CSCE 658: Randomized Algorithms

Lecture 3

Samson Zhou

Class Logistic Updates

- Course materials: https://samsonzhou.github.io/CSCE658-S24/csce658-s24

Class Logistic Updates

- Problem Set 1 posted, due next Thursday, February 1, 2024, 5 pm CT
- PS1 groups sent via e-mail, please confirm receipt by reply-all

- Submit PS1 via e-mail as a PDF, typeset in LaTeX
- LaTeX template for PS1 available on class webpage, for your convenience

Trivia Question #1 (Birthday Paradox)

- Suppose we have a fair n-sided die. How many times should we roll the die before the probability we see a repeated outcome among the rolls is at least $\frac{1}{2}$? Example: 1, 5, 2, 4, 5

- $\Theta(1)$
- $\Theta(\log n)$
- $\Theta(\sqrt{n})$
- $\Theta(n)$

Trivia Question #2 (Limits)

- Let $c > 0$ be a constant. What is $\lim_{n\to\infty}\left(1 - \frac{c}{n}\right)^n$?

- 0
- $\frac{1}{c}$
- $\frac{1}{2c}$
- $\frac{1}{e^c}$
- 1

Birthday Paradox

- Suppose we have a room with 367 people. What is the probability that two people share the same birthday?

Birthday Paradox

- Suppose we have a room with 367 people. What is the probability that two people share the same birthday?

- Suppose we have a room with 23 people. What is the probability that two people share the same birthday?

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4$,... times. What is the probability we DO NOT see a repeated outcome among the rolls?

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4$,... times. What is the probability we DO NOT see a repeated outcome among the rolls?

$$\left(1 - \frac{0}{n}\right)$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4,...$ times. What is the probability we DO NOT see a repeated outcome among the rolls?

$$\left(1 - \frac{0}{n}\right)\left(1 - \frac{1}{n}\right)$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4$,... times. What is the probability we DO NOT see a repeated outcome among the rolls?

$$\left(1 - \frac{0}{n}\right)\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4,\ldots$ times. What is the probability we DO NOT see a repeated outcome among the rolls?

$$\left(1-\frac{0}{n}\right)\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4, \ldots$ times. What is the probability we DO NOT see a repeated outcome among the rolls?

$$\left(1 - \frac{0}{n}\right)\left(1 - \frac{1}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right) < \frac{1}{2}$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4$,... times. What is the probability we DO NOT see a repeated outcome among the rolls?

$$\left(1 - \frac{0}{n}\right)\left(1 - \frac{1}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right) < \frac{1}{2} \qquad \text{for} \qquad k = O(\sqrt{n})$$

Birthday Paradox

- Suppose we have a fair n-sided die. “On average”, how many times should we roll the die before we see a repeated outcome among the rolls?
- $O(\sqrt{n})$
- But is it $\Theta(\sqrt{n})$?

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4$,... times. What is the probability we DO NOT see a repeated outcome among the rolls?

- Let S_i be the event that the i-th roll is a repeated outcome, conditioned on the previous rolls not being a repeated outcome
- $\Pr[S_i] = \frac{i-1}{n}$
- $\Pr[S_1 \cup \cdots \cup S_k] \leq ???$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4,\ldots$ times. What is the probability we DO NOT see a repeated outcome among the rolls?

- Let S_i be the event that the i-th roll is a repeated outcome, conditioned on the previous rolls not being a repeated outcome
- $\Pr[S_i] = \frac{i-1}{n}$
- $\Pr[S_1 \cup \cdots \cup S_k] \leq \frac{0}{n} + \ \ldots + \frac{k-1}{n} \leq \frac{k^2}{n}$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4$,... times. What is the probability we DO NOT see a repeated outcome among the rolls?

