Ch9: Probabilistic Analysis and Randomized Algorithms

305234

Algorithm Analysis and Design

Jiraporn Pooksook
Naresuan University

Worst-case vs. Average-case running time

- We are normally interested in the worst case running time with input size n of an algorithm.
- For example, the worst case running time of insertion sort is when an input is in descending order for sorting from the lowest to the highest numbers.
- However, we could be interested in average case running time by measuring typical inputs.
- Typical inputs are assumed that all permutations of input are equally likely.
- Can we improve worst-case by adding randomization?

The Hiring Problem

- Suppose you need to hire a new office assistant.
- One candidate walks in each day.
- You will interview that person and decide to either hire that person or not.
- After interviewing, if that person is better qualified than your current assistant, you will fire the current assistant and hire the new applicant.
- You must pay a small fee if you don’t hire that applicant.
- You must pay a large fee if you hire that applicant.

The Hiring Problem

Pseudo code: Hire-Assistant(n)

```
best = 0 // dummy candidate
for i = 1 to n
        do interview candidate i
        if candidate i is better than candidate best
                then best = i
                        hire candidate i
```

We do not focus on the running time but more on the costs incurring by interviewing and hiring.

Cost of the Hiring Problem

- Let c_i is denoted as an interviewing cost.
- Let c_h is denoted as a hiring cost.
- Let m be the number of people hired.
- Total cost is $O(n\ c_i + m\ c_h)$
- In the worst-case, we hire every candidate that we interview. A total hiring cost will be $O(n\ c_h)$.
- It is reasonable to expect that the candidates do not always come in increasing order of quality.

Probabilistic Analysis

- Probabilistic analysis is the use of probability in the analysis of algorithm.
- It is commonly used to analyze the running time of algorithms.
- It can be used to analyze other quantities such as the cost of procedures.
- We must use knowledge of, or make assumptions about the **distributions of inputs** for using probabilistic analysis.
- We can then make an average-case analysis, averaging the cost over all possible inputs.

Probabilistic Analysis

- For the hiring problem, we can assume that the candidates walk in a random order. This means that we assume that we can compare any two candidates and decide which one is better; there is a total order on the candidates.
- Then we can rank each candidate with a unique number from 1 to n. We use rank(i) to denote the rank of applicant i.
- A higher rank corresponds to a better qualified applicant.
- The order list (rank(1), rank(2),…,rank(n)) is a permutation of the list (1,2,…,n).
- Therefore saying that the applicants come in a random order is equivalent to saying that this list of ranks is equally likely to be any one of the n! permutations of 1 to n (the ranks form a **uniform random permutation**; each of the possible n! permutations appears with equal probability).

Randomized Algorithms

- In the hiring problem, it may seem as if the candidates walks in a random order, but we cannot be sure about it.
- In order to develop a randomized algorithm for the hiring problem, we must control over the order in which we interview the candidates.
- Hence, we change the model by giving the list of candidates in advance. On each day we choose randomly which candidate to interview.

Randomized Algorithms: The Hiring Problem

```
Pseudo code: Randomized-Hire-Assistant(n)
randomly permute the list of candidates
best = 0 // dummy candidate
for i = 1 to n
        do interview candidate i
        if candidate i is better than candidate best
                then best = i
                        hire candidate i
```

Randomized Algorithms

- We call an algorithm **randomized** if its behavior is determined not only by its input but also by values produced by a **random-number generator**.
- For example, random(0,1) produces 0 or 1 with probability ½. Each integer returned by random is independent of the integers returned on previous calls.
- Most programming environments offer a (deterministic) **pseudorandom-number generator**: it returns numbers that "look" statistically random.

Randomized Algorithms

- We typically refer to the analysis of randomized algorithms by talking about the expected cost (ex: the expected running time).
- We can use probabilistic analysis to analyse randomized algorithms.

Basic of Probabilistic: Permutation

- A **permutation** of a finite set S is an ordered sequence of all the elements of S, with each element appearing exactly once.
- If S={a,b,c}, then there are 6 permutations of S:
  - abc, acb, bac, bca, cab, cba
- A **k-permutation** of S is an ordered sequence of k elements of S, with no element appearing more than once in the sequence.
- If S = {a,b,c,d}, then there are 12 2-permutations of S:
  - ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc

Basic of Probabilistic : Permutation

- The number of k-**permutation** of an n-set is

$$n(n-1)(n-2)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

- Since there are n ways of choosing the 1st element, n-1 ways of choosing the 2nd element and so on until k elements are selected, the last being a selection from n-k+1 elements.

Basic of Probabilistic: Combination

- A **k-combination** of an n-set S is a k-subset of S.
- If S={a,b,c,d}, then there are 6 2-combinations of S:
  - {a,b}, {a,c}, {a,d}, {b,c}, {b,d}, {c,d} or
  - ab, ac, ad, bc, bd, cd
- The number of k-combinations of an n-set can be expressed in terms of the number of k-permutations of an n-set:

$$\frac{n!}{k!(n-k)!}$$

Basic of Probabilistic Analysis: Binomial coefficient

- We use the notation "n choose k" to denote the number of k-combinations of an n-set.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- This formula is symmetric in k and n-k: $\binom{n}{k} = \binom{n}{n-k}$

- These numbers are known as **binomial coefficients** due to their appearrence in the binomial expansion:

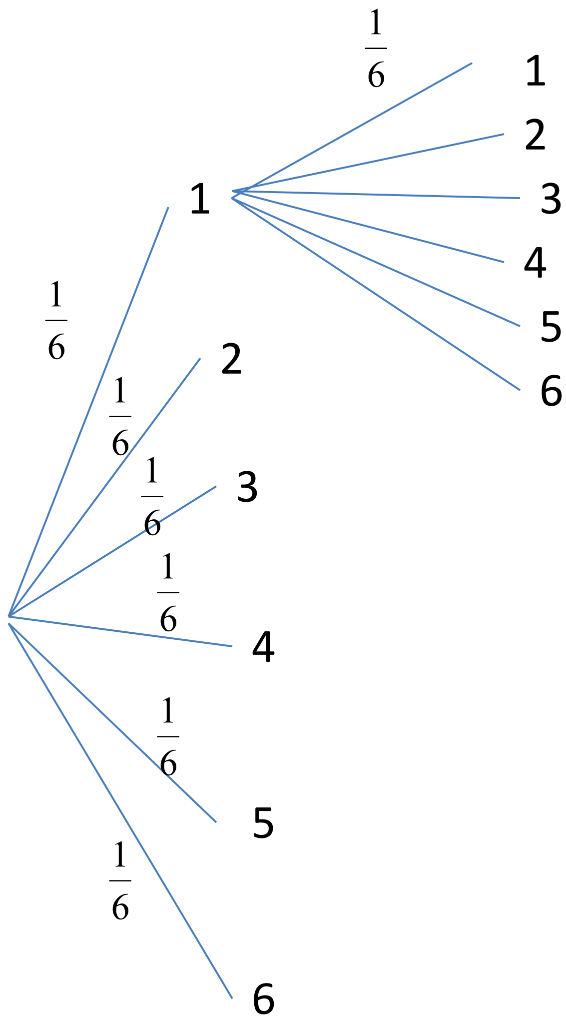
$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Basic of Probabilistic Analysis

- Consider rolling a dice and observing the results.
- We call this an experiment.
- It has 6 possible outcomes: 1,2,3,4,5,6
- Each of these outcomes has probability 1/6 (assuming fair dice)
- Again we roll two dice and there is 36 possible outcomes: 1-1, 1-2, 1-3,1-4,1-5,1-6,2-1,…,6-5,6-6.
- Each of these outcomes has probability 1/6 (assuming fair dice)
- What is the probability of the sum of dice being 7?

Add the probabilities of all the outcomes satisfying this condition: 1-6, 2-5, 3-4, 4-3, 5-2, 1-6 (probability is 1/6)

Basic of Probabilistic Analysis



Basic of Probabilistic Analysis

- A **sample space** S is a set whose elements are called elementary events, all possible outcomes.
- Each **elementary event** can be viewed as a possible outcome of an experiment.
- An **event** is a subset of the sample space S.
- For example, rolling two dice:
- A sample space S = {1-1, 1-2, 1-3, 1-4, 1-5, 1-6, 2-1,…,6-5,6-6}
- The event of obtaining same number of both dice is {1-1,2-2,3-3,4-4,5-5,6-6}.

