SAMPLING DISTRIBUTIONS FROM A NORMAL POPULATION

Let $X_1, X_2, \ldots, X_n$ be a sample from a normal population having mean μ and variance σ^2. That is, they are independent and $X_i \sim \mathcal{N}(\mu, \sigma^2), i = 1, \ldots, n$. Also let

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

and

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

denote the sample mean and sample variance, respectively. We would like to compute their distributions.

Distribution of the sample mean

Since the sum of independent normal random variables is normally distributed, it follows that $\bar{X}$ is normal with mean

$$E[\bar{X}] = \sum_{i=1}^{n} \frac{E[X_i]}{n} = \mu$$

and variance

$$Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

That is, $\bar{X}$, the average of the sample, is normal with a mean equal to the population mean ?but with a variance reduced by a factor of $1/n$. It follows from this that

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

is a standard normal random variable.

Joint Distribution of $\bar{X}$ and S^2

For number $x_1, \ldots, x_n$, let $y_i = x_i - \mu, i = 1, \ldots, n$. Then as $\bar{y} = \bar{x} - \mu$, it follows that

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

that

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2$$

Now, if $X_1, \ldots, X_n$ is a sample from a normal population having mean μ variance σ^2, then we obtain from the preceding identity that

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

Equivalently

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

- $\frac{(X_i-\mu)}{\sigma}, i=1,\ldots,n$ are independent standard normals. Hence LHS is a chi-square random variable with n degrees of freedom.
- $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ is a standard normal variable hence a chi-square random variable with 1 degree of freedom.
- Therefore, $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$ is a chi-square random variable with $n-1$ degrees of freedom.

Theorem

If $X_1,\ldots,X_n$ is a sample from a normal population having mean μ and variance σ^2, then $\bar{X}$ and S^2 are independent random variables, with $\bar{X}$ being normal with mean μ and variance σ^2/n and $(n-1)S^2/\sigma^2$ being chi-square with $n-1$ degrees of freedom.

Sampling Distribution of the Variance (Chi-Square)

Concept:

- When a sample is drawn from a normal population, the sample variance S^2 fluctuates around the true variance σ^2.
- The quantity $\frac{(n-1)S^2}{\sigma^2}$ follows a χ^2 distribution with $(n-1)$ degrees of freedom.

Formula:

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

Example:

- A machine produces bolts with $\sigma^2 = 0.01$ mm^2.
- From $n = 10$ bolts, $S^2 = 0.012$ mm^2.
- $\chi^2 = \frac{9 \times 0.012}{0.01} = 10.8$.
- Compare with χ^2_9 distribution to test consistency.

Independence of $\bar{X}$ and S^2

Concept:

- For a normal population, the sample mean $\bar{X}$ and the sample variance S^2 are independent.
- $\bar{X}$ measures location; S^2 measures spread.

Intuition (short):

- The center of the data and the spread vary separately in the Gaussian model.
- This property is unique to the normal distribution.

Example:

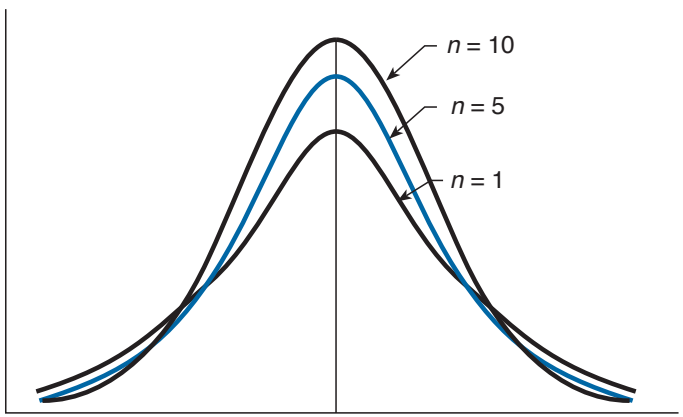
- Stock returns over 30 days: mean daily return ($\bar{X}$) and volatility (S^2) fluctuate independently.

