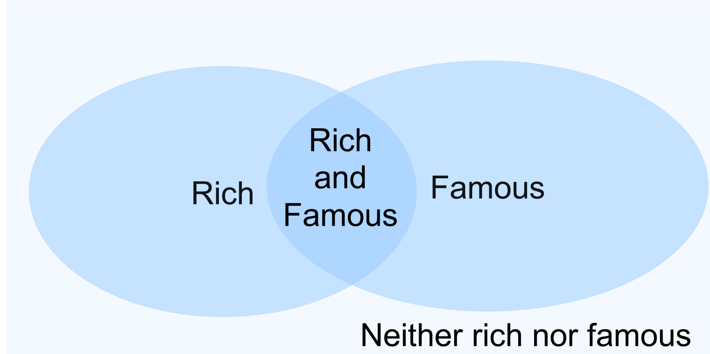
Aalto University
School of Business

Business Analytics 2
Lecture 2: Review of basic probability theory cont'd

- *Law of Total Probability*
- *Bayes theorem*
- *Random variables*
- *CDF and PDF*
- *Expectation*
- *Some probability distributions*

Last time



Concepts related to probability:

- Sample space, simple event, event
- Complement, union, intersection
- Mutually exclusive / collectively exhaustive events

Properties of probability:

I. $P(\emptyset) = 0$
II. $P(A^c) = 1 - P(A)$
III. If $A \subset B$, then $P(A) \leq P(B)$
IV. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability of A given that B has happened: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

If A and B are independent: $P(A \cap B) = P(A)P(B) \rightarrow P(A|B) = P(A)$

Law of total probability and Bayes rule

- **Law of total probability:** If $E_1, E_2, \ldots, E_n$ are mutually exclusive and collectively exhaustive, then

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_n)P(E_n)$$

- **Bayes rule:** If A and B are events with a positive probability, then

$$P(B|A) = \frac{P(\mathrm{A}|B)P(B)}{P(A)}$$

- Common usage:
  - We know $\mathrm{P}(A|B)$ and $\mathrm{P}(B)$ (base rate)
  - We need to deduce $\mathrm{P}(B|A)$

Law of total probability and Bayes rule: Example

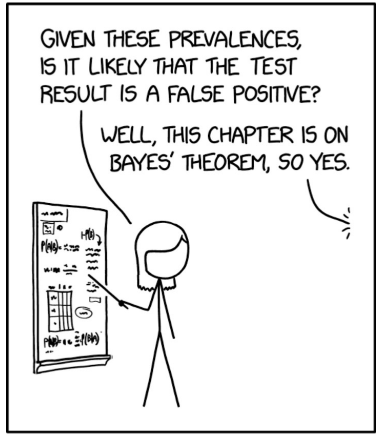
- A certain incurable rare disease affects 1 out of every 100,000 people.
  - There is a test for the disease which is quite accurate:
    - For infected people the test gives a positive result in 99% of cases
    - For non-infected people the test gives a positive result in 1% of cases
- Suppose you are tested positive for the disease. What is the probability that you actually have the disease?
  - Solution: Write down events
    - TP = Tested Positive, TP^C = Tested Negative
    - D = Having Disease, D^C = Not Having Disease
    - We need to figure out $P(D|TP)$

Bayes: $\mathrm{P}(D|TP) = \frac{P(TP|D)P(D)}{P(TP)} = \frac{0.99 \cdot 0.00001}{P(TP)} = \frac{0.0000099}{P(TP)} = 0.000989031 \approx \frac{1}{1000}$

Law of total probability: $P(TP) = P(TP|D)P(D) + P(TP|D^C)P(D^C)$

$$= 0.99 \cdot 0.00001 + 0.01 \cdot 0.99999 = 0.0100098$$

Law of total probability and Bayes rule



SOMETIMES, IF YOU UNDERSTAND BAYES' THEOREM WELL ENOUGH, YOU DON'T NEED IT.

