Thursday, 14. May 2020

Foundations of Data science

Foundations of data science, summer 2020

JONATHAN LENNARTZ, MICHAEL NÜSKEN, ANNIKA TARNOWSKI

4. Exercise sheet
Hand in solutions until Thursday, 14 May 2020, 12:00

Exercise 4.1 (Random point in a ball). (8+3 points)

Write a Python routine (eg. in a jupyter notebook) to choose a uniformly random point in a unit ball of dimension d. $\boxed{5}$

Use your routine with $d = 2$ to pick $1\,000$ points and plot them, so that we can visually check the correctness of the algorithm. $\boxed{3}$

Can you also plot a sample with $d = 3$? $\boxed{+3}$

Exercise 4.2 (Random exit). (10 points)

Consider an experiment for which some general indicator variable X has

$$\text{prob}\,(X = 1) = p \neq 0.$$

We create new random variables as following: Repeat the experiment to obtain X_i for all $i \in \mathbb{N}'$. That is, ... (fill in!)

Then we set the exit time

$$T = \min\,\{i \mid X_i = 1\}\,.$$

We simply may assume that this minimum always exists.

(i) Prove that $\boxed{2}$

$$T = i \iff X_i = 1 \wedge X_{i-1} = 0 \wedge \cdots \wedge X_1 = 0.$$

(ii) Compute(!) prob $(T = 1)$, prob $(T = 2)$, prob $(T = 3)$. $\boxed{1}$

(iii) Prove a formula for prob $(T = i)$. $\boxed{2}$

(iv) Express $\mathrm{E}(T)$ as an infinite sum. $\boxed{1}$

(v) Use the geometric series to derive a formula for $\sum_{i=1}^{\infty} ix^{i-1}$. $\boxed{1}$

(vi) Compute the expected exit time $\mathrm{E}(T)$. $\boxed{1}$

(vii) Consider the method 0 algorithm from class. What is its expected runtime? $\boxed{2}$

i) Given : $T=i \Rightarrow X_i=1 \wedge X_{i-1}=0 \wedge \cdots \wedge X_1=0$

since we know by definition of indicator variable:

$X \leftarrow \{0, 1\}$, $P(1)$ is the probability of the indicated event. $A = \{X = 1\}$.

iii) $T_i = X_i = 1 \wedge X_{i-1} = 0 \wedge \cdots \wedge X_1 = 0$

$\text{prob}(T_i) = \text{Prob}(X_i = 1 \ldots \wedge X_{i-1} = 0 \ldots X_1 = 0)$

$= \text{prob}(X_i = 1) \cdot \text{prob}(X_{i-1} = 0) \cdot \ldots \text{prob}(X_1 = 0)$

$= P(1-P) \ldots, (1-P)$

ii) $T(1) = P$, $T(2) = \text{prob}(X_2 = 1) \cdot \text{Prob}(X_1 = 0)$

$= P(1-P)$

$T(3) = \text{prob}(X_3 = 1) \cdot \text{prob}(X_2 = 0) \cdot \text{prob}(X_1 = 0)$

$= P \cdot (1-P)(1-P) = P(1-P)2$

Foundations of Data science

iv) By the definition we know

$$E(D) = \sum_{x \in D} x \, prob(D = x)$$

Here, $E(T) = \sum_{x=1}^{\infty} x \left(prob \; T = x \right)$

$$= \sum_{i=1}^{\infty} x \left(p \cdot (1-p) \right)^{x-1}$$