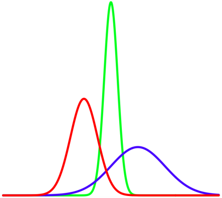
Introduction to Statistics

LV Nr. 105.692
Summer Semester 2021

Efstathia Bura

ASTAT
E105 Institute of Statistics and Mathematical Methods in Economics



Introduction to Statistics

LV Nr. 105.692

Summer Semester 2021

Efstathia Bura

ASTAT
E105 Institute of Statistics and Mathematical Methods in Economics

Overview

- This course is an introduction to Statistics.

Overview

- This course is an introduction to Statistics.
- It comprises of
  - Relevant Concepts of Probability theory
  - Statistics
    - Descriptive und Inferential Statistics
    - Implementation of statistical theory
      (basic ideas of data analysis).

Overview

- This course is an introduction to Statistics.
- It comprises of
  - Relevant Concepts of Probability theory
  - Statistics
    - Descriptive und Inferential Statistics
    - Implementation of statistical theory (basic ideas of data analysis).
  - TISS: The course announcement and description is at `https://tiss.tuwien.ac.at/course/courseDetails.xhtml?dswid=9952&dsrid=970&courseNr=105692&semester=2021S`

Overview

- This course is an introduction to Statistics.
- It comprises of
  - Relevant Concepts of Probability theory
  - Statistics
    - Descriptive und Inferential Statistics
    - Implementation of statistical theory (basic ideas of data analysis).
  - TISS: The course announcement and description is at `https://tiss.tuwien.ac.at/course/courseDetails.xhtml?dswid=9952&dsrid=970&courseNr=105692&semester=2021S`
  - TUWEL: Course material: `https://tuwel.tuwien.ac.at/course/view.php?id=36761`.
  - Course registration in TISS and TUWEL.

Accompanying exercises

UE Exercises LV 105.693

- Exercises will be on topics covered in lectures.
- TISS: `https://tiss.tuwien.ac.at/course/courseDetails.xhtml?dswid=4993&dsrid=707&courseNr=105693&semester=2021S`
- TUWEL: Course material and further course information: `https://tuwel.tuwien.ac.at/course/view.php?idnumber=105693-2021S`.

Course dates/times

- Lectures
  - LV Nr. 105.692
  - Wed 10:00–12:15 (without break)

- Exercise sessions
  - LVA-Nr. 107.693
  - 3 Groups, Tuesdays (please see TISS)
  - The first exercise session is on Tuesday 9.03.2021.

Requirements

- Basic knowledge of probability theory, linear algebra and calculus.
- For the accompanying exercises, you should install
  1. the statistical software R https://cran.r-project.org
  2. the interface RStudio https://www.rstudio.com/
  3. `R` is a free open source software.
  4. No prior knowledge of `R` is needed!

Textbooks

- F. Abramovich and Y. Ritov
  Statistical theory : a concise introduction,
  CRC Press, 2013.

- G. Casella, R. L. Berger
  Statistical Inference (2nd Edition).
  Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 2002.

- M. Messer, G. Schneider
  Statistik: Theorie und Praxis im Dialog.
  Springer, Berlin
  (available as an online resource in the TU library)

- L. Wasserman
  All of Statistics: A Concise Course in Statistical Inference.
  Springer, 2004.

- U. Krengel
  Einfuührung in die Wahrscheinlichkeitstheorie und Statistik.
  Vieweg Wiesbaden

- Y. Pawitan
  In All Likelihood: Statistical Modelling And Inference Using Likelihood.
  Oxford University Press, 2001.

Grading

- One final comprehensive written exam (based on the course material).
    1. Multiple-Choice questions
    2. Exactly one correct answer.
    3. The exam will be 90 minute long
    4. A non-programmable calculator and a two-sided handwritten A4 sheet may be used during the exam.
    5. Computers, smartphones, tablets, further notes, books, etc., as well as discussions and consultations are prohibited during the exam.

