Homework 5 Solutions

Note: These solutions are not necessarily model answers. Rather, they are designed to be tutorial in nature, and sometimes contain a little more explanation than an ideal solution. Also, bear in mind that there may be more than one correct solution. The maximum total number of points available is 42.

1. (a) $\binom{n}{k}(\frac{1}{365})^k(\frac{364}{365})^{n-k}$. *2pts*

(b) The probabilities for $k = 0, 1, 2, 3$ are respectively $0.3337, 0.3667, 0.2010, 0.0733$ *2pts*

(c) We use a Poisson approximation with parameter $\lambda = 400/365$. The probabilities for $k = 0, 1, 2, 3$ are respectively $0.3342, 0.3663, 0.2007, 0.0733$. The approximation is accurate up to 3 decimal places. *3pts*

2. (a) Using $X_i^{(k)}$ to denote the distribution of balls in bins when k balls are thrown, as in MU (and omitting the superscript when $k = m$), we have *5pts*

$$\begin{aligned}
\mathrm{E}\Big[f(Y_1,\ldots,Y_n)\Big] &= \sum_{k=0}^{\infty}\mathrm{E}\Big[f(Y_1,\ldots,Y_n)\mid\sum_{i=1}^{n}Y_i=k\Big]\Pr\Big[\sum_{i=1}^{n}Y_i=k\Big]\\
&\geq \sum_{k=m}^{\infty}\mathrm{E}\Big[f(Y_1,\ldots,Y_n)\mid\sum_{i=1}^{n}Y_i=k\Big]\Pr\Big[\sum_{i=1}^{n}Y_i=k\Big] \qquad (f \text{ is non-negative})\\
&= \sum_{k=m}^{\infty}\mathrm{E}\Big[f(X_1^{(k)},\ldots,X_n^{(k)})\Big]\Pr\Big[\sum_{i=1}^{n}Y_i=k\Big]\\
&\qquad\qquad (\text{joint distribution of the } Y_i \text{ given } \textstyle\sum_{i=1}^{n}Y_i=k \text{ is exactly that of the } X_i^{(k)})\\
&\geq \mathrm{E}\Big[f(X_1,\ldots,X_n)\Big]\sum_{k=m}^{\infty}\Pr\Big[\sum_{i=1}^{n}Y_i=k\Big] \qquad (f \text{ is monotonically increasing})\\
&= \mathrm{E}\Big[f(X_1,\ldots,X_n)\Big]\Pr\Big[\sum_{i=1}^{n}Y_i\geq m\Big].
\end{aligned}$$

(b) First, we show that *4pts*

$$\Pr[Z=\lambda+i]\geq\Pr[Z=\lambda-i-1], \quad 0\leq i\leq\lambda-1 \qquad (**)$$

This is equivalent to showing $\frac{\lambda^{\lambda+i}}{(\lambda+i)!}\geq\frac{\lambda^{\lambda-i-1}}{(\lambda-i-1)!}$, or $\lambda^{2i+1}\geq\frac{(\lambda+i)!}{(\lambda-i-1)!}$. Now observe that the right-hand side can be written as $(\lambda+i)(\lambda+i-1)\ldots(\lambda-i+1)(\lambda-i)=\lambda\cdot\prod_{j=1}^{i}(\lambda+j)(\lambda-j)=\lambda\cdot\prod_{j=1}^{i}(\lambda^2-j^2)$. And this is less than or equal to λ^{2i+1}, as required.
Next, summing $(**)$ over $i=0,\ldots,\lambda-1$, we have

$$\Pr[\lambda\leq Z\leq 2\lambda-1]\geq\Pr[Z\leq\lambda-1].$$

The term on the left is bounded from above by $\Pr[Z\geq\lambda]$, so $\Pr[Z\geq\lambda]\geq 1-\Pr[Z\geq\lambda]$ and thus $\Pr[Z\geq\lambda]\geq 1/2$.

