CS630 Homework 3

Qilin Ye

February 26, 2025

Solution to problem 1. As a starter we consider the expected number of fixed points in a permutation: given $\pi : \{1, \cdots, k\} \to \{1, \cdots, n\}$, we let $Z_i = \mathbf{1}[\pi(i) = i]$ and $Z = \sum_{i=1}^{k} Z_i$ to be the total number of fixed points. Immediately $\mathbb{E}Z_i = 1/k$ for each i, so $\mathbb{E}Z = k \cdot 1/k = 1$, proving that $\mathbb{E}[X_i \mid X_1, \cdots, X_{i-1}] = 1$.

Now, going back to martingales, for notational simplicity let $\mathcal{F}_i$ denote the filtration by $\{X_1, \cdots, X_i\}$, effective for the entire problem set whenever context is clear. Since $X_i - 1$ yields random variables with zero mean, we can define a process whose new increment has expected value 0 given the past:

$$M_i = \sum_{j=1}^{i} (X_j - 1).$$

Indeed, $\{M_i\}$ forms a maringale, since

$$\mathbb{E}[M_i \mid \mathcal{F}_{i-1}] = \sum_{j=1}^{i-1} (X_j - 1) + \mathbb{E}[X_i - 1 \mid \mathcal{F}_{i-1}] = M_{i-1}.$$

Now we define the stopping time T to be when all owners have received their keys, i.e., when $\sum_{i=1}^{T} X_i = n$. Correspondingly $M_T = n - T$. By the optional stopping theorem, $\mathbb{E}[M_T] = \mathbb{E}M_0 = 0$, which implies $\mathbb{E}T = n$. Thus, the expected number of rounds required for all owners to receive their keys is n.

Solution to problem 2. Let Y_t denote the number of red balls drawn from the first t draws and $\mathcal{F}_t$ the corresponding filtration. Define M_t to be the fraction of fraction of red balls remaining w.r.t. to all remaining balls in the bag after t draws. It follows that among the $(r+g)-t$ balls remaining, $r - Y_t$ are red, so $M_t = (r - Y_t)/((r+g)-t)$. Assuming $\{M_t\}$ forms a martingale, the martingale property implies $\mathbb{E}M_n = M_0 = r/(r+g)$, i.e.,

$$\frac{r - \mathbb{E}Y_n}{(r+g) - n} = \frac{r}{r+g} \implies \mathbb{E}Y_n = r - \frac{r(r+g-n)}{r+g} = \frac{nr}{r+g}.$$

Now, to conclude the proof, we show that $\{M_t\}$ is indeed a martingale. To do so, observe that

$$\mathbb{E}[r - Y_{t+1} \mid \mathcal{F}_t] = r - Y_t - M_t$$

since the number of red balls after $t+1$ draws depends on that after t draws, in conjunction with the outcome of the $(t+1)^{\text{th}}$ draw. By definition $r - Y_t = M_t((r+g) - t)$, so

$$\mathbb{E}[r - Y_{t+1} \mid \mathcal{F}_t] = M_g((r+g) - t) - M_t = M_t((r+g) - t - 1)$$

which, by dividing both sides by $(r + g - t - 1)$, gives the desired claim $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = M_t$.

For the last part, consider $N_t = \mathbb{E}(Y_n \mid \mathcal{F}_t)$, which tracks the evaluation of the expected final count of red balls as more draws become known. It follows immediately that $|N_k - N_{k-1}| \leqslant 1$ for all k. Since $N_0 = \mathbb{E}Y_n$ and $N_n - Y_n$,

$$\mathbb{P}(|N_n - N_0| \geqslant \epsilon) = \mathbb{P}(|Y_n - \mathbb{E}Y_n| \geqslant \epsilon) \leqslant 2\exp(-\epsilon^2/(2n)).$$

To get a high probability concentration we may set $\epsilon = \sqrt{2n \log n}$ so $\exp(-\epsilon^2/(2n)) = 1/n$. Now, the Y_n follow a hypergeomoetric distribution with variance

$$\sigma^2 = n \cdot \frac{r}{r+g} \cdot \frac{g}{r+g} \cdot \frac{r+g-n}{r+g-1}.$$

In order for our bound obtained from Azuma's inequality to be nontrivial, this variance must be bounded away from 0. In other words, the last fraction needs to be $\Theta(1)$. One assumption that ensures this is by imposing some $\delta > 0$ such that $n \leqslant (1-\delta)(r+g)$.

