Introduction to
Data Structures and Algorithms

Chapter: **Probabilistic Analysis & Randomized Algorithms**

Friedrich-Alexander-Universität
Erlangen-Nürnberg

Lehrstuhl Informatik 7 (Prof. Dr.-Ing. Reinhard German)
Martensstraße 3, 91058 Erlangen

The hiring problem

- Suppose you need to hire a new office assistant: the employment agency sends you the candidates
  - You interview that person – result either to hire or not
  - The employment agency gets a small fee for applicant interview
  - Actually hiring an applicant is much more costly (to fire the current office assistant and to pay a hiring fee to the agency)

- Aim: at all times to have the best possible person for the job
  - Your decision after interviewing each applicant: is that applicant better qualified than the current office assistant
  - If so, fire the current and hire the new applicant

- Wish: estimation of the price of this strategy
  - Assume the candidates are numbered 1 through n

(this scenario – a model for computational paradigms)

The hiring problem

```
Hire-Assistant(n)

best := 0      // ▹ Candidate 0 is a least-
                  qualified dummy candidate
for i:= 1 to n
  do interview candidate i
     if candidate i is better than cand. best
        then best := i
             hire candidate i
```

The hiring problem

- **<u>Cost model</u>** here differs from models of previous chapters
  - Important is not the running time T(n) of algorithm Hire-Assistant (H-A), but the costs caused by interviewing and hiring
  - The cost of H-A may seem different from T(n) of Merge sort e.g., the analytical methods, however, are identical whether analyzing cost or running time: important - always the number of times of basic operations is estimated
- Be c_i - the low cost of interviewing and c_h - the much higher cost of hiring
  - n - the number of interviewed people
  - m - the number of people hired $\Rightarrow$ total cost of algorithm H-A is $O(c_i n + c_h m)$
  - The cost $c_i n$ is fixed, but $c_h m$ varies with each run of algorithm H-A: therefore concentration on analyzing $c_h m$

The hiring problem

- **<u>Worst case analysis</u>**
  - Every candidate is hired that we interview, i.e. the candidates come in increasing order of quality $\Longrightarrow$ we hire n times with
  - Total hiring cost of $O(n \cdot c_h)$
  - In fact, there is no idea about the order of quality of arrived candidates, nor do we have any control over this order

- Our interest: what can we expect to happen in a typical or <u>average</u> case ??

The hiring problem

- <u>**Probabilistic analysis**</u>
  - Use of probability in the analysis of problems
  - To analyze the running time T(n) of an algorithm or other quantities like hiring cost of H-A
  - Based on probabilistic distribution of the inputs we compute an *average-case running time* (taking the average over the distribution of the possible inputs)
- For hiring problem: assume the applicants come in random order
  - We can compare any two candidates an decide which one is 'better' $\Rightarrow$ ranking with a unique number from *1* to *n*: *rank(i)* to mark the rank of applicant *i*
  - Convention: a higher rank means a better qualified applicant
  - $\Rightarrow$ ordered list [*rank(1), rank(2, …, rank(n)*] as permutation of the list (1, 2, …, n)
  - <u>Random order</u> of applicants: each of the $n!$ permutations is equally likely $\Rightarrow$ the ranks form a **uniform random permutation,** permutations with equal probability

The hiring problem

Indicator random variables

- For analysing many algorithms – including the hiring problem – we use so-called Indicator random variables

- Aim: Easy conversion between probabilities and expectations

- Given is a probability space (S, Φ, P) with event $A \in \Phi$. Then Indicator random variable $I\{A\}$ w.r.t. event A is defined:

$$I\{A\} = \begin{cases} 1 & \text{if} \quad \text{A occurs} \\ 0 & \text{if} \quad \text{A does not occur} \end{cases}$$

- <u>Expl</u>: Expected number of heads when flipping a fair coin

fair coin: head H / tail T $\Rightarrow \quad S = \{H, T\} \quad \text{with} \quad \mathrm{P}\{H\} = \mathrm{P}\{T\} = \frac{1}{2}$

Indicator random variables

- We define the indicator random variable (r.v.) X_H

$$X_H = I\{H\} = \begin{cases} 1 & \text{if H} \\ 0 & \text{if T} \end{cases} \quad \Rightarrow$$

The expected number of heads is simply the expected value of indicator variable $X_H : E[X_H] = E[I\{H\}] = 1 \cdot P\{H\} + 0 \cdot P\{T\} = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}$

- Lemma: Given a sample space S and an event A in sample space S.
  Let $X_A = I\{A\}$. Then $E[X_A] = P\{A\}$.

[Proof: By definition of indicator r.v. we get: $E[X_A] = E[I\{A\}] = 1 \cdot P\{A\} + 0 \cdot P\{\overline{A}\} = P\{A\}$,
where $\overline{A} \in \Phi$ denotes event $S \setminus A$, the complement of A (the event “not A”)]

Indicator random variables

- **Extended Expl**: Be X_i - indicator r.v. to event in which the *i*-th flip comes up heads:

  $X_i = I\{\text{the } i\text{-th flip results in the event } H\}$. Let

  X - r.v. denoting the total number of heads in case of *n* times to flip the coin $X = \sum_{i=1}^{n} X_i$

What is the expected number of heads?

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right]$$

By linearity of expectation we get:

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \tfrac{1}{2} = \tfrac{n}{2}$$

The hiring problem

Indicator random variables

- **Analysis of the hiring problem** by indicator random variables
- Aim: computation of the expected number of times that we hire a new office assistant

Assumption: candidates arrive in a random order

Let X - r.v. to describe the number of times of hiring a new applicant

$$\Rightarrow \quad E[X] = \sum_{x=1}^{n} x \cdot P\{X = x\}.$$

Let X_i - Indicator r.v. with:

$$X_i = I\{\text{candidate } i \text{ is hired}\} = \begin{cases} 1 & \text{if} \quad \text{cand. i is hired} \\ 0 & \text{if} \quad \text{cand. i is not hired} \end{cases}$$

$$\Rightarrow \quad X = X_1 + \cdots + X_n$$

- <u>By Lemma:</u>

$\Rightarrow$ $E[X_i] = P\{\text{candidate } i \text{ is hired}\}$, that means, we compute the probability, that lines 5 - 6 of Hire-Assistant algorithm (H-A) are executed

- Candidate *i* is hired in line 5: he is better than each of candidates $1 \cdots i-1$. All candidates arrive in random order $\Rightarrow$ any one of these first *i* candidates is equally likely to be best-qualified so far

- Candidate i has probability $1/i$ of being better than *1* through $i-1$ and thus probability of $1/i$ of being hired