(c) $\sum_{i=1}^{n}Y_i$ has the Poisson distribution with parameter m. Hence, $\Pr[\sum_{i=1}^{n}Y_i\geq m]\geq 1/2$ by part (b), and upon substitution into the result of part (a), we obtain the desired bound. *1pt*

3. (a) Let X_i denote the r.v. for the number of balls in bin i and let $Y_1,\ldots,Y_n$ be independent Poisson r.v.'s each with parameter $m/n=\ln n+c$. Then, we may approximate the joint distribution of the X_i by that of the Y_i. In this Poisson approximation, we have $\Pr_{\text{Poisson}}[\mathcal{E}]=\Pr_Y[Y_i>0,\forall i]=\prod_{i=1}^{n}(1-\Pr[Y_i=0])=(1-e^{-m/n})^n=(1-\frac{e^{-c}}{n})^n$. *3pts*

(b) Clearly, the probability of the event $\mathcal{E}$ ("no bin is empty") is monotonically increasing with the number of balls m. Hence, by inequality $(*)$ from the previous problem, $\Pr[\mathcal{E}] \leq 2\Pr_{\text{Poisson}}[Y_i > 0, \forall i] = 2(1-\frac{e^{-c}}{n})^n \leq 2e^{-e^{-c}}$. *4pts*

Similarly, the probability of the event $\overline{\mathcal{E}}$ ("some bin is empty") is monotonically decreasing, so by inequality $(*)$ from the previous problem (applied to decreasing events) we have $\Pr[\mathcal{E}] \leq 2\Pr_{\text{Poisson}}[Y_i = 0 \text{ for some } i] = 2(1-\Pr_{\text{Poisson}}[Y_i > 0, \forall i]) = 2(1-(1-\frac{e^{-c}}{n})^n) \sim 2(1-e^{-e^{-c}})$.

[NOTE: Both of the above statements hold for all c. However, the first one is useful only when $c<0$ and the second one is useful only when $c>0$ (because otherwise the probability bounds are greater than 1 and thus vacuous). This follows from the fact that $e^{-e^{-c}}$ tends to 0 as $c \to -\infty$, and tends to 1 as $c \to \infty$.]

(c) Part (b) implies that the number of cereal boxes we need to purchase in order to get a copy of all n coupons is tightly concentrated around the value $n\ln n$. To see this, note first that coupon collecting is equivalent to tossing balls into bins, where each bin corresponds to a coupon and each ball to a cereal box. Suppose first that we purchase $m = n\ln n + cn$ boxes with $c<0$, i.e., a small amount *less than* $n\ln n$ (note that cn is a lower order term). Then the first part of part (b) says that the probability we have all the coupons is asymptotically at most $2e^{-e^{-c}}$; and this value tends to zero as $c \to -\infty$. So just a little bit less than $n\ln n$ boxes is *not* enough with high probability. On the other hand, suppose that we purchase $m = n\ln n + cn$ boxes with $c>0$, i.e., a small amount *more than* $n\ln n$. Then the second part of part (b) says that the probability we fail to get all the coupons is asymptotically at most $2(1-e^{-e^{-c}})$; and this value tends to zero as $c \to \infty$. So just a little bit more than $n\ln n$ boxes is enough with high probability. *3pts*

This is an example of a *sharp threshold*: the behavior of the process is determined by *lower order* terms (on the order of n) around the threshold value $n\ln n$. So, if we measure boxes purchased on the $n\ln n$ scale, there is a sudden transition from all coupons collected with probability close to 0 to all coupons collected with probability close to 1 at the point $1\times n\ln n$.

A number of students had trouble making this connection: you are encouraged to carefully review the above argument.

