A Crash Course on Probabilistic Analysis and Randomized Algorithms

Tingjian Ge

These slides are from/based on Prof. Eli Upfal's course on Probability and Computing.

"It is remarkable that this science, which originated in the consideration of games and chances, should have become the most important object of human knowledge... The most important questions of life are, for the most part, really only problems of probability"

Pierre Simons, Marquis de Laplace

(1749–1827).

Why Probability in Computing?

- Almost any advance computer application today has some randomization/statistical components:
  - Network security
  - Cryptography
  - Web search and Web advertising
  - Spam filtering
  - Recommendation systems Amazon, Netfix,..
  - Machine learning
  - Communication protocols
  - Computational finance
  - System biology
  - DNA sequencing

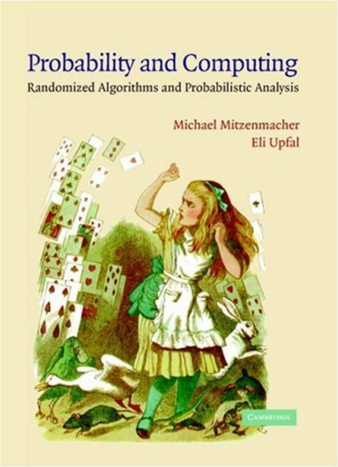
Probability and Computing

- Randomized algorithms - random steps help!
- Probabilistic analysis of algorithms - average case, almost always case, worst case.
- Statistical inference - Machine learning, data mining...

Course Outline

- Basic (discrete) probability theory, moments, tail bounds.
- Randomized algorithms, probabilistic analysis, average and almost sure performance.
- Applications: sorting, selection, routing, graph algorithms,...
- Random walks - Markov chains.
- The Monte-Carlo method.
- Continuous random variables, uniform, exponential, Poisson process.
- Point and interval estimate, hypothesis testing.
-

Textbook



Probability and Computing
Randomized Algorithms and Probabilistic Analysis
Michael Mitzenmacher
Eli Upfal

Mitzenmacher and Upfal:
Probability and Computing: Randomized Algorithms and Probabilistic Analysis. Cambridge Press 2005.

Verifying Polynomial Identities

Problem: Verify that $P(x) \equiv Q(x)$.
Example: Check if

$$(x+1)(x-2)(x+3)(x-4)(x+5)(x-6) \equiv x^6 - 7x^3 + 25.$$

We use $\equiv$ for polynomial identities, $=$ for numerical equality.
$P(x) \equiv Q(x)$ iff for any value $r \in R$, $P(r) = Q(r)$.

Deterministic solution:

$$\begin{aligned} H(x) &\equiv (x+1)(x-2)(x+3)(x-4)(x+5)(x-6) \\ G(x) &\equiv x^6 - 7x^3 + 25. \end{aligned}$$

Transform $H(x)$ to a "canonical" form

$$H(x) \equiv \sum_{i=0}^{6} c_i x^i$$

$H(x) \equiv G(x)$ iff the coefficients of all monomials are equal.

A Randomized Solution:

1. Choose a random integer r in the range $[1, ..., 600]$.
2. Compute $H(r)$ and $G(r)$.
3. If $H(r) = G(r)$ then output CORRECT else output FALSE.

Randomized Algorithm

We augment the standard RAM operations with a new operation:
Choose a random number uniformly from the set $\{a_1, a_2, ..., a_k\}$.
We assume that this operation takes 1 step.

The output (might) depend on the choice of r, thus it is a **random variable**.
What is the “chance” that the algorithm actually gives the correct answer???

$$\begin{aligned} H(x) &\equiv (x+1)(x-2)(x+3)(x-4)(x+5)(x-6) \\ G(x) &\equiv x^6 - 7x^3 + 25. \end{aligned}$$

Assume $r = 2$
$H(2) = 3 \times 0 \times 5 \times -2 \times 7 \times -4 = 0$.
$G(2) = 2^6 - 72^3 + 25 = 64 - 56 + 25 = 33.$
Since $H(2) \neq G(2)$ we proved that $H(x) \not\equiv G(x)$.

What happens if we have equality?

