Math 554

Hemanshu Kaul

kaul@iit.edu

A quick review of elementary probability

A **discrete** probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a sample space Ω that is finite (or countable), and $\mathcal{F} = 2^{\Omega}$, the family of allowable events which are all subsets of Ω, and a probability function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ such that

- $0 \leq \mathbb{P}(A) \leq 1 \quad \forall A \subseteq \Omega$
- $\mathbb{P}(\Omega) = 1$
- If $A_1, A_2, \dots$ are pairwise disjoint subsets of Ω then $\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i)$

Union Bound For any sequence of events $A_1, A_2, \ldots$

$$\mathbb{P}\left(\bigcup_i A_i\right) \leq \sum_i \mathbb{P}(A_i)$$

Principle of Inclusion-Exclusion Let $A_1, A_2, A_3, \ldots A_n$ be any events.

$$\mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \mathbb{P}[A_i] - \sum_{i<j} \mathbb{P}[A_i \cap A_j] + \sum_{i<j<k} \mathbb{P}[A_i \cap A_j \cap A_k]$$

$$- \cdots + (-1)^{\ell+1} \sum_{i_1 < i_2 < \cdots < i_\ell} \mathbb{P}\left[\bigcap_{r=1}^{\ell} A_{i_r}\right] + \cdots$$

What does PIE say for $n=3$? $n=4$?

Defn Events $A_1, \dots, A_k$ are <u>mutually independent</u>

if for every $I \subseteq [k] = \{1, \dots, k\}$

$$\mathbb{P}\left[\bigcap_{i \in I} A_i\right] = \prod_{i \in I} \mathbb{P}[A_i]$$

What is this definition for $k=2$ (pairwise independent)?

Does "pairwise independence" $\Rightarrow$ "mutual independence"?

of all pairs

<u>Defn</u> Events $A_1, \dots, A_k$ are <u>mutually independent</u> if for every $I \subseteq [k] = \{1, \dots, k\}$

$$\mathbb{P}\left[\bigcap_{i \in I} A_i\right] = \prod_{i \in I} \mathbb{P}[A_i]$$

What is this definition for $k=2$ (pairwise independent)?

Does "pairwise independence" $\Rightarrow$ "mutual independence"?
of all pairs

<u>Defn</u> For events A, B with $\mathbb{P}[B] \neq 0$, the <u>conditional probability of A given B</u> is

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

So, $\mathbb{P}[A \cap B] = \mathbb{P}[B]\, \mathbb{P}[A \mid B]$ "Principle of deferred decisions"
Do B then A

Law of Total Probability

$A_i \cap A_j = \emptyset \;\forall i \neq j$

$\Omega = \bigcup_{i=1}^{n} A_i$

Let the events $A_1, \dots, A_n$ partition the sample space Ω.

Then $\mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B \cap A_i] = \sum_{i=1}^{n} \mathbb{P}[B \mid A_i]\, \mathbb{P}[A_i]$

Law of Total Probability

$A_i \cap A_j = \phi \ \forall i \neq j$

$\Omega = \bigcup_{i=1}^{n} A_i$

Let the events $A_1, \dots, A_n$ partition the sample space Ω.

Then $\mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B \cap A_i] = \sum_{i=1}^{n} \mathbb{P}[B \mid A_i]\, \mathbb{P}[A_i]$

Defn A random variable is a function $X : \Omega \to \mathbb{R}$

A discrete random variable has range in $\mathbb{N} \cup \{0\}$

"$X = k$" denotes the event $\{a \in \Omega : X(a) = k\}$.

Expectation of X $\quad \mathbb{E}[X] = \sum_{k} k\, \mathbb{P}[X = k]$

Pigeonhole Property $\exists$ element of the probability space for which X has the value as large as (or as small as) $\mathbb{E}[X]$

Linearity of Expectation If $X_1, \ldots, X_n$ are random variables on Ω then $\mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i]$

and $\mathbb{E}[c \sum X_i] = c\, \mathbb{E}[\sum X_i]$

Linearity of Expectation If $X_1, \ldots, X_n$ are random variables on Ω then $\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$

and $\mathbb{E}[c \sum X_i] = c\, \mathbb{E}[\sum X_i]$

Notational Review

For positive functions with an underlying parameter n $(\to \infty)$

- $f = O(g)$ or $g = \Omega(f)$ or $f \lesssim g$ means $f \leq Cg$ for some constant $C > 0$
- $f = o(g)$ or $g = \omega(f)$ or $f \ll g$ means $f/g \to 0$
- $f = \Theta(g)$ means $f = O(g)$ and $g = O(f)$ i.e., $C_1 g \leq f \leq C_2 g$
- $f \sim g$ means $f/g \to 1$ i.e., $f = (1 + o(1))\, g$
- whp (with high probability) means with probability $1 - o(1)$.

How to verify matrix multiplication efficiently?