Question: Ca. 5 % of the population of 45-65-year-old Finns will have a coronary heart disease (CHD) within the next 10 years. Consider two prognostic tests:

Test 1 (500 €)	**+**	**–**
CHD event	0.75	0.25
No CHD event	0.15	0.85

Test 2 (1000 €)	**+**	**–**
CHD event	0.95	0.05
No CHD event	0.05	0.95

Assume that a person gets a negative result from test 1. What is the probability of them having a CHD event anyway (false negative)?

What about with test 2?

Law of total probability and Bayes rule

Test 1:

$$P(\mathrm{CHD}|-) = \frac{P(-|CHD)P(CHD)}{P(-)}$$

$$= \frac{P(-|CHD)P(CHD)}{P(-|CHD)P(CHD) + P(-|No\ CHD)P(No\ CHD)}$$

$$= \frac{0.25 \cdot 0.05}{0.25 \cdot 0.05 + 0.85 \cdot 0.95} = 0.015244$$

Test 2:

$$P(\mathrm{CHD}|-) = \frac{0.05 \cdot 0.05}{0.05 \cdot 0.05 + 0.95 \cdot 0.95} = 0.002762$$

Test 1 (500 €)	+	–
CHD event	0.75	0.25
No CHD event	0.15	0.85

Test 2 (1000 €)	+	–
CHD event	0.95	0.05
No CHD event	0.05	0.95

Random Variables (r.v.)

- Random variable X is a mapping from sample space S to real numbers
  - Discrete random variable: set of possible values is countable
    - E.g. Sum of two dice $X(s)=s_1+s_2,\ s \in S = \{(i,j)|i,j = 1,\dots,6\}$
  - Continuous random variable: set of possible values is infinite
    - E.g. payoff from a stock option with an exercise price of 42 euros:
      $$X(s) = \max\{s - 42,0\}, s \in S = \mathbb{R}_+$$

- The probability measure P on the sample space defines the random variable's **probability distribution** over real numbers
  - This distribution can be represented by using

1. The Cumulative Distribution Function (CDF) or
2. The Probability Density Function (PDF)

Contain the same information, i.e., one can be derived from the other

Probability density function (PDF)*

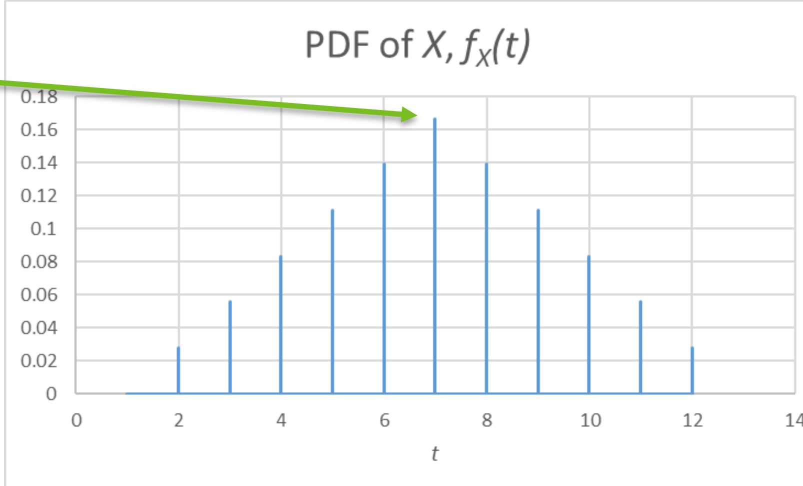
- The PDF of random variable X is denoted by f_X
  - Verbal definition: $f_X(t)$ is the probability that random variable X obtains a value equal to t
  - Formal definition: $f_X(t) = \mathrm{P}(\{s \in S | X(s) = t\})$
  - Properties: Non-negative, "sums up to one"
- Example:

$$f_X(7) = 6 \cdot \frac{1}{36} = 0.167$$

  - Sample space:

$$S = \begin{Bmatrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{Bmatrix}$$