Example 1: Check if $(x+1)(x-1) \equiv x^2 - 1$.
Since the two sides of the equation are identical - any number that we try would give equality.

Is this algorithm **always** correct?

Example 2: Check if $x^2 + 7x + 1 = (x+2)^2$.
If we try $r = 2$ we get

$$LHS = 4 + 14 + 1 = 19, \qquad RHS = 4^2 = 16$$

showing that the two sides are not identical.

But for $r = 1$ we get equality:

$$1 + 7 + 1 = (1+2)^2 = 9.$$

A bad choice of r may lead to a wrong answer!

How many of the possible choices or r are **bad**?

Some Algebra

Assume that $G(x) \not\equiv H(x)$, and that the sum of the degrees of x in H and G is bounded by d.
$F(x) \equiv G(x) - H(x)$ is a polynomial in one variable of degree bounded by d.

Theorem

If

$$F(x) = G(x) - H(x) \not\equiv 0$$

then the equation

$$F(x) = G(x) - H(x) = 0$$

has no more than d roots (solutions).

Analysis of the Algorithm

If the identity is correct, the algorithm always outputs a correct answer.

If the identity is NOT correct, the algorithm outputs the WRONG answer only if we randomly picked r which is a root of the polynomial $F(x) = G(x) - H(x) = 0$.

If we choose r in the range $[1, ..., 100d]$, the "chance" of returning a wrong answer is no more than 1%.

A randomized technique gives a significantly simpler algorithm - at a cost of a small probability of error.

Getting an arbitrary small error probability

We can reduce the “error probability” at the expense of increasing the run-time of the algorithm:

1. Run the algorithm 10 times.
2. Output “CORRECT” if got “CORRECT” in all the 10 runs.

If the new algorithm outputs “CORRECT” The “chance” that $G(x) \not\equiv H(x)$ is less than $10^{-20} < 2^{-64}$.

Probability Space

Definition

A probability space has three components:

1. A sample space Ω, which is the set of all possible outcomes of the random process modeled by the probability space;
2. A family of sets $\mathcal{F}$ representing the allowable events, where each set in $\mathcal{F}$ is a subset of the sample space Ω;
3. A probability function $\Pr : \mathcal{F} \to \mathbf{R}$, satisfying the definition below.

In a discrete probability space the we use $\mathcal{F} = 2^{\Omega}$.

Probability Function

Definition

A probability function is any function $\Pr : \mathcal{F} \rightarrow \mathbf{R}$ that satisfies the following conditions:

1. For any event E, $0 \leq \Pr(E) \leq 1$;
2. $\Pr(\Omega) = 1$;
3. For any finite or countably infinite sequence of pairwise mutually disjoint events $E_1, E_2, E_3, \ldots$

$$\Pr\left(\bigcup_{i \geq 1} E_i\right) = \sum_{i \geq 1} \Pr(E_i).$$

The probability of an event is the sum of the probabilities of its simple events.

Examples:

Consider the random process defined by the outcome of rolling a dice.

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

We assume that all “facets” have equal probability, thus

$$Pr(1) = Pr(2) = ...Pr(6) = 1/6.$$

The probability of the event “odd outcome”

$$= Pr(\{1, 3, 5\}) = 1/2$$

Assume that we roll two dice:
$\mathcal{S} =$ all ordered pairs $\{(i, j)$, $1 \leq i, j \leq 6\}$.
We assume that each (ordered) combination has probability $1/36$.
Probability of the event "sum $= 2$" $=$

$$Pr((1,1)) = 1/36.$$

Probability of the event "sum $= 3$"

$$Pr(\{(1,2),(2,1)\}) = 2/36.$$

Let $E_1 =$ “sum bounded by 6”,

$$E_1 \; = \; \{(1,1),(1,2),(1,3),(1,4),(1,5),(2,1),(2,2),$$

$$(2,3),(2,4),(3,1),(3,2),(3,3),(4,1),(4,2),(5,1)\}$$

$$Pr(E_1) = 15/36$$

Let $E_2 =$ “both dice have odd numbers”, $Pr(E_2) = 1/4$.

$$Pr(E_1 \cap E_2) =$$

$$Pr(\{(1,1),(1,3),(1,5),(3,1),(3,3),(5,1)\}) =$$

$$6/36 = 1/6.$$

Back to the Polynomial Identity Checking Algorithms

- A simple event = a choice of r.
- Sample Space = all integers in $[1, ..., 100d]$.
- We assume that all integers in the range have equal probability, thus the probability of a simple event r is $Pr(r) = \frac{1}{100d}$.
- The **“bad”** events: choosing a root of the polynomial. There are no more than d simple events in the bad event.
- $Pr(\text{“bad” event}) \leq \frac{d}{100d}$.

