18-847F: Special Topics in Computer Systems

Foundations of Cloud and Machine Learning Infrastructure

Lecture 3: Review of Probability Theory

Foundations of Cloud and Machine Learning Infrastructure

TO DO

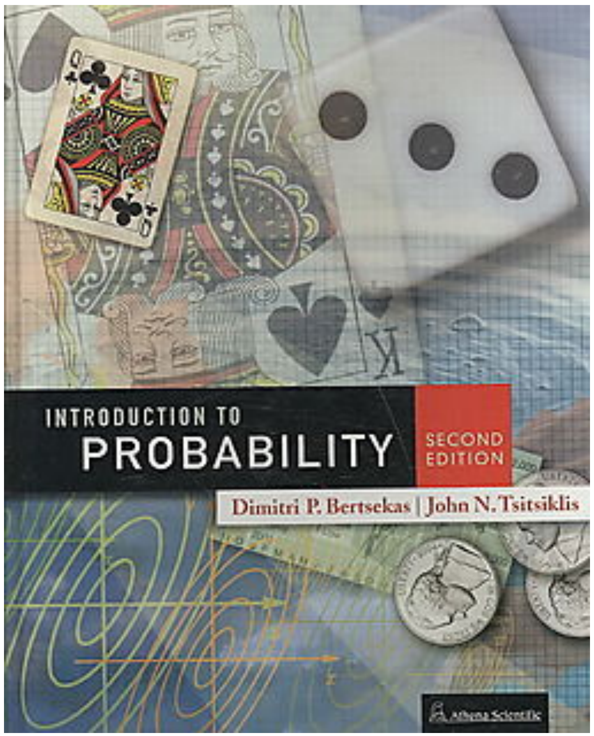
- Sign-up for Class Presentations

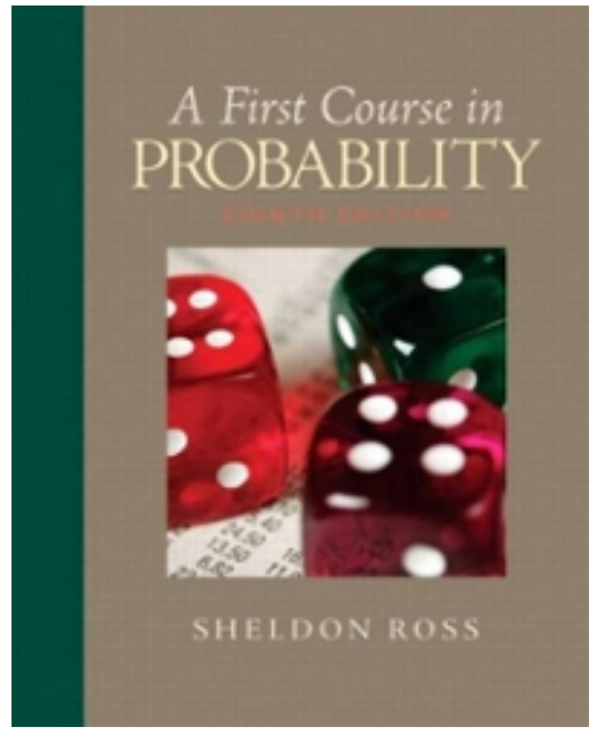
- First 2 assignments (due Sept 10th and Sept 12th) are on Canvas

Aim of this Lecture

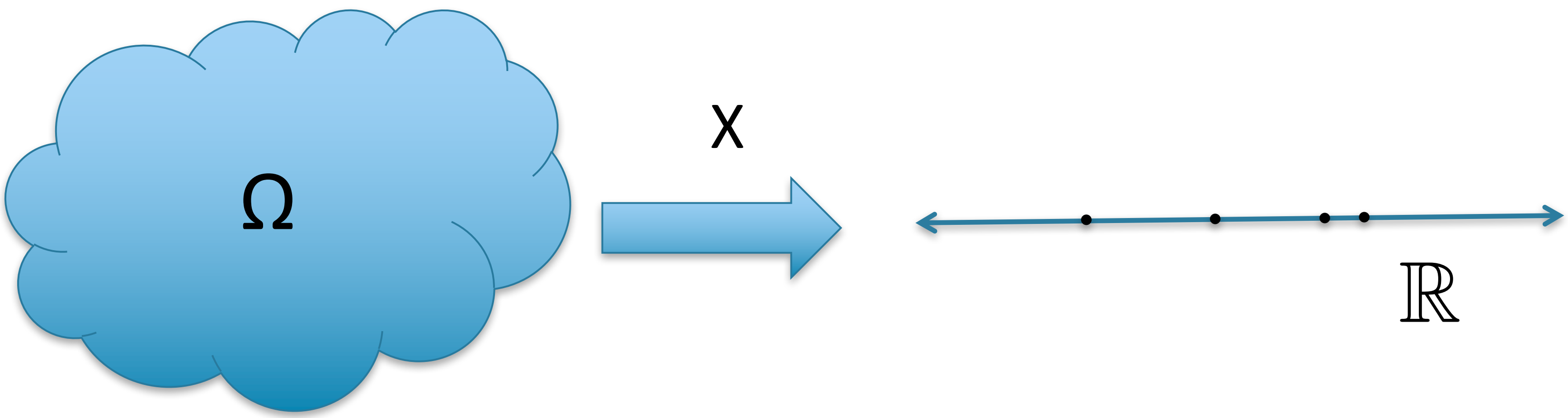
- Review undergraduate probability relevant to this class
- We will solve practice exercises during the class
- I will go fast, assuming knowledge of undegraduate probability. Please stop and ask questions if you don't follow !

