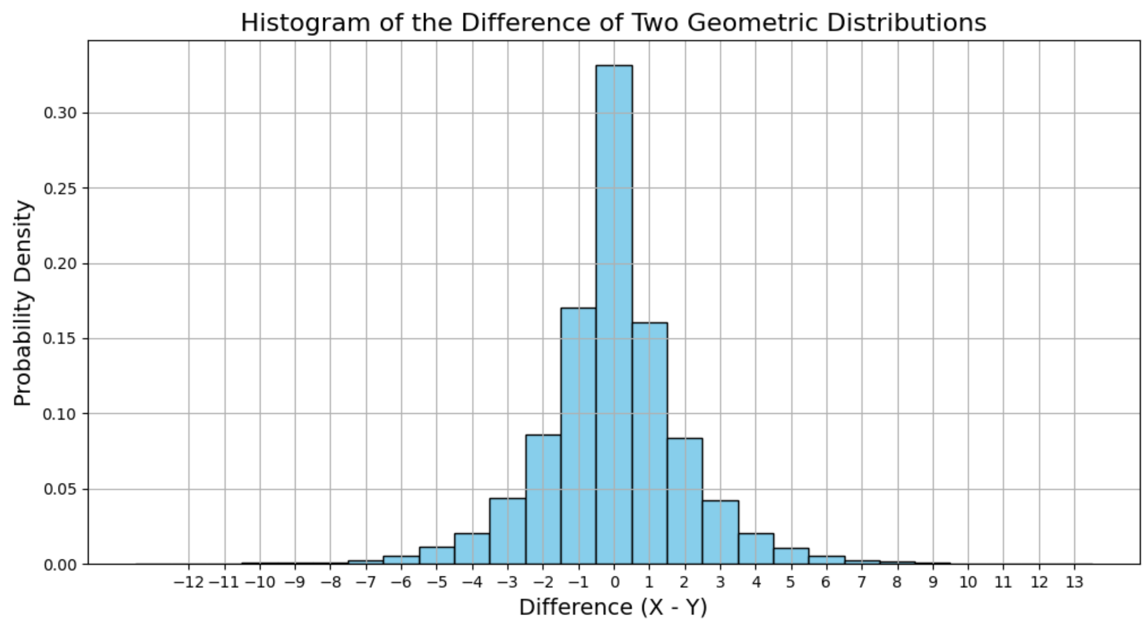
202410270542 Correlated Geometric vs Binomial Distribution

Difference Between Two Geometric Distributions



Introduction to the Geometric Distribution

The geometric distribution models the number of Bernoulli trials needed to achieve the first success. It is a discrete probability distribution with the following properties:

Support:

$$k = 1, 2, 3, \ldots$$

Probability Mass Function (PMF):

$$P(X = k) = p(1 - p)^{k-1}$$

where:

- p is the probability of success on each trial.
- $(1 - p)$ is the probability of failure on each trial.

Mean:

$$\mu = E[X] = \frac{1}{p}$$

Variance:

$$\sigma^2 = \mathrm{Var}(X) = \frac{1-p}{p^2}$$

Properties:

- The geometric distribution is memoryless, meaning the probability of success on the next trial is independent of the number of failures that have already occurred.
- It is a discrete distribution over positive integers.
- The distribution is skewed to the right, especially for smaller values of p.

Difference Between Two Geometric Distributions

Consider two independent geometric random variables X and Y, each representing the number of trials until the first success, with the same probability of success p. We are interested in the distribution of their difference:

Difference:

$$D = X - Y$$

Properties of D

- **Support:**

$$D \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

- **Mean:**
  Since X and Y are independent and identically distributed (i.i.d.), their expected values are the same:

$$E[D] = E[X - Y] = E[X] - E[Y] = \frac{1}{p} - \frac{1}{p} = 0$$

- **Variance:**
  Because X and Y are independent:

$$\mathrm{Var}(D) = \mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 2 \times \frac{1-p}{p^2} = \frac{2(1-p)}{p^2}$$

- **Distribution of D:**
  The distribution of D is symmetric around zero when X and Y are identically distributed.

Probability Mass Function of D

Computing the exact PMF of D involves convolution of the PMFs of X and Y. Since X and Y are independent, the PMF of D can be expressed as:

$$P(D = k) = \sum_{n=1}^{\infty} P(X = n + k)P(Y = n), \quad \text{for all integers } k$$

This calculation can be complex due to the infinite summation. However, for practical purposes, we can approximate the distribution of D via simulation.