Elementary event

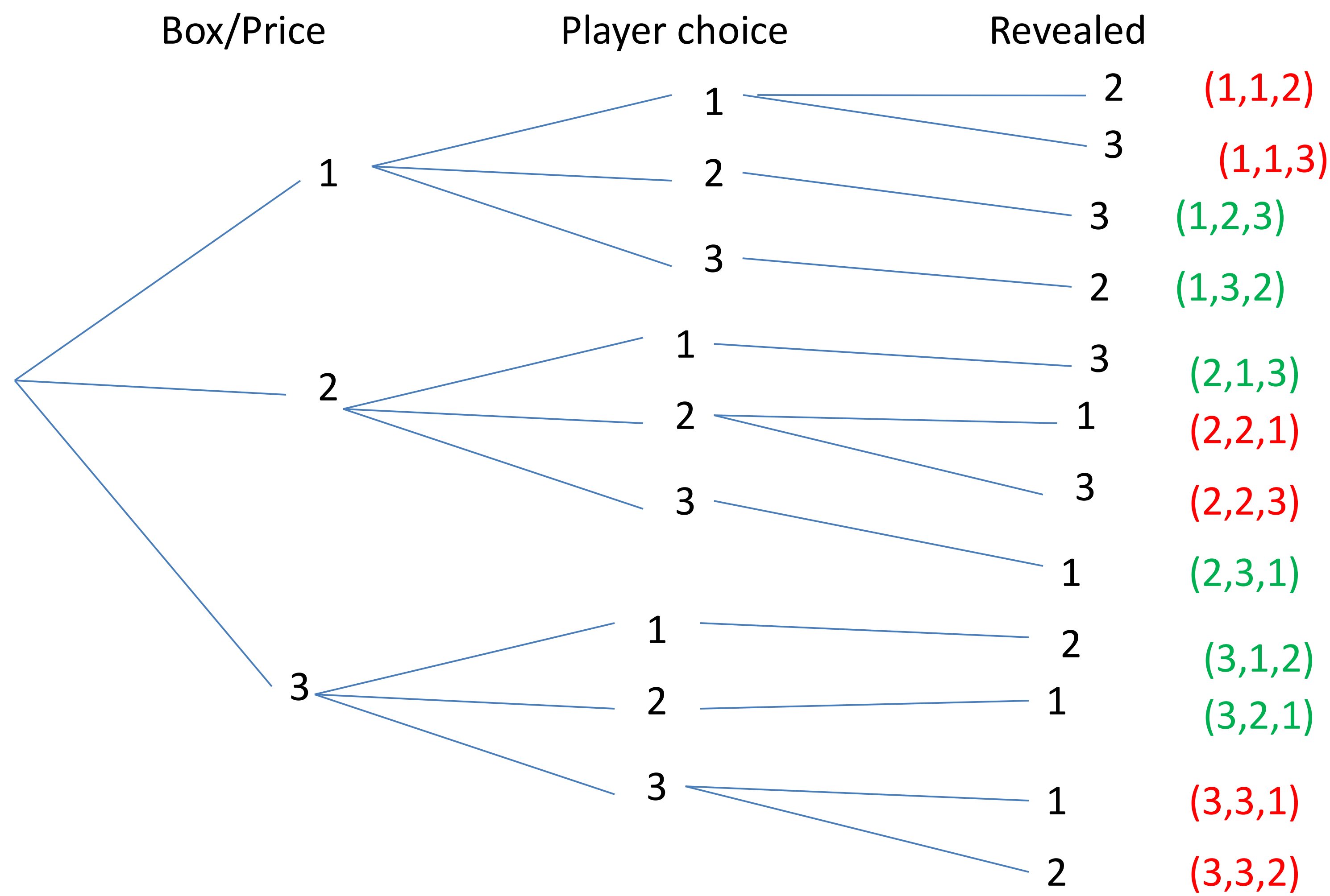
event

Example: Monty Hall Problem

- There are 3 doors and the big price is behind 1 door out of 3.
- The player chooses 1 door.
- One door is revealed.
- The player is asked to stay or change his/her choice.
- Question: The chance to win the price will be ½ or 2/3 if the player chooses to change?

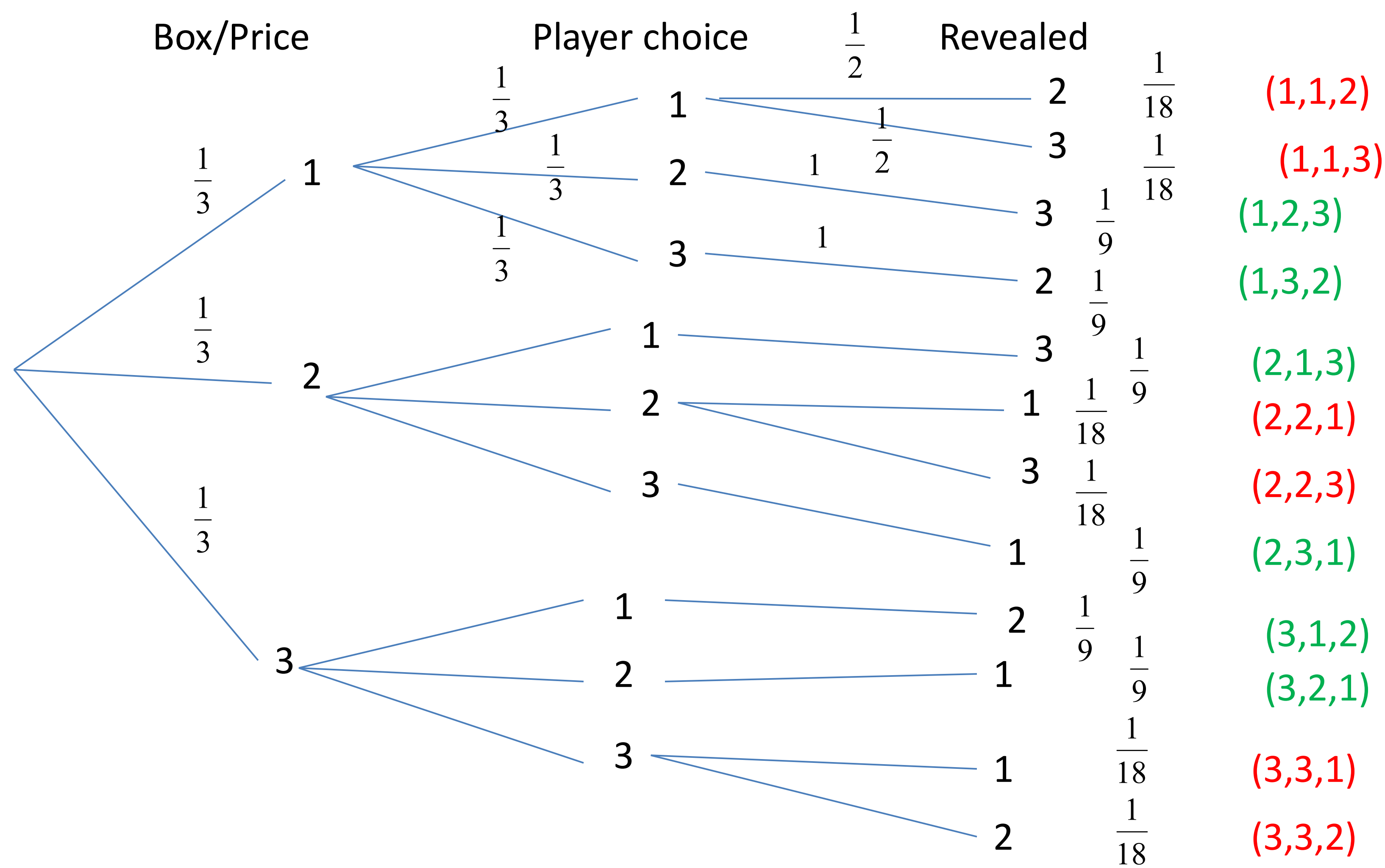
Example: Monty Hall Problem

If switch, pr(win)= ½?



