Relationship Between Bernoulli and Binomial Distributions
With Extensions to Z-Scores and the Students t-Distribution

General Statement

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables where

$$X_i \sim \text{Bernoulli}(p), \quad \text{for all } i \in \{1, \ldots, n\}.$$

Define the aggregate sum:

$$S_n = X_1 + X_2 + \cdots + X_n.$$

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:

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

which is the pmf of a 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 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:

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

This integral cannot be expressed in terms of elementary functions, but it is closely related to the error function:

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

where:

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

The Students t-Distribution

The Students t-distribution arises when estimating the mean of a normally distributed population with unknown variance, especially for small sample sizes.

Definition (via random variables)

Let:

- $Z \sim \mathcal{N}(0,1)$,
- $V \sim \chi^2_\nu$ (chi-squared with ν degrees of freedom),

with Z and V independent. Then the random variable

$$T = \frac{Z}{\sqrt{V/\nu}}$$

has the **Students t-distribution** with ν degrees of freedom:

$$T \sim t_\nu.$$

Probability Density Function (PDF)

The PDF of the t-distribution with ν degrees of freedom is:

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad t \in \mathbb{R}.$$

Properties

- Symmetric about 0.
- Heavier tails than the normal distribution.
- $\lim_{\nu\to\infty} t_\nu = \mathcal{N}(0,1)$
- Mean $= 0$ (for $\nu > 1$), Variance $= \nu/(\nu-2)$ (for $\nu > 2$)

Connection to the t-Statistic

Suppose $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ are i.i.d. Let:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Then the statistic:

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

follows a Student's t-distribution with $n-1$ degrees of freedom.

Computing t-Scores in Hypothesis Testing

In the context of a one-sample t-test, we test $H_0 : \mu = \mu_0$ using the statistic:

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

Under the null hypothesis and assuming normality, this follows a t_{n-1} distribution.

As $n \to \infty$, $S \to \sigma$ and:

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

so the t-distribution approaches the standard normal.

Z-Tables and t-Tables

For practical computations:

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

Tables are widely available online:

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