Solution to problem 3. For a vertex to remain isolated, none of the cn edges can be incident to it. By linearity of expectation, the expected number of isolated vertices equals n times the probability that any fixed v is isolated, so we pick a vertex v and analyze $\mathbb{P}(v \text{ is isolated})$, which can be easily written as

$$\mathbb{P}(v \text{ is isolated}) = \binom{\binom{n}{2} - (n-1)}{cn} \Big/ \binom{\binom{n}{2}}{cn}^{-1} = \prod_{i=0}^{cn-1} \left(1 - \frac{n-1}{n(n-1)/2 - i}\right). \tag{*}$$

Upper bounding (*) is simple:

$$(*) \leqslant \prod_{i=0}^{n-1} \left(1 - \frac{n-1}{n(n-1)/2}\right) = \left(1 - \frac{2}{n}\right)^{cn} \leqslant e^{-2c}.$$

To lower bound, we note that the denominators can be replaced by $n(n-1)/2 - cn$ since $i \leqslant cn - 1$:

$$(*) \geqslant \left(1 - \frac{n-1}{n(n-1)/2 - cn}\right)^{cn} = \left(1 - \frac{2}{n-2c}\right)^{cn}.$$

When $n \geqslant 4c$, we have $n/(n-2c) \leqslant 2$, and so

$$\left(1 - \frac{2}{n-2c}\right)^{cn} = \left(1 - \frac{2}{n-2c}\right)^{(n-2c)\cdot cn/(n-2c)} \geqslant e^{-2c\cdot n/(n-2c)} \geqslant e^{-c},$$

and so the total expected number of isolated vertices is at least ne^{-4c}. Combining both parts, we conclude with bound $[ne^{-4c}, ne^{-2c}]$.

The second part of this problem is a simple application of Azuma's inequality. Let X be the number of isolated vertices eventually. For each $i \leqslant cn$, let $\mathcal{F}_i$ be the σ-algebra generated by the first i edges chosen, and we consider the Doob martingale $M_i = \mathbb{E}[X \mid \mathcal{F}_i]$ for $0 \leqslant i \leqslant cn$. Clearly $M_0 = \mathbb{E}X$ and $M_{cn} = X$. When a new edge is added, the most dramatic effect it can cause is if it connects two previously isolated vertices. Therefore $|M_i - M_{i-1}| \leqslant 2$. Applying Azuma's inequality gives

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{2 \cdot cn \cdot 2^2}\right) = 2\exp\left(-\frac{t^2}{8cn}\right).$$

Replacing t with $2\lambda\sqrt{cn}$ we precisely recover

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant 2\lambda\sqrt{cn}) \leqslant 2\exp\left(-\frac{4\lambda^2 cn}{8cn}\right) = 2e^{-\lambda^2/2}.$$

For the sake of completeness, we show that $\{M_t\}$ forms a martingale, as tower property implies

$$\mathbb{E}[M_i \mid \mathcal{F}_{i-1}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_i] \mid \mathcal{F}_{i-1}] = \mathbb{E}[X \mid \mathcal{F}_{i-1}] = M_{i-1}.$$

Solution to problem 4. Let X_i be the individual process (bet), so $\mu = \mathbb{E}X_i = 0$ and $\sigma^2 = \operatorname{var}(X_i) = 1$. Write $S_n = \sum_{i=1}^n X_i$. Then $Z_n = S_n^2 - n\sigma^2$ is known to be the quadratic martingale:

$$\begin{aligned}
\mathbb{E}[S_{n+1}^2 - (n+1)\sigma^2 \mid \mathcal{F}_n] &= \mathbb{E}[S_n^2 + 2S_nX_{n+1} + X_{n+1}^2 - (n+1)\sigma^2 \mid \mathcal{F}_n] \\
&= \mathbb{E}[S_n^2 \mid \mathcal{F}_n] + \mathbb{E}[2S_nX_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[X_{n+1}^2 \mid \mathcal{F}_n] - (n+1)\sigma^2 \\
&= S_n^2 + 0 + \sigma^2 - (n+1)\sigma^2 = S_n^2 - n\sigma^2,
\end{aligned}$$

since

- $\mathbb{E}[S_n^2 \mid \mathcal{F}_n] = S_n^2$ due to total information; and

- $\mathbb{E}[X_{n+1}^2 \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1}^2] = 1$ also due to independence from $\mathcal{F}_n$.
- $\mathbb{E}[2S_nX_{n+1} \mid \mathcal{F}_n] = 2\mathbb{E}[S_n \mid \mathcal{F}_n]\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]$ due to independence of S_n, X_{n+1}, and that the last term $= \mathbb{E}X_{n+1} = 0$ since X_{n+1} is independent of $\mathcal{F}_n$;

