Stochastic Abstraction

Represent the execution behavior of events and actions by properties chosen from a distribution.

Examples - execution times
- branching probabilities

Distributions are usually derived from statistical analysis of data given the values of properties measured on realized systems.

For example, the execution time of each action is represented as the average of the execution times of all events.

The execution time of each action is chosen from a distribution derived from the execution times of the individual actions.

If the behavior of the actual systems is structured or patterned then the distributions should be chosen to capture the metric determining characteristics of the property.

Example - File Access on a web server.

- measurement shows that there is a frequent access to a small set of files.

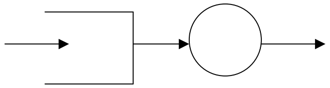
- measurement shows there is a wide span of sizes in the files which are accessed in a given period.

- measurement shows that a small number (2%) of large files require 50% of the total transfer time.

- measurement shows that the characteristics of the set of files which is accessed differs at different times of the day.

What kinds of distributions (stochastic abstractions) would be used to capture and return the "important properties" of these file accesses?

Example - Arrival of Requests at a web server where the web server is represented as a queue/server pair.



- The log file for the web server will give a trace of the requests. This is the most accurate representation for the current workload and system.

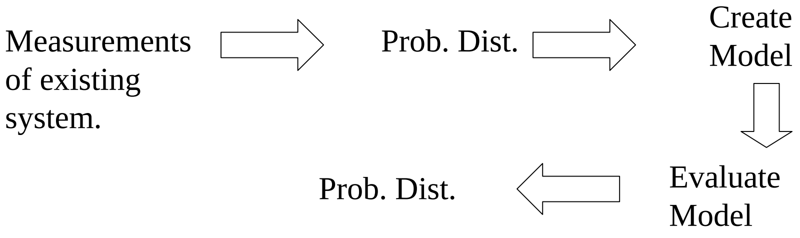
- But log files are bulky, varying the workload by modifying the log files is awkward, processing of them serially is effort intensive.

- So we take the log file and generate distributions for interarrival times for different types of jobs.

- Then we run the simulation by sampling from the distribution function.

The Use of Probability/Statistics

- *Probability distributions* are the basic building blocks that allow us to generate complicated sequences of behavior from simple descriptions
- *Statistics* take us from observed behavior back to probability distributions (or measures of distributions)



Measurements of existing system.

Prob. Dist.

Create Model

Evaluate Model

Prob. Dist.

Probability/Statistics

- Suppose we want to represent a system as a single service center. How should we describe customer arrivals (say, in a simulation of the system)?
- Most faithful representation: a *trace* of actual arrival times

  $(e,t_1), (e,t_2), \ldots\ldots (e,t_k),$

- Problems
  - ∗ This can be a large amount of data
  - ∗ Was the interval traced representative of “typical behavior”?
  - ∗ How do you modify the trace to represent the projected system?

What to Do?

- If you’re not going to replay a trace, then you must generate the input data somehow. How?
- One simple possibility :
  - ∗ Define x = t_i-t_{i+1} (the i th inter-arrival time)
  - ∗ Compute the average inter-arrival time
  - ∗ Create the i th arrival at time as i*$x_{ave.}$
- Problem : Arrivals are as evenly spread out as possible (so contention for the server is most likely underestimated)
- Another possibility
  - * Find a distribution function by a fitting procedure.
- Problem
  - * What distribution?

(Discrete) Probability

- An *experiment* is the process of making an observation.
  - ∗ Running Example: Examining the number of processes in the CPU run queue
- A *sample spac*e, S, is the set of all possible outcomes of an experiment.
  - ∗ S = {0, 1, 2, …}
- A sample space is *discrete* if it is countable (which for our purposes means contains no real-valued intervals)

(Discrete) Probability

- A *random variable* is a variable whose value is the outcome of (some not yet performed) experiment.
  - ∗ X = 4 is the outcome that 4 runable processes are found in the run queue
- A *probability distributio*n, f(x), for a random variable x is a mapping from the possible values of x to probabilities. (We require that f(x) =>0 and sum(f(x) = 1)
  - ∗ f(0) = .2, f(1) = .2, f(2) = .1, f(3) = .3, f(4) = .2
- Terminology:
- {probability} density function (pdf)
- “x is distributed according to f(x)”
- “x has distribution f(x)”

