Basic Probability Definitions

COMPCSI 753: Algorithms for Massive Data

Ninh Pham

University of Auckland

Parts of this material are modifications of the lecture slides of Kevin Wayne

(https://www.cs.princeton.edu/~wayne/kleinberg-tardos/)

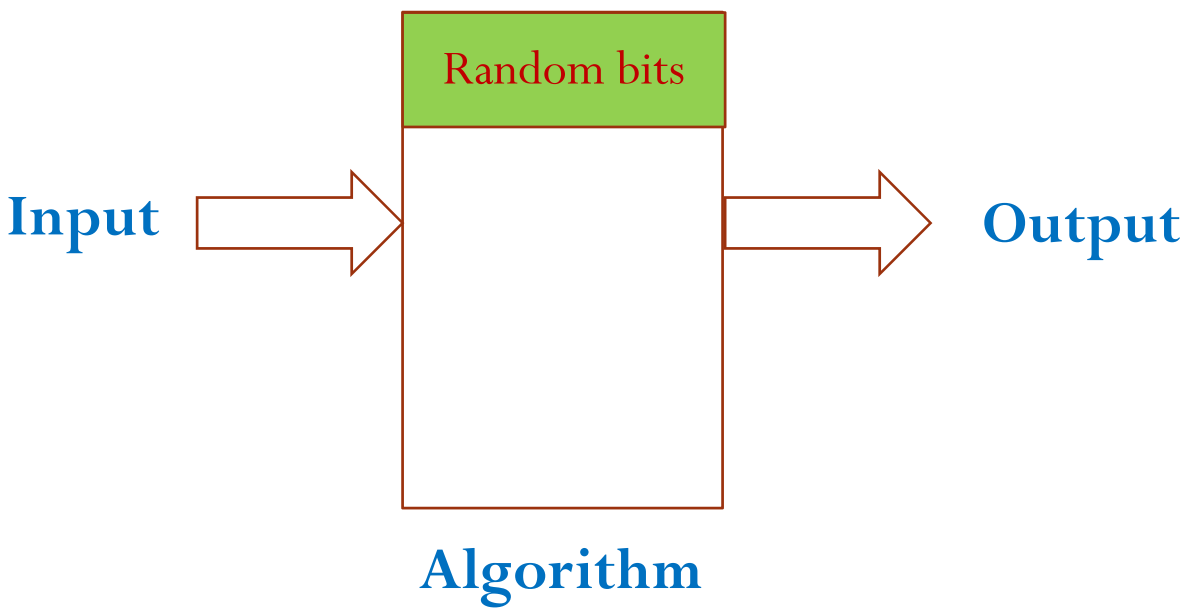
Designed for the textbook **Algorithm Design**

by Jon Kleinberg and Eva Tardos.

Auckland, Aug 3, 2020

Randomized algorithms

- The output and running time of algorithm are functions of both inputs and random bits chosen.
- Example of random bits: flipping a coin, i.e. Bernoulli(0.5), random integer, i.e. randint(n)…



Outline

- Basic probability definitions

- Warm up exercises

Finite probability space

- Simple probability statements:
  - S1: "If a fair coin is flipped, the probability of heads is **1/2**."
  - S2: "If a fair dice is rolled, the probability of **6** is **1/6**."
- Finite probability space: all possible outcomes
  - S1: $\boldsymbol{\Omega} = \{$**'heads', 'tails'**$\}$
  - S2: $\boldsymbol{\Omega} = \{\mathbf{1, 2, 3, 4, 5, 6}\}$
- Compute the probability of a particular outcome:
  - **Pr [coin = 'heads'] = 1/2**
  - **Pr [dice = 6] = 1/6**

Event definition

- Event:
  - A specific event **E** is defined by a subset of outcomes from the finite probability space.
  - We are interested in **Pr [E]**, i.e. the probability that the event **E** occurs.
  - Complement event: $\overline{E}$ - when the event **E** does not occur, i.e. **Pr** $[\overline{E}]$ **= 1 – Pr [E]**

- Example:
  - Rolling two fair dices, what is the probability that the event **2** dices are the same occurs?
  - $\Omega = \{(1,1), \ldots, (1,6), \ldots, (6,1), \ldots, (6,6)\}$
  - **Pr** [2 dices are the same] = **Pr** [(1,1)] + **Pr** [(2,2)] + … + **Pr** [(6,6)]
  - **Pr** [2 dices are different] = **1 – Pr** [2 dices are the same]

Independent events

- Two events **A** and **B** are independent if their outcomes do not affect each other.
  - **A**: The event that flipping the 1st fair coin returns 'heads'.
  - **B**: The event that flipping the 2nd fair coin returns 'heads'.

