Review of Statistics I

Hüseyin Taştan[1]

[1]Department of Economics
Yildiz Technical University

April 17, 2010

Review of Distribution Theory

- Random variables, discrete vs continuous
- Probability distribution functions
- Probability density function
- Expectation operator
- Normal distribution and related distributions: t-distribution, Chi-square distribution, F-distribution

Review of Distribution Theory

- Random variables, discrete vs continuous
- Probability distribution functions
- Probability density function
- Expectation operator
- Normal distribution and related distributions: t-distribution, Chi-square distribution, F-distribution

Review of Distribution Theory

- Random variables, discrete vs continuous
- Probability distribution functions
- Probability density function
- Expectation operator
- Normal distribution and related distributions: t-distribution, Chi-square distribution, F-distribution

Review of Distribution Theory

- Random variables, discrete vs continuous
- Probability distribution functions
- Probability density function
- Expectation operator
- Normal distribution and related distributions: t-distribution, Chi-square distribution, F-distribution

Review of Distribution Theory

- Random variables, discrete vs continuous
- Probability distribution functions
- Probability density function
- Expectation operator
- Normal distribution and related distributions: t-distribution, Chi-square distribution, F-distribution

Random Variables

- A random variable is a function from the sample space S to the real numbers. We may denote a random variable as $X(S)$ which is a usual, measurable function
- The value of a random variable cannot be known before the realization of event - uncertainty-
- Notation: use capital letters X, Y, or Z or any other letter to denote random variables
- Specific values of random variable will be denoted using letters x, y or z
- Using this notation we may calculate $P(X = x)$ for discrete random variables
- Two types of rv: discrete vs. continuous

Random Variables

- A random variable is a function from the sample space S to the real numbers. We may denote a random variable as $X(S)$ which is a usual, measurable function
- The value of a random variable cannot be known before the realization of event - uncertainty-
- Notation: use capital letters X, Y, or Z or any other letter to denote random variables
- Specific values of random variable will be denoted using letters x, y or z
- Using this notation we may calculate $P(X = x)$ for discrete random variables
- Two types of rv: discrete vs. continuous

Random Variables

- A random variable is a function from the sample space S to the real numbers. We may denote a random variable as $X(S)$ which is a usual, measurable function
- The value of a random variable cannot be known before the realization of event - uncertainty-
- Notation: use capital letters X, Y, or Z or any other letter to denote random variables
- Specific values of random variable will be denoted using letters x, y or z
- Using this notation we may calculate $P(X = x)$ for discrete random variables
- Two types of rv: discrete vs. continuous

Random Variables

- A random variable is a function from the sample space S to the real numbers. We may denote a random variable as $X(S)$ which is a usual, measurable function
- The value of a random variable cannot be known before the realization of event - uncertainty-
- Notation: use capital letters X, Y, or Z or any other letter to denote random variables
- Specific values of random variable will be denoted using letters x, y or z
- Using this notation we may calculate $P(X = x)$ for discrete random variables
- Two types of rv: discrete vs. continuous

Random Variables

- A random variable is a function from the sample space S to the real numbers. We may denote a random variable as $X(S)$ which is a usual, measurable function
- The value of a random variable cannot be known before the realization of event - uncertainty-
- Notation: use capital letters X, Y, or Z or any other letter to denote random variables
- Specific values of random variable will be denoted using letters x, y or z
- Using this notation we may calculate $P(X = x)$ for discrete random variables
- Two types of rv: discrete vs. continuous

Random Variables

- A random variable is a function from the sample space S to the real numbers. We may denote a random variable as $X(S)$ which is a usual, measurable function
- The value of a random variable cannot be known before the realization of event - uncertainty-
- Notation: use capital letters X, Y, or Z or any other letter to denote random variables
- Specific values of random variable will be denoted using letters x, y or z
- Using this notation we may calculate $P(X = x)$ for discrete random variables
- Two types of rv: discrete vs. continuous

Discrete Random Variable

- Can take a countable number of distinct possible values
- Whenever all possible values a random variable can assume can be listed (or counted) the random variable is discrete
- Example: the sum of two numbers on the top when two dices are rolled: $x = 2, 3, \ldots, 12$
- Example: the number of sales made by a salesperson per week: $x = 0, 1, 2, ...$
- Example: the number of errors made by an accountant

Discrete Random Variable

- Can take a countable number of distinct possible values
- Whenever all possible values a random variable can assume can be listed (or counted) the random variable is discrete
- Example: the sum of two numbers on the top when two dices are rolled: $x = 2, 3, \ldots, 12$
- Example: the number of sales made by a salesperson per week: $x = 0, 1, 2, ...$
- Example: the number of errors made by an accountant

Discrete Random Variable

- Can take a countable number of distinct possible values
- Whenever all possible values a random variable can assume can be listed (or counted) the random variable is discrete
- Example: the sum of two numbers on the top when two dices are rolled: $x = 2, 3, \ldots, 12$
- Example: the number of sales made by a salesperson per week: $x = 0, 1, 2, ...$
- Example: the number of errors made by an accountant

Discrete Random Variable

- Can take a countable number of distinct possible values
- Whenever all possible values a random variable can assume can be listed (or counted) the random variable is discrete
- Example: the sum of two numbers on the top when two dices are rolled: $x = 2, 3, \ldots, 12$
- Example: the number of sales made by a salesperson per week: $x = 0, 1, 2, ...$
- Example: the number of errors made by an accountant

Discrete Random Variable

- Can take a countable number of distinct possible values
- Whenever all possible values a random variable can assume can be listed (or counted) the random variable is discrete
- Example: the sum of two numbers on the top when two dices are rolled: $x = 2, 3, \ldots, 12$
- Example: the number of sales made by a salesperson per week: $x = 0, 1, 2, ...$
- Example: the number of errors made by an accountant

Continuous Random Variable

- Can take uncountable number of values within an interval on the real line
- Many variables from business and economics belong to this category
- Example: Average disposable income in a city
- Example: Closing value of a stock exchange index
- Example: Inflation rate for a given month

Continuous Random Variable

- Can take uncountable number of values within an interval on the real line
- Many variables from business and economics belong to this category
- Example: Average disposable income in a city
- Example: Closing value of a stock exchange index
- Example: Inflation rate for a given month

Continuous Random Variable

- Can take uncountable number of values within an interval on the real line
- Many variables from business and economics belong to this category
- Example: Average disposable income in a city
- Example: Closing value of a stock exchange index
- Example: Inflation rate for a given month

Continuous Random Variable

- Can take uncountable number of values within an interval on the real line
- Many variables from business and economics belong to this category
- Example: Average disposable income in a city
- Example: Closing value of a stock exchange index
- Example: Inflation rate for a given month

Continuous Random Variable

- Can take uncountable number of values within an interval on the real line
- Many variables from business and economics belong to this category
- Example: Average disposable income in a city
- Example: Closing value of a stock exchange index
- Example: Inflation rate for a given month

Probability Distributions for Discrete R.V.

Probability Distribution Function

$$f(x) \geq 0$$

$$f(x) = P(X = x)$$

$$\sum_x f(x) = 1$$

Cumulative Distribution Function

$$P(X \leq x_0) = F(x_0) = \sum_{x \leq x_0} f(x)$$

- The probability distribution of a discrete random variable is a formula, graph or table that specifies the probability associated with each possible value the random variable can assume

Example

- Three coins are tossed. Let X denote the number of Heads
- Sample space contains 8 points: (HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), ve (TTT)
- These 8 events are mutually exclusive and each has the same probability: $1/8$
- X can take four values: 0, 1, 2, 3

Example

- Three coins are tossed. Let X denote the number of Heads
- Sample space contains 8 points: (HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), ve (TTT)
- These 8 events are mutually exclusive and each has the same probability: $1/8$
- X can take four values: 0, 1, 2, 3

Example

- Three coins are tossed. Let X denote the number of Heads
- Sample space contains 8 points: (HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), ve (TTT)
- These 8 events are mutually exclusive and each has the same probability: $1/8$
- X can take four values: 0, 1, 2, 3

Example

- Three coins are tossed. Let X denote the number of Heads
- Sample space contains 8 points: (HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), ve (TTT)
- These 8 events are mutually exclusive and each has the same probability: $1/8$
- X can take four values: 0, 1, 2, 3

Example continued

Distribution of X

Results	x	$f(x)$
TTT	0	1/8
TTH	1	
THT	1	3/8
HTT	1	
THH	2	
HTH	2	3/8
HHT	2	
HHH	3	1/8

Example continued

Distribution Function of X

x	0	1	2	3
$f(x) = P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Note that

$$\sum_x f(x) = 1$$

Find

- $P(X \leq 1) = ?$
- $P(1 \leq X \leq 3) = ?$

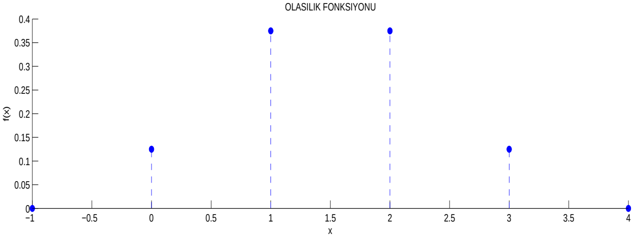
Example continued

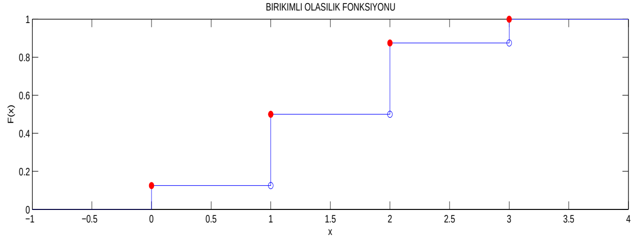
Cumulative Distribution Function of X

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0; \\ \frac{1}{8}, & 0 \leq x < 1; \\ \frac{1}{2}, & 1 \leq x < 2; \\ \frac{7}{8}, & 2 \leq x < 3; \\ 1, & x \geq 3. \end{cases}$$

Example continued

Distribution Function of X





Expectations of Discrete R.V.

- Expected Value of discrete r.v. X

$$E(X) = \sum_x x f(x)$$

- If $g(x)$ is a function of X then the expected value of $g(x)$ is

$$E(g(X)) = \sum_x g(x) f(x)$$

- If $g(x) = x^2$ then

$$E(g(X)) = E(X^2) = \sum_x x^2 f(x)$$

Expectations of Discrete R.V.