Correlated Geometric Differences Using Copula

Generating Correlated Differences X_1 and X_2

To generate correlated differences $X_1 = Z_{11} - Z_{12}$ and $X_2 = Z_{21} - Z_{22}$ between pairs of geometric distributions, we apply a copula method. This allows us to introduce a specified correlation directly between X_1 and X_2 without altering the underlying distributions. Here's the step-by-step process:

Steps

1. **Define the Independent Geometric Variables:**

   For each difference, we define two independent geometric distributions. The parameters are as follows:
   - p_1: Probability of success for the first pair (Z_{11}, Z_{12})
   - p_2: Probability of success for the second pair (Z_{21}, Z_{22})
2. **Compute Differences:**

   Define:

   $$X_1 = Z_{11} - Z_{12}$$

   $$X_2 = Z_{21} - Z_{22}$$

3. **Copula for Correlation:**

   To impose correlation between X_1 and X_2, use a copula approach:

- **Step 3a**: Generate standard normal correlated variables N_1 and N_2 with the desired correlation $ \rho$.
- **Step 3b**: Transform these normals to uniform variables using the cumulative distribution function (CDF) of the normal distribution:

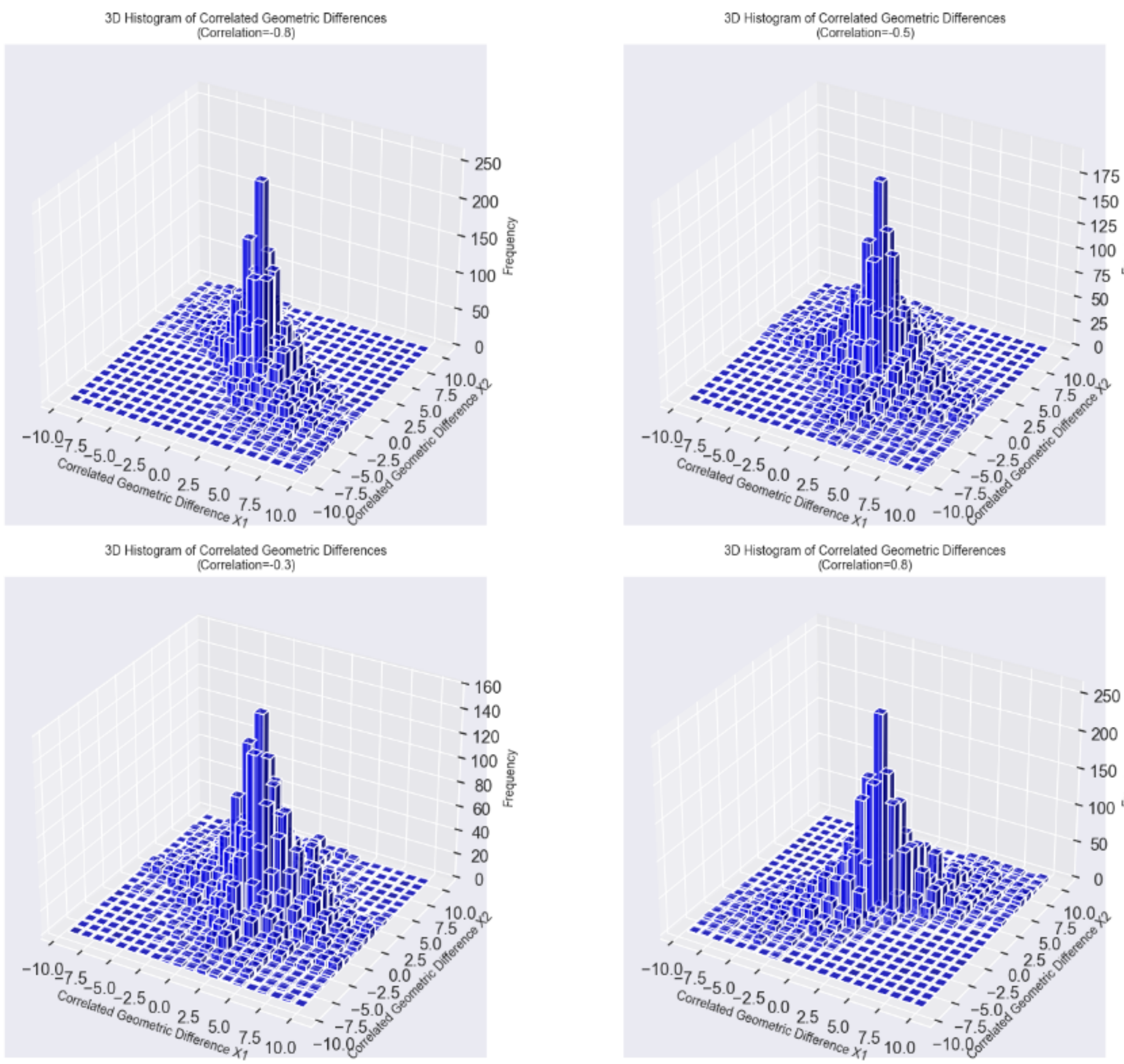
$$U_1 = \Phi(N_1), \quad U_2 = \Phi(N_2)$$