- <u>By Lemma:</u>

$$E[X_i] = 1/i \quad \Rightarrow$$

$$E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} 1/i = \ln n + O(1)$$

- <u>Resumė:</u>
  - Even though we interview *n* people, we only actually hire approximately $\ln n$ of them, on average $\Rightarrow$
  - If candidates come in random order, algorithm Hire-Assistant has a **total hiring cost** of $O(c_h \cdot \ln n)$

($\Rightarrow$ The average-case hiring cost is a significant improvement over the worst-case hiring cost of $O(c_h \cdot n)$)

- **<u>Randomized algorithms</u>**
  - Case: No knowledge of the distribution on the inputs $\Rightarrow$ average-case analysis is not possible, but <u>randomized algorithms:</u>
  - Probability and randomnes as tools for algorithm itself by making the behavior of part of the algorithm random
  - Regarding Hiring problem: independent on any input (-distribution) of applicants the <u>randomized Hire-Assistant algorithm</u> (r. H-A) impose a distribution of inputs as
  - <u>First action</u> of algorithm: randomly permute the candidates to enforce the property, that every permutation is equally likely

- For this algorithm (and other randomized algorithm): *no particular input elicits its worst-case behavior*

The hiring problem

```
Randomized-hire-Assistant(n)

Randomly permute the list of candidates
best := 0      // ▻ Candidate 0 is a least-
                   qualified dummy candidate
for i:= 1 to n
  do interview candidate i
    if candidate i is better than cand. best
      then best := i
           hire candidate i
```

The expected hiring cost of **Randomized Hire-Assistant** is $O(c_h \cdot \ln n)$

Probabilistic Analysis & randomized algorithms

- Distinction between probabilistic analysis and randomized algorithms

  - **Probabilistic Analysis**: probability distribution of the input data
  - Mostly, probability analysis to analyze the running time of an algorithm:
  - Averaging the running time over all possible inputs
  - The algorithm itself is deterministic
  - Speech: **average-case running time**

  - **Randomized algorithm**: randomization in the algorithm, not in input
  - no assumption about the input
  - Running time: expectation of running time over the distribution of values produced by random-number generator (RANDOM)
  - Speech: **expected running time**

- **Random-number generator** RANDOM
  - RANDOM(a,b) returns an integer between *a* and *b*
  - Each integer with equal probability
  - <u>Example:</u> RANDOM(0,1) produces both *0* and *1* with probability 1/2
    RANDOM(3,7) returns 3,4,5,6 or 7 with with probability 1/5

  - Each integer given by RANDOM is independent of integers returned on previous calls
  - Imagine: RANDOM as rolling of a (*b*-*a*+*1*)-sided die to get its output

- In practice: most programming environments use a ***pseudorandom-number generator***: a deterministic algorithm returning numbers that "look" statistically random

Quick Sort

Pseudo code for Randomized Quick Sort

```
Randomized-Quicksort (A,p,r)
  if p < r   then
     q := Randomized-Partition (A,p,r)
     Randomized-Quicksort (A,p,q-1)
     Randomized-Quicksort (A,q+1,r)

Randomized-Partition (A,p,r)
     i := Random (p,r)
     exchange A[r] ⟷ A[i]
     return Partition (A,p,r)
```

- **Now:** Expected running time $T(n) = \Theta(n \cdot \lg n)$

Basics of Probability Theory

Goal

- recall the basic concepts of probability theory

Contents

- Randomness and Probability
- Random Variables and their Distributions
- Moments and Quantiles
- Some Distributions
- Dependence of Random Variables

Why probability theory (and statistics) ?

- often random input in simulation application areas and analytical models
  - manufacturing: processing times, machine failure/repair times,...
  - communications: interarrival times of messages, packet sizes,...
  - ...
- output analysis
  - statistical methods for random simulation output

Why Probability Theory and Statistics ?

- Characterization of single random quantities
  - random variables
  - probability distribution
  - expectation, variance, quantiles, ...
  - dependence on other random variables
- Simulation and analytical model
  - often is a stochastic process
  - allows mathematical characterization
  - sometimes possible to analyze
- Statistical methods
  - to find probability distributions and their parameters (input modeling)
  - to generate random numbers/variates during simulation run (random number generation)
  - to analyze simulation output (output analysis)

Randomness and Probabilities

- Probability theory
  - concerned with the study of random phenomena
  - not predictable in a deterministic fashion
  - mathematical descriptions to deduce patterns of future outcomes
- Random experiments and their outcomes
  - random experiment: a process whose outcome is not known with certainty (properties: reproducible with same possible outcomes)
  - sample space: the set S of all possible outcomes of a random experiment
  - sample point or elementary event: a possible single outcome of a random experiment, an element of S
  - event: a set A of elementary events, a subset of S
- Examples
  - toss of a die
  - time to failure of a hard disk/machine

Randomness and Probabilities

- Intuitive interpretations of probability
  - P(A) denotes the probability of an event A
  - a measure of how likely a performance of the random experiment results in an elementary event in A
  - relative-frequency interpretation
    - repeat the experiment a large number of n times
    - count the number m of occurrences of elementary events in A
    - $P(A) \sim m/n$
    - experience: the quotient fluctuates less for increasing n
  - interpretation based on equiprobability of elementary events
    - if S finite: $m = |A|$, $n = |S|$, then $P(A) = m/n$
    - $|\cdot|$ denotes the cardinality of a set
  - intuitive interpretations sufficient for most engineering applications
    - some subtle mathematical difficulties lead to paradoxes with this interpretation

⇨ alternative: axiomatic definition of probability

Randomness and Probabilities

- Axiomatic definition of probability
  - by Kolmogorov 1933
  - probabilities, i.e., real numbers, can be assigned to events so as to satisfy the three basic axioms of probability:
    1. for any event A: $P(A) \geq 0$
    2. $P(S) = 1$ (the universal event has probability one)
    3. $P(A \cup B) = P(A) + P(B)$, whenever A and B are disjoint, i.e., when $A \cap B = \varnothing$
       for infinite sample spaces and disjoint events: $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$
  - consistent with our intuition
  - axioms of probability allow to derive a number of calculation rules (together with conventional set theory)
  - to avoid mathematical difficulties:
    only those events can be considered, which are "measurable" in the sense of measure theory

Randomness and Probabilities (cont.)