t-Distribution

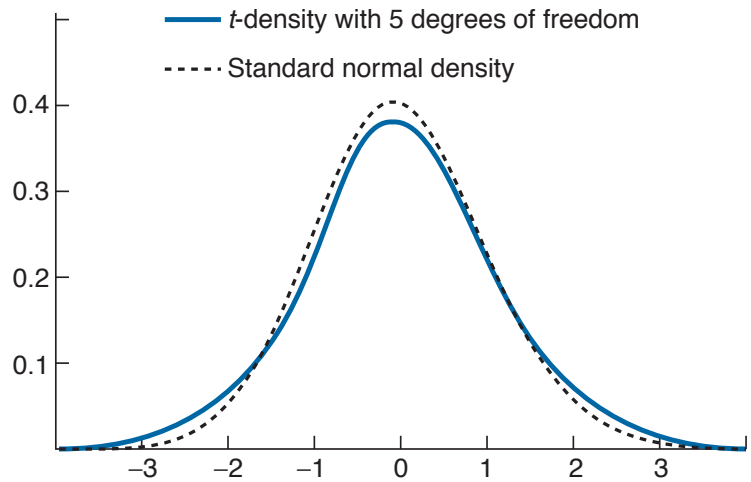
If Z and χ^2_n are independent random variables, with Z having a standard normal distribution and χ^2_n having a chi-square distribution with n degrees of freedom, then the random variable T_n defined by

$$T_n = \frac{Z}{\sqrt{\chi^2_n/n}}$$

is said to have a t-distribution with n degrees of freedom



Density function of T_n.



Comparing standard normal density with the density of T_5

The t-Distribution

Definition:

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

When to Use:

- When population variance σ^2 is unknown.
- Small samples (n ¡ 30).

Properties:

- Symmetric like normal, but with heavier tails.
- As n increases, $t_{n-1} \to N(0, 1)$.

Intuition: Why t Has Heavier Tails

- **Random denominator:** Replacing σ with S makes the standard error random, inflating tail probabilities.
- **Small-n volatility:** With few points, S can under/overestimate σ, producing extreme T values more often.
- **Convergence:** As n grows, $S \to \sigma$ and $t \to Z$.

t-Distribution

The mean and variance of Tn can be shown to equal

$$E[T_n] = 0, \qquad n > 1$$

$$Var(T_n) = \frac{n}{n-2}, \qquad n > 2$$

Thus the variance of T_n decreases to 1 – the variance of a standard normal random variable – as n increases to ∞.

Corollary

Let $X_i, \ldots, X_n$ be a sample from a normal population with mean μ. If $\bar{X}$ denotes the sample mean and S the sample standard deviation, then

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

Corollary

Let $X_i, \ldots, X_n$ be a sample from a normal population with mean μ. If $\bar{X}$ denotes the sample mean and S the sample standard deviation, then

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

Proof: A t–random variable with n degrees of freedom is defined as the distribution of

$$\frac{Z}{\sqrt{\chi^2_n/n}}$$

where Z is a standard normal random variable that is independent of χ^2_n, a chi-square random variable with n degrees of freedom. It follows from previous theorem,

$$\frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{S^2/\sigma^2} = \sqrt{n}}$$

is a t-random variable with $n - 1$ degrees of freedom.

Corollary Linking Normal, Chi-Square, and t

Given:

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

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

Z, S^2 are independent.

Therefore:

$$T = \frac{Z}{\sqrt{\chi^2_{n-1}/(n-1)}} = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

Interpretation:

- Accounts for uncertainty in S.
- Leads to wider confidence intervals for small n.

Real-World Applications of the t-Distribution

Engineering

- Quality control (blade strength test)
- Sensor calibration (voltage consistency)

Medicine

- Drug trials with few patients
- Paired t-test for pre/post measurements

Finance

- Testing mean return > 0 with few data points

Education

- Comparing class averages (Class A vs B)

Psychology

- Testing mean stress reduction after meditation

Retail/Business

- Before/after marketing campaign sales comparison

Example

Suppose that 45 percent of the population favors a certain candidate in an upcoming election. If a random sample of size 200 is chosen, find

(a) The expected value and standard deviation of the number of members of the sample that favor the candidate;

(b) The probability that more than half the members of the sample favor the candidate.

Example

Suppose that 45 percent of the population favors a certain candidate in an upcoming election. If a random sample of size 200 is chosen, find

(a) The expected value and standard deviation of the number of members of the sample that favor the candidate;

(b) The probability that more than half the members of the sample favor the candidate.

