Probability Theory 2 – Proposed solution of exam A

1. In the solution of this question we will use the central limit theorem which states the following:

   Theorem 1 *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite expectation μ and finite variance σ^2. For every positive integer n, let $Y_n = \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$ and let $F_n(a) = Pr(Y_n \leq a)$ be the cumulative probability function of Y_n. Let $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx$. Then*

   $$\lim_{n \to \infty} F_n(a) = \Phi(a)$$

   for every $a \in \mathbb{R}$.

   For every $1 \leq i \leq 360000$ let X_i be the indicator random variable for the event "the ith roll of the dice resulted in a 6", that is, $X_i = 1$ if the ith dice roll was a 6 and $X_i = 0$ otherwise. We then have $X = \sum_{i=1}^{360000} X_i$. Since the X_i's are indicator random variables, it follows that for every $1 \leq i \leq 360000$ we have

   $$\mathbb{E}(X_i) = Pr(X_i = 1) = 1/6$$

   and

   $$Var(X_i) = \mathbb{E}(X_i^2) - [\mathbb{E}(X_i)]^2 = Pr(X_i^2 = 1) - 1/36 = 1/6 - 1/36 = 5/36.$$

   Since the X_i's are independent and identically distributed random variables with finite expectation and variance, we can apply Theorem 1 to obtain

   $$\begin{aligned}
   Pr(54000 \leq X \leq 63000) &= Pr\left(\frac{54000 - 360000 \cdot 1/6}{\sqrt{5/36} \cdot \sqrt{360000}} \leq \frac{X - 360000 \cdot 1/6}{\sqrt{5/36} \cdot \sqrt{360000}} \leq \frac{63000 - 360000 \cdot 1/6}{\sqrt{5/36} \cdot \sqrt{360000}}\right) \\
   &= Pr\left(\frac{-6000}{100\sqrt{5}} \leq \frac{X - 360000 \cdot 1/6}{\sqrt{5/36} \cdot \sqrt{360000}} \leq \frac{3000}{100\sqrt{5}}\right) \\
   &= Pr\left(-12\sqrt{5} \leq \frac{X - 360000 \cdot 1/6}{\sqrt{5/36} \cdot \sqrt{360000}} \leq 6\sqrt{5}\right) \\
   &= Pr\left(\frac{X - 360000 \cdot 1/6}{\sqrt{5/36} \cdot \sqrt{360000}} \leq 6\sqrt{5}\right) - Pr\left(\frac{X - 360000 \cdot 1/6}{\sqrt{5/36} \cdot \sqrt{360000}} < -12\sqrt{5}\right) \\
   &\approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{6\sqrt{5}} e^{-x^2/2} dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-12\sqrt{5}} e^{-x^2/2} dx \\
   &= \frac{1}{\sqrt{2\pi}} \int_{-12\sqrt{5}}^{6\sqrt{5}} e^{-x^2/2} dx.
   \end{aligned}$$

   We conclude that $a = -12\sqrt{5}$ and $b = 6\sqrt{5}$ satisfy the requirements of the question.

2. The algorithm does the following:

   (a) For every $1 \leq i \leq 200$ draw an element x_i of A independently and uniformly at random with replacement.

   (b) Return $\max\{x_1, x_2, \ldots, x_{200}\}$.

   It is evident that the algorithm runs in constant time (as usual we assume that sampling one element from a set of size n takes constant time). It remains to prove that the algorithm outputs a correct answer with high probability.

   Suppose that the algorithm returns an incorrect answer, that is, it returns an element $x \in A$ for which there is a set $B_x \subseteq A$ of size $|B_x| \geq n/3$ such that $y > x$ for every $y \in B_x$. Since the algorithm returns $\max\{x_1, x_2, \ldots, x_{200}\}$, it follows that $x_i \in A \setminus B_x$ for every $1 \leq i \leq 200$. However, $Pr(x_i \in A \setminus B_x) \leq 2/3$ for every $1 \leq i \leq 200$. Moreover, since we sampled elements of A independently and uniformly at random with replacement, we have

$$Pr(x_i \in A \setminus B_x \text{ for every } 1 \leq i \leq 200) \leq (2/3)^{200} = (4/9)^{100} < 2^{-100}$$

   as required.

3. By the definition of the relative entropy we have

$$\begin{aligned} H(p|u) &= \sum_{i=1}^{n} p_i \log_2 \left(\frac{p_i}{1/n} \right) = \sum_{i=1}^{n} p_i \log_2(np_i) = \sum_{i=1}^{n} p_i [\log_2 n + \log_2 p_i] \\ &= \log_2 n \sum_{i=1}^{n} p_i + \sum_{i=1}^{n} p_i \log_2 p_i = \log_2 n - H(p_1, \ldots, p_n), \end{aligned}$$

   where in the last equality we used the fact that $\sum_{i=1}^{n} p_i = 1$. Rearranging we obtain $H(p_1, \ldots, p_n) = \log_2 n - H(p|u)$ as required.

4. Colour the elements of $\{1, \ldots, n\}$ independently and uniformly at random with the four colours c_1, c_2, c_3 and c_4, that is, for every $1 \leq i \leq n$ and every $1 \leq j \leq 4$

$$Pr(i \text{ is coloured with colour } c_j) = 1/4.$$

   Fix some $1 \leq i \leq m$. The probability that at most 3 of the 4 colours were used to colour the elements of A_i is at most

$$4 \cdot \frac{3^k}{4^k} = \frac{3^k}{4^{k-1}}$$

   where the 4 term is for the choice of one colour we are not allowed to use, the 3^k term is for all the possible ways of colouring the elements of A_i with the three colours we are allowed to use, and the 4^k term is for all the possible ways of colouring the elements of A_i without any restrictions.

Therefore, applying a union bound we see that the probability that there exists an index $1 \leq i \leq m$ such that at most three colours are used to colour the elements of A_i is at most

$$\sum_{i=1}^{m} Pr(\text{at most three colours are used to colour the elements of } A_i) \leq m \cdot \frac{3^k}{4^{k-1}}$$
$$< \frac{4^{k-1}}{3^k} \cdot \frac{3^k}{4^{k-1}} = 1.$$

We conclude that, with positive probability, our random colouring uses all four colours in every A_i and thus there exists a colouring with this property.