- Probability systems (S, Φ, P)
  - S sample space
  - Φ is a Borel field of subsets of S
    - $\Phi \subseteq 2^S$, where power set 2^S is the set of all subsets
    - example: $S = \{x, y, z\}$

      $2^S = \{\emptyset, \{x\}, \{y\}, \{z\}, \{x,y\}, \{x,z\}, \{y,z\}, \{x,y,z\}\}$

      $\Phi = 2^S$ (in this example)
  - P is a probability measure on Φ satisfying the three axioms of probability
    - example: assume all events in S are equiprobable

      $P: \Phi \rightarrow [0,1]$, where

      $P(\emptyset) = 0$

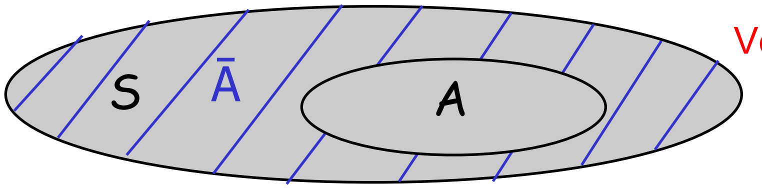
      $P(\{x\}) = P(\{y\}) = P(\{z\}) = 1/3$

      $P(\{x,y\}) = P(\{x,z\}) = P(\{y,z\}) = 2/3$

      $P(\{x,y,z\}) = 1$
  - we do not need to enter measure theory in more detail

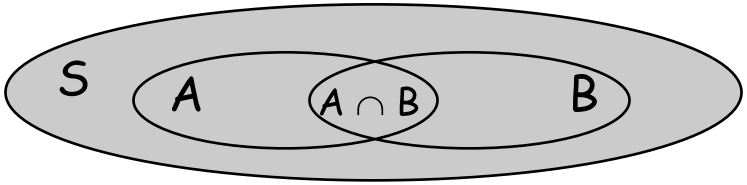
Randomness and Probabilities

- Two rules
  - $P(\emptyset) = 0$ (impossible event)
  - $P(\bar{A}) = P(S \setminus A) = 1-P(A)$ (complementary event)



Venn diagram

- Conditional probability
  - P(A|B) = the probability of event A, given that event B has occurred
  - important: $P(B) > 0$!
  - if A occurs on the condition B, we have the additional information that the outcome of this random experiment is contained in subset B:

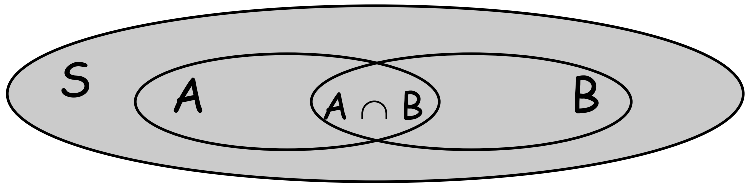


  - intuitively: event B plays now the role of the sample space
  - $P(A|B) = P(A \cap B)/P(B)$

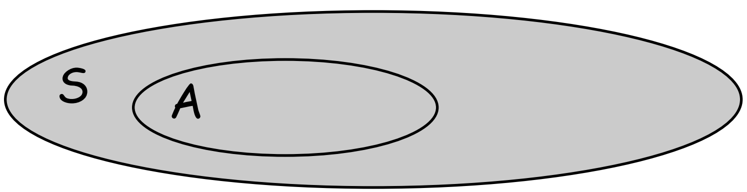
Randomness and Probabilities

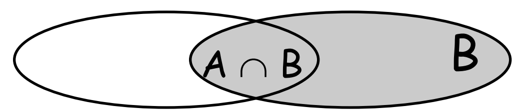
Independence

- two events A and B are independent if $P(A \mid B) = P(A)$



- intuitively: the relation of the areas of A and S is the same as the relation of the areas of $A \cap B$ and B





- an equivalent criterion: $P(A \cap B) = P(A) \cdot P(B)$

Random Variables and their Distributions

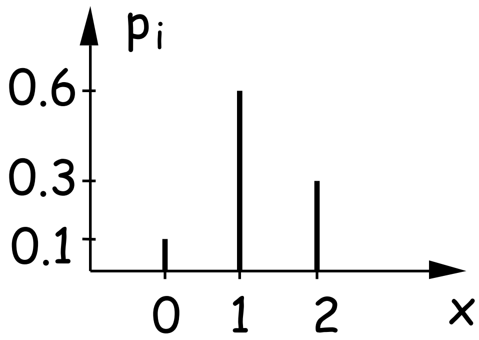
- From sample spaces to random variables
  - need a more compact representation than the sample space and its elementary events
  - random variable: a function $X: S \rightarrow \mathbb{R}^+_0$, that assigns a real number X(s) to each possible elementary event or sample point $s \in S$
  - the term “random variable” is thus misleading
  - convention: capital letters for the random variable and lowercase letters for their values
  - example: rolling a pair of dice
    - S = {(1, 1), (1, 2), ..., (6, 6)}
    - (i, j) means that i appeared on the first and j on the second die
    - X((i, j)) = i + j
  - example: time to failure of a hard disk
    - $S = \mathbb{R}^+_0$
    - X(s) = s (here X is just the identity)

Random Variables and their Distributions

- Discrete and continuous random variables (RVs)
  - the image of X is X(S), the set of values the random variable can assume
  - discrete RV: $|X(S)| \leq |\mathbb{N}_0|$
    - the random variable assumes values from a discrete set of numbers, hence the image is either finite or countable
    - example: rolling a pair of dice (finite image)
  - continuous RV: $|X(S)| > |\mathbb{N}_0|$
    - the random variable assumes values from a continuous set of numbers, hence the image is uncountable
    - example: time to failure of a hard disk
  - mixtures of discrete and continuous RVs are possible

Random Variables and their Distributions

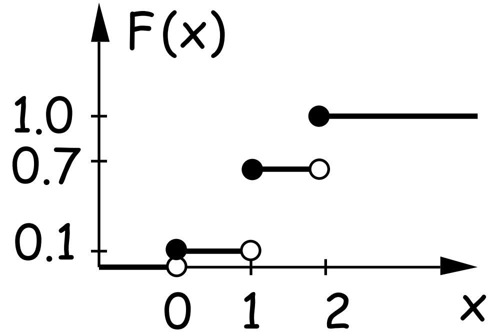
- Distribution of a discrete RV
  - let x_1, x_2, ... be the discrete values the RV can assume
  - $p_i = P(X = x_i)$ is called the probability mass function (pmf)
  - intuitive interpretation: the pmf describes how the probability mass is distributed over the different values of the RV
  - example:
    - $x_1 = 0$, $x_2 = 1$, $x_3 = 2$
    - $p_1 = 0.1$, $p_2 = 0.6$, $p_3 = 0.3$



Random Variables and their Distributions

Distribution of a discrete RV (cont.)