  - Random variable: $X(s) = s_1 + s_2$



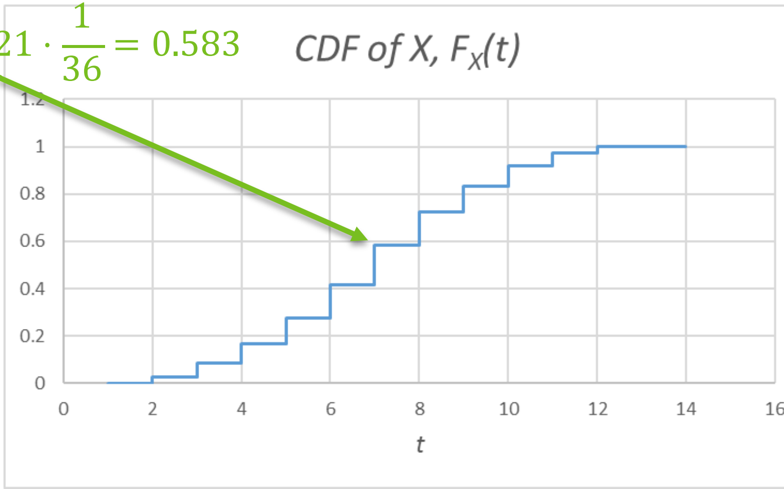
*) PDF of a discrete random variable is often called probability mass function

Cumulative distribution function (CDF)

- The CDF of random variable X is denoted by F_X
  - Verbal definition: $F_X(t)$ is the probability that random variable X obtains a value less than or equal to t
  - Formal definition: $F_X(t) = \mathrm{P}(\{s \in S | X(s) \leq t\})$
  - Properties: Non-decreasing, ranges from zero to one
- Example:
  - Sample space:

$$S = \begin{Bmatrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{Bmatrix}$$

$$F_X(7) = 21 \cdot \frac{1}{36} = 0.583$$



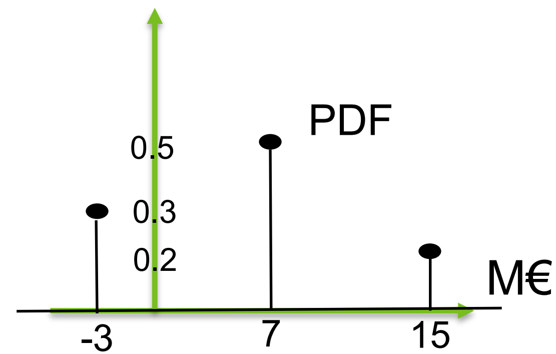
  - Random variable: $X(s) = s_1 + s_2$

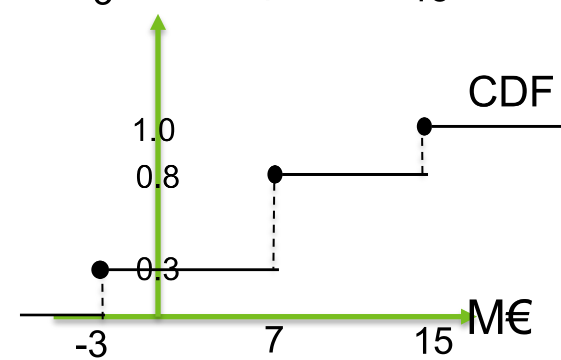
Example: PDF/ CDF of a discrete random variable

- Flame Consulting has prepared three scenarios for future sales

	Up, s^1	Stable, s^2	Down, s^3
Sales, $X(s)$	? M€	? M€	? M€
Probability, $P(\{s\})$	?	?	?

