Probability Theory 2
Proposed solution of moed aleph exam 2021

1. **(a)** If f is indeed a probability density function, then $\int_{-\infty}^{\infty} f(x)dx = 1$. Solving this equation yields

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_0^1 c(1-x)(1+x)dx = c\int_0^1 (1-x^2)dx = c(x - x^3/3)\big|_0^1 \\ &= c(1-1/3) - c(0-0) = 2c/3,\end{aligned}$$

implying that $c = 3/2$.

(b) Let F_X denote the cumulative distribution function of X. Observe that if $a \leq 0$, then $F_X(a) = \mathbb{P}(X \leq a) = 0$, and if $a \geq 1$, then $F_X(a) = \mathbb{P}(X \leq a) = 1$. Fix some $0 < a < 1$. Then

$$F_X(a) = \mathbb{P}(X \leq a) = \int_{-\infty}^{a} f(x)dx = 3/2 \cdot \int_0^a (1-x^2)dx = 3/2 \cdot (x - x^3/3)\big|_0^a = 3/2 \cdot (a - a^3/3).$$

We conclude that

$$F_X(a) = \begin{cases} 0 & \text{if } a \leq 0 \\ 3/2 \cdot (a - a^3/3) & \text{if } 0 < a < 1 \\ 1 & \text{if } a \geq 1 \end{cases}$$

(c) Recall that if X is a random variable with probability density function f, then $\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$. Hence

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 3/2 \cdot x(1-x)(1+x)dx = 3/2 \cdot \int_0^1 (x - x^3)dx = c(x^2/2 - x^4/4)\big|_0^1 \\ &= 3/2 \cdot [(1/2 - 1/4) - (0-0)] = 3/8.\end{aligned}$$

2. We will present a randomized algorithm and then prove that it meets all the requirements of the question.

Algorithm:

(i) For 100 times, draw an element of $\{1, \ldots, n\}$ independently and uniformly at random (with replacement); denote the set of all drawn numbers by J.

(ii) For every $i \in J$, compare x_i and y_i.

(iii) If $x_i = y_i$ for every $i \in J$, then output $\bar{x} = \bar{y}$. Otherwise, output $\bar{x} \neq \bar{y}$.

First, it is evident that the running time of the algorithm is constant as we make 100 random element draws and then compare 100 pairs of bits.

Next, we prove the (randomized) correctness of the algorithm. If $\bar{x} = \bar{y}$, then for any random subset $J \subseteq \{1, \ldots, n\}$, we will always have $x_i = y_i$ for every $i \in J$. Hence, in this case, with probability 1 the algorithm will output $\bar{x} = \bar{y}$. Assume then that $\bar{x} \neq \bar{y}$. Let $D = \{1 \leq i \leq n : x_i \neq y_i\}$; we claim that $|D| \geq n/2$. Indeed

$$\begin{aligned}
|D| &= |\{1 \leq i \leq n : x_i = 1 \text{ and } y_i = 0\}| + |\{1 \leq i \leq n : x_i = 0 \text{ and } y_i = 1\}\}| \\
&= |\{1 \leq i \leq n : x_i = 1\} \setminus \{1 \leq i \leq n : x_i = 1 \text{ and } y_i = 1\}| \\
&+ |\{1 \leq i \leq n : y_i = 1\} \setminus \{1 \leq i \leq n : x_i = 1 \text{ and } y_i = 1\}| \\
&\geq (n/2 - n/4) + (n/2 - n/4) = n/2,
\end{aligned}$$

where the inequality holds by the assumed properties of the family $\mathcal{F}$.

Hence, for each of the 100 random number draws, the probability of this number being in $\{1, \ldots, n\} \setminus D$ is at most $1/2$. It follows by independence that

$$\mathbb{P}(J \subseteq \{1, \ldots, n\} \setminus D) \leq (1/2)^{100} = 2^{-100}.$$

By the description of the algorithm we conclude that

$$\mathbb{P}(\text{the algorithm outputs } \bar{x} = \bar{y}) = \mathbb{P}(J \subseteq \{1, \ldots, n\} \setminus D) \leq 2^{-100}.$$

3. Let $G \sim G(n, \ln n/n)$ and let $r = \lceil n/100 \rceil$. Observe that if there is an edge in G between any two disjoint subsets of $V(G)$ of size r each, then there is an edge in G between any two disjoint subsets of $V(G)$ of size t for any $n/100 \leq t \leq n$. Let $\mathcal{E}$ denote the event "there exist disjoint sets $A, B \subseteq V(G)$ of size $|A| = |B| = r$ such that no edge of G has one endpoint in A and the other in B". Then, considering any pair of disjoint subsets of $V(G)$ of size r and assuming all potential edges connecting these sets are missing in G, we see that

$$\begin{aligned}
\mathbb{P}(\mathcal{E}) &\leq \binom{n}{r}\binom{n-r}{r}(1 - \ln n/n)^{r^2} \leq \binom{n}{r}^2 e^{-r^2 \ln n/n} \leq \left(\frac{en}{r}\right)^{2r} e^{-n \ln n/10000} \\
&\leq (101e)^{2n/100} e^{-n \ln n/10000} \leq e^{n - n \ln n/10000} \leq e^{-n \ln n/20000},
\end{aligned}$$

where the second inequality holds since $\binom{n-1}{i} \leq \binom{n}{i}$ for every $0 \leq i < n$ and since $1 - x \leq e^{-x}$ for every $x \in \mathbb{R}$, and the third inequality holds since $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$. We conclude that $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}) = 0$ as required.