Assume that we repeat the algorithm k times:

- If any iteration returns FALSE output FALSE, else output CORRECT.
- A simple event = A sequence of k choices $r_1, ..., r_k$.
- The sample space = All sequences of r numbers in the range $[1, ..., 100d]$.
- The probability of a simple event = $(\frac{1}{100d})^k$.
- The **bad** event = all k choices are roots of the polynomial, there are no more than d^k such simple events.
- Probability of the bad event $\leq d^k(\frac{1}{100d})^k$.

Independent Events

Definition

Two events E and F are independent if and only if

$$\Pr(E \cap F) \;=\; \Pr(E) \cdot \Pr(F).$$

More generally, events $E_1, E_2, \ldots E_k$ are mutually independent if and only if for any subset $I \subseteq [1, k]$,

$$\Pr\left(\bigcap_{i \in I} E_i\right) \;=\; \prod_{i \in I} \Pr(E_i).$$

- The probability of picking a root in one round is $\leq \frac{d}{100d}$.
- The events in different round are independent
- The probability of picking roots in k successive independent rounds $\leq (\frac{d}{100d})^k$.

Conditional Probability

What is the probability that a student at UMass Lowell was born in MA.

E1 = the event "born in MA".

E2 = the event "a student at UML".

The conditional probability that a student at UML was

born in MA is denoted

Pr(E1 | E2).

Computing Conditional Probabilities

Definition

The conditional probability that event E occurs given that event F occurs is

$$\Pr(E \mid F) \;=\; \frac{\Pr(E \cap F)}{\Pr(F)}.$$

The conditional probability is only well-defined if $\Pr(F) > 0$.

By conditioning on F we restrict the sample space to the set F. Thus we are interested in $Pr(E \cap F)$ “normalized” by $Pr(F)$.

Example

What is the probability that in rolling two dice the sum is 8 given that the sum was even?
$E_1 =$ "sum is 8",

$$Pr(E_1) = Pr((2,6),(3,5),(4,4),(5,3),(6,2)) = 5/36$$

$E_2 =$ "sum even",
$Pr(E_2) = 1/2 = 18/36.$
$Pr(E_1 \mid E_2) = \frac{Pr(E_1 \cap E_2)}{Pr(E_2)} = \frac{5/36}{1/2} = 5/18.$

Example - a posteriori probability

We are given 2 coins. One is a fair coin A, the other coin, B, has head on both sides B.
We choose a coin at random, i.e. each coin is chosen with probability $1/2$.
Given that we got head, what is the probability that we chose the fair coin A???

Define a sample space of ordered pairs $(coin, outcome)$.
The sample space has three points

$$\{(A, h), (A, t), (B, h)\}$$

$$Pr((A, h)) = Pr((A, t)) = 1/4$$

$$Pr((B, h)) = 1/2$$

Define two events:
$E_1 =$ "Chose coin A".
$E_2 =$ "Outcome is head".

$$Pr(E_1 \mid E_2) = \frac{Pr(E_1 \cap E_2)}{Pr(E_2)} = \frac{1/4}{1/4 + 1/2} = 1/3.$$