Grading

- One final comprehensive written exam (based on the course material).
  1. Multiple-Choice questions
  2. Exactly one correct answer.
  3. The exam will be 90 minute long
  4. A non-programmable calculator and a two-sided handwritten A4 sheet may be used during the exam.
  5. Computers, smartphones, tablets, further notes, books, etc., as well as discussions and consultations are prohibited during the exam.
- Three examination dates:
  1. June 29, 2021, 10:00–12:00, online
  2. October 1, 2021, time and place TBA
  3. December 2, 2021, time and place TBA

Grading

- One final comprehensive written exam (based on the course material).
    1. Multiple-Choice questions
    2. Exactly one correct answer.
    3. The exam will be 90 minute long
    4. A non-programmable calculator and a two-sided handwritten A4 sheet may be used during the exam.
    5. Computers, smartphones, tablets, further notes, books, etc., as well as discussions and consultations are prohibited during the exam.
- Three examination dates:
    1. June 29, 2021, 10:00–12:00, online
    2. October 1, 2021, time and place TBA
    3. December 2, 2021, time and place TBA
- Exam Registration in TISS

Probability vs. Statistics

Different subjects: both about random processes

- Non-probabilistic: Tomorrow's date: we are certain of the value of this variable

Probability vs. Statistics

Different subjects: both about random processes

- Non-probabilistic: Tomorrow's date: we are certain of the value of this variable
- Probability
  - A quantity takes values without certainty
    - We know the price of a stock today, do we know its price tomorrow, in a week, in a month?
    - Due to a variety of (partially unknown) factors affecting stock prices, there will always be uncertainty in its future value
  - If the random mechanism that generates the stock price were known exactly, calculating probabilities of its possible values, etc, is what probability theory does

Probability vs. Statistics

- Statistics
  - The random mechanism is **unknown**. In order to learn and understand the random mechanism, one can collect data and analyze them
  - Three stages in statistical studies
    - Data collection with the aim of drawing probabilistic conclusions
    - Data analysis (initial descriptive statistics, fitting an appropriate statistical model and estimating its parameters)
    - Statistical inference

Probability vs. Statistics

- Statistics
  - The random mechanism is **unknown**. In order to learn and understand the random mechanism, one can collect data and analyze them
  - Three stages in statistical studies
    - Data collection with the aim of drawing probabilistic conclusions
    - Data analysis (initial descriptive statistics, fitting an appropriate statistical model and estimating its parameters)
    - Statistical inference
- Probability vis-a-vis Statistics:
  - Probability: you know the population (the whole) and compute probabilities about members of the population (subset)
  - Statistics: you observe members of the population (data, subset) and you infer about the population (whole)

Probability vs. Statistics: Examples

1. Probability example

   We toss a fair coin (equal probability of heads or tails) 100 times.
   What is the probability of obtaining 60 or more heads?

Probability vs. Statistics: Examples

1. Probability example

   We toss a fair coin (equal probability of heads or tails) 100 times.
   What is the probability of obtaining 60 or more heads?

   1. There is only one answer (about 0.028444).
      We will see how to compute it.

Probability vs. Statistics: Examples

1. Probability example

   We toss a fair coin (equal probability of heads or tails) 100 times.
   What is the probability of obtaining 60 or more heads?

   1. There is only one answer (about 0.028444).
      We will see how to compute it.

2. Statistics example

   We have a coin of unknown source.
   In order to investigate whether it is fair we toss it 100 times and count the number of heads. Assume we count 60 heads.
   The goal is to draw a conclusion (inference) from these data.

Probability vs. Statistics: Examples

1. Probability example

   We toss a fair coin (equal probability of heads or tails) 100 times.
   What is the probability of obtaining 60 or more heads?

   1. There is only one answer (about 0.028444).
      We will see how to compute it.

2. Statistics example

   We have a coin of unknown source.
   In order to investigate whether it is fair we toss it 100 times and count the number of heads. Assume we count 60 heads.
   The goal is to draw a conclusion (inference) from these data.

   1. There are many ways to proceed.
      Different statisticians might draw different conclusions.

Probability vs. Statistics: Examples

1. Probability example

   We toss a fair coin (equal probability of heads or tails) 100 times.
   What is the probability of obtaining 60 or more heads?

   1. There is only one answer (about 0.028444).
      We will see how to compute it.
   2. The random process is fully known (probability of heads is 0.5).