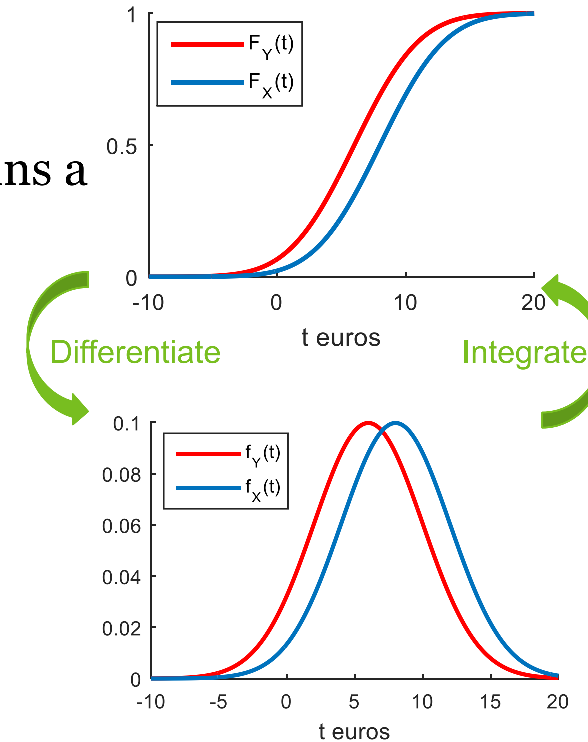
  - To visualize the probability distribution of future sales they also prepared the PDF and CDF
  - Computer bug destroyed the original data ☹
- Task: Use the CDF and the PDF to restore the destroyed data





PDF of continuous random variables

- The same definition of the CDF can be used for continuous random variables:
  - "$F_X(t)$ is the probability that random variable X obtains a value less than or equal to t"



- However, the probability of a continuous random variable being *exactly* equal to t is zero
  - Hence, PDF is defined as the derivative of the CDF:

$$f_X(t) = \frac{d}{dt} F_X(t)$$

  - Properties:
    - f_X is non-negative
    - Area between the horizontal axis and f_X is equal to one

Expectation

- The expected value of a random variable is the average of all possible values the variable can take weighted by their probabilities (densities)

Discrete r.v.

$$\mathrm{E}(X) = \sum_t t f_X(t),$$

Continuous r.v.

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} t f_X(t) dt,$$

 – A function *g(X)* of a random variable *X* is a random variable, whereby

$$\mathrm{E}(g(X)) = \sum_t g(t) f_X(t),$$

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t) dt,$$

- Example: Consider a game where you throw a die and get the outcome squared in euros. What is the expected return?

$$g(t) = t^2\text{:}\ E[g(X)] = \sum_{t=1}^{6} g(t) f_X(t) = \sum_{t=1}^{6} t^2 f_X(t) = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + \cdots + 36 \cdot \frac{1}{6} \approx 15.16$$

Expectation

Question: Consider a game where you throw a die and get the outcome doubled in euros. What is the expected return?

$$g(t) = 2t\text{: } E[g(X)] = \sum_{t=1}^{6} g(t) f_X(t) = \sum_{t=1}^{6} 2t f_X(t)$$

$$= 2 \cdot 1 \cdot \frac{1}{6} + 2 \cdot 2 \cdot \frac{1}{6} + 2 \cdot 3 \cdot \frac{1}{6} + \cdots + 2 \cdot 6 \cdot \frac{1}{6} = 7$$

Note: The expected outcome is $E[X] = \sum_{t=1}^{6} t f_X(t) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3.5$; in this case, $E[g(X)] = E[2X] = 2E[X] = g(E[X])$.