Expected Value of a Distribution

- "Mean of a distribution" or "the expected value, E(x), of a random variable x"
  - sum(xf(x))
  - ∗ "Average queue length" = 0*.2 + 1*.2 + 2*.1 + 3*.3 + 4*.2 = 2.1
- The mean is often denoted as μ
- You can also take expected values of functions of random variables

∗ $E(x-2) = -2*.2 + -1*.2 + 0*.1 + 1*.3 + 2*.2 = 0.1$

Properties of Expected Value

- E(c) = c for any constant c
- Let g(X) be a function of the random variable X, and let c be a constant. Then E(*c*g(X)) = cE(g(X))
  - ∗ $E(4(X+2)^2) = 4E((X+2)^2)$
- Let *g* 1 (X), *g* 2 (X), ¼, *g* K (X) be K functions of random variable X. Then

E(*g* 1 (X)+¼+ *g* K (X)) = E(*g* 1 (X))+¼+ E(*g* K (X))

  - ∗ $E((X+2)^2) = E(X^2 +2X+4)$ = E(X 2)+2E(X)+4

Moments of Distributions

- *n th moment of X* is $E(X^n)$
  - ∗ The mean is also called "the first moment"
- The 2nd central moment is called *varianc*e, and is often denoted as σ^2
  - ∗ $\sigma^2 = E((X- \mu)^2) = \text{sum } (((X- \mu)^2) *f(x))$
- The square root of variance, s, is called the *standard deviation*

What Does Variance Tell You?

- Variance is the "mean squared error" of the mean as an estimate of the outcome of a single observation
- It is a (not very intuitive) measure of how much the outcomes are spread from the mean
- Is a variance of 10 big? It depends on the mean.
- It's often useful to use a normalized form of variance:
  - ∗ *Coefficient of variatio*n: $CV = \sigma / \mu$
  - ∗ *Squared coefficient of variatio*n: $CV^2 = \sigma^2 / \mu^2$

An Example

f(0) = .2 , f(1) = .2, f(2) = .1, f(.3) = .3 , f(4) = .2

$\mu = 2.1$

σ^2 = .2*(0-2.1)2 +.2*(1-2.1)2 +.1*(2-2.1)2 +.3*(3-2.1)2 + .2*(4-2.1)2

= 2.09

$CV^2 = 2.09/4.41 = 0.473923$

CV = 0.688421

If f(x) = 1 for all x then, $\sigma^2 = 0$, deterministic behavior

Let there be two outcomes, $A = \mu - \varepsilon$ (with probability p and $B = \mu + \varepsilon/(1-p)$ with probability (1-p).

Then $\sigma^2 = \varepsilon^2 (2-p)/(1-p)$

Moral

- Means are in the range (-infinity,infinity)

.

- Variances are in the range [0,infinity).
- "Deterministic" behavior leads to low variance.
- Many frequently occurring values that are not close to each other lead to high variance.
- However, high variance can also result from rare, very abnormal behavior.
- We'll often use CV instead of variance.

Conditional Distributions

- It is sometimes useful to consider restricted sets of outcomes for a random variable
  - * What is the mean number of processes in the run queue *given that* the run queue is not empty?
- Imagine running an infinite number of trials, throwing out the results that don't meet the conditional criterion, and computing the distribution of the trials that do
  - * We'll write $f_{X|criterion}(x)$ to denote the conditional distribution

  In general, Prob[X=x|Y] = Prob[X=x & Y]/P[Y]

Independence

- Two random variables are independent iff

Prob[X=x & Y=y] = Prob[X=x]*Prob[Y=y]

- This does not correspond to intuition about "independence"!
- If X is independent of Y, then

Prob[X=x|Y=y] = Prob[X=x & Y=y] / Prob[Y=y]
= Prob[X=x]Prob[Y=y]/Prob[Y=y]
= Prob[X=x]

Independence

x =	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
y = 0	.2	.2	.1	.3	.2	0	0	0	0
1	.2	.2	.1	.3	.2	0	0	0	0
2	.2	.2	.1	.3	.2	0	0	0	0
3	.2	.2	.1	.3	.2	0	0	0	0
4	.2	.2	.1	.3	.2	0	0	0	0
5	.2	.2	.1	.3	.2	0	0	0	0
6	.2	.2	.1	.3	.2	0	0	0	0
7	.2	.2	.1	.3	.2	0	0	0	0
8	.2	.2	.1	.3	.2	0	0	0	0

This is unlikely for our example, but possible.

