CS 6323
Complex Networks and Systems

Continuous-Valued Probability Review

Prof. Gregory Provan
Department of Computer Science
University College Cork

Overview

- Review of discrete distributions
- Continuous distributions

Discrete Distributions: Review

- Key discrete distributions
  - Bernoulli
  - Binomial
  - Geometric
  - Poisson

Bernoulli distribution

- The Bernoulli distribution is the “coin flip” distribution.

- X is Bernoulli if its probability function is:

$$X = \begin{cases} 1 & w.p. \quad p \\ 0 & w.p. \quad 1-p \end{cases}$$

X=1 is usually interpreted as a “success.” E.g.:

X=1 for heads in coin toss

X=1 for male in survey

X=1 for defective in a test of product

X=1 for “made the sale” tracking performance

Binomial distribution

- The binomial distribution is just *n* independent Bernoulli events "added up"

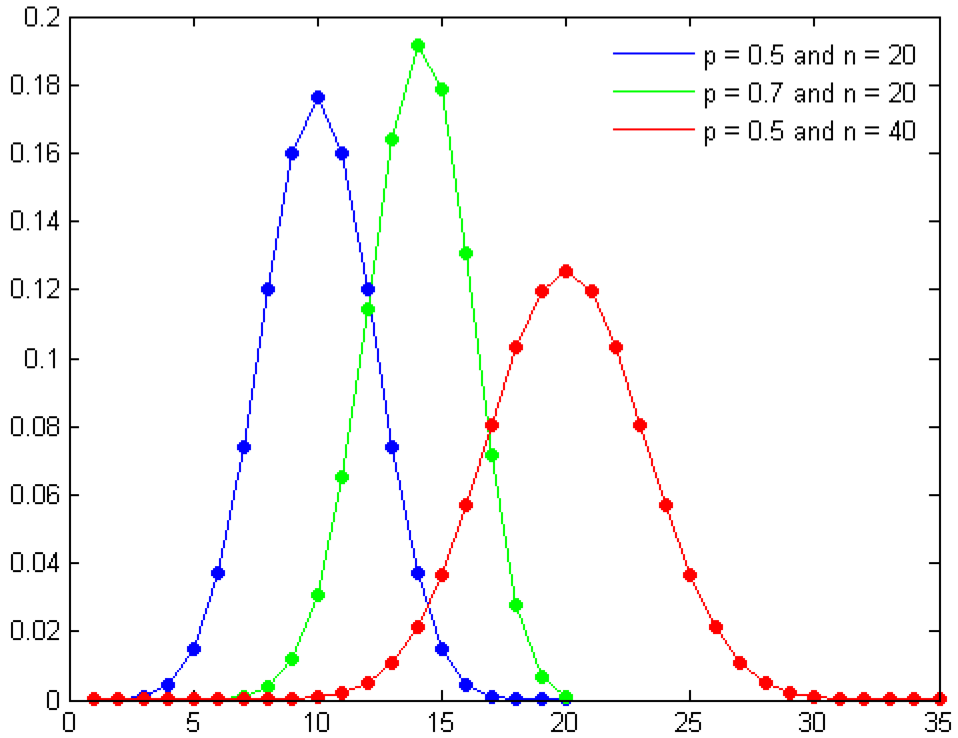
- It is the number of "successes" in n trials.

- If $Z_1, Z_2, \ldots, Z_n$ are Bernoulli, then X is binomial:

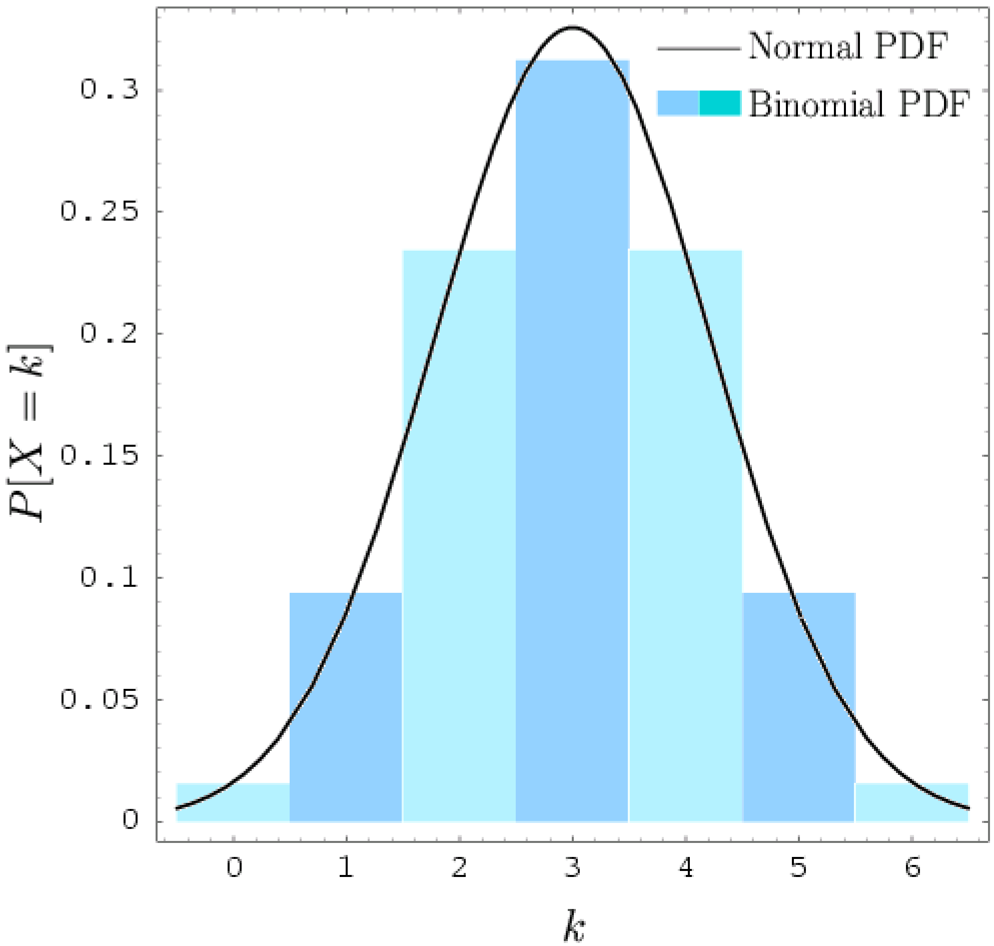
$$X = Z_1 + Z_2 + \ldots + Z_n$$

Binomial distribution

- Typical shape of binomial:
  - Symmetric



Binomial and normal / Gaussian distribution



The normal distribution is a good approximation to the binomial distribution. ("large" n, small skew.)

B(n, p)

$$\mathrm{N}(np, np(1-p)).$$

Prob. density function:

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Geometric Distribution

- A geometric distribution is usually interpreted as number of ***time periods*** *until* a failure occurs.

- Imagine a sequence of coin flips, and the random variable X is the flip number on which the first tails occurs.

- The probability of a head (a success) is p.

Geometric

- The probability function for the geometric distribution:

$$P_X(1) = (1 - p)$$

$$P_X(2) = p(1 - p)$$

$$P_X(3) = p(p)(1 - p) = p^2(1 - p)$$

etc.

So, $P_X(x) = ?$ (x is a positive integer)

$$P_X(x) = p^{x-1}(1 - p)$$

Poisson distribution

- The Poisson distribution is typical of random variables which represent counts.

    - Number of requests to a server in 1 hour.
    - Number of sick days in a year for an employee.

- The Poisson distribution is derived from the following underlying arrival time model:
  - The probability of an unit arriving is uniform through time.
  - Two items never arrive at exactly the same time.
  - Arrivals are independent --- the arrival of one unit does not make the next unit more or less likely to arrive quickly.

