MTL108
Sampling Distributions-II

Rahul Singh

Continuous Mapping Theorem (CMT)

Theorem 1 (Without proof)**.** *Let $\{X_n\}$ be a sequence of random variables and let X be a random variable. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function at every point of a set C such that $\mathbb{P}(X \in C) = 1$.*
If X_n converges to X, then $g(X_n)$ converges to $g(X)$. Then,

1. ***Convergence in Probability:*** *If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{p} g(X)$.*
2. ***Convergence in Distribution:*** *If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$.*

Example 1 (Consistency of the Sample Standard Deviation)**.** We have already proven that the sample variance, S^2, is a consistent estimator of the population variance, σ^2. That is:

$$S^2 \xrightarrow{p} \sigma^2$$

However, in practice, we often need to estimate the population standard deviation, σ (for instance, when constructing confidence intervals). We use the sample standard deviation, $S = \sqrt{S^2}$. Is S a consistent estimator of σ?
Application of CMT: We define our function as $g(t) = \sqrt{t}$. This function is strictly continuous for all strictly positive numbers ($t > 0$). Since the population variance σ^2 is a positive constant (assuming a non-degenerate distribution), the function $g(t)$ is continuous at the limit point σ^2.
Because $S^2 \xrightarrow{p} \sigma^2$ and $g(t) = \sqrt{t}$ is continuous at σ^2, we can directly apply the Continuous Mapping Theorem, we have

$$g(S^2) \xrightarrow{p} g(\sigma^2)$$
$$\sqrt{S^2} \xrightarrow{p} \sqrt{\sigma^2}$$
$$S \xrightarrow{p} \sigma$$

Conclusion: Thanks to the Continuous Mapping Theorem, we can instantly prove that the sample standard deviation S is a consistent estimator of the population standard deviation σ, without having to calculate expected values or variances of square roots.

Joint Convergence in Probability

When working with multiple sequences of random variables, we often need to establish that they converge together as a vector. A fundamental and highly useful property of convergence in probability is that marginal convergence implies joint convergence.

Theorem 2 (Without proof)**.** *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables defined on the same probability space, and let X and Y be random variables (or constant values). If:*

$$X_n \xrightarrow{p} X \quad \text{and} \quad Y_n \xrightarrow{p} Y$$

as $n \to \infty$, then the random vector (X_n, Y_n) converges in probability to the random vector (X, Y):

$$(X_n, Y_n) \xrightarrow{p} (X, Y)$$

Definition (bivariate): Formally, convergence in probability for random vectors is defined using the Euclidean norm (distance). For any given $\epsilon > 0$:

$$\lim_{n\to\infty} P\left(\sqrt{(X_n - X)^2 + (Y_n - Y)^2} > \epsilon\right) = 0$$

Application Note: This theorem is the crucial stepping stone for the Multivariate Continuous Mapping Theorem. By establishing that the vector $(X_n, Y_n) \xrightarrow{p} (X, Y)$, we can instantly prove that any continuous multivariable function $g(X_n, Y_n) \xrightarrow{p} g(X, Y)$. This is the exact mechanism that formally justifies the algebraic operations (addition, multiplication, division).

Theorem 3. *Let $X_1, X_2, \ldots, X_n$ be a random sample of IID RVs with mean μ, variance σ^2 and finite fourth moment. Let $\bar{X} = \dfrac{1}{n}\sum_{i=1}^n X_i$, $S_n^2 = \dfrac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$, and $S_{n-1}^2 = \dfrac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Then,*

$$S_n^2 \xrightarrow{p} \sigma^2 \quad \text{and} \quad S_{n-1}^2 \xrightarrow{p} \sigma^2.$$

Proof. First, we expand the formula for S_n^2 into its computational form.

$$\begin{aligned} S_n^2 &= \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\bar{X}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) + \frac{1}{n}\sum_{i=1}^n \bar{X}^2. \end{aligned}$$

Since $\frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$, this simplifies beautifully:

$$S_n = \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\bar{X}(\bar{X}) + \bar{X}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2.$$

Now, we evaluate the convergence of these two terms separately using the Weak Law of Large Numbers (WLLN). Because $X_1, \ldots, X_n$ are IID, their squares $X_1^2, \ldots, X_n^2$ are also IID. The expected value of the squared observations is $\mathbb{E}[X^2] = \text{Var}(X) + (\mathbb{E}[X])^2 = \sigma^2 + \mu^2$.