Useful identities:

$$Pr(A \mid B) \;=\; \frac{Pr(A \cap B)}{Pr(B)}$$

$$Pr(A \cap B) = Pr(A \mid B)Pr(B)$$

$$Pr(A \cap B \cap C) = Pr(A \mid B \cap C)Pr(B \cap C)$$

$$= Pr(A \mid B \cap C)Pr(B \mid C)Pr(C)$$

Let $A_1, ..., A_n$ be a sequence of events. Let $E_i = \bigcap_{j=1}^{i} A_i$

$$Pr(E_n) = Pr(A_n \mid E_{n-1})Pr(E_{n-1}) =$$

$$Pr(A_n \mid E_{n-1})Pr(A_{n-1} \mid E_{n-2}) .... P(A_2 \mid E_1)Pr(A_1)$$

Independence

Two events A and B are independent if

$$Pr(A \cap B) = Pr(A) \times Pr(B),$$

or

$$Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = Pr(A).$$

Independent events do not have to be related to independent physical processes.
Example: the probability that the outcome of a dice roll is *even* $(= \frac{3}{6})$ is independent of the event "the outcome is ≤ 4"$(= \frac{4}{6})$.
The probability of "an even outcome ≤ 4" is

$$\frac{2}{6} = \frac{12}{36} = \frac{3}{6} \cdot \frac{4}{6}$$

The "intuition" here is that there are the same number of odd and even outcomes that are ≤ 4. Thus, the "information" that the outcome is ≤ 4 does not "help" in deciding if it is odd or even.

Example

A family has two children:
1. Given that the first child is a girl, what is the probability that the second child is a girl? - $\frac{1}{2}$.
2. Given that one of the children is a girl, what is the probability that the other child is a girl?

A - “at least one child is a girl”
B - “both children are girls”

$$Prob(A) = 1 - (\frac{1}{2})^2 = \frac{3}{4}$$

$$Prob(B) = \frac{1}{4}$$

$$Prob(B \mid A) = \frac{Pr(A \cap B)}{Pr(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Verifying Matrix Multiplication

Given three $n \times n$ matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in a Boolean field, we want to verify

$$\mathbf{AB} \quad = \quad \mathbf{C}.$$

Standard method: Matrix multiplication - takes $\Theta(n^3)$ ($\Theta(n^{2.37})$) operations.

Randomized algorithm:

1. Chooses a random vector $\bar{r} = (r_1, r_2, \ldots, r_n) \in \{0,1\}^n$.
2. Compute $\mathbf{B}\bar{r}$;
3. Compute $\mathbf{A}(\mathbf{B}\bar{r})$;
4. Computes $\mathbf{C}\bar{r}$;
5. If $\mathbf{A}(\mathbf{B}\bar{r}) \neq \mathbf{C}\bar{r}$ return $\mathbf{AB} \neq \mathbf{C}$, else return $\mathbf{AB} = \mathbf{C}$.

The algorithm takes $\Theta(n^2)$ time.

Theorem

If $\mathbf{AB} \neq \mathbf{C}$, and $\bar{r}$ is chosen uniformly at random from $\{0,1\}^n$, then

$$\Pr(\mathbf{AB}\bar{r} = \mathbf{C}\bar{r}) \leq \frac{1}{2}.$$

Lemma

Choosing $\bar{r} = (r_1, r_2, \ldots, r_n) \in \{0, 1\}^n$ *uniformly at random is equivalent to choosing each* r_i *independently and uniformly from* $\{0, 1\}$*.*

Proof.

If each r_i is chosen independently and uniformly at random, each of the 2^n possible vectors $\bar{r}$ is chosen with probability 2^{-n}, giving the lemma. □

Proof:

Let $\mathbf{D} = \mathbf{AB} - \mathbf{C} \neq 0$.
$\mathbf{AB}\bar{r} = \mathbf{C}\bar{r}$ implies that $\mathbf{D}\bar{r} = 0$.
Since $\mathbf{D} \neq 0$ it has some non-zero entry; assume d_{11}.
For $\mathbf{D}\bar{r} = 0$, it must be the case that

$$\sum_{j=1}^{n} d_{1j} r_j \quad = \quad 0,$$

or equivalently

$$r_1 \quad = \quad -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}. \tag{1}$$

Here we use $d_{11} \neq 0$.

Principle of Deferred Decision

Assume that we fixed $r_2, \ldots, r_n$.
The RHS is already determined, the only variable is r_1.

$$r_1 \quad = \quad -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}. \tag{2}$$

Probability that $r_1 =$ RHS is no more than $1/2$.

