* \subseteq \Omega^*$ by setting

$$\phi^*(A^*) = \begin{cases} \phi(A^*) & \text{if } t \notin A^*, \\ -\phi(A) & \text{if } A^* = A \cup \{t\}. \end{cases}$$

Then we define

$$k^*(A^*, B^*) := \langle \phi^*(A^*), \phi^*(B^*) \rangle.$$

It follows that $(2^{\Omega^*}, k^*)$ is a balanced semantic space.

Another way to perform this construction is by using the balanced matrix

$$B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

and the Kronecker product $\otimes$. Define

$$K^* := B \otimes K,$$

i.e. the block matrix

$$K^* = \begin{pmatrix} K & -K \\ -K & K \end{pmatrix},$$

where K is the Gram matrix of $k(A, B)$. Here the elements of 2^{Ω^*} are arranged so that first all subsets $A \subseteq \Omega$ appear, and then, in the same order, all sets of the form $A \cup \{t\}$.

This extension from $(2^{\Omega}, k)$ to $(2^{\Omega^*}, k^*)$ satisfies:

1. For all $A, B \in 2^{\Omega}$, we have $k^*(A, B)^2 = k(A, B)^2$,
2. $\sum_{A,B \in 2^{\Omega^*}} k^*(A, B) = 0$, and
3. The original σ-algebra $\Sigma := 2^{\Omega}$ is a subset of the extended σ-algebra $\Sigma^* := 2^{\Omega^*}$.

12.1 Connection to "Negative Probabilities"

We can recover the probabilities from the modified kernel k^* defined above on the extended space as follows. For any subset $A^* \subseteq \Omega^*$, define

$$B(A^*) := k^*(A^*, \Omega).$$

In particular, for subsets that originally belonged to Ω, i.e. if $A^* = A \subseteq \Omega$, we have

$$B(A^*) = k^*(A, \Omega) = k(A, \Omega) = P(A),$$

and for subsets of the form $A^* = A \cup \{t\}$ we have

$$B(A^*) = -k(A, \Omega) = -P(A).$$

For any subcollection $S \subseteq \Sigma^* = 2^{\Omega^*}$, define

$$B(S) := \sum_{A^* \in S} B(A^*).$$

Then it holds that

$$B(\Sigma^*) = \sum_{A^* \in \Sigma^*} B(A^*) = 0.$$

Thus, $(\Sigma^*, 2^{\Sigma^*}, B)$ forms a *balance space* with the following properties:

1. $\Sigma \subseteq \Sigma^*$,
2. For every $A \in \Sigma$, $B(A) = P(A)$.

13 The Dedekind-Frobenius matrix

Let $(\Omega, 2^{\Omega}, f)$ be a finite balanced space such that $\Omega = G$ is a finite group.

Let $G = \{g_1, \ldots, g_n\}$ be a finite group, and let

$$f\colon G \to \mathbb{R}$$

be a real-valued zero-sum function on G, i.e.

$$\sum_{g \in G} f(g) = 0.$$

Then we will show, how to construct a balance matrix from this space:

(Remark: In the literature, the "balance" property seems to be called "zero-sum function". Examples of such functions are non-trivial real-valued characters χ of finite abelian groups G.)

We define the *Dedekind–Frobenius f-valued matrix*

$$M_f \;=\; \big(m_{i,j}\big)_{1\le i,j\le n},$$

by the rule

$$m_{i,j} \;=\; f\big(g_i\, g_j^{-1}\big).$$

We wish to prove that M_f is *balanced*, meaning that the sum of its entries in each row and in each column is zero.

13.1 Row Sums

Fix an index j. Then the sum of the entries in the j-th column is

$$\sum_{i=1}^{n} m_{i,j} \;=\; \sum_{i=1}^{n} f\big(g_i\, g_j^{-1}\big).$$

Because the map $g \mapsto g\, g_j^{-1}$ is a bijection (permutation) on G, the set $\{\, g_i\, g_j^{-1} : i = 1, \ldots, n \,\}$ is simply a re-labeling of all elements of G. Hence

$$\sum_{i=1}^{n} f\big(g_i\, g_j^{-1}\big) \;=\; \sum_{h\in G} f(h) \;=\; 0,$$

since f is a zero-sum function on G. This shows that each *column* of M_f sums to zero.

