Exercise sheet 4

Exercises for the exercise session on 12 November 2020

Problem 4.1. Let $h = h(n)$ be a positive-valued function with $h(n) = \omega(1)$ (i.e. $h \to \infty$ as $n \to \infty$), but $h(n) = o(\ln n)$. Given

$$m = (\ln n - h) \cdot n$$

balls and n bins, place each ball into a bin chosen uniformly at random, independently for each ball. Prove that

$$\mathbb{P}[\exists \text{ empty bin}] \overset{n\to\infty}{\longrightarrow} 1.$$

Problem 4.2. Suppose that an urn contains one red ball and one blue ball. A ball is drawn from the urn uniformly at random. After that, the ball is put back into the urn and another ball of the same colour is added to the urn. This process is repeated n times. Denote by X_n the proportion of red balls in the urn after these n steps (i.e. number of red balls divided by total number of balls). Use Azuma's inequality to prove that

$$\mathbb{P}\left[\left|X_n - \frac{1}{2}\right| \geq \varepsilon\right] < 2\exp\left(-\frac{6\varepsilon^2}{2\pi^2 - 15}\right).$$

Problem 4.3. Let $S_1, \ldots, S_m$ be finite sets. Independently for each S_i, consider an arbitrary probability distribution $\mathbb{P}_i$. Let $\mathbb{P}$ be the probability distribution on

$$\Omega := \{(s_1, \ldots, s_m) \mid \forall 1 \leq i \leq m\colon s_i \in S_i\}$$

in which the coordinate s_i is chosen according to $\mathbb{P}_i$, independently from the other coordinates. Let $f\colon \Omega \to \mathbb{R}$ be a function. For every $\sigma = (s_1, \ldots, s_m) \in \Omega$, we choose $\tau = (t_1, \ldots, t_m) \in \Omega$ according to $\mathbb{P}$ and set, for $i = 0, \ldots, m$,

$$X_i(\sigma) := \mathbb{E}[f(\tau) \mid \forall 1 \leq j \leq i\colon s_j = t_j].$$

Then X_i is a random variable on Ω.

(a) Prove that $X_0, \ldots, X_m$ is a martingale. Deduce from this that in particular, the edge exposure martingale and the vertex exposure martingale are indeed martingales.

(b) Prove that if $|f(\sigma_1) - f(\sigma_2)| \leq 1$ holds for all $\sigma_1, \sigma_2 \in \Omega$ that differ in only one coordinate, then we have

$$|X_i(\sigma) - X_{i-1}(\sigma)| \leq 1$$

for every $\sigma \in \Omega$ and all $i = 1, \ldots, m$.

Problem 4.4. For an integer $n \geq 1$, let G be the graph with vertex set $V(G) = (\mathbb{Z}_7)^n$ and with $\{u, v\} \in E(G)$ if and only if u and v differ in only one coordinate. Suppose that $U \subset V(G)$ with $|U| = 7^{n-1}$ is given. For every $c > 0$, we define W_c to be the set of vertices of G with distance at least $(2 + c)\sqrt{n}$ from U. Show that

$$|W_c| < 7^n e^{-\frac{c^2}{2}}.$$

(*Hint.* Define a martingale $X_0, \dots, X_n$ for which X_0 is the average distance—taken over all vertices of G—from U and $X_n(v)$ is the distance of v from U. Apply Azuma's inequality twice: once to prove that X_0 is 'small' and once to deduce the desired upper bound for W_c.)

Problem 4.5. Let $0 < p < \frac{1}{2}$ be a constant and let $X_1, X_2, \dots$ be i.i.d. random variables with

$$X_i = \begin{cases} 1 & \text{with probability } p, \\ -1 & \text{with probability } 1 - p. \end{cases}$$

Set $Y_0 := 1$ and

$$Y_i := \left(\frac{1-p}{p}\right)^{X_1 + \cdots + X_i} \qquad \text{for } i \geq 1.$$

(a) Verify that $Y_0, \dots, Y_n$ is a martingale (for fixed n).

(b) Show that there exists a constant $0 < q < 1$ (depending on p) such that

$$\mathbb{P}\left[X_1 + \cdots + X_n \geq 0\right] < q^n$$

for all n and prove that this implies

$$\lim_{k \to \infty} \mathbb{P}\left[\exists n \geq k \colon Y_n \geq 1\right] = 0.$$

(*Hint.* The existence of q can either be proved via Chernoff bounds or with the help of an 'exposure' martingale.)

Problem 4.6. Let n be a positive integer and let $p = p(n) \in (0, 1)$. Let X be the sum of n i.i.d. random variables $Y_1, \dots, Y_n$, which are 1 with probability p and 0 with probability $1-p$. Define a martingale $X_0, \dots, X_n$ that satisfies $X_0 = \mathbb{E}[X]$ and $X_n = X$. Compare the bound that Azuma's inequality gives for

$$\mathbb{P}[X > \mathbb{E}[X] + t]$$

with the bounds from Chernoff's inequality (i.e. Chernoff 1 and Chernoff 2). Which one is better? Does the answer depend on the choice of $p(n)$ and t?