

Questions: Find a simple function f such that the following expressions are $\Theta(f(n))$ as $n \rightarrow \infty$. Justify your answer.

① $\sqrt{n+1} - \sqrt{n}$

Answer:

$$\begin{aligned}\sqrt{n+1} &= \sqrt{n\left(1+\frac{1}{n}\right)} = \sqrt{n} \cdot \sqrt{1+\frac{1}{n}} \\ &= \sqrt{n} \cdot \left(1 + \frac{1}{2n} - \frac{1}{8n^2} + \dots\right)\end{aligned}$$

(Taylor series)

$$\begin{aligned}\therefore \sqrt{n+1} - \sqrt{n} &= \frac{1}{2\sqrt{n}} - \left(\frac{1}{8n\sqrt{n}} - \frac{1}{16n^2\sqrt{n}} + \dots\right) \\ &\leq \frac{1}{2\sqrt{n}} \rightarrow \textcircled{1}\end{aligned}$$

$$\begin{aligned}\text{Also } \sqrt{n+1} - \sqrt{n} &\geq \frac{1}{2\sqrt{n}} - \frac{1}{8n^{1.5}} \quad \left(\because \text{the remaining terms are +ve}\right) \\ &= \frac{2\sqrt{n}(4n-1)}{16n^2} \\ &= \frac{(4n-1)}{8n\sqrt{n}} = \frac{1}{2\sqrt{n}} - \frac{1}{8n\sqrt{n}} \\ &\geq \frac{3}{8\sqrt{n}} \rightarrow \textcircled{2}\end{aligned}$$

① & ② $\Rightarrow \sqrt{n+1} - \sqrt{n} = \Theta\left(\frac{1}{\sqrt{n}}\right)$

(2) $\sum_{i=n+1}^{n^2} \frac{1}{i}$

Answer: $\sum_{i=n+1}^{n^2} \frac{1}{i} = H_{n^2} - H_n$ (H_n is the n -th harmonic number)

Recall, we showed in class that for $n \rightarrow \infty$

$$\ln n \leq H_n \leq \ln n + 1$$

Use this to get

$$\ln(n^2) - \ln(n) - 1 \leq H_{n^2} - H_n \leq \ln(n^2) + 1 - \ln n$$

$$\Rightarrow \ln(n) - 1 \leq H_{n^2} - H_n \leq \ln(n) + 1$$

$$\therefore H_{n^2} - H_n = \Theta(\ln(n))$$

$$\Rightarrow \sum_{i=n+1}^{n^2} \frac{1}{i} = \Theta(\ln(n))$$

Question 2:

Suppose there is an algorithm whose runtime depends on a flexible parameter t , and is $O(n^2/t) + O(t \log^2 t)$. What is the best t to

choose? What is the run-time of this algorithm for this best t ?

Solution

Sketch: In general for functions $f(n)$ & $g(n)$

$$\max(f, g) \leq f + g \leq 2 \cdot \max(f, g).$$

$$\therefore O(n \ln t) + O(t \log^2 t) = O\left(\max\left(\frac{n \ln}{t}, t \log^2 t\right)\right)$$

$\therefore$ Intuitively, the optimal value of t is such that (I am skipping writing O below)
 $\frac{n \ln}{t} = t \log^2 t$ (Actual need: $t^2 \log^2 t = O(n \ln)$)

$$\Rightarrow t \log t = \sqrt{n \ln} \rightarrow \text{find this } t!$$

Show that if $y = x \log x$ for $x \geq 1$, then $x \sim \frac{y}{\log y}$

$$\Rightarrow t = \frac{(n \ln)^{1/2}}{\log(n \ln)}, \text{ now use Stirling to get}$$

$$t \log t = O\left(\frac{n^{1/2}}{e^{1/2}} \cdot n^{1/4}\right)$$

$$\therefore \text{Runtime of algo for best } t = O\left(\frac{n^{1/2}}{e^{1/2}} \cdot n^{1/4}\right)$$

Question 3

Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be two independent random variables. Prove that for $a_1, a_2 \in \mathbb{R}$, $X = (a_1 \cdot X_1 + a_2 \cdot X_2) \sim \mathcal{N}(a_1 \mu_1 + a_2 \mu_2, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)$

Solution (Sketch): Remark: Note that it is not sufficient to just find mean & variance of X to prove this. You need to find the pdf of X or alternately you can show the moment generating function corresponding to X is that of $\mathcal{N}(a_1 \mu_1 + a_2 \mu_2, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)$

Recall, from class, we define moment generating of a random variable X as

$$\phi_X(t) = \mathbb{E}[e^{tX}]$$

Consider for $X = a_1 X_1 + a_2 X_2$

$$\begin{aligned}\phi_X(t) &= \mathbb{E}[e^{t(a_1 X_1 + a_2 X_2)}] \\ &= \mathbb{E}[e^{t(a_1 X_1)}] \mathbb{E}[e^{t(a_2 X_2)}] \quad (\text{using independence}) \\ &= \phi_{X_1}(a_1 t) \cdot \phi_{X_2}(a_2 t) \rightarrow \textcircled{1}\end{aligned}$$