- (cumulative) distribution function (CDF): $F(x) = P(X \le x) = \sum_{x_i \le x} p_i$
- F(x) is a step function jumping with height p_i at the discrete values x_i of the RV
- it contains the same information as the pmf
- in the example:



Random Variables and their Distributions

- A distribution function F(x) has the following properties

1) $0 \leq F(x) \leq 1$

2) F(x) is non-decreasing: if $x_1 \leq x_2$ then $F(x_1) \leq F(x_2)$

3) $\lim_{x \to -\infty} F(x) = 0, \lim_{x \to +\infty} F(x) = 1$

4) F(x) is continuous from the right

- any function satisfying these properties is a distribution function

Random Variables and their Distributions

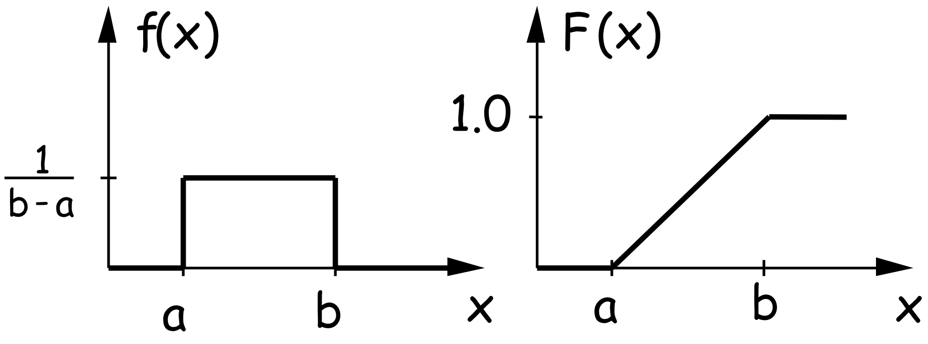
- Distribution of a continuous RV
  - the probability distribution $F(x) = P(X \leq x)$ is defined as before
  - the probability density function (pdf) is defined as its derivation:

$$f(x) = \frac{d}{dx} F(x)$$

  - intuitive interpretation: the pdf describes how the probability is distributed over the different values of the RV
  - example: the uniform distribution from a to b:

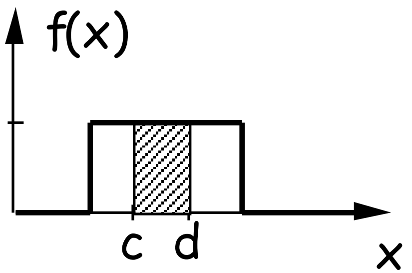
$$f(x) = \begin{cases} \frac{1}{b-a} & a < x \leq b \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & \text{otherwise} \end{cases}$$



Random Variables and their Distributions

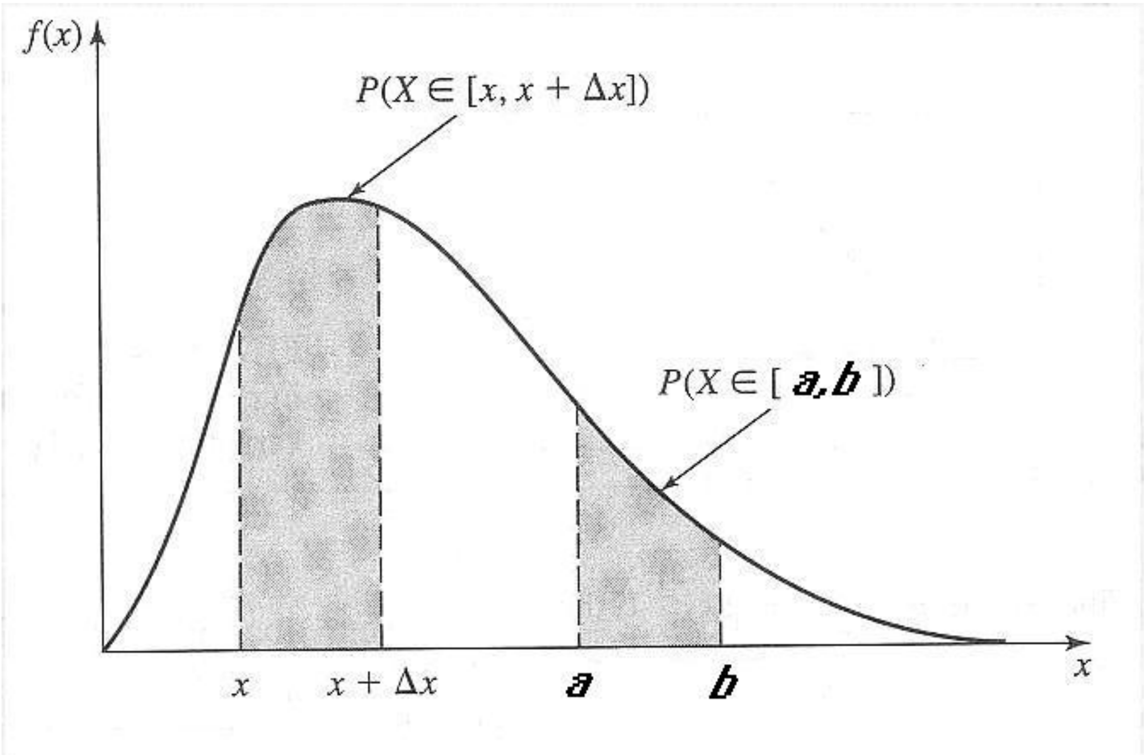
- Distribution of a continuous RV (cont.)
  - integration of a pdf yields a CDF: $F(x) = \int_{-\infty}^{x} f(y)\,dy$
  - probability $P(c < X \leq d)$ of an interval (c, d]: $\int_{c}^{d} f(y)\,dy = F(d) - F(c)$
  - example:



  - probability $P(X = a)$ of a single value: $\int_{a}^{a} f(y)\,dy = F(a) - F(a) = 0$
  - areas under f(x) are probabilities
  - f(x) is not a probability!

Random Variables and their Distributions

- Distribution of a continuous RV (cont.)
  - only the area under the f(x) in the neighborhood [x, x+dx] is a probability



  - thus x is more likely in neighborhoods where f(x) is large

Moments and Quantiles

- Need for a concise description
  - the CDF $F(x)$ or the pdf $f(x)$ (pmf p_i in the discrete case) completely characterizes the behavior of a RV
  - a function is often too complex
  - we need a simpler description: a single number or a few numbers
- Expectation
  - the expectation (or mean) $\mu = E[X]$ of a RV X is defined as

$$E[X] = \begin{cases} \sum_{i=1}^{\infty} x_i p_i & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

provided the sum or integral converges absolutely (otherwise the expectation does not exist)

Moments and Quantiles

- Expectation (cont.)
  - intuitive interpretation: a measure of central tendency in the sense that it is the center of gravity
  - example (discrete case)
    - $x_1 = 0$, $x_2 = 1$, $x_3 = 2$
    - $p_1 = 0.1$, $p_2 = 0.6$, $p_3 = 0.3$
    - $E[X] = 0 \cdot 0.1 + 1 \cdot 0.6 + 2 \cdot 0.3 = 1.2$
  - example (continuous case)
    - uniform distribution from a to b
    - $$E[X] = \int_a^b \frac{x}{b-a} dx = \left. \frac{x^2}{2(b-a)} \right|_a^b = \frac{a+b}{2}$$

Moments and Quantiles

Expectation (cont.)