By the WLLN, the sample average of these squared terms converges in probability to their expected value, that is,

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} \sigma^2 + \mu^2.$$

Similarly, by the WLLN, the sample mean converges in probability to the population mean,

$$\bar{X} \xrightarrow{p} \mu.$$

We now apply the Continuous Mapping Theorem (CMT). The theorem states that for any continuous function $g(u, v)$, if $(U_n, V_n) \xrightarrow{p} (u, v)$, then $g(U_n, V_n) \xrightarrow{p} g(u, v)$. Let $U_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ and $V_n = \bar{X}$. Our function is $g(u, v) = u - v^2$, which is strictly continuous. Applying the CMT to the limits, we have

$$S_n = U_n - V_n^2 \xrightarrow{p} (\sigma^2 + \mu^2) - (\mu)^2 = \sigma^2$$

Next, we can express the unbiased sample variance S_{n-1} algebraically in terms of the unadjusted variance S_n:

$$S_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{n}{n-1}S_n.$$

We know that $S_n \xrightarrow{p} \sigma^2$. The multiplier $\frac{n}{n-1}$ is a deterministic sequence that converges to 1 as $n \to \infty$. We can define another continuous function $h(c, x) = c \cdot x$. Because the sequence $\frac{n}{n-1} \to 1$ and $S_n \xrightarrow{p} \sigma^2$, we apply the Continuous Mapping Theorem again, we get

$$S_{n-1} = \frac{n}{n-1}S_n \xrightarrow{p} 1 \cdot \sigma^2 = \sigma^2$$

□

Slutsky's Lemma

Slutsky's Lemma is a fundamental theorem in asymptotic statistics. It allows us to combine sequences of random variables that converge in different modes (specifically, convergence in distribution and convergence in probability) using basic algebraic operations.

Theorem 4. *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables. Suppose that as $n \to \infty$:*

- $X_n \xrightarrow{d} X$ *(The sequence X_n converges in distribution to a random variable X)*
- $Y_n \xrightarrow{p} c$ *(The sequence Y_n converges in probability to a finite constant c)*

Then, the following properties hold as $n \to \infty$:

1. ***Addition:*** $X_n + Y_n \xrightarrow{d} X + c$
2. ***Multiplication:*** $X_n Y_n \xrightarrow{d} cX$
3. ***Division:*** $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$, *provided that* $c \neq 0$.

Replacing σ with $S_n =$ in the Central Limit Theorem

By the Central Limit Theorem (CLT), we know that for a random sample of size n drawn from a population with mean μ and finite variance σ^2, the standardized sample mean converges in distribution to a Standard Normal distribution,

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1).$$

In practice, the true population standard deviation σ is almost always unknown and must be estimated using the sample standard deviation S_n. We want to determine the asymptotic distribution of the test statistic when σ is replaced by S_n, that is,

$$T_n = \frac{\bar{X} - \mu}{S_n/\sqrt{n}} \xrightarrow{d} N(0,1).$$

Theorem 5. *Let $X_1, X_2, \ldots, X_n$ be a random sample of IID RVs with mean μ, variance σ^2 and finite fourth moment. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Then,*

$$T_n = \frac{\bar{X} - \mu}{S_n/\sqrt{n}} \xrightarrow{d} N(0,1).$$

Proof. We can rewrite T_n by multiplying and dividing by σ as

$$T_n = \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)\left(\frac{\sigma}{S_n}\right) = Z_n \cdot Y_n, \quad \textit{where} \quad Z_n = \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right) \quad \textit{and} \quad Y_n = \left(\frac{\sigma}{S_n}\right).$$

Now, we evaluate the asymptotic behavior of the two components.