- **Pr [A ∧ B] = Pr [A] . Pr [B].**
  - **Pr** [1st coin = 'heads' ∧ 2nd coin = 'heads'] =**1/2 . 1/2 = 1/4**

- If $\mathbf{E_1}, \ldots, \mathbf{E_n}$ are independent events, then

$$\mathbf{Pr}\ [E_1 \wedge \ldots \wedge E_n] = \mathbf{Pr}\ [E_1] \ldots \mathbf{Pr}\ [E_n]$$

Monte Carlo algorithm

- Problem:
  - Given an array of **n** elements **A[1, … , n]** where **n** is even. Half of elements are **0s**, the other half are **1s**.
- Goal:
  - Find an index containing **1s**.

- Monte Carlo solutions:

```
1) Repeat 100 times:
    1.1) k = randint(n)
    1.2) If A[k] == 1
            Return k
2) Return "Failed"
```

- Probability of failure in the 1st times:
- Probability of failure:

Monte Carlo algorithm

- Problem:
  - Given an array of **n** elements **A[1, … , n]** where **n** is even. Half of elements are **0s**, the other half are **1s**.
- Goal:
  - Find an index containing **1s**.

- Monte Carlo solutions:

```
1) Repeat 100 times:
    1.1) k = randint(n)
    1.2) If A[k] == 1
             Return k
2) Return "Failed"
```

- Probability of failure in the 1st times: ½
- Probability of failure: $\mathbf{(\frac{1}{2})^{100}}$

Random variables

- Informally, **X** is a random variable if
  - **X** is an outcome of a random process.
  - **X** presents a value which we are uncertain.

- Examples:
  - Random process: rolling a fair dice, define **X** as the value of the dice.
  - **X = 1/2/3/4/5/6** with the same probability **1/6** (i.e. randint(6)).

  - Random process: flipping a fair coin, if 'heads', we return **1**; otherwise, we return **0**. Define **X** as the random variable of this process outcome.
  - **X = 0/1** with the same probability **1/2** (i.e. Bernoulli(0.5)).

Expectation of a random variable

- Expectation: Given a discrete random variable **X** with value in {**0, 1, … ,** ∞}, its expectation **E [X]** (e.g. the average value of **X**) is define by:

$$\mathrm{E}[X] = \sum_{j=0}^{\infty} j \ \mathrm{Pr}[X = j]$$

- Example:
  - Rolling a fair dice, expectation of the dice value is:
    1*1/6 + 2*1/6 + 3*1/6 + 4*1/6 + 5*1/6 + 6*1/6 = 7/2.
  - Given a random variable **X** generated by Bernoulli(0.5):
    - **Pr [X = 1] = 0.5** and **Pr [X = 0] = 0.5**.
    - **E [X] = 1 * 0.5 + 0 * 0.5 = 0.5**.

Roulette game in casino

- Game rule:
  - If you bet on Reds and the ball lands on red, you double your bet.
  - Otherwise, you lose your bet.
- Problem:
  - P1: Given that you bet **$1B** on Reds, what is the expectation of your outcome?
  - P2: Which game should we play?

Expectation properties

- If **X** is a **0/1** random variable, then **E [X] = Pr [X = 1]**
  - **E [X] = 1 * Pr [X = 1] + 0 * Pr [X = 0] = Pr [X = 1]**
- Given two random variables **X** and **Y** defined over the same finite probability space, we have **E [X + Y] = E [X] + E [Y]**
- Given **n** random variables $\mathbf{X_1}, \ldots, \mathbf{X_n}$ defined over the same finite probability space, we have:

$$\mathbf{E\,[X_1 + \ldots + X_n] = E\,[X_1] + \ldots + E\,[X_n]}$$

Guessing cards

- Game: Shuffle **n** different cards and turn them over one at a time. Try to guess each card without remembering the previous cards (sampling with replacement).