- Expected Value of discrete r.v. X

$$E(X) = \sum_x x f(x)$$

- If $g(x)$ is a function of X then the expected value of $g(x)$ is

$$E(g(X)) = \sum_x g(x) f(x)$$

- If $g(x) = x^2$ then

$$E(g(X)) = E(X^2) = \sum_x x^2 f(x)$$

- Expected Value of discrete r.v. X

$$E(X) = \sum_x x f(x)$$

- If $g(x)$ is a function of X then the expected value of $g(x)$ is

$$E(g(X)) = \sum_x g(x) f(x)$$

- If $g(x) = x^2$ then

$$E(g(X)) = E(X^2) = \sum_x x^2 f(x)$$

Example

- Let's find the expected value of X in the previous example

$$E(X) = 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{3}{2}$$

- Find the expected value of $g(x) = x^2$

$$E(X^2) = 0\frac{1}{8} + 1\frac{3}{8} + 4\frac{3}{8} + 9\frac{1}{8} = 3$$

- Find the expected value of $g(x) = 2x + 3x^2$

Example

- Let's find the expected value of X in the previous example

$$E(X) = 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{3}{2}$$

- Find the expected value of $g(x) = x^2$

$$E(X^2) = 0\frac{1}{8} + 1\frac{3}{8} + 4\frac{3}{8} + 9\frac{1}{8} = 3$$

- Find the expected value of $g(x) = 2x + 3x^2$

Example

- Let's find the expected value of X in the previous example

$$E(X) = 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{3}{2}$$

- Find the expected value of $g(x) = x^2$

$$E(X^2) = 0\frac{1}{8} + 1\frac{3}{8} + 4\frac{3}{8} + 9\frac{1}{8} = 3$$

- Find the expected value of $g(x) = 2x + 3x^2$

Variance of a Discrete RV

- Definition:

$$\begin{aligned} Var(X) &= E\left[(X - E(X))^2\right] \\ &= E\left[(X^2 - 2XE(X) + (E(X))^2)\right] \\ &= E\left(X^2\right) - 2E\left(XE(X)\right) + E\left((E(X))^2\right) \\ &= E\left(X^2\right) - 2E(X)^2 + (E(X))^2 \\ &= E\left(X^2\right) - E(X)^2 \end{aligned}$$

- Let $\mu_x = E(X)$ then the variance can be written as

$$Var(X) = E\left(X^2\right) - \mu_x^2$$

Variance of a Discrete RV

- Definition:

$$
\begin{aligned}
Var(X) &= E\left[(X - E(X))^2\right] \\
&= E\left[(X^2 - 2XE(X) + (E(X))^2)\right] \\
&= E\left(X^2\right) - 2E\left(XE(X)\right) + E\left((E(X))^2\right) \\
&= E\left(X^2\right) - 2E(X)^2 + (E(X))^2 \\
&= E\left(X^2\right) - E(X)^2
\end{aligned}
$$

- Let $\mu_x = E(X)$ then the variance can be written as

$$Var(X) = E\left(X^2\right) - \mu_x^2$$

Moments of a Discrete RV

- Definiton: kth moment of a discrete rv is defined as

$$\mu_k = E(X^k) = \sum_x x^k f(x) \quad k = 0, 1, 2, \ldots$$

$$\begin{array}{llcll}
\text{1. moment} & \mu_1 & = & E(X) & \Longrightarrow \text{population mean} \\
\text{2. moment} & \mu_2 & = & E(X^2) & = Var(X) + \mu_1^2 \\
\text{3. moment} & \mu_3 & = & E(X^3) & \\
\text{4. moment} & \mu_4 & = & E(X^4) &
\end{array}$$

Central Moments of a Discrete RV

- Definiton: kth central moment of a discrete rv is defined as

$$m_k = E((X-\mu_1)^k) = \sum_x (x-\mu_1)^k f(x) \quad k = 0, 1, 2, \ldots$$

$$\begin{array}{llcll}
\text{1. central moment} & m_1 & = & 0 & \\
\text{2. central moment} & m_2 & = & E((X-\mu_1)^2) & = Var(X) \\
\text{3. central moment} & m_3 & = & E((X-\mu_1)^3) & \\
\text{4. central moment} & m_4 & = & E((X-\mu_1)^4) &
\end{array}$$

Standard Moments of a Discrete RV

- Definiton: kth standard moment of a discrete rv is defined as

$$\gamma_k = \frac{m_k}{\sigma^k} \quad k = 0, 1, 2, \ldots$$

σ is the population standard deviation:

$$\sigma = \sqrt{Var(X)} = \sqrt{E\left[(X - \mu_1)^2\right]}$$

1. standart moment $\gamma_1 = 0$
2. standart moment $\gamma_2 = 1$ why?
3. standart moment $\gamma_3 = \frac{m_3}{\sigma^3}$ skewness
4. standart moment $\gamma_4 = \frac{m_4}{\sigma^4}$ kurtosis

Some Discrete Distributions

- Bernoulli
- Binomial
- Hypergeometric
- Poisson

Some Discrete Distributions

- Bernoulli
- Binomial
- Hypergeometric
- Poisson

Some Discrete Distributions

- Bernoulli
- Binomial
- Hypergeometric
- Poisson

Some Discrete Distributions

- Bernoulli
- Binomial
- Hypergeometric
- Poisson

Bernoulli(p) Distribution

$$f(x) = \begin{cases} p, & \text{if } X = 1 \\ 1 - p, & \text{if } X = 0 \end{cases}$$

Expected Value:

$$E(X) = \sum_x x f(x) = p \cdot 1 + (1 - p) \cdot 0 = p$$

Second Moment:

$$E(X^2) = \sum_x x^2 f(x) = p \cdot 1 + (1 - p) \cdot 0 = p$$

Variance (second central moment):

$$Var(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p)$$

Binomial Distribution

Let X be the number of successes in an n independent trials of Bernoulli random experiment. If Y is distributed as Bernoulli(p) then $X = \sum Y$ has a Binomial(n, p) distribution. X denotes total number of successes.

$$f(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x(1-p)^{n-x} = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

$$E(X) = np$$

$$Var(X) = np(1-p)$$

If Bernoulli trials are not independent then the total number of successes does not have a Binomial distribution but follows hypergeometric distribution. In an N-element population containing B successes total number of successes X in a random sample n has the following distribution

$$f(x) = \frac{\left(\begin{array}{c} B \\ x \end{array}\right)\left(\begin{array}{c} N-B \\ n-x \end{array}\right)}{\left(\begin{array}{c} N \\ n \end{array}\right)}$$

Here x can take integer values between $\max(0, n-(N-B))$ and $\min(n, B)$

$$E(X) = np, \quad Var(X) = \frac{N-n}{N-1}np(1-p), \quad p = \frac{B}{N}$$

Poisson Distribution

Describes the distribution of the number of realizations of a certain event in a given period of time

Notation: $X \sim Poisson(\lambda)$

Probability DF:

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

$$skewness = \frac{1}{\sqrt{\lambda}}$$

$$excess\ kurtosis = \frac{1}{\lambda}$$

Let X and Y be two discrete rv. Then the joint pd is

$$f(x, y) = P(X = x \cap Y = y)$$

More generally, joint pd of k discrete rv denoted $X_1, X_2, \ldots, X_k$ is

$$f(x_1, x_2, \ldots, x_k) = P(X_1 = x_1 \cap X_2 = x_2, \cap, \ldots, \cap X_k = x_k)$$

Example

X: number of customers waiting in line for Counter 1 in a bank,
Y: number of customers waiting in line for Counter 2 in a bank.
Joint pd for these two rv is given by

Joint probability distribution of X and Y

$y \setminus x$	0	1	2	3	Total
0	0.05	0.21	0	0	0.26
1	0.20	0.26	0.08	0	0.54
2	0	0.06	0.07	0.02	0.15
3	0	0	0.03	0.02	0.05
Total	0.25	0.53	0.18	0.04	1.00

Marginal Distribution Function

Marginal DF for X:

$$f(x) = \sum_y f(x, y)$$

Marginal DF Y:

$$f(y) = \sum_x f(x, y)$$

Conditional Distribution Function

Conditional DF for X given $Y = y$:

$$f(x|y) = \frac{f(x,y)}{f(y)}$$

Conditional DF for Y given $X = x$:

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

Statistical Independence

Discrete rvs X and Y are said to be **independent if and only if**

$$f(x,y) = f(x)f(y)$$

or, in other words

$$f(x|y) = f(x),$$

and

$$f(y|x) = f(y)$$

Covariance between two discrete rv

Let $g(X, Y)$ be a function of X and Y. The expected value of this function is:

$$E\left[g(X,Y)\right] = \sum_x \sum_y g(x,y) f(x,y)$$

Now let $g(X, Y) = (X - \mu_x)(Y - \mu_y)$. The expected value of this function is called **covariance**:

$$Cov\left(X,Y\right) = \sum_x \sum_y (x - \mu_x)(y - \mu_y) f(x,y)$$

It measures the linear association between two random variables. Statistical independence implies that the covariance is zero but the reverse is not true.

- A continuous rv can assume any value within some interval.
- Probability distribution for a continuous rv is a smooth function $f(x)$ called probability density function (pdf).
- The probability associated with a certain value of X is 0.

Properties of pdf $f(x)$:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $Pr(a < X < b) = \int_a^b f(x)dx$

Probability Distributions for Continuous Random Variables

- A continuous rv can assume any value within some interval.
- Probability distribution for a continuous rv is a smooth function $f(x)$ called probability density function (pdf).
- The probability associated with a certain value of X is 0.

Properties of pdf $f(x)$:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $Pr(a < X < b) = \int_a^b f(x)dx$

- A continuous rv can assume any value within some interval.
- Probability distribution for a continuous rv is a smooth function $f(x)$ called probability density function (pdf).
- The probability associated with a certain value of X is 0.

Properties of pdf $f(x)$:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $Pr(a < X < b) = \int_a^b f(x)dx$

Probability Distributions for Continuous Random Variables

- A continuous rv can assume any value within some interval.
- Probability distribution for a continuous rv is a smooth function $f(x)$ called probability density function (pdf).
- The probability associated with a certain value of X is 0.

Properties of pdf $f(x)$:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $Pr(a < X < b) = \int_a^b f(x)dx$

Probability Distributions for Continuous Random Variables

- A continuous rv can assume any value within some interval.
- Probability distribution for a continuous rv is a smooth function $f(x)$ called probability density function (pdf).
- The probability associated with a certain value of X is 0.

