Въпрос 4. Тестват се за Ковид n души по следната прцедура: пробите се събират заедно и ако тестът е отрицателен, всички се декларират здрави, а ако е положителен, всеки се тества отново с индивидуален тет. В популация има 4 кръвни групи (КГ$_i$, $i = \overline{1{,}4}$) с равна представителност от $1/4$. Сред КГ$_1$ има $10\,\%$ заразени, сред КГ$_2$ има $1\,\%$ заразени, сред КГ$_3$ има $5\,\%$ заразени и сред КГ$_4$ има $4\,\%$ заразени. Всеки от $n-$те тествани индивида се допуска, че е с кръвна група, независима от тази на всички останали $n-1$ души и падаща се с вероятността на представителността на кръвната група в популацията.

Нека $N = \Phi \mod 3$, където Φ е последната цифра на вашия факултетен номер.

Цената на общия тест е 1 лев, а цената на всеки повторен, индивидуален тест е $1 + 0.1 \times N$.

- Съставете модел, който отразава очакваната цена $\rho(n)$ на един тестван по тази процедура човек. *(5 точки)*
- Третирайки n като непрекъсната променлива x, изведете уравнение за x (без да го решавате), което ви задава x^*, такова че $\rho(x^*) = \min\limits_{x>0} \left\{\rho(x)\right\}$. *(2 точки)*
- Ако имате време, намерете с помощта на компютър за кое n^* се получава минимална единична цена и нейната стойност *(3 бонус точки)*

Решение.

а) Въвеждаме събитието $\underset{infected}{I} = \{$случайно избран човек е заразен$\}$.

Вероятността то да се сбъдне, може да изчислим по формулата за пълната вероятност (събираме вероятностите на всички възможни сценарии, в които даден човек може да е заразен – съответно за всяка една от кръвните групи):

$$\mathbb{P}(I) = \sum_{i=1}^{4} \mathbb{P}(I \cap \text{КГ}_i) =$$

от формулата за условна вероятност имаме, че $\mathbb{P}(A\,|\,B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, в случая $A \equiv I$, КГ$_i \equiv B$

$$= \sum_{i=1}^{4} \mathbb{P}(I\,|\,\text{КГ}_i) \times \mathbb{P}(\text{КГ}_i) =$$

$\mathbb{P}(\text{КГ}_i) = \dfrac{1}{4}$, за $i = \overline{1{,}4}$ по условие

$$= \frac{1}{4}\sum_{i=1}^{4} \mathbb{P}(I\,|\,\text{КГ}_i) =$$

$$= \frac{1}{4}\left(0.1 + 0.01 + 0.05 + 0.04\right) = 0.05.$$

Трябва да изчислим средната цена за тест за един човек. За улеснение ще въведем случайната величина $X_n = \{$цената, която се заплаща при тестването на n души за Ковид$\}$.

$$X_n = \begin{cases} 1 \text{ лв.,} & \text{ако всичките } n \text{ не са заразени;} \\ 1 + (1 + 0.1 \times N) \times n \text{ лв.,} & \text{ако има поне един заразен от всички } n \text{ души в извадката} \\ & \text{(+1, т.к. вече сме изхабили един тест за цялата извадка)} \end{cases}$$

Ние вече пресметнахме каква е вероятността произволен човек да бъде заразен и знаем, че $\mathbb{P}(I) = 0.05$. Следователно, вероятността да не бъде заразен е равна на $\mathbb{P}(\bar{I}) = \mathbb{P}(I^c) = 1 - \mathbb{P}(I) = 0.95$, тъй като $I \cup I^c = \Omega$, $I \cap I^c = \emptyset$ (I и допълнението му I^c образуват пълна група от събития).

Следователно, за X_n имаме следната таблица на разпределение:

X_n	1	$1 + (1 + 0.1 \times N) \times n$
$\mathbb{P}$	$(1-p)^n$	$1 - (1-p)^n$

, където $p = \mathbb{P}(I) = 0.05$.

Ще се заплати 1 лв. за тестване на цялата извадка от n души, тогава и само тогава, когато **нито един** от тях **не е** бил заразен. Т.е. това се случва с вероятност $(1-p)^n$.