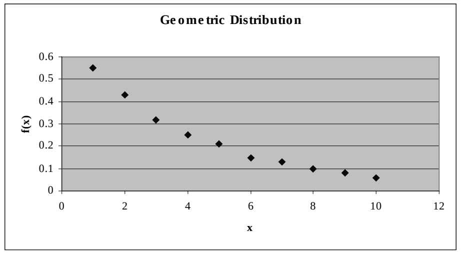
Cumulative Distribution Function

- The *cumulative distribution functio*n, cdf, is defined by
  F(x) = sum(f(i)), i = -infinity, x
- F(x)-F(x-1) = f(x)
- Clearly, 0<=F(x)<=1, and F(infinity)=1

f(0) = 0.2 F(0) = = .2
f(1) = 0.2 F(1) = = .4
f(2) = 0.1 F(2) = = .5
f(3) = 0.3 F(3) = = .8
f(4) = 0.2 F(4) = = 1.0

Geometric

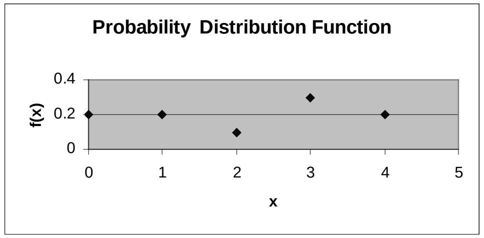
- **Example: Number of coin tosses until a heads is tossed**
- **If p is the probability “of success”, then f(x) = p(1-p)x-1**
- **μ = 1/p**
- **$\sigma^2 = (1-p)/p^2$**
- **The geometric is “memoryless” -- distribution of number of future throws is unaffected by how many have taken place so far**

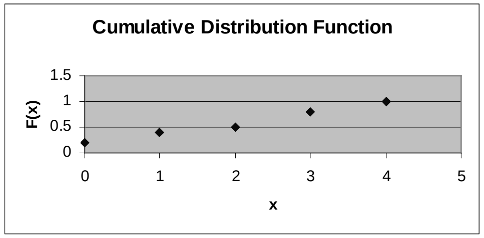


Poisson

- Example: Number of radioactive decays during an interval of length t
- Intuitively, imagine that
  - time is partitioned into small intervals of length ε
  - either 0 or 1 event takes place in each interval (but not more than one)
  - each interval (independently) has an event with probability p
- If λ is the mean rate at which events take place (so $1/\lambda$ is the mean time between events), then

Cumulative Distribution Function





F(x) = Sum(f(i), all i)
F(x) = Prob[X>=x]
0<=F(x) => 1, F(infinity) = 1
F(x) - F(x-1) = f(x)
E[X] = sum(i*f(i), all i)
E[X]= f(1) + f(2) + ..
E[X] = sum(i-F(i), all i)

Uniform Distribution

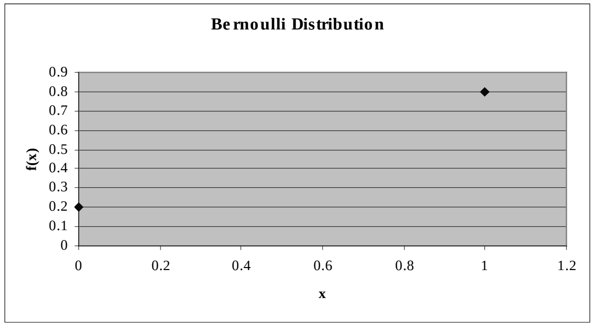
Uniform: U(A,B)
f(x) = 1/(B-A+1)

