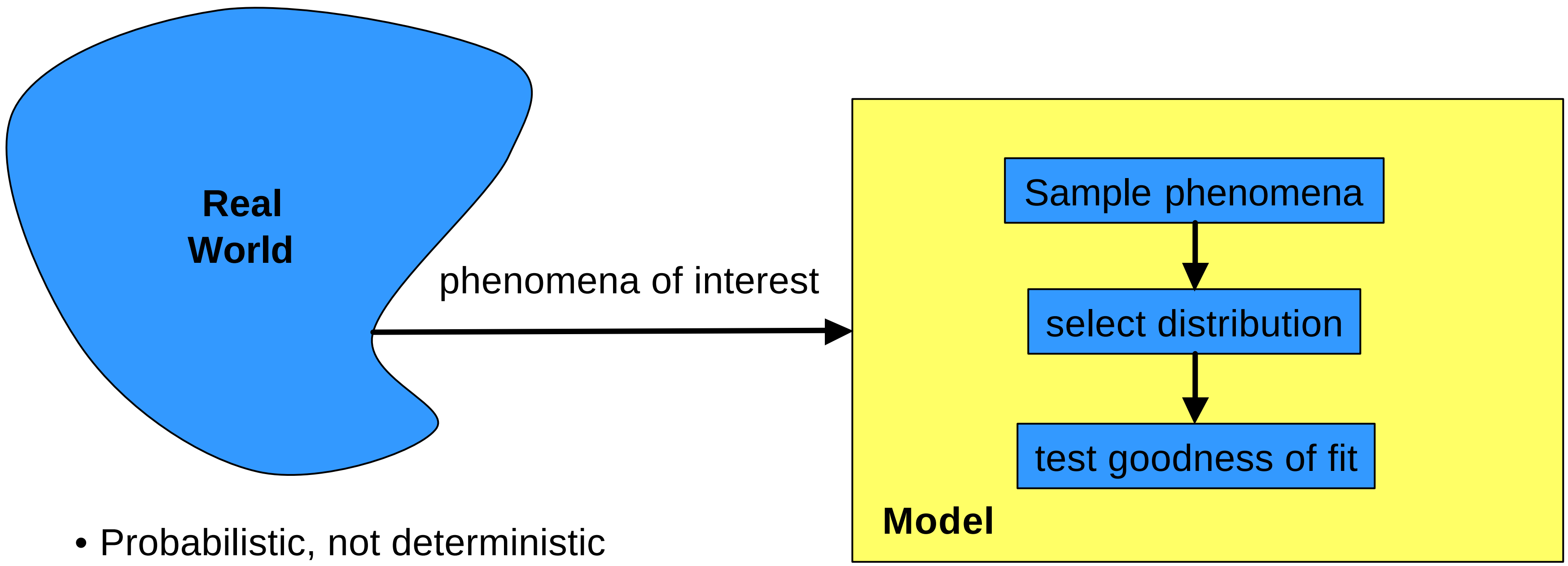
EE/CpE 345

Modeling and Simulation

Spring 2003

Class 4

Statistical Models in Simulation



• Probabilistic, not deterministic

Terminology/Concepts in Probability and Statistics

- Discrete Random Variables (X)
  - The number of possible values of X is finite or countably infinite
  - Example 5.1: The number of jobs arriving at a shop each week
    - Number of jobs in a given week: *X*
    - Possible values of *X*: The range space of *X*,

      $R_X = \{1,2,3,\dots\}$
    - The probability that the R.V. *X* takes on the value x_i