Given $n \times n$ matrices A, B, C (over integers modulo 2)

How to verify $AB = C$?

How to verify matrix multiplication efficiently?

Given $n \times n$ matrices A, B, C (over integers modulo 2)

How to verify $AB = C$?

Matrix multiplication takes $\Theta(n^3)$ operations (simple algo)

$\Theta(n^{2.38})$ operations (sophisticated algo)

How to verify matrix multiplication efficiently?

Given $n \times n$ matrices A, B, C (over integers modulo 2)

How to verify $AB = C$?

Matrix multiplication takes $\Theta(n^3)$ operations (simple algo)

$\Theta(n^{2.38})$ operations (sophisticated algo)

A Randomized Algorithm

Pick a random vector $\vec{r} = (r_1, r_2, \ldots, r_n) \in \{0,1\}^n$

Compute $A(B\vec{r})$ and $C\vec{r}$ $\longleftarrow$

If $A(B\vec{r}) \neq C\vec{r}$ then $AB \neq C$.

Otherwise, return $AB = C$ $\longleftarrow$

How to verify matrix multiplication efficiently?

Given $n \times n$ matrices A, B, C (over integers modulo 2)

How to verify $AB = C$?

Matrix multiplication takes $\Theta(n^3)$ operations (simple algo)

$\Theta(n^{2.38})$ operations (sophisticated algo)

A Randomized Algorithm

Pick a random vector $\vec{r} = (r_1, r_2, \ldots, r_n) \in \{0,1\}^n$

Compute $A(B\vec{r})$ and $C\vec{r}$ $\longleftarrow$ three matrix-vector multiplications: $\Theta(n^2)$

If $A(B\vec{r}) \neq C\vec{r}$ then $AB \neq C$.

Otherwise, return $AB = C$ $\longleftarrow$ Is this always true?

Theorem If $AB \neq C$ and $\vec{r}$ is chosen uniformly at random from $\{0,1\}^n$ then $\mathbb{P}[AB\vec{r} = C\vec{r}] \leq 1/2$

Theorem If $AB \neq C$ and $\vec{r}$ is chosen uniformly at random from $\{0,1\}^n$ then $\mathbb{P}[AB\vec{r} = C\vec{r}] \leq 1/2$

Proof

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

Theorem If $AB \neq C$ and $\vec{r}$ is chosen uniformly at random from $\{0,1\}^n$ then $\mathbb{P}[AB\vec{r} = C\vec{r}] \leq 1/2$

Proof

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

Let $D = AB - C \neq 0$. Then $AB\vec{r} = C\vec{r} \iff D\vec{r} = 0$

Since $D \neq 0$, it must have a nonzero entry, say d_{11} (WLOG)

$D\vec{r} = 0 \Rightarrow \sum_{j=1}^{n} d_{1j} r_j = 0$ (1^{st} entry) $\iff r_1 = -\dfrac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}$

u.a.r.

Choosing $\vec{r}$ unif at random from $\{0,1\}^n$ is equivalent to each r_k independently u.a.r from $\{0,1\}$ in the order from r_n down to r_1.

Since r_1 is determined by the choice of $r_2, r_3, \ldots, r_n$,

$\mathbb{P}[AB\vec{r} = C\vec{r}] =$

u.a.r.

Choosing $\vec{r}$ <u>unif at random</u> from $\{0,1\}^n$ is equivalent to each r_k independently u.a.r from $\{0,1\}$ in the order from r_n down to r_1.

Since r_1 is determined by the choice of $r_2, r_3, \dots, r_n$,

$$\mathbb{P}[AB\vec{r} = C\vec{r}] = \sum_{(x_2,\dots,x_n)\in\{0,1\}^{n-1}} \mathbb{P}[(AB\vec{r} = C\vec{r}) \cap ((r_2,\dots,r_n) = (x_2,\dots,x_n))]$$

$$\leq$$

Choosing $\vec{r}$ unif at random from $\{0,1\}^n$ (u.a.r.)
is equivalent to each r_k independently u.a.r from $\{0,1\}$
in the order from r_n down to r_1.

Since r_1 is determined by the choice of $r_2, r_3, \dots, r_n$,

$$\mathbb{P}[AB\vec{r} = C\vec{r}] = \sum_{(x_2,\dots,x_n)\in\{0,1\}^{n-1}} \mathbb{P}[(AB\vec{r}=C\vec{r}) \cap ((r_2,\dots,r_n)=(x_2,\dots,x_n))]$$

$$\leq \sum \mathbb{P}\left[\left(r_1 = -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}\right) \cap ((r_2,\dots,r_n)=(x_2,\dots,x_n))\right]$$

Why?

$$= \sum \mathbb{P}\left[r_1 = -\frac{\sum d_{1j} r_j}{d_{11}}\right] \mathbb{P}[(r_2,\dots,r_n)=(x_2,\dots,x_n)]$$

$$\leq \sum \frac{1}{2}\, \mathbb{P}[(r_2,\dots,r_n)=(x_2,\dots,x_n)] = 1/2$$

why?