- two properties of the expectation operator, needed for other derivations:
  - linearity of the expectation: $E[aX + bY] = aE[X] + bE[Y]$
  - function of a RV: let $Y = g(X)$, then

$$E[Y] = E[g(X)] = \begin{cases} \sum_{i=1}^{\infty} g(x_i) p_X(i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f_X(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

    (where $p_X(i)$ / $f_X(x)$ are the pmf / density of RV X)

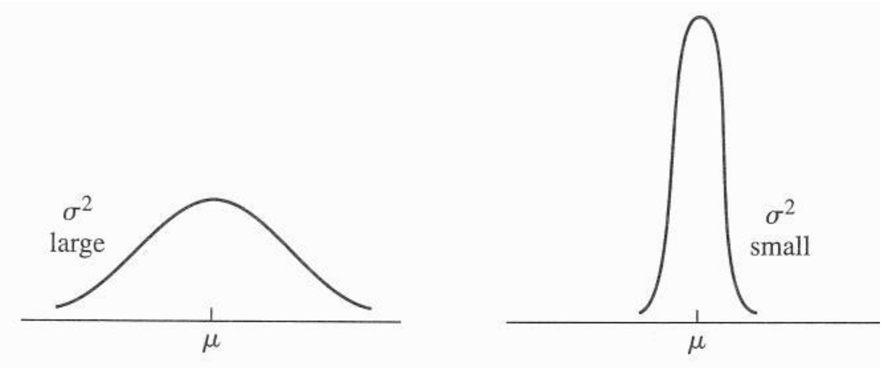
  - what about $E[X \cdot Y]$?

    $E[X \cdot Y] = E[X] \cdot E[Y]$, only if X and Y are independent

    (for independence of RVs see slide 33)

Moments and Quantiles

- Variance
  - the variance σ^2 = Var[X] of a RV X is defined as Var[X] = E[(X-E[X])^2], provided it exists
  - the expectation of the square of the deviation from the mean or the "second central moment"
  - intuitive interpretation: a measure of the dispersion of a RV about its mean; the larger the variance, the more likely the RV is to take on values far from its mean
  - σ is known as the standard deviation
  - $C_X = \sigma/E[X]$ is the coefficient of variation, a normalized measure



Moments and Quantiles

- Variance (cont.)
  - with the linearity of the expectation we can derive

$$\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]^2 + E[X]^2$$
$$= E[X^2] - E[X]^2$$

  - in the discrete example:
    $\mathrm{Var}[X] = 0^2 \cdot 0.1 + 1^2 \cdot 0.6 + 2^2 \cdot 0.3 - 1.2^2 = 0.36$
  - in the uniform example:

$$\mathrm{Var}[X] = \int_a^b \frac{x^2}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2 = \cdots = \frac{(b-a)^2}{12}$$

  - properties of the variance:
    - $\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$
    - $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$, if X and Y are uncorrelated

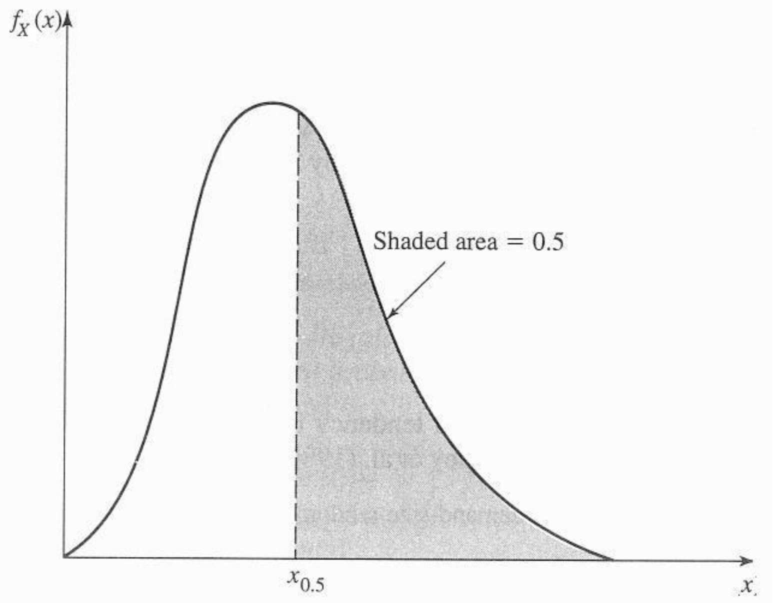
(for uncorrelated RVs see slide 34)

Moments and Quantiles

- Moments
  - n'th moment is $E[X^n]$, $n \geq 1$
  - n'th central moment is $E[(X-E[X])^n]$
  - first moment is the expectation
  - second central moment is the variance
  - third central moment allows to define the skewness : $\nu = E[(X-E[X])^3] / \sigma^3$
    a measure of symmetry
    - $\nu = 0$ for symmetric distributions as the normal or uniform
    - $\nu < 0$: skewed to the left; $\nu > 0$: skewed to the right
  - fourth central moment allows to define kurtosis : $\eta = E[(X-E[X])^4] / \sigma^4$
    a measure of the tail weight
    - $\eta = 3$ for normal distribution
    - $\eta < 3$: platykurtic; $\eta > 3$: leptokurtic (more peaked in center, fatter tails)
  - hard to find interpretations for the higher moments (with larger n)
  - a distribution can also be represented by the sequence of its moments (if they exist)

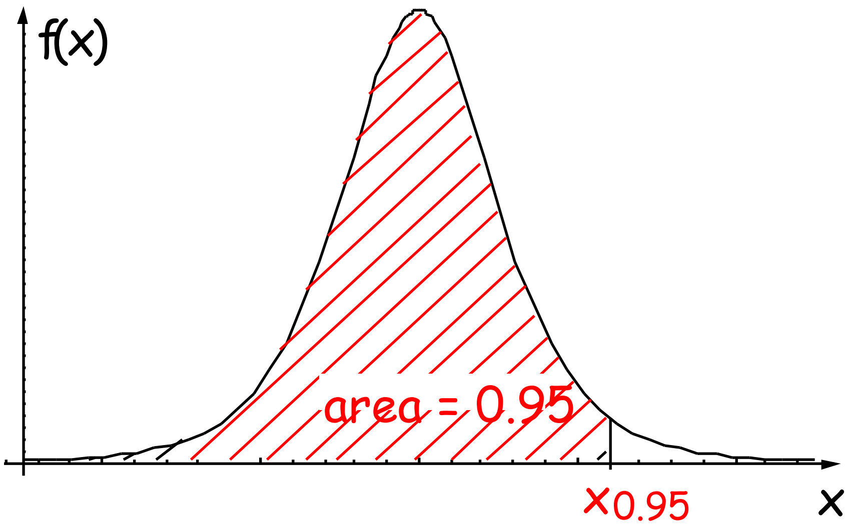
Moments and Quantiles

- Median
  - the median is the smallest value $x_{0.5}$ such that $F(x_{0.5}) \geq 0.5$
  - an alternative measure of central tendency
  - may be better when X can assume extreme values, since they can greatly affect the mean even if they are unlikely to occur