Expectation: Properties

- If $X_1, \dots, X_n$ and $Y = \sum_{i=1}^{n} X_i$ are random variables, then

$$\mathrm{E}(Y) = \sum_{i=1}^{n} E[X_i]$$

- If r.v. $Y = aX + b$, where a and b are constants, then

$$\mathrm{E}(Y) = aE(X) + b$$

  - Note: this applies only if Y is a linear function of X (cf. slide 13)
  - I.e., in general $E[g(X)] = g(E(X))$ does **NOT** hold
  - Example: Outcome squared in euros.

$$\mathrm{E}[g(X)] \approx 15.16, \text{but } g(\mathrm{E}[X]) = g(3.5) = 3.5^2 = 12.25$$

See slide 12

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6}$$

Random variables vs. sample space

- Rules for event probabilities apply with distributions
  - E.g. Random variables X and Y are independent if

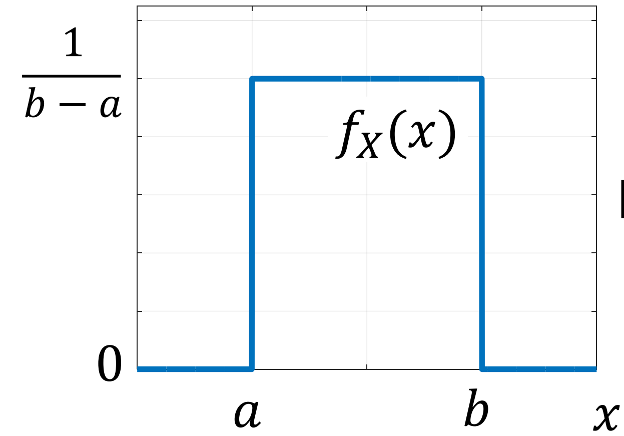
$$P(X = x \text{ and } Y = y) = P(X = x)P(Y = y) \text{ for all } x, y$$

- Models are often built by directly defining distributions (C/PDFs) of random variables rather than starting with a sample space
  - E.g., alternative models of a die:
    1. Sample space is S={1,…,6} and its probability measure P(s)=1/6
    2. PDF is given by $f_X(t) = \frac{1}{6}, t = 1, \dots, 6$, and $f_X(t) = 0$ elsewhere

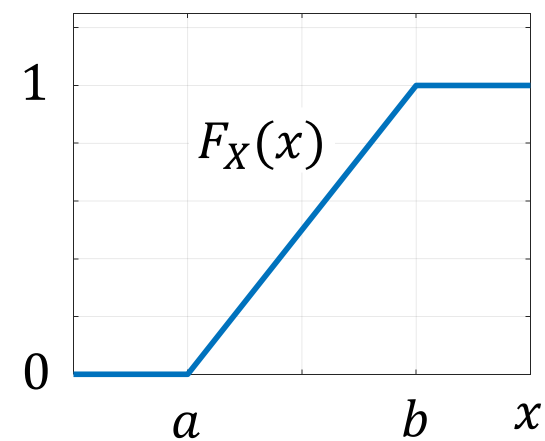
- There is a vast literature on distributions for discrete and continuous random variables (e.g. Wikipedia)
  - Some examples on the next slides

Uniform Distribution

- X follows the uniform distribution with parameters a and b: $X \sim \mathrm{UNI}(\mathrm{a}, \mathrm{b})$
  - PDF: $f_X(x) = \begin{cases} \frac{1}{b-a}, x \in [a, b] \\ 0, \text{otherwise} \end{cases}$
  - CDF: $F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$
  - The expected value of X is $\mathrm{E}(X) = \frac{a+b}{2}$
  - The variance of X is $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$