- **$\mu = (A+B)/2$**
- **$\sigma^2 = (B-A)^2 /12 + (B-A)/6$**

A = 1, B = 5

Bernoulli (Selection)

- **Outcome 1 with probability p, 0 with probability 1-p**
- **$\mu = p$**
- **$\sigma^2 =p(1-p)$**

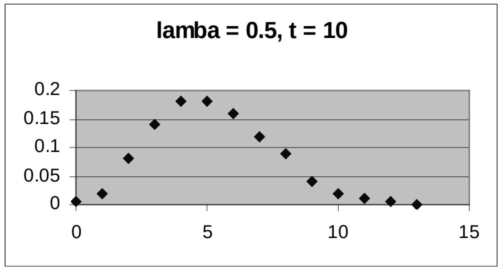


Poisson

- **Example: Number of radioactive decays during an interval of length t**
- **Intuitively, imagine that**
    - ∗ **time is partitioned into small intervals of length ε**
    - ∗ **either 0 or 1 event takes place in each interval (but not more than one)**
    - ∗ **each interval (independently) has an event with probability p**
- **If λ is the mean rate at which events take place (so $1/\lambda$ is the mean time between events), then**

$$f(x,t) = ((\lambda t)/(x!)) * e^{-\lambda x}$$

$$\mu = \lambda\, \mathbf{t}$$
$$\sigma^2 = \lambda\, \mathbf{t}$$



Binomial

- Number of heads in n flips
- In general, given n trials and probability p of outcome 1 (and 1-p of outcome 0), $f_{n,p}(x) = (n \text{ chose } x) p^x (1-p)^{n-x}$
- $\mu = np$
- $\sigma^2 = np(1-p)$

Continuous Random Variables

- Take on a non-countable set of values

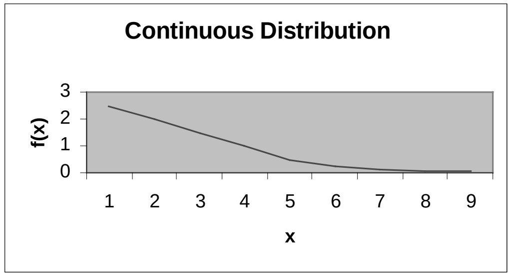
- For us, this will mean the interval [0,infinity)

- Examples:
  - ∗ The time between successive arrivals (inter-arrival time)
  - ∗ The time in service (service time)
  - ∗ The total time in system (response time)

- All the concepts are “more or less the same” as in the discrete case, except...

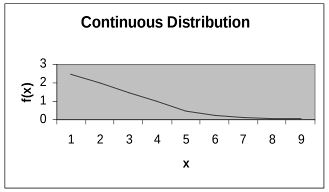
(Continuous) Probability Distributions

- Suppose X has the following distribution



- What does it mean?

What It Doesn’t Mean



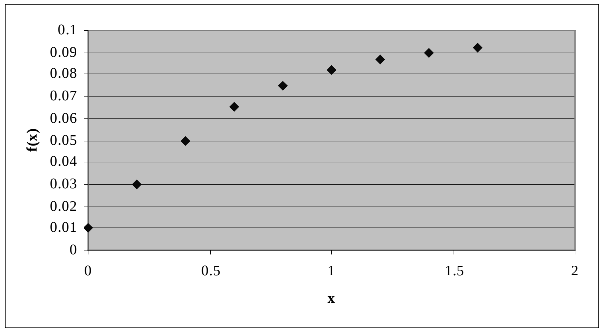
- f(x) is NOT the probability that x occurs…
- because f(x) is not a probability at all
- To talk about probabilities, you must integrate
  Prob[$0 \le X \le 2$] = $\int_0^2$ f(x)dx

Other Properties/Peculiarities

- **We require (only) that $f(x) \ge 0$ for all x**
  $\int$f(x)dx =1 over all x
- **Prob[X=x] = $\int_x^x$ f(y)dy = 0] (at least for the distributions we’ll consider)**
- **Note that Prob[X=x]=0 does NOT mean x can’t occur**
- **Prob[x contained in S]=0 for any countable set S**
- **Prob[x contained in S]=1 means an outcome in S occurs *almost surely***