Solution:

(a) The expected value and standard deviation of the proportion that favor the candidate are

$$E[X] = 200(0.45) = 90, \qquad SD(X) = \sqrt{200(0.45)(1-0.45)} = 7.0356$$

Example

Suppose that 45 percent of the population favors a certain candidate in an upcoming election. If a random sample of size 200 is chosen, find

(a) The expected value and standard deviation of the number of members of the sample that favor the candidate;

(b) The probability that more than half the members of the sample favor the candidate.

Solution:

(a) The expected value and standard deviation of the proportion that favor the candidate are

$$E[X] = 200(0.45) = 90, \qquad SD(X) = \sqrt{200(0.45)(1-0.45)} = 7.0356$$

(b) Using the normal approximation to the binomial with parameters 200 and 0.45

$$P\{X \geq 101\} = P\{X \geq 100.5\} \quad \text{(the continuity correction)}$$
$$= P\left\{\frac{X-90}{7.0356} \geq \frac{100.5-90}{7.0356}\right\} \approx P\{Z \geq 1.4924\} \approx 0.0678$$

Example

According to the U.S. Department of Agriculture's World Livestock Situation, the country with the greatest per capita consumption of pork is Denmark. In 1994, the amount of pork consumed by a person residing in Denmark had a mean value of 147 pounds with a standard deviation of 62 pounds. If a random sample of 25 Danes is chosen, approximate the probability that the average amount of pork consumed by the members of this group in 1994 exceeded 150 pounds.

Example

According to the U.S. Department of Agriculture's World Livestock Situation, the country with the greatest per capita consumption of pork is Denmark. In 1994, the amount of pork consumed by a person residing in Denmark had a mean value of 147 pounds with a standard deviation of 62 pounds. If a random sample of 25 Danes is chosen, approximate the probability that the average amount of pork consumed by the members of this group in 1994 exceeded 150 pounds. **Solution:** If we let X_i be the

amount consumed by the ith member of the sample, $i = 1, \ldots, 25$, then the desired probability is

$$P\left\{\frac{X_1 + \cdots + X_{25}}{25} > 150\right\} = P\{\bar{X} > 150\}$$

Since we can regard the X_i as being independent random variables with mean 147 and standard deviation 62, it follows from the central limit theorem that their sample mean will be approximately normal with mean 147 and standard deviation $62/5$. Thus, with Z being a standard normal random variable, we have

$$P\{\bar{X} > 150\} = P\left\{\frac{\bar{X} - 147}{12.4} > \frac{150 - 147}{12.4}\right\} \approx P\{Z > 0.242\} \approx 0.404$$

Summary

- S^2 follows a χ^2 distribution for normal data.
- $\bar{X}$ and S^2 are independent in normal populations.
- The t-distribution arises when σ is unknown.
- As n grows, t becomes normal.

Key Formula:

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

Introduction to MLE

- MLE is a method to estimate unknown parameters by maximizing the likelihood function.
- **Why MLE?** Used in various industries including manufacturing, finance, and machine learning.
- **Goal:** Find the parameter value that makes the observed data most likely.

MLE in Predictive Maintenance

Example: Machine Failure Rate

- Data: Time-to-failure for n machines.
- Model: Exponential distribution with mean θ.
- **MLE Estimate:** $\hat{\theta} = \frac{\sum x_i}{n}$
- **Outcome:** Optimized maintenance schedules, reduced downtime.

MLE in Quality Control

Example: Estimating Defect Rate

- Data: Binary defective/non-defective product classification.
- Model: Bernoulli distribution with probability p.
- **MLE Estimate:** $\hat{p} = \frac{\sum x_i}{n}$
- **Impact:** Process improvement and cost reduction.

MLE in Customer Analytics

Example: Call Center Optimization

- Data: Number of customer calls per hour.
- Model: Poisson distribution with rate λ.
- **MLE Estimate:** $\hat{\lambda} = \frac{\sum x_i}{n}$
- **Business Impact:** Improved staffing and efficiency.

Maximum Likelihood Estimators

- Any statistic used to estimate the value of an unknown parameter θ is called an **estimator** of θ.
- The observed value of the estimator is called the estimate.
- Suppose that the random variables $X_1, \ldots, X_n$, whose joint distribution is assumed given except for an unknown parameter θ, are to be observed.
- The problem of interest is to use the observed values to estimate θ.