Reference Textbooks





Discrete Random Variable



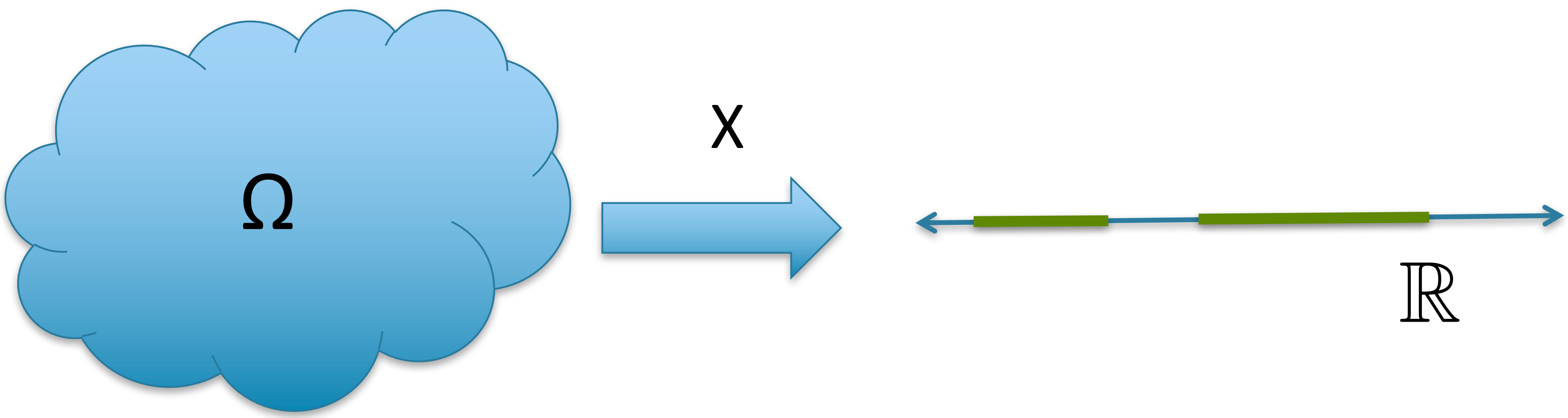
Probability Mass Function $P_X(x)$ = Probability that X takes value x

1. Bernoulli: Ω = {heads, tails}, X(heads) = 1 , X(tails) = 0

2. Two-Dice Roll, Ω = {(1,1), (1,2), .. (6,6)}}

   $X(n_1, n_2) = n_1 + n_2$

Continuous Random Variable



Examples of Continuous Random Variables: $f_X(x)$

1. Exponential $\quad f_X(x) = \mu e^{-\mu x} \quad \text{for all } x \geq 0$
2. Gaussian $\quad f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Cumulative and Tail Distribution Functions

Cumulative Dist (CDF)

$$F_X(x) = \Pr(X \leq x)$$

Tail Dist (CCDF)

$$\bar{F}_X(x) = \Pr(X > x)$$

Q1: What is the CDF and CCDF of X with the following PMF?

$$X = \begin{cases} 1 & \text{w.p. } 0.5 \\ 2 & \text{w.p. } 0.3 \\ 3 & \text{w.p. } 0.2 \end{cases}$$

Cumulative and Tail Distribution Functions

Cumulative Dist (CDF)

$$F_X(x) = \Pr(X \leq x)$$

Tail Dist (CCDF)

$$\bar{F}_X(x) = \Pr(X > x)$$

Q2: What is the CDF and CCDF of X ~ exp(μ)?

$$f_X(x) = \mu e^{-\mu x} \quad \text{for all } x \geq 0$$

Cumulative and Tail Distribution Functions

Cumulative Dist (CDF)

$$F_X(x) = \Pr(X \leq x)$$

Tail Dist (CCDF)

