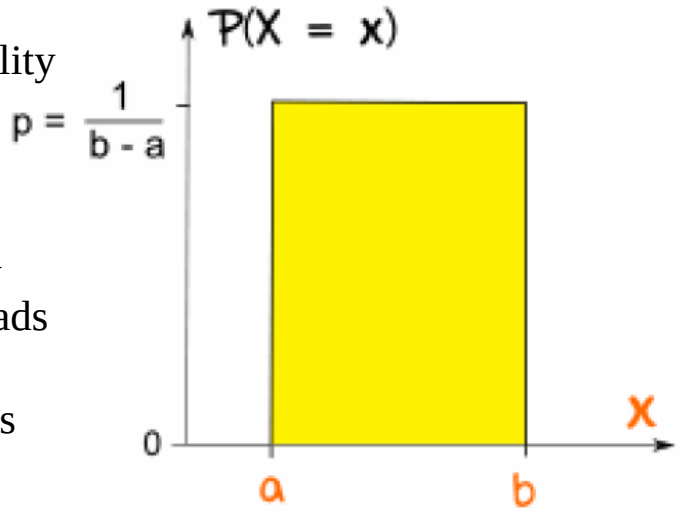
Probability/Statistics Review

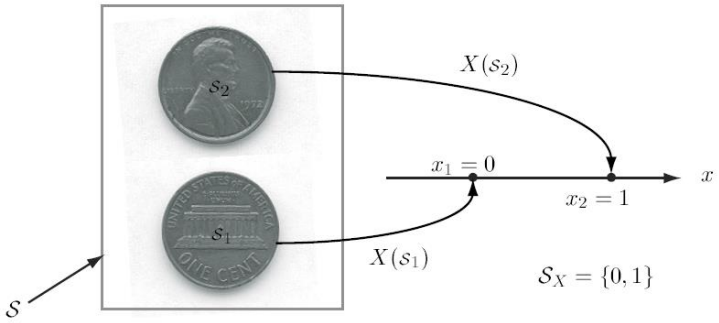
- A random experiment:
  - Outcome varies in an unpredictable fashion when the experiment is repeated under the same conditions.
  - Requires an experimental procedure and a set of observations
- The sample space, S, of a random experiment is the set of all possible outcomes.
  - An event, A, consists of any subset of outcomes in the sample space.
- Well known distributions (like a uniform) provide us with simple math functions that generate the probabilities of outcomes
  - Probability theory provides the mathematics to apply probability to model real-world problems!
- Example random experiments
  - flip a coin 3 times (discrete example).
    - S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
    - Define an event, A, : A: number of times we get 3 heads
      - P[A]= 1 / 8
  - Select a number from a random number generator (continuous example):
    - S = [0,1]. The numbers drawn will be of distribution uniform



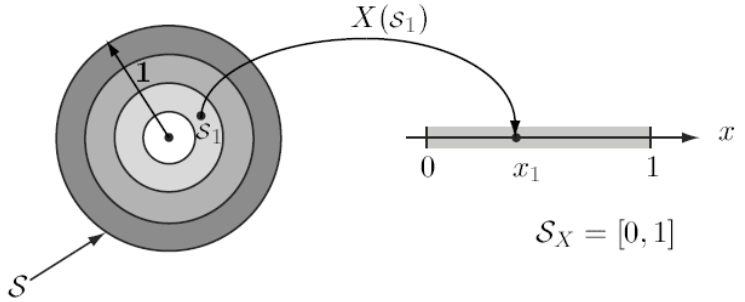
Random Variable

•The concept of RV's formalizes the task of assigning a numeric value to the outcome of random experiments.
•The domain of a random variable (S) is to assign real numbers to outcomes in the sample space of a random experiment
•The range of a RV X (Sx) represents the possible values of X
•RV's are either

- Discrete: Sx takes on finite or countably infinite set of values
- Continuous: Sx is infinite AND uncountable