Maximum Likelihood Estimators

- Any statistic used to estimate the value of an unknown parameter θ is called an **estimator** of θ.
- The observed value of the estimator is called the estimate.
- Suppose that the random variables $X_1, \ldots, X_n$, whose joint distribution is assumed given except for an unknown parameter θ, are to be observed.
- The problem of interest is to use the observed values to estimate θ.

For example, the X_i's might be independent, exponential random variables each having the same unknown mean θ. In this case, the joint density function of the random variables would be given by

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n} \\ &= \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \frac{1}{\theta} e^{-\frac{x_2}{\theta}} \ldots \frac{1}{\theta} e^{-\frac{x_n}{\theta}}, \quad 0 < x_i < \infty, i = 1, \ldots, n \\ &= \frac{1}{\theta^n} exp\left\{ -\sum_{i=1}^{n} \frac{x_i}{\theta} \right\}, \quad 0 < x_i < \infty, i = 1, \ldots, n \end{aligned}$$

and the objective would be to estimate θ from the observed data $X_1, X_2, \ldots, X_n$.

- Let $f(x_1, \ldots x_n|\theta)$ denote the joint probability mass function of the random variables $X_1, X_2, \ldots, X_n$ when they are discrete, and let it be their joint probability density function when they are jointly continuous random variables.

- Let $f(x_1, \ldots x_n|\theta)$ denote the joint probability mass function of the random variables $X_1, X_2, \ldots, X_n$ when they are discrete, and let it be their joint probability density function when they are jointly continuous random variables.
- Because θ is assumed unknown, we also write f as a function of θ. Now since $f(x_1, \ldots, x_n|\theta)$ represents the likelihood that the values $x_1, x_2, \ldots, x_n$ will be observed when θ is the true value of the parameter, it would seem that a reasonable estimate of θ would be that value yielding the largest likelihood of the observed values.

- Let $f(x_1, \ldots x_n|\theta)$ denote the joint probability mass function of the random variables $X_1, X_2, \ldots, X_n$ when they are discrete, and let it be their joint probability density function when they are jointly continuous random variables.
- Because θ is assumed unknown, we also write f as a function of θ. Now since $f(x_1, \ldots, x_n|\theta)$ represents the likelihood that the values $x_1, x_2, \ldots, x_n$ will be observed when θ is the true value of the parameter, it would seem that a reasonable estimate of θ would be that value yielding the largest likelihood of the observed values.
- In other words, the maximum likelihood estimate $\hat{\theta}$ is defined to be that value of θ maximizing $f(x_1, \ldots, x_n|\theta)$ where $x_1, \ldots, x_n$ are the observed values. The function $f(x_1, \ldots, x_n|\theta)$ is often referred to as the likelihood function of θ.

Maximum Likelihood Estimator for Bernoulli Parameter

Suppose that n independent trials, each of which is a success with probability p, are performed.

The data consist of the values $X_1, \ldots, X_n$ where

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases}$$

Now

$$P\{X_1 = 1\} = p = 1 - P\{X_i = 0\}$$

which is

$$P\{X_i = x\} = p^x(1-p)^{1-x}, \quad x = 0, 1$$

Maximum Likelihood Estimator for Bernoulli Parameter

Suppose that n independent trials, each of which is a success with probability p, are performed.
The data consist of the values $X_1, \ldots, X_n$ where

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases}$$

Now

$$P\{X_1 = 1\} = p = 1 - P\{X_i = 0\}$$

which is

$$P\{X_i = x\} = p^x(1-p)^{1-x}, \quad x = 0, 1$$

Since each trial is independent, the likelihood or the joint probability mass function of data is

$$\begin{aligned} f(x_1, \ldots, x_n|p) &= P\{X_1 = x_1, \ldots, X_n = x_n|p\} \\ &= p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, \quad x_i = 0, 1, \quad i = 1, \ldots, n \end{aligned}$$

To determine the value of p that maximizes the likelihood, first take logs to obtain

$$log\ f(x_1, \ldots, x_n|p) = \sum_{i=1}^{n} x_i log(p) + \left(n - \sum_{i=1}^{n}\right) log(1-p)$$

To determine the value of p that maximizes the likelihood, first take logs to obtain

$$log\ f(x_1, \ldots, x_n|p) = \sum_{i=1}^{n} x_i log(p) + \left(n - \sum_{i=1}^{n}\right) log(1-p)$$

