Tutorial I

20 February 2020 13:57

1.) A CRASH COURSE IN PROBABILITY THEORY

2.) RANDOMIZED QSORT

1.) PROBABILITY SPACE ~ A set of all possible outcomes of a random experiment.

S - set

finite
countable → in this course

uncountable → not in this course

EXAMPLE — a set of all n-bit strings

2.) EVENTS — $E \subseteq S$

Example - a string with exactly 3 symbols '1'

3.) PROBABILITY FUNCTION

$$p: S \to [0,1]$$

$$\sum_{i \in S} p(i) = 1$$

Example: - Uniform distribution $\forall x \in \{0,1\}^n$ $p(x) = \frac{1}{2^n}$

$$P(E) = \sum_{i \in E} p(i)$$

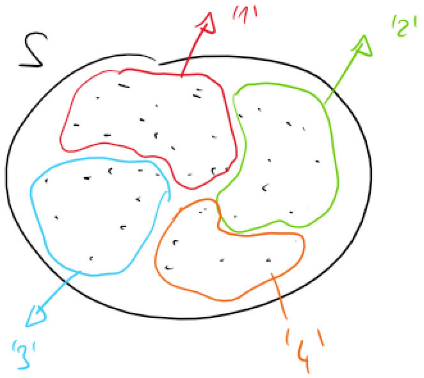
What is the probability to obtain a 5-bit string with exactly 3 symbols '1' (call it event E)

$$P(E) = \binom{5}{3} = 10 \cdot \frac{1}{2^5} = \frac{10}{32} \checkmark$$

RANDOM VARIABLES

(X, Y, Z)

$X: S \to \mathbb{R}$



Essentially X is a division of probability space into mutually exclusive and collectively exhaustive set of events.

EXAMPLE X - is the number of symbols '1' in n-bit string

Q For $n = 4$, What is the distribution of X

$Pr[X=0] = 1/16$ $\qquad Pr[X > 1] = 11/16$

$Pr[X=1] = 4/16$ $\qquad Pr[X \geq 2 \wedge X < 4] = \frac{10}{16}$

$Pr[X=2] = 6/16$

$Pr[X=3] = 4/16$

$Pr[X=4] = 1/16$

EXPECTATION OF RANDOM VARIABLES

$$E(X) = \sum_{\substack{i \in \mathbb{R} \\ \in Im(X)}} i \cdot Pr(X = i)$$

EXAMPLE: $E(X) = \sum_{i=0}^{4} i \cdot Pr[X=i] = 0 \cdot Pr[X=0] + 1 \cdot Pr[X=1] + 2 \cdot Pr[X=2] + 3 \cdot Pr[X=3] + 4 \cdot Pr[X=4]$

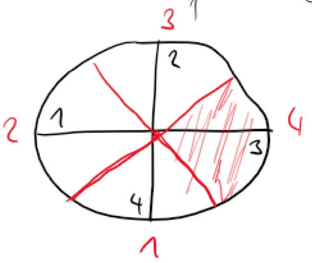
$= 0 + 4/16 + 12/16 + 12/16 + 4/16$

$= 2$

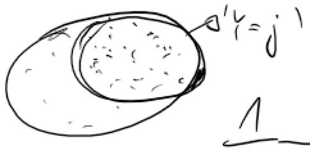
CONDITIONAL PROBABILITIES

GIVEN 2 random variables X and Y (over the same random experiment) define conditional probability of X given $Y = y$

$$\Pr(X=i \mid Y=j) = \boxed{\Pr(X=i, Y=j) \,/\, \Pr(Y=j)} \qquad [\Pr(Y=j)\neq 0]$$



Intuitively we are creating a new probability space equal to an event $(Y=j)$



$\dfrac{1}{\Pr(Y=j)}$ renormalizes the original probabilities

Examples

$S = \{0,1\}^4$

X – number of '1'

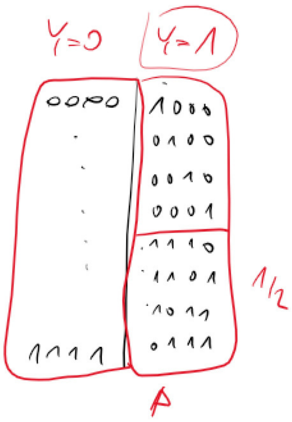
Y – parity of the string $\begin{pmatrix} \text{even} = 0 \\ \text{odd} = 1 \end{pmatrix}$

$\Pr[Y=0] = 1/2$

$\Pr[Y=1] = 1/2$

$$\Pr(X=3 \mid Y=0) = 0 = \overset{=0}{P(X=3, Y=0)} \,/\, \overset{=1/2}{\Pr[Y=0]}$$

