CS-174 Combinatorics and Discrete Probability, Fall 98

Solutions for Sample Midterm 1

1. 11/5. We will have seen two different outcomes when, and only when, we have seen an outcome different from that of the first toss. Let X be the number of tosses (after the first toss) till we see a value different from the first outcome. Then X is a geometric random variable with "heads" probability $p = 5/6$, and hence $\mathrm{E}(X) = 1/p = 6/5$. Thus the total expected number of tosses is $1 + \mathrm{E}(X) = 11/5$.

2. e^{-1}. This is just the number of fixed points in a random permutation, which as $n \to \infty$ has a Poisson distribution (see Note 3).

3. $\sqrt{n}$. This is the number of balls we can throw with a small probability of getting a collision.

4. **(a)** np; **(b)** $np(1-p)$. This is because $X = \sum_{i=1}^{n} X_i$, where the X_i are independent Bernoulli random variables with probability p of being 1. The expectation of each X_i is p, and its variance is $p - p^2$. Thus the expectation of X is np (since expectations sum) and the variance of X is $np(1-p)$ (since the variance of the sum of *independent* random variables is the sum of the individual variances).

5. $\Pr[X > 10] < 1/10$ is true by Markov's inequality.

$\Pr[X = 1] > 0$ may be false. For example, let $X = 0$ w.p. 3/4 and $X = 4$ w.p. 1/4.

$\Pr[X \geq 2] > 0$ is true, since if X were always in the interval $[0, 2)$ with $\mu = \mathrm{E}(X) = 1$, then $(X - \mu)^2$ would always be in the range $[0, 1]$, and hence its expectation $\mathrm{Var}(X) = \mathrm{E}((X - \mu)^2)$ would be at most 1, which contradicts $\mathrm{Var}(X) = 3$.

$\mathrm{E}(X^2) = \mathrm{Var}(X) + \mathrm{E}(X)^2 = 4$ is true.

$\Pr[X \leq 1] = \Pr[X \geq 1]$ may be false (see the example above).

$\Pr[X \geq 3] = 0$ may be false (again see the example above).

6. **(a)** e^{-2}. This is simply $(1 - 2/n)^n \to e^{-2}$.

(b) e^{-2}. One way to see this is to let $\mathcal{E}_1$ be the event that the first bin has exactly one ball, $\mathcal{E}_2$ be the event that the second bin is empty and use $\Pr[\mathcal{E}_1 \wedge \mathcal{E}_2] = \Pr[\mathcal{E}_1]\Pr[\mathcal{E}_2|\mathcal{E}_1] = \binom{n}{1}\frac{1}{n}(1 - \frac{1}{n})^{n-1} \times (1 - \frac{1}{n-1})^{n-1} \to e^{-2}$.

7. **(a)** 7; **(b)** 6. Use the inclusion-exclusion principle (or just draw a Venn diagram).

8. All three methods in this question involve a simple experiment that is repeated until a number is generated. Clearly the whole method will be unbiased if and only if the simple experiment is unbiased.

(a) Only for $p = \frac{1}{2}$, since we need $p^2 = p(1-p)$, i.e., $p = 1 - p$.

(b) No values, since in each experiment the probability of getting 1 is $(1-p)^2$, that of 2 is $2p(1-p)$ and that of 3 is p^2, but these values can never all be equal.

(c) For all values $0 < p < 1$, since in each experiment the probability of each number in $\{1, 2, 3\}$ is $p(1-p)^2$.

(d) Let X be the number of rounds required for Alice to succeed, and Y be the number of tosses. Then $Y = 2X$, and X is a geometric random variable with probability of success being $1 - (1-p)^2$. Hence $\mathrm{E}(Y) = 2 \times [1 - (1-p)^2]^{-1} = 2/(2p - p^2)$ $(= 8/3$ for $p = 1/2)$.

(e) Let X be the number of rounds required for Charlie to succeed, and Y be the number of tosses. Then $Y = 3X$, and hence $\mathrm{Var}(Y) = 9\mathrm{Var}(X)$. But X is a geometric random variable with success probability $q = 3p(1-p)^2$, and so $\mathrm{Var}(X) = 1/q^2 - 1/q = (1-q)/q^2$. Thus $\mathrm{Var}(Y) = 9(1-q)/q^2$ with q as above. (This can't really be simplified much.)

(f) Since we can flip each coin twice, we use von Neumann's trick (from Homework 2, qun. 2(c)) to simulate an unbiased coin, and then we can use Alice's algorithm using these (simulated) unbiased coins.

9. **(a)** $2^{\binom{n}{2}}$.

(b) $p^m(1-p)^{\binom{n}{2}-m}$.

(c) Let X_e be the indicator variable for the inclusion of the edge e (i.e., $X_e = 1$ if e is present, and $X_e = 0$ otherwise). Then $\mathrm{E}(X_e) = p$. Also $X = \sum_e X_e$, and so $\mathrm{E}(X) = \binom{n}{2}p$.

(d) S is a clique if and only if each of its $\binom{k}{2}$ internal edges is present. Since edges are independent, the probability of this is $p^{\binom{k}{2}}$.

(e) Let X_S be the indicator variable for the event that S is a clique, for any k-vertex subgraph S. Then, from part (d), $\mathrm{E}(X_S) = p^{\binom{k}{2}}$. If X is the number of cliques of size k in G, we have $X = \sum_S X_S$, and so $\mathrm{E}(X) = \sum_S \mathrm{E}(X_S) = \binom{n}{k}p^{\binom{k}{2}}$, since G has exactly $\binom{n}{k}$ k-vertex subgraphs.

10. **(a)** There are more bit-strings of length 1Gb than those of length no more than 0.9Gb. However, the number of compressible files cannot be greater than the number of bit-strings of length no more than 0.9Gb (consider compressing a file and then uncompressing it — you should get the original file). Thus some files will not be compressible, contrary to the company's claim.

(b) *Since we can't assume anything about the way the software works*, the sensible thing to do is to just choose a 1Gb bit-string uniformly at random, and hope that it will not compress.

(c) The number of length-k bit-strings is 2^k. Thus the number of different bit strings of length at most m is $\sum_{k=0}^{m} 2^k = 2^{m+1} - 1 < 2^{m+1}$. As argued in part (a), the number of correctly compressed files cannot be greater than the number of bit-strings of length at most 0.9Gb, which is less than $2^{0.9\times 2^{33}+1}$. Since there are $2^{2^{33}}$ 1Gb bit-strings, we have Pr[file compressed correctly] $< 2^{0.9\times 2^{33}+1} \times 2^{-2^{33}} = 2^{1-0.1\times 2^{33}}$, which is minuscule. Thus the string chosen in part (b) will almost certainly not compress correctly.