Differentiating and equating to zero

$$\frac{d}{dp} log\ f(x_1, \ldots, x_n|p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

$$\frac{\sum_{i=1}^{n} x_i}{\hat{p}} = \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-\hat{p}}$$

To determine the value of p that maximizes the likelihood, first take logs to obtain

$$log\ f(x_1, \ldots, x_n|p) = \sum_{i=1}^{n} x_i log(p) + \left(n - \sum_{i=1}^{n}\right) log(1-p)$$

Differentiating and equating to zero

$$\frac{d}{dp} log\ f(x_1, \ldots, x_n|p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

$$\frac{\sum_{i=1}^{n} x_i}{\hat{p}} = \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-\hat{p}}$$

Hence

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The maximum likelihood estimator of the unknown mean of a Bernoulli distribution is given by

$$d(X_1, \ldots, X_n) = \frac{\sum_{i=1}^{n} x_i}{n}$$

Example

Two proofreaders were given the same manuscript to read. If proofreader 1 found n_1 errors, and proofreader 2 found n_2 errors, with $n_{1,2}$ of these errors being found by both proofreaders, estimate N, the total number of errors that are in the manuscript.

Example

Two proofreaders were given the same manuscript to read. If proofreader 1 found n_1 errors, and proofreader 2 found n_2 errors, with $n_{1,2}$ of these errors being found by both proofreaders, estimate N, the total number of errors that are in the manuscript.

Solution: Let us assume that the results of the proofreaders are independent, and that each error in the manuscript is independently found by proofreader i with probability $p_i, i = 1, 2$.

Example

Two proofreaders were given the same manuscript to read. If proofreader 1 found n_1 errors, and proofreader 2 found n_2 errors, with $n_{1,2}$ of these errors being found by both proofreaders, estimate N, the total number of errors that are in the manuscript.

Solution: Let us assume that the results of the proofreaders are independent, and that each error in the manuscript is independently found by proofreader i with probability $p_i, i = 1, 2$.
To estimate N, we will start by deriving an estimator of p_1. To do so, note that each of the n_2 errors found by reader 2 will, independently, be found by proofreader 1 with probability p_i . Because proofreader 1 found $n_{1,2}$ of those n_2 errors, a reasonable estimate of p_1 is given by

$$\hat{p}_1 = \frac{n_{1,2}}{n_2}$$

Example

Two proofreaders were given the same manuscript to read. If proofreader 1 found n_1 errors, and proofreader 2 found n_2 errors, with $n_{1,2}$ of these errors being found by both proofreaders, estimate N, the total number of errors that are in the manuscript.

Solution: Let us assume that the results of the proofreaders are independent, and that each error in the manuscript is independently found by proofreader i with probability $p_i, i = 1, 2$.
To estimate N, we will start by deriving an estimator of p_1. To do so, note that each of the n_2 errors found by reader 2 will, independently, be found by proofreader 1 with probability p_i . Because proofreader 1 found $n_{1,2}$ of those n_2 errors, a reasonable estimate of p_1 is given by

$$\hat{p}_1 = \frac{n_{1,2}}{n_2}$$

However, because proofreader 1 found n_1 of the N errors in the manuscript, it is reasonable to suppose that p_1 is also approximately equal to $\frac{n_1}{N}$. Equating this to $\hat{p}_1$ gives that

$$\frac{n_{1,2}}{n_2} \approx \frac{n_1}{N} \quad \text{or} \quad N \approx \frac{n_1 n_2}{n_{1,2}}$$

Example 2: Estimating Fish Population

The Capture-Recapture Method

Marine biologists want to estimate the total number of salmon in a lake.

Day 1 - Capture:

- Catch and tag 150 salmon
- Release them back

Day 2 - Recapture (one week later):

- Catch 120 salmon
- 18 of them have tags

Fish Population: The Calculation

Logic:

- Proportion tagged in Day 2 sample $= \frac{18}{120} = 0.15 = 15\%$
- This should equal proportion tagged in entire lake
- We tagged 150 fish initially

MLE Formula:

$$\hat{N} = \frac{n_1 \times n_2}{n_{12}} = \frac{150 \times 120}{18} = 1000 \text{ salmon}$$

where:

- n_1 = number tagged on Day 1 (150)
- n_2 = number caught on Day 2 (120)
- n_{12} = number tagged in Day 2 sample (18)

Estimate

Total salmon population $\approx$ 1000 fish