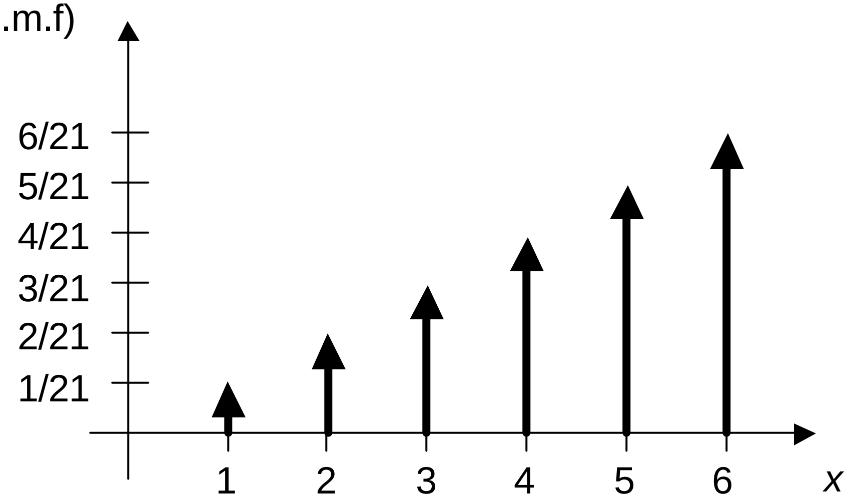
  - *$p(x_i)$ is the probability that the R.V. X equals x_i* $\quad p(x_i) = P(X = x_i)$
  - *all event probabilities are non-negative* $\quad p(x_i) \geq 0 \quad \forall i$
  - *probabilities measure proportion of event occurrences* $\quad \sum_{i=1}^{\infty} p(x_i) = 1$

Discrete Random Variables

- Example 5.2: Testing a single die
  - R_X={1,2,3,4,5,6}
  - Assume the die is loaded, with the probability of a given face showing proportional to the number of spots

x_i	1	2	3	4	5	6
$p(x_i)$	1/21	2/21	3/21	4/21	5/21	6/21

probability mass function (p.m.f)
p(x)



Continuous Random Variables

- The Range Space R_X of the random variable X is an interval or collection of intervals

$$P(a \le X \le b) = \int_a^b f(x)dx$$

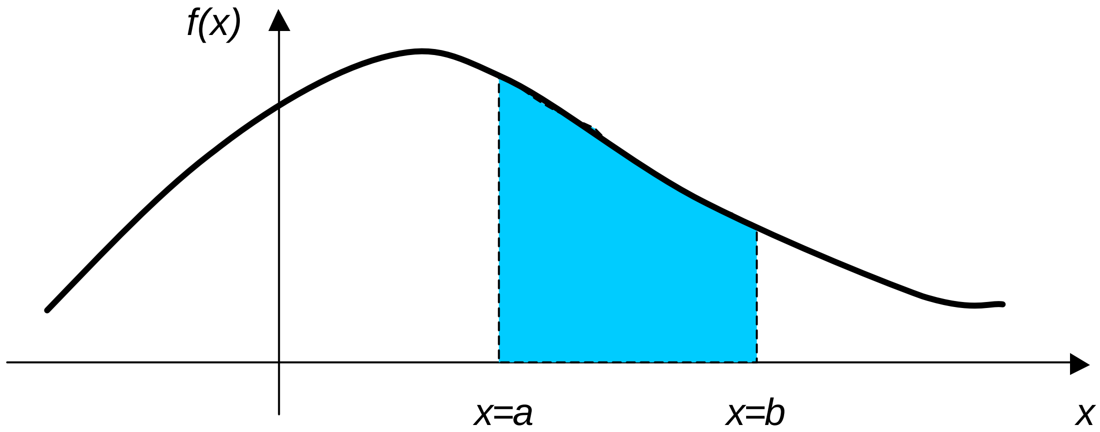
- *f(x)* is the probability density function (p.d.f.)

  – The p.d.f. is nonzero within the Range Space $\quad f(x) \ge 0 \quad \forall x \in R_X$

  – The total area under the p.d.f. represents all possible events $\quad \int_{R_X} f(x)dx = 1$

  – The p.d.f. is zero outside the Range Space $\quad f(x) = 0 \quad \forall x \notin R_X$

Probability of events in an interval



$$P(a < X < b) = \int_{a}^{b} f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$$

Continuous Random Variables

- Exponential distribution with mean m

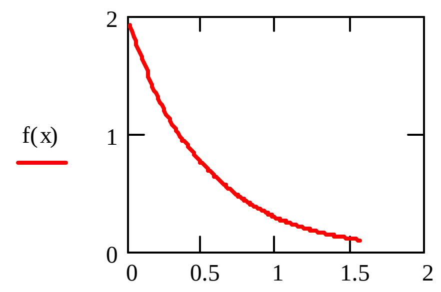
$$f(x) = \begin{cases} \frac{1}{m} e^{-x/m} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Exponential distribution