PDF



CDF

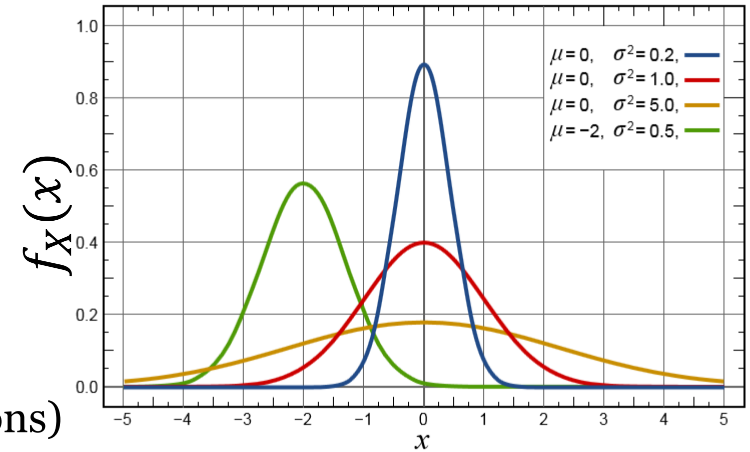
Normal distribution

- X follows the normal distribution with parameters μ and σ^2: $X \sim \mathrm{N}(\mu, \sigma^2)$
  - PDF: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
  - CDF: $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$

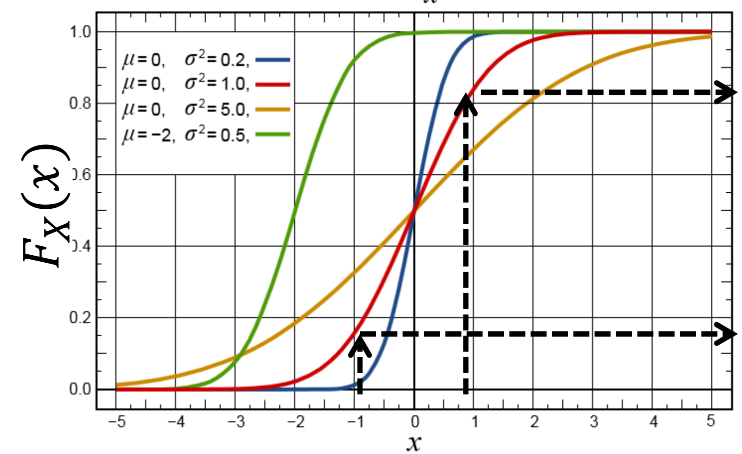
    (cannot be expressed with elementary functions)
  - The expected value of X is $\mathrm{E}(X) = \mu$
  - The variance of X is $\mathrm{Var}(X) = \sigma^2$

- For any normal distribution

$$P(\mu - \sigma \le X \le \mu + \sigma) \approx 0.68$$
$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.95$$
$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.997$$