Properties of pdf $f(x)$:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $Pr(a < X < b) = \int_a^b f(x)dx$

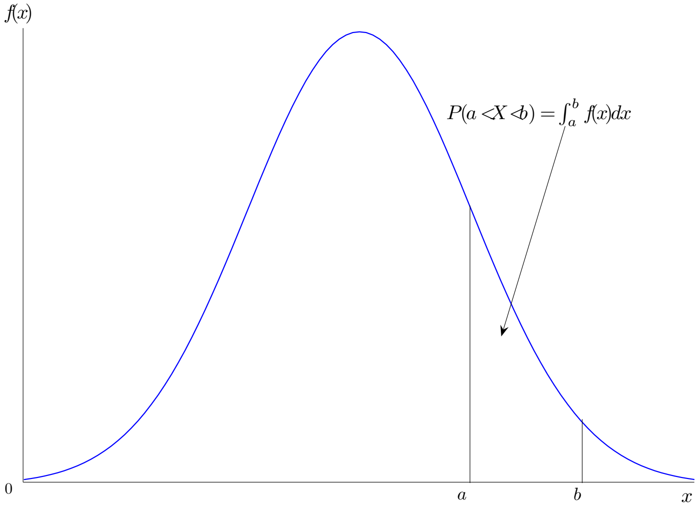
Probability Distributions for Continuous Random Variables

- A continuous rv can assume any value within some interval.
- Probability distribution for a continuous rv is a smooth function $f(x)$ called probability density function (pdf).
- The probability associated with a certain value of X is 0.

Properties of pdf $f(x)$:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $Pr(a < X < b) = \int_a^b f(x)dx$

A Probability Density Function



Cumulative Density Function

- Cumulative density function (cdf), denoted $F(x)$, shows the probability that a random variable does not exceed a given value x
- Definition:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

- From this definition cdf and pdf are related with

$$f(x) = \frac{dF(x)}{dx}$$

Cumulative Density Function

- Cumulative density function (cdf), denoted $F(x)$, shows the probability that a random variable does not exceed a given value x
- Definition:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

- From this definition cdf and pdf are related with

$$f(x) = \frac{dF(x)}{dx}$$

Cumulative Density Function

- Cumulative density function (cdf), denoted $F(x)$, shows the probability that a random variable does not exceed a given value x
- Definition:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

- From this definition cdf and pdf are related with

$$f(x) = \frac{dF(x)}{dx}$$

Properties of CDF

- $$F(-\infty) = 0, \quad F(+\infty) = 1$$

  $F(x)$ is a nondecreasing function of x: $x_1 \leq x_2$, $F(x_1) \leq F(x_2)$.

- $$P(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx$$

- $$P(-\infty < X < +\infty) = P(-\infty < X < a) + P(a < X < b) + P(b < X < +\infty)$$

  $$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_a^b f(x)dx + \int_b^{+\infty} f(x)dx$$

- $$F(+\infty) - F(-\infty) = [F(a) - F(-\infty)] + P(a < X < b) + [F(+\infty) - F(b)]$$

  $$1 = F(a) - 0 + P(a < X < b) + 1 - F(b)$$

  $$P(a < X < b) = F(b) - F(a)$$

Properties of CDF

- $$F(-\infty) = 0, \quad F(+\infty) = 1$$

  $F(x)$ is a nondecreasing function of x: $x_1 \le x_2$, $F(x_1) \le F(x_2)$.

- $$P(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx$$

- $$P(-\infty < X < +\infty) = P(-\infty < X < a) + P(a < X < b) + P(b < X < +\infty)$$

  $$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_a^b f(x)dx + \int_b^{+\infty} f(x)dx$$

- $$F(+\infty) - F(-\infty) = [F(a) - F(-\infty)] + P(a < X < b) + [F(+\infty) - F(b)]$$

  $$1 = F(a) - 0 + P(a < X < b) + 1 - F(b)$$

  $$P(a < X < b) = F(b) - F(a)$$

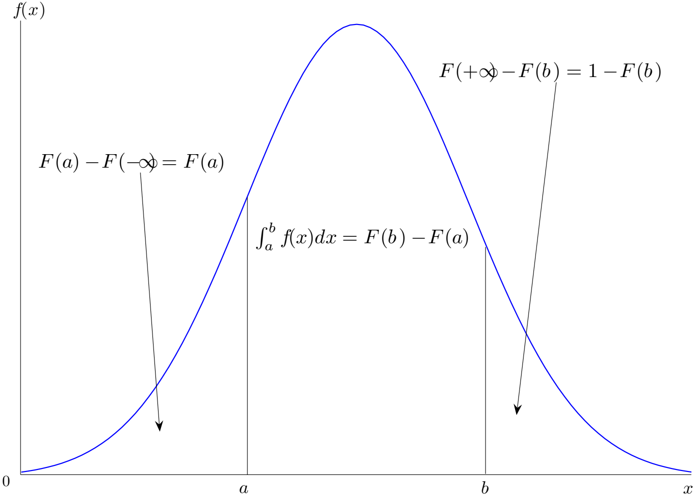
Properties of CDF

- $$F(-\infty) = 0, \quad F(+\infty) = 1$$
  $F(x)$ is a nondecreasing function of x: $x_1 \leq x_2$, $F(x_1) \leq F(x_2)$.
- $$P(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx$$
- $$P(-\infty < X < +\infty) = P(-\infty < X < a) + P(a < X < b) + P(b < X < +\infty)$$
  $$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_a^b f(x)dx + \int_b^{+\infty} f(x)dx$$
- $$F(+\infty) - F(-\infty) = [F(a) - F(-\infty)] + P(a < X < b) + [F(+\infty) - F(b)]$$
  $$1 = F(a) - 0 + P(a < X < b) + 1 - F(b)$$
  $$P(a < X < b) = F(b) - F(a)$$

Properties of CDF

- $$F(-\infty) = 0, \quad F(+\infty) = 1$$
  $F(x)$ is a nondecreasing function of x: $x_1 \leq x_2$, $F(x_1) \leq F(x_2)$.
- $$P(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx$$
- $$P(-\infty < X < +\infty) = P(-\infty < X < a) + P(a < X < b) + P(b < X < +\infty)$$
  $$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_a^b f(x)dx + \int_b^{+\infty} f(x)dx$$
- $$F(+\infty) - F(-\infty) = [F(a) - F(-\infty)] + P(a < X < b) + [F(+\infty) - F(b)]$$
  $$1 = F(a) - 0 + P(a < X < b) + 1 - F(b)$$
  $$P(a < X < b) = F(b) - F(a)$$

CDF



Expectations of Continuous Random Variables

$$E(X) \equiv \mu_x = \int_{-\infty}^{\infty} x f(x) dx$$

If $g(x)$ is a continuous function of X, then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx$$

Variance of Continuous RV

$$
\begin{aligned}
Var(X) &= E\left[(X - E(X))^2\right] \equiv \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx + \mu_x^2 \int_{-\infty}^{\infty} f(x)dx - 2\mu_x \int_{-\infty}^{\infty} xf(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} xf(x)dx\right)^2 \\
&= E(X^2) - \mu_x^2
\end{aligned}
$$

Note that we used $\int_{-\infty}^{\infty} f(x)dx = 1$ and $\int_{-\infty}^{\infty} xf(x)dx = E(X) \equiv \mu_x$

Moments of Continuous RV

Definition: kth moment of a continuous rv X is given by

$$\mu_k = E(X^k) = \int_{x \in \mathcal{X}} x^k f(x) dx \quad k = 0, 1, 2, \dots$$

$$\begin{array}{lllll}
1.\ \text{moment} & \mu_1 & = & E(X) & \Longrightarrow \text{population mean} \\
2.\ \text{moment} & \mu_2 & = & E(X^2) & \quad = Var(X) + \mu_1^2 \\
3.\ \text{moment} & \mu_3 & = & E(X^3) & \\
4.\ \text{moment} & \mu_4 & = & E(X^4) &
\end{array}$$

Definition: kth central moment of X is

$$m_k = E((X-\mu_1)^k) = \int_{x \in \mathcal{X}} (x-\mu_1)^k f(x)dx \quad k = 0, 1, 2, \ldots$$

1. central moment $\quad m_1 = 0$
2. central moment $\quad m_2 = E((X-\mu_1)^2) = Var(X)$
3. central moment $\quad m_3 = E((X-\mu_1)^3)$
4. central moment $\quad m_4 = E((X-\mu_1)^4)$

Properties of Expectation Operator

- Linearity: Let $Y = a + bX$ then the expected value of Y is:

$$E[Y] = E[a + bX] = a + bE(X)$$

- More generally let $X_1, X_2, \ldots, X_n$ be n continuous rvs and let Y be a linear combination of these rvs:

$$Y = b_1X_1 + b_nX_2 + \ldots + b_nX_n$$

Expected value of Y is given by:

$$E[Y] = b_1E[X_1] + b_2E[X_2] + \ldots + b_nE[X_n]$$

or more compactly

$$E(Y) = E\left(\sum_{i=1}^{n} b_iX_i\right) = \sum_{i=1}^{n} b_iE(X_i)$$

Properties of Expectation Operator

- Linearity: Let $Y = a + bX$ then the expected value of Y is:

$$E[Y] = E[a + bX] = a + bE(X)$$

- More generally let $X_1, X_2, \ldots, X_n$ be n continuous rvs and let Y be a linear combination of these rvs:

$$Y = b_1 X_1 + b_n X_2 + \ldots + b_n X_n$$

Expected value of Y is given by:

$$E[Y] = b_1 E[X_1] + b_2 E[X_2] + \ldots + b_n E[X_n]$$

or more compactly

$$E(Y) = E\left(\sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} b_i E(X_i)$$

Properties of Expectation Operator

- For a nonlinear function of X we have (in general)

$$E[h(X)] \neq h(E(X))$$

- For example, $E(X^2) \neq (E(X))^2$, $E(\ln(X)) \neq \ln(E(X))$
- For two continuous rvs X and Y

$$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$$

Properties of Expectation Operator

- For a nonlinear function of X we have (in general)

$$E[h(X)] \neq h(E(X))$$

- For example, $E(X^2) \neq (E(X))^2$, $E(\ln(X)) \neq \ln(E(X))$
- For two continuous rvs X and Y

$$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$$

Properties of Expectation Operator

- For a nonlinear function of X we have (in general)

$$E[h(X)] \neq h(E(X))$$

- For example, $E(X^2) \neq (E(X))^2$, $E(\ln(X)) \neq \ln(E(X))$
- For two continuous rvs X and Y

$$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$$

Properties of Variance

- Let c be any constant then

$$Var(c) = 0$$

- Variance of $Y = bX$ where b is a constant

$$Var(Y) = Var(bX) = b^2 Var(X)$$

- Variance of $Y = a + bX$

$$Var(Y) = Var(a + bX) = b^2 Var(X)$$

- For two independent continuous rvs X and Y

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

This can easily be generalized to n independent rvs.