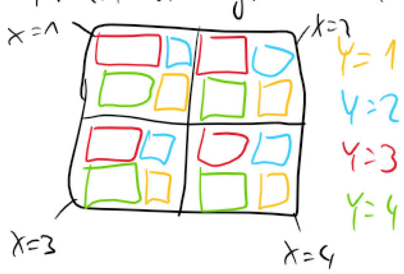
$$\Pr(X=3 \mid Y=1) = \frac{\Pr[X=3, Y=1]}{\Pr[Y=1]} = \frac{4/16}{1/2} = \frac{1}{2}$$



INDEPENDENCE

X and Y are independent if for all i, j

$$\Pr(X=i \mid Y=j) = \Pr(X=i)$$



EXAMPLE

ARE Y and X from the previous example independent?

$n=3$ $\qquad$ $x=4$

EXAMPLE

ARE Y and X from the previous example independent?

X - number of '1'

Y - parity

$$\# \; Pr(X=3) = 1/4$$

$$Pr(X=3 \mid Y=0) = 0$$

Z is the value of the first bit of $\{0,1\}^4$

is Z and Y independent?

$$Pr(Z=1) = 1/2$$

$$Pr(Z=0) = 1/2$$

$$Pr(Z=1 \mid Y=1) = 1/2$$

$$Pr(Z=0 \mid Y=1) = 1/2$$

$$Pr(Z=1 \mid Y=0) = 1/2$$

$$Pr(Z=0 \mid Y=0) = 1/2$$

LINEARITY OF EXPECTATION

$$W = X + Y + Z$$

W is a well defined r.v.

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

$$E\left(\sum_i a_i X_i\right) = \sum_i a_i E(X_i)$$

$$E\left(\sum_i X_i\right) = \sum_i E(X_i)$$

this is often much easier to calculate

$$E(W) = \sum_{i,j,k} (i+j+k) \cdot Pr(X=i, Y=j, Z=k)$$

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

$$= 2 + 1/2 + 1/2 = 3$$

$E(X_1 \cdot X_2) \neq E(X_1) \cdot E(X_2)$

THE LAW OF TOTAL PROBABILITY

r.v. X and Y

$$\Pr(X=i) = \sum_{j \in \mathrm{Im}(Y)} \underbrace{\Pr(X=i \mid Y=j) \cdot \Pr(Y=j)}_{= P(X=i,\, Y=j)}$$

$\boxed{\text{RQUICKSORT}}$

IN: Collection of numbers S

OUT: Ordered list of elements in S

0.) if S contains a single element output it

1.) Choose a pivot $y \in S$ uniformly at random

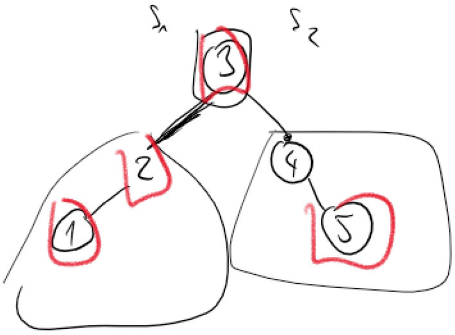
2.) Create S_1 which contains all $s \in S$ $s < y$

Create S_2 which contains all $s \in S$ $s > y$

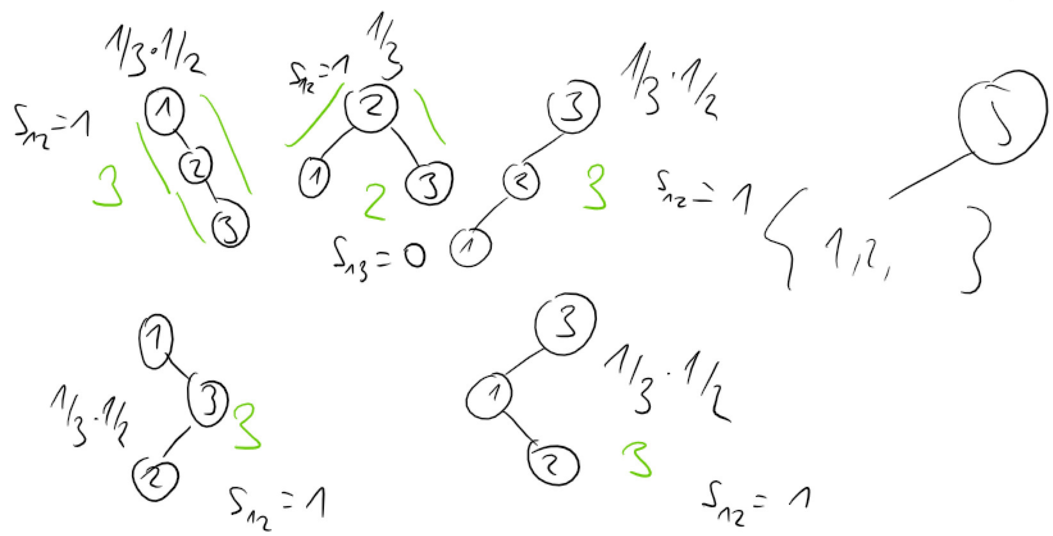
3.) Output $(\text{rquicksort}(S_1), y, \text{rquicksort}(S_2))$