- Let S_i be the event that the i-th roll is a repeated outcome, conditioned on the previous rolls not being a repeated outcome
- $\Pr[S_i] = \frac{i-1}{n}$
- $\Pr[S_1 \cup \cdots \cup S_k] \leq \frac{0}{n} + \ \ldots + \frac{k-1}{n} \leq \frac{k^2}{n}$

Union Bound

Birthday Paradox

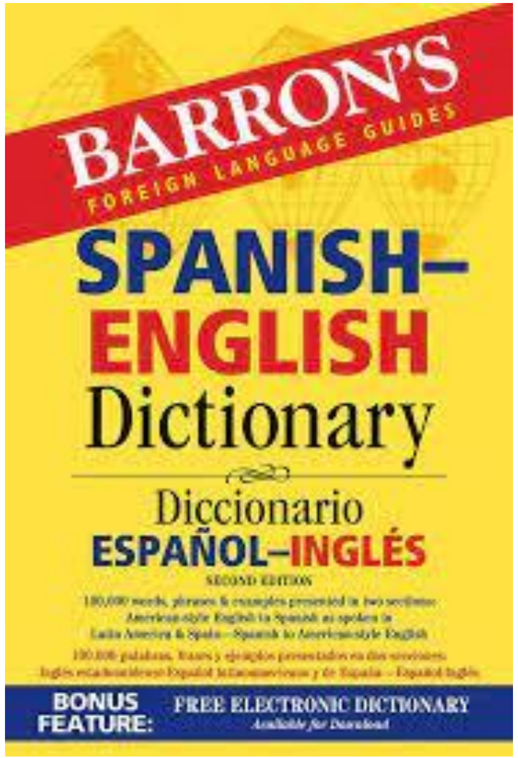
- Suppose we have a fair n-sided die. “On average”, how many times should we roll the die before we see a repeated outcome among the rolls?
- $\Theta(\sqrt{n})$

Case Study

- We are trying to learn a new language on an app, which claims to have a database of *1 million words*

- Each time we ask the app, it gives us a random word in the database

- We want to verify the claim

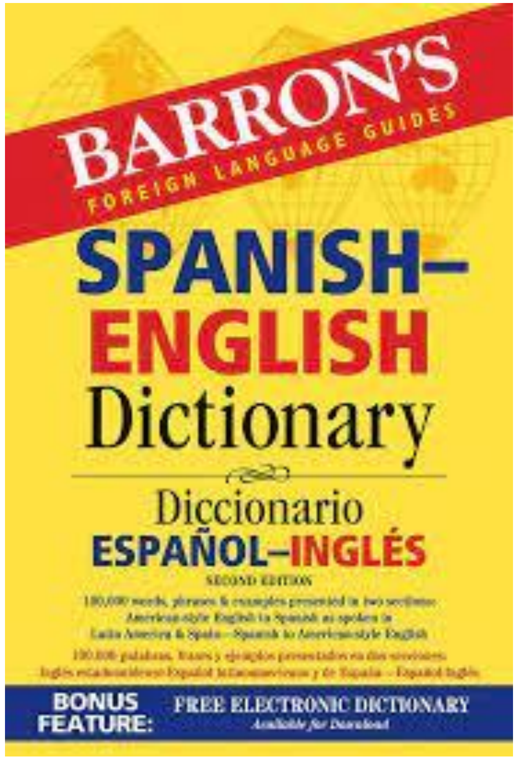


Case Study

- We could use the app until we see 1 million unique words, but that would take at least *1 million checks*

- Instead, we use the app for *1000 times* and count the number of pairwise duplicates

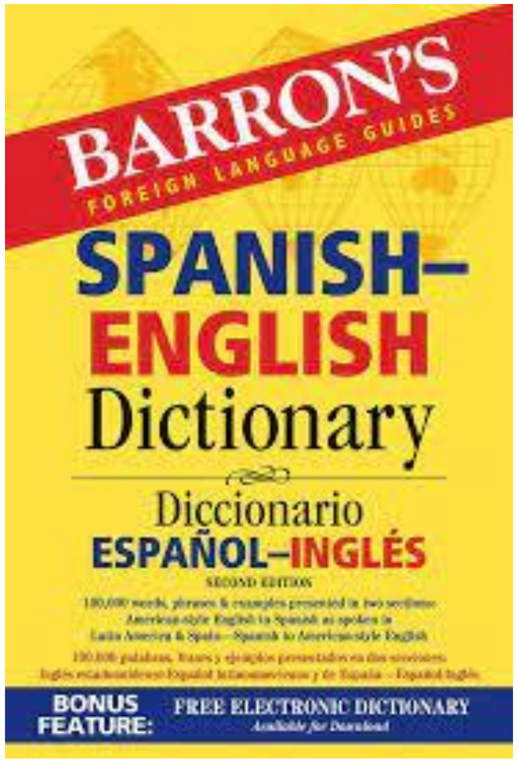
- If there are many duplicates, the database is probably not very large