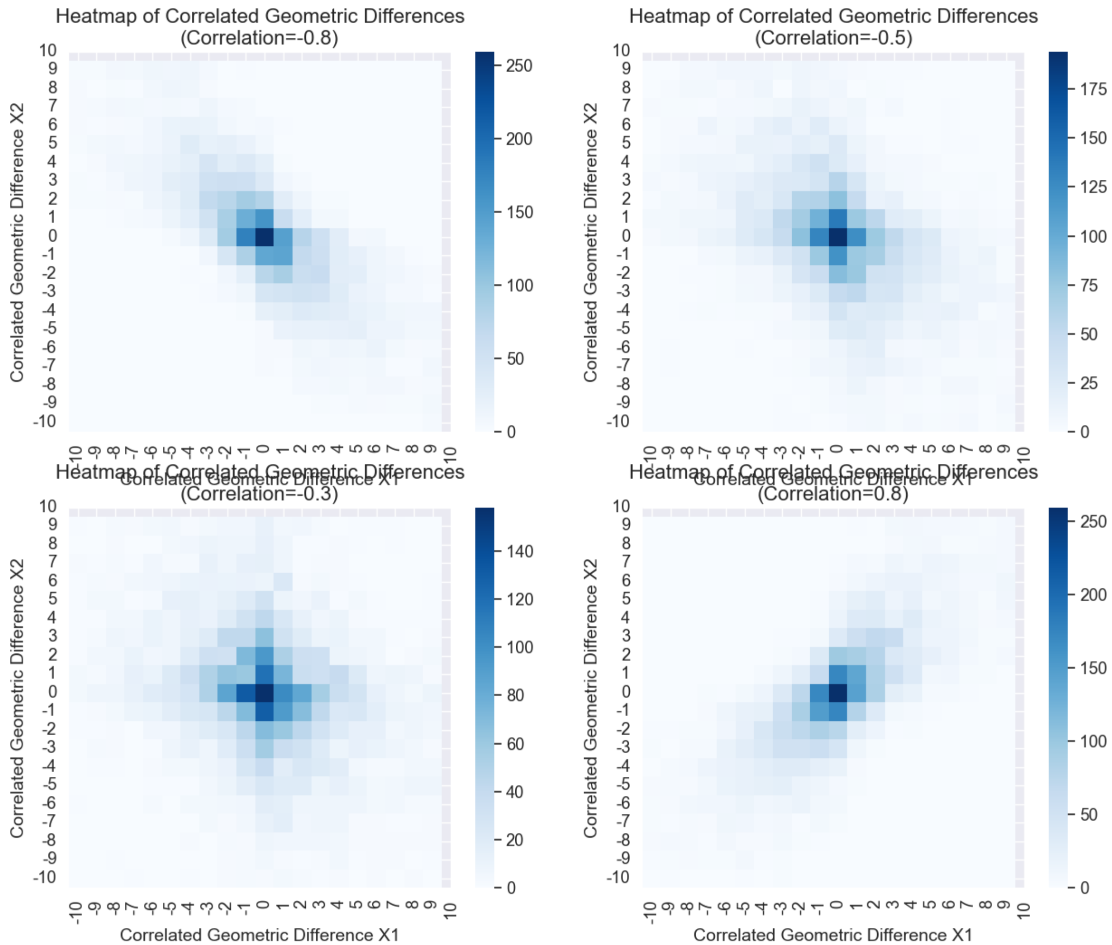
4. **Rank Transformation:**

   Sort X_1 and X_2 based on the ranks of U_1 and U_2 to introduce the desired correlation structure directly on X_1 and X_2 without any additional rescaling.

Visualization

The final correlated variables X_1 and X_2 can be visualized using both a 2D heatmap and a 3D histogram to represent the joint distribution and observe the effect of the copula on the correlation.





Heatmap of Correlated Geometric Differences (Correlation=-0.8)
Heatmap of Correlated Geometric Differences (Correlation=-0.5)
Heatmap of Correlated Geometric Differences (Correlation=-0.3)
Heatmap of Correlated Geometric Differences (Correlation=0.8)
Correlated Geometric Difference X1
Correlated Geometric Difference X2