Ще се заплати $1 + (1 + 0.1 \times N) \times n$ лв. за тестване на цялата извадка от n души, тогава и само тогава, когато **поне един** от тях **е** бил заразен. Т.е. това се случва с вероятност $1 - (1-p)^n$.

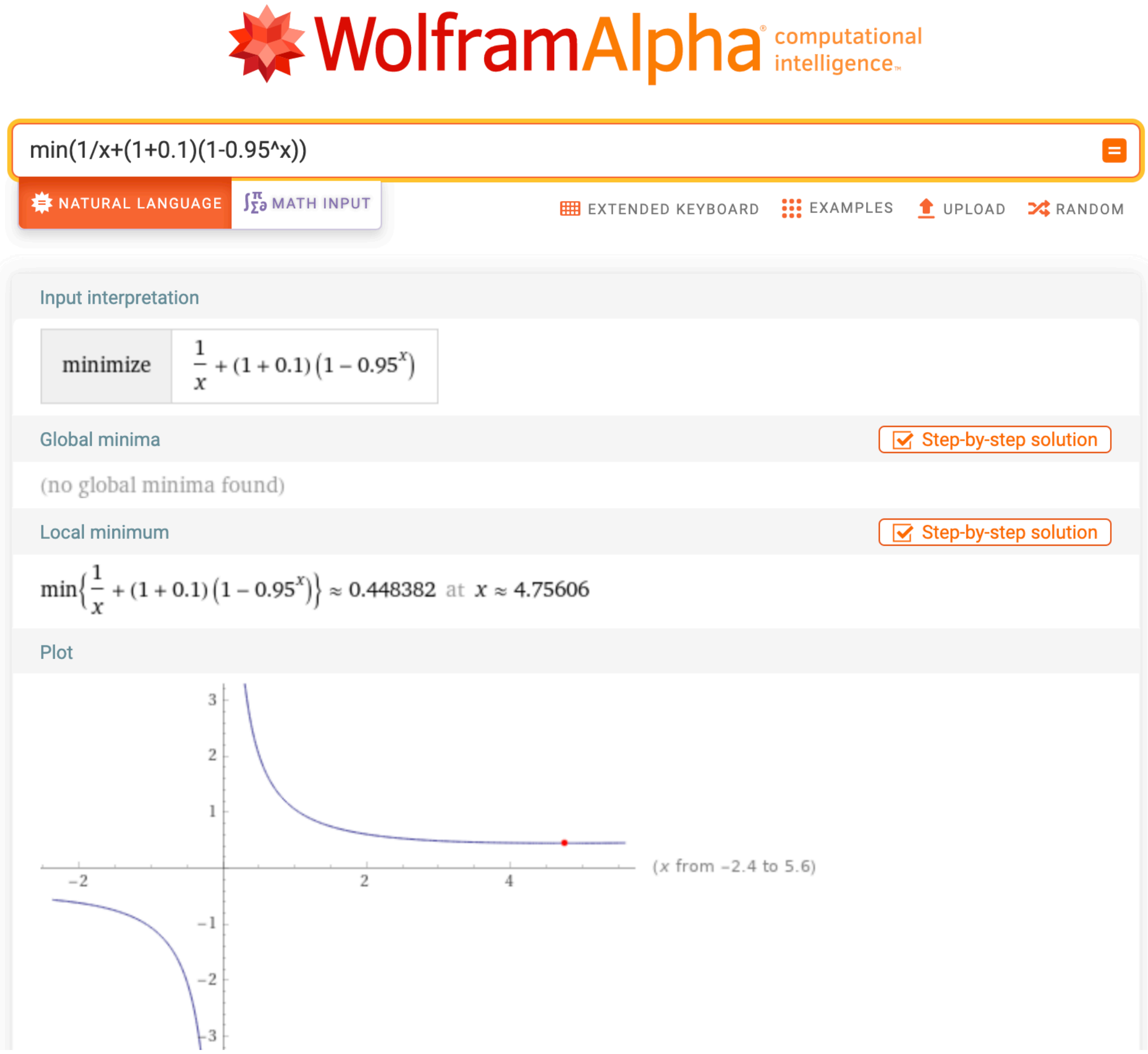
Следователно $\mathbb{E}X_n = 1 \times (1 - p^n) + \big(1 + (1 + 0.1 \times N) \times n\big) \times \big(1 - (1-p)^n\big)$.