Case Study

- We use the app for k times and count the number of pairwise duplicates

- If we see the same word on the 3-rd time, the 100-th time, and the 205-th time, there are 3 pairwise duplicates: $(3, 100)$, $(3, 205)$, $(100, 205)$



Expected Value

- The expected value of a random variable X over a sample space Ω is:

$$\mathrm{E}[X] = \sum_{x \in \Omega} \Pr[X = x] \cdot x$$

- The “average value of the random variable"

- Linearity of expectation: $\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

Expected Value

- Suppose we roll a 6-sided die

- Let X be the outcome of the roll

- What is $\mathrm{E}[X]$?

Linearity of Expectation

- Linearity of expectation: $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

$$\mathrm{E}[X+Y] = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} \Pr[X = x, Y = y] \cdot (x + y)$$

Linearity of Expectation

- Linearity of expectation: $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

$$\mathrm{E}[X+Y] = \sum_{x \in \Omega_\mathrm{X}} \sum_{y \in \Omega_\mathrm{Y}} \Pr[X = x, Y = y] \cdot (x + y)$$

$$= \sum_{x \in \Omega_\mathrm{X}} \sum_{y \in \Omega_\mathrm{Y}} \Pr[X = x, Y = y] \cdot x + \sum_{x \in \Omega_\mathrm{X}} \sum_{y \in \Omega_\mathrm{Y}} \Pr[X = x, Y = y] \cdot y$$

Linearity of Expectation

- Linearity of expectation: $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

$$
\begin{aligned}
\mathrm{E}[X+Y] &= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} \Pr[X = x, Y = y] \cdot (x + y) \\
&= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} \Pr[X = x, Y = y] \cdot x + \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} \Pr[X = x, Y = y] \cdot y \\
&= \sum_{x \in \Omega_X} x \sum_{y \in \Omega_Y} \Pr[X = x, Y = y] + \sum_{y \in \Omega_Y} y \sum_{x \in \Omega_X} \Pr[X = x, Y = y]
\end{aligned}
$$

Linearity of Expectation

- Linearity of expectation: $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

$$\begin{aligned}
\mathrm{E}[X+Y] &= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} \Pr[X=x, Y=y] \cdot (x+y) \\
&= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} \Pr[X=x, Y=y] \cdot x + \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} \Pr[X=x, Y=y] \cdot y \\
&= \sum_{x \in \Omega_X} x \sum_{y \in \Omega_Y} \Pr[X=x, Y=y] + \sum_{y \in \Omega_Y} y \sum_{x \in \Omega_X} \Pr[X=x, Y=y] \\
&= \sum_{x \in \Omega_X} x \cdot \Pr[X=x] + \sum_{y \in \Omega_Y} y \cdot \Pr[Y=y] = \mathrm{E}[X] + \mathrm{E}[Y]
\end{aligned}$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4$,... times. What is the probability we DO NOT see a repeated outcome among the rolls?

$$\left(1 - \frac{0}{n}\right)\left(1 - \frac{1}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right)$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4,\ldots$ times. What is the expected number of pairwise collisions among the rolls?

- Let X_i be the number of pairwise collisions on the i-th roll

- We have $\mathrm{E}[X_i] = \frac{i-1}{n}$

Birthday Paradox

- Let X be the number of pairwise collisions after k rolls

- What is $\mathrm{E}[X]$?

