Solutions will be posted on the web on Sunday, February 23.

Practice Midterm

- This is a **closed book** exam. Calculators **are** permitted. One single-sided sheet of notes is also permitted.
- This practice midterm consists of 10 questions. The first seven questions are multiple choice; the remaining three require written answers. (The actual midterm will contain only five multiple-choice questions and two questions requiring written answers.)
- Answer the multiple choice questions by circling the correct answer. You should be able to answer all of these from memory, by inspection, or with a very small calculation. Incorrect answers attract a negative score, so if you do not know the answer you should **not** guess.
- Write your answers to the other questions in the spaces provided below them. None of these questions requires a long answer, so you should have enough space; if not, continue on the back of the page and state clearly that you have done so. **Show all your working**.
- The questions vary in difficulty: if you get stuck on some part of a question, leave it and go on to the next one.

---

1. A fair 6-sided die is tossed repeatedly. The expected number of tosses until **two** different outcomes are observed is

$$2 \qquad \frac{11}{5} \qquad \frac{49}{40} \qquad \infty \qquad \frac{6}{5} \qquad \frac{3}{2}$$

---

2. An absent-minded secretary places n letters in n addressed envelopes in a completely random manner. The probability that none of the letters is placed in the right envelope is asymptotically

$$\frac{1}{n} \qquad e^{-1} \qquad 1-e^{-1} \qquad 1-\frac{1}{n} \qquad \frac{e}{n} \qquad \mathrm{e}^{-n}$$

---

3. The largest number of balls that can be tossed into n bins, if the probability that any bin contains more than one ball is to be kept small, is on the order of

a constant $\qquad \ln n \qquad \sqrt{n} \qquad n \qquad \frac{n}{e} \qquad \ln\ln n$

---

4. The random variable X has the Binomial distribution with parameters n and p, that is, $\mathbf{Pr}[X = k] = \binom{n}{k} 2^{-n}$.

   (a) The expectation of X is

   $0 \qquad 1 \qquad \frac{n}{p} \qquad \frac{p}{n} \qquad np$ $\qquad$ not determined

   (b) The variance of X is

   $1 \qquad \sqrt{np(1-p)} \qquad np \qquad \frac{np}{1-p} \qquad np(1-p)$ $\qquad$ not determined

---

5. A non-negative random variable X has expectation $\mathbf{E}[X] = 1$ and variance $\mathbf{Var}[X] = 3$. Circle those **three** of the following statements that **must** be true about X:

$$\mathbf{Pr}[X \geq 10] \leq \frac{1}{10} \qquad \mathbf{Pr}[X \geq 2] > 0 \qquad \mathbf{Pr}[X \leq 1] = \mathbf{Pr}[X \geq 1]$$

$$\mathbf{Pr}[X = 1] > 0 \qquad \mathbf{E}[X^2] = 4 \qquad \mathbf{Pr}[X \geq 3] = 0$$

---

6. n balls are tossed at random into n bins.

   (a) As $n \to \infty$, the probability that the first **two** bins are empty is

   $0 \qquad \frac{2}{n} \qquad \frac{1}{2n} \qquad e^{-2} \qquad e^{-1} \qquad e^{-1/2}$

   (b) As $n \to \infty$, the probability that the first bin contains exactly one ball **and** the second bin is empty is

   $0 \qquad \frac{2}{n} \qquad \frac{1}{2n} \qquad e^{-2} \qquad e^{-1} \qquad e^{-1/2}$

---

7. In a group of 10 people, 6 can play the piano, 5 can play the saxophone, 4 can play the violin, 4 can play the piano and the saxophone, 3 