Poisson distribution

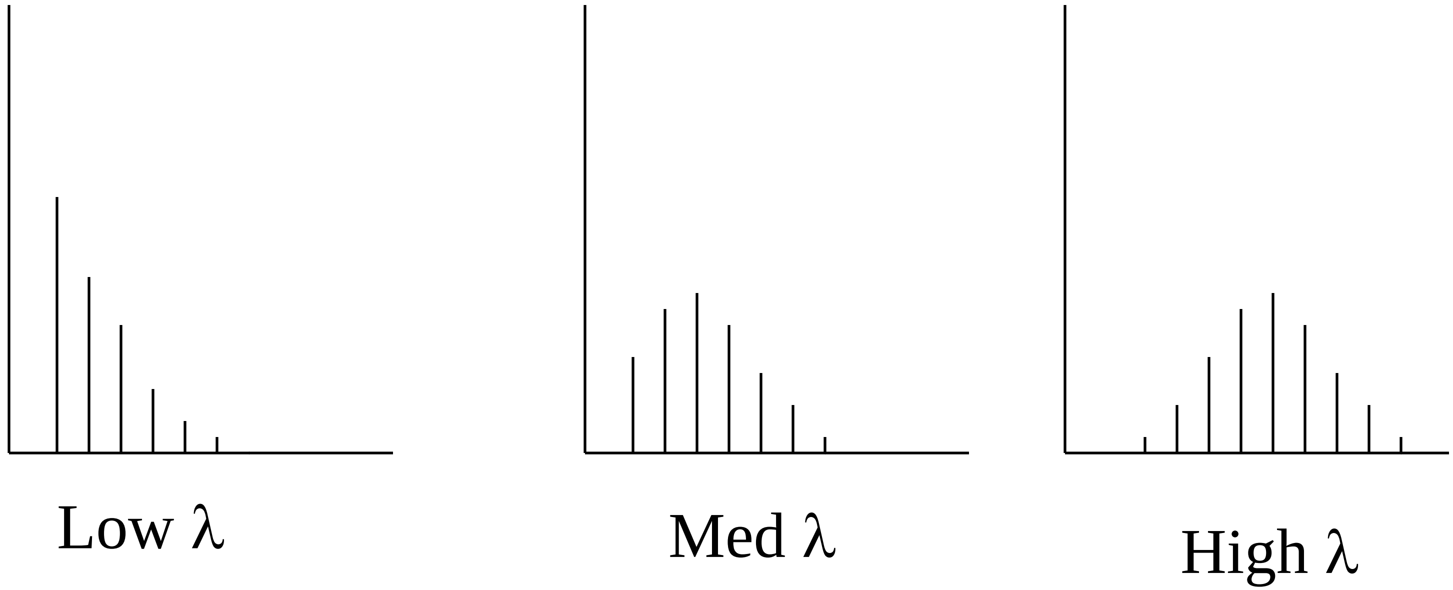
- The probability function for the Poisson distribution with parameter λ is:
- λ is the arrival rate --- higher means more/faster arrivals

$$P_x(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for x} = 0,1,2,3,\ldots$$

$$E(X) = V(X) = \lambda$$

Poisson distribution

- Shape



Exponential and Geometric Distributions

- The exponential distribution is the continuous-valued analogue of the geometric
- Exponential
  - constrained only to be non-negative real numbers instead of integers (geometric)

Continuous Random Variables

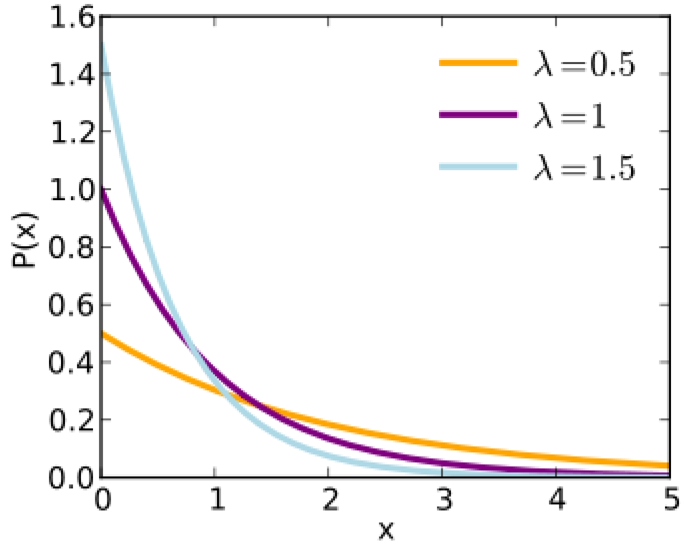
- Continuous RV
  - The values that the random variable can take are continuous
- Examples:
  - The failure time of a system
  - The value of a circuit resistance
- CDF F(X): cumulative distribution function
- The density function *f(X)* is given by the derivative of the cumulative distribution function
  - $f(X) = F'(X)$
- Example:

"The failure time of a system is exponentially distributed"

Exponential Distribution

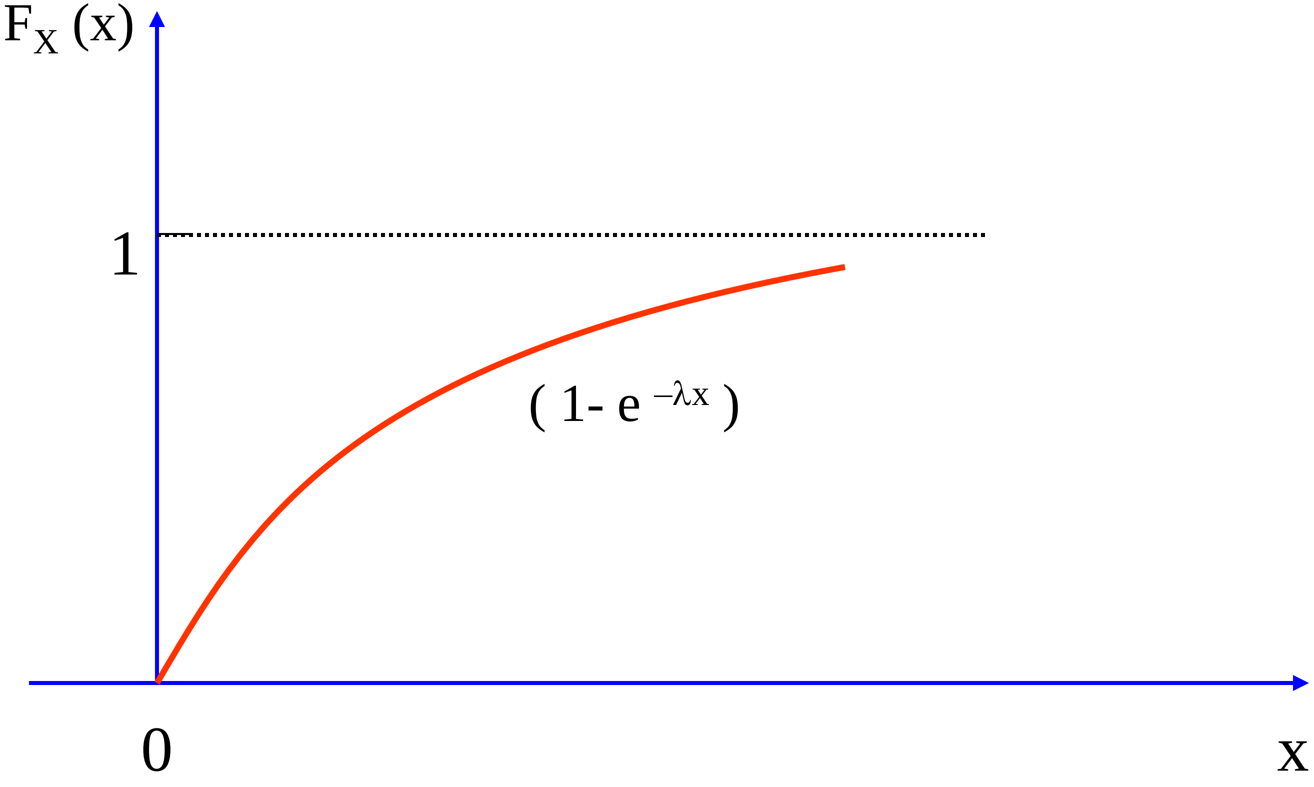
- The cumulative distribution function: $F(X \leq t) = 1 - e^{-\lambda t}$
- $F(X>t) = 1-F(X\leq t) = e^{-\lambda t}$
- The exponential density function is $f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$
  - The parameter λ is constant