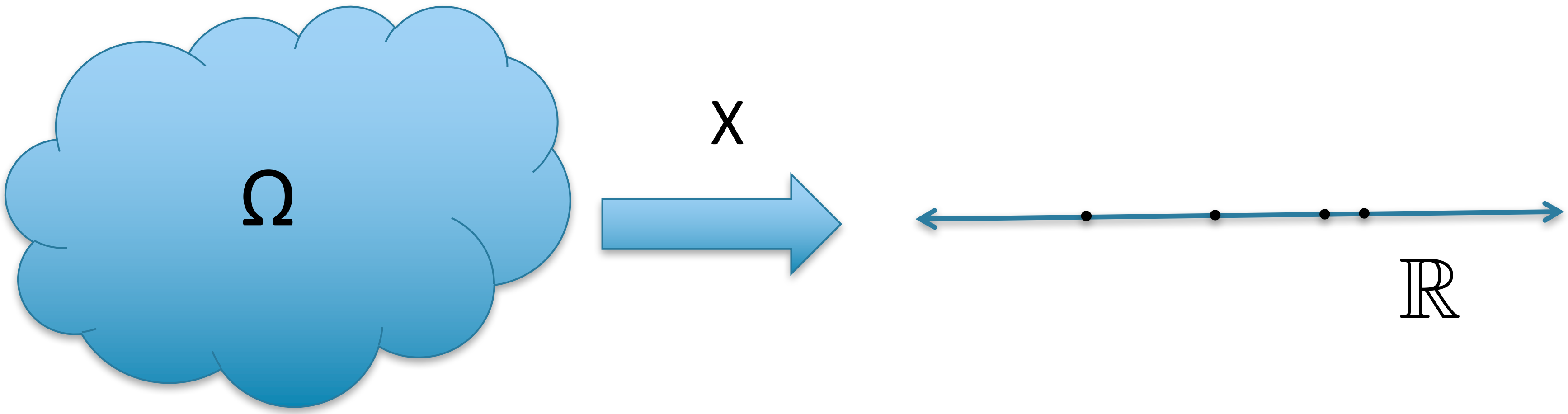
$$\bar{F}_X(x) = \Pr(X > x)$$

Q2: What is the CDF and CCDF of X ~ exp(μ)?

$$F_X(x) = 1 - e^{-\mu x} \qquad \text{for all } x \geq 0$$

$$\bar{F}_X(x) = e^{-\mu x} \qquad \text{for all } x \geq 0$$

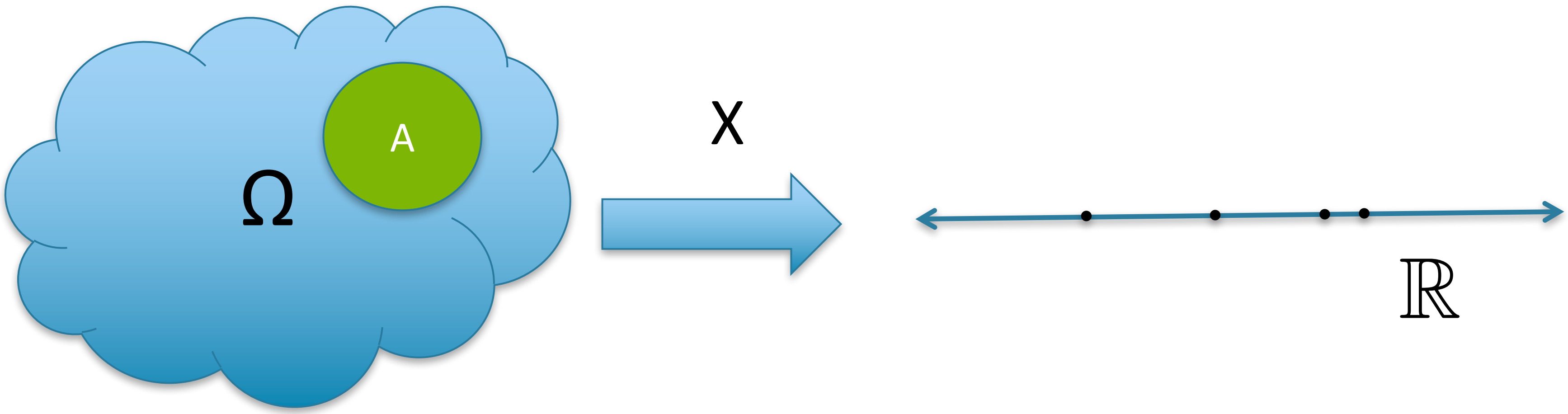
Expectation and Variance



$$\mathbb{E}[X] = \sum_x x \Pr(X = x) = \mu_X$$

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

Conditional Probability



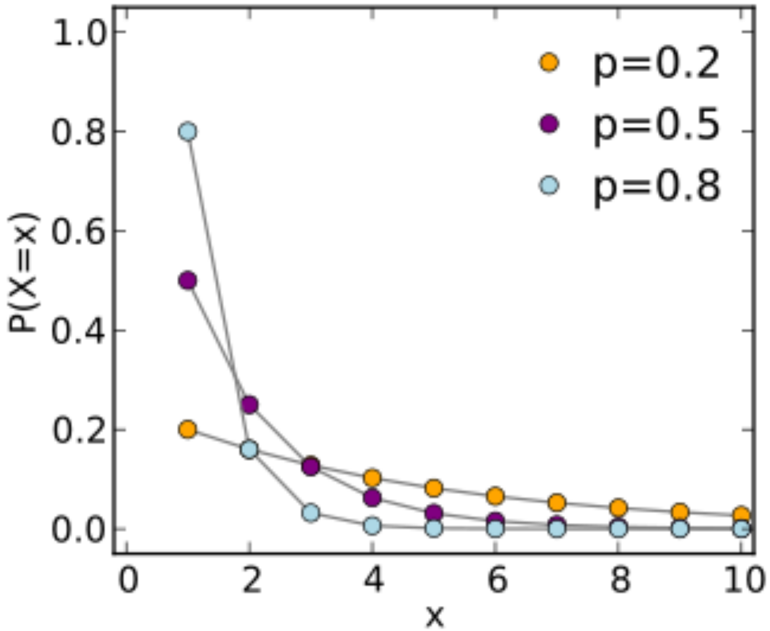
$$\Pr(X = x|A) = \frac{\Pr(X \cap A)}{\Pr(A)}$$