1. Let $Z_n = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$. By the CLT, $Z_n \xrightarrow{d} Z$, where $Z \sim N(0,1)$.
2. Let $Y_n = \frac{\sigma}{S_n}$. We know that the sample variance is a consistent estimator of the population variance, meaning $S_n^2 \xrightarrow{p} \sigma^2$. By the Continuous Mapping Theorem (CMT), since $g(t) = \sigma/\sqrt{t}$ is continuous at $\sigma^2 > 0$, we have $Y_n = \frac{\sigma}{S_n} \xrightarrow{p} \frac{\sigma}{\sigma} = 1$.

Because $X_n \xrightarrow{d} N(0,1)$ and $Y_n \xrightarrow{p} 1$, we apply the multiplication rule of Slutsky's Lemma ($c = 1$):

$$T_n = X_n Y_n \xrightarrow{d} 1 \cdot Z = Z \sim N(0,1)$$

□

Conclusion: Estimating the variance from the sample does not change the asymptotic distribution. As $n \to \infty$, the T_n statistic converges to a Standard Normal distribution, just like the Z_n statistic.

Example 2 (Application: Large-Sample Confidence Intervals)**.** The most immediate practical application of this result is the construction of large-sample confidence intervals for a population mean when the population variance is unknown.
Because $T_n \xrightarrow{d} N(0,1)$, for a sufficiently large sample size (typically $n \geq 30$), we can approximate the distribution of T_n with the standard normal distribution. Let $z_{\alpha/2}$ be the critical value such that $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$. We can write the approximate probability statement as

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{S_n/\sqrt{n}} \leq z_{\alpha/2}\right) \approx 1 - \alpha$$

Rearranging the inequalities to isolate the unknown parameter μ in the center we get

$$P\left(\bar{X} - z_{\alpha/2}\frac{S_n}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{S_n}{\sqrt{n}}\right) \approx 1 - \alpha.$$

This yields the standard $(1-\alpha)100\%$ large-sample confidence interval for μ as

$$\bar{X} \pm z_{\alpha/2}\frac{S_n}{\sqrt{n}}$$

Without Slutsky's Lemma, we would have no rigorous mathematical justification for plugging S_n into this formula.

Independence of the Sample Mean and Sample Variance for normal RVs

A unique and incredibly powerful property of the Normal distribution is that the sample mean ($\bar{X}$) and the sample variance ($S^2 := S^2_{n-1}$) are completely independent random variables. This property is crucial for the later development of the t-distribution and hypothesis testing.

Theorem 6. *Let $X_1, X_2, \ldots, X_n$ be a random sample of IID random variables from $\mathcal{N}(\mu, \sigma^2)$. Let $\bar{X}$ be the sample mean and S^2 be the sample variance. Then, $\bar{X}$ and S^2 are independent.*

Proof. Establish the joint normality of the sample mean and the deviations. First, we define the sample mean and the i-th deviation from the sample mean as follows:

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, \quad D_i = X_i - \bar{X}$$

Because the original variables $X_1, \ldots, X_n$ are independent and normally distributed, they are jointly normal. Since both $\bar{X}$ and D_i are strictly linear combinations of these jointly normal random variables, the set containing the sample mean and all deviations $(\bar{X}, D_1, D_2, \ldots, D_n)$ also follows a multivariate normal distribution.

