Random, Average, Probability

Some algorithms have terrible worst-case times, but:

- on random input, they are fast on average
- or, they contain randomizing steps, and are fast on average

We will study their “average case” times.

(Harder, rarely done in this course: what is the probability that it is fast?)

First, we review probability theory.

Example of Probability

Setting: toss a coin twice.

The set of all outcomes: \(\{HH, HT, TH, TT\}\)

Assumption: those 4 outcomes have equal probability.

Equivalently:

- the coin is unbiased
- no relation between two tosses

We want to say, for example:

- Probability of \(\{HH\}\) is \(\frac{1}{4}\).
- Probability of first toss being \(H\), \(\{HH, HT\}\), is \(\frac{1}{2}\).

We want a probability for each subset of the outcomes.

Terminology and Definitions

Sample space refers to the set of all outcomes.

Example: The sample space was \(\{HH, HT, TH, TT\}\).

An event refers to a subset of the sample space.

Example: \(\{HH\}\) was an event, \(\{HH, HT\}\) was another event.

To each subset \(A\) of the sample space, assign a probability \(\Pr(A)\).

Example: We assigned \(\Pr(\{HH\}) = \frac{1}{4}\).

\(\Pr\) must satisfy these:

- \(0 \leq \Pr(A) \leq 1\)
- \(\Pr(\emptyset) = 0\), \(\Pr(\text{sample space}) = 1\)
- if \(A \cap B = \emptyset\), then \(\Pr(A \cup B) = \Pr(A) + \Pr(B)\)

Example: \(\Pr(\{HH, HT\}) = \Pr(\{HH\}) + \Pr(\{HT\}) = \frac{1}{4} + \frac{1}{4}\)

Some Theorems

Notation: \(\bar{A}\) stands for \((\text{sample space}) \setminus A\).

\(\Pr(\bar{A}) + \Pr(A) = 1\)

\(\Pr(A) + \Pr(B) = \Pr(A \cap B) + \Pr(A) + \Pr(B)\)

Exercise

Setting: toss a coin three times.

Sample space: \(\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}\)

Assumption: those 8 outcomes have equal probability.

\(\Pr(\{HHH\}) = \frac{1}{8}\)

\(\Pr(2\text{nd toss is }H) = \Pr(\{HHH, HHT, THH, TTH\}) = \frac{4}{8}\)

\(\Pr(\text{number of }H\text{'s is }2) = \Pr(\{HHT, HTT, TTH\}) = \frac{3}{8}\)

Independence: Motivation

Let \(A = \{HHH, HHT, HTH, HTT\}\) “1st toss is \(H\).

Let \(B = \{HHH, HHT, THH, THT\}\) “2nd toss is \(H\).

\(\Pr(A) = \frac{4}{8}\)

\(\Pr(B) = \frac{4}{8}\)

\(\Pr(1\text{st toss is }H \text{ and } 2\text{nd toss is }H) = \Pr(A \cap B)\)

\[\begin{align*}
\frac{4}{8} & = \Pr(\{HHH, HHT\}) \\
& = 2/8 \\
& = (4/8) \cdot (4/8)
\end{align*}\]

This seems to say: \(\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)\)

It seems to say: Those two events are unrelated, so to compute the probability that both happens, just multiply their respective probabilities.
Independence: Definition

Two events $A, B$ are independent iff

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B).$$

If we talk about more events, it gets more complicated.

Events $A_1, A_2, A_3$ are mutually independent iff

- $\Pr(A_1 \cap A_2) = \Pr(A_1) \cdot \Pr(A_2)$
- $\Pr(A_2 \cap A_3) = \Pr(A_2) \cdot \Pr(A_3)$
- $\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1) \cdot \Pr(A_2) \cdot \Pr(A_3)$

Independence is usually an assumption, e.g., we assume that a bunch of coin tosses are mutually independent.

Random Variable: Motivation

We are most interested in events of this kind:

- number of $H$'s is 2
- number of $H$'s is at least 2
- number of $H$'s is at most 2

Define $X: \{HHH, \ldots, TTT\} \to \mathbb{R}$, $X(o)$ is the number of $H$'s in outcome $o$.

Example: $X(HHT) = 2$.

$$\Pr(\text{number of } H \text{'s is 2}) = \Pr(o \mid X(o) = 2)$$

$$= \Pr(HHT, HTH, THH)$$

$$= \frac{3}{8}$$

Random Variable: Definition

A random variable is a function from the sample space to $\mathbb{R}$.

The notation $\Pr(X = k)$ means $\Pr(o \mid X(o) = k)$.

Similarly $\Pr(X \leq k)$, $\Pr(X < k)$, etc.

Example: in the previous slide, $\Pr(X = 2) = \frac{3}{8}$.

Random Variable: Convention

They call it random variable because:

- "number of $H$'s" sounds like a variable
- its value is random

It's sloppy wording, probably sloppy thinking too, but it sticks.

Following that convention, we will speak like "let the random variable $X$ be the number of $H$'s" instead of "let $X(o)$ be the number of $H$'s in outcome $o$".

We also have expressions in random variables like $2 \cdot X$ and $X + Y$.

Technically they mean $2 \cdot X(o)$ and $X(o) + Y(o)$.

Example: Roll a dice twice. Let $X$ be the points of the 1st roll, $Y$ be the points of the 2nd roll. $X + Y$ is the total.

Exercise

Setting: toss a coin many times, until getting an $H$, then stop.