- Let c be any constant then

$$Var(c) = 0$$

- Variance of $Y = bX$ where b is a constant

$$Var(Y) = Var(bX) = b^2 Var(X)$$

- Variance of $Y = a + bX$

$$Var(Y) = Var(a + bX) = b^2 Var(X)$$

- For two independent continuous rvs X and Y

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

This can easily be generalized to n independent rvs.

Properties of Variance

- Let c be any constant then

$$Var(c) = 0$$

- Variance of $Y = bX$ where b is a constant

$$Var(Y) = Var(bX) = b^2 Var(X)$$

- Variance of $Y = a + bX$

$$Var(Y) = Var(a + bX) = b^2 Var(X)$$

- For two independent continuous rvs X and Y

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

This can easily be generalized to n independent rvs.

Properties of Variance

- Let c be any constant then

$$Var(c) = 0$$

- Variance of $Y = bX$ where b is a constant

$$Var(Y) = Var(bX) = b^2 Var(X)$$

- Variance of $Y = a + bX$

$$Var(Y) = Var(a + bX) = b^2 Var(X)$$

- For two independent continuous rvs X and Y

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

This can easily be generalized to n independent rvs.

Notation: $X \sim U(0,1)$, pdf:

$$f(x) = \begin{cases} 1, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

$$E(X) = \frac{1}{2}$$

$$Var(X) = \frac{1}{12}$$

General Uniform Distribution

Notation: $X \sim U(a,b)$, pdf:

$$f(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0, & \text{otherwise.} \end{cases}$$

$$E(X) = \frac{b+a}{2}$$

$$Median = \frac{b+a}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

$$Skewness = 0$$

$$Excess\ kurtosis = -\frac{6}{5}$$

General Uniform Distribution

Expectation of $X \sim U(a, b)$:

$$\begin{aligned} E(X) &= \int_a^b \frac{x}{b-a} dx \\ &= \frac{1}{b-a} \left[\frac{b^2 - a^2}{2} \right] \\ &= \frac{(b-a)(b+a)}{2(b-a)} \\ &= \frac{a+b}{2} \end{aligned}$$

General Uniform Distribution

$X \sim U(a,b)$ Expected value of $g(x) = x^2$

$$
\begin{aligned}
E[g(x)] &= \int_a^b x^2 \frac{1}{b-a} \\
&= \frac{b^3 - a^3}{3(b-a)} \\
&= \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} \\
&= \frac{a^2 + ab + b^2}{3} = E[X^2].
\end{aligned}
$$

Variance of $X \sim U(a, b)$:

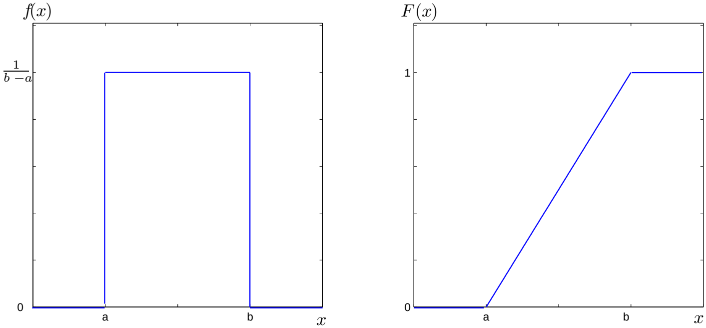
$$\begin{aligned} Var(X) &= E[(X - E(X))^2] = E(X^2) - [E(X)]^2 \\ &= \frac{(a^2 + ab + b^2)}{3} - \frac{(a+b)^2}{4} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Cumulative density function (cdf) of $U \sim (a, b)$:

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \int_a^x \frac{1}{b-a} dt \\ &= \left. \frac{t}{b-a} \right|_a^x \\ &= \frac{x-a}{b-a}, \quad \text{for } a \leq x \leq b \end{aligned}$$

$$F(x) = \begin{cases} 0, & \text{for } x < a; \\ \frac{x-a}{b-a}, & \text{for } a \leq x \leq b; \\ 1, & \text{for } x > b. \end{cases}$$

Uniform Distribution



Example

$$f(x) = \begin{cases} e^{-x}, & 0 < x < \infty \text{ ise;} \\ 0, & \text{deilse.} \end{cases}$$

1. Show that this function is a pdf.
2. Draw the graph of this function and mark the area for $X > 1$.
3. Calculate probability: $P(X > 1)$.
4. Find cdf.

Example

$$f(x) = \begin{cases} e^{-x}, & 0 < x < \infty \text{ ise;} \\ 0, & \text{deilse.} \end{cases}$$

1. Show that this function is a pdf.
2. Draw the graph of this function and mark the area for $X > 1$.
3. Calculate probability: $P(X > 1)$.
4. Find cdf.

Example

$$f(x) = \begin{cases} e^{-x}, & 0 < x < \infty \text{ ise;} \\ 0, & \text{deilse.} \end{cases}$$

1. Show that this function is a pdf.
2. Draw the graph of this function and mark the area for $X > 1$.
3. Calculate probability: $P(X > 1)$.
4. Find cdf.

Example

$$f(x) = \begin{cases} e^{-x}, & 0 < x < \infty \text{ ise;} \\ 0, & \text{deilse.} \end{cases}$$

1. Show that this function is a pdf.
2. Draw the graph of this function and mark the area for $X > 1$.
3. Calculate probability: $P(X > 1)$.
4. Find cdf.

Example continued

1. Let us see if the conditions for pdf are satisfied:
   1. (i) First of all, $f(x) \geq 0$ condition is satisfied for all values within the interval $0 < x < \infty$.
   2. (ii) Should integrate to 1:

$$
\begin{aligned}
\int_0^\infty e^{-x} dx &= 1 \\
-e^{-x}\Big|_0^\infty &= 1 \\
-e^{-\infty} - (-e^0) &= 1 \\
0 + 1 &= 1
\end{aligned}
$$

$-e^{-\infty} = \lim_{x\to\infty} -e^{-x} = 0$

Example continued

1. Let us see if the conditions for pdf are satisfied:
    1. (i) First of all, $f(x) \geq 0$ condition is satisfied for all values within the interval $0 < x < \infty$.
    2. (ii) Should integrate to 1:

$$
\begin{aligned}
\int_0^\infty e^{-x} dx &= 1 \\
-e^{-x}\Big|_0^\infty &= 1 \\
-e^{-\infty} - (-e^0) &= 1 \\
0 + 1 &= 1
\end{aligned}
$$

$-e^{-\infty} = \lim_{x \to \infty} -e^{-x} = 0$

Example continued

1. Let us see if the conditions for pdf are satisfied:
   1. (i) First of all, $f(x) \geq 0$ condition is satisfied for all values within the interval $0 < x < \infty$.
   2. (ii) Should integrate to 1:

$$\begin{aligned} \int_0^\infty e^{-x} dx &= 1 \\ -e^{-x}\Big|_0^\infty &= 1 \\ -e^{-\infty} - (-e^0) &= 1 \\ 0 + 1 &= 1 \end{aligned}$$

$-e^{-\infty} = \lim_{x \to \infty} -e^{-x} = 0$

1. $P(X > 1)$ area shown on the figure.
2. $$\begin{aligned} P(X>1) &= \int_1^\infty e^{-x}dx \\ &= -e^{-x}\Big|_1^\infty \\ &= e^{-1} \\ &\approx 0.36787 \end{aligned}$$
3. $$\begin{aligned} F(x) &= \int_0^x e^{-t}dt \\ &= -e^{-t}\Big|_0^x \\ &= -e^{-x}+e^0 \\ &= 1-e^{-x} \end{aligned}$$

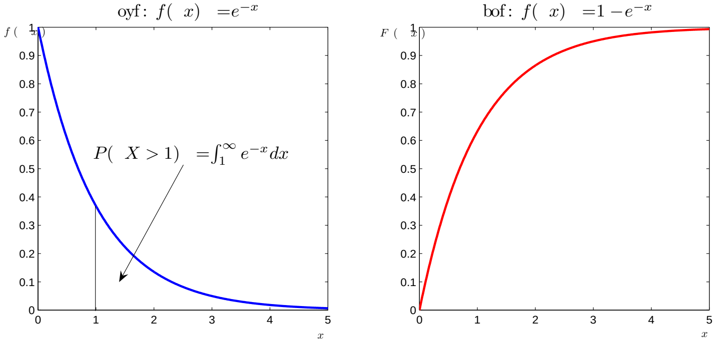
1. $P(X > 1)$ area shown on the figure.
2. $$\begin{aligned} P(X > 1) &= \int_1^\infty e^{-x} dx \\ &= -e^{-x}\Big|_1^\infty \\ &= e^{-1} \\ &\approx 0.36787 \end{aligned}$$
3. $$\begin{aligned} F(x) &= \int_0^x e^{-t} dt \\ &= -e^{-t}\Big|_0^x \\ &= -e^{-x} + e^0 \\ &= 1 - e^{-x} \end{aligned}$$

Example continued

1. $P(X > 1)$ area shown on the figure.
2. $$\begin{aligned} P(X > 1) &= \int_1^\infty e^{-x} dx \\ &= -e^{-x}\Big|_1^\infty \\ &= e^{-1} \\ &\approx 0.36787 \end{aligned}$$
3. $$\begin{aligned} F(x) &= \int_0^x e^{-t} dt \\ &= -e^{-t}\Big|_0^x \\ &= -e^{-x} + e^0 \\ &= 1 - e^{-x} \end{aligned}$$

CDF is

$$F(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-x}, & 0 < x < \infty. \end{cases}$$



Joint Probability Density Function

Let X and Y be two continuous random variables defined within the intervals $-\infty < X < +\infty$ and $-\infty < Y < +\infty$. The joint pdf for X and Y, denoted $f(x, y)$ is defined as

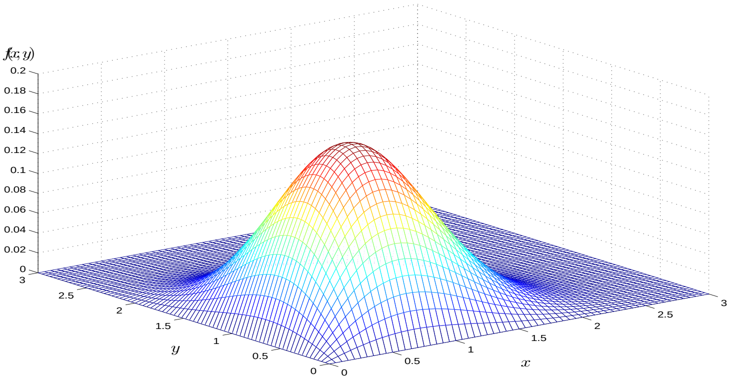
$$f(x,y) \geq 0,$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1,$$

$$Pr(a < X < b, c < Y < d) = \int_{c}^{d} \int_{a}^{b} f(x,y)dxdy.$$

Joint Probability Density Function: Example

Graph of joint density $f(x, y) = xye^{-(x^2+y^2)}$, $x > 0$, $y > 0$,



Example

$$f(x,y) = \begin{cases} k(x+y), & 0 < x < 1,\ 0 < y < 2 \text{ ise;} \\ 0, & \text{deilse.} \end{cases}$$

Find the constant k

Find $P\left(0 < X < \frac{1}{2},\ 1 < Y < 2\right)$.