Probability space is a set of ordered trees with $|S| = n$ nodes

$S = \{\underbrace{1, 2}_{S_1}, 3, \underbrace{4, 5}_{S_2}\}$



$1/5 \cdot 1/2 \cdot 1/2 =$



$1/3 \cdot 1/2 \cdot 1/2 =$

$S_{12} = 1$ $\quad$ 3 $\quad$ $S_{12} = 1$

X = r.v. – number of comparisons in a given tree

$$E(X \text{ (for } n=3)) = \underbrace{4 \cdot 1/3 \cdot 1/2 \cdot 3}_{Pr\{x=3\} \cdot 3} + \underbrace{1/3 \cdot 2}_{Pr\{x=2\}} = \frac{8}{3}$$

$E(X_3)$ $E(X_4)$ ... $E(X_{100})$

$f(n) = E(X_n)$

Q How does scale with n? $\quad O(n \cdot \log n)$

$\forall i,j \in S$ define a random variable S_{ij}

$S_{ij} = 1$ if i and j are compared in a given run

$= 0$ otherwise

$X = \sum_{i<j} S_{ij}$ → total number of comparisons

$$E(X) = E\left(\sum_{i<j} S_{ij}\right) = \sum_{i<j} E(S_{ij}) \leq \sum_{i<j} P_{ij}^{*} \in O(n \lg n)$$

$$E(S_{ij}) = 0 \cdot Pr(S_{ij}=0) + 1 \cdot Pr(S_{ij}=1) = Pr(S_{ij}=1) < P_{ij}^{*}$$

Step 1 What is the probability that i and j get compared?

$$Pr(S_{ij} = 1) = \frac{2}{|S|}$$

Step 2 $\quad |S_1|, m, |S_2|$ $\quad$ Probability of i,j to get compared in step 2

Step 2 $|S_1|, y, |S_2|$ of i,j to get compared in step 2

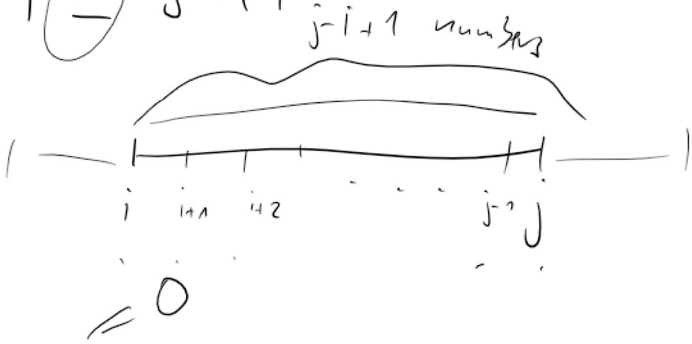
$Pr(S_{ij}=1 \mid \boxed{i<y<j}) = 0$ depends on y.

$Pr(S_{ij}=1 \mid \boxed{y<i<j}) = \boxed{\frac{2}{|S_2|}} < \frac{2}{j-i-1}$ $i,j \in S_1$

$Pr(S_{ij}=1 \mid \boxed{i<j<y}) = \boxed{\frac{2}{|S_1|}} < \frac{2}{j-i-1}$

Step k   let's bound the probability of comparing i,j conditioned on the event of both i,j being in a same subset S_k $|S_k| \geq j-i+1$

$$\frac{2}{|S_k|} < \boxed{\frac{2}{j-i+1}}$$



$$Pr(S_{ij}=1) = Pr(S_{ij}=1 \mid \text{they are not in the same set}) \cdot Pr(\text{not in the same set})$$
$$+ Pr(S_{ij}=1 \mid \text{they are in the same set}) \cdot Pr(\text{they are in the same set})$$
$$= Pr(S_{ij}=1 \mid \text{they are in the same set}) \cdot Pr(\text{they are in the same set})$$
$$\leq Pr(S_{ij}=1 \mid \text{they are in the same set})$$
$$\leq \frac{2}{j-i+1}$$

$$E(X) \leq \sum_{i<j} \frac{2}{j-i+1} = \dots = n \cdot \underbrace{\sum_{i=1}^{n} \frac{1}{i}}_{\Theta(\log n)} = \Theta(n \log n)$$

in the slides