13.2 Column Sums

A similar argument applies when we fix an index i and sum down the i-th row. Specifically,

$$\sum_{j=1}^{n} m_{i,j} \;=\; \sum_{j=1}^{n} f\big(g_i\, g_j^{-1}\big).$$

Again, as g_j runs through all elements of G, g_j^{-1} also runs through all elements of G (just in a different order), so $\{\, g_i\, g_j^{-1} : j = 1, \ldots, n \,\} = G$. Thus

$$\sum_{j=1}^{n} f\big(g_i\, g_j^{-1}\big) \;=\; \sum_{h\in G} f(h) \;=\; 0.$$

Therefore, each *row* of M_f also sums to zero.

13.3 Remark

Since both the row sums and the column sums of M_f vanish, M_f is a **balanced matrix**. In symbols:

$$\sum_{i=1}^{n} m_{i,j} \;=\; 0 \quad \text{and} \quad \sum_{j=1}^{n} m_{i,j} \;=\; 0, \quad \text{for all } i, j.$$

Hence M_f belongs to the class of matrices whose row and column sums are all zero.

14 Example of "negative probabilities": Coin Transitions in a Wallet

In a wallet U_1, there are four coin types:

- 1 cent and 2 cent coins,
- 1 euro and 2 euro coins.

Their distribution is as follows:

$U_1 =$

	1	2	Total
Cent	1	5	6
Euro	5	25	30
Total	6	30	36

We simultaneously remove (–) and add (+) the following number of coins from U_1:

$dU =$

	1	2	Total
Cent	-1	2	1
Euro	2	3	5
Total	1	5	6

Thus, the wallet U_2 contains the following number of coins:

$U_2 =$

	1	2	Total
Cent	0	7	7
Euro	7	28	35
Total	7	35	42

The probabilities of drawing a specific coin type (with replacement) are given by:

In U_1:

$P_1 =$

	1	2	Total
Cent	$\frac{1}{36}$	$\frac{5}{36}$	$\frac{1}{6}$
Euro	$\frac{5}{36}$	$\frac{25}{36}$	$\frac{5}{6}$
Total	$\frac{1}{6}$	$\frac{5}{6}$	1

During the removal and addition process:

$dP =$

	1	2	Total
Cent	$-\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$
Euro	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{5}{6}$
Total	$\frac{1}{6}$	$\frac{5}{6}$	1

In U_2:

$P_2 =$

	1	2	Total
Cent	0	$\frac{7}{42}$	$\frac{1}{6}$
Euro	$\frac{7}{42}$	$\frac{28}{42}$	$\frac{5}{6}$
Total	$\frac{1}{6}$	$\frac{5}{6}$	1

It holds that

$$P_1 + dP = P_2 \quad \text{or equivalently} \quad dP = P_2 - P_1,$$

and

$$U_1 + dU = U_2 \quad \text{or equivalently} \quad dU = U_2 - U_1.$$

Here, P_1 and P_2 are standard probability matrices, whereas dP is a quasi-probability matrix. So we see here in this example how "negative probabilities" can occur naturally although the marginals probabilities are ≥ 0.

14.1 Discussion and Interpretation of the Example

Below is an informal, step-by-step interpretation of the coin-wallet example and why it illustrates "negative probabilities" (or quasi-probabilities) in a simple setting.

Overview of the Example We have a wallet U_1 containing four types of coins:

- **Cent coins**: 1 cent and 2 cent.
- **Euro coins**: 1 euro and 2 euro.

Their initial distribution (i.e., how many of each coin type the wallet holds) is given by a 2×2 table, broken down by "Cent" vs. "Euro" along one axis and "1" vs. "2" along the other:

$U_1 =$

	1	2	Total
Cent	1	5	6
Euro	5	25	30
Total	6	30	36

- Row-wise, we see 6 cent coins total ($1 + 5 = 6$) and 30 euro coins total ($5 + 25 = 30$).
- Column-wise, we see 6 coins of denomination "1" ($1\,\text{cent} + 5\,\text{euro} = 6$) and 30 coins of denomination "2" ($5\,\text{cent} + 25\,\text{euro} = 30$).
- In total, there are 36 coins in U_1.