$$f(x) = \lambda e^{-\lambda x}$$



Exponential CDF

- The CDF is shown below:



Interpretation-Exponential Distribution

- Exponential distribution occurs in reliability work over and over again, in the way used as the distribution of the time to failure for a great number of electronic-system parts
- The parameter λ is constant and is usually called the **failure rate** (with the units fraction failures per hour)
- The cumulative distribution function: $F(t) = 1 - e^{-\lambda t}$
- The success probability (probability of no failure): $1 - F(t) = e^{-\lambda t}$
- expected value (Mean Time Between Failures): $1/\lambda$ (MTBF)
- The most commonly used distribution in reliability and performance modeling

Exponential Distribution CDF

- ***Problem 1:*** *The transmission time X of messages in a communication system obeys the exponential law with parameter λ ,that is $P[X>x] = e^{-\lambda x}$ $x > o$*

  Find the cdf of X. Find $P[T<X \leq 2T]$, where $T=1/\lambda$.

Solution: *The cdf of X is $F_X(x) = P[X \leq x] = 1 - P[X > x]$:*

$$f_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Continued...

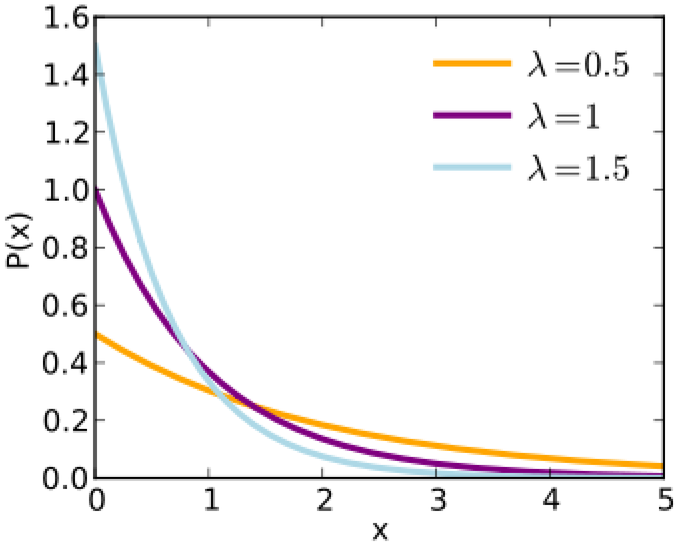
Exponential Example

- *Compute P[T < X ≤ 2T]*

$$P[T < X \le 2T] = (1-e^{-2}) - (1 - e^{-1}) = e^{-1} - e^{-2} = .233$$

$F_X(x)$ is continuous for all x. Note also that its derivative exits everywhere except at x=0:

$$F'_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases}$$



Continued…

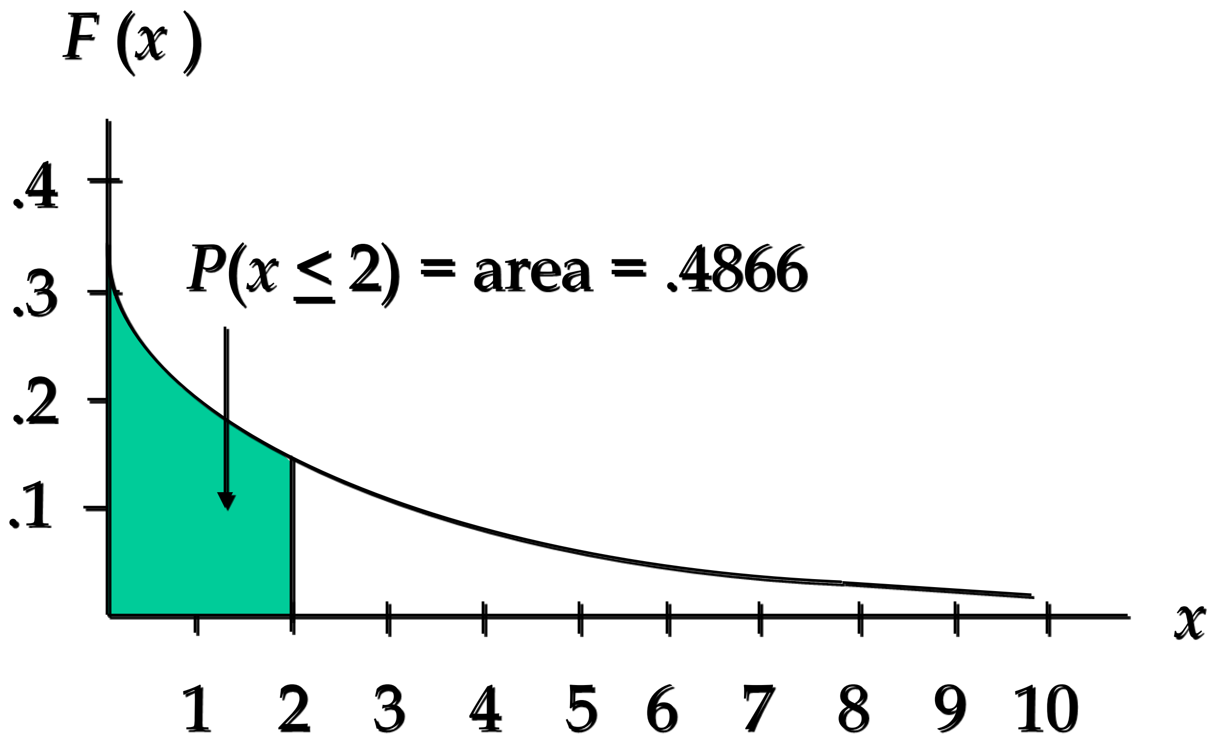
Example: Al's Website

- The time between arrivals of orders at Al's Website follows an exponential probability distribution with a mean time between arrivals of 3 minutes.
- Al would like to know the probability that the time between two successive arrivals will be 2 minutes or less.

$$P(x \leq 2) = 1 - 2.71828^{-2/3} = 1 - .5134$$
$$= .4866$$

Example: Al's Website

- Graph of the Probability Density Function



Memory-less property

A random variable X is said to be without memory, or *memory-less*, if

$P\{X>s+t|X>t\} = P\{X>s\}$ for all $s, t \geq 0$

Interpretation

If an item is alive at time t, then the distribution of the remaining amount of time that it survives is the same as the original lifetime distribution.

The item does not remember that it has already been in use for a time t

More on Memoryless Property

P{X>s+t|X>t} = P{X>s} for all s, t≥0

$$\frac{P\{X > s+t, X > t\}}{P\{X > t\}} = P\{X > s\}$$

$$P\{X > s+t\} = P\{X > s\}P\{X > t\}$$

When X is exponentially distributed, it follows that $e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t}$.

Hence, exponentially distributed random variable are memoryless.

Example

- Suppose the amount of time one spends in a bank is exponentially distributed with mean 10 minutes, that is, $\lambda=1/10$.
  - What is the probability that a customer will spend more than fifteen minutes in the bank?
  - What is the probability that a customer will spend more than fifteen minutes in the bank given that s/he is still in the bank after ten minutes?

Solution

$$P\{X > 15\} = e^{-15\lambda}$$

$$P\{X > 10\} = e^{-10\lambda}$$

$$P\{X > 5\} = e^{-5\lambda}$$

$$P\{X > 15 | X > 10\} =$$

It turns out that the exponentially distribution is the unique distribution possessing the memory-less property (proof omitted here)