$m := .52$

$f(x) := \left(\frac{1}{m}\right) \cdot e^{\frac{-x}{m}}$

$x := 0, .01 .. 3 \cdot m$



MathCad

Cumulative Distribution Function

- Measures probability that the R.V. *X* assumes a value less than or equal to *x*
  - for discrete R.V.

$$F(x) = \sum_{\forall x_i \le x} p(x_i)$$

  - for continuous R.V.

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

  - Properties of F( ):

$$a < b \quad \longrightarrow \quad F(a) \le F(b)$$

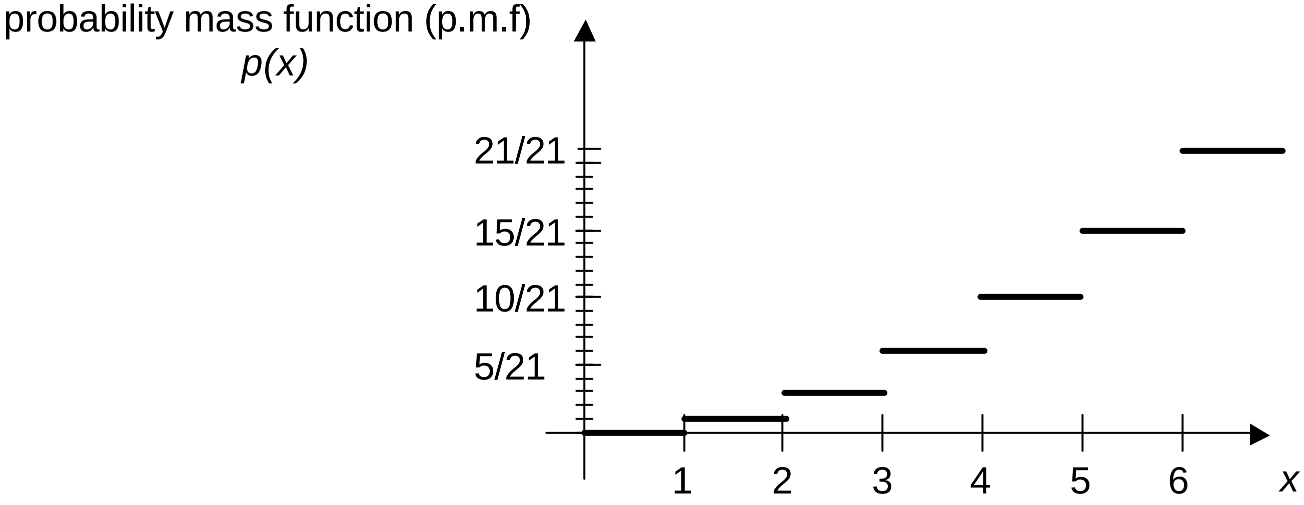
$$\lim_{x \to \infty} F(x) = 1$$

$$\lim_{x \to -\infty} F(x) = 0$$

C.D.F. Examples

- Loaded die

x_i	(-8,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6, 8)
$p(x_i)$	0	1/21	3/21	6/21	10/21	15/21	21/21

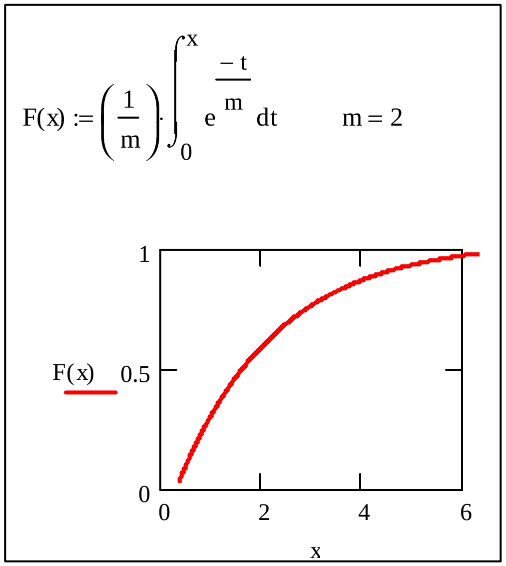


C.D.F. Examples

- Exponential distribution

$$F(x) = \frac{1}{m}\int_0^x e^{-t/m}dt = 1 - e^{-x/m}$$

  - exponential distribution is easily evaluated, making it very popular for closed form solutions



