Homework 2 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 35.

1. For $i = 1, 2, \ldots, 2^{20} - 2$, let X_i be an indicator random variable that assumes the value 1 if the 3 letters beginning at the i'th letter in the text form the sequence "ape", and 0 otherwise. Clearly, $\Pr[X_i = 1] = 26^{-3}$, so $\mathrm{E}[X_i] = 26^{-3}$. Now, let the random variable X be the number of times the sequence "ape" appears in the text. Then, *5pts*

$$X = X_1 + X_2 + \ldots + X_{2^{20}-2}.$$

By linearity of expectation, the expected number of times the sequence "ape" appears in the text is

$$\mathrm{E}[X] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \ldots + \mathrm{E}[X_{2^{20}-2}] = \frac{2^{20} - 2}{26^3} \approx 59.7.$$

2. (a) The distribution of X is given by *2pts*

$$\begin{aligned}
\Pr[X = t] &= \Pr[\text{exactly } k-1 \text{ of the first } t-1 \text{ tosses land heads}] \cdot \Pr[\text{the } t\text{'th toss lands heads}] \\
&= \binom{t-1}{k-1} p^{k-1}(1-p)^{t-k} \cdot p \\
&= \binom{t-1}{k-1} p^{k}(1-p)^{t-k}.
\end{aligned}$$

(b) Let Y_1 be the number of tosses up to, and including, the first head. For $i = 2, \ldots, k$, let Y_i be the number of tosses up to, and including, the i'th head, starting immediately after the $i-1$'th head. Then clearly $X = Y_1 + Y_2 + \ldots + Y_k$. Each of $Y_1, \ldots, Y_k$ is a geometric r.v. with mean p, so from class $\mathrm{E}[Y_1] = \mathrm{E}[Y_2] = \cdots = \mathrm{E}[Y_k] = \frac{1}{p}$. Hence $\mathrm{E}[X] = \frac{k}{p}$. *3pts*

3. (a) Andrew's scheme involves repeatedly flipping 1000 coins until exactly 500 heads and 500 tails are obtained. The probability of this happening is $p = \binom{1000}{500}\frac{1}{2^{1000}} \sim \frac{1}{\sqrt{500\pi}}$, using Stirling's approximation as explained below. The number of trials until we succeed is a geometric r.v. with parameter p, so the expected number of trials is $\frac{1}{p} \sim \sqrt{500\pi}$, giving a total of $1000\sqrt{500\pi} \approx 39600$ coin flips. *3pts*
It remains to explain the calculation with Stirling's formula. Note that, for any n, we have

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \sim \frac{(2n/e)^{2n}\sqrt{4\pi n}}{(n/e)^{2n}2\pi n} = \frac{2^{2n}}{\sqrt{\pi n}}.$$

Plugging in $n = 500$ we get $\binom{1000}{500} \approx \frac{2^{1000}}{\sqrt{500\pi}}$, which is exactly what we used in the previous paragraph.
Note: Some students didn't use Stirling's formula to simplify the calculation here. You should get into the habit of using it, since it gives you a much clearer idea of the magnitudes of various quantities than the raw binomial coefficients do.

(b) Betty's scheme involves tossing coins until either 500 heads or 500 tails have been obtained, and then padding the sequence to get 1000 tosses with exactly 500 heads and 500 tails. This is a bad scheme because it does not give a uniform distribution. To see this, consider for example the sequence consisting of 500 heads followed by 500 tails. Since there are $\binom{1000}{500}$ valid sequences, this particular one should have probability $1/\binom{1000}{500} \approx \sqrt{500\pi}2^{-1000}$ using Stirling's approximation as in part (a). But in Betty's scheme, this sequence has probability 2^{-500} (i.e., the chance of getting 500 heads in a row), which is clearly way too high since $\sqrt{500\pi} \ll 2^{500}$. *3pts*

(c) A better scheme involves randomly selecting the positions of the 500 heads in the sequence of 1000 tosses. One way to do this is to repeatedly choose random positions in the range 1 to 1000 until 500 *distinct* random positions have been obtained. A random number in the range 1 to 1024 can be picked using 10 coin tosses (since $2^{10} = 1024$; if we pick a number larger than 1000 we just try again–we ignore this detail). So say we're picking our k^{th} position, where $1 \leq k \leq 500$. The probability that a random number in the range 1 to 1024 has not been previously selected and is ≤ 1000 is $\frac{1001-k}{1024} \geq \frac{501}{1024} \geq \frac{1}{2.05}$, so the expected number of trials (each involving 10 coin tosses) till such a value is found is ≤ 2.05. The expected total number of coin tosses is then at most $\sum_{k=1}^{500} 2.05 \times 10 = 10250$. (Actually, we have been quite pessimistic here in taking the worst case over k. The exact expectation is $10 \times \sum_{k=1}^{500} \frac{1024}{1001-k} = 10240(\frac{1}{501} + \frac{1}{502} + \cdots + \frac{1}{1000}) \approx 10240 \ln(\frac{1000}{501}) \approx 7100$.) *4pts*

An even better method is the following (but we did not necessarily expect anybody to come up with this!). Suppose our first random coin toss is H. Then for our next toss, we should use a biased coin with heads probability $499/999$. (Why? Think about the conditional probabilities.) By the same argument, if the first i flips have h_i heads (and $i - h_i$ tails), then at the next step we should flip a coin with the probability of heads being $(500 - h_i)/(1000 - i)$, and so on. But how do we simulate these biased coins with our fair coin? We claim that *any* bias p can be realized using an expected number of only <u>two</u> fair coin flips!!! To see this, write $p = 0.p_1p_2\ldots$ in binary. Now generate a random number $r = 0.r_1r_2\ldots$ between 0 and 1 by successively choosing each binary digit r_i using an independent fair coin flip. We can stop when we know that $r > p$ or that $r < p$ (corresponding to an outcome of H or T respectively for our biased coin). And when do we know this? It is when we reach the first i for which $r_i \neq p_i$. But the expected number of tosses for this event to happen is just 2. (Why?) Putting all this together, the expected number of coin flips needed to generate the entire random sequence of 1000 flips is only 2000.