2. Statistics example

   We have a coin of unknown source.
   In order to investigate whether it is fair we toss it 100 times and count the number of heads. Assume we count 60 heads.
   The goal is to draw a conclusion (inference) from these data.

   1. There are many ways to proceed.
      Different statisticians might draw different conclusions.

Probability vs. Statistics: Examples

1. Probability example

   We toss a fair coin (equal probability of heads or tails) 100 times.
   What is the probability of obtaining 60 or more heads?

   1. There is only one answer (about 0.028444).
      We will see how to compute it.
   2. The random process is fully known (probability of heads is 0.5).
   3. The aim is to find the probability of a certain outcome (at least 60 heads) arising from the random process.

2. Statistics example

   We have a coin of unknown source.
   In order to investigate whether it is fair we toss it 100 times and count the number of heads. Assume we count 60 heads.
   The goal is to draw a conclusion (inference) from these data.

   1. There are many ways to proceed.
      Different statisticians might draw different conclusions.

Probability vs. Statistics: Examples

1. Probability example

   We toss a fair coin (equal probability of heads or tails) 100 times.
   What is the probability of obtaining 60 or more heads?

   1. There is only one answer (about 0.028444).
      We will see how to compute it.
   2. The random process is fully known (probability of heads is 0.5).
   3. The aim is to find the probability of a certain outcome (at least 60 heads) arising from the random process.

2. Statistics example

   We have a coin of unknown source.
   In order to investigate whether it is fair we toss it 100 times and count the number of heads. Assume we count 60 heads.
   The goal is to draw a conclusion (inference) from these data.

   1. There are many ways to proceed.
      Different statisticians might draw different conclusions.
   2. The result is known (60 heads).

Probability vs. Statistics: Examples

1. Probability example

   We toss a fair coin (equal probability of heads or tails) 100 times.
   What is the probability of obtaining 60 or more heads?

   1. There is only one answer (about 0.028444).
      We will see how to compute it.
   2. The random process is fully known (probability of heads is 0.5).
   3. The aim is to find the probability of a certain outcome (at least 60 heads) arising from the random process.

2. Statistics example

   We have a coin of unknown source.
   In order to investigate whether it is fair we toss it 100 times and count the number of heads. Assume we count 60 heads.
   The goal is to draw a conclusion (inference) from these data.

   1. There are many ways to proceed.
      Different statisticians might draw different conclusions.
   2. The result is known (60 heads).
   3. The aim is to explain the unknown random process (the probability of heads).

Probability theory and Statistics

Wide range of applications

- medicine, physical sciences, engineering, the social sciences, the life sciences, economics and computer science
  - tests of one medical treatment against another (or a placebo)
  - measures of genetic linkage
  - the search for elementary particles
  - machine learning for vision or speech
  - gambling probabilities and strategies
  - climate modeling
  - economic forecasting
  - epidemiology
  - Marketing
  - ...

Probability theory: Topics

1. Basic probability theory
    1. probability space, the calculus with probabilities conditional probability and independence, Law of total probability, Bayes' Rule
2. Random variables
    1. discrete and continuous random variables, distribution functions, pdfs, pmfs, common families of distributions
    2. Moments (expectation and variance), moment generating functions, properties, transformations, independence, covariance and correlation
3. Properties of a random sample
    1. sum of random variables from random sample, sampling from normal distribution, properties of sample mean and sample variance, order statistics
    2. Convergence concepts, Law of large numbers, Central limit theorem, Normal approximations

Statistics: Topics

1. Descriptive statistics
    1. numerical summaries, empirical distribution, graphical summaries (histograms, boxplots, scatterplots)
2. Inferential statistics
    1. Estimation and inference: point estimation, interval estimation, methods of finding estimators, methods of evaluating estimators (mean squared error, unbiased estimators), sufficient statistics
    2. Hypothesis testing, methods of finding tests, methods of evaluating tests (error probabilities, power function), the p-value for these tests.
    3. Nonparametric tests, Analysis of Variance
    4. Linear regression
        1. simple linear models, regression line, coefficient of determination

Sample spaces and events

Suppose that we conduct an experiment.
An experiment is a measurement of a random (stochastic) process. Our measurements take values in some set Ω: this is the sample space.

Definition

The *sample space* is the collection of all possible outcomes of a random experiment.