Discrete experiment: Flip a coin, RV X maps S to two possible values Sx= {0,1}



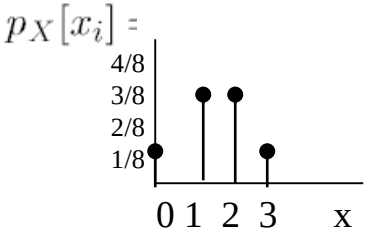
Continuous experiment: Throw a dart, the score is the distance of S1 from the center

Random Variables : Discrete

- Flip a coin 3 times, define X to represent the number of heads in three sequential tosses.
  - X assigns each possible outcome (i.e., the S from the previous page) to a number in Sx : {0,1,2,3}
  - With a little bit of math, we can find the probability generator for Sx.
    - Probability we don't see ANY heads:
      - P[X=0]=P[{TTT}] = 1/8
- It is easy to obtain the probabilities for each possibility.

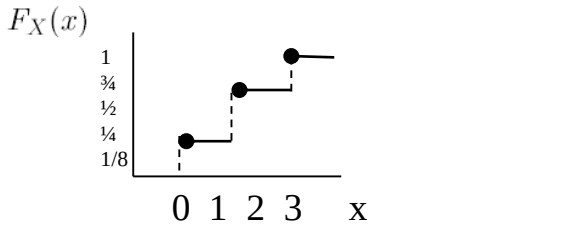
Probability Mass Function

$$p_X[x_i] = P[X(s) = x_i]$$



Cumulative Distribution Function

$$F_X(x) = P[X \le x] \quad -\infty < x < \infty.$$

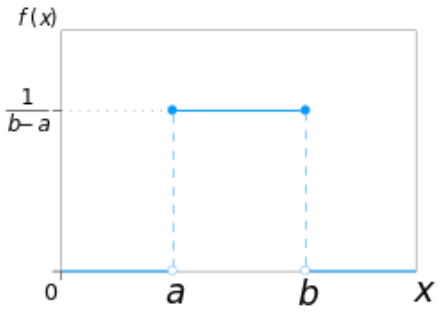


Random Variables : Continuous

- Uniform random variable : randomly select a number in the range of : Sx=[a,b]
  - The term random number generally means a uniform in the range of [0,1].

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

$$\begin{aligned} P[a \le X \le b] &= \int_a^b p_X(x)dx \\ &= \int_a^b 1\,dx \end{aligned}$$



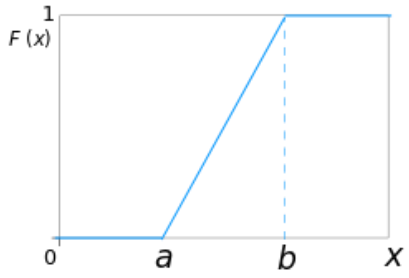
Uniform RV probability density function (PDF)

$$F_X(x) = P[X \le x] \qquad -\infty < x < \infty$$

$$F_X(x) = \int_{-\infty}^{x} p_X(t)dt \qquad -\infty < x < \infty.$$

$$F_X(x) = \begin{cases} 0 & x \le a \\ \int_a^x \frac{1}{b-a}dt & a < x < b \\ 1 & x \ge b \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x \le a \\ \frac{1}{b-a}(x-a) & a < x < b \\ 1 & x \ge b\,. \end{cases}$$



Cumulative distribution function (CDF)

Well Known Random Variables

•Discrete: binomial, geometric, poisson
•Continuous: Uniform, exponential, normal
•Exponential:

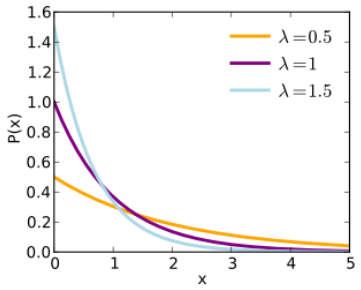
- Definition: Exponential distribution with parameter λ:

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

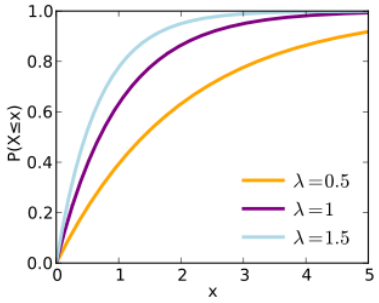
- The cdf:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- Mean $E(X) = 1/\lambda$.