Useful Parameters of R.V.s

- Expected Value
  - if X is a R.V., the Expected Value (Expectation) is:
    - Discrete case:

$$E(X) = \sum_{\forall i} x_i \, p(x_i)$$

    - Continuous case:

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

    - *E(X)* is also called the mean, μ, or the 1st moment of *X*
    - $E(X^n)$ is the nth moment of *X*

$$E(X^n) = \sum_{\forall i} x_i^{\,n} \, p(x_i)$$

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

More Useful Parameters of R.V.s

- Variance

$$V(X) = E\left[\left(X - E[X]\right)^2\right] = \mathrm{var}(X) = s^2$$

$$V(X) = E(X^2) - \left[E(X)\right]^2$$

- Standard deviation

$$s = \sqrt{V(X)}$$

- Mode - the value that occurs most often (discrete) or peak of p.d.f. (continuous).

- Bimodality - two peaks in the p.d.f.

Statistical Models

To address the statistical model to consider:

- What is the problem being addressed?
  - Queuing
  - Reliability/failure
  - Inventory
  - Communications systems behavior

- What is known about the process?
  - Completely random
  - Essentially constant with a random component
  - Constrained
    - non-negative
    - limited tails of distribution

- Tractable mathematical analysis

Statistical Models

To address the statistical model to consider:

- What is the problem being addressed?
  - Queuing
  - Reliability/failure
  - Inventory
  - Communications systems behavoir
- What is known about the process?
  - Completely random
  - Essentially constant with a random component
  - Constrained
    - non-negative
    - limited tails of distribution
- Tractable mathematical analysis

Exponential distribution

unlimited tails

Statistical Models

To address the statistical model to consider:

- What is the problem being addressed?
  - Queuing
  - Reliability/failure
  - Inventory
  - Communications systems behavoir
- What is known about the process?
  - Completely random
  - Essentially constant with a random component
  - Constrained
    - non-negative
    - limited tails of distribution
- Tractable mathematical analysis

Gaussian/Normal distribution

unlimited precursors

unlimited tails

Statistical Models

To address the statistical model to consider:

- What is the problem being addressed?
  - Queuing
  - Reliability/failure
  - Inventory
  - Communications systems behavoir

- What is known about the process?
  - Completely random
  - Essentially constant with a random component ← Truncated Gaussian/Normal distribution or small variance
  - Constrained
    - non-negative
    - limited tails of distribution

- Tractable mathematical analysis

Statistical Models

To address the statistical model to consider:

- What is the problem being addressed?
  - Queuing
  - Reliability/failure
  - Inventory
  - Communications systems behavoir

- What is known about the process?
  - Completely random
  - Essentially constant with a random component
  - Constrained
    - non-negative
    - limited tails of distribution

- Tractable mathematical analysis

Normal
Exponential
Poisson

Discrete Distributions

- Bernoulli trials and binomial distributions
  - Consider an experiment (e.g., flipping a coin, receiving a bit) consisting of *n* trials which can be a success (1) or failure (0)
  - The *n* Bernoulli trials are called a Bernoulli process if the trials are independent
  - For one trial, the Bernoulli distribution is:

$$p(x) = \begin{cases} p & \text{x=1} \\ 1-p=q & \text{x=0} \\ 0 & \text{otherwise} \end{cases}$$

- Binomial distribution:
  - The number of successes in a Bernoulli process has a binomial distribution

$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{x=0,1,2,...n} \\ 0 & \text{otherwise} \end{cases}$$