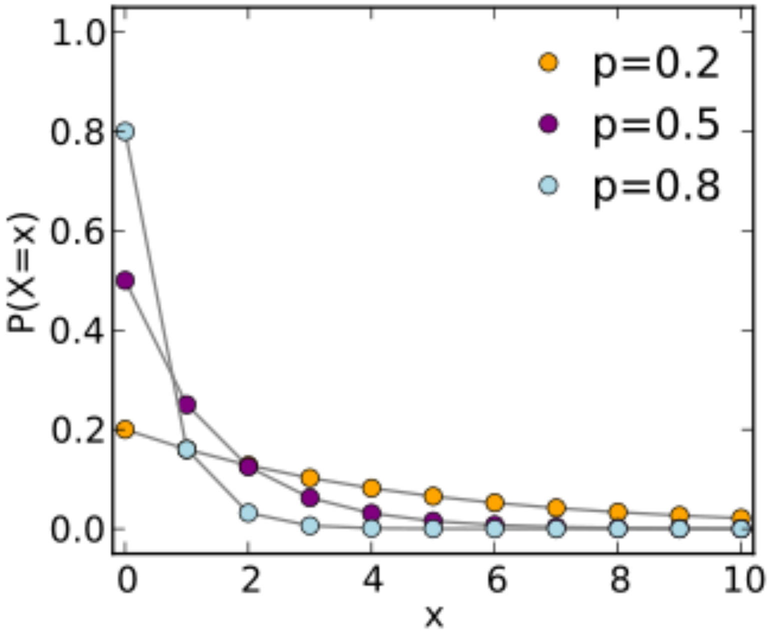
Two-dice Example: Given that both die rolls are <= 3, what is the probability that the sum is 5?

Geometric Random Variable

$$\Pr(X = k) = (1-p)^{k-1}p \quad \text{for all } k = 1, 2, \ldots$$

Example: Number of tosses of a coin with bias p (probability of H), until we get the first H





Geometric Random Variable

$$\Pr(X = k) = (1-p)^{k-1}p \quad \text{for all } k = 1, 2, \ldots$$

Example: Number of tosses of a coin with bias p (probability of H), until we get the first H

Q: What is E[X]?

Geometric Random Variable

$$\Pr(X = k) = (1-p)^{k-1}p \quad \text{for all } k = 1, 2, \ldots$$

Example: Number of tosses of a coin with bias p (probability of H), until we get the first H

Q: What is E[X]?

A: E[X] = 1/p

Geometric Random Variable

$$\Pr(X = k) = (1-p)^{k-1} p \quad \text{for all } k = 1, 2, \ldots$$

Example: Number of tosses of a coin with bias p (probability of H), until we get the first H

Q: What is E[X]?

A: E[X] = 1/p

Memoryless Property:

$$\Pr(X > x + s | X > s) = \Pr(X > x) \quad \text{for all } x, s \geq 0$$

Exercise: Coupon Collector's Problem

Given n different coupons, how many coupons do you expect you need to draw with replacement before having drawn each coupon at least once?

Exercise: Coupon Collector's Problem

Given n different coupons, how many coupons do you expect you need to draw with replacement before having drawn each coupon at least once?

# of Draws until we get the i^{th} unique coupon

$$\Pr(X_i = k) = \left(\frac{i-1}{n}\right)^{k-1}\left(\frac{n-i+1}{n}\right) \quad \text{for all } k = 1, 2, \ldots$$

Exercise: Coupon Collector's Problem

Given n different coupons, how many coupons do you expect you need to draw with replacement before having drawn each coupon at least once?

# of Draws until we get the i^{th} unique coupon

$$\Pr(X_i = k) = \left(\frac{i-1}{n}\right)^{k-1}\left(\frac{n-i+1}{n}\right) \quad \text{for all } k = 1, 2, \ldots$$

$$\mathbb{E}[X_i] = \frac{n}{n-i+1}$$

Exercise: Coupon Collector's Problem

Given n different coupons, how many coupons do you expect you need to draw with replacement before having drawn each coupon at least once?

Expected # of Draws until we get n unique coupons

$$\mathbb{E}[T] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]$$

Exercise: Coupon Collector's Problem

Given n different coupons, how many coupons do you expect you need to draw with replacement before having drawn each coupon at least once?

Expected # of Draws until we get n unique coupons

$$
\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] \\
&= \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} \\
&= nH_n \\
&\simeq n \log n
\end{aligned}
$$

Exercise: Review of Discrete r.v.s

Example: Suppose you have n disks. Each dies independently with probability p every year

-State of a particular disk after one year?

Is it a random variable? If yes, which type?

-Number of disks that die in year one?

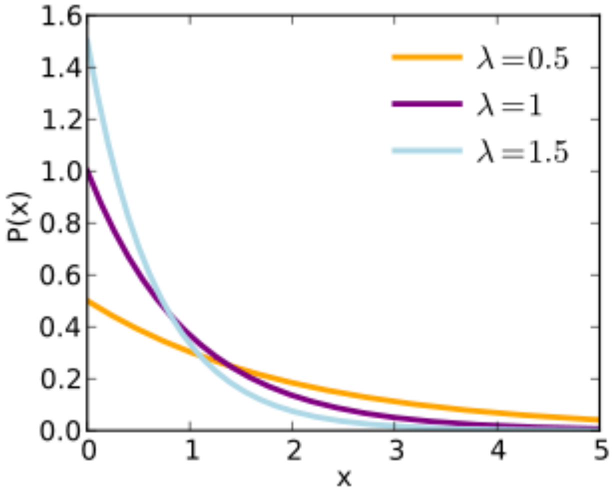
Is it a random variable? If yes, which type?