Birthday Paradox

- Let X be the number of pairwise collisions after k rolls

$$
\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[X_1 + \cdots + X_k] \\
&= \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_k] \\
&= \frac{0}{n} + \cdots + \frac{k-1}{n} \\
&= \frac{k(k-1)}{2n}
\end{aligned}
$$

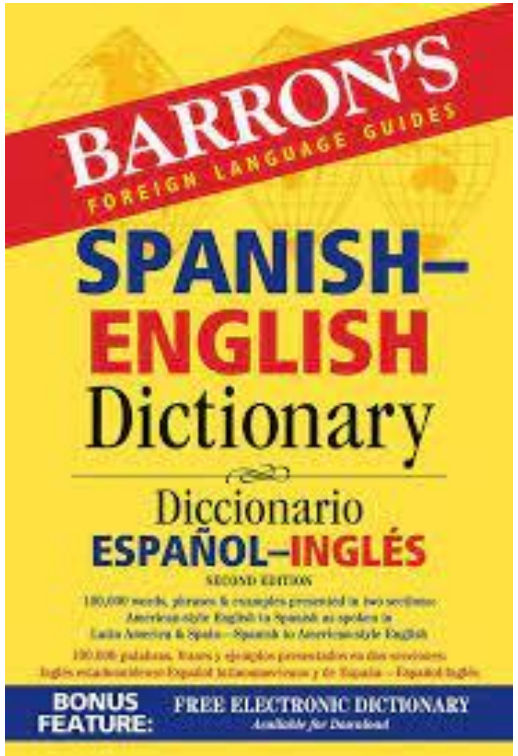
Birthday Paradox

- $\mathrm{E}[X] = \frac{k(k-1)}{2n}$
- $\frac{(k-1)^2}{2n} \leq \mathrm{E}[X] \leq \frac{k^2}{2n}$
- $k = 2\sqrt{n} + 1$ implies $\mathrm{E}[X] \geq 1$
- $k = \frac{\sqrt{n}}{2}$ implies $\mathrm{E}[X] \leq \frac{1}{4}$

Case Study

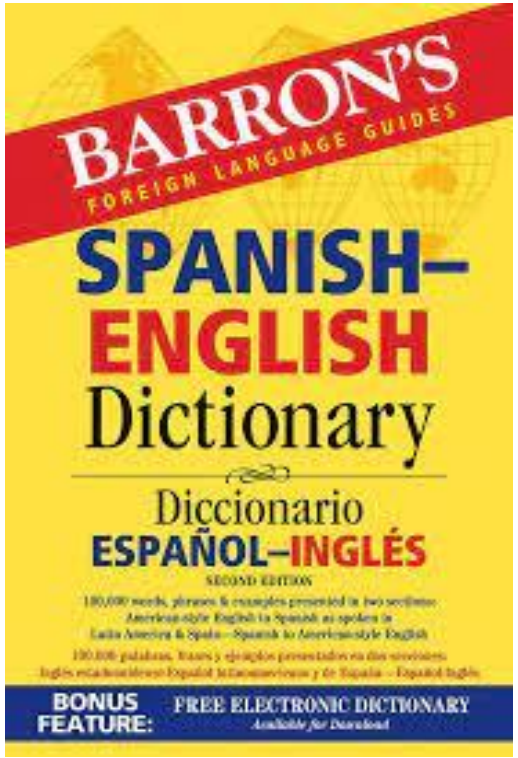
- We use the app for $k = 1000$ times and count the number of pairwise duplicates

- If the database contains *1 million words*, the expected number of pairwise duplicates is $\mathrm{E}[X] = \frac{k(k-1)}{2n} < 0.5$



Case Study

- If the database contains *1 million words*, the expected number of pairwise duplicates is $\mathrm{E}[X] = \frac{k(k-1)}{2n} < 0.5$
- …We see 20 duplicates
- We think the claim is incorrect, but how can we be sure?