Example: Monty Hall Problem

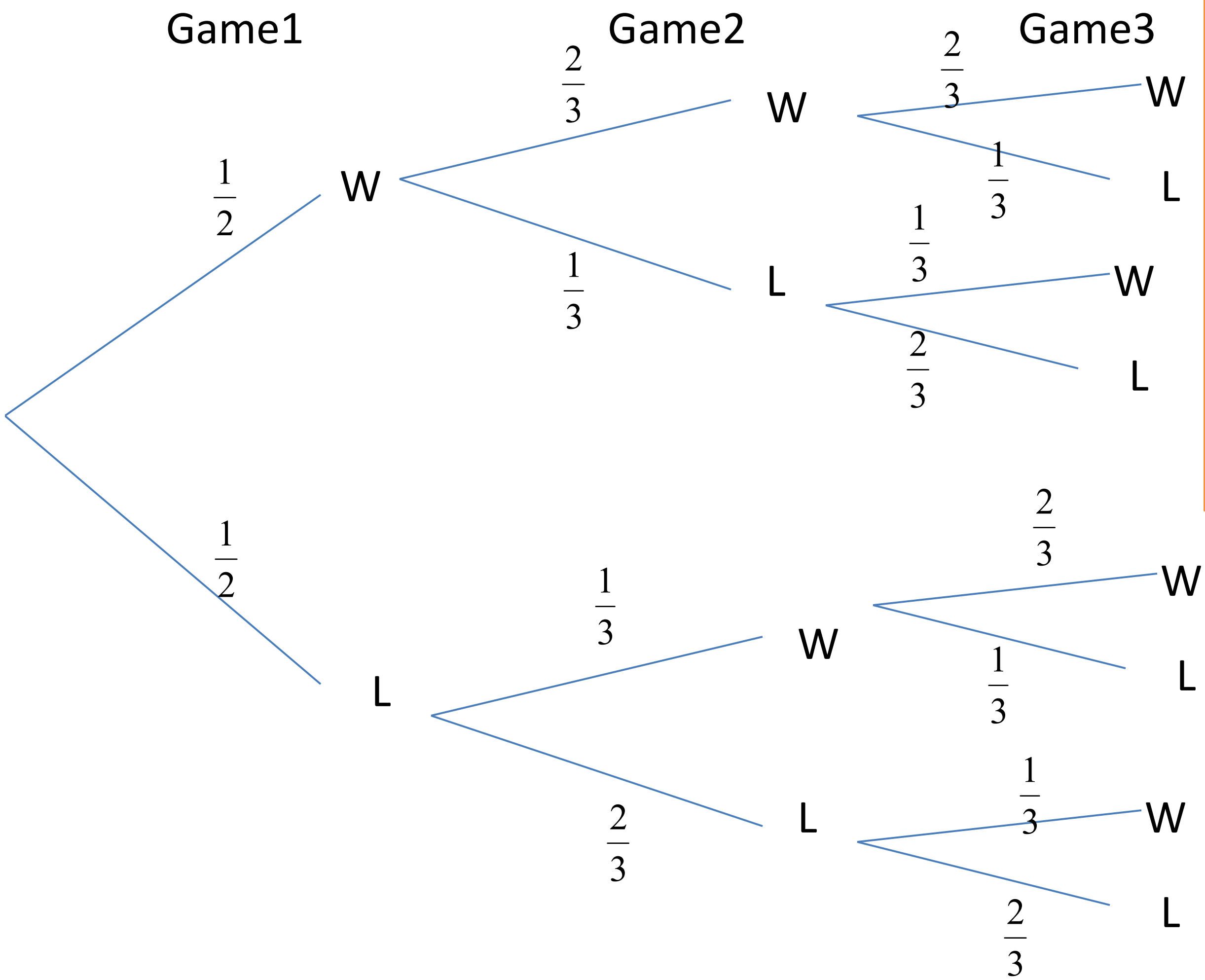
If switch, pr(win)= $\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}=\frac{2}{3}$



Example: 3 Games

- In a best 2 out of 3 series, the probability of winning the 1st game is ½. The probability of winning a game following a win is 2/3. The probability of winning a game after a loss is 1/3.

