Assignment 3: Solution Sketch

Itay Hazan
Random Graphs, Technion, 2017

March 5, 2018

1 Generalized Bonferroni Inequalities

1.a. Let $S \in \binom{[n]}{r}$ be a set of r indices.

$$\begin{aligned}
\Pr\left[\bigwedge_{i\in S} E_i \wedge \bigwedge_{i\notin S} \neg E_i\right] &= \Pr\left[\bigwedge_{i\in S} E_i\right] \cdot \Pr\left[\bigwedge_{i\notin S} \neg E_i \,\middle|\, \bigwedge_{i\in S} E_i\right] \\
&= \Pr\left[\bigwedge_{i\in S} E_i\right] \cdot \sum_{t=0}^{n-r} (-1)^t \sum_{T\in\binom{[n]\setminus S}{t}} \Pr\left[\bigwedge_{j\in T} E_j \,\middle|\, \bigwedge_{i\in S} E_i\right] \\
&= \sum_{t=0}^{n-r} (-1)^t \sum_{T\in\binom{[n]\setminus S}{t}} \Pr\left[\bigwedge_{i\in S\cup T} E_i\right].
\end{aligned}$$

Observe that $\Pr[X=r] = \sum_{S\in\binom{[n]}{r}} \Pr\left[\bigwedge_{i\in S} E_i \wedge \bigwedge_{i\notin S} \neg E_i\right]$, and therefore:

$$\begin{aligned}
\Pr[X=r] &= \sum_{S\in\binom{[n]}{r}} \Pr\left[\bigwedge_{i\in S} E_i \wedge \bigwedge_{i\notin S} \neg E_i\right] \\
&= \sum_{S\in\binom{[n]}{r}} \sum_{t=0}^{n-r} (-1)^t \sum_{T\in\binom{[n]\setminus S}{t}} \Pr\left[\bigwedge_{i\in S\cup T} E_i\right] \\
&= \sum_{t=r}^{n} (-1)^{t-r} \binom{t}{r} \sum_{T\in\binom{[n]}{t}} \Pr\left[\bigwedge_{i\in T} E_i\right].
\end{aligned}$$

1.b. Following similar calculations to those in (1a) proves the following inequalities:

$$\Pr[X=r] \le \sum_{t=r}^{k} (-1)^{t-r} \binom{t}{r} \sum_{T\in\binom{[n]}{t}} \Pr\left[\bigwedge_{i\in T} E_i\right] \qquad \text{when } 2 \mid k-r,$$

$$\Pr[X=r] \ge \sum_{t=r}^{k} (-1)^{t-r} \binom{t}{r} \sum_{T\in\binom{[n]}{t}} \Pr\left[\bigwedge_{i\in T} E_i\right] \qquad \text{when } 2 \nmid k-r.$$

1.c. Let $p = \frac{\log n + c}{n}$. For every $S \subset [n]$, let E_S be the event that all vertices in S are isolated. It holds that

$$\begin{aligned}
\Pr[E_S] &= \left(1 - \frac{\log n + c}{n}\right)^{\binom{|S|}{2} + |S|(n-|S|)} \\
&= \left(1 - \frac{\log n + c}{n}\right)^{|S|n} \cdot \left(1 - \frac{\log n + c}{n}\right)^{-O\left(|S|^2\right)} \\
&= (1 \pm o(1)) e^{-|S|(\log n + c)} \cdot (1 - o(1)) \\
&= (1 \pm o(1)) \frac{e^{-|S|c}}{n^t}.
\end{aligned}$$

Using the equality in (1a), and since $\binom{m}{k} = (1 \pm o(1)) \frac{m^k}{k!}$, then:

$$\begin{aligned}
\Pr[X_n = r] &= \sum_{t=r}^{n} (-1)^{t-r} \binom{t}{r} \sum_{S\in\binom{[n]}{t}} (1 \pm o(1)) \frac{e^{-|S|c}}{n^t} \\
&= (1 \pm o(1)) \sum_{t=r}^{n} (-1)^{t-r} \binom{t}{r} \binom{n}{t} \frac{e^{-tc}}{n^t} \\
&= (1 \pm o(1)) \sum_{t=r}^{n} (-1)^{t-r} \frac{t!}{r!(t-r)!} \frac{n^t}{t!} \frac{e^{-tc}}{n^t} \\
&= (1 \pm o(1)) \sum_{t=r}^{n} (-1)^{t-r} \frac{e^{-tc}}{r!(t-r)!} \\
&= (1 \pm o(1)) \sum_{k=0}^{n} (-1)^{k} \frac{e^{-(k+r)c}}{r!k!} && \text{(renaming } k = t-r\text{)} \\
&= (1 \pm o(1)) \frac{e^{-rc}}{r!} \sum_{k=0}^{n} (-1)^{k} \frac{e^{-kc}}{k!} \\
&\xrightarrow{n\to\infty} \frac{e^{-rc}}{r!} e^{e^{-c}} = \Pr[P = r].
\end{aligned}$$

1.d. Let $\epsilon > 0$. Since $\sum_{r=0}^{\infty} \Pr[P = r] = 1$ is a series of nonnegative and monotonically decreasing elements, then there exists R for which $\sum_{r=R}^{\infty} \Pr[P = r] \le \frac{\epsilon}{4}$. Furthermore, since $\Pr[X_n = r] \overset{n\to\infty}{\Longrightarrow} \Pr[P = r]$, then there exists N such that $|\Pr[X_n = r] - \Pr[P = r]| \le \frac{\epsilon}{4R}$ for any $n \ge N$ and any $r \le R$. Therefore, for any $n \ge N$,

$$\begin{aligned}
\sum_{r=0}^{\infty} |\Pr[X_n = r] - \Pr[P = r]| &= \sum_{r=0}^{R-1} |\Pr[X_n = r] - \Pr[P = r]| + \\
&+ \sum_{r=R}^{\infty} |\Pr[X_n = r] - \Pr[P = r]| \\
&\le R \cdot \frac{\epsilon}{4R} + \sum_{r=R}^{\infty} \Pr[X_n = r] + \sum_{r=R}^{\infty} \Pr[P = r] \\
&\le \frac{\epsilon}{4} + \left(1 - \sum_{r=0}^{R-1} \Pr[X_n = r]\right) + \sum_{r=R}^{\infty} \Pr[P = r] \\
&\le \frac{\epsilon}{4} + 1 - \sum_{r=0}^{R-1} \left(\Pr[P = r] - \frac{\epsilon}{4R}\right) + \sum_{r=R}^{\infty} \Pr[P = r] \\
&\le \frac{\epsilon}{4} + \left(1 - \sum_{r=0}^{R-1} \Pr[P = r]\right) + R \cdot \frac{\epsilon}{4R} + \sum_{r=R}^{\infty} \Pr[P = r] \\
&\le \frac{\epsilon}{4} + \sum_{r=R}^{\infty} \Pr[P = r] + R \cdot \frac{\epsilon}{4R} + \sum_{r=R}^{\infty} \Pr[P = r] \\
&\le \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon.
\end{aligned}$$

2 Google Interview Question

2.a. For any $i \in n$, let $X_k^i(m)$ denote the number of balls in the ith bin after m steps. Clearly, $X_k^i(m) \sim \text{Bin}(m, {}^1\!/_n)$, and using the Poisson approximation for the Binomial distribution, we get that $X_k^i(m) \approx \text{Pois}({}^m\!/_n)$. Therefore, by linearity of expectationa we get that:

$$\mathbb{E}\left[X_k(m)\right] = \sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}_{\{X_k^i(m)=k\}}\right] \approx n \cdot e^{-m/n} \frac{\left({}^m\!/_n\right)^k}{k!}.$$

A more careful analysis, that follows the footsteps of the proof of the Poisson limit theorem, proves that $\mathbb{E}\left[X_k(m)\right] = n \cdot e^{-m/n} \frac{({}^m\!/_n)^k}{k!} + O(1)$, as required. For the variance of $X_k(m)$: for every i, the random variable $\mathbf{1}_{\{X_k^i(m)=k\}}$ is a Bernoulli indicator, and hence $\mathbb{V}\left[\mathbf{1}_{\{X_k^i(m)=k\}}\right] \le 1$. Furthermore, it is fairly easy to see that

$$\mathbb{C}\text{ov}\left(\mathbf{1}_{\{X_k^i(m)=k\}}, \mathbf{1}_{\{X_k^j(m)=k\}}\right) < 0.$$