- Question: What is the expected number of correct guesses if you guess **n** times?

- Proof:
  - Define $\mathbf{X_i = 1}$ if we guess correctly at the ith time and **0** otherwise.
  - We have: $\mathbf{E\,[X_i] = 1/n}$.
  - Define $\mathbf{X = X_1 + \ldots + X_n}$ as the number of correct guesses.
  - $\mathbf{E\,[X] = E\,[X_1 + \ldots + X_n] = E\,[X_1] + \ldots + E\,[X_n] = n * 1/n = 1}$

Outline

- Basic probability definitions
- Warm up exercises

Guessing cards

- Game: Shuffle **n** different cards and take one of them out. Try to guess the card with memory of the previous cards (sampling without replacement).
- Question: What is the expected number of correct guesses if you guess **n** times?
- Proof:
  - Define $\mathbf{X_i = 1}$ if we guess correctly at the **i**th time and **0** otherwise.
  - We have: $\mathbf{E\,[X_i] = 1/(n - i + 1)}$.
  - Define $\mathbf{X = X_1 + \ldots + X_n}$ as the number of correct guesses.
  - $\mathbf{E\,[X] = E\,[X_1 + \ldots + X_n] = E\,[X_1] + \ldots + E\,[X_n]}$
  - $\mathbf{E\,[X] = 1/n + 1/(n\text{-}1) + \ldots + 1/1 = H(n) = O(\log n)}$.

Harmonic series ≈ ln(n) + 0.577

Las Vegas algorithm

- Problem:
  - Given an array of **n** elements **A[1, … , n]** where **n** is even. Half of elements are **0s**, the other half are **1s**.
- Goal:
  - Find an index containing **1s**.
- Las Vegas solutions:

```
Repeat:
    1) k = randint(n)
    2) If A[k] == 1
          Return k
```

- Exercise: Expected running time is **O(1) (2 iterations)**
  - Define **X** as number of iterations.
  - $\mathbf{E[X] = 1 * 1/2 + 2 * (1/2)^2 + \ldots + n * (1/2)^n \rightarrow 2 \text{ when } n \rightarrow \infty}$

Waiting for a first success

- Game: Flipping a biased coin where it is heads with probability **p** and tails with probability **1 - p**.

- Question: How many independent flips **X** until first heads?

- Proof:

$$E[X] = \sum_{j=0}^{\infty} j \cdot \Pr[X = j] = \sum_{j=0}^{\infty} j\,\underbrace{(1-p)^{j-1}}_{\textbf{j-1}\text{ tails}}\,\underbrace{p}_{\textbf{1}\text{ head}} = \frac{p}{1-p}\sum_{j=0}^{\infty} j\,(1-p)^{j} = \frac{p}{1-p}\cdot\frac{1-p}{p^2} = \frac{1}{p}$$

Hard trick:
Taylor Series of **f(x) = 1/x²**
where **x = p** and **a = 1**.

Coupon collector

- Game: Each box of cereal contains a coupon. There are **n** different types of coupons. Assuming all boxes are equally likely to contain each coupon, how many boxes before you have ≥ **1** coupon of each type?

- Proof:
  - Phase **j** = time between **j** and **j + 1** distinct coupons.
  - Let $\mathbf{X_j}$ = number of boxes you open in phase **j**.
  - Let $\mathbf{X = X_0 + X_1 + \ldots + X_{n-1}}$ = number of boxes in total.

$$E[X] = \sum_{j=0}^{n-1} E[X_j] = \sum_{j=0}^{n-1} \frac{n}{n-j} = n\sum_{i=1}^{n} \frac{1}{i} = nH(n)$$

prob. of success = **(n-j)/n**
→ expected waiting time = **n/(n-j)**