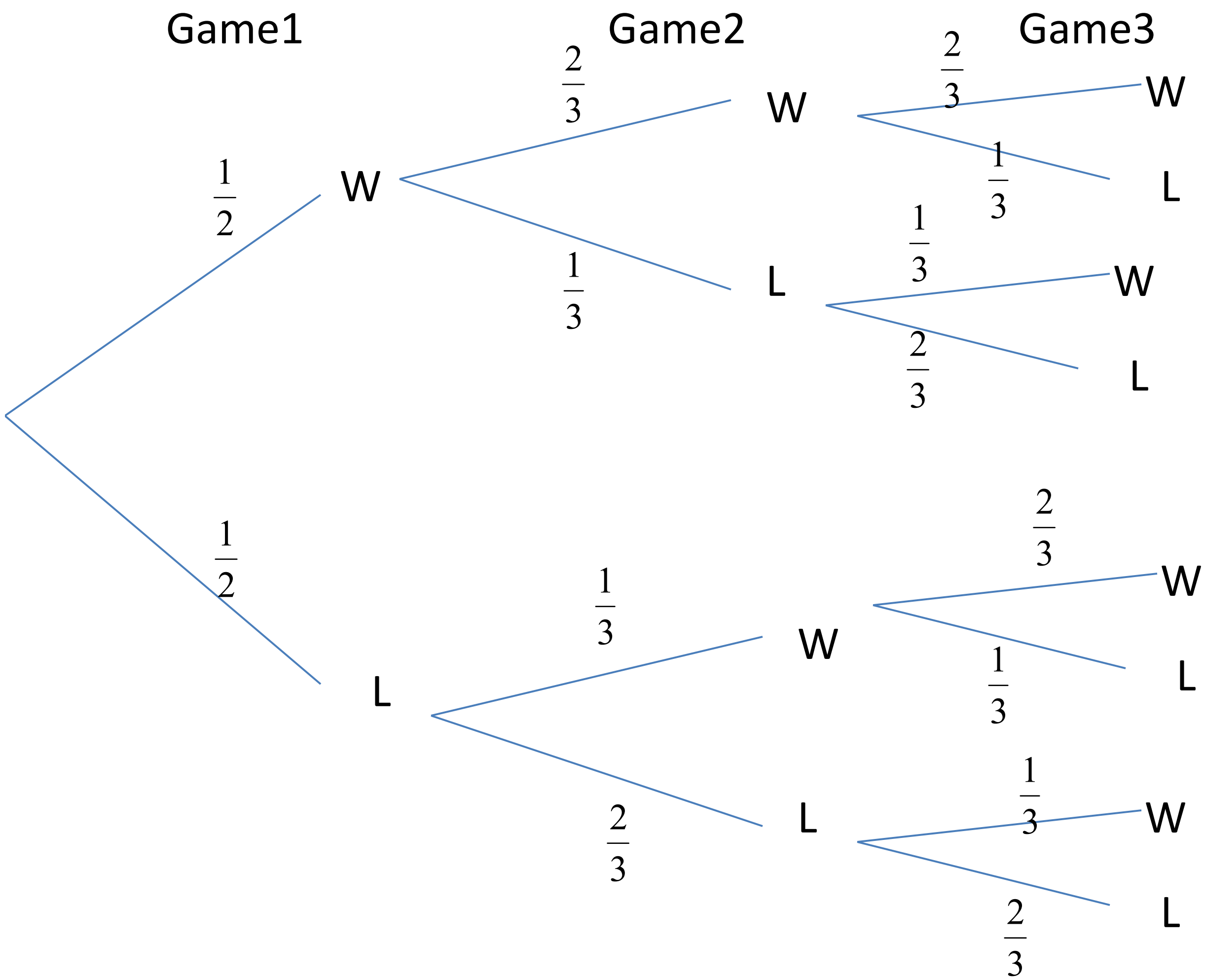
Example: 3 Games

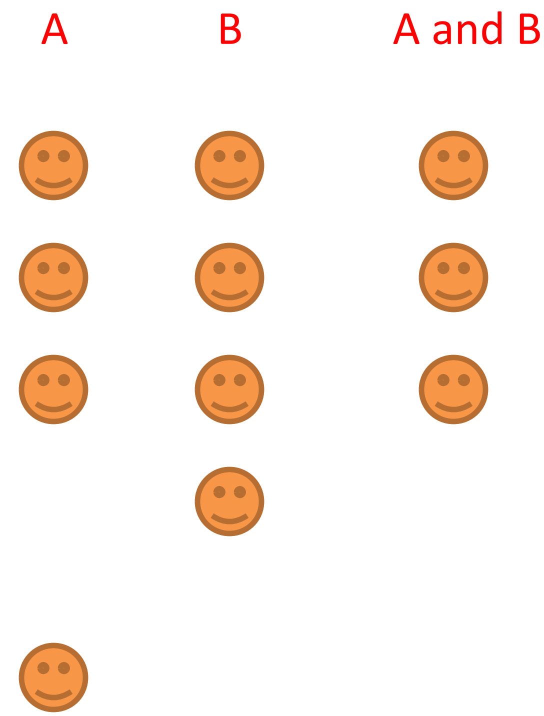


General product rule:
Pr(WW)= Pr(W1).Pr(W2|W1)
= ½ . 2/3 = 1/3
Pr(WLW)=Pr(W1).pr(L2|W1)
.Pr(W3|W1L2)
=Pr(W1).pr(L2|W1)
.Pr(W3|L2)
=1/2.1/3.1/3 = 1/18

Example: 3 Games

A = event win series
B = event win 1st game
Pr(A|B)=?





Pr(A|B)= P(A and B) / P(B)
= (2/9+1/9+1/18)/ (9/18)
= 7/9

Basic of Probabilistic Analysis

- We say that two events A and B are **mutually exclusive** if $A \cap B = \emptyset$.
- A **probability distribution Pr{}** on a sample space S is a mapping from events of S to real numbers such that the following **probability axioms** are satisfied:
  - Pr{A} ≥ 0 for any event A.
  - Pr{S} = 1
  - Pr{ $A \cup B$ } = Pr{A} + Pr{B} for any two mutually exclusive events A and B.