Calculate the covariance between the sample mean and any deviation. For jointly normal random variables, if their covariance is exactly zero, they are completely independent. Let us calculate the covariance between $\bar{X}$ and the i-th deviation D_i,

$$\mathrm{Cov}(\bar{X}, D_i) = \mathrm{Cov}(\bar{X}, X_i - \bar{X})$$

Using the properties of covariance, we can rewrite this expression as

$$\mathrm{Cov}(\bar{X}, X_i - \bar{X}) = \mathrm{Cov}(\bar{X}, X_i) - \mathrm{Cov}(\bar{X}, \bar{X})$$

By definition, the covariance of a variable with itself is its variance, consequently,

$$\mathrm{Cov}(\bar{X}, \bar{X}) = \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Next, we evaluate the covariance between the sample mean and a single observation X_i,

$$\mathrm{Cov}(\bar{X}, X_i) = \mathrm{Cov}\left(\frac{1}{n}\sum_{j=1}^{n} X_j, X_i\right)$$

Because the observations are independent, $\text{Cov}(X_j, X_i) = 0$ for all $j \neq i$. The only non-zero term in the sum occurs when $j = i$, so

$$\text{Cov}(\bar{X}, X_i) = \frac{1}{n}\text{Cov}(X_i, X_i) = \frac{1}{n}\text{Var}(X_i) = \frac{\sigma^2}{n.}$$

Substituting both results back into our original covariance equation, we get

$$\text{Cov}(\bar{X}, D_i) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0.$$

Because the covariance is zero and the variables are jointly normal, the sample mean $\bar{X}$ is perfectly independent of every single deviation $D_i = (X_i - \bar{X})$. The sample variance is defined entirely as a function of these deviations, that is,

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

A fundamental principle of probability states that if a random variable Y is independent of a set of variables $Z_1, \ldots, Z_n$, it is also independent of any function $g(Z_1, \ldots, Z_n)$. Since $\bar{X}$ is independent of all D_i, it must also be independent of S^2. This completes the proof. □

Sampling Distribution of the Sample Variance (S^2)

While the sample mean tracks the center of the data, the sample variance tracks the dispersion. To understand how $S^2 := S^2_{n-1}$ behaves from sample to sample, we must introduce the Chi-Square (χ^2) distribution.

The Chi-Square (χ^2) Distribution

If $Z_1, Z_2, \ldots, Z_k$ are IID $\mathcal{N}(0, 1)$ random variables, their sum of squares follows a Chi-Square distribution with k degrees of freedom. It is denoted by

$$Y = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k$$

Properties:

- **Support:** bounded to the right of zero ($Y \geq 0$).
- **Shape:** Right-skewed for small k, becoming more symmetric as k increases.
- **Moments:** Mean $= k$, Variance $= 2k$.
- **PDF:** For a continuous random variable $X \sim \chi^2_k$, the PDF is given by

$$f(x) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

  where $\Gamma(\cdot)$ represents the Gamma function, defined as $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$.
- **MGF:** The MGF of $X \sim \chi^2_k$ is given by

$$M_Y(t) = (1 - 2t)^{-k/2} \quad \text{for } t < \frac{1}{2}.$$

The Additive Property of Chi-Square Variables

A powerful feature of the Chi-Square distribution is that the sum of independent Chi-Square variables is itself a Chi-Square variable.

Theorem 7. *Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables such that each $X_i \sim \chi^2_{k_i}$ with k_i degrees of freedom. Then, their sum $Y = \sum_{i=1}^n X_i$ follows a Chi-Square distribution with degrees of freedom equal to the sum of the individual degrees of freedom, precisely,*

$$Y \sim \chi^2_{\sum_{i=1}^n k_i}$$

Proof. We will prove this using the properties of MGFs. The MGF of a sum of independent random variables is equal to the product of their individual MGFs. Therefore, the MGF of Y is

$$M_Y(t) = E[e^{tY}] = E\left[\exp\left(t\sum_{i=1}^n X_i\right)\right] = E\left[\prod_{i=1}^n e^{tX_i}\right].$$