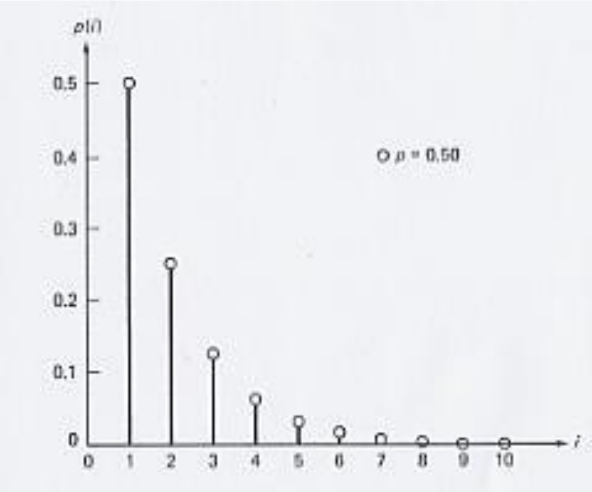
Moments and Quantiles

- Quantile
  - for $0 < q < 1$, the q-quantile is the smallest value x_q such that $F(x_q) \geq q$
    - median for q=0.5: median is 0.5-quantile
    - quartiles for q=0.25 or q=0.75
    - octiles for q=0.125 or q=0.875
  - when X is continuous and F(x) is strictly increasing for $0 < F(x) < 1$: $F(x_q) = q$, $x_q = F^{-1}(q)$
  - quantiles also called fractiles or percentiles
  - quantiles are not probabilities! (found on x-axis)
  - for distribution of discrete RV: the quantile must also be one of the discrete values x_1, x_2, ... that the RV can assume



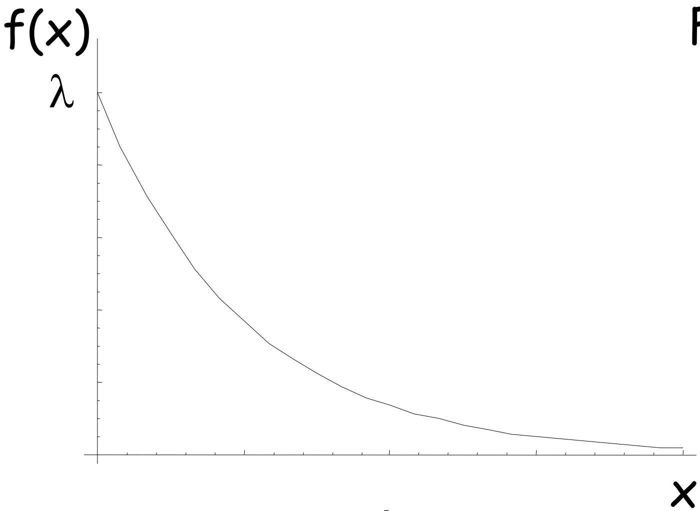
Some Distributions

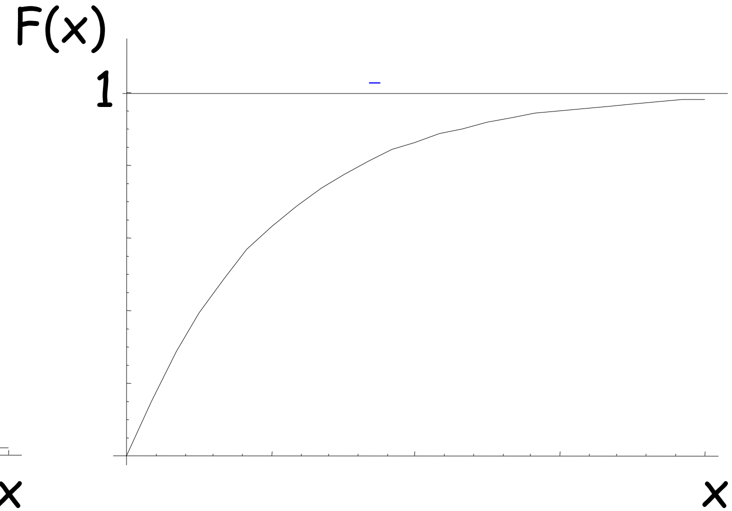
- Geometric distribution
  - Bernoulli trial
    - a random experiment with two possible outcomes: success or failure
    - the probability of success is p
    - example: tossing a die
      success: obtaining a six, failure: the other numbers → p = 1/6
  - new experiment: repeat Bernoulli trials until the first success
  - discrete RV X: the number of trials
  - pmf: $p_i = (1-p)^{i-1} p, \; i = 1, 2, \ldots$
  - CDF: $F(i) = \Sigma_{j=1}^{i} p_j = \Sigma_{j=1}^{i} (1-p)^{j-1} p = \ldots = 1 - (1-p)^i$
  - expectation and variance:
    - $E[X] = \Sigma_{j=1} j p_j = p\, \Sigma_{j=1} j (1-p)^{j-1} = \ldots = 1/p$
    - $Var[X] = E[(X-E[X])^2] = \ldots = (1-p)/p^2$



Some Distributions

- Exponential distribution
  - pdf: $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$ CDF: $F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$





  - one parameter: the rate λ
  - expectation: $E[X] = \int_0^\infty x\lambda e^{-\lambda x}\,dx = \cdots = 1/\lambda$
  - variance: $Var[X] = \int_0^\infty (x - 1/\lambda)^2 \lambda e^{-\lambda x}\,dx = \cdots = 1/\lambda^2$

Some Distributions

- The exponential distribution is memoryless
  - the memoryless property: $P(X \le y + z \mid X > y) = P(X \le z)$
  - interpretation
    - let X be the time to failure of a system
    - given the system has not failed until time y, the probability that the system fails in the coming z time units, i.e., until time y + z,
      is the same as
      the probability that the system fails until time z starting at time 0
      (thus, the memory y plays no role)
  - proof for the exponential distribution:

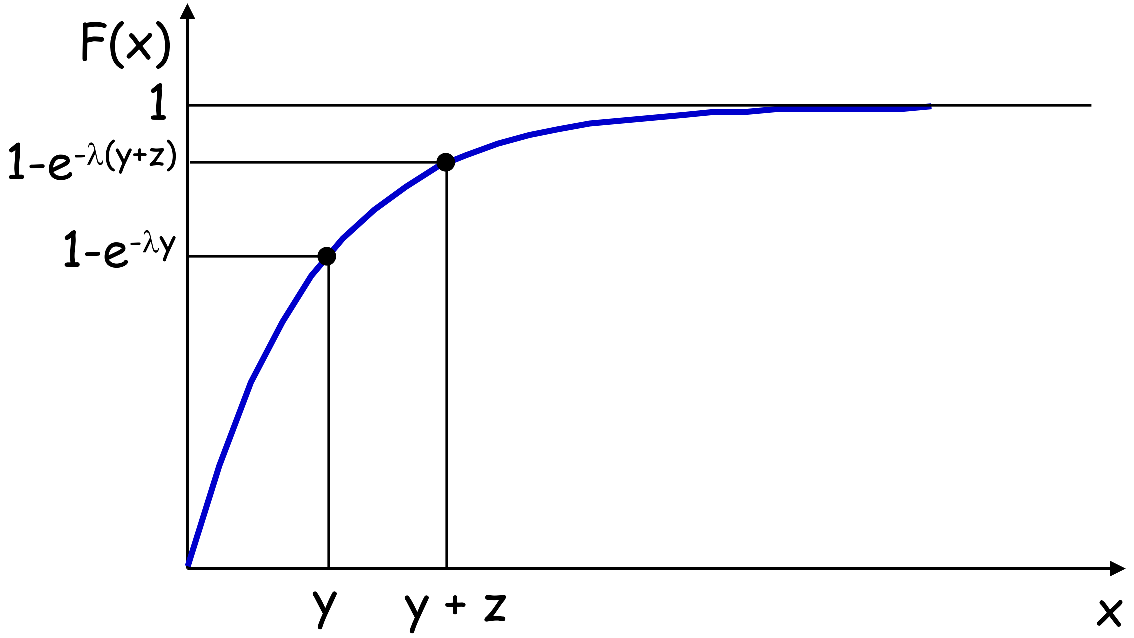
def. of conditional probability

$$P(X \le y + z \mid X > y) = \frac{P(y < X \le y + z)}{P(X > y)} = \frac{F(y+z) - F(y)}{1 - F(y)}$$

$$= \frac{1 - e^{-\lambda(y+z)} - (1 - e^{-\lambda y})}{e^{-\lambda y}} = 1 - e^{-\lambda z} = P(X \le z)$$

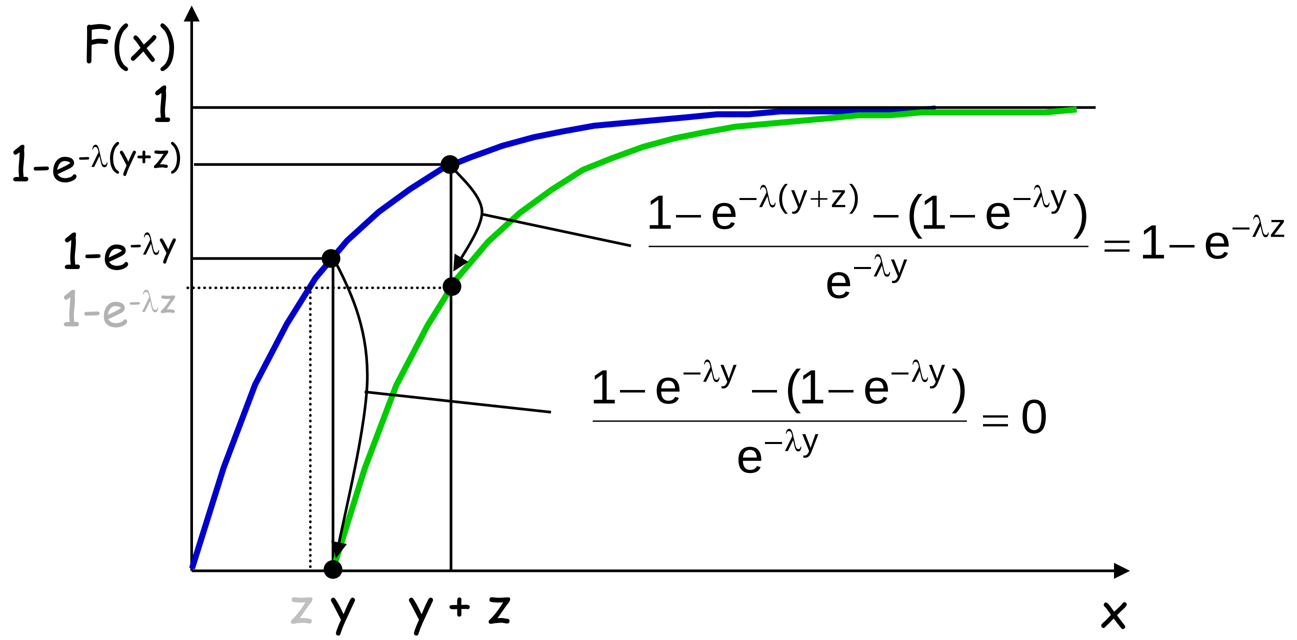
Some Distributions (cont.)

- Illustration of the memoryless property
  - the curve of the exponential distribution function starting at zero with the corresponding values at y and y + z:



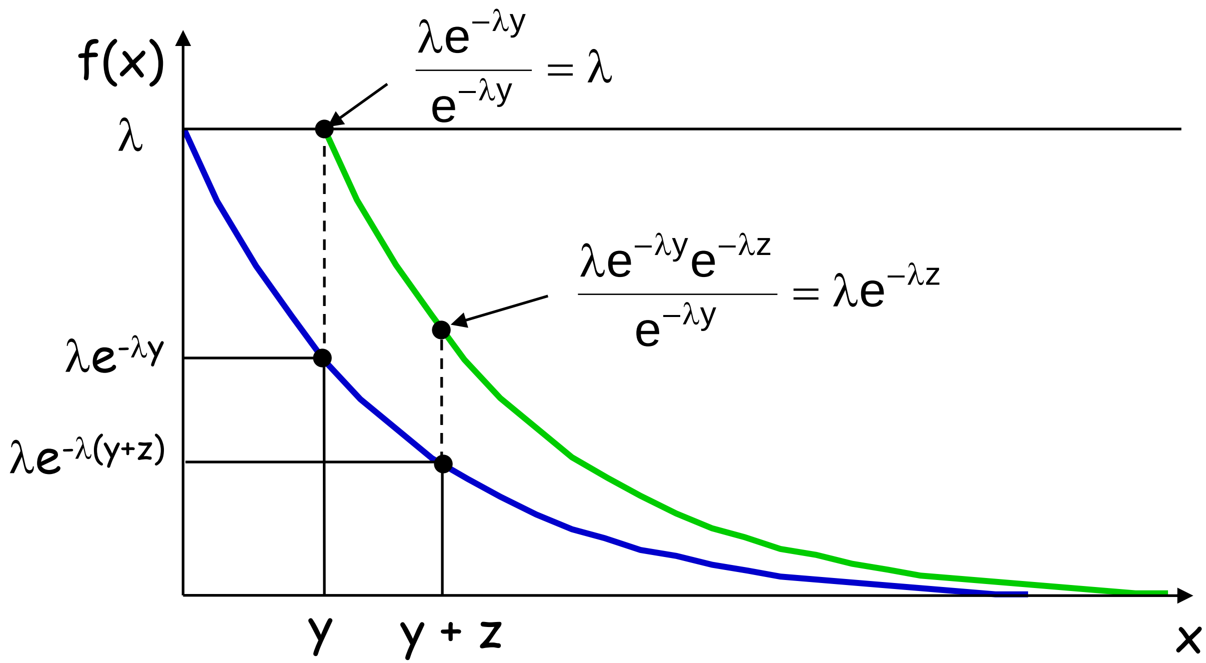
Some Distributions

- Illustration of the memoryless property (cont.)
  - given that X > y, subtract $1-e^{-\lambda y}$ and divide by $e^{-\lambda y}$, the result is the same curve shifted to the right by y:



Some Distributions

- Illustration of the memoryless property (cont.)
  - equivalently, scaling the density with the factor $e^{-\lambda y}$ for all values equal to or greater than y leads to the same curve shifted to the right by y:



Some Distributions

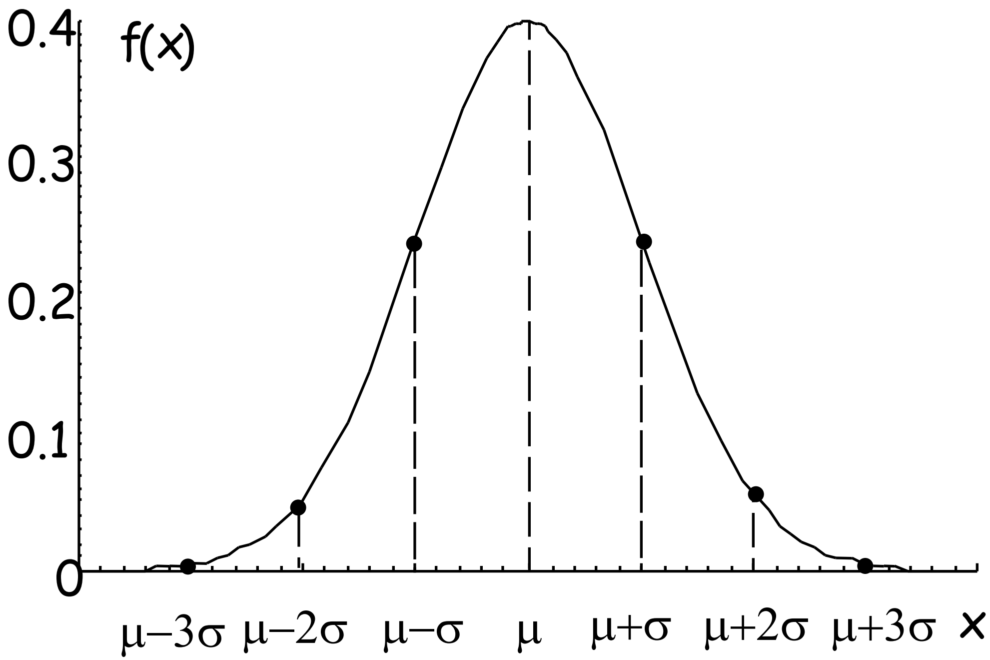
- The exponential distribution is memoryless (cont.)
  - it can be shown that the exponential distribution is the only continuous distribution which is memoryless
  - analogously, the geometric distribution is the only discrete distribution which is memoryless

  - memoryless property (of exponential distribution) accounts for high tractability in analysis
  - Markovian systems are built from exponential phases:
    - phase-type distributions (dense class of probability distributions)
    - Markovian arrival processes (with correlated interarrival times)
    - Continuous-Time Markov Chains (CTMC)
    - Markovian queues

    (more details in Chapter *Analytical Modeling*)

Some Distributions

- Normal distribution
  - pdf: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
  - μ is the mean
  - σ^2 is the variance
  - bell-shape curve
  - notation: $X \sim N(\mu, \sigma^2)$
  - F(x) has no closed form
  - $Z \sim N(0, 1)$, the standard normal distribution is recorded in tables
  - relation: $F_X(x) = F_Z(\frac{x-\mu}{\sigma})$
  - $P(|x-\mu| < \sigma) = 0.683$, $P(|x-\mu| < 2\sigma) = 0.955$, $P(|x-\mu| < 3\sigma) = 0.997$
  - common for describing measurement errors
  - common for quantities that are the sum of a large number of other quantities, therefore it plays a central role in statistics (central limit theorem)



Dependence of Random Variables

- Joint distribution functions
  - Let X and Y be two RVs
  - $F(x, y) = P(X \le x, Y \le y)$ is the joint distribution function
  - marginal distribution functions:

$$\lim_{x\to\infty} F(x,y) = F_Y(y), \quad \lim_{y\to\infty} F(x,y) = F_X(x)$$

  - X and Y are independent if $F(x, y) = F_X(x)\, F_Y(y)$ for all x,y
  - analogous definitions of independence can be given based on the pmf (in the discrete case) or the pdf (in the continuous case):
    $p(x, y) = p_X(x)\, p_Y(y)$, $f(x, y) = f_X(x)\, f_Y(y)$

Relationship between Random Variables

- Measures of dependence between two RVs X_i and X_j
  - covariance $C_{ij} = E[(X_i-\mu_i)(X_j-\mu_j)] = E[X_iX_j] - \mu_i\mu_j$
  - if X_i and X_j are independent, $C_{ij} = 0$, the converse is not true in general
  - if $C_{ij} > 0$, X_i and X_j are positively correlated ($X_i > \mu_i$ and $X_j > \mu_j$ tend to occur together and $X_i < \mu_i$ and $X_j < \mu_j$ tend to occur together)
  - if $C_{ij} < 0$, X_i and X_j are negatively correlated ($X_i > \mu_i$ and $X_j < \mu_j$ tend to occur together and $X_i < \mu_i$ and $X_j > \mu_j$ tend to occur together)
  - the covariance is not dimensionless, difficult to use, especially for comparing degree of dependency
  - correlation $\rho_{ij} = \dfrac{C_{ij}}{\sqrt{\sigma_i^2\sigma_j^2}}$
  - normalized to values $-1 \le \rho_{ij} \le 1$
    - $\rho_{ij} = 0$ (then also $C_{ij} = 0$) ⇨ X_i, X_j uncorrelated (but not necessarily independent)
    - $\rho_{ij} = \pm 1$ ⇨ X_i, X_j linearly dependent (i.e., $X_j = a X_i + b$ with $a > 0$ or $a < 0$, respectively)