Now for Gaussian R.V. $Z \sim \mathcal{N}(\mu, \sigma^2)$ show that

$$\phi_Z(t) = \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right)$$

Using this in $\textcircled{1}$ for $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ & $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ we get:

$$\begin{aligned}\phi_{X_1}(a_1 t) \cdot \phi_{X_2}(a_2 t) &= \exp\left(\frac{\sigma_1^2 a_1^2 t^2}{2} + \mu_1 a_1 t\right) \exp\left(\frac{\sigma_2^2 a_2^2 t^2}{2} + \mu_2 a_2 t\right) \\ &= \exp\left(\frac{(a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2) t^2}{2} + (a_1 \mu_1 + a_2 \mu_2) \cdot t\right)\end{aligned}$$

$$\therefore X \sim \mathcal{N}(a_1 \mu_1 + a_2 \mu_2, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)$$

($\because$ moment generating function uniquely determines the distribution)

Question 4:

Berry-Essen Theorem: Let $X_1, X_2, \dots, X_n$ be independent with $E[X_i] = 0 \ \forall i = 1, \dots, n$ and $\sum_{i=1}^n \sigma_i^2 = 1$, where $\sigma_i^2 = \text{Var}[X_i]$.

Let $S = X_1 + X_2 + \dots + X_n$ & let $Z \sim N(0,1)$. Then $\forall u \in \mathbb{R}$

$$|\Pr[S \leq u] - \Pr[Z \leq u]| \leq C \gamma$$

where $\gamma = \sum_{i=1}^n E[|X_i|^3]$ and C is universal constant.

Find an asymptotic upper bound for γ , showing that it is small!

① T : random variable that denotes # tails in n fair coin flips.

Find γ in Berry-Essen while estimating $\Pr[T \leq (n/2 + u \cdot \sqrt{n}/2)]$.

Solution (Sketch)

Take $X_i = \begin{cases} 1 & \text{if } i\text{th coin flip is tails} \\ 0 & \text{o.w.} \end{cases}$ (w.p. $1/2$)

$$\text{the } \sum_{i=1}^n X_i = T$$

Now, to use Berry-Essen we need to standardize X_i 's to get:

$$Y_i = \frac{(X_i - 1/2)2}{\sqrt{n}} = \frac{2X_i - 1}{\sqrt{n}}$$

& use Berry-Essen to estimate

$$\Pr[S \leq u] \text{ for } S = Y_1 + Y_2 + \dots + Y_n$$

$$\text{Also, } \Pr[T \leq (n/2 + u \cdot \sqrt{n}/2)] = \Pr[S \leq u]$$

$$\text{Now, show that } \gamma = \sum_{i=1}^n E[|X_i|^3] = O\left(\frac{1}{\sqrt{n}}\right)$$

② Consider Galton board experiment

$$X_i = \begin{cases} -1 & \text{w.p. } 1/3 \\ +1 & \text{w.p. } 2/3 \end{cases}$$

Find δ in Berry-Essen corresponding to

$$\Pr[S \leq n/3 + u \cdot \sqrt{8n/9}] \text{ for } S = \sum_{i=1}^n X_i$$

Solution (sketch):

Again to use Berry-Essen we need to standardize the X_i 's, to get:

$$Y_i = \frac{(X_i - 1/3)}{\sqrt{8/9}} \cdot \sqrt{\frac{9}{8}}$$

* use Berry-Essen to estimate

$$\Pr\left[\sum_{i=1}^n Y_i \leq u\right] \text{ which is } = \Pr\left[S \leq \frac{n}{3} + u \cdot \sqrt{\frac{8n}{9}}\right]$$

Now, show that

$$\gamma = \sum_{i=1}^n \mathbb{E}[|Y_i|^3] = O\left(\frac{1}{\sqrt{n}}\right).$$

Question 5

Order these functions in ascending order of growth.

1) $(n!)!$ 2) $n^{(\log n)!}$ 3) $((n-1)!)^n$ 4) $(\log n)^n$

Easy to see that $(\log n)^n$ is slowest and $(n!)!$ is fastest (since you know Stirling's)
for 2) & 3), the trick (as mentioned in class) is to take log, until you get more info.

$$\begin{aligned} \log(n^{(\log n)!}) &= ((\log n)! \cdot \log n) \\ (\text{Stirling's}) &= \Theta\left(\frac{(\log n)^{\log n}}{e^{\log n}} \cdot \sqrt{\log n} \cdot \log n\right) \end{aligned}$$

$$\begin{aligned} &\& \log((n-1)!^n) = n \cdot \log((n-1)!) \\ (\text{Stirling's}) &= \Theta\left(n \cdot ((n-1) \log(n-1) - (n-1) + \frac{1}{2} \log(n-1))\right) \end{aligned}$$

Simplify these, now you can visualize the limit as $n \rightarrow \infty$ of $2^{1/2}$.

Final order here:

$$(\log n)^n < n^{(\log n)!} < ((n-1)!^n) < (n!)!$$