Basic of Probabilistic Analysis

- Suppose each of elementary events of tossing two dice has probability 1/36. Then the probability of getting same number on both dice is
- Pr{1-1,2-2,3-3,4-4,5-5,6-6} =

  Pr{1-1} + Pr{2-2}+Pr{3-3}+

  Pr{4-4}+Pr{5-5}+Pr{6-6}

  = 1/36*6 = 1/6

Basic of Probabilistic Analysis

- A **probability distribution is discrete** if it is defined over a finite or countably infinite sample space.
- Let S be a sample space. Then for any event A,

$$\Pr\{A\} = \sum_{s \in A} \Pr\{s\}$$

- Since elementary events are mutually exclusive. If S is finite and every elementary s ϵ S has probability

Pr{S} = 1/|S|

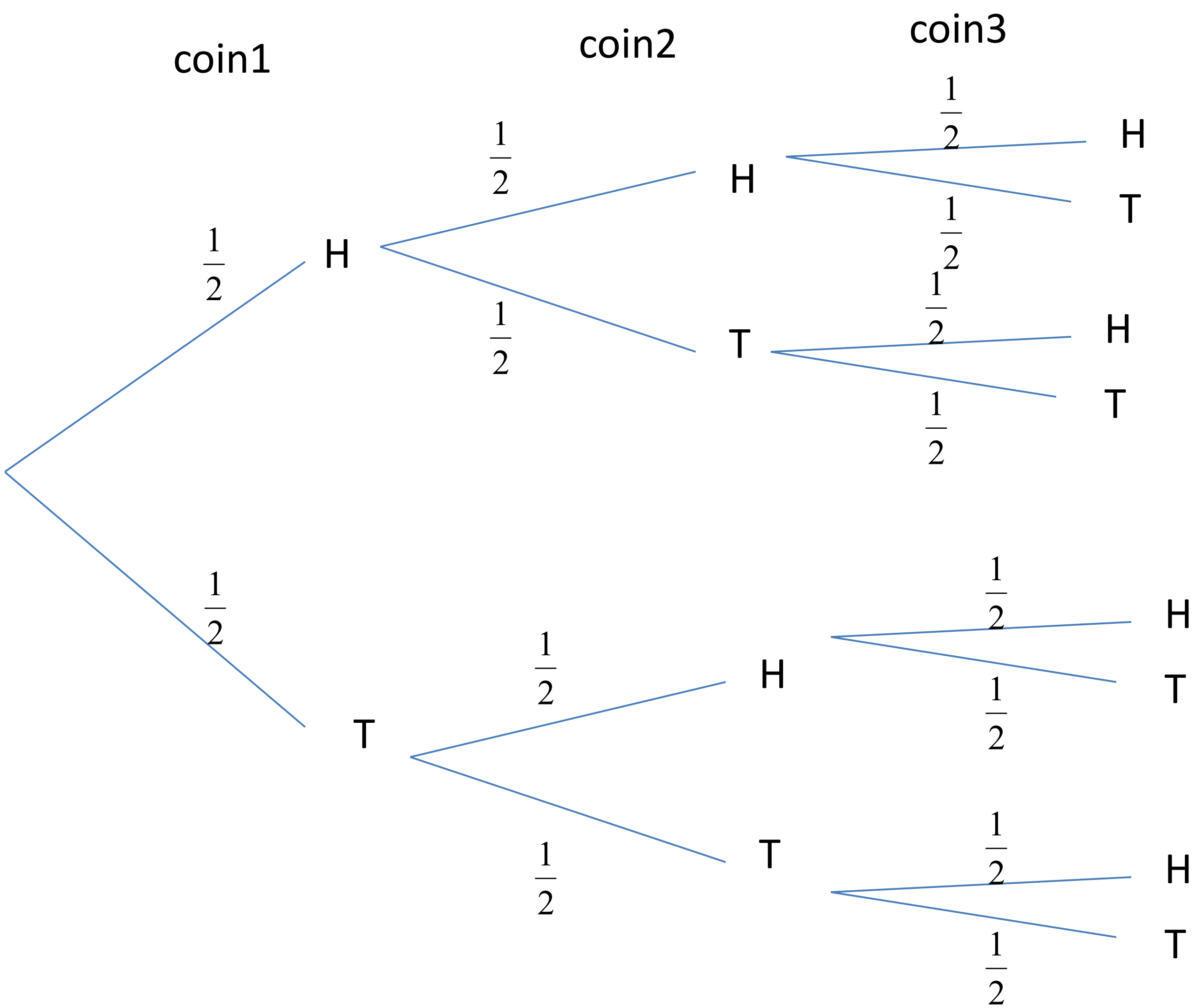
- Then we have the uniform probability distribution on S, as “picking an element of S at random”.