More formally, summing over all collections of values $(x_2, x_3, x_4, \ldots, x_n) \in \{0,1\}^{n-1}$, we have

$$
\begin{aligned}
& \Pr(\mathbf{AB}\bar{r} = \mathbf{C}\bar{r}) \\
= & \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr\left(\mathbf{AB}\bar{r} = \mathbf{C}\bar{r} \mid (r_2,\ldots,r_n) = (x_2,\ldots,x_n)\right) \\
& \qquad\qquad\qquad \cdot \Pr\left((r_2,\ldots,r_n) = (x_2,\ldots,x_n)\right) \\
= & \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr\left((\mathbf{AB}\bar{r} = \mathbf{C}\bar{r}) \cap ((r_2,\ldots,r_n) = (x_2,\ldots,x_n))\right) \\
\leq & \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr\Bigg(\Bigg(r_1 = -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}\Bigg) \cap ((r_2,\ldots,r_n) = (x_2,\ldots \\
= & \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr\Bigg(r_1 = -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}\Bigg) \cdot \Pr((r_2,\ldots,r_n) = (x_2,\ldots \\
\leq & \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \frac{1}{2} \Pr((r_2,\ldots,r_n) = (x_2,\ldots,x_n)) \\
= & \frac{1}{2}.
\end{aligned}
$$

Theorem (Law of Total Probability)

Let $E_1, E_2, \ldots, E_n$ be mutually disjoint events in the sample space Ω, and $\cup_{i=1}^n E_i = \Omega$, then

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B \mid E_i) \Pr(E_i).$$

Proof.

Since the events $B \cap E_i$, $i = 1, \ldots, n$ are disjoint and cover the entire sample space Ω,

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B \mid E_i) \Pr(E_i).$$

□

Bayes' Law

Theorem (Bayes' Law)

Assume that $E_1, E_2, \ldots, E_n$ *are mutually disjoint sets such that* $\cup_{i=1}^n E_i = E$, *then*

$$\Pr(E_j \mid B) = \frac{\Pr(E_j \cap B)}{\Pr(B)} = \frac{\Pr(B \mid E_j)\Pr(E_j)}{\sum_{i=1}^n \Pr(B \mid E_i)\Pr(E_i)}.$$

Application: Finding a Biased Coin

- We are given three coins, two of the coins are fair and the third coin is biased, landing heads with probability $2/3$. We need to identify the biased coin.
- We flip each of the coins. The first and second coins come up heads, and the third comes up tails.
- What is the probability that the first coin is the biased one?

Let E_i be the event that the i-th coin flipped is the biased one, and let B be the event that the three coin flips came up heads, heads, and tails.

Before we flip the coins we have $\Pr(E_i) = 1/3$ for $i = 1, \ldots, 3$, thus

$$\Pr(B \mid E_1) = \Pr(B \mid E_2) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{6},$$

and

$$\Pr(B \mid E_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12}.$$

Applying Bayes' law we have

$$\Pr(E_1' \mid B) = \frac{\Pr(B \mid E_1)\Pr(E_1)}{\sum_{i=1}^{3} \Pr(B \mid E_i)\Pr(E_i)} = \frac{2}{5}.$$

The outcome of the three coin flips increases the probability that the first coin is the biased one from $1/3$ to $2/5$.

Bayesian approach

A test shows that a machine is working correctly. The test has 10% error rate. What is the probability that the machine is functioning correctly.

A - test result positive.

B - machine is working correctly.

$$Pr(B\ |A) = \frac{Pr(B \cap A)}{Pr(A)} = \frac{Pr(B \cap A)}{Pr(A \mid B)Pr(B) + Pr(A \mid \bar{B})Pr(\bar{B})}$$

What is $Pr(B)$?

Bayesian approach

- Start with an *prior* model, giving some initial value to the model parameters.
- This model is then modified, by incorporating new observations, to obtain a *posterior* model that captures the new information.

Bayesian approach

A test shows that a machine is working correctly. The test has 10% error rate. What is the probability that the machine is functioning correctly.

A - test result positive.

