Exercise 2. Let X; Bin(ni, Pi), I = 1,...,n, Where X1,..., Xn Are Assumed To Be Independent. Derive The

Exercise 2. Let X; Bin(ni, Pi), I = 1,...,n, Where X1,..., Xn Are Assumed To Be Independent. Derive The

---

Introduction

In the realm of probability theory and statistical inference, understanding the properties of various distributions is fundamental. One such distribution, the binomial distribution, plays a crucial role in modeling the number of successes in a sequence of independent Bernoulli trials. The exercise at hand involves a collection of binomial random variables, each characterized by parameters n_i and p_i, and assumed to be independent. The goal is to derive certain properties or distributions related to these variables, such as their joint distribution, marginal distributions, or expectations.

This article aims to provide a comprehensive, step-by-step derivation and explanation of the problem presented, making it accessible for students and practitioners alike. We will delve into the definitions, assumptions, and mathematical techniques necessary to arrive at the desired result, emphasizing clarity and depth.

Understanding the Setup

The Random Variables

We are given a collection of random variables:

$$
X_i \sim \text{Binomial}(n_i, p_i), \quad i = 1, \dots, n
$$

where:

- n_i is the number of Bernoulli trials in the i-th binomial experiment.
- p_i is the probability of success in each trial for the i-th experiment.
- X_i counts the number of successes in those trials.

Assumption of Independence

The key assumption made is that the X_i are independent:

$$
X_1, X_2, \dots, X_n \text{ are independent}
$$

This independence implies that the joint distribution of the vector \(

$\mathbf{X} = (X_1, X_2, \dots, X_n)$ factors into the product of individual binomial distributions:

$$
f_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)
$$

where each $f_{X_i}(x_i)$ is the binomial probability mass function (pmf):

$$
f_{X_i}(x_i) = \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i}
$$

for $x_i = 0, 1, \dots, n_i$.

Objective of the Derivation

The exercise prompts to derive the—likely the joint distribution, marginal distribution, or some other property. Based on typical exercises involving multiple independent binomial variables, common objectives include:

- Deriving the joint distribution of the vector $\mathbf{X}$.
- Finding the distribution of a sum of the X_i.
- Computing the expected value or variance of a combined statistic.
- Deriving the distribution of a linear combination of the X_i.

For the purpose of this article, we will focus on deriving the joint distribution of the vector $\mathbf{X}$. This is foundational, as it enables understanding of how these variables interact under independence, and it forms the basis for further statistical inference.

Deriving the Joint Distribution

Step 1: Recall the Binomial PMF

For each X_i, the pmf is:

$$
f_{X_i}(x_i) = \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i}
$$

with $x_i \in \{0, 1, \dots, n_i\}$.

Step 2: Independence and Joint Distribution

Since the X_i are independent, the joint pmf is the product:

$$
f_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i}
$$

This is straightforward but important, as it confirms that the joint distribution is a product of the individual binomial pmfs.

Step 3: Expressing the Joint Distribution

The joint distribution can thus be written explicitly as:

$$
\boxed{
f_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n \left[\binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i} \right]
}
$$

for all $(x_1, \dots, x_n)$ such that $x_i \in \{0, 1, \dots, n_i\}$.

Deriving Distributions of Sums and Other Statistics

Sum of the Binomial Variables

Often, an important quantity is the sum:

$$
S = \sum_{i=1}^n X_i
$$

which counts the total number of successes across all experiments.

Distribution of S

Since the X_i are independent, the distribution of S is the convolution of the individual binomial distributions. However, because the parameters p_i are generally different, S does not follow a simple binomial distribution unless all p_i are equal.

Case 1: All $p_i = p$ and $n_i = n$

In this special case, the sum $S \sim \text{Binomial} \left(\sum_{i=1}^n n_i, p \right)$.

Case 2: Different p_i and n_i

In the general case, the distribution of S is a Poisson-binomial distribution, which is the distribution of the sum of independent Bernoulli trials with different success probabilities.

Poisson-Binomial Distribution

Definition

Given independent Bernoulli variables:

$$
Y_i \sim \text{Bernoulli}(p_i), \quad i=1,\dots,n
$$

then:

$$
S = \sum_{i=1}^n Y_i
$$

has the Poisson-binomial distribution.