Because the X_i variables are independent, the expectation of the product is the product of the expectations, so

$$M_Y(t) = \prod_{i=1}^n E[e^{tX_i}] = \prod_{i=1}^n M_{X_i}(t)$$

Substituting the known MGFs for each $X_i \sim \chi^2_{k_i}$, we get

$$M_Y(t) = \prod_{i=1}^n (1-2t)^{-k_i/2}.$$

Observe that

$$M_Y(t) = (1-2t)^{-k_1/2} \cdot (1-2t)^{-k_2/2} \cdots (1-2t)^{-k_n/2} = (1-2t)^{-\frac{1}{2}\sum_{i=1}^n k_i}$$

Now, by the Uniqueness Theorem of Moment Generating Functions, since $M_Y(t)$ has the exact functional form of a Chi-Square MGF with parameter $\sum_{i=1}^n k_i$, the random variable Y must follow a Chi-Square distribution with $\sum_{i=1}^n k_i$ degrees of freedom. This completes the proof. □

Theorem 8. *Let $X_1, X_2, \ldots, X_n$ be a random sample of IID random variables from $\mathcal{N}(\mu, \sigma^2)$. Let $\bar{X}$ be the sample mean and S^2 be the sample variance. Then,*

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Proof: Since each $X_i \sim N(\mu, \sigma^2)$, the standardized variable $Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$. By the definition of the Chi-Square distribution, the sum of n independent squared standard normal variables follows a χ^2_n distribution, consequently,

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n.$$

We algebraically decompose the numerator by strategically adding and subtracting the sample mean $\bar{X}$,

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\mu)^2 &= \sum_{i=1}^{n}((X_i-\bar{X})+(\bar{X}-\mu))^2 \\ &= \sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{i=1}^{n}(\bar{X}-\mu)^2 + 2\sum_{i=1}^{n}(X_i-\bar{X})(\bar{X}-\mu) \\ &= \sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2.\end{aligned}$$

Dividing by the population variance σ^2 to standardize the terms:

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2}$$

Let us carefully identify each of these three components:

1. **Left-Hand Side (LHS):** As established, this is the sum of n squared standard normal variables, so LHS $\sim \chi_n^2$.
2. **First RHS Term:** Noting that $\sum_{i=1}^{n}(X_i-\bar{X})^2 = (n-1)S^2$, this term becomes $\frac{(n-1)S^2}{\sigma^2}$. This is the target variable whose distribution we wish to find.
3. **Second RHS Term:** We can rewrite this term as $\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$. Because $\bar{X} \sim N(\mu, \sigma^2/n)$, the term inside the parenthesis is a single standard normal variable Z. Therefore, its square follows a χ_1^2 distribution.

Let $W = \frac{(n-1)S^2}{\sigma^2}$. Our equation is now analytically reduced to

$$\chi_n^2 = W + \chi_1^2$$

Exploiting Independence via MGFs We previously proved that for a normal distribution, the sample mean $\bar{X}$ and the sample variance S^2 are independent. Therefore, the random variables W and χ_1^2 are independent.

The MGF of a sum of independent random variables is the product of their individual MGFs, so

$$M_{\chi_n^2}(t) = M_W(t) \cdot M_{\chi_1^2}(t)$$

Recall the MGF for a Chi-Square distribution with k degrees of freedom is $M(t) = (1-2t)^{-k/2}$ for $t < 1/2$. Substituting the known MGFs into our equation:

$$(1-2t)^{-n/2} = M_W(t) \cdot (1-2t)^{-1/2}$$

We solve for $M_W(t)$ by dividing both sides:

$$M_W(t) = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-n/2-(-1/2)} = (1-2t)^{-(n-1)/2}$$

Step 4: Conclusion The function $(1-2t)^{-(n-1)/2}$ is exactly the recognizable Moment Generating Function of a Chi-Square distribution with $n-1$ degrees of freedom. Because MGFs uniquely identify probability distributions, we conclude that:

$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

This completes the proof.

Remark: This theorem strictly requires the underlying population to be Normally distributed. S^2 is sensitive to departures from normality.