Discrete Distributions

- Geometric distribution
  - The number of Bernoulli trials before the first success

$$p(x) = \begin{cases} q^{x-1} p & \text{x=1,2,...} \\ 0 & \text{otherwise} \end{cases}$$

- Poisson distribution
  - Models arrival processes in queuing systems quite well
  - One of the few distributions that makes the mathematics tractable
  - mean and variance = α

$$p(x) = \begin{cases} \dfrac{e^{-a} a^{\text{x}}}{\text{x}!} & \text{x=0,1,2,...} \\ 0 & \text{otherwise} \end{cases}$$

Continuous Distributions

- Exponential distribution
  - Model completely random interarrival times
  - Model highly variable service times
  - λ is a rate: service rate, arrival rate
  - Distribution has long tail, useful for modeling component lifetime. λ is failure rate

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{x} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Continuous Distributions

- Uniform distribution
  - Easiest distribution to generate, serves as a basis for other R.V.s in simulation
  - Uniform p.d.f.:

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

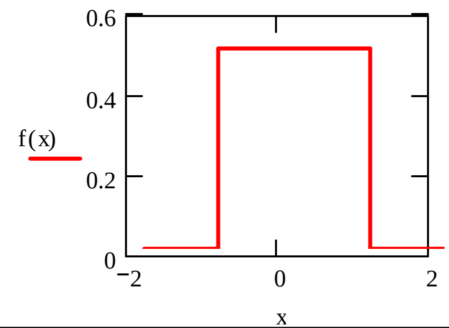
Uniform distribution

$a := -1$

$b := 1$

$f(x) := \text{if}\left(a \le x \le b, \frac{1}{b-a}, 0\right)$

$x := a - 1, a - 1 + .01 .. b + 1$



Continuous Distributions

- Gamma distribution

  Gamma function: $\Gamma(b) = \int_0^\infty x^{b-1} e^{-x} dx$

  $\Gamma(b) = (b-1)\Gamma(b-1)$

  If b is an integer, $\Gamma(b)=(b-1)!$

  – Gamma pdf:

$$f(x) = \begin{cases} \dfrac{bq}{\Gamma(b)}(bq\,x)^{b-1} e^{-bq\,x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

  – β is shape parameter, θ is scale parameter

  – when β is integer, Gamme distribution is related to exponential distribution: If the R.V. *X* is the sum of β independent, exponentially distributed R.V.s, each with parameter βθ, then *X* has a Gamma distribution with parameters β and θ

Continuous Distributions

- Erlang distibution
  - The Gamma p.d.f. is also known as the Erlang distribution of order k when β=k, an integer
  - With the "block calls dropped, exponential arrival rates and exponential holding times" assumptions, the Erlang distribution is used to predict the number of busy trunks in a telephone system.

Continuous Distributions

- Normal distribution (Gaussian distribution)
  - mean μ, variance σ^2

$$f(x) = \frac{1}{s\sqrt{2p}} e^{\left(-\frac{1}{2}\left[\frac{x-m}{s}\right]^2\right)}$$

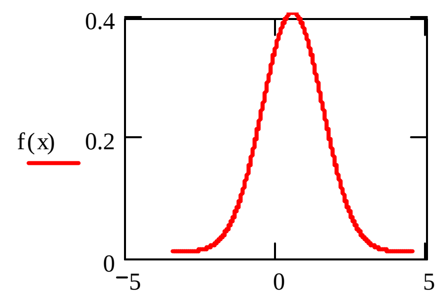
Normal distribution

$\sigma := 1$

$\mu := 0$

$$f(x) := \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{\left[\frac{-1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2\right]}$$

$x := -\mu - 4 \cdot \sigma, -\mu - 4 \cdot \sigma + .01 .. \mu + 4 \cdot \sigma$



Continuous Distributions

- Weibull distribution

$$f(x) = \begin{cases} \frac{b}{a}\left(\frac{x-u}{a}\right)^{b-1} e^{-\left(\frac{x-u}{a}\right)^{b}} & x \geq u \\ 0 & \text{otherwise} \end{cases}$$

  - location parameter ν, scale parameter α, shape parameter β
  - This distribution is a general case of many other distributions, e.g.,
    - with $\beta=1$ this is the exponential distribution with $\lambda=1/\alpha$