PMF of S

The pmf can be expressed as:

$$
P(S = k) = \sum_{A \subseteq \{1,\dots,n\} : |A|=k} \prod_{i \in A} p_i \prod_{j \notin A} (1 - p_j)
$$

which involves summing over all subsets of size k. This formula, while exact, becomes computationally intensive for large n.

Characteristic Function Approach

An alternative is to use the moment-generating function (MGF):

$$
M_S(t) = \prod_{i=1}^n \left(1 - p_i + p_i e^{t} \right)
$$

which can be inverted to find the pmf via Fourier transforms or recursive algorithms.

Expectations and Variances

Expected Value

Since $X_i \sim \text{Binomial}(n_i, p_i)$:

$$
E[X_i] = n_i p_i
$$

and the linearity of expectation gives:

$$
E[S] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n n_i p_i
$$

Variance

Because the X_i are independent:

$$
\operatorname{Var}(X_i) = n_i p_i (1 - p_i)
$$

and

$$
\operatorname{Var}(S) = \sum_{i=1}^n \operatorname{Var}(X_i) = \sum_{i=1}^n
$$

$$
n_i p_i (1 - p_i)
$$

Practical Applications and Implications

Quality Control

In manufacturing, multiple quality checks (each modeled by a binomial variable) can be aggregated to assess the overall defect rate.

Clinical Trials

Different treatment groups with varying success probabilities can be modeled independently, and the total number of successes across all groups can be analyzed.

Risk Analysis

Assessing the probability of a certain total number of failures or successes across multiple independent processes.

Summary and Key Takeaways

- The collection of independent binomial random variables $X_i \sim \text{Binomial}(n_i, p_i)$ have a joint distribution obtained by multiplying their individual pmfs.
- The joint distribution is explicitly:

$$
f_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i}
$$

- The sum $S = \sum_{i=1}^n X_i$ follows a Poisson-binomial distribution if p_i differ; otherwise, it simplifies to a binomial distribution when parameters are identical.
- The expected value of the sum is $\sum n_i p_i$, and its variance is \(\sum n_i p_i (

Frequently Asked Questions

What is the distribution of the sum of independent Binomial random variables X_i with parameters (n_i, P_i)?

The sum $X = \sum_{i=1}^n X_i$ is itself a binomial distribution only when all P_i are equal; otherwise, it is a generalized distribution that can be expressed as a convolution of the individual binomials. Specifically, if the X_i are independent Bin(ni, Pi), then the distribution of X depends on the joint distribution of these variables and may require a compound or mixture approach for derivation.

How can we derive the probability mass function (PMF) of the sum X = ∑ X_i where each X_i ~ Bin(n_i, P_i) independently?

The PMF of X can be derived using the convolution of the individual binomial pmfs. Since the X_i are independent, the probability that the sum equals k is obtained by summing over all possible combinations of counts that sum to k, involving multinomial coefficients and the probabilities P_i. In some cases, generating functions or recursive formulas can simplify this derivation.

What role do probability generating functions (PGFs) play in deriving the distribution of the sum of independent binomials?

Probability generating functions are useful because the PGF of a sum of independent random variables is the product of their individual PGFs. For Bin(n_i, P_i), the PGF is (1 - P_i + P_i t)^{n_i}. Multiplying these PGFs and expanding allows us to find the distribution of the sum X by extracting coefficients from the resulting polynomial.

Can the sum of independent binomial variables be expressed as another binomial distribution?

Only in the special case where all P_i are equal, the sum of independent Bin(n_i, P) variables is Bin(∑ n_i, P). If the P_i differ, the sum does not follow a binomial distribution but can be represented as a mixture or compound distribution derived from the convolution of their pmfs.

How does the assumption of independence among X_i affect the derivation of the distribution of their sum?

Independence allows the joint distribution to be expressed as the product of individual distributions, simplifying the derivation of the sum's distribution. It enables the use of generating functions and convolution techniques without additional dependence considerations, making the mathematical analysis more straightforward.

