Homework 5

Out: 17 Feb. Due: 24 Feb.

Submit your solutions in pdf format on Gradescope by **5pm on Friday, February 24***. Solutions may be written either in LaTeX (with either machine-drawn or hand-drawn diagrams) or* **legibly** *by hand. (The LaTeX source for this homework is provided in case you want to use it as a template.) Please be sure to begin the solution for each problem on a new page, and to tag each of your solutions to the correct problem! Per course policy, no late solutions will be accepted. Take time to write* **clear** *and* **concise** *answers; confused and long-winded solutions may be penalized. You are encouraged to form small groups (two to four people) to work through the homework, but you* **must** *write up all your solutions on your own. Depending on grading resources, we reserve the right to grade a random subset of the problems and check off the rest; so you are advised to attempt all the problems.*

1. (a) In a company of n people, what is the probability that exactly k of them have a birthday on Christmas Day? (Your answer should contain a binomial coefficient, and should be given as a function of n and k. Ignore the detail of leap years.)

(b) Now suppose $n = 400$. Compute the above probabilities (accurate to four decimal places) for $k = 0, 1, 2, 3$.

(c) Now use the Poisson approximation to estimate the probabilities in part (b), again to four decimal places. How good is the approximation?

2. Suppose m balls are thrown at random into n bins. In class we proved that, for any non-negative function f of the bin loads,

$$\mathrm{E}[f(X_1, \ldots, X_n)] \leq \mathrm{e}\sqrt{m}\mathrm{E}[f(Y_1, \ldots, Y_n)],$$

where the X_i are the actual bin loads and the Y_i are independent Poisson loads each with parameter m/n. (See also MU, Theorem 5.7.)

Suppose now that the function f is *monotonically increasing* in the sense that $\mathrm{E}[f(X_1, \ldots, X_n)]$ does not decrease when the number of balls m is increased. We will improve the above statement to

$$\mathrm{E}[f(X_1, \ldots, X_n)] \leq 2\mathrm{E}[f(Y_1, \ldots, Y_n)]. \qquad (*)$$

(By a symmetrical argument, the same holds when f is monotonically decreasing.)

(a) When f is monotonically increasing, show that

$$\mathrm{E}[f(Y_1, \ldots, Y_n)] \geq \mathrm{E}[f(X_1, \ldots, X_n)] \Pr[\textstyle\sum_i Y_i \geq m].$$

[NOTE: Recall that the original bound above relied on the weaker fact that $\mathrm{E}[f(Y_1, \ldots, Y_n)] \geq \mathrm{E}[f(X_1, \ldots, X_n)] \Pr[\sum_i Y_i = m]$.]

(b) For any Poisson r.v. Z with parameter $\lambda \geq 1$, show that $\Pr[Z \geq \lambda] \geq 1/2$. You may assume for simplicity that λ is an integer. [HINT: Show that $\Pr[Z = \lambda + i] \geq \Pr[Z = \lambda - i - 1]$ for each $0 \leq i \leq \lambda - 1$.]

(c) Deduce the desired bound $(*)$ from parts (a) and (b).

[Turn over for problems 3 & 4]

3. Suppose $m = n \ln n + cn$ balls are thrown into n bins, where c is a constant (and may be positive or negative). Let $\mathcal{E}$ denote the event that no bin is empty.

(a) Show that, under the Poisson approximation, the probability of event $\mathcal{E}$ is $\left(1 - \frac{\mathrm{e}^{-c}}{n}\right)^n \sim \mathrm{e}^{-\mathrm{e}^{-c}}$.

(b) Deduce that, when $c < 0$, the actual probability of event $\mathcal{E}$ is asymptotically at most $2\mathrm{e}^{-\mathrm{e}^{-c}}$, and when $c > 0$ the probability of event $\overline{\mathcal{E}}$ is asymptotically at most $2(1 - \mathrm{e}^{-\mathrm{e}^{-c}})$. [HINT: Use the result of the previous problem.]

(c) What does part (b) imply about the coupon collecting problem? Be as precise as you can.

4. Mitzenmacher & Upfal, Ex. 4.26. [NOTES: In part (a), show a lower bound of $\max\{c, d\})$, which is $\Omega(c+d)$. In part (b), you should assume that α is chosen sufficiently large (as in part (c)). In part (d), your algorithm should succeed with probability $1 - O(1/N)$; also, the "length" of a schedule is just the total time taken to route all N packets. For this part, carefully explain how to replace each time step in the unconstrained schedule by $O(\log(Nd))$ time steps in the real schedule.]