Example 3 (Bowling Machine Consistency)**.** A mechanical bowling machine used for cricket practice is designed to deliver balls with a highly consistent speed. The variance in delivery speed is known to be $\sigma^2 = 4\ (\text{km/h})^2$. The speeds are normally distributed. A coach records a sample of $n = 16$ deliveries. What is the probability that the sample variance S^2 exceeds $7.44\ (\text{km/h})^2$?

Solution: We know that the statistic $\frac{(n-1)S^2}{\sigma^2}$ follows a χ^2 distribution with $n-1 = 15$ degrees of freedom. We want to find $P(S^2 > 7.44)$. Let's convert S^2 into our χ^2 test statistic,

$$\chi^2 = \frac{(16-1)(7.44)}{4} = \frac{15 \times 7.44}{4} = \frac{111.6}{4} = 27.9$$

We now look up $P(\chi^2_{15} > 27.9)$. Looking at a Chi-Square distribution table for $df = 15$, the critical value for an upper-tail area of 0.025 is 27.488, and for 0.01 it is 30.578. *Conclusion:* Since 27.9 falls between these values, the probability of observing a sample variance this high is approximately between 1% and 2.5%.

Practice Problem 2

The lifespan of a certain type of sensor used in spatial data collection is normally distributed with a population variance of $\sigma^2 = 25$ months^2. A research team tests a random sample of $n = 10$ sensors. Find the probability that the sample variance S^2 will be less than 11.6 months^2. *(Hint: Find the corresponding χ^2 value for $df = 9$ and use the table.)*

Student's t-Distribution

While the Central Limit Theorem and Slutsky's Lemma guarantee that $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is asymptotically normal for large samples ($n \geq 30$), we frequently encounter scenarios where our sample size is small, the population variance σ^2 is unknown, and the underlying population is Normal. In these finite-sample cases, the normal approximation fails because substituting S for σ introduces additional variability. To handle this, we use the Student's t-distribution.

Definition (t-distribution)

Let Z be a standard normal random variable, $Z \sim N(0,1)$. Let V be a Chi-Square random variable with k degrees of freedom, $V \sim \chi^2_k$. If Z and V are independent, then the random variable T

defined by

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

follows a Student's t-distribution with k degrees of freedom, denoted as $T \sim t_k$.

Key Properties of the t-Distribution

- **Symmetry:** Like the standard normal distribution, the t-distribution is perfectly symmetric around a mean of zero ($\mathbb{E}[T] = 0$ for $k > 1$).
- **Variance:** The variance of the t-distribution is $\text{Var}(T) = \frac{k}{k-2}$ (for $k > 2$). Notice that the variance is strictly greater than 1, unlike the standard normal.
- **Asymptotic Normality:** As the degrees of freedom increase ($k \to \infty$), the t-distribution converges exactly to the Standard Normal distribution, $N(0, 1)$.
- **Heavier Tails:** Because the denominator $\sqrt{V/k}$ introduces extra random variation, the t-distribution has heavier (fatter) tails than the normal distribution. This reflects the increased uncertainty of having to estimate the population variance.

An application

The most profound application of the t-distribution is deriving the exact sampling distribution of the standardized sample mean when the population standard deviation σ is estimated by the sample standard deviation S.

Theorem 9. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a Normal distribution, $N(\mu, \sigma^2)$. Then the statistic*

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

follows a t-distribution with $n-1$ degrees of freedom.

Proof: Because the underlying population is Normal, the sample mean is exactly normally distributed: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. Standardizing this gives our Z variable:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

We previously proved that the scaled sample variance follows a Chi-Square distribution with $n-1$ degrees of freedom:

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Here, our degrees of freedom parameter is $k = n - 1$.

We also previously proved that for a sample from a Normal distribution, the sample mean $\bar{X}$ and the sample variance S^2 are independent. Therefore, our variables Z and V are independent, satisfying the prerequisite for the t-distribution.

Now, we plug Z, V, and k directly into the formal definition of the t-distribution,

$$T = \frac{Z}{\sqrt{V/k}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}} = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)\left(\frac{\sigma}{S}\right) = \frac{\bar{X}-\mu}{S/\sqrt{n}}.$$

By showing that $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ maps perfectly to the definition of the t-distribution, therefore $T \sim t_{n-1}$.

