Relationship Between Bernoulli and Binomial Distributions With Extensions to Z-Scores and the Student's t-Distribution

General Statement

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables where $X_i \sim \text{Bernoulli}(p)$ for all $i \in \{1, \ldots, n\}$. Define the aggregate sum:

$$S_n = X_1 + X_2 + \ldots + X_n$$

Then the distribution of S_n is:

$$S_n \sim \text{Binomial}(n, p)$$

This means that the total number of successes in n independent Bernoulli trials, each with success probability p, follows the Binomial distribution with parameters n and p.

Proof

To show $S_n \sim \text{Binomial}(n, p)$, we compute the probability mass function (pmf) of S_n.

Let $P(S_n = k)$ be the probability of obtaining exactly k successes in n trials. A specific sequence with k successes and $(n-k)$ failures has probability:

$$p^k (1-p)^{n-k}$$

The number of such sequences is given by the binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Therefore:

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

This is the pmf of a $\text{Binomial}(n, p)$ distribution.

Simulation Algorithms

Simulating Bernoulli(p)

- Input: probability p

- Output: 1 (success) with probability p, 0 (failure) otherwise
- Algorithm:
  a. Generate $u \sim \text{Uniform}(0,1)$
  b. If $u \leq p$, return 1; else return 0

Simulating Binomial(n, p)

- Input: number of trials n, probability p
- Output: total number of successes in n trials
- Algorithm:
  a. Initialize `count = 0`
  b. Repeat n times:
     - Simulate $\text{Bernoulli}(p)$
     - If result is 1: increment `count`
  c. Return `count`

Computing Z-Scores Using the Normal CDF

Given a normal distribution with mean μ and standard deviation σ, the z-score of a value x is defined as:

$$z=\frac{x-\mu}{\sigma}$$

The cumulative distribution function (CDF) of the standard normal distribution is given by:

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-t^2/2}\,dt$$

This integral is not expressible in elementary functions and is usually evaluated numerically or using error functions:

$$\Phi(z)=\frac{1}{2}\left[1+\text{erf}\left(\frac{z}{\sqrt{2}}\right)\right)$$

where the error function is:

$$\text{erf}(x)=\frac{2}{\sqrt{\pi}}\int_{0}^{x} e^{-t^2}\,dt$$

The Student's t-Distribution

The Student's t-distribution arises when estimating the mean of a normally distributed population with unknown variance, particularly in small samples. It is defined as:

$$T = \frac{\acute{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$$

where:

- $\acute{X}$ is the sample mean
- S is the sample standard deviation
- n is the sample size

As $n \to \infty$, the t-distribution converges to the standard normal distribution due to the Central Limit Theorem.

Computing z-scores in the Student's t-test

In the context of a t-test, the t-statistic plays the role analogous to the z-score:

$$t = \frac{\acute{X} - \mu_0}{S / \sqrt{n}}$$

Here, μ_0 is the null hypothesis value of the population mean. This statistic follows a t_{n-1} distribution. For large n, since $S \to \sigma$, we get:

$$t \approx z = \frac{\acute{X} - \mu_0}{\sigma / \sqrt{n}}$$

Hence, the t-distribution converges to the normal distribution.

Z-tables and t-tables

For practical computations:

- Z-tables (or standard normal tables) provide values of $\Phi(z)$, the CDF of the standard normal distribution.
- t-tables provide critical values for the t-distribution based on degrees of freedom.

Both are widely available in textbooks and online repositories, including:

- https://www.ztable.net/
- https://stattrek.com/statistics/t-distribution.aspx