- Examples:
  - Suppose I toss a coin: in this case the sample space $\Omega = \{H, T\}$
  - If I measure the reaction time to some stimulus the sample space $\Omega = (0, \infty)$
  - Suppose I toss a coin twice: what is the sample space?

Sample spaces and events

Suppose that we conduct an experiment.
An experiment is a measurement of a random (stochastic) process. Our measurements take values in some set Ω: this is the sample space.

Definition

The *sample space* is the collection of all possible outcomes of a random experiment.

- Examples:
  - Suppose I toss a coin: in this case the sample space $\Omega = \{H, T\}$
  - If I measure the reaction time to some stimulus the sample space $\Omega = (0, \infty)$
  - Suppose I toss a coin twice: what is the sample space?
    - $\Omega = \{HH, HT, TH, TT\}$

Sample spaces and events

Definition

An event is a subset A of Ω; i.e., $A \subset \Omega$. That is, A is a set of possible outcomes of the random experiment.

We say that an event A *occurs* if the outcome of our experiment lies in the set A.

- Some basic notation for set operations.
  - $A \subset B$ means that all elements in A are also in B.
  - (Complement) A^c: elements that are not in A.
  - (Empty set) $\Omega^c = \emptyset$
  - (Union) $A \cup B$: elements that are either in A or in B, or both.
  - (Intersection) $A \cap B$: elements that are both in A and B. Sometimes we use AB for brevity.
  - (Set difference) $A - B = A \cap (B^c)$: elements that are in A but not in B.
  - (Cardinality) $|A|$ denotes the number of elements in A.
  - **Exercise:** Show that $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$

Probability

Definition

A probability distribution is a mapping from events to real numbers, $\mathbb{P} : A \to \mathbb{R}$ that satisfies the following axioms.

1. Non-negativity: $\mathbb{P}(A) \geq 0,\ \forall A \subset \Omega$
2. Unity of Ω: $\mathbb{P}(\Omega) = 1$
3. Countable additivity: For a collection of disjoint sets $A_1, A_2, \ldots,$

$$\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Probability

We can use these axioms to show several useful and intuitive properties of probability distributions:

1. $\mathbb{P}(\emptyset) = 0$
2. $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
3. $0 \leq \mathbb{P}(A) \leq 1$
4. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
5. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Example: Suppose I toss a fair coin twice, and denote by A the event that the first coin lands heads, and B the event that the second coin lands heads. Calculate $\mathbb{P}(A \cup B)$.

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.5 + 0.5 - 0.25 = 0.75$$

Exercise: Derive an analogous formula for $\mathbb{P}(\cup_{i=1}^{n} A_i)$.

Counting and the uniform distribution on discrete sets

Suppose we toss a die twice. There are 36 possible outcomes:

$$\Omega = \{(o_1, o_2) : o_1, o_2 = 1, 2, 3, 4, 5, 6\}.$$

If the die is fair then each outcome is equally likely. This is an example of a uniform distribution on a discrete set.
The general rule of calculating probability of an event A under the uniform distribution in finite sample space is

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

For example, let A be the event that the sum of two rolls is less than five. Then,

$$A = \{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\}$$

so that $\mathbb{P}(A) = 6/36 = 1/6$.

Example: There are two black balls and three white balls in a bag. Two balls are randomly drawn, without replacement, from the bag. What is the probability of the two balls have different color? What is the probability if the balls are drawn with replacement?

- When drawing without replacement, the sample space has cardinality $|\Omega| = 5 \times 4 = 20$, and the event A has cardinality $|A| = 2 \times 3 + 3 \times 2 = 12$ (first white and second black, or first black second white). So $\mathbb{P}(A) = 12/20 = 0.6$.
- When drawing with replacement, the sample space has cardinality $|\Omega| = 5\times = 25$, and A still has cardinality 12. Then $\mathbb{P}(A) = 12/25 = 0.48$.

Even more generally, probabilities under non-uniform distributions can be calculated by adding the probabilities of the individual singleton events.

Example: Suppose we toss an unfair die twice. List events and associated probabilities.