What are some practical applications of deriving the distribution of the sum of independent binomial variables?

This derivation is useful in areas like clinical trials, quality control, and risk assessment, where multiple independent sources or processes contribute successes or failures. Understanding the distribution helps in making probabilistic inferences, calculating confidence intervals, and performing hypothesis tests on aggregate data.

What are common methods to approximate the distribution of the sum of independent binomial

random variables when exact calculation is complex?

Common methods include using normal approximation via the Central Limit Theorem, especially for large n; Poisson approximation if P_i are small; or using moment-generating functions to approximate the distribution. These methods simplify calculations and provide useful estimates when exact pmf derivation is computationally intensive.

Additional Resources

Exercise 2. Let $X_1, X_2, \ldots, X_n$ be independent random variables where each X_i follows a binomial distribution with parameters n_i and p_i, denoted as $X_i \sim \mathrm{Bin}(n_i, p_i)$. The task is to derive the distribution of the sum $S = X_1 + X_2 + \ldots + X_n$. This problem is fundamental in probability theory and statistical inference, especially in contexts where multiple independent binomial processes are aggregated, such as in pooled data analysis or combined experiment results. Understanding how these independent binomial variables sum up provides insight into the overall success probability across combined trials, an essential step in various applications including quality control, clinical trials, and survey analysis.

Understanding the Binomial Distribution and Independence

Before delving into the derivation, it’s critical to revisit the foundational concepts of the binomial distribution and independence.

The Binomial Distribution

The binomial distribution models the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success p. Formally, if $X \sim \mathrm{Bin}(n, p)$, then the probability mass function (pmf) is:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k=0,1,...,n$$

This reflects the probability of observing exactly k successes out of n trials.

Independence of Random Variables

The assumption that $X_1, \ldots, X_n$ are independent means that the occurrence of successes in one experiment does not influence the outcomes of others. This independence simplifies the analysis of their sum, as joint probabilities factor into the product of individual probabilities.

Deriving the Distribution of the Sum $S = X_1 + X_2 + ... + X_n$

The core goal is to determine the distribution of S, which is the sum of the independent binomial variables. The key insight is that the sum of independent binomial random variables, each with potentially different parameters, is itself a binomial distribution under certain conditions.

Case 1: Identical Success Probabilities ($p_i = p$ for all i)

If all individual binomial variables have the same success probability p, then:

- Each $X_i \sim \text{Bin}(n_i, p)$

- The sum $S = \sum X_i$

Because the binomial distribution is closed under addition when success probabilities are identical, the sum S follows a binomial distribution with parameters:

$$S \sim \text{Bin}\left(\sum_{i=1}^{n} n_i, p\right)$$

This is a well-known result: the sum of independent binomials with the same p is binomial with the total number of trials equal to the sum of individual trials.

Formal derivation:

1. Since each X_i counts the successes in n_i independent Bernoulli trials with success probability p, their sum counts successes across all trials:

$$S = \sum_{i=1}^{n} X_i$$

2. The total number of trials is:

$$N = \sum_{i=1}^{n} n_i$$

3. Each trial is independent, with the same success probability p, so the combined process is equivalent to N independent Bernoulli trials with probability p.

4. Therefore, $S \sim \text{Bin}(N, p)$.

Case 2: Different Success Probabilities (p_i vary)

When the success probabilities differ among the variables, the distribution of S becomes more complex. In this scenario, the sum S does not follow a simple binomial distribution but can be characterized as a Poisson binomial distribution.

Poisson Binomial Distribution:

- It describes the distribution of a sum of independent Bernoulli trials with different success probabilities.

- The pmf is given by:

$$ P(S = k) = \sum_{A \subseteq \{1,...,n\}, |A|=k} \prod_{i \in A} p_i \prod_{j \notin A} (1 - p_j) $$

- This sum involves all subsets of size k, making direct calculation computationally intensive for large n.

Extension to Binomial Variables:

- Since each $X_i \sim Bin(n_i, p_i)$, the sum S involves more complex convolution processes, and its distribution is a Poisson binomial of weighted sums of Bernoulli variables.