Example 4 (Confidence Intervals for small sample)**.** Used to construct exact interval estimates for μ when samples are small and the population is assumed to be normal: $\bar{X} \pm t_{\alpha/2,n-1}\left(\frac{S}{\sqrt{n}}\right)$.

The F-Distribution

To complete our toolkit of sampling distributions, we introduce the F-distribution (named in honor of Sir Ronald A. Fisher). While the t-distribution is primarily used for making inferences about population means, the F-distribution is the fundamental tool for comparing the variances of two independent populations.

Definition (F-Distribution)

Let U and V be two independent random variables such that:

- $U \sim \chi^2_{d_1}$ (a Chi-Square distribution with d_1 degrees of freedom)
- $V \sim \chi^2_{d_2}$ (a Chi-Square distribution with d_2 degrees of freedom)

The random variable F, defined as the ratio of these two independent Chi-Square variables, each divided by its respective degrees of freedom, follows an F-distribution with d_1 and d_2 degrees of freedom:

$$F = \frac{U/d_1}{V/d_2} \sim F_{d_1,d_2}$$

Here, d_1 is called the *numerator degrees of freedom* and d_2 is the *denominator degrees of freedom*.

Key Properties of the F-Distribution

- **Support:** Because it is derived from Chi-Square variables (which represent sums of squares and are strictly non-negative), the F-distribution is bounded at zero: $F \geq 0$.
- **Shape:** It is heavily right-skewed, especially for small degrees of freedom. As both d_1 and d_2 approach infinity, the distribution slowly converges to a normal distribution.
- **Reciprocal Property:** A unique mathematical feature of the F-distribution is that if $F \sim F_{d_1,d_2}$, then its reciprocal follows an F-distribution with reversed degrees of freedom: $\frac{1}{F} \sim F_{d_2,d_1}$. This property is highly useful when calculating critical values for lower tails using statistical tables.

Application: The Sampling Distribution of the Ratio of Variances

The primary application of the F-distribution in sampling theory arises when we want to compare the dispersion of two independent Normal populations.

Theorem 10. *Let $X_1, \ldots, X_{n_1}$ be a random sample from $N(\mu_1, \sigma_1^2)$, and let $Y_1, \ldots, Y_{n_2}$ be an independent random sample from $N(\mu_2, \sigma_2^2)$. Let S_1^2 and S_2^2 be their respective sample variances. Then the ratio*

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

follows an F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

Proof: We previously established that for a random sample from a normal population, the scaled sample variance follows a Chi-Square distribution. For the first sample:

$$U = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1 - 1}$$

For the second sample:

$$V = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2 - 1}$$

Because the two original random samples (X and Y) are drawn from completely independent populations, any statistics calculated from them (including the sample variances S_1^2 and S_2^2, and therefore U and V) are also independent. This satisfies the core requirement of the F-distribution.

We plug U and V into the formal definition of the F-distribution, using their respective degrees of freedom $d_1 = n_1 - 1$ and $d_2 = n_2 - 1$, we have

$$F = \frac{U/(n_1 - 1)}{V/(n_2 - 1)} = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1 - 1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2 - 1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}.$$

Therefore, this ratio perfectly matches the definition of an F_{n_1-1, n_2-1} random variable.

Example 5 (Equality of Two Variances)**.** In hypothesis testing, we often test the null hypothesis that two population variances are equal ($H_0 : \sigma_1^2 = \sigma_2^2$). Under this null hypothesis, the ratio σ_1^2/σ_2^2 becomes exactly 1, and our test statistic collapses beautifully to simply the ratio of the sample variances: $F = \frac{S_1^2}{S_2^2}$.

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

[2] Ross, Sheldon M. (2020). *Introduction to probability and statistics for engineers and scientists.* Academic press.

Disclaimer