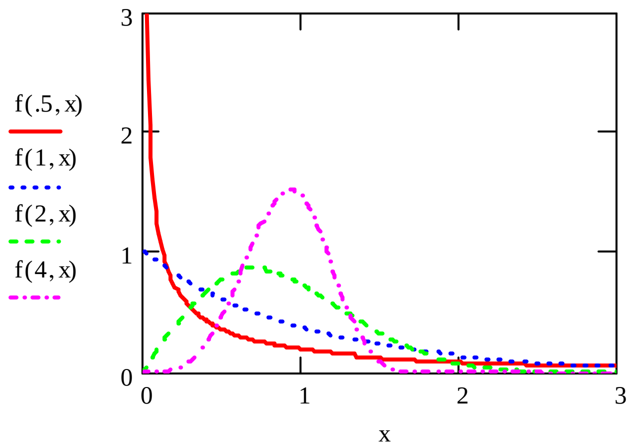
Continuous Distributions

Weibull distribution with ν=0

$\alpha := 1$

$$f(\beta, x) := \frac{\beta}{\alpha} \cdot \left(\frac{x}{\alpha}\right)^{\beta-1} \cdot e^{\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]}$$

$x := 0, .01 .. 3$



Error in text book, figure 5.20: α=1, not .5

Continuous Distributions

- Rayleigh Distribution
  - Special case of Weibull distribution with α=2

$$f(x) = \begin{cases} \frac{b}{2}\left(\frac{x-n}{2}\right)^{b-1} \exp\left[-\left(\frac{x-n}{2}\right)^{b}\right] & x \geq n \\ 0 & \text{otherwise} \end{cases}$$

  - used to model multipath fading, radiation, wind speeds

Continuous Distributions

- Triangular distribution

$$f(x)=\begin{cases}\dfrac{2(x-a)}{(b-a)(c-a)} & a\leq x\leq b\\[2ex] \dfrac{2(c-x)}{(c-b)(c-a)} & b<x\leq c\\[2ex] 0 & \text{elsewhere}\end{cases}$$

Triangular distribution

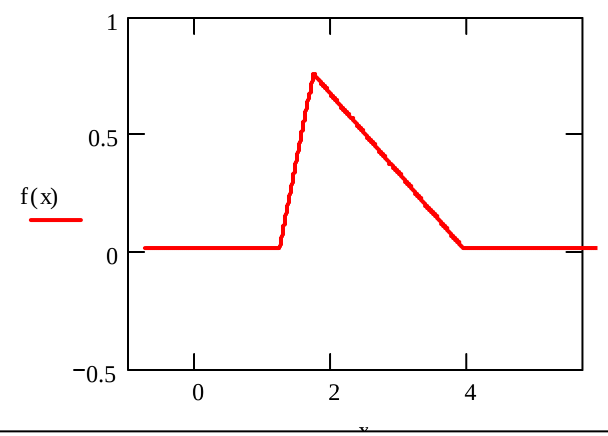
a := 1

b := 1.5

c := 3.7

$$f(x) := \text{if}\left[a\leq x\leq b,\ 2\cdot\frac{(x-a)}{(b-a)(c-a)},\ \text{if}\left[b<x\leq c,\ 2\cdot\frac{(c-x)}{(c-b)(c-a)},\ 0\right]\right]$$

x := a − 2, a − 2 + .01 .. c + 2



Continuous Distributions

- Lognormal distribution

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2p}\, s\, x} e^{\left[-\frac{(\ln x - m)^2}{2s^2}\right]} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