-Number of years until a particular disk dies?

Is it a random variable? If yes, which type?

Exponential Random Variable

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{for all } x \geq 0$$



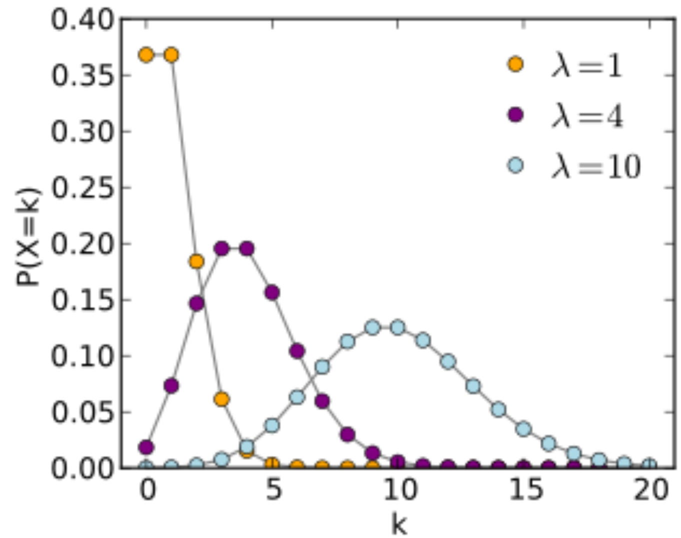
Expected Value and Variance: E[X] = 1/λ , Var[X] = 1/λ^2

Memoryless Property:

$$\Pr(X > x + s | X > s) = \Pr(X > x) \quad \text{for all } x, s \geq 0$$

Poisson Random Variable

$$\Pr(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$



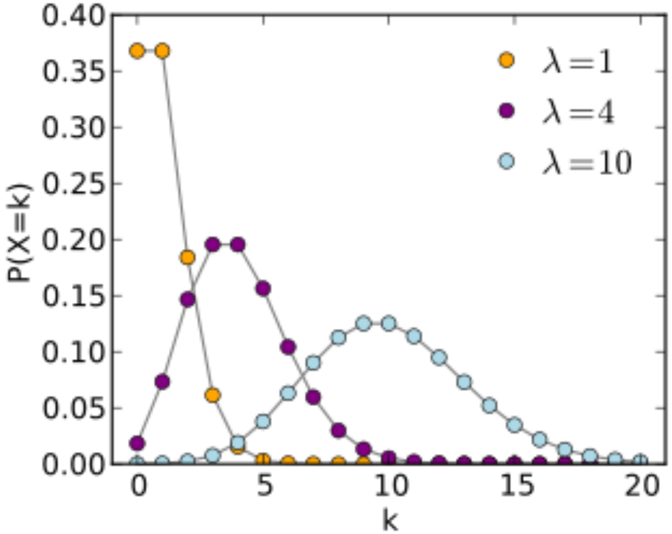
Expected Value and Variance: E[X] = λ , Var[X] = λ

Poisson Process: Given a time interval T and rate λ, the number of Poisson events that occur in that interval is X such that $\Pr(X = k) = e^{-\lambda T} \frac{(\lambda T)^k}{k!}$

Inter-arrival times of a Poisson process are exponential

Exercise: Poisson Random Variable

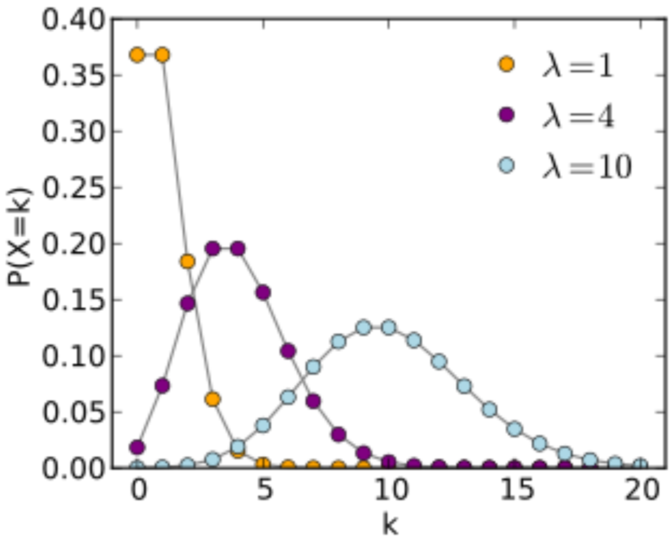
$$\Pr(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$



Overflow floods occur on a river once every 100 years on average. Calculate the probability of k = 0, 1, 2 floods in a 100-year interval given that floods follow a Poisson Process

Exercise: Poisson Random Variable

$$\Pr(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$



Overflow floods occur on a river once every 100 years on average. Calculate the probability of k = 0, 1, 2 floods in a 100-year interval given that floods follow a Poisson Process

Pr(X=0) = e^{-1} ~ 0.368

Pr(X = 1) = e^{-1} ~ 0.368

Pr(X = 2) = $e^{-1}/2$ = ~ 0.184

Order Statistics

Suppose we have n random variables $X_1, X_2, ..X_n$

Order-statistics are obtained by ordering the random variables

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

$$\begin{array}{ccc} || & & || \\ \min(X_1, \ldots X_n) & & \max(X_1, \ldots X_n) \end{array}$$