B - machine is working correctly.

$$Pr(B \mid A) = \frac{Pr(B \cap A)}{Pr(A)} = \frac{Pr(B \cap A)}{Pr(A \mid B)Pr(B) + Pr(A \mid \bar{B})Pr(\bar{B})}$$

What is $Pr(B)$? Without any *prior* knowledge we set $Pr(B) = Pr(\bar{B}) = 1/2$.

$$Pr(B \mid A) = \frac{\frac{1}{2}\frac{9}{10}}{\frac{1}{2}\frac{9}{10} + \frac{1}{2}\frac{1}{10}} = \frac{9}{10}$$

This estimate is dominates by the reliability of the test. Can we do better?

Assume that we know that this machine fails only $1/5$ of the time. We set $Pr(B) = 4/5$.

$$Pr(B\ |A) = \frac{Pr(B \cap A)}{Pr(A)} = \frac{Pr(B \cap A)}{Pr(A \mid B)Pr(B) + Pr(A \mid \bar{B})Pr(\bar{B})}$$

$$= \frac{\frac{4}{5}\frac{9}{10}}{\frac{4}{5}\frac{9}{10} + \frac{1}{5}\frac{1}{10}} = \frac{36}{37} \approx 0.97\%$$

Example: randomized matrix multiplication test

- We want to evaluate the increase in confidence through repeated tests.
- If we have no information about the process that generated the identity, a reasonable prior assumption is that the identity is correct with probability $1/2$.
- If we run the randomized test once and it returns that the matrix identity is correct, how does it change our confidence in the identity?

E - the identity is correct
B the test returns that the identity is correct.
We start with $\Pr(E) = \Pr(\bar{E}) = 1/2$, and since the test has a one side error bounded by $1/2$, we have $\Pr(B \mid E) = 1$, and $\Pr(B \mid \bar{E}) \leq 1/2$.
Applying Bayes' law we have

$$\begin{aligned} \Pr(E' \mid B) &= \frac{\Pr(B \mid E)\Pr(E)}{\Pr(B \mid E)\Pr(E) + \Pr(B \mid \bar{E})\Pr(\bar{E})} \\ &\geq \frac{1/2}{1/2 + 1/2 \cdot 1/2} = 2/3. \end{aligned}$$

- Assume now that we run the randomized test again and it again returns that the identity is correct.
- After the first test, the prior model was revised, so $\Pr(E) \geq 2/3$, and $\Pr(\bar{E}) \leq 1/3$.
- $\Pr(B \mid E) = 1$ and $\Pr(B \mid \bar{E}) \leq 1/2$.

Applying Bayes' law we have

$$\Pr(E' \mid B) \geq \frac{2/3}{2/3 + 1/3 \cdot 1/2} = 4/5.$$

In general, if before running the test our prior model is that $\Pr(E) \geq 2^i/(2^i+1)$, and the test returns that the identity is correct (event B), then

$$\Pr(E' \mid B) \geq \frac{\frac{2^i}{2^i+1}}{\frac{2^i}{2^i+1} + \frac{1}{2}\frac{1}{2^i+1}} = \frac{2^{i+1}}{2^{i+1}+1} = 1 - \frac{1}{2^i+1}.$$

Thus, if all 100 calls to the matrix identity test return that the identity is correct, then our confidence in the correctness of this identity is at least $1 - \frac{1}{2^{100}+1}$.

Counterintuitive?

- $1/1000$ of tourists who visit tropical country X return with a dangerous virus Y.
- There is a test to check for the virus. The test has 5% *false positive* rate and no *false negative* error.
- You returned from country X, took the test, and it was positive. Should you take the painful treatment for the virus?
- A - has the virus. B - positive in the test.

$$Pr(A \mid B) = \frac{\frac{1}{1000}}{\frac{1}{1000} + \frac{999}{1000}\frac{5}{100}} = \frac{20}{1019} \approx 2\%$$

Explanation: Out of 1000 tourist, 1 will have the virus and another 50 will be false positive in the test.

Min-Cut Algorithm

Input: An n-node graph G.

Output: A minimal set of edges that disconnects the graph.

1. **Repeat $n-2$ times:**
    1. Pick an edge uniformly at random.
    2. Contract the two vertices connected by that edge, eliminate all edges connecting the two vertices.
2. Output the set of edges connecting the two remaining vertices.

