Probability Theory 2
Proposed solution of concluding assignment 2019

1. Fix an arbitrary $n \in \mathbb{N}$.

(a) We aim to apply the following version of Chernoff's inequality which was introduced in Lecture 1.

Theorem 1 *Let $X_1, \ldots, X_n$ be independent random variables that return values in $[0,1]$, and let $X = \sum_{i=1}^{n} X_i$. Then, for every $t > 0$, it holds that*

$$Pr(X \geq \mathbb{E}(X) + t) \leq e^{-2t^2/n} \text{ and } Pr(X \leq \mathbb{E}(X) - t) \leq e^{-2t^2/n}.$$

We first observe that $X_i \sim U(\{-1, 0, 1\})$, whereas Theorem 1 is only applicable to random variables that return values in $[0,1]$. Therefore, we start by introducing, for every $1 \leq i \leq n$, the auxiliary random variable $Y_i := (X_i + 1)/2$. Then, the random variables $Y_1, \ldots, Y_n$ are mutually independent, and $Y_i \sim U(\{0, 1/2, 1\})$ for every $1 \leq i \leq n$. In particular, $\mathbb{E}(Y_i) = (0 + 1/2 + 1)/3 = 1/2$ holds for every $1 \leq i \leq n$. Let $Y = \sum_{i=1}^{n} Y_i$. It then follows by the linearity of expectation that $\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{E}(Y_i) = n/2$. Moreover

$$Y - \mathbb{E}(Y) = \sum_{i=1}^{n} Y_i - n/2 = \sum_{i=1}^{n} (X_i + 1)/2 - n/2 = S_n/2.$$

Therefore

$$\begin{aligned} Pr(|S_n| > 2\sqrt{n}) &= Pr(|Y - \mathbb{E}(Y)| > \sqrt{n}) \\ &\leq Pr(Y \geq \mathbb{E}(Y) + \sqrt{n}) + Pr(Y \leq \mathbb{E}(Y) - \sqrt{n}). \end{aligned} \tag{1}$$

Applying Theorem 1 then implies that

$$Pr(Y \geq \mathbb{E}(Y) + \sqrt{n}) \leq e^{-2(\sqrt{n})^2/n} = e^{-2} \tag{2}$$

and

$$Pr(Y \leq \mathbb{E}(Y) - \sqrt{n}) \leq e^{-2}. \tag{3}$$

Combining (1), (2) and (3), we conclude that $Pr(|S_n| > 2\sqrt{n}) \leq 2e^{-2}$.

(b) We will first use CLT to estimate the probability of the complementary event $-2\sqrt{n} \leq S_n \leq 2\sqrt{n}$. Note that, for every $1 \leq i \leq n$, it holds that $\mathbb{E}(X_i) = 0$ and that $Var(X_i) = \mathbb{E}\left(X_i^2\right) - (\mathbb{E}(X_i))^2 = 2/3$. Therefore

$$\begin{aligned} Pr(-2\sqrt{n} \leq S_n \leq 2\sqrt{n}) &= Pr(\frac{-2\sqrt{n}}{\sqrt{2/3} \cdot \sqrt{n}} < \frac{X}{\sqrt{2/3} \cdot \sqrt{n}} < \frac{2\sqrt{n}}{\sqrt{2/3} \cdot \sqrt{n}}) \\ &\approx \Phi(\sqrt{6}) - \Phi(-\sqrt{6}) = 2\Phi(\sqrt{6}) - 1 \approx 0.9858. \end{aligned}$$

We conclude that $Pr(|S_n| > 2\sqrt{n}) \leq 0.0142$.

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

Algorithm:

(i) Given $\varepsilon > 0$ and $\delta > 0$, let $t = \left\lceil \frac{\ln(2/\delta)}{2\varepsilon^2} \right\rceil$.

(ii) For every integer $1 \leq i \leq t$ choose an element of $\{a_1, \ldots, a_n\}$ uniformly at random, all choices being mutually independent.

(iii) Let $b_1, b_2, \ldots, b_t$ be the elements that were chosen at step (ii) of the algorithm. Output $\left(\sum_{j=1}^{t} b_j\right)/t$.