For i.i.d. random variables $X_1, X_2, ..X_n$ we denote the order statistics by

$$X_{1:n}, X_{2:n}, \ldots X_{n:n}$$

Order Statistics

Consider n i.i.d. exponentially distributed random variables X_1, X_2, ..X_n with rate μ where

$$f_X(x) = \mu e^{-\mu x} \quad \text{for } x > 0$$

- What is $\mathbb{E}[X_{1:n}] = \mathbb{E}[\min(X_1, X_2, \ldots X_n)]$?
- What is $\mathbb{E}[X_{n:n}] = \mathbb{E}[\max(X_1, X_2, \ldots X_n)]$?
- What is $\mathbb{E}[X_{k:n}]$?

Order Statistics

Consider n i.i.d. exponentially distributed random variables X_1, X_2, ..X_n with rate μ where

$$f_X(x) = \mu e^{-\mu x} \quad \text{for } x > 0$$

- What is $\mathbb{E}[X_{1:n}] = \mathbb{E}[\min(X_1, X_2, \ldots X_n)]$? $\frac{1}{n\mu}$
- What is $\mathbb{E}[X_{n:n}] = \mathbb{E}[\max(X_1, X_2, \ldots X_n)]$?

$$\frac{1}{n\mu} + \frac{1}{(n-1)\mu} + \ldots \frac{1}{\mu} \sim \frac{\log n}{\mu}$$

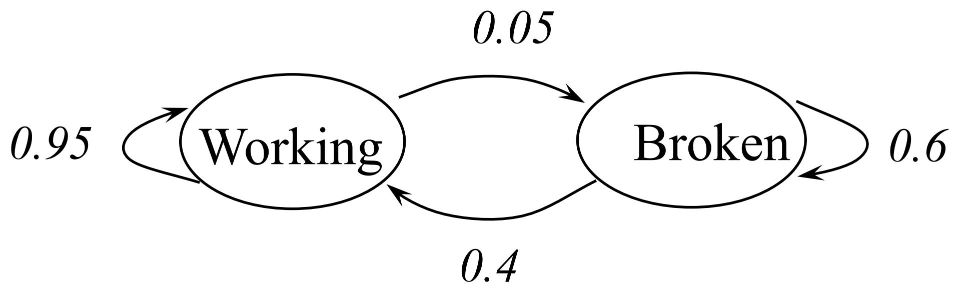
- What is $\mathbb{E}[X_{k:n}]$?

$$\frac{1}{n\mu} + \frac{1}{(n-1)\mu} + \ldots \frac{1}{(n-k+1)\mu} \sim \frac{1}{\mu} \log \frac{n}{n-k}$$

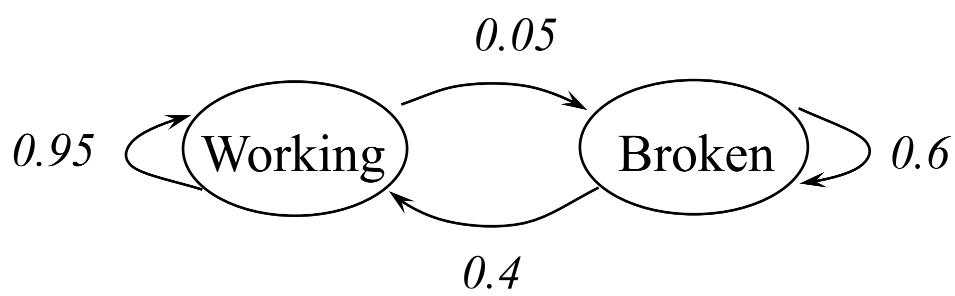
Discrete-time Markov Chains

A DTMC is a stochastic process $X_1, .. X_n$ where the n-th state X_n is independent of the past history, given the state X_{n-1}

Example: A machine is 'working' or 'broken'. If it is working today, it will work tomorrow 95% of times. If it is broken today, there is a 40% chance that it will be working tomorrow



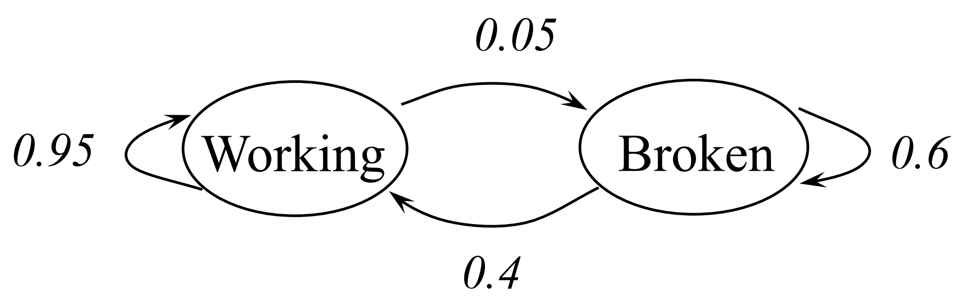
Transition Probability Matrix



$$(P_n(W)\ P_n(B)) = (P_{n-1}(W)\ P_{n-1}(B)) \boxed{\begin{pmatrix} 0.95 & 0.05 \\ 0.4 & 0.6 \end{pmatrix}}$$