Example continued

First for $f(x, y) > 0$ we must have $k > 0$. From the second condition we have

$$\begin{aligned}\int_0^2 \int_0^1 k(x+y)dxdy &= 1 \\ &= k\int_0^2 \left(\frac{1}{2} + y\right) dy = k\left(\frac{1}{2}y + \frac{y^2}{2}\right)\Bigg|_0^2 \\ &= 3k = 1\end{aligned}$$

Thus $k = \frac{1}{3}$. The joint pdf is

$$f(x, y) = \begin{cases} \frac{1}{3}(x+y), & 0 < x < 1,\ 0 < y < 2 \text{ ise;} \\ 0, & \text{deilse.}\end{cases}$$

Example continued

Probability is found as a volume measure:

$$
\begin{aligned}
P\left(0 < X < \frac{1}{2},\ 1 < Y < 2\right) &= \int_1^2 \int_0^{\frac{1}{2}} \frac{1}{3}(x+y)\,dx dy \\
&= \frac{1}{3}\int_1^2 \left(\frac{1}{8} + \frac{1}{2}y\right) dy \\
&= \frac{1}{3}\left(\frac{1}{8}y + \frac{y^2}{4}\right)\Bigg|_1^2 \\
&= \frac{7}{24}
\end{aligned}
$$

Marginal Density Function

Definition: marginal pdf of X:

$$f(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Bounds on the interval is the interval of Y.

Marginal pdf of Y:

$$f(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Bounds on the interval is the interval of X.

Example

Using the joint pdf from the previous example find the marginal pdfs of X and Y.

$$\begin{aligned} f(x) &= \int_0^2 \frac{1}{3}(x+y)dy \\ &= \frac{1}{3}\left(xy + \frac{y^2}{2}\right)\Big|_0^2 \\ &= \frac{2}{3}(x+1) \end{aligned}$$

Example continued

Marginal pdf of X :

$$f(x) = \begin{cases} \frac{2}{3}(x+1), & 0 < x < 1 \ ; \\ 0, & \text{otherwise.} \end{cases}$$

Marginal pdf of Y:

$$g(y) = \begin{cases} \frac{1}{3}(y+\frac{1}{2}), & 0 < y < 2 \ ; \\ 0, & \text{otherwise.} \end{cases}$$

Conditional Density Function

Given $Y = y$ the conditional density of X is defined as :

$$f(x|y) = \frac{f(x,y)}{f(y)}$$

Similarly given $X = x$ the conditional density of Y is defined as

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

Independence

Recall that two events A and B are independent if only if:

$$P(A \cap B) = P(A)P(B)$$

Similarly X and Y are independent if and only if the following condition is satisfied:

$$f(x, y) = f(x)f(y)$$

In other words if the joint density function can be written as the product of marginal densities then the two random variables are statistically independent.

More generally let $X_1, X_2, \ldots, X_n$ be n continuous random variables. If the joint density can be written as the product of marginal densities then these random variables are independent:

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= f_1(x_1) \cdot f_2(x_2) \cdot, \ldots, \cdot f_n(x_n) \\ &= \prod_{j=1}^{n} f_j(x_j) \end{aligned}$$

This property is useful for obtaining Maximum Likelihood estimators for certain parameters.

Independence: Example

Determine if the random variables are independent in the previous example.

$$\begin{aligned} f(x)g(x) &= \frac{2}{3}(x+1)\frac{1}{3}(y+\frac{1}{2}) \\ &\neq f(x,y) \end{aligned}$$

Thus X and Y are not independent.

Independence: Another Example

Determine if X and Y are independent using the following joint pdf:

$$f(x, y) = \begin{cases} \frac{1}{9}, & \text{for } 1 < x < 4,\ 1 < y < 4; \\ 0, & \text{otherwise.} \end{cases}$$

Marginal densities are

$$f(x) = \int_1^4 \frac{1}{9} dy = \frac{1}{3}$$

$$g(y) = \int_1^4 \frac{1}{9} dx = \frac{1}{3}$$

Thus

$$f(x, y) = \frac{1}{9} = f(x)g(y) = \left(\frac{1}{3}\right)\left(\frac{1}{3}\right)$$

Therefore X and Y are statistically independent.

Normal Distribution

Notation: $X \sim N(\mu, \sigma^2)$. pdf is given by

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \quad -\infty < x < \infty$$

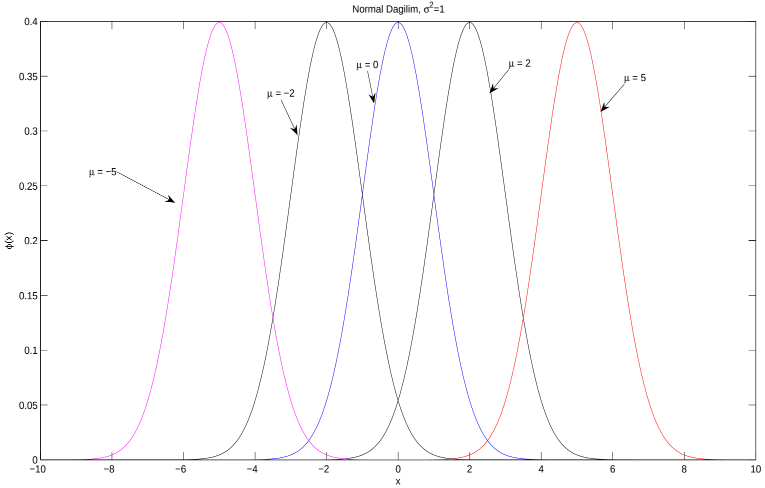
$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

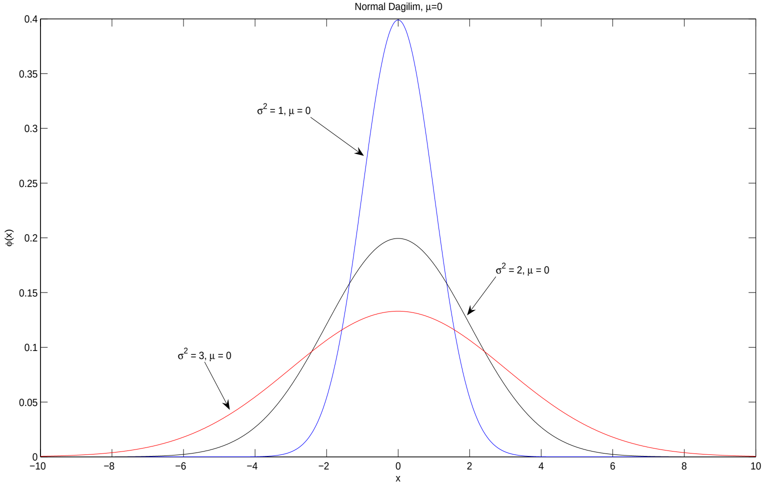
$$skewness = 0$$

$$kurtosis = 3$$

Normal Distribution, $\sigma^2 = 1$ Different location parameters



Normal Distribution, $\mu = o$ Different scale parameters



Define $Z = \frac{X-\mu}{\sigma}$, pdf is given by

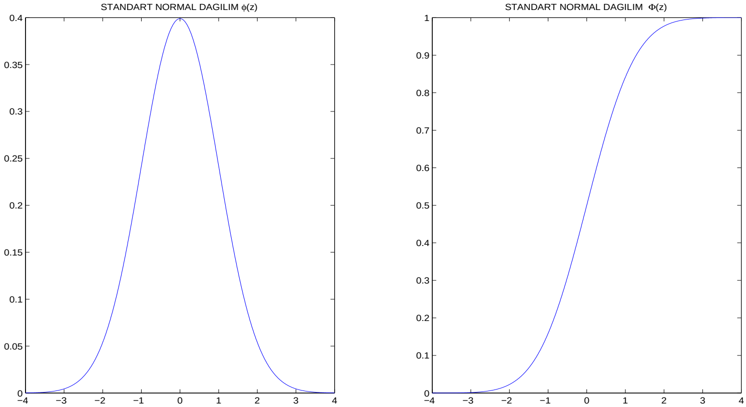
$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right), \quad -\infty < z < \infty$$

$$E(Z) = 0$$

$$Var(Z) = 1$$

CDF is:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt$$


STANDART NORMAL DAGILIM ϕ(z)
STANDART NORMAL DAGILIM Φ(z)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−4
−3
−2
−1
0
1
2
3
4

Let $X \sim N(\mu, \sigma^2)$. We want to calculate the following probability:

$$P(a < X < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx$$

There is no closed-form solution to this integral. It can only be calculated using numerical methods. This requires using computational methods each time we need to evaluate a probability. Instead of this cumbersome method we can use normal distribution tables. Let us write the desired probability as follows:

$$P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

Calculating Normal Distribution Probabilities

$$P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Here $\Phi(z) = P(Z \leq z)$ is the value of standard normal cdf at z. This can easily found by using standard normal tables.