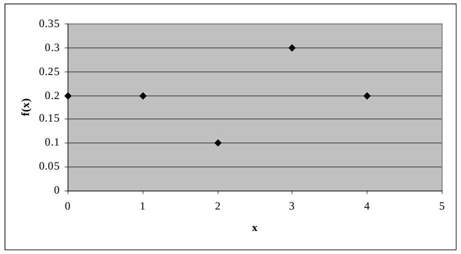
The Cumlative Distribution Function - CDF

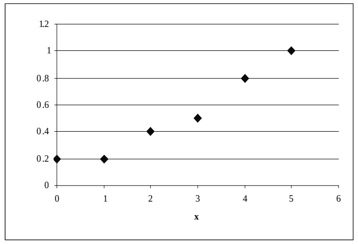
- **F(x) = Prob[X<=x]**
- **f(x) = (dF(y)/dy) at y = x**
  (almost always



Why the CDF?

- Any function that is a CDF • has domain (-infinity, +infinity)
- has range [0,1]
- is monotone non-decreasing





Expectation / Moments For Continuous Distributions

- Mean

$E(X) = X = \mu = \int xf(x)dx = \int(1-F(x))dx$,
$0 <= x => \infty$

- All properties of means we saw for discrete distributions still hold

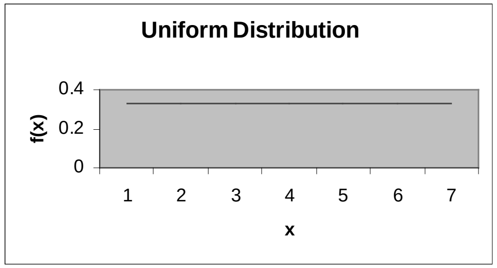
- All definitions we saw for moments still hold, e.g.,

$$\text{Variance} = \sigma^2 = E[(X- \mu)^2] = \int (x- \mu)^2 f(x)dx$$

Some Common Continuous Distributions

Uniform: U(A,B)
- **$f(x) = 1/(B-A)$**
- **$\mu = (A+B)/2$**
- **$\sigma^2 = (B-A)^2 /12$**

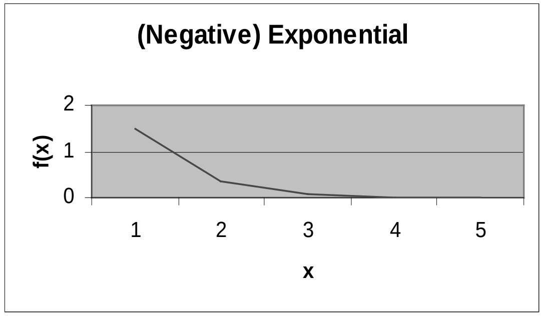
"chosen at random"



Some Common Continuous Distributions

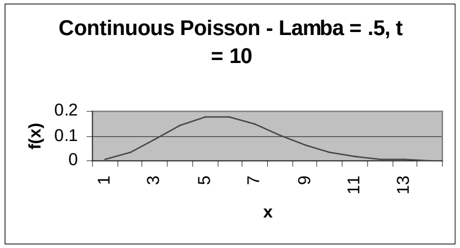
Negative Exponential Distribution

$f(x) = \lambda * \exp(- \lambda x)$
$F(x) = 1- \exp(- \lambda x)$
$\mu = 1/ \lambda$
$\sigma^2 = 1/ \lambda^2$
$CV^2 = 1$



Poisson
- **$\mu = \lambda t$**
- **$\sigma^2 = \lambda t$**
- **Number of events with exponential inter-event times during an interval of length t**
- **If λ is the mean rate at which events take place (so $1/\lambda$ is the mean time between events), then**

$f(x,t) = ((\lambda t)/(x!)) * e^{-\lambda x}$



Discrete/Continuous

	Discrete Time	Continuous Time
Inter-arrival time distribution	Geometric	Exponential
Distribution of Number of Arrivals in (0,t)	Binomial	Poisson