Basic of Probabilistic Analysis

- Consider the example of tossing a dice, the probability of obtaining each number is 1/6.
- If we roll the dice n times, we have the uniform probability distribution defined on the sample space $S = \{1,2,3,4,5,6\}^{n}$, a set of size 6^n.
- Each elementary event occurs with probability $1/6^n$.
- Therefore, the event A = {exactly n number of i occur for i = 1…6 } is a subset of S of size $|A| = 6$. The probability of event A is thus

$$Pr\{A\}=6/6^n = 1/6^{n-1}.$$

Example: Toss 3 coins

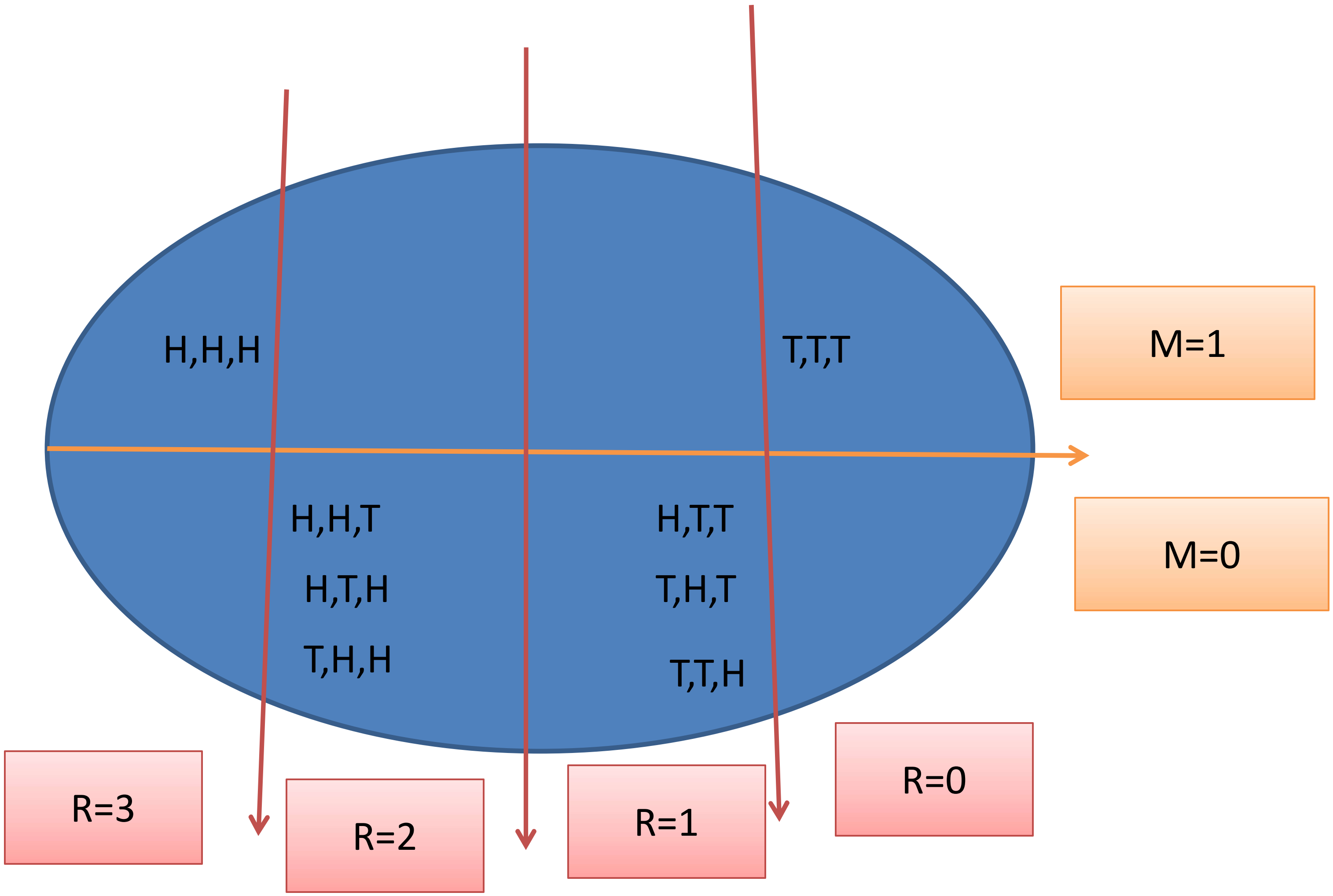


Random variable
R = number of H
R(H,T,H) = 2

Random variable
(Indicator random var or
Bernoulli random var)

M = 1 if all coins match
= 0 otherwise
M(H,H,T) = 0
M(T,T,T) =1

Random var vs Indicator random var



Basic of Probabilistic: Random variable

- A (discrete) **random variable** X is a function from a finite or countably infinite sample space S to the real numbers.
- For example, let X_1 be a random variable representing the result of the 1st dice, and X_2 represent the result of the 2nd dice.
- Let X be a random variable representing the sum of two dice: $X = X_1 + X_2$.

Basic of Probabilistic: Random variable

- For a random variable X and a real number x, we define the event X = x to be {s ϵ S: X(s) = x}, thus

$$\Pr\{X = x\} = \sum_{s \in S : X(s) = x} \Pr\{s\}$$

- The function f(x) = Pr{X=x} is the **probability density function** of the random variable X. From the probability axioms,

$$\Pr\{X=x\} \geq 0 \text{ and } \sum_{x} \Pr\{X = x\} = 1$$

Basic of Probabilistic: Random variable

- For example, rolling two dice there are 36 possible elementary events in the sample space. We assume that the probability distribution is uniform, so that each elementary event $s \in S$ is equally likely: $Pr\{s\} = 1/36$.
- Let X be the random variable representing the maximum of two values showing on the dice.
- We have $Pr\{X=3\} = 5/36$ since the possible elementary events are {1-3, 2-3, 3-3, 3-2, 3-1}.