Concentration Inequalities

- Concentration inequalities bound the probability that a random variable is “far away” from its expectation

- Often used in understanding the performance of statistical tests, the behavior of data sampled from various distributions, and for our purposes, the guarantees of randomized algorithms

Markov's Inequality

- Let $X \geq 0$ be a non-negative random variable. Then for any $t > 0$:

$$\Pr[X \geq t \cdot E[X]] \leq \frac{1}{t}$$

Proof of Markov's Inequality

- Let $X \geq 0$ be a non-negative random variable. Then for any $t > 0$:

$$
\begin{aligned}
\mathrm{E}[X] &= \textstyle\sum_{x \in \Omega} \Pr[X = x] \cdot x \\
&= \textstyle\sum_{x \geq t \cdot \mathrm{E}[X]} \Pr[X = x] \cdot x + \sum_{x < t \cdot \mathrm{E}[X]} \Pr[X = x] \cdot x \\
&\geq \textstyle\sum_{x \geq t \cdot \mathrm{E}[X]} \Pr[X = x] \cdot x \\
&\geq t \cdot \mathrm{E}[X] \textstyle\sum_{x \geq t \cdot \mathrm{E}[X]} \Pr[X = x] \\
&= t \cdot \mathrm{E}[X] \cdot \Pr[X \geq t \cdot \mathrm{E}[X]]
\end{aligned}
$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4, \ldots$ times. What is the probability we DO NOT see a repeated outcome among the rolls?

$$\left(1 - \frac{0}{n}\right)\left(1 - \frac{1}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right)$$

Birthday Paradox

- Suppose we have a fair n-sided die that we roll $k = 1, 2, 3, 4$,... times. What is the expected number of pairwise collisions among the rolls?

- Let X_i be the number of pairwise collisions on the i-th roll

- We have $\mathrm{E}[X_i] = \frac{i-1}{n}$

Birthday Paradox

- $\mathrm{E}[X] = \frac{k(k-1)}{2n}$
- $\frac{(k-1)^2}{2n} \leq \mathrm{E}[X] \leq \frac{k^2}{2n}$
- $k = 2\sqrt{n} + 1$ implies $\mathrm{E}[X] \geq 1$
- $k = \frac{\sqrt{n}}{2}$ implies $\mathrm{E}[X] \leq \frac{1}{4}$

Birthday Paradox

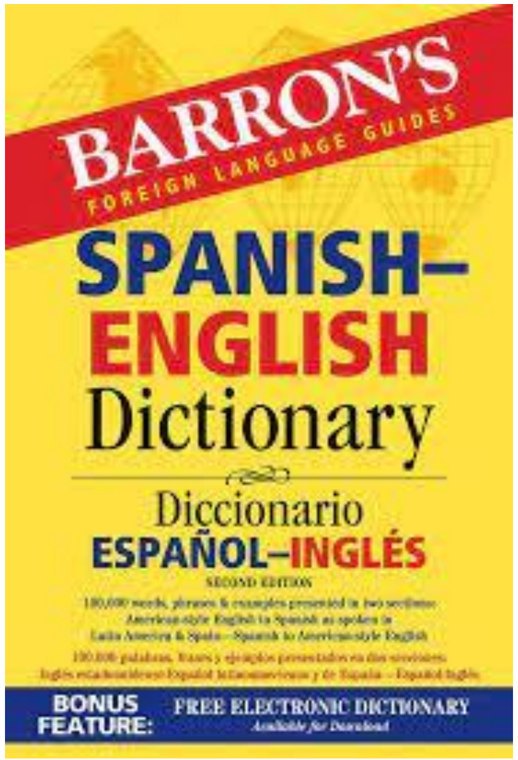
- $\mathrm{E}[X] = \frac{k(k-1)}{2n}$
- $\frac{(k-1)^2}{2n} \leq \mathrm{E}[X] \leq \frac{k^2}{2n}$
- $k = 2\sqrt{n} + 1$ implies $\mathrm{E}[X] \geq 1$
- $k = \frac{\sqrt{n}}{2}$ implies $\mathrm{E}[X] \leq \frac{1}{4}$, and by Markov's inequality, $\Pr[X \geq 1] \leq \frac{1}{4}$

Case Study

- If the database contains *1 million words*, the expected number of pairwise duplicates is $\mathrm{E}[X] = \frac{k(k-1)}{2n} < 0.5$

- …We see 20 duplicates

- We think the claim is incorrect, but how can we be sure?



Case Study

- If the database contains *1 million words*, the expected number of pairwise duplicates is $\mathrm{E}[X] = \frac{k(k-1)}{2n} < 0.5$
- …We see 20 duplicates
- $\Pr[X \geq 20] \leq \frac{1}{40}$