Sample space: $\{H, TH, TTH, TTT, \ldots\}$

Assumption: fair coin; the tosses are independent.

Let the random variable $X$ be the number of tosses until getting an $H$.

$$\Pr(X = 1) = \Pr(1 \text{st toss is } H) = \frac{1}{2}$$

$$\Pr(X = 2) = \Pr(1 \text{st toss is } T) \cdot \Pr(2 \text{nd toss is } H) = (1/2) \cdot (1/2)$$

$$\Pr(X = 3) = \Pr(1 \text{st toss is } T) \cdot \Pr(2 \text{nd toss is } T) \cdot \Pr(3 \text{rd toss is } H)$$

$$\Pr(X = k) = (1/2)^k$$

Expected Value: Motivation

What is the "average number of tosses" until getting an $H$?

- The number could be 1, but it happens only half of the time.
- The number could be 2, but it happens only a quarter of the time.
- The number could be 3, but it happens only one-eighth of the time.

Idea: Imagine doing this experiment many times, find the arithmetic mean.

It suggests:

$$\text{average } = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + \cdots$$

$$= \sum_{k=1}^{\infty} k \cdot \Pr(X = k)$$
Expected Value: Definition

The expected value of random variable $X$ is:

$$E(X) = \sum_k k \cdot \Pr(X = k)$$

($k$ ranges over all possible values of $X$)

Average means expected value. E.g., Average time means expected value of running time.

Theorems:

$$E(\text{constant} \cdot X) = \text{constant} \cdot E(X)$$

$$E(X + Y) = E(X) + E(Y)$$

Distribution

Consider $f(k) = \Pr(X = k)$. This is a function.

The distribution of $X$ refers to that function.

We also say: $X$ follows that distribution.

If you know the distribution, you know how to calculate various things, since it tells you about $\Pr(X = k)$.

There are several very common distributions you should know about.

Bernoulli Distribution

This models one single success-or-failure, with probability of success $p$.

$X$ follows a Bernoulli distribution iff:

$$\Pr(X = 0) = 1 - p$$

$$\Pr(X = 1) = p$$

In other words, $f(0) = 1 - p$, $f(1) = p$

E.g., tossing a fair coin: $p = 1/2$

E.g., tossing a biased coin: $p = 1/3$ for example.

$$E(X) = p$$

Binomial Distribution

This models, e.g., tossing a coin $n$ times, the tosses are mutually independent, each toss has probability $p$ of head. Now count the number of heads.

$X$ follows a binomial distribution iff:

$$\Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$k$ ranges from 0 to $n$ inclusive.

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1 - p)^{n-k} = np$$

Geometric Distribution

This models, e.g., tossing a coin until a head, the tosses are mutually independent, each toss has probability of $p$ head. Now count how many tosses.

$X$ follows a geometric distribution iff:

$$\Pr(X = k) = (1 - p)^{k-1} p$$

$k \geq 1$

$$E(X) = \sum_{k=1}^{\infty} k (1 - p)^{k-1} p$$

$$= \frac{1}{(1 - (1 - p))^2} p$$

$$= \frac{1}{p}$$

Randomized Quick-Sort

To sort array $A$: if $|A| \geq 2$:

1. randomly pick an element $p$ “pivot”
2. partition $A$ into:
   - $L$: array of elements $< p$
   - $E$: array of elements $= p$
   - $G$: array of elements $> p$

($O(|A|)$ time)

3. recurse: sort $L$
4. recurse: sort $G$
5. concatenate $L$ sorted, $E$, $G$ sorted

I write it like $L$, $E$, $G$ are new arrays for simplicity.

In reality, “partition” is done in-place in $A$, so $L$ occupies the left side, $E$ is in the middle, $G$ occupies the right side. Nothing to do for “concatenate”.
Quick-Sort Worse-Case Time
(Assume distinct elements in A.)
If we always happen to pick the max for the pivot:

|L| = |A| − 1
|E| = 1
|G| = 0

1st call takes n steps partitioning
2nd call takes n − 1 steps partitioning
... all the way down to 2.
This is \(\Theta(n^2)\) time.

Quick-Sort Average-Case Time
Draw the call-history as a binary tree.
Each level takes \(O(n)\) time for partitioning.
What is the expected value of the number of levels?
What is the expected value of path length from root to a bottom node?

Quick-Sort Average-Case Time
Define “good invocation”: a call during which

|L| \(\leq 3|A|/4\) or
|G| \(\leq 3|A|/4\).

Among all \(|A|\) possible pivots, half of them makes this happen.
Probability of a call being a good invocation is 1/2, like a fair coin.
A good invocation also ensures that each child call deals with at most \(3|A|/4\) elements.

On a path from root to bottom, if there are \(\log_{4/3}(n)\) good invocations, the path is done, no more elements to sort.
A path makes calls until \(\log_{4/3}(n)\) good invocations at most, then must end.
What does this tell us about expected path length?

Quick-Sort Average-Case Time
Toss a fair coin until \(k\) heads. Expected number of tosses is \(2k\) (2 for each head) (recall \(E(X + Y) = E(X) + E(Y)\)).
Follow a path until \(k = \log_{4/3}(n)\) good invocations. Expected number of nodes is \(2 \cdot \log_{4/3}(n), O(\log(n))\).
Expected number of levels is \(O(\log(n))\), each level takes \(O(n)\) time for partitioning.
Expected running time is \(O(n \log(n))\). This is a good speed.