Conditional probability

Definition

If $\mathbb{P}(B) > 0$, the conditional probability of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Example: Consider tossing a fair die. Let A be the event that the result is an odd number, and $B = \{1, 2, 3\}$. Then $\mathbb{P}(A \mid B) = 2/3$, and $\mathbb{P}(A) = 1/2$.

- In general, $\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$
- The *chain rule*: A simple re-writing of the above expression yields

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A \mid B) = \mathbb{P}(A)\mathbb{P}(B \mid A)$$

- More generally,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap \ldots) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1, A_2) \ldots$$

Independence of Events

Independence means whether one event provides any information about another.

For example, if we toss a fair coin twice and let A_i be the event that the ith toss is heads $(i = 1, 2)$, then intuitively knowing if the event A_1 occurred or not does not provide any information about A_2. Formally,

Definition

Two events A and B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

A set of events A_i, $i \in I$, are called mutually independent if

$$\mathbb{P}\left(\cap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

for any finite subset J of I.

Independence of Events

Conditional probability gives another interpretation of independence: A and B are independent if the unconditional probability is the same as the conditional probability.

- **Example:** We can formally verify the coin toss example: $\mathbb{P}(A_1) = 1/2$, $\mathbb{P}(A_2) = 1/2$, $\mathbb{P}(A_1 \cap A_2) = 1/4$
- A less obvious example: toss a fair die once. Let $A = \{1, 2, 3, 4\}$, $B = \{2, 4, 6\}$. Then $A \cap B = \{2, 4\}$. $\mathbb{P}(A) = 2/3$, $\mathbb{P}(B) = 1/2$, and $\mathbb{P}(A \cap B) = 1/3 = \mathbb{P}(A)\mathbb{P}(B)$.

Some simple facts about independence.

- Ω is independent of any other event. The same holds for $\emptyset$
- If A, B are disjoint, both having positive probability, then A and B cannot be independent.
- If A and B are independent, then A^c and B are also independent.

Law of total probability

Theorem

(Law of total probability) Let $A_1 \ldots, A_k$ be a partition of Ω. Then for any $B \in \Omega$,

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)$$

Proof: The claim follows by observing $A_i \cap B$ forms a partition of B and $\mathbb{P}(A_i \cap B) = \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)$

The law of total probability is a combination of additivity and conditional probability. It leads to the very useful Bayes' theorem.

Bayes' Rule

Theorem

(Bayes' Rule) Let $A_1 \ldots, A_k$ be a partition of Ω. Then, for any $B \in \Omega$,

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{i=1}^{k} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}$$

- Bayes' rule is useful when $\mathbb{P}(A_i \mid B)$ is not obvious to calculate but $\mathbb{P}(B \mid A_i)$ and $\mathbb{P}(A_i)$ are easy to find.
- A typical application is classification

Classification Example

- Suppose there are three types of emails: $A_1 =$"spam," $A_2 =$ "low priority," $A_3 =$ "high priority." Based on previous experience, $\mathbb{P}(A_1) = 0.85$, $\mathbb{P}(A_2) = 0.1$, $\mathbb{P}(A_3) = 0.05$. Let B be the event that an email contains the word "free," and $\mathbb{P}(B \mid A_1) = 0.9$, $\mathbb{P}(B \mid A_2) = 0.1$, $\mathbb{P}(B \mid A_3) = 0.1$. Now a new incoming email contains the word "free," what is the probability that it is spam?
- Answer:

$$\begin{aligned}\mathbb{P}(A_1 \mid B) &= \frac{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1)}{\sum_{i=1}^{3} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)} \\ &= \frac{0.85 \times 0.9}{0.85 \times 0.9 + 0.1 \times 0.1 + 0.05 \times 0.1}\end{aligned}$$

Random Variables

Often we are interested in summaries of experiments rather than the actual outcome. For instance, suppose we are repeating an experiment 100 times, and each time it either succeeds or fails. We are not really interested in which of the 2^{100} possible outcomes occurred but rather just in some simple summary statistic of the experiment, such as the number of successes. These summaries are called random variables.

Definition

A random variable X is a function from the sample space Ω to $\mathbb{R}$, $X : \Omega \to \mathbb{R}$.