- One way to approach this is to interpret each X_i as the sum of n_i independent Bernoulli trials with parameters p_i, then convolve their distributions.

Practical implications:

- Exact calculations can be cumbersome for large n, but recursive algorithms and generating functions are available for efficient computation.

Using Generating Functions to Derive the Distribution

Generating functions are powerful tools in probability for deriving the distribution of sums of independent variables.

Probability Generating Function (PGF)

- For a binomial random variable $X \sim Bin(n, p)$, the PGF is:

$$ G_X(t) = E[t^X] = (1 - p + p t)^n $$

- For independent variables $X_1, ..., X_n$, their PGFs multiply:

$$ G_S(t) = \prod_{i=1}^{n} (1 - p_i + p_i t)^{n_i} $$

- The distribution of S can be recovered by expanding $G_S(t)$ as a polynomial in t, where the coefficient of t^k gives $P(S=k)$.

Key steps:

1. Write the PGF as a product of individual PGFs.

2. Expand the product to find coefficients.

3. Coefficients correspond to probabilities $P(S=k)$.

This approach handles both the identical and differing p_i cases, though computational complexity grows with n.

Summary of the Distribution of the Sum S

Based on the previous sections, the distribution of the sum $S = \sum X_i$, where each $X_i \sim \text{Bin}(n_i, p_i)$ and the X_i are independent, can be summarized as follows:

- If all p_i are equal:

$$S \sim \text{Bin}\left(\sum_{i=1}^{n} n_i, p\right)$$

- If p_i differ:

$$P(S = k) = \sum_{A \subseteq \{1,...,n\}, |A|=k} \prod_{i \in A} p_i^{n_i} \prod_{j \notin A} (1 - p_j)^{n_j}$$

which constitutes a Poisson binomial distribution over the weighted sum of Bernoulli trials.

Applications and Practical Considerations

Understanding the distribution of the sum of independent binomial variables is more than a theoretical exercise. It has practical relevance across various fields:

- Quality Control: Combining defect counts from multiple production lines, each with different batch sizes and defect probabilities.

- Clinical Trials: Summing successes across different study centers with varying patient responses.

- Survey Sampling: Aggregating yes/no responses from different subpopulations.

Computational techniques:

- When dealing with large n and varying p_i, exact calculation may be computationally infeasible.

- Approximation methods include:

- Normal approximation: Using the central limit theorem when n_i are large.

- Poisson approximation: When success probabilities are small.

- Recursive algorithms and Fourier transforms: Efficient for exact probabilities of Poisson binomial distributions.

Software tools:

- Several statistical software packages and libraries (e.g., R's `poibin` package) facilitate computation of Poisson binomial probabilities.

Conclusion: The Significance of the Derivation

Deriving the distribution of the sum of independent binomial variables is foundational in statistical theory and practice. When the success probabilities are identical, the sum simplifies elegantly to a binomial distribution with parameters summing the individual trials. However, when probabilities differ, the distribution becomes a Poisson binomial, requiring more sophisticated methods to analyze.

This derivation not only deepens our understanding of compound binomial processes but also equips statisticians and data analysts with the tools to model complex, real-world scenarios where multiple independent Bernoulli or binomial processes combine. By leveraging generating functions and understanding the nuanced differences between identical and varying success probabilities, practitioners can accurately assess probabilities, make informed decisions, and interpret aggregated data with confidence.

In essence, this exercise underscores the importance of foundational probability concepts in tackling complex, layered problems—an essential skill in the arsenal of any statistician or data scientist.

Exercise 2 Let X Bin Ni Pi I 1 N Where X1 Xn Are Assumed To Be Independent Derive The

Exercise 2 Let X Bin Ni Pi I 1 N Where X1 Xn Are Assumed To Be Independent Derive The

Related Articles

- Fill Out The Blanks With Appropriate Words In The Following Sentences UPI A. FETs Usually Are Sensitive
- FILL IN THE BLANK. When There Are A Few Unusual Or Extreme Scores, It Is Best To Use The _______ Or The
- Imagine That The Year Is 1970.Elizabeth Is A Housekeeper In Pennsylvania. She Plans To Go To School And

Back to Home