Therefore:

$$\mathbb{V}\left[X_k(m)\right] = \sum_{i=1}^{n} \mathbb{V}\left[\mathbf{1}_{\{X_k^i(m)=k\}}\right] + 2\sum_{i<j} \mathbb{C}\mathrm{ov}\left(\mathbf{1}_{\{X_k^i(m)=k\}}, \mathbf{1}_{\{X_k^j(m)=k\}}\right) = O(n).$$

2.b. Let $\tilde{\delta}(n) = n^{2/3}$. Let $t = n^{1/6}$, and define a set of $t+1$ evenly-spread m_is as follows: $m_i = \epsilon n + \frac{(C-\epsilon)n}{t} i$ for every $0 \le i \le t$. For every such m_i, Chebyshev's inequality yields

$$\Pr[|X_k(m_i) - \mathbb{E}[X_k(m_i)]| \ge \tilde{\delta}] \le \frac{\mathbb{V}[X_k(m_i)]}{\tilde{\delta}^2} = \frac{O(n)}{n^{4/3}} = \frac{O(1)}{n^{1/3}}.$$

Now, applying a union bound, we get that

$$\Pr\left[\bigwedge_{i=0}^{t} |X_k(m_i) - \mathbb{E}[X_k(m_i)]| \ge \tilde{\delta}\right] \le (t+1)\frac{O(1)}{n^{1/3}},$$

and since $t = o\left(n^{1/3}\right)$ then this probability is $o(1)$; namely, with high probability, $|X_k(m_i) - \mathbb{E}[X_k(m_i)]| \le \tilde{\delta}$ for all our m_is. Now, let m be any natural number in the range $[\epsilon n, Cn]$. Let $i \in \{0, \ldots, t\}$ be the index for which $m_i \le m \le m_{i+1}$. We therefore get:

$$\begin{aligned} |X_k(m) - \mathbb{E}[X_k(m)]| &= |X_k(m) - \mathbb{E}[X_k(m)] + X_k(m_i) - \mathbb{E}[X_k(m_i)] - \\ &\quad - X_k(m_i) + \mathbb{E}[X_k(m_i)]| \\ &\le |X_k(m) - X_k(m_i)| + |\,\mathbb{E}[X_k(m)] - \mathbb{E}[X_k(m_i)]| + \\ &\quad + |X_k(m_i) - \mathbb{E}[X_k(m_i)]|. \end{aligned}$$

Since we throw a single ball in each step, then $|X_k(m) - X_k(m_i)| \le |m - m_i| \le 2\frac{(C-\epsilon)n}{t}$. Similarly, $|\,\mathbb{E}[X_k(m)] - \mathbb{E}[X_k(m_i)]| \le 2\frac{(C-\epsilon)n}{t}$. Taking $t = \omega(1)$ ensures that these bounds are $o(n)$. Furthermore, we know that $|X_k(m_i) - \mathbb{E}[X_k(m_i)]| \le \tilde{\delta}$ with high probability. Therefore, with high probability,

$$|X_k(m) - \mathbb{E}[X_k(m)]| \le 2\frac{(C-\epsilon)n}{t} + n^{2/3} = o(n).$$

Therefore, for $\delta(n) = 2\frac{(C-\epsilon)n}{t} + n^{2/3} = o(n)$, we get that with high probability $|X_k(m) - \mathbb{E}[X_k(m)]| \le \delta$ for all $m \in [\epsilon n, Cn]$.

2.c. We first remark that in this sub-exercise we are not given than $\epsilon n \le m^* \le Cn$ for some constant $\epsilon, C > 0$, and we thus cannot immediately apply (2a) and (2b). Nevertheless, it can be easily shown that $\frac{1}{2}n \le m^* \le 2n$ with high probability.