P

Transition Probability Matrix



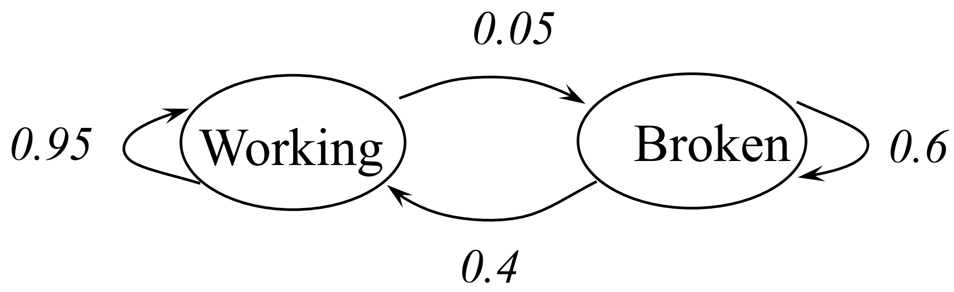
$$(P_n(W)\ P_n(B)) = (P_{n-1}(W)\ P_{n-1}(B)) \boxed{\begin{pmatrix} 0.95 & 0.05 \\ 0.4 & 0.6 \end{pmatrix}}$$

P

$$(P_n(W)\ P_n(B)) = (P_0(W)\ P_0(B)) \begin{pmatrix} 0.95 & 0.05 \\ 0.4 & 0.6 \end{pmatrix}^n$$

Steady-state Distribution

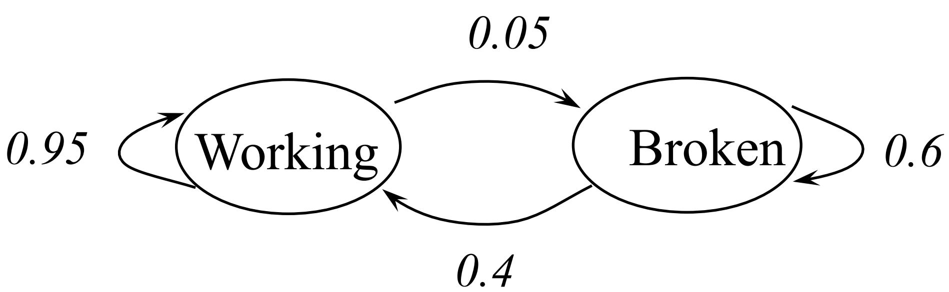
If we go away for a long time, and then come back and look at the machine, what is the probability that it will be working?



$$(P_n(W)\ P_n(B)) = (P_0(W)\ P_0(B)) \begin{pmatrix} 0.888 & 0.111 \\ 0.888 & 0.111 \end{pmatrix}$$

How to Find the Steady-state Distribution?

If we go away for a long time, and then come back and look at the machine, what is the probability that it will be working?



$$(\pi_W \ \pi_B) = (\pi_W \ \pi_B) \begin{pmatrix} 0.95 & 0.05 \\ 0.4 & 0.6 \end{pmatrix}$$

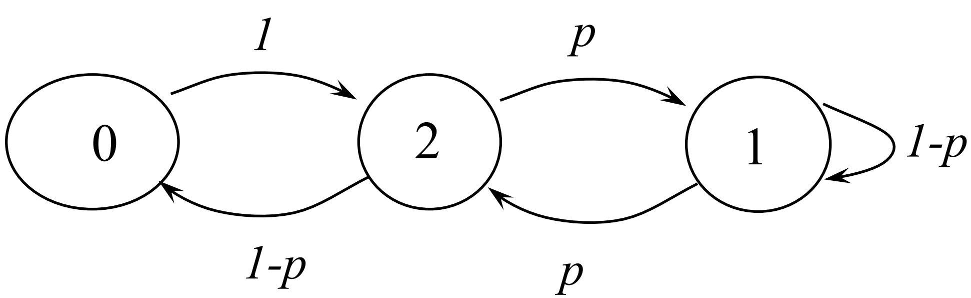
$$\pi_W + \pi_B = 1$$

Exercise: Two Umbrella Problem

A professor has 2 umbrellas that she uses when commuting between home and office. If it is raining and an umbrella is available at her location she takes it. If it is not raining, she does not take the umbrella. Suppose it rains with probability p every time she commutes, what is the probability that she won't have an umbrella during her commute.

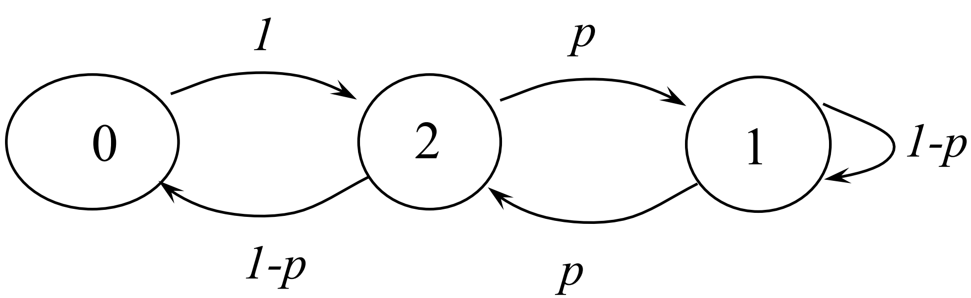
Exercise: Two Umbrella Problem

A professor has 2 umbrellas that she uses when commuting between home and office. If it is raining and an umbrella is available at her location she takes it. If it is not raining, she does not take the umbrella. Suppose it rains with probability p every time she commutes, what is the probability that she won't have an umbrella during her commute.