∎

Choosing $\vec{r}$ $\underbrace{\text{unif at random}}_{}$ from $\{0,1\}^n$ (u.a.r.)
is equivalent to each r_k independently u.a.r from $\{0,1\}$
in the order from r_n down to r_1.

Since r_1 is determined by the choice of $r_2, r_3, \ldots, r_n$,

$$\mathbb{P}[AB\vec{r} = C\vec{r}] = \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \mathbb{P}[(AB\vec{r} = C\vec{r}) \cap ((r_2,\ldots,r_n) = (x_2,\ldots,x_n))]$$

Law of Total Pr.

$D\vec{r} = 0 \Rightarrow r_1 = \cdots$

$$\leq \sum \mathbb{P}\left[\left(r_1 = -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}\right) \cap ((r_2,\ldots,r_n) = (x_2,\ldots,x_n))\right]$$

Independence of r_i's

$$= \sum \mathbb{P}\left[r_1 = -\frac{\sum d_{1j} r_j}{d_{11}}\right] \mathbb{P}[(r_2,\ldots,r_n) = (x_2,\ldots,x_n)]$$

atmost one choice out two possible values for r_1 will make "$r_1 = -\frac{\sum\cdots}{d_{11}}$"

$$\leq \sum \frac{1}{2} \mathbb{P}[(r_2,\ldots,r_n) = (x_2,\ldots,x_n)] = 1/2$$

$\mathbb{P}(\Omega) = 1$ ∎

How can we improve the probability of error (failure) of this randomized algorithm?

How can we improve the probability of error (failure) of this randomized algorithm?

Run the algorithm k times with $\vec{r}$ chosen ind. u.a.r. each time.

If we find $\vec{r}$ s.t. $AB\vec{r} \neq C\vec{r}$ then algo gives $AB \neq C$ correctly.

If $AB\vec{r} = C\vec{r}$ for all the runs then probability of error is at most 2^{-k}

while running time is $\Theta(kn^2)$

e.g. $k=100$ running time is still $\Theta(n^2)$ much faster than $\Theta(n^{2.38})$ for large n.

while probability of making a mistake is atmost 2^{-100}

(computer is more likely to crash than getting a wrong answer)

Probabilistic Method To prove an object exists, define an appropriate probability space where in a random construction of the object works with positive probability.

Theorem Every graph G contains a bipartite subgraph with at least $|E(G)|/2$ edges.

Probabilistic Method To prove an object exists, define an appropriate probability space where in a random construction of the object works with positive probability.

Theorem Every graph G contains a bipartite subgraph with at least $|E(G)|/2$ edges.

Proof Randomly color each vertex of G with 0 or 1 independently u.a.r. ← what does this mean here?

Let E' = set of edges with one endpt. 0 and other 1.

Then $(V(G), E')$ is a bipartite subgraph of G.

Each edge belongs to E' with probability ½

$\therefore \mathbb{E}[|E'|] = \frac{1}{2}|E(G)|$ by lin. of exp. Hence $\exists$ a coloring with $|E'| \geq \frac{1}{2}|E(G)|$ as needed

Probabilistic Method To prove an object exists, define an appropriate probability space where in a random construction of the object works with positive probability.

Theorem Every graph G contains a bipartite subgraph with at least $|E(G)|/2$ edges.

Proof Randomly color each vertex of G with 0 or 1 ind. w. prob. $1/2$

Let $X_e = \begin{cases} 1 & \text{if endpoints of } e \text{ have different colors} \\ 0 & \text{otherwise} \end{cases}$ Indicator r.v. for "good" edges

Then $X = \sum_{e \in E(G)} X_e$ counts the number of edges in the bipartite subgraph.

Probabilistic Method To prove an object exists, define an appropriate probability space where in a random construction of the object works with positive probability.

Theorem Every graph G contains a bipartite subgraph with at least $|E(G)|/2$ edges.

Proof Randomly color each vertex of G with 0 or 1 ind. w. prob. ½

Let $X_e = \begin{cases} 1 & \text{if endpoints of } e \text{ have different colors} \\ 0 & \text{otherwise} \end{cases}$ Indicator r.v. for "good" edges

Then $X = \sum_{e \in E(G)} X_e$ counts the number of edges in the bipartite subgraph.

$$\mathbb{E}[X] = \sum_e \mathbb{E}[X_e] = \sum_e \mathbb{P}[X_e = 1] = \sum_e \left(\tfrac{1}{4} + \tfrac{1}{4}\right) = \sum_e \tfrac{1}{2} = \tfrac{1}{2}|E(G)|$$

$\therefore \exists$ coloring with $X \geq \frac{1}{2}|E(G)|$ by pigeonhole property.