PDF



0.84

CDF

0.16

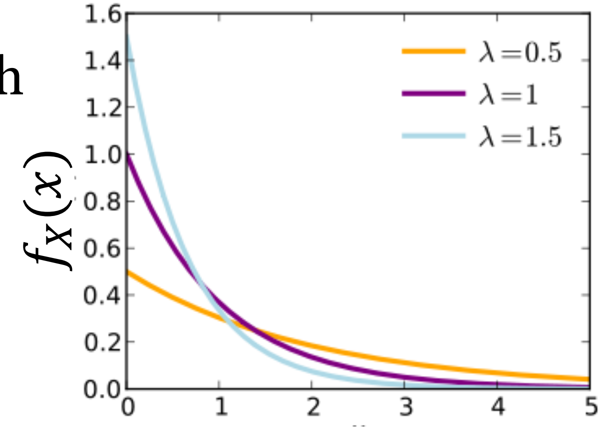
0.84-0.16=0.68

Exponential distribution

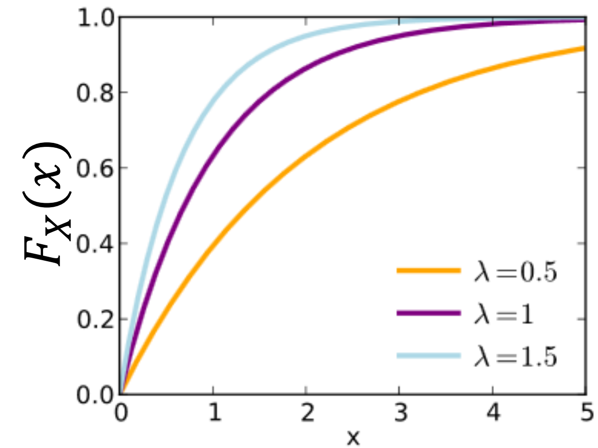
- X follows the Exponential distribution with parameter λ: $X \sim \mathrm{Exp}(\lambda)$
  - PDF: $f_X(x) = \lambda e^{-\lambda x}$
  - CDF: $F_X(x) = 1 - e^{-\lambda x}$
  - The expected value of X is $\mathrm{E}(X) = 1/\lambda$
  - The variance of X is $\mathrm{Var}(X) = 1/\lambda^2$
- Exponential distribution has no memory:

$$P(X > x + y | X > y) = P(X > x)$$

- Question:
  - Time between customer arrivals in minutes follows Exp(0.5)
  - What is the probability that a new customer arrives at most 4 min after the previous one?



PDF

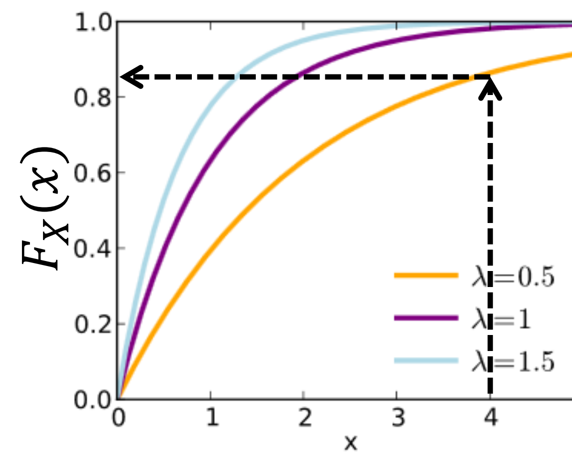


CDF

Exponential distribution

- Time between customer arrivals in minutes follows exp(0.5)
- What is the probability that a new customer arrives at most 4 min after the previous one?

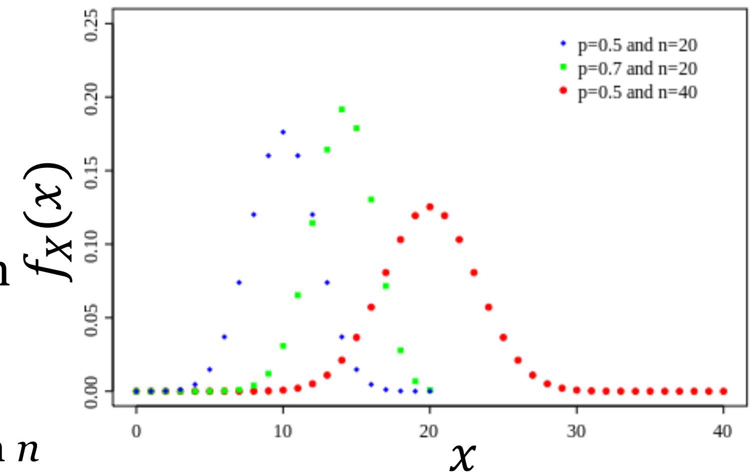
$$P(X \leq x) = F_X(x) = 1 - e^{-\lambda x}$$
$$P(X \leq 4) = F_X(4) = 1 - e^{-0.5 \cdot 4} \approx 0.86$$