  – Shadow fading is often modeled as a lognormal process

Poisson Process

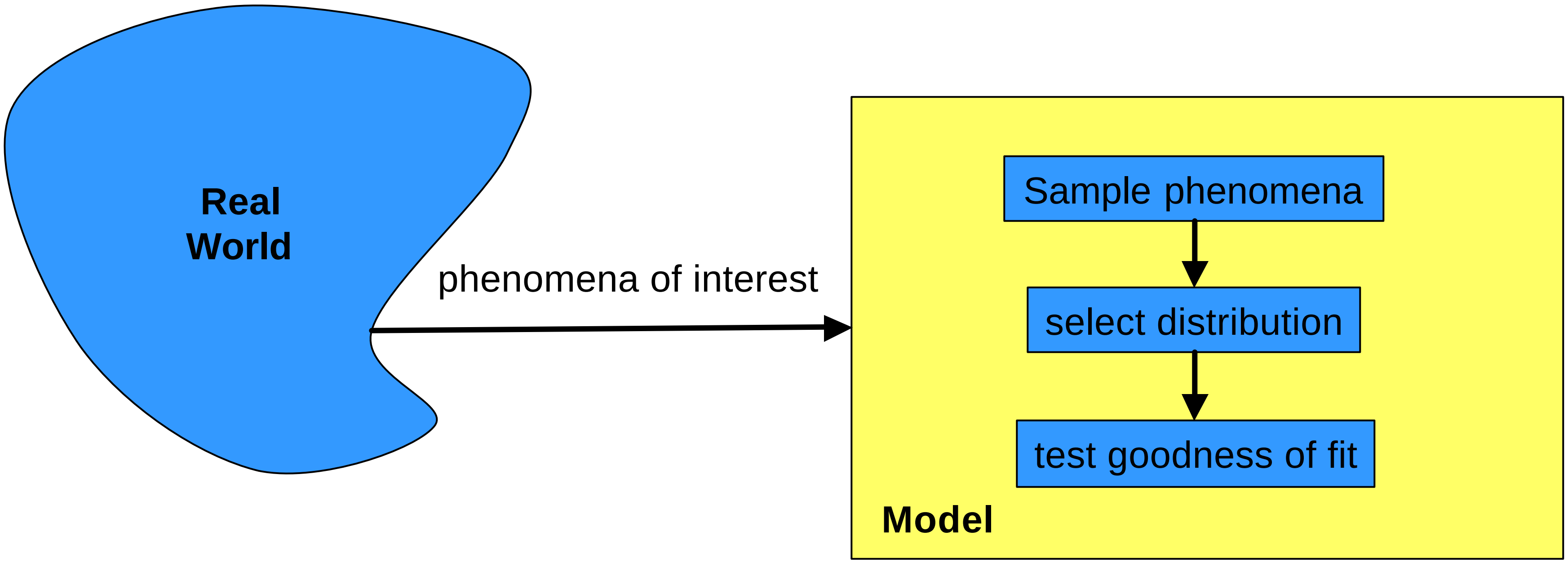
- A counting process {N(t), t=0} is a Poisson process with mean rate λ if
  - arrivals occur singly
  - {N(t), t=0} has stationary increments: The distribution of the number of arrivals between t and t+s depends only on s, the length of the interval, and not on t, the starting point
  - {N(t), t=0} has independent increments: The number of arrivals during non-overlapping time intervals are independent R.V.s

$$P\left[N(t)=n\right]=\frac{e^{-lt}(lt)^n}{n!} \quad \text{for } t \geq 0 \quad \text{and } n=0,1,2,\ldots$$

Poisson Process

- Properties of a Poisson Process
  - Random splitting
    - A Poisson process {N(t), t=0} having rate λ
    - Each time an event occurs, it is arbitrarily be classified as either a type I (with probability p, $N_1(t)$) or type II (with probability 1-p, $N_2(t)$) event
    - $N_1(t)$ and $N_2(t)$ are retain the property of being Poisson processes having rates λp and $\lambda(1-p)$
  - Pooling of two arrival streams
    - Given two independent Poisson processes
    - $N(t)=N_1(t)+N_2(t)$ is a Poisson process with rate $\lambda_1+\lambda_2$

Known vs. Empirical Distributions



- Situation A: You understand the processes that create random variability in observed phenomena. Pick the proper distribution, adjust parameters, and verify fit to data

- Situation B: You do not fully understand (or don't really care, or don't have the time to analyze) the processes that create random variability. Either
  - use sampled data to form an empirical distribution
  - pick a known distribution that is the best approximation

Homework 4

1. Chapter 5, p.198 problems 15 and 25