Now let T be the stopping time when either of the absorbing barriers at l_1, l_2 is hit. The stopping time theorem says $\mathbb{E}[Z_T] = \mathbb{E}[Z_0] = 0$.
It is well-known that for a SSRW, the probability of the walk reaching l_1 before $-l_2$ is given by $p = l_2/(l_1 + l_2)$. Therefore, at stopping time T,

$$Z_T = S_T^2 - T = \begin{cases} l_1^2 - T & \text{w.p. } l_2/(l_1 + l_2) \\ l_2^2 - T & \text{w.p. } l_1/(l_1 + l_2). \end{cases}$$

Given $\mathbb{E}[Z_T] = 0$, we see

$$\mathbb{E}[Z_T] = \frac{l_2 \cdot l_1^2}{l_1 + l_2} + \frac{l_1 \cdot l_2^2}{l_1 + l_2} - \mathbb{E}T = 0 \implies \mathbb{E}[T] = \frac{l_1 l_2 (l_1 + l_2)}{l_1 + l_2} = l_1 l_2.$$

Solution to problem 5. For any vertex v in the n-cube and any radius $r > 0$, define the ball w.r.t. Hamming distance $d(x, y)$ to be

$$B_r(v) = \{x \in \{0,1\}^n : d(x, v) \leqslant r\}.$$

As any element in $B_r(v)$ must have hamming distance $\leqslant r$ to v, for $r \leqslant n/2$, the volume can be monotonically bounded by

$$\Delta_r := |B_r(v)| = \sum_{i=0}^{r} \binom{n}{i}.$$

It follows that the r-cover of S defined by $\bigcup_{v \in S} B_r(v)$ can contain at most $|S|\Delta_r$ elements. If we can bound this quantity by 2^{n-1}, then w.p. $\geqslant 1/2$, $D(x, S) > r$, and by Markov's inequality, $\mathbb{E}[D(x, S)] > r/2$.
By this post, if we define $H(p) = -\log_2 p - (1-p)\log_2(1-p)$ the binary entropy function, then $\Delta_r \leqslant 2^{nH(r/n)}$. This in conjunction with taking log of $|S|\Delta_r \leqslant 2^{n-1}$ gives $\log|S| + nH(r/n) \leqslant n - 1$.
When S is sufficiently small, the term involving $\log|S|$ is insignificant, and we need $r/n \approx 1/2$ as $H(1/2) = 1$ and $H(\cdot)$ monotonically increases on $[1/2, 1]$. In this case we obtain $\mathbb{E}[D(x, S)] \geqslant n/4$.
In the nontrivial case where $|S|$ is exponential in n, i.e. there exists $\alpha \in (0, 1)$ such that $|S| = 2^{\alpha n}$. In this case we want $\alpha n + nH(r/n) \leqslant n - 1$, which asymptotically reduces to $H(r/n) = 1 - \alpha$. Heuristically this correpsonds to $r \approx (n - \alpha n)/2 = (n - \log|S|)/2$, and so

$$\mathbb{E}[D(x, S)] > \frac{n - \log|S|}{4} = \Omega(n - \log|S|).$$

As for the second part, define

$$Y_i = \mathbb{E}[D(x, S) \mid \mathcal{F}_i]$$

where Y_i is the expected distance after we reveal the first i bits. Going from Y_i to Y_{i+1} we are only changing one bit, so the set Hamming distance changes by at most 1, i.e., $|Y_i - Y_{i-1}| \leqslant 1$. Azuma's inequality therefore implies

$$\mathbb{P}\left[|D(x, S) - \mathbb{E}[D(x, S)]| \geqslant \lambda\right] = \mathbb{P}\left[|Y_n - Y_0| \geqslant \lambda\right] \leqslant 2\exp(-\lambda^2/(2n)).$$