It is evident that the running time of the algorithm is a function of t, that is, it depends on ε and δ but not on n. It remains to prove that with sufficiently high probability, the output of the algorithm is not too far from $(a_1 + \ldots + a_n)/n$.

For every $1 \leq i \leq t$, let X_i be the random variable whose value is the ith number chosen at Step (ii) of the algorithm. Then $X_i \sim U(\{a_1, \ldots, a_n\})$ for every $1 \leq i \leq t$, and $X_1, \ldots, X_t$ are independent. Moreover, $\mathbb{E}(X_i) = (a_1 + \ldots + a_n)/n$ for every $1 \leq i \leq t$. Let $X = \sum_{i=1}^{t} X_i$. It then follows by the linearity of expectation that $\mathbb{E}(X) = \sum_{i=1}^{t} \mathbb{E}(X_i) = t(a_1 + \ldots + a_n)/n$. Applying Chernoff's bound as seen in Theorem 1 (recall that $0 \leq a_1, \ldots, a_n \leq 1$ by assumption), implies that

$$Pr\left(\left|\frac{X}{t} - \frac{a_1 + \ldots + a_n}{n}\right| \geq \varepsilon\right) = Pr\left(\left|X - \frac{t(a_1 + \ldots + a_n)}{n}\right| \geq \varepsilon t\right) = Pr\left(|X - \mathbb{E}(X)| \geq \varepsilon t\right)$$
$$\leq 2e^{-2\varepsilon^2 t^2/t} \leq \delta,$$

where the last inequality holds by the choice of t.

3. Let red, blue and green denote the three colours at our disposal. For every $1 \leq i \leq n$, colour i independently and uniformly at random, i.e.,

$$Pr(i \text{ is red}) = Pr(i \text{ is blue}) = Pr(i \text{ is green}) = 1/3.$$

Call a subset of $\{1, \ldots, n\}$ *monochromatic* if all its elements have the same colour. Since the colours of every $1 \leq i \leq n$ were chosen independently, for every $1 \leq j \leq m$ we have

$$Pr(A_j \text{ is monochromatic}) = 3 \cdot (1/3)^{|A_j|} = 3^{1-k}.$$

Hence, the probability that some A_j is monochromatic is at most

$$\sum_{j=1}^{m} Pr(A_j \text{ is monochromatic}) \leq m \cdot 3^{1-k} < 2^{k-1} \cdot 3^{1-k} < 1.$$

Therefore, the probability that our random colouring satisfies the requirements of the question is positive, i.e.,

$$Pr(A_j \text{ contains elements of at least two different colours for every } 1 \leq j \leq m) > 0.$$

We conclude that there exists a colouring which satisfies the requirements of the question.

4. **(a)** Since f_X is a density function, it follows that

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 (ax^2 + b)dx = (ax^3/3 + bx)\Big|_0^1 = a/3 + b. \tag{4}$$

Moreover

$$3/5 = \mathbb{E}(X) = \int_{-\infty}^{\infty} xf_X(x)dx = \int_0^1 (ax^3 + bx)dx = (ax^4/4 + bx^2/2)\Big|_0^1 = a/4 + b/2. \tag{5}$$

Solving equations (4) and (5) for a and b yields $a = 6/5$ and $b = 3/5$.

(b) For any $y \in \mathbb{R}$, the equation $4z^2 + 4yz + y + 2 = 0$ has two distinct real roots if and only if $(4y)^2 - 4 \cdot 4 \cdot (y + 2) > 0$. The latter inequality holds if and only if $0 < y^2 - y - 2 = (y - 2)(y + 1)$ which is clearly satisfied if and only if $y > 2$ or $y < -1$. Since Y is uniform on $(0, 5)$, we conclude that

$$Pr(4z^2 + 4zY + Y + 2 = 0 \text{ has two distinct real roots}) = Pr(Y > 2) = 3/5.$$

(c) Using the equality $\Phi(-x) = 1 - \Phi(x)$ which was proved in Lecture 12, we obtain

$$\begin{aligned} Pr(|Z| > a) &= Pr(Z > a) + Pr(Z < -a) = Pr(Z > a) + \Phi(-a) \\ &= Pr(Z > a) + (1 - \Phi(a)) = 2Pr(Z > a). \end{aligned}$$