Exercise: Two Umbrella Problem

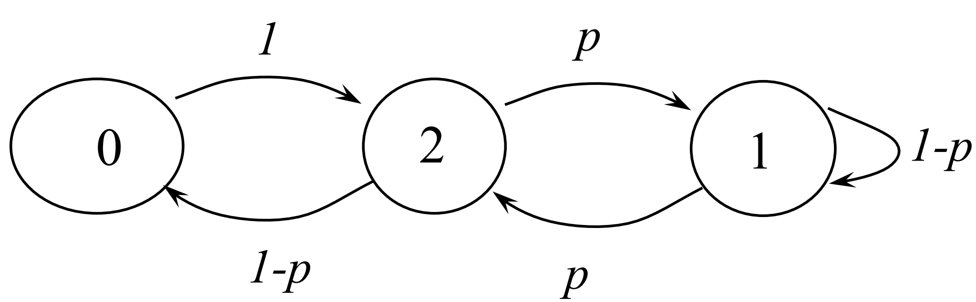
A professor has 2 umbrellas that she uses when commuting between home and office. If it is raining and an umbrella is available at her location she takes it. If it is not raining, she does not take the umbrella. Suppose it rains with probability p every time she commutes, what is the probability that she won't have an umbrella during her commute.



ANSWER: $$p \times \frac{1-p}{3-p}$$

Exercise: Two Umbrella Problem

A professor has 2 umbrellas that she uses when commuting between home and office. If it is raining and an umbrella is available at her location she takes it. If it is not raining, she does not take the umbrella. Suppose it rains with probability p every time she commutes, what is the probability that she won't have an umbrella during her commute.

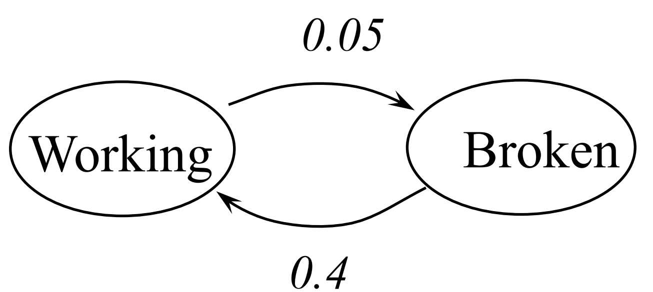


ANSWER: $$p \times \frac{1-p}{3-p}$$

Google's PageRank algorithm is also a DTMC problem

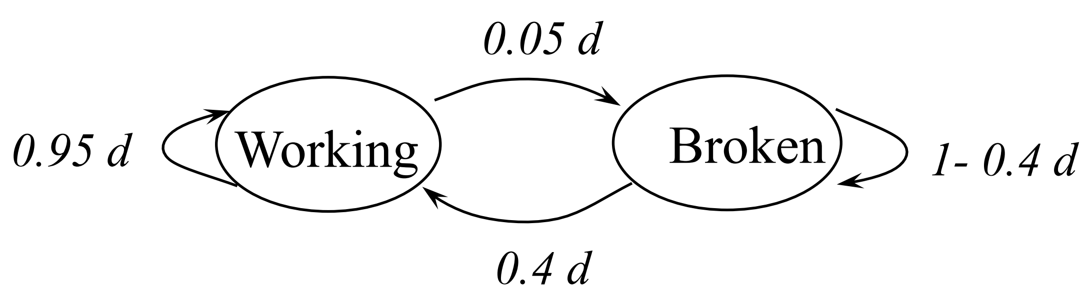
Continuous-time Markov Chains

Instead of discrete-time transitions X_1, .. X_n, here the transitions can occur at any time t. The time spent in any state is exponential with rate v_i ,which is the rate of exiting that state



Continuous-time Markov Chains

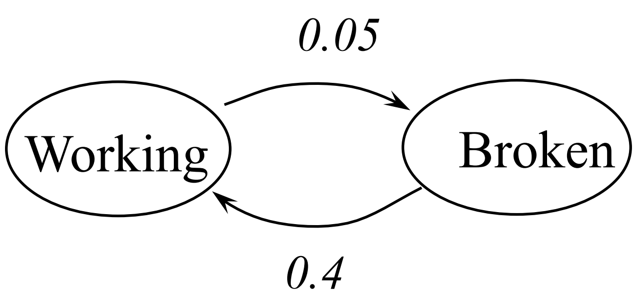
Instead of discrete-time transitions X_1, .. X_n, here the transitions can occur at any time t. The time spent in any state is exponential with rate v_i ,which is the rate of exiting that state



Equivalent Discrete-time Model

Steady-state Analysis

Instead of discrete-time transitions X_1, .. X_n, here the transitions can occur at any time t. The time spent in any state is exponential with rate v_i ,which is the rate of exiting that state



Steady-state Rate of exiting a state = Steady-state Rate of entering the state

Probability Review Summary

- Random variables, CDF, CCDF, Expectation, Variance
- Geometric Random Variable, Coupon Collector Problem
- Exponential Random Variable, Poisson Random Variable
- Order Statistics
- Discrete-time Markov Chains, Steady-state analysis
- Continuous-time Markov Chains