Търсената средна цена за един човек е равна на $\dfrac{\mathbb{E}X_n}{n}$.

Окончателно,

$$\begin{aligned} \frac{\mathbb{E}X_n}{n} &= \frac{0.95^n + \big(1 + (1 + 0.1 \times N) \times n\big) \times \big(1 - 0.95^n\big)}{n} = \\ &= \frac{\cancel{0.95^n} + 1 - \cancel{0.95^n} + (1 + 0.1 \times N) \times n \times (1 - 0.95^n)}{n} \\ &= \frac{1}{n} + (1 + 0.1 \times N) \times (1 - 0.95^n) \end{aligned}$$

б) и в) Функцията, която получаваме за $\dfrac{\mathbb{E}X_n}{n}$ е $f(x) = \dfrac{1}{x} + (1 + 0.1 \times N) \times (1 - 0.95^x)$, която е трансцендентна функция и може да я минимизираме най-лесно чрез услугите на WilframAlpha (допускаме, че $N = 1$):



Следователно за извадка от $n \approx 4.7$ души (в контекста на $n = x > 0$ непрекъсната величина) ще имаме средно най-евтина цена на човек за тестване от Ковид. Но тъй като n в крайна сметка е дискретна величина, то може да пресметнем стойността на функцията за $n = 4$ и $n = 5$ и да проверим, за коя от стойностите ще имаме минимум.

Тъй като търсим минимума, то производната на уравнението от модела ще се нулира. Следователно $\rho(x^*) = f'(x) = 0$

$$0 = f'(x) = -\frac{1}{x^2} + 0 - (0.95^x)'(1 + 0.1 \times N) = -\frac{1}{x^2} - \frac{(\ln(20) - \ln(19)) \times 19^x}{20^x} \times 1.1$$

(Допуснали сме, че $N = 1$)

Забележка: Разгледай лекция номер 5.
Извод: При извадка от $n = 5$ човека и прилагане на посочената в условието стратегия за правене на тестове за вирус, ще излиза средно по $f(5) = 1/5 + 1.1 \times (1 - 0.95^5) \approx 0.44$ лв. цена на човек, което е грубо два пъти по-евтино от стандартната стратегия – всеки човек да се тества отделно.

Въпрос 6. Нека $\xi \in \mathscr{U}(0,\, 2)$, $\eta \in \mathscr{U}(0,\, 1)$.

- Намерете $\mathbb{E}[\xi]$, $\mathbb{E}[\eta]$, $\mathbb{E}[\xi^2]$, $\mathbb{E}[\eta^2]$. *(3 точки)*

Нека е дадено, че $\operatorname{cov}(\xi, \eta) = -\dfrac{1}{10}$.

(0) Намерете $D(\xi - 2\eta)$. *(4 точки)*
(1) Намерете $D(\xi - 3\eta)$. *(4 точки)*

Решение.

Нека за улеснение преименуваме $\xi \equiv X$ и $\eta \equiv Y$. По условие имаме, че $X \in \mathscr{U}(0,\, 2)$ и $Y \in \mathscr{U}(0,\, 1)$. Търси се $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\mathbb{E}[X^2]$, $\mathbb{E}[Y^2]$.

За равномерно разпределена случайна величина (виж лекция 10) знаем, че имаме плътност $f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{за } x \in (a,b) \\ 0, & \text{за } x \notin (a,b) \end{cases}$, където в нашия случай $a = 0$, $b = 2$ за X и $a = 0$, $b = 1$ за Y.

От дефиницията за очакването на случайна величина (което очакване се нарича още „първи централен момент“) знаем, че $\mathbb{E}[X] = \int_{-\infty}^{\infty} x \times f_X(x) \mathrm{d}\, x$, където $f_X(x)$ е плътността на случайната величина X.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \times f_X(x) \mathrm{d}\, x = \int_0^2 x \times \frac{1}{b-a} \mathrm{d}\, x = \int_0^2 x \times \frac{1}{2-0} \mathrm{d}\, x = \int_0^2 \frac{x}{2} \mathrm{d}\, x = \left.\frac{x^2}{4}\right|_0^2 = 1$$

Може да използваме и наготово формулата $\underset{X \in \mathscr{U}nif(a,b)}{\mathbb{E}[X]} = \dfrac{b-a}{2} = \dfrac{2-0}{2} = \mathbf{1}$, която се извежда с аналогични на по-горните сметки.