Calculating Normal Distribution Probabilities

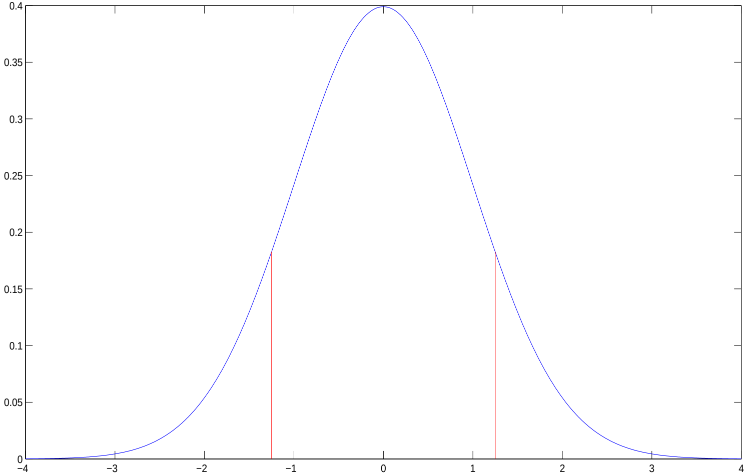
Since the standard normal distribution is symmetric only positive values are listed in the tables. Using the symmetry property one can find the probabilities involving negative values by:

$$\begin{aligned}\Phi(-z) &= P(Z \leq -z) \\ &= P(Z \geq z) \\ &= 1 - P(Z \leq z) \\ &= 1 - \Phi(z)\end{aligned}$$

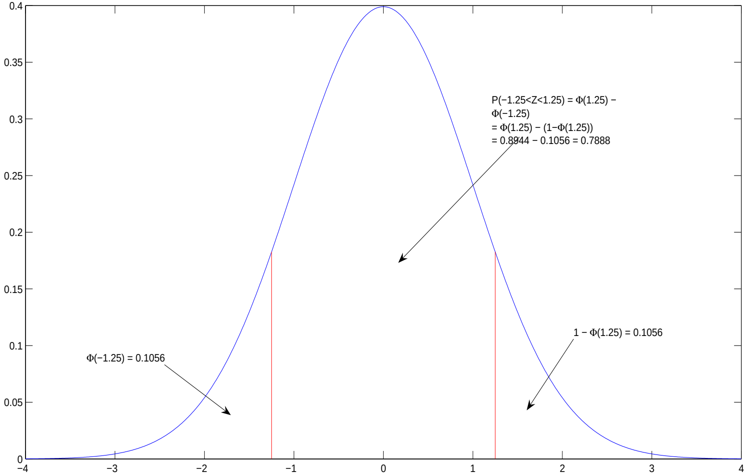
e.g.:

$$\begin{aligned}P(Z \leq -1.25) &= \Phi(-1.25) \\ &= 1 - \Phi(1.25) \\ &= 1 - 0.8944 = 0.1056\end{aligned}$$

Calculating Normal Distribution Probabilities



Calculating Normal Distribution Probabilities



CENTRAL LIMIT THEOREM - CLT

Let $X_1, X_2, \ldots, X_n$ be a random sample of n iid random variables each having the same mean μ and variance σ^2. More compactly:

$$X_i \sim \ i.i.d\ (\mu, \sigma^2), \quad i = 1, 2, \ldots, n$$

iid means: **i**dentically and **i**ndependently **d**istributed
Note that we do not mention the name of their distribution.
Expected value and variance of the sum of these n iid variables will be:

$$E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = n\mu$$

$$Var[X_1 + X_2 + \ldots + X_n] = Var[X_1] + Var[X_2] + \ldots + Var[X_n] = n\sigma^2$$

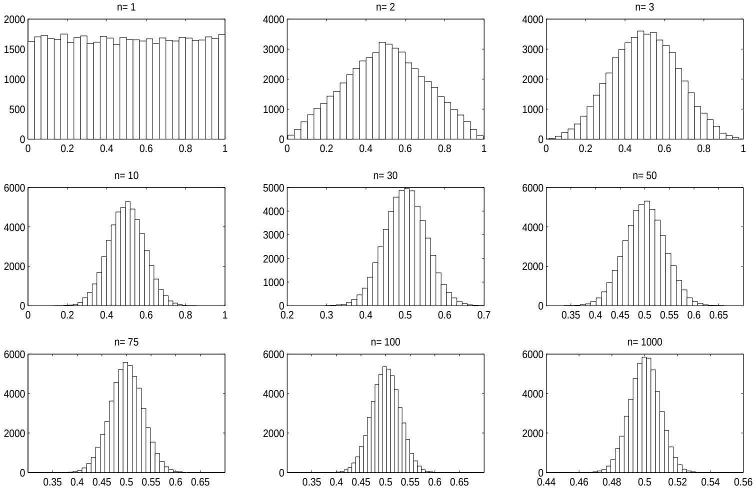
Let X denote the sum of these rvs: $X = X_1 + X_2 + \ldots + X_n$, then

$$\begin{aligned} Z &= \frac{X - E(X)}{\sqrt{Var(X)}} = \frac{X - n\mu}{\sqrt{n\sigma^2}} \\ &= \frac{\frac{X}{n} - \mu}{\frac{n^{1/2}}{n}\sigma} \\ &= \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \end{aligned}$$

According to CLT as $n \to \infty$, the expression above converges to standard normal distribution:

$$Z \to \ N(0,1)$$

CENTRAL LIMIT THEOREM - CLT



According to LLN, the sample mean of n iid random variables converges to the population mean as the sample size n increases. Let $\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ be the sample mean then the LLN says that

$$n \longrightarrow \infty, \quad \overline{X}_n \longrightarrow \mu$$

In other words, for a positive number ϵ that we can arbitrarily choose as small as possible we can write:

$$\lim_{n \to \infty} P\left[|\overline{X}_n - \mu| < \epsilon\right] = 1$$

- Let $X_1, X_2, \ldots, X_{12}$ be an iid random sample each distributed as $U \sim (0, b)$, $b > 0$. Using CLT show that the probability $P(\frac{b}{4} < \overline{X} < \frac{3b}{4})$ is approximately 0.9973.
- Answer: These 12 iid rvs come from the same population. So we need to find the population mean and variance first. For Uniform(a, b) distribution these quantities are

$$\mu_x = \frac{b+a}{2}, \quad \sigma_x^2 = \frac{(b-a)^2}{12}$$

Hence in our example

$$\mu_x = \frac{b}{2}, \quad \sigma_x^2 = \frac{b^2}{12}$$

CLT - Example

- Let $X_1, X_2, \ldots, X_{12}$ be an iid random sample each distributed as $U \sim (0, b)$, $b > 0$. Using CLT show that the probability $P(\frac{b}{4} < \overline{X} < \frac{3b}{4})$ is approximately 0.9973.
- Answer: These 12 iid rvs come from the same population. So we need to find the population mean and variance first. For Uniform(a, b) distribution these quantities are

$$\mu_x = \frac{b+a}{2}, \quad \sigma_x^2 = \frac{(b-a)^2}{12}$$

Hence in our example

$$\mu_x = \frac{b}{2}, \quad \sigma_x^2 = \frac{b^2}{12}$$

CLT - Example, cont'd

$$Var(\overline{X}) = \frac{\sigma_x^2}{n} = \frac{b^2}{144}$$

Using CLT we have:

$$\begin{aligned}
P\left(\frac{b}{4} < \overline{X} < \frac{3b}{4}\right) &= P\left(\frac{\frac{b}{4} - \frac{b}{2}}{\frac{b}{12}} < \frac{\overline{X} - \mu_x}{\sqrt{\sigma_x^2/n}} < \frac{\frac{3b}{4} - \frac{b}{2}}{\frac{b}{12}}\right) \\
&= P(-3 < Z < 3) = \Phi(3) - (1 - \Phi(3)) \\
&= 0.99865 - (1 - 0.99865) = 0.9973
\end{aligned}$$

Chi-square Distribution

- Square of a standard normal random variable is distributed as chi-square with 1 degrees of freedom:

$$\text{If } Z \sim N(0,1) \text{ then } Z^2 \sim \chi^2_1$$

- Sum of squares of n independent standard normal variables is distributed as a chi-squared with n degrees of freedom:

$$\text{If } Z_i \sim i.i.d.\ N(0,1) \quad i = 1, 2, \ldots, n \text{ then } \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$$

- Chi-square distribution has one parameter: ν (nu) degrees of freedom. For χ^2_ν the pdf is

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad x > 0, \ \nu > 0$$

where Γ is the gamma distribution:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0$$

Chi-square Distribution

- Square of a standard normal random variable is distributed as chi-square with 1 degrees of freedom:

$$\text{If} \;\; Z \sim N(0,1) \;\; \text{then} \;\; Z^2 \sim \chi_1^2$$

- Sum of squares of n independent standard normal variables is distributed as a chi-squared with n degrees of freedom:

$$\text{If} \;\; Z_i \sim i.i.d. \; N(0,1) \;\; i = 1, 2, \ldots, n \;\; \text{then} \;\; \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

- Chi-square distribution has one parameter: ν (nu) degrees of freedom. For χ_ν^2 the pdf is

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad x > 0, \;\nu > 0$$

where Γ is the gamma distribution:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0$$

- Square of a standard normal random variable is distributed as chi-square with 1 degrees of freedom:

$$\text{If } \; Z \sim N(0,1) \; \text{ then } \; Z^2 \sim \chi^2_1$$

- Sum of squares of n independent standard normal variables is distributed as a chi-squared with n degrees of freedom:

$$\text{If } \; Z_i \sim i.i.d. \; N(0,1) \;\; i = 1, 2, \ldots, n \; \text{ then } \; \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$$

- Chi-square distribution has one parameter: ν (nu) degrees of freedom. For χ^2_ν the pdf is

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad x > 0, \; \nu > 0$$

where Γ is the gamma distribution:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0$$

Chi-square Distribution

- Let χ^2_ν be a chi-square random variable with ν degrees of freedom. Then the expected value and variance are:

$$E(\chi^2_\nu) = \nu \ \text{ ve } \ Var(\chi^2_\nu) = 2\nu$$

- If the population is normal then the ratio of the sum of squared deviations from the sample mean to population variance has a chi-square distribution with $n-1$ degrees of freedom:

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

- Meaning of dof: the knowledge of n quantities $(X_i - \overline{X})$ is equivalent to $(n-1)$ mathematically independent components. Since we first estimate the sample mean from the observation set there are $n-1$ mathematically independent terms.