From (2a), we know that $\mathbb{E}[X_0(m)] = e^{-m/n}n + O(1)$ and $\mathbb{E}[X_1(m)] = e^{-m/n}\frac{m}{n}n + O(1) = e^{-m/n}m + O(1)$. Therefore,

$$\begin{aligned} |\mathbb{E}[X_1(m^*)] - \mathbb{E}[X_0(m^*)]| &\ge \left|e^{-m/n}m - e^{-m/n}n\right| - O(1) \\ &\ge e^{-m^*/n}|m^* - n| - O(1). \end{aligned}$$

On the other hand, we get that:

$$
\begin{aligned}
&|\mathbb{E}[X_1(m^*)] - \mathbb{E}[X_0(m^*)]| \leq \\
&\qquad |\mathbb{E}[X_1(m^*)] - X_1(m^*) - \mathbb{E}[X_0(m^*)] + X_0(m^*) + X_1(m^*) - X_0(m^*)| \leq \\
&\qquad |\mathbb{E}[X_1(m^*)] - X_1(m^*)| + |\mathbb{E}[X_0(m^*)] - X_0(m^*)| + |X_1(m^*) - X_0(m^*)| \\
&\qquad\qquad \leq \delta(n) + \delta(n) + O(1),
\end{aligned}
$$

where the last inequality holds due to (2b) and since $X_1(m^*) = X_0(m^*) + O(1)$. We can thus conclude the following:

$$
e^{-m^*/n}|m^* - n| - O(1) \leq |\mathbb{E}[X_1(m^*)] - \mathbb{E}[X_0(m^*)]| \leq \delta(n) + \delta(n) + O(1),
$$

and namely $|m^* - n| \leq 2e^{m^*/n}\delta(n) + O(1) \cdot e^{m^*/n}$ whp, as desired. Taking $\beta = e^{-m^*/n}$ concludes the exercise:

$$
\begin{aligned}
|X_0(m^*) - \beta n| &\leq |X_0(m^*) - \mathbb{E}[X_0(m^*)]| + |\,\mathbb{E}[X_0(m^*)] - \beta n| \leq \delta(n) + O(1), \\
|X_1(m^*) - \beta n| &\leq |X_1(m^*) - \mathbb{E}[X_1(m^*)]| + |\,\mathbb{E}[X_1(m^*)] - \beta n| \\
&\leq \delta(n) + \left|\mathbb{E}[X_1(m^*)] - \beta\left(m^* + 2e^{m^*/n}\delta(n) + O(1) \cdot e^{m^*/n}\right)\right| \\
&\leq \delta(n) + |\,\mathbb{E}[X_1(m^*)] - \beta m^*| + 2\beta e^{m^*/n}\delta(n) + O(1) \cdot \beta e^{m^*/n} \\
&\leq \delta(n) + \delta(n) + 2\delta(n) + O(1) = 4\delta(n) + O(1).
\end{aligned}
$$

2.d. We consider an equivalent game, in which the pills are uniquely labeled $\{1, \ldots, n\}$, and whenever we consume the second half of a pill, we insert a new "dummy pill" with the same label into the bag. This way, the total number of elements (either pills or half-pills) in the bag is always n, and there is always an element in the bag with each label in $[n]$. Suppose we perform t steps of the new experiment, and every time we consume a pill whose label is i, we throw a ball in the ith bin. Let $\tilde{H}(\tilde{m}), \tilde{W}(\tilde{m})$ denote the number of half/whole pills in the game we propose (excluding dummy pills), and let $\tilde{m}^*$ denote the first step at which $\tilde{H}(\tilde{m}^*) \geq \tilde{W}(\tilde{m}^*)$. Clearly, our mapping satisfies $\tilde{H}(\tilde{m}^*) = H(m^*)$ and $\tilde{W}(\tilde{m}^*) = W(m^*)$.

Due to the equivalence described above between the proposed game and the balls-in-bins game, we get that whp $|W(m^*) - \beta n| \leq \eta(n)$ and $|H(m^*) - \beta n| \leq \eta(n)$ for some β, η as described in (2c). Furthermore, since $m^* = 2n - 2W(m^*) - H(m^*)$, then with high probability $m^* = 2n - (3\beta n \pm 3\eta(n))$, namely $|m^* - (2 - 3\beta)n| \leq 3\eta(n)$ with high probability, and that concludes the exercise.