Аналогично, $\mathbb{E}[Y] = \dfrac{b-a}{2} = \dfrac{1-0}{2} = \mathbf{\dfrac{1}{2}}$.

$\mathbb{E}[X^2]$ и $\mathbb{E}[Y^2]$ са вторите централни моменти. За тях също има готова формула, която може да се изведе, но не е толкова известна: $\mathbb{E}[X^2] = \dfrac{b^3 - a^3}{3b - 3a}$. Нека все пак го пресметнем:

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \times f_X(x) \mathrm{d}\, x = \int_0^2 \frac{x^2}{b-a} \mathrm{d}\, x = \int_0^2 \frac{x^2}{2} \mathrm{d}\, x = \left.\frac{x^3}{6}\right|_0^2 = \mathbf{\frac{4}{3}}.$$

Аналогично за $\mathbb{E}[Y^2] = \int_{-\infty}^{\infty} y \times f_Y(y)\mathrm{d}\,y = \int_0^1 \frac{y^2}{1-0}\,\mathrm{d}\,y = \frac{y^3}{3}\Big|_0^1 = \mathbf{\frac{1}{3}}.$

За **(0)** и **(1)** е необходимо само да прилагаме дефиниции и да използваме факта, че очакването $\mathbb{E}$ е линеен функционал. Докато при дисперсията D, умножението по константа от нея излиза повдигната на квадрат. Например $\mathbb{E}[3X] = 3\mathbb{E}[X]$, а $D[3X] = 9D[X]$.

Нека първо си пресметнем DX и DY.

$D[X] \overset{\text{деф.}}{=} \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2 = \frac{4}{3} - 1 = \frac{1}{3}$ (виж лекция 5 – Дисперсия, стр. 5);

$D[Y] \overset{\text{деф.}}{=} \mathbb{E}[Y^2] - \left(\mathbb{E}[Y]\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$

(0)

$$
\begin{aligned}
D(X-2Y) &= \mathbb{E}\left[\left(X - 2Y - \mathbb{E}[X-2Y]\right)^2\right] = \\
&= \mathbb{E}\left[\left(X - \mathbb{E}[X] - 2(Y - \mathbb{E}[Y])\right)^2\right] = \\
&= \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2 + 4\left(Y - \mathbb{E}[Y]\right)^2 - 4\left(X - \mathbb{E}[X]\right)\left(Y - \mathbb{E}[Y]\right)\right] = \\
&= \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right] + 4\mathbb{E}\left[\left(Y - \mathbb{E}[Y]\right)^2\right] - 4\mathbb{E}\left[\left(X - \mathbb{E}[X]\right)\left(Y - \mathbb{E}[Y]\right)\right] = \\
&= D[X] + 4D[Y] - 4\,\mathrm{cov}(X,Y) = \\
&= \frac{1}{3} + \frac{4}{12} + \frac{4}{10} = \frac{20+20+24}{60} = \mathbf{\frac{16}{15}}.
\end{aligned}
$$

Аналогично,

$$
\begin{aligned}
D(X-3Y) &= \mathbb{E}\left[\left(X - 3Y - \mathbb{E}[X-3Y]\right)^2\right] = \\
&= \mathbb{E}\left[\left(X - \mathbb{E}[X] - 3(Y - \mathbb{E}[Y])\right)^2\right] = \\
&= \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2 + 9\left(Y - \mathbb{E}[Y]\right)^2 - 6\left(X - \mathbb{E}[X]\right)\left(Y - \mathbb{E}[Y]\right)\right] = \\
&= \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right] + 9\mathbb{E}\left[\left(Y - \mathbb{E}[Y]\right)^2\right] - 6\mathbb{E}\left[\left(X - \mathbb{E}[X]\right)\left(Y - \mathbb{E}[Y]\right)\right] = \\
&= D[X] + 9D[Y] - 6\,\mathrm{cov}(X,Y) = \\
&= \frac{1}{3} + \frac{9}{12} + \frac{6}{10} = \frac{20+45+36}{60} = \mathbf{\frac{101}{60}}.
\end{aligned}
$$

Въпрос 7. Работата на централа за конкретен ден зависи от отклонението от очакването на параметър, моделиран със случайна величина X с $\mathbb{E}[X] = 1$, $DX = 0.01$, т.е. от $Y = |X - 1|$. Поради съществуваща екологична опасност при големи отклонения от средното, регулаторен орган налага глоба от $g(n)$ лева, ако $Y \in (n, n+1], n \geq 2$.