4. (a) Fix any $r \in \{1, 2\ldots, n\}$. The algorithm outputs r if and only if $c_n, \ldots, c_{r+1}$ land tails, and c_r lands heads. This happens with probability *2pts*

$$\frac{1}{r} \cdot \prod_{j=n}^{r+1} \left(1 - \frac{1}{j}\right) = \frac{1}{r} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot \cdots \cdot \frac{r}{r+1} = \frac{1}{n}.$$

Hence, the scheme generates $r \in \{1, 2, \ldots, n\}$ u.a.r.

(b) Following the hint, suppose we pick each s_i using the method of part (a). (This thought experiment is OK because we know from part (a) that this is a valid way of generating an integer u.a.r.) Following through the whole process of generating the s_i, we see that it is equivalent to flipping each coin c_j, $j = n, n-1, \ldots, 1$ until it lands tails, and then moving on to the next coin. The value m_j is the number of times c_j lands heads before it lands tails. The probability that c_j lands heads exactly m_j times before landing tails is $\left(\frac{1}{j}\right)^{m_j}\left(1 - \frac{1}{j}\right)$. Hence, the probability we generate the sequence (s_i) is *3pts*

$$\prod_{j=2}^{n} \left(\frac{1}{j}\right)^{m_j} \left(1 - \frac{1}{j}\right).$$

(Note that $j = 1$ is missed out because $m_1 = 1$ always.)

(c) Write $r \in \{1, \ldots, n\}$ in factored form as $\prod_p p^{\beta_p}$. (Recall that this factorization is unique.) The probability that the coin tosses of the algorithm produce the number r is exactly the probability that $m_p = \beta_p$ for all primes $p \leq n$. The values of the other m_j are irrelevant. Summing the probabilities in part (b) over all sequences satisfying this condition, we get *3pts*

$$\Pr\Big[m_p = \beta_p \text{ for all primes } p \leq n\Big] = \prod_p \left(\frac{1}{p}\right)^{\beta_p} \left(1 - \frac{1}{p}\right) = \frac{\alpha_n}{r}.$$

Note that, since all values of the other m_j are allowed, their contributions sum to 1. Finally, taking into account the last line of the algorithm, the probability that r is output is $\frac{r}{n} \cdot \frac{\alpha_n}{r} = \frac{\alpha_n}{n}$.

(d) The probability that the algorithm does not fail is $\sum_{r=1}^{n} \Pr[\text{algorithm outputs } r] = n \cdot \frac{\alpha_n}{n} = \alpha_n$. Hence, the expected number of trials is $\alpha_n^{-1} \sim 1.8 \ln n$. *2pts*

(e) Let X denote the number of primality tests performed by one trial of the algorithm. We can write this as $X = X_1 + \ldots + X_n$, where X_j is the indicator r.v. of the event that j is tested for primality. Note that this happens iff $m_j \geq 1$, so by our analysis in part (b) we have $\mathrm{E}[X_j] = \Pr[j \text{ is tested}] = \frac{1}{j}$. Hence $\mathrm{E}[X] = \sum_{j=1}^{n} \mathrm{E}[X_j] = H_n$, as required. *2pts*

(f) The total number of primality tests performed is $X_1 + X_2 + \ldots + X_T$, where X_i is the number of tests performed in the ith trial and T is the total number of trials needed until success is achieved. Since T is a stopping time for the X_i (in the sense that the event $T = t$ depends only on the outcomes of the first t trials), and the X_i are iid, we have by Wald's equation as given in the Note: *2pts*

$$\mathrm{E}(X_1 + X_2 + \ldots + X_T) = \mathrm{E}(X_1)\mathrm{E}(T) = O((\log n)^2),$$

using parts (d) and (e).

The following alternative argument avoids appealing to Wald's equation by using conditional expectations in a careful way. (Actually, this argument amounts to a proof of Wald's equation in this case.) First note that we can write $\sum_{i=1}^{T} X_i$ as $\sum_{i=1}^{\infty} X_i I_{\{T \geq i\}}$, where $I_{\{T \geq i\}}$ is the indicator r.v. of the event that $T \geq i$. Taking expectations:

$$\mathrm{E}\Big(\sum_{i=1}^{\infty} X_i I_{\{T \geq i\}}\Big) = \sum_{i=1}^{\infty} \mathrm{E}(X_i I_{\{T \geq i\}}) = \sum_{i=1}^{\infty} \mathrm{E}(X_i) \Pr[T \geq i] = \mathrm{E}(X_1) \sum_{i=1}^{\infty} \Pr[T \geq i] = \mathrm{E}(X_1)\mathrm{E}(T).$$

The key step here is the second equality, which follows from the fact that X_i is independent of the event $T \geq i$ (since this event depends only on the outcomes of the first $i - 1$ trials).

Note: We penalized students who simply multiplied $\mathrm{E}(X_1)$ *by* $\mathrm{E}(T)$ *without justification, since this is not always valid. You need to say that* T *is a stopping time and that the* X_i *are independent, as given in the Hint.*