Basic of Probabilistic: Random variable

- If X, Y are random variables, the function F(x,y) = Pr{X=x and Y=y} is the **joint probability density function** of X and Y.
- For a fix value y,

$$\Pr\{Y=y\} = \sum_x \Pr\{X = x \text{ and } Y = y\}$$

- And similarly for a fix value x,

$$\Pr\{X=x\} = \sum_y \Pr\{X = x \text{ and } Y = y\}$$

Two random variables X, Y are independent if for all x and y, the event X=x and Y=y are independent:

$$\Pr\{X = x \text{ and } Y = y\} = \Pr\{X = x\} \Pr\{Y = y\}$$

Basic of Probabilistic: Expectation

- The expected value(or, expectation or mean) of a discrete random variable X is

$$E[X] = \sum_x x \Pr\{X = x\}$$

- For example, in a game of flipping two fair coins. You earn \$3 for each head but lose \$2 for each tail. The expected value of X is
- E[X] = 6.Pr{HH} + 1.Pr{1H,1T} – 4. Pr{TT}
  = 6(1/4) + 1(1/2) – 4(1/4) = 1

In average you will earn \$1 for this game.

Basic of Probabilistic: Expectation

- The linearity of expectation property: the expectation of the sum of two random variables is the sum of their expectations:

$$\mathrm{E}[\mathrm{X}+\mathrm{Y}] = E[X] + E[Y]$$

- If X is any random variable, any function g(x) defines a new random variable g(X). If the expectation of g(X) is defined, then

E[g(X)] = $$\mathrm{E}[\mathrm{g}(\mathrm{X})] = \sum_{x} g(x) \Pr\{X = x\}$$

Basic of Probabilistic: Expectation

- Let $g(x) = ax$, we have for any constant a,

  $$E[aX] = aE[X]$$

- When two random variables X,Y are independent and each has a defined expectation,

  $$E[XY] = E[X]\,E[Y]$$

- The **variance**, expressing how far from the mean, of a random variable X with mean E[X] is

  $$Var[X] = E[X^2] - E^2[X]$$

  $$Var[X+Y] = Var[X]+Var[Y] \quad \text{if X,Y are independent}$$

  The **standard deviation** of a random variable X is the nonnegative square root of the variance of X.

Basic of Probabilistic: Expectation

- For example, if we have random variables X,Y for which $Pr\{X=1/4\}=Pr\{X=3/4\}= ½$ and $Pr\{Y=0\}=Pr\{Y=1\} = ½$.

X	1/4	3/4
Pr(X)	1/2	1/2

- Then $E[X] = ¼.1/2 + ¾.1/2 = ½$

  $E[Y] = 0.1/2 + 1.1/2 = ½$.
- However, the actual values taken on by Y are farther from the mean than the actual values taken on by X.
- Compute its variance :

  $E[X^2] = (1/4)^2.1/2 + (3/4)^2.1/2 = 5/16$

  $Var[X] = 5/16 - (1/2)^2 = 1/16$

  $E[Y^2] = 0.1/4 + 1^2.1/2 = 1/2$

  $Var[Y] = 1/2 - (1/2)^2 = 1/4$

Y	0	1
Pr(Y)	1/2	1/2

Basic of Probabilistic: geometric dist.

- A **Bernoulli trial** is defined as an experiment with only two possible outcomes: **success**, which occurs with probability p, and **failure**, which occurs with probability q=1-p.
- Suppose we have a sequence of Bernoulli trials. **How many trails occur before we obtain a success?**
- Let the random variable X be the number of trails needed to obtain a success. Then

$$\Pr\{X=k\}=q^{k-1}\,p\ ,$$

since we have k-1 failures before success

- This probability distribution is called the **geometric distribution**.

Basic of Probabilistic: geometric dist.

- Assuming that $q < 1$, the expectation of a geometric distribution is:

$$\mathrm{E[X]} = \sum_{k=1}^{\infty} kq^{k-1}p$$
$$= \frac{p}{q}\sum_{k=1}^{\infty} kq^{k}$$
$$= \frac{p}{q}\cdot\frac{q}{(1-q)^2}$$
$$= 1/p$$

- Thus on average, it takes 1/p trails before we obtain a success.
- The variance is $\mathrm{Var[X]} = q/p^2$

Basic of Probabilistic: geometric dist.

- For example, suppose we repeatedly roll two dice until we obtain either 7 or 11.
- There are 6 possible outcomes yielding 7 and 2 possible outcomes yielding 11.
- Thus, the probability of success is $p = 8/36 = 2/9$.
- We must roll $1/p = 9/2 = 4.5$ times on average to obtain 7 or 11.

Basic of Probabilistic: binomial dist.

- Suppose we have a sequence of Bernoulli trials. **How many successes occur during n Bernoulli trails** where a success occurs with probability p and a failure with probability q = 1-p**?**
- Let the random variable X be the number of successes in n trails. Then

$$\Pr\{X = k\} = \binom{n}{k} p^k q^{n-k}$$

  since there are $\binom{n}{k}$ ways to pick which k of the n trails are successes, and the probability that each occurs is $p^k q^{n-k}$.
- This probability distribution is called the **binomial distribution**.

Basic of Probabilistic: binomial dist.

- For convenience, the family of binomial distributions use the notation:

$$\mathrm{b(k;n,p)} = \binom{n}{k} p^k (1-p)^{n-k}$$

- The expectation of a random variable having a binomial distribution is : E[X] = np.
- Hence its variance is :

Var[X] = npq.