- Намерете горна граница за стойностите на $f(a) = \mathbb{P}(Y > a)$. *(3 точки)*
- Според независими експерти за $g(n) = n^{3/2}$ е вярно, че $\lim_{n\to\infty} g(n)\mathbb{P}(Y[n, n+1)) = 0$ и глобите не са достатъчно ефективни. Вярно ли е тяхното твърдение за границата? *(4 точки)*
- Ако X е случайна величина, приемаща стойности в $\{0, 1, 2, 3, \dots\}$, докажете, че за всяко $\lambda > 0$

$$\mathbb{P}(X \leq n) \leq \mathbb{E}\left[e^{-\lambda X}\right] e^{\lambda n}.$$ *(4 бонус точки)*

Решение.

а) Директно следва от неравенството на Чебишев:

$$\mathbb{P}(Y > a) = \mathbb{P}(|X - 1| > a) = \underbrace{\mathbb{P}(|X - \mathbb{E}X| > a) \leq \frac{DX}{a^2}}_{\text{Чебишев}}$$

б) Ще обработим израза и ще приложим неравенството на Чебишев:

$$\begin{aligned}
\lim_{n\to\infty} g(n)\mathbb{P}(Y[n, n+1)) &= \lim_{n\to\infty} n^{3/2}\mathbb{P}(Y[n, n+1))) \leq \\
&\overset{(1)}{\leq} \lim_{n\to\infty} n^{3/2}\mathbb{P}(Y \geq n) = \\
&= \lim_{n\to\infty} n^{(3/2)}\mathbb{P}(|X - \underbrace{1}_{=\mathbb{E}X}| \geq n) \leq \\
&\overset{(2)}{\leq} \lim_{n\to\infty} n^{(3/2)} \times \frac{DX}{n^2} = \lim_{n\to\infty} n^{3/2} \times \frac{0.01}{n^2} = \\
&= \lim_{n\to\infty} \frac{0.01}{\sqrt{n}} = \mathbf{0}.
\end{aligned}$$

(1) Следва от факта, че вероятността $\mathbb{P}(Y \in [n, n+1))$ ще е винаги по-малка в сравнение с вероятността за Y, в която сме пренебрегнали горната граница на интервала $[n, n+1)$ и по този наичн сме го разширили. Тоест казваме, че е по-вероятно $Y \geq n$ отколкото $\underbrace{Y \geq n \text{ и } Y < n+1)}$ едновременно;

(2) Следва от директното прилагане на неравенството на Чебишев.

Следователно твърдението на независимите експерти е вярно и глобите не са достатъчно ефективни, тъй като с течение на времето ще стават (статистически доказано) все по-незначими от финансова гледна точка за централата. Т.е. от централата няма да имат стимул да се съобразяват с екологичната опасност.

в) За всяко $\lambda > 0$ имаме, че

$$
\begin{aligned}
\mathbb{P}(X \leq n) &= \mathbb{P}(-\lambda X \geq -\lambda n) = \\
&\overset{(1)}{=} \mathbb{P}(e^{-\lambda X} \geq e^{-\lambda n}) \leq \\
&\overset{(2)}{\leq} \frac{\mathbb{E}\left[e^{-\lambda X}\right]}{e^{-\lambda n}} = \\
&= \mathbb{E}\left[e^{-\lambda X}\right] e^{\lambda X}.
\end{aligned}
$$

(1) Следва от факта, че $e \approx 2.71828... > 1$;
(2) Следва от директното прилагане на неравенството на Марков.

