02-21-2020

HOMEWORK-2

DIPENDRA NATH BAGCHI
NetId: dbagchi2

1) Given.

$m = \#$ of candy bars

$L =$ length of each candy bar.

$L \sim$ Uniform $(0, L)$

Each candy bar is broken into two pieces $x, L-x$

where $x =$ length of the one piece

$L-x =$ length of the other piece.

let $l_1, l_2 \dots l_m$ be the longest of the two pieces,

This implies,

$$l_1 = l_2 \dots l_m \sim \text{Uniform}\left(L/2, L\right)$$

All of them have to have length of "at-least" $L/2$ to become the longest of the two.

$\therefore$ the density func is $f_L(l) = \dfrac{1}{L - 1/2} = \dfrac{2}{L}$

$$\boxed{f_L(l) = \frac{2}{L}}$$

Now, we have to find the minimum of the maximum lengths.

$\therefore$ From order statistics we know that, the cumulative distribution fun^n (cdf) of the minimum value is given by,

$$F_{x_{(1)}}(x) = 1 - [1 - F(x)]^n \quad \text{— (A)}$$

where $x_{(1)}$ is the smallest value of the sample

n : # of data points.

$F_X(x)$: cdf of the original data.

$F_{x_{(1)}}(x)$: cdf of the minimum value, $x_{(1)}$

Using this formulae (A) in our case we can write,

$$F_{L_{(1)}}(x) = \min\{L_1, L_2 \ldots L_m\} = 1-\left[1-F_L(x)\right]^m \quad \text{—— (i)}$$

∴ we need to find cdf of L, $F_L(l)$

cdf of L (the longer of all the divided parts), we get.

$$F_L(x) = \int_{L/2}^{x} \left(\frac{2}{L}\right) \cdot dx.$$

$$= \frac{2}{L}\left[x \Big|_{L/2}^{x} \right]$$

$$= \frac{2}{L}\left(x - \frac{L}{2}\right)$$

$$\boxed{F_L(x) = \left(\frac{2x}{L} - 1\right)}$$

→ cdf of $F_L(x)$: the longer of the lengths.

(B)

using eq. (B) in (i) we get,

$$F_{L(1)}(x) = 1-\left[1-\frac{2x}{L}+1\right]^{m}$$

$$= 1-\left[2-\frac{2x}{L}\right]^{m}$$

Now, the density fcn of the smallest of the m- longest value, is definded by,

$$f_{L(1)}(x) = \frac{d}{dx} F_{L(1)}(x)$$

$$= -m\left[2-\frac{2x}{L}\right]^{m-1} \cdot \left(-\frac{2}{L}\right)$$

$$\boxed{f_{L(1)}(x) = \frac{2m}{L}\left[2-\frac{2x}{L}\right]^{m-1}}$$

Now, to calculate the average we find the expected value of all the larger lengths and integrate over $L/2, L$

$$\therefore E(L) = \frac{2m}{L}\int_{L/2}^{L} x\left(2-\frac{2x}{L}\right)^{m-1} dx.$$

using IBP

$u = x,$ $\quad v = \frac{-L}{2m}\left(2-2x/L\right)^{m}$

$du = dx$ $\quad dv = \left(2-\frac{2x}{L}\right)^{m-1}$

$$= \frac{2m}{L}\left[\frac{-xL}{2m}\left(2-\frac{2x}{L}\right)^m \Bigg|_{L/2}^{L} + \frac{L}{2m}\int_{L/2}^{L}\left(2-\frac{2x}{L}\right)^m \cdot dx \right]$$

$$= -x\left(2-\frac{2x}{L}\right)^m\Bigg|_{L/2}^{L} - \frac{L\left(2-\frac{2x}{L}\right)^{m+1}}{2(m+1)}\Bigg|_{L/2}^{L}$$

$$= -L\cancel{(2-2)}^{0} + \frac{L}{2}(2-1) - \frac{L}{2}\cancel{\frac{(2-2)}{m+1}}^{0} + \frac{L}{2}\frac{(2-1)}{m+1}$$

$$= \frac{L}{2} + \frac{L}{2(m+1)} = \frac{[(m+1)+1]L}{2(m+1)}$$

$$= \frac{(m+2)L}{2(m+1)}$$

Hence,

$$\boxed{E(L) = \frac{L}{2}\frac{(m+2)}{(m+1)}}$$

2 Randomized Selection Algorithm,

i) m: m-many pivots to split the original array of length n.

ii) The recursion happens on the smallest of the longest size array (for worst case scenario)

~~for~~ For worst case running time,

Let's assume there are 'n' many data points (~~i.e.~~ i.e array of length n)

and 'm' : # of random pivots to distribute the data.

This, parts take $m\,c(n-1)$-time as for each m there has to be $(n-1)$ many computations to create the buckets of array $< m$, array $> m$, array $= m$.

In the worst case scenario, we know. we will always be recursing on the shortest of ~~longest~~ of the two arrays. The average number of the elements cannot be greater than $\frac{n}{2}\left(\frac{m+2}{m+1}\right)$.

$\because$ we get this value from question 1.

$\therefore$ The total running-time for worst case scenario is

$$A_m(n) \leq c(n-1)\,m + A\left(\frac{n}{2}\left(\frac{m+2}{m+1}\right)\right)$$

$$\leq cnm + A\left(\frac{n}{2}\left(\frac{m+2}{m+1}\right)\right)$$

$$\leq cn + A\left(\frac{n}{2}\left(\frac{m+2}{m+1}\right)\right) \qquad [\because m \lll n]$$

Now, using the recursion logic, we have,

$$A_m(n) \leq cn\left[1 + \frac{1}{2}\left(\frac{m+2}{m+1}\right) + \left[\frac{1}{2}\left(\frac{m+2}{m+1}\right)\right]^2 \cdots\right]$$

Using geometric series we know.

$$cn\left[\frac{1}{1-\frac{1}{2}\left(\frac{m+2}{m+1}\right)}\right]$$

$$= cn\left[\frac{2(m+1)}{2m+\cancel{2}-m-\cancel{2}}\right]$$

$$= cn\left[\frac{2(m+1)}{m}\right]$$

$$\therefore \boxed{A_m(n) \leq cn\left(\frac{2(m+1)}{m}\right)}$$

Hence, even in Randomized Selection Alogorithm with m pivots, the running time is linear.