Random Variables

- One way of thinking about a random variable is as a mapping between a distribution on Ω to a distribution on the reals, the range of the random variable.
- Formally, we have that for some subset $A \subset \mathbb{R}$,

$$\mathbb{P}_X(X \in A) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A\}\right)$$

- $\mathbb{P}_X$ is called the induced probability distribution.

Random Variables: Coin tossing example

- Let us consider an experiment of tossing a fair coin three times, and define a random variable X to be the number of heads.
  - The original sample space is $\{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$.
  - The induced sample space is $\{0, 1, 2, 3\}$.
  - We can also calculate the induced probability distribution:

$$\mathbb{P}_X(X = 0) = \frac{1}{8}, \quad \mathbb{P}(X = 1) = \frac{3}{8}, \quad \mathbb{P}(X = 2) = \frac{3}{8}, \quad \mathbb{P}(X = 3) =$$

  - This distribution of counting number of successes in $n = 3$ identical and independent trials is the Binomial distribution:

$$\mathbb{P}_X(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n$$

    where $p = 1/2 = \mathbb{P}(H)$ for a fair coin.

Distribution Functions

Every random variable is associated with a cumulative distribution function (cdf).

Definition

The cdf of a random variable X is defined to be the function F_X such that

$$F_X(x) = \mathbb{P}(X \leq x), \; \forall x$$

Quantiles:

- The point x for which $F_X(x) = 0.5$ is called the median.
- The point x for which $F_X(x) = 0.75$ is called the third quartile.
- The point x for which $F_X(x) = 0.25$ is called the first quartile.

Distribution Functions

A function F is a cdf if and only if

1. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
2. F is a non-decreasing function of x.
3. F is right-continuous; i.e., for every number x_0

$$\lim_{x\to x_0^+} F(x) = F(x_0)$$

Any function that satisfies these three conditions is the distribution function of some random variable.

Distribution Functions

Example: Suppose we toss a coin repeatedly until we see a head, and let the random variable of interest X be the number of tosses. Then, X has induced distribution:

$$\mathbb{P}_X(X = x) = (1-p)^{x-1}p, \ \forall x = 1, 2, \ldots$$

For any positive integer x:

$$\begin{aligned} F_X(x) = \mathbb{P}_X(X \leq x) &= \sum_{i=1}^{x} (1-p)^{i-1} p \\ &= \frac{1-(1-p)^x}{1-(1-p)} p = 1-(1-p)^x \end{aligned}$$

What does this cdf look like?

Distribution Functions

Example: Suppose

$$F_X(x) = \frac{1}{1 + \exp(-x)}$$

Is this a valid cdf? We need to verify the three conditions.

1. Since $\exp(-x)$ tends to ∞ as $x \to -\infty$ and 0 as $x \to \infty$, it is clear that the first property holds.
2. Differentiate $F_X(x)$

$$\frac{d}{dx} F_X(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2} > 0$$

   so that F_X is a non-decreasing function of x.
3. Since, it is differentiable it is clear that F_X is continuous not just right-continuous.

Discrete and continuous random variables

- A random variable X can be continuous or discrete, depending on whether its values are continuous (intervals in $\mathbb{R}$) or discrete (countable)
- Formally, a random variable X is continuous if its cdf $F_X(x)$ is a continuous function of x, and
- X is discrete if its cdf $F_X(x)$ is a step function of x; i.e., it can be written as a finite linear combination of indicators of intervals.

i.i.d.

- An important concept is that of identically distributed random variables.
- Two random variables X and Y are identically distributed if for any (measurable) set A,

$$\mathbb{P}_X(X \in A) = \mathbb{P}_Y(Y \in A)$$

- Identically distributed does not mean equal
  - If I toss a fair coin n times, where n is odd, and let X be the number of heads and Y be the number of tails, these are identically distributed random variables but are clearly always unequal.

One of the most heavily used results about distribution functions is that the following two statements are equivalent:

- The random variables X and Y are identically distributed.

$$\mathbb{P}_X(X \in A) = \mathbb{P}_Y(Y \in A)$$

- Their distribution functions are equal; i.e.,

$$F_X(x) = F_Y(x),\ \forall x.$$

One of these implications is easy to verify, while the other is substantially more involved.

- It is easy to check that if (1) holds then (2) holds since we can just use (1) with the sets $(-\infty, x]$.