- Let χ^2_ν be a chi-square random variable with ν degrees of freedom. Then the expected value and variance are:

$$E(\chi^2_\nu) = \nu \text{ ve } Var(\chi^2_\nu) = 2\nu$$

- If the population is normal then the ratio of the sum of squared deviations from the sample mean to population variance has a chi-square distribution with $n-1$ degrees of freedom:

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

- Meaning of dof: the knowledge of n quantities $(X_i - \overline{X})$ is equivalent to $(n-1)$ mathematically independent components. Since we first estimate the sample mean from the observation set there are $n-1$ mathematically independent terms.

Chi-square Distribution

- Let χ^2_ν be a chi-square random variable with ν degrees of freedom. Then the expected value and variance are:

$$E(\chi^2_\nu) = \nu \text{ ve } Var(\chi^2_\nu) = 2\nu$$

- If the population is normal then the ratio of the sum of squared deviations from the sample mean to population variance has a chi-square distribution with $n-1$ degrees of freedom:

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

- Meaning of dof: the knowledge of n quantities $(X_i - \overline{X})$ is equivalent to $(n-1)$ mathematically independent components. Since we first estimate the sample mean from the observation set there are $n-1$ mathematically independent terms.

Chi-square Distribution

- Skewed to right, i.e., right tail is longer than the left. This means that the 3rd standard moment is positive.
- As the degrees of freedom parameter ν gets bigger distribution becomes more symmetric. In the limit, it converges to the normal distribution.
- Chi-square probabilities can easily be calculated using appropriate tables.

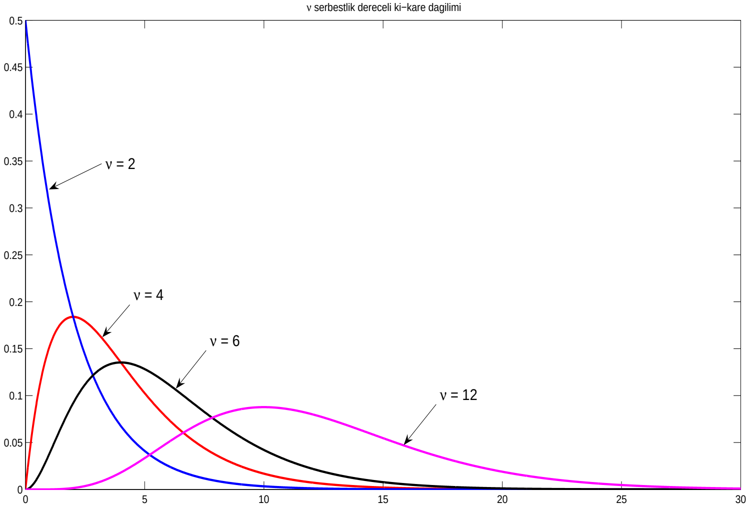
Chi-square Distribution

- Skewed to right, i.e., right tail is longer than the left. This means that the 3rd standard moment is positive.
- As the degrees of freedom parameter ν gets bigger distribution becomes more symmetric. In the limit, it converges to the normal distribution.
- Chi-square probabilities can easily be calculated using appropriate tables.

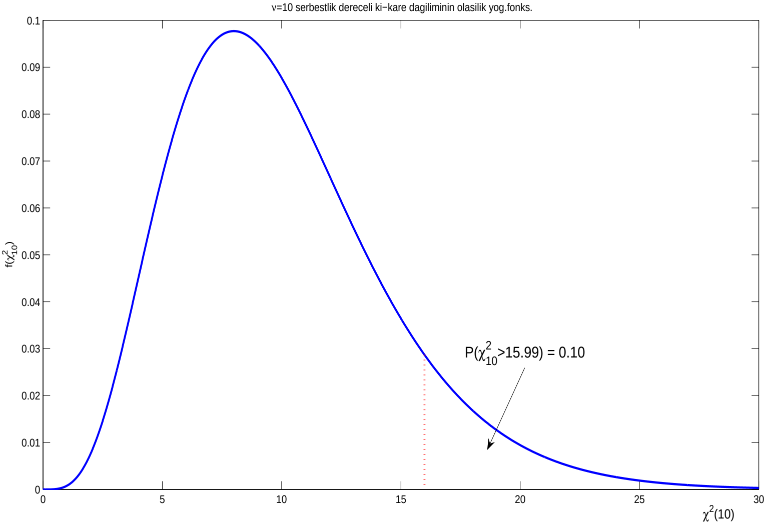
Chi-square Distribution

- Skewed to right, i.e., right tail is longer than the left. This means that the 3rd standard moment is positive.
- As the degrees of freedom parameter ν gets bigger distribution becomes more symmetric. In the limit, it converges to the normal distribution.
- Chi-square probabilities can easily be calculated using appropriate tables.

Chi-square Distribution



Chi-square Distribution



Student t Distribution

Let Z and Y be two rvs defined as follows: $Z \sim N(0,1)$, and $Y \sim \chi^2_\nu$. The random variable defined below follows a t distribution with ν degrees of freedom:

$$t_\nu = \frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

The rv t_ν has a t distribution with ν dof defined in the denominator.

- PDF is given by:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \frac{1}{(1+(t^2/\nu))^{(\nu+1)/2}}, \quad -\infty < t < \infty$$

- It has one parameter (ν) and it has symmetric shape. For $E(t_\nu) = 0$ and $\nu \geq 3$ $Var(t_\nu) = \nu/(\nu-2)$
- $\nu \to \infty$, $t_\nu \to N(0,1)$

Student t Distribution

Let Z and Y be two rvs defined as follows: $Z \sim N(0,1)$, and $Y \sim \chi^2_\nu$. The random variable defined below follows a t distribution with ν degrees of freedom:

$$t_\nu = \frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

The rv t_ν has a t distribution with ν dof defined in the denominator.

- PDF is given by:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \frac{1}{(1+(t^2/\nu))^{(\nu+1)/2}}, \quad -\infty < t < \infty$$

- It has one parameter (ν) and it has symmetric shape. For $E(t_\nu) = 0$ and $\nu \geq 3$ $Var(t_\nu) = \nu/(\nu-2)$
- $\nu \to \infty$, $t_\nu \to N(0,1)$

Student t Distribution

Let Z and Y be two rvs defined as follows: $Z \sim N(0,1)$, and $Y \sim \chi^2_\nu$. The random variable defined below follows a t distribution with ν degrees of freedom:

$$t_\nu = \frac{Z}{\sqrt{Y/\nu}} \sim t_\nu$$

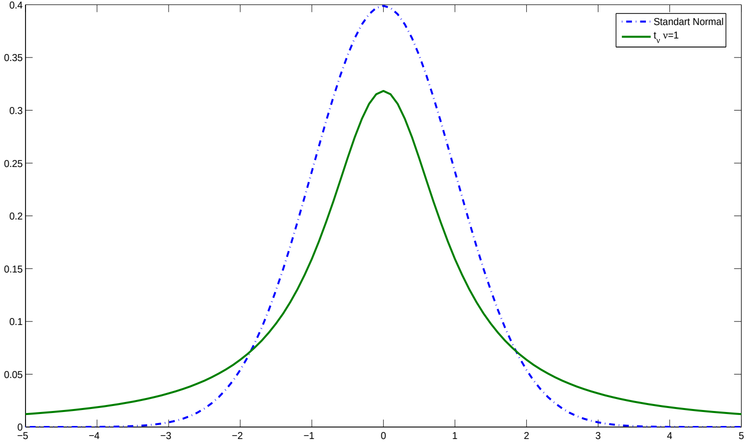
The rv t_ν has a t distribution with ν dof defined in the denominator.

- PDF is given by:

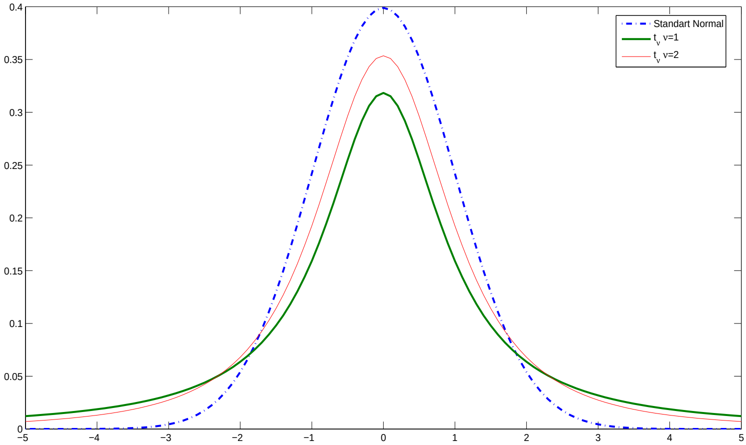
$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \frac{1}{(1+(t^2/\nu))^{(\nu+1)/2}}, \quad -\infty < t < \infty$$

- It has one parameter (ν) and it has symmetric shape. For $E(t_\nu) = 0$ and $\nu \geq 3$ $Var(t_\nu) = \nu/(\nu-2)$
- $\nu \to \infty$, $t_\nu \to N(0,1)$