Exponential RV PDF



Exponential RV CDF

•Expected Value (also referred to as the mean or average)
•Discrete RV X :

$$E[X] = \sum_i x_i p_X[x_i]$$

•Continuous RV X: $$E[X] = \int_{-\infty}^{\infty} x p_X(x) dx$$

•Easy to show that for a uniform RV X,
- Mean E[X] = integ x*(a-b)/2 dx = (a-b)/2

•For an Exponential:

- Mean $E(X) = 1/\lambda$.

Random Number Generators

- Rather than generating random variables, we generate random variates.
- Random variates generated from the uniform distribution U[0,1] are called random numbers.
- Random variates from other distributions (e.g., exponential) and realizations from various random processes (e.g., Poisson process) can be obtained by transforming IID RN's in particular ways.

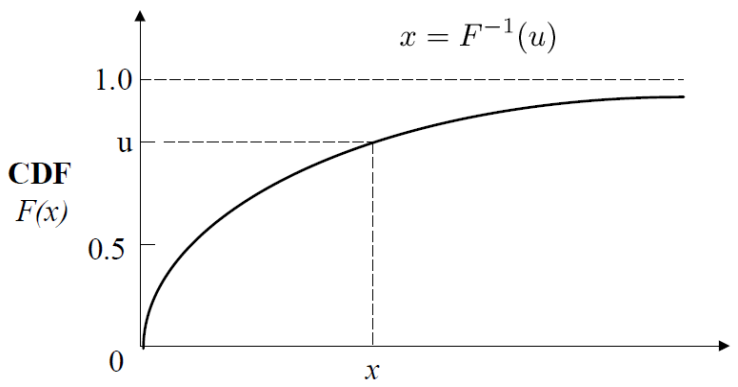
Random Number Generators

- Three important topics:
  - Elemental RNG's (i.e., U[0,1]) ← This is what we focus on – how to build a RNG
  - Testing RNG's –
    - How do we know if a RNG is 'good' ?
    - Properties include
      - infinite sample space, probability of generating the same number twice should be 0
      - Each selected random number is independent from previous selections
  - Generating Random Variates – one method Inverse Transform

- Used when F^{-1} can be determined either analytically or empirically.

$$u = F(x) \sim U(0,1)$$

$$x = F^{-1}(u)$$



- For exponential variates:

The pdf $f(x) = \lambda e^{-\lambda x}$
The CDF $F(x) = 1 - e^{-\lambda x} = u$ or, $x = -\frac{1}{\lambda}\ln(1-u)$

- If u is U(0,1), 1-u is also U(0,1)
- Thus, exponential variables can be generated by:

$$x = -\frac{1}{\lambda}\ln(u)$$

Random Number Generators

- At least three methods for obtaining RNs:
  - Read from a table of RN's.
  - Use a physical device
  - Employ a recursive equation which generates the (i+1)st RN from previous RNs.
    - Deterministic therefore 'pseudorandom number'
    - The seed represents the location within the repeatable set of numbers

Random Number Generators

- Properties of a U(0,1) RN Generator:
  - Uniformly distributed in the interval (0,1)
  - The RNs should be independent….we will look at the mathematical definition of ‘correlation’ later
  - Many RNs should be generated before the cycle repeats (ideally it exhibits full period)
  - Reproducible and allow multiple streams
  - Consumes minimal cpu and memory resources

RC4 RNG

Input: a key (seed) of length up to 256 bytes
Output: a random byte (to be XORed with plaintext)

Initialization: Let K0, K1, K2, ... K255 be the bytes of the key, repeating the key as necessary
j = 0
for i = 0 to 255
    Si = i
for i = 0 to 255
    j = (j + Si + Ki) mod 256
    swap Si and Sj
i = 0
j = 0

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i.$$

Algorithm:
i = (i + 1) mod 256
j = (j + Si) mod 256
swap Si and Sj
t = (Si + Sj) mod 256
output St

Reference:

Applied Cryptography, by Bruce Schneier, Wiley, 1996