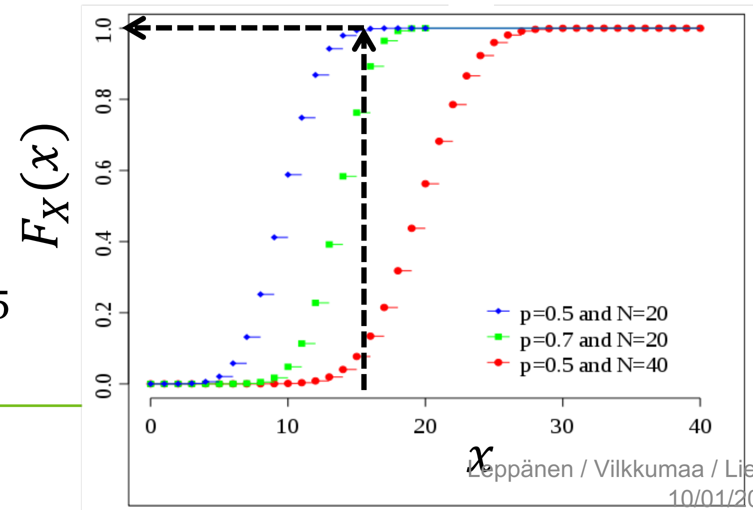
CDF

Binomial Distribution

- n independent binary (0/1, yes/no) experiments, each with the success probability p
- If X is the number of success then it follows the binomial distribution with parameters n and p,
  - Denoted $X \sim \text{Bin}(n, p)$.
  - The probability of getting exactly t successes in n trials $P(X = x) = f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$
  - The expected value of X is $\text{E}(X) = np$
  - The variance of X is $\text{Var}(X) = np(1\text{-}p)$
- Question:
  - What is the probability of getting more than 15 tails when tossing a coin 20 times?



PDF



CDF

http://commons.wikimedia.org/wiki/File%3ABinomial_distribution_pmf.svg
http://commons.wikimedia.org/wiki/File%3ABinomial_distribution_cdf.svg
y Tayste (Own work) [Public domain], via Wikimedia Commons

Binomial distribution

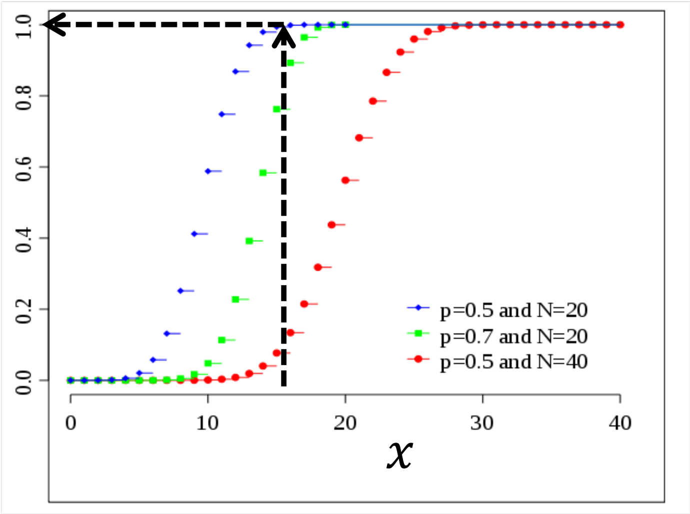
What is the probability of getting more than 15 tails when tossing a coin 20 times?

$$P(X = x) = f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$P(X > 15) = 1 - P(X \le 15)$$

$$= 1 - \sum_{t=0}^{15} \binom{20}{t} 0.5^t (1-0.5)^{20-t}$$

$$= 1 - \sum_{t=0}^{15} \binom{20}{t} 0.5^{20} = 1 - 0.5^{20} \sum_{t=0}^{15} \binom{20}{t} \approx 0.0059$$



CD

In Excel, $\binom{n}{x}$ is computed using function = COMBIN(n;x)

Binomial distribution?

Question: A bag contains 100 balls, of which 25 are green and 75 are red. Four balls are drawn from the bag such that after each draw, the ball is placed back in the bag. What is the probability of drawing at most one red ball?

$$P(X \leq 1) = \sum_{t=0}^{1} \binom{4}{t} 0.75^t 0.25^{4-t} = \binom{4}{0} 0.25^4 + \binom{4}{1} 0.75 \cdot 0.25^3 \approx 0.05078$$