Theorem

The algorithm outputs a min-cut set of edges with probability $\geq \frac{1}{n(n-1)}$.

Lemma

Vertex contraction does not reduce the size of the min-cut set. (Contraction can only increase the size of the min-cut set.)

Proof.

Every cut set in the new graph is a cut set in the original graph. □

Analysis of the Algorithm

Assume that the graph has a min-cut set of k edges.
We compute the probability of finding one such set C.

Lemma

If the edge contracted does not belong to C, no other edge eliminated in that step belongs to C.

Proof.

A contraction eliminates a set of parallel edges (edges connecting one pair of vertices).
Parallel edges either all belong, or don't belong to C. □

Let $E_i =$ "the edge contracted in iteration i is not in C."
Let $F_i = \cap_{j=1}^{i} E_j =$ "no edge of C was contracted in the first i iterations".
We need to compute $Pr(F_{n-2})$

Since the minimum cut-set has k edges, all vertices have degree $\geq k$, and the graph has $\geq nk/2$ edges.
There are at least $nk/2$ edges in the graph, k edges are in C.
$Pr(E_1) = Pr(F_1) \geq 1 - \frac{2k}{nk} = 1 - \frac{2}{n}$.

Assume that the first contraction did not eliminate an edge of C (conditioning on the event $E_1 = F_1$).

After the first vertex contraction we are left with an $n-1$ node graph, with minimum cut set, and minimum degree $\geq k$.

The new graph has at least $k(n-1)/2$ edges.

$Pr(E_2 \mid F_1) \geq 1 - \frac{k}{k(n-1)/2} \geq 1 - \frac{2}{n-1}$.

Similarly,

$Pr(E_i \mid F_{i-1}) \geq 1 - \frac{k}{k(n-i+1)/2} = 1 - \frac{2}{n-i+1}$.

We need to compute

$$Pr(F_{n-2})$$

We use

$$Pr(A \cap B) = Pr(A \mid B)Pr(B)$$

$$Pr(F_{n-2}) \ =$$

$$Pr(E_{n-2} \cap F_{n-3}) = Pr(E_{n-2} \mid F_{n-3})Pr(F_{n-3}) =$$

$$Pr(E_{n-2} \mid F_{n-3})Pr(E_{n-3} \mid F_{n-4}) \ldots . Pr(E_2 \mid F_1)Pr(F_1) \geq$$

$$\geq \Pi_{i=1}^{n-2}(1 - \frac{2}{n-i+1}) = \Pi_{i=1}^{n-2}(\frac{n-i-1}{n-i+1})$$

$$= (\frac{n-2}{n})(\frac{n-3}{n-1})(\frac{n-4}{n-2}) \ldots (\frac{4}{6})(\frac{3}{5})(\frac{2}{4})(\frac{1}{3}) = \frac{2}{n(n-1)}.$$

Useful identities:

$$Pr(A \mid B) \; = \; \frac{Pr(A \cap B)}{Pr(B)}$$

$$Pr(A \cap B) = Pr(A \mid B)Pr(B)$$

$$Pr(A \cap B \cap C) = Pr(A \mid B \cap C)Pr(B \cap C)$$

$$= Pr(A \mid B \cap C)Pr(B \mid C)Pr(C)$$

Let $A_1, ..., A_n$ be a sequence of events. Let $E_i = \bigcap_{j=1}^{i} A_i$

$$Pr(E_n) = Pr(A_n \mid E_{n-1})Pr(E_{n-1}) =$$

$$Pr(A_n \mid E_{n-1})Pr(A_{n-1} \mid E_{n-2})....P(A_2 \mid E_1)Pr(A_1)$$

Theorem

Assume that we run the randomized min-cut algorithm $n(n-1)\log n$ *times and output the minimum size cut-set found in all the iterations. The probability that the output is not a min-cut set is bounded by*

$$(1 - \frac{2}{n(n-1)})^{n(n-1)\log n} \leq e^{-2\log n} = \frac{1}{n^2}.$$

Proof.

The algorithm has a one side error: the output is never smaller than the min-cut value. □

The Taylor series expansion of e^{-x} gives

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \ldots\ldots$$

Thus, for $x < 1$,

$$1 - x \leq e^{-x}.$$