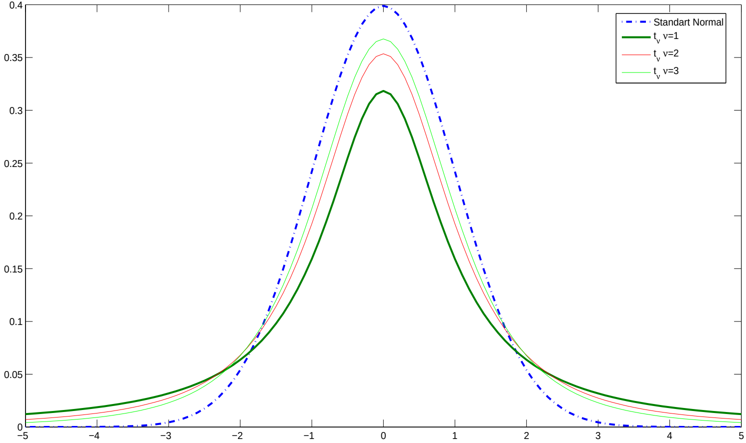
Student t Distribution



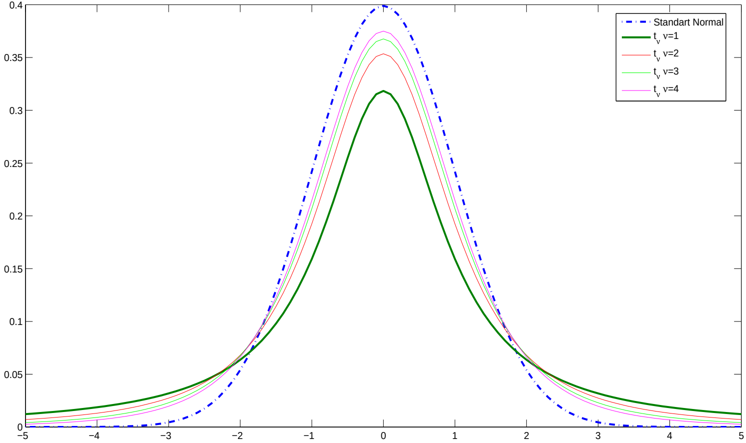
Student t Distribution



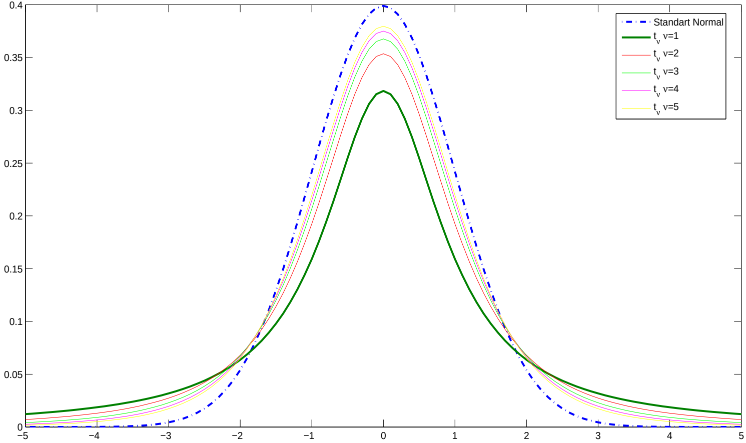
Student t Distribution



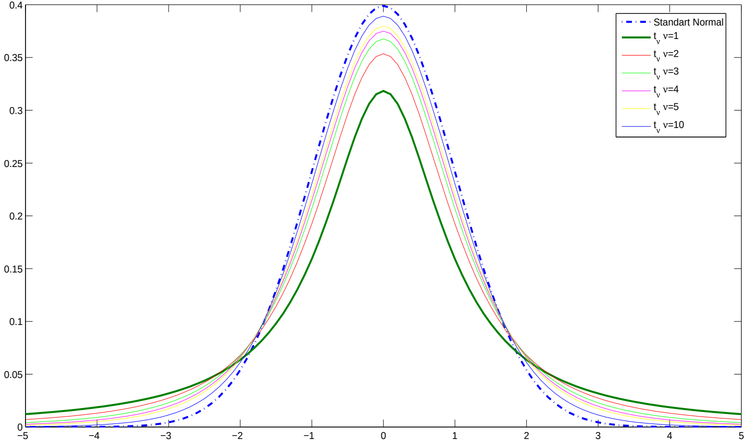
Student t Distribution



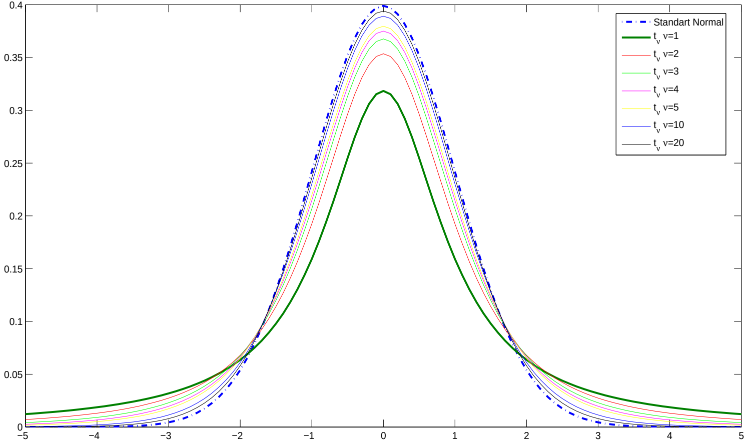
Student t Distribution



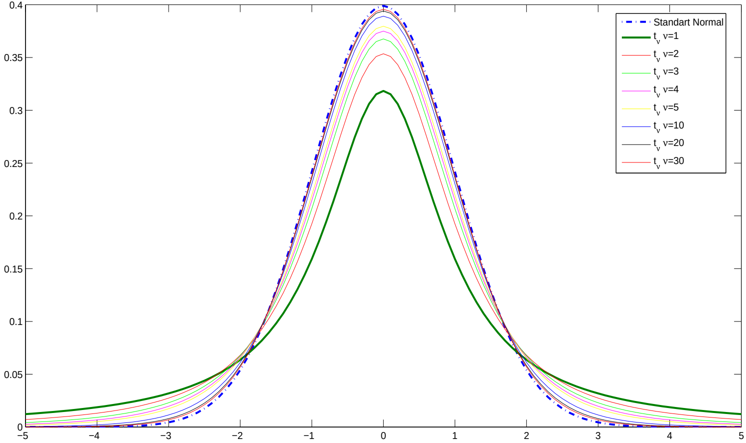
Student t Distribution



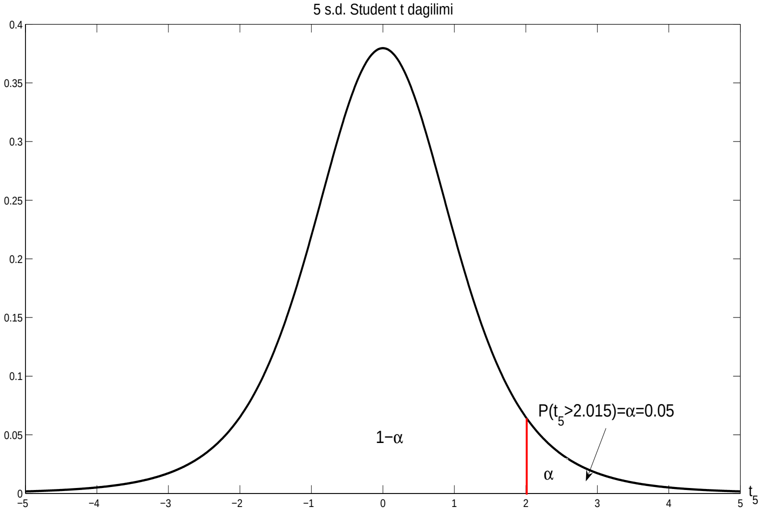
Student t Distribution



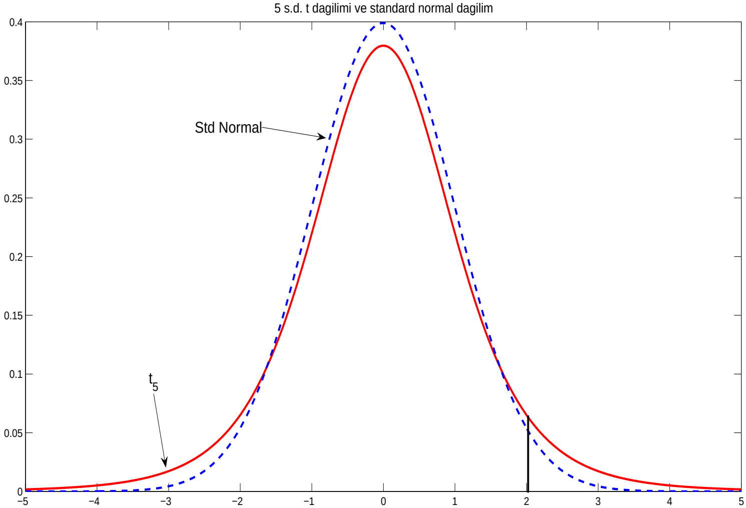
Student t Distribution



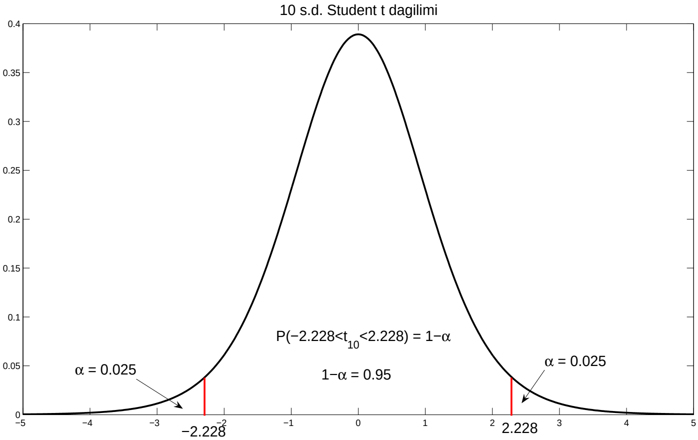
Calculating Student t Probabilities



Calculating Student t Probabilities



$$P(-2.228 < t_{10} < 2.228) = 1 - 0.025 - 0.025 = 0.95$$



F Distribution

- Let $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$ be two independent chi-square random variables.
- Then the following random variable has an F distribution

$$F = \frac{X_1/k_1}{X_2/k_2} \sim F(k_1, k_2)$$

- F distribution has two parameters:
- k_1 is the numerator degrees of freedom
- k_2 is denominator degrees of freedom

F Distribution

- Let $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$ be two independent chi-square random variables.
- Then the following random variable has an F distribution

$$F = \frac{X_1/k_1}{X_2/k_2} \sim F(k_1, k_2)$$

- F distribution has two parameters:
- k_1 is the numerator degrees of freedom
- k_2 is denominator degrees of freedom

F Distribution

- Let $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$ be two independent chi-square random variables.
- Then the following random variable has an F distribution

$$F = \frac{X_1/k_1}{X_2/k_2} \sim F(k_1, k_2)$$

- F distribution has two parameters:
- k_1 is the numerator degrees of freedom
- k_2 is denominator degrees of freedom

F Distribution

- Let $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$ be two independent chi-square random variables.
- Then the following random variable has an F distribution

$$F = \frac{X_1/k_1}{X_2/k_2} \sim F(k_1, k_2)$$

- F distribution has two parameters:
- k_1 is the numerator degrees of freedom
- k_2 is denominator degrees of freedom

F Distribution

- Let $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$ be two independent chi-square random variables.
- Then the following random variable has an F distribution

$$F = \frac{X_1/k_1}{X_2/k_2} \sim F(k_1, k_2)$$

- F distribution has two parameters:
- k_1 is the numerator degrees of freedom
- k_2 is denominator degrees of freedom

F Distribution