Неравенство на Марков: (виж „Разширена версия за монотонно увеличаващи се функции“ и интуитивното доказателство на опростената версия)
https://bg.jejakjabar.com/wiki/Markov%27s_inequality

Въпрос 9. Зар с шест стени се хвърля 3×10^{12} пъти и се образува броя X на падналите се от тези хвърляния единици или тройки. **Нека $N = \Phi \mod 3$, където Φ е предпоследната цифра на вашия факултетен номер.** Слаб студент трябва да отговори на следните въпроси:

- Каква е вероятността (приблизително) за $X > 10^{12}$?
- Каква е вероятността (приблизително) за $X > 10^{12} + 10^{1+N}$?
- Каква е вероятността (приблизително) за $X > 10^{12} + 10^{7+N}$?

Студентът не знаел какво да прави и отговорил навсякъде 50 на 50 или $1/2$. На колко и на кои въпроси е отговорил правилно? Обосновете математически отговора си. *(7 точки)*

Решение.

Възможните елементарни изходи от хвърлянето на зара са $\{1, 2, 3, 4, 5, 6\}$. Според условието благоприятните изходи са $\{1, 3\}$, а останалите са неблагоприятни. Следователно, $\mathbb{P}(\text{успех}) = \frac{1}{3}$, $\mathbb{P}(\text{неуспех}) = \frac{2}{3}$.

Нека $X_i = \{\text{на } i\text{-тото хвърляне се пада } 1 \text{ или } 3\}$, $i = \overline{1,n}$.

X_i са бернулиево разпределени случайни величини с вероятност за успех $\frac{1}{3}$, т.е. $X_i \in Ber\left(\frac{1}{3}\right)$.

Тъй като $X = \{\text{броя успехи от бернулиеви случайни величини}\}$, то X е биномно разпределена случайна величина, т.е. $X \in Bin\left(n, \frac{1}{3}\right)$.

$X = \sum_{i=1}^{n} X_i$. Но ние знаем, че $\mu = \mathbb{E}X_1 = p = \frac{1}{3}$ и $\sigma^2 = DX_1 = pq = \frac{2}{9}$.

$X_i \sim X_1$, $i = \overline{1,n}$ са прототипи (еднакви експерименти).

Тъй като имаме много хвърляния на зара (много повтаряния на един и същ експеримент $3 \times 10^{12} \gg 30$), то от ЦГТ знаем, че

$$\frac{X - n\mu}{\sqrt{n\sigma^2}} \xrightarrow[n \to \infty]{d} \mathcal{N}(0, 1)$$

Т.е. $X \sim \sqrt{n\sigma^2} \times \mathcal{N}(0, 1) + n\mu$, но от свойствата на нормалното разпределение $\mathcal{N}$ знаем, че $X \sim \mathcal{N}(n\mu, n\sigma^2)$ (дисперсията влиза на квадрат, докато средното влиза линейно в параметрите на $\mathcal{N}$).

Следователно,

$$\mathbb{P}(X > a) = 1 - \mathbb{P}(X \leq a) - 1 - \Phi\left(\frac{a - n\mu}{\sqrt{n\sigma^2}}\right) =$$

$$= 1 - \Phi\left(\frac{a - 3 \times 10^{12} \times \frac{1}{3}}{\sqrt{3 \times 10^{12} \times \frac{2}{9}}}\right).$$

За $a = 10^{12}$, числителя ще нулира дробта и ще получим $1 - \Phi(0) = 1 - \frac{1}{2} = \frac{1}{2}$, тоест това ще е първия верен отговор от студента.

За $a = 10^{12} + 10^1$ (допуснали сме, че $N = 0$),
$1 - \Phi\left(\frac{10^1}{10^6}.\frac{\sqrt{3}}{\sqrt{2}}\right) \sim 1 - \Phi(0.0000122) \sim 1 - \Phi(0) = \frac{1}{2}$. Това ще е втория верен отговор даден от студента.

За $a = 10^{12} + 10^7$,

$$1 - \Phi\left(\frac{10^{12} + 10^7 - 10^{12}}{10^6} \times \frac{\sqrt{3}}{\sqrt{2}}\right) = 1 - \Phi\left(\frac{10\sqrt{3}}{\sqrt{2}}\right) \sim 1 - \Phi(12.26) \sim 1 - 0 = 1$$

Това означава, че последният отговор на студента ще е грешен.