4. (a) Fix any schedule, and suppose the schedule has length T. By definition of dilation, there exists a packet that travels a distance d, and it takes at least d time steps to travel a distance d, so $T \geq d$. Let e be the edge with congestion c. Since at each time step at most one packet can pass through e, it must take c time steps for all c packets passing through e to go through, so $T \geq c$. Therefore, $T \geq \max\{c,d\} = \Omega(c+d)$ and this holds for every schedule. *2pts*

(b) Fix a time step t and an edge e. At most c packets use the edge e at some time, and we may assume WLOG that exactly c packages use the edge e at some time (since this is the worst case). Let X be the r.v. for the number of packets traversing e at time t. We write $X = \sum_i X_i$, where the indicator r.v. X_i is 1 if packet i passes through e at time t, and 0 otherwise. Clearly, $\mathrm{E}[X_i] = \frac{\log(Nd)}{\alpha c}$, so $\mathrm{E}[X] = \frac{\log(Nd)}{\alpha}$. Also, the X_i are independent because the packet delays are independent. So we may apply the Chernoff bound in the form $\Pr[X \geq (1+\delta)\mu] \leq \exp(-\frac{\delta^2}{2+\delta}\mu)$, with $\mu = \frac{\log(Nd)}{\alpha}$ and $\delta = b\alpha - 1$ to get *5pts*

$$\Pr\Big[X \geq b\log(Nd)\Big] \leq \exp\Big(-\frac{(b\alpha-1)^2}{\alpha(b\alpha+1)}\log(Nd)\Big).$$

(Here b is a constant that we can choose.) Now if we set (for example) $b=5$ and $\alpha=2$ the exponent in the above bound becomes $\frac{81}{22}\log(Nd) \geq 3\log(Nd)$. Thus we have

$$\Pr\Big[X \geq 5\log(Nd)\Big] \leq \exp\Big(-3\log(Nd)\Big) = \frac{1}{(Nd)^3}.$$

Hence, the probability that more than $O(\log(Nd))$ packages use e at time step t is at most $\frac{1}{(Nd)^3}$. [Note: We chose $\frac{1}{(Nd)^3}$ here for use in the next part. More generally, we can achieve an upper bound of any poly$(\frac{1}{Nd})$ by replacing b and α with correspondingly larger constants.]

(c) We need to take a union bound over all edges e that are used and over all time steps t. To do this, we need upper bounds on both the number of edges and the number of time steps: *5pts*

- Each packet uses at most d distinct edges, so the total number of edges used is at most Nd.
- The total number of time steps is at most $d + \frac{\alpha c}{\log(Nd)}$. The congestion c is bounded by N, so this total number of time steps is at most $d + \frac{\alpha N}{\log(Nd)} \leq d + N \leq Nd$, for sufficiently large N. (Recall that $\alpha = 2$ is a constant.)

Now, we may apply a union bound to deduce that the probability that there exists some e, t such that more than $5\log(Nd)$ packets use the edge e at time step t is at most $Nd \cdot Nd \cdot \frac{1}{(Nd)^3} \leq \frac{1}{Nd}$. Therefore, with probability $1 - 1/(Nd)$, we obtain a schedule in the unconstrained model with low congestion, namely one wherein at every time step, at most $5\log(Nd)$ packets traverse any particular edge.

(d) Note that it suffices to handle the case where the schedule in the unconstrained model has low congestion (i.e., .at every time step, at most $5\log(Nd)$ packets traverse any edge), since by part (c) this occurs with probability $1-O(1/N)$. (With probability $O(1/N)$, our schedule will do arbitrarily poorly, which is OK.) We turn such an unconstrained schedule into a real schedule by replacing every time step in the unconstrained schedule by $s = 5\log(Nd)$ time steps in the real schedule; we want it to be the case that for each $i = 1, 2, \ldots$, the locations of the packets in the real schedule after the (is)th time step will be the same as that in the unconstrained schedule after the ith time step. (This ensures that there is no interference between steps in the unconstrained schedule, so the analysis of parts (b) and (c) still holds.) *3pts*

Since in the unconstrained schedule at most s packets traverse any particular edge in a single time step, all of these packets can traverse this edge in s time steps in the real schedule without violating the constraint that at most one packet crosses an edge per time step. Once a packet crosses an edge, it waits at the other end of the edge until the next time step on the unconstrained schedule. Clearly, we only need queues of size $s = O(\log(Nd))$ to implement this scheme. The length of the unconstrained schedule is $d + \frac{\alpha c}{\log(Nd)}$, so the length of the real schedule is s times that, which is $O(c + d\log(Nd))$ (recall that $\alpha = 2$ is a constant).