Next, we simultaneously *remove* some coins from U_1 (these will appear as negative entries) and *add* other coins to U_1 (these appear as positive entries). This is shown in a change matrix dU:

$dU =$

	1	2	Total
Cent	-1	2	1
Euro	2	3	5
Total	1	5	6

- For the cent coins in the "1" column, -1 means we remove one 1-cent coin.
- For the cent coins in the "2" column, $+2$ means we add two 2-cent coins.

- Similarly, we add two 1-euro coins and three 2-euro coins.

After this simultaneous removal and addition, the new wallet is denoted U_2. It is simply given by $U_2 = U_1 + dU$:

$U_2 =$

	1	2	Total
Cent	0	7	7
Euro	7	28	35
Total	7	35	42

- For instance, $1 + (-1) = 0$ cent coins of type "1" remain.
- $25 + 3 = 28$ euro coins of type "2," etc.

Probabilities of Drawing Each Coin Type In U_1:
Since U_1 has 36 coins total, the probability matrix P_1 indicates the chance of drawing each type if we pick a coin at random:

$P_1 =$

	1	2	Total
Cent	$\frac{1}{36}$	$\frac{5}{36}$	$\frac{1}{6}$
Euro	$\frac{5}{36}$	$\frac{25}{36}$	$\frac{5}{6}$
Total	$\frac{1}{6}$	$\frac{5}{6}$	1

- The "Cent" row sums to $\frac{1}{6}$ (i.e., $\frac{6}{36}$).
- The "Euro" row sums to $\frac{5}{6}$ (i.e., $\frac{30}{36}$).
- The column sums $\frac{1}{6}$ and $\frac{5}{6}$ reflect the distribution of coins with denominations "1" and "2."

During the Removal and Addition (dP):
While we move from U_1 to U_2, we can consider a "difference" in probabilities:

$dP =$

	1	2	Total
Cent	$-\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$
Euro	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{5}{6}$
Total	$\frac{1}{6}$	$\frac{5}{6}$	1

Notice that the upper-left entry is $-\frac{1}{6}$. This negative value is not interpretable as a probability in the classical sense; instead, it represents a *quasi-probability*—an intermediate value that captures the process of removing probability mass from that category.

In U_2:
The final probability matrix P_2 (for the 42 coins in U_2) is:

$P_2 =$

	1	2	Total
Cent	0	$\frac{7}{42}$	$\frac{1}{6}$
Euro	$\frac{7}{42}$	$\frac{28}{42}$	$\frac{5}{6}$
Total	$\frac{1}{6}$	$\frac{5}{6}$	1

The fractions reflect the new composition of coin types in U_2 while maintaining consistent marginal totals.

Key Point: Negative (Quasi-)Probabilities

1. **Relation $P_1 + dP = P_2$:**
   The matrix dP (the "difference" in probabilities) can have negative entries even though P_1 and P_2 are valid probability matrices. This shows that the intermediate process of removing and adding coins leads to a temporary negative entry in dP.

2. **Interpretation:**

   - P_1 represents the probabilities of drawing each coin type before any changes.
   - P_2 represents the probabilities after the coin changes.
   - dP captures the *change* from P_1 to P_2. Because the process involves literally "taking away" coins (removing probability mass) and "adding" coins, some entries in dP become negative.

3. **Relevance:**
   This example demonstrates how negative (or quasi-)probabilities can arise in practical scenarios when modeling transitions. Even though the initial and final states (P_1 and P_2) are standard probability distributions, the intermediate difference dP may include negative values. Such quasi-probabilities can be useful for intermediate calculations in areas like quantum mechanics, game design, or financial models.

Bottom Line

- **Before the change:** We have a valid probability matrix P_1.
- **After the change:** We have another valid probability matrix P_2.
- **Difference:** The matrix $dP = P_2 - P_1$ can be viewed as a quasi-probability matrix. Although it may contain negative entries, its row and column sums are balanced so that the transition from P_1 to P_2 is correctly captured.

This coin-wallet example thus provides a tangible illustration of how negative or quasi-probabilities can naturally appear when modeling a transition (such as removing and adding coins) via a single matrix operation, even though the initial and final distributions are valid (nonnegative) probability distributions.

14.2 Extended Polya urn model

In the extended Pólya urn model the urn (analogous to the wallet) contains objects that can be classified by two attributes—say, color (with m distinct colors) and size (with n distinct sizes). The initial composition of the urn is given by the matrix

$$U_1 = \left(u_{ij}\right)_{1\le i\le m,\ 1\le j\le n},$$

with total number of objects

$$T_1 = \sum_{i=1}^{m}\sum_{j=1}^{n} u_{ij}.$$

Thus, the probability of drawing an object of color i and size j is

$$P_1(i,j) = \frac{u_{ij}}{T_1}.$$

Simultaneous Addition and Removal:
We now allow for the possibility of *simultaneously* removing some objects (represented by negative entries) and adding others (positive entries). Let the change be represented by the delta matrix

$$dU = \big(d_{ij}\big)_{1\le i\le m,\ 1\le j\le n}.$$

The new composition is then

$$U_2 = U_1 + dU,$$

with total number of objects

$$T_2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(u_{ij} + d_{ij}).$$

The updated probability of drawing an object from category (i,j) is

$$P_2(i,j) = \frac{u_{ij} + d_{ij}}{T_2}.$$

Preservation of Marginal Probabilities:
To ensure that the marginal probabilities (for example, the overall probability for each color or for each size) remain the same after the transition, the change matrix dU must be chosen so that the net effect on the row and column sums of U_1 is proportional. Let

$$r_i = \sum_{j=1}^{n} u_{ij} \quad \text{and} \quad c_j = \sum_{i=1}^{m} u_{ij}$$

be the row and column totals of U_1. After the change, if the new row and column totals are

$$r_i' = \sum_{j=1}^{n}(u_{ij} + d_{ij}) \quad \text{and} \quad c_j' = \sum_{i=1}^{m}(u_{ij} + d_{ij}),$$

then to preserve the marginal probabilities it is sufficient to have

$$\frac{r_i'}{T_2} = \frac{r_i}{T_1} \quad \text{for each } i,$$

and similarly for the column sums.

14.2.1 Conditions for the Appearance of Negative Probabilities

If we only **add** objects (Addition), then $d_{ij} > 0$ and consequently $dP_{ij} > 0$, meaning that the probability of drawing objects from those categories increases.

If we only **remove** objects (Removal), then $d_{ij} < 0$, but the total number of objects also decreases, meaning $T < 0$. As a result, the probability change is given by

$$dP_{ij} = \frac{d_{ij}}{T},$$

which remains **positive** ($dP_{ij} > 0$) because both d_{ij} and T are negative, leading to an overall increase in relative probability.

The **effect of negative probabilities** only arises when we **simultaneously remove and add** objects. In this case:

- Some entries of d_{ij} will be **negative** (representing removed objects),
- But hopefully, the total sum T remains **positive**, ensuring that the probability difference matrix dP contains some **negative entries** ($dP_{ij} < 0$ in some places).

This observation aligns with the phenomenon seen in **quantum mechanics (QM)**, where a state that involves both "removal" and "addition" is a **mixture** of the "pure states" of **only adding** and **only removing**. In QM, negative probability-like effects emerge in interference phenomena, where the transition between states involves both positive and negative contributions to probability amplitudes. Similarly, in our model, negative quasi-probabilities appear when an **intermediate state is formed by a combination of adding and removing** objects, rather than from pure addition or pure removal alone.

14.3 Possible applications of the extended Pólya urn model

The extended Pólya urn model generalizes the classical urn model by allowing simultaneous addition and removal of objects while preserving the marginal probabilities of the original urn distribution. This framework could have several practical applications in areas where the overall relative proportions of categories must be preserved despite fluctuations in absolute counts. For example:

- **Inventory Management:** In retail, products are sold (removed) and restocked (added) simultaneously. Using an extended Pólya urn model ensures that the product mix (marginal probabilities) remains constant even as absolute quantities vary.
- **Population Dynamics:** In biological systems, individuals in different subpopulations (e.g., age groups or species) may be born and die concurrently, but the overall structure (relative proportions) is maintained.
- **Evolutionary Game Theory:** Agents might switch strategies (moving from one category to another) while the overall distribution of strategies remains in equilibrium.
- **Marketing and Consumer Behavior:** Consumers may shift preferences among products without altering the overall market share distribution.

Discussion of Applications: These examples illustrate how the extended Pólya urn model can be applied to systems in which objects (or agents) are simultaneously added and removed while maintaining fixed marginal proportions. Such scenarios are prevalent in:

- **Inventory Management:** Stock levels are adjusted by simultaneously selling (removing) and restocking (adding) items while preserving the product mix.
- **Population Dynamics:** In ecosystems or cell populations, births and deaths occur concurrently, yet the relative proportions of subpopulations remain stable.
- **Evolutionary Game Theory:** Agents may switch strategies (i.e., move from one category to another) in a manner that keeps the overall distribution of strategies unchanged.
- **Consumer Behavior:** Shifts in consumer preferences can be modeled by simultaneous transitions between product categories, while overall market shares are maintained.

In each case, the extended Pólya urn model provides a framework for understanding how internal changes (captured by the quasi-joint-probability matrix dP) can occur without affecting the observable marginal distributions. This insight could be helpful for both theoretical analyses and practical applications where the maintenance of certain proportions is important.

In summary, the extended Pólya urn model offers a powerful tool for modeling systems with simultaneous additions and removals, ensuring that key marginal probabilities are preserved. This characteristic makes it highly relevant in diverse fields ranging from inventory control and biological population studies to economic and social systems.

15 Applications of the extension with a Balanced Semantic Space

The extension of a finite semantic space to a balanced semantic space (i.e. extending Ω to Ω^* so that the extended kernel k^* satisfies

$$\sum_{x,y\in\Omega^*} k^*(x,y) = 0,$$

while preserving the squared values on the original set) offers several intriguing applications across different fields. We briefly outline some of these potential applications below:

1. Natural Language Processing and Distributional Semantics

In many models of word meaning, words are represented as vectors in a high-dimensional space and semantic similarity is measured via inner products. However, such spaces may possess a nonzero mean that can bias similarity measures.

The extension provides a principled way to "center" the semantic space by extending the vocabulary so that the overall representation is balanced. This can lead to more accurate similarity computations, improved clustering, and enhanced performance in tasks such as analogical reasoning and semantic role labeling.

2. Kernel Methods in Machine Learning

Kernel-based techniques (e.g., kernel principal component analysis, spectral clustering) often benefit from centering the kernel matrix. A balanced semantic space naturally induces a centered kernel, ensuring that the sum of all pairwise similarities is zero. The extension method, preserves key pairwise relationships while eliminating systemic bias. This can improve the performance and interpretability of dimensionality reduction and clustering algorithms.

4. Graph and Network Analysis

In many network and graph-based applications, nodes (e.g., individuals in social networks or entities in relational databases) are embedded into a semantic space for tasks such as community detection or link prediction. If the underlying similarity (or kernel) matrix is biased, it can obscure the true structure of the network. An extension to a balanced semantic space removes this bias, yielding a representation where the overall interaction is neutral. This can lead to improved detection of communities or clusters and a clearer interpretation of network dynamics.

5. Financial Modeling and Risk Management

Financial models often rely on probability measures to assess risk and model asset returns. In practice, observed data may induce quasi-probability distributions that are biased. By applying the extension to create a balanced semantic space, one can obtain a corrected representation that is centered (i.e., has zero net bias). Such balanced representations are beneficial in risk-neutral pricing, portfolio optimization, and in the identification of systematic deviations that could lead to market anomalies.

6. Signal Processing and Time Series Analysis

In signal processing, it is common to remove the DC (zero-frequency) component of a signal so that the residual signal is centered around zero. Similarly, when constructing feature spaces or embedding signals into a semantic space, a balanced representation (one with a zero mean) can improve filtering, compression, and noise reduction techniques. The extension method can be used to adjust the feature space so that it becomes balanced, thus ensuring that further analysis is not influenced by an overall bias.

7. Data Visualization and Dimensionality Reduction

When visualizing high-dimensional data, a balanced embedding can improve interpretability. For example, in techniques such as multidimensional scaling

(MDS) or t-SNE, a balanced semantic space ensures that the origin corresponds to a natural center of the data, allowing for better separation of clusters and more meaningful visualization of relationships.

In summary, the extension with a balanced semantic space is not only a mathematically elegant solution to the problem of nonzero biases in quasi-probability and kernel representations but also has the potential to impact diverse fields—from natural language processing and machine learning to quantum physics and financial modeling. Each application benefits from the elimination of systemic biases, leading to improved performance and interpretability in both theoretical analyses and practical implementations.

16 Speculative Applications

Although our investigation has focused on the mathematical structure underlying negative probabilities and balance matrices, the methods developed herein have intriguing potential applications beyond pure mathematics.

Quantum Physics

In quantum mechanics, quasi-probability distributions—such as the Wigner function—are used to describe quantum states. These distributions often take on negative values, reflecting the non-classical behavior of quantum systems. The balance matrix framework could offer a new perspective on these quasi-probabilities by decomposing them into a conventional probability part and a bias term. Such an approach might clarify the role of interference effects and entanglement in quantum measurements and contribute to a better understanding of quantum-to-classical transitions.

Financial Modeling and Risk Management

Financial markets are rife with uncertainties and asymmetric risks, and classical probability models sometimes fail to capture extreme events (the so-called "black swan" phenomena). By applying balance matrices, one might model market probabilities in a way that incorporates hidden biases or risk factors. For instance, a quasi-probability model of asset returns could be decomposed to isolate the systematic deviations that lead to market crashes or bubbles, thus providing a novel tool for risk assessment and management.

Cognitive Science and Decision Theory

Human decision-making often deviates from the predictions of classical probability theory. In psychology and behavioral economics, observed choices sometimes reflect negative probabilities or biases that are not easily captured by standard models. By decomposing these quasi-probabilities, researchers could identify the underlying biases in perception or judgment. This may lead to improved models of human cognition, allowing for better predictions of behavior in situations involving uncertainty.

Artificial Intelligence and Machine Learning

In domains where systems must make decisions under uncertain or conflicting information, such as in autonomous agents or recommendation systems, incorporating a balance matrix approach could enhance robustness. Decomposing uncertain data into a traditional probability component and an adjustment term may allow AI systems to better manage ambiguous inputs, leading to improved decision-making and performance in complex environments.

Other Interdisciplinary Areas

Beyond the fields mentioned above, the principles of balance matrices and quasi-probabilities may find applications in any domain where uncertainty and hidden biases play a role. This includes areas such as epidemiology (modeling the spread of diseases with imperfect data), social sciences (analyzing opinion dynamics), and even art and design (where probabilistic models can inform generative processes).

17 Conclusion

In conclusion, we have presented a comprehensive framework for interpreting negative probabilities using balance matrices. By decomposing a quasi-probability matrix Q into the sum $Q = P + B$, where P is a proper joint probability matrix and B is a balance matrix with zero row and column sums, our approach preserves key marginal properties while isolating the non-classical components. The theoretical development—including results on ring isomorphisms, Moore-Penrose inverses, and iterative proportional fitting—provides a robust mathematical foundation for this decomposition.

Moreover, the versatility of this framework is evident in its potential applications beyond mathematics. Whether it is in quantum physics, financial risk management, cognitive science, or artificial intelligence, the balance matrix method opens new avenues for interpreting and managing uncertainties that classical probability theory alone cannot adequately address.

Future work may further explore these interdisciplinary applications, refine the computational algorithms, and extend the theory to infinite-dimensional spaces. In doing so, the balance matrix approach promises to bridge the gap between abstract mathematical theory and practical problems in a wide array of fields.

References

[1] S. N. Cohen, R. J. Elliott, and C. E. M. Pearce, *A Ring Isomorphism and Corresponding Pseudoinverses*, 2008.

[2] Székely, G. J., *Half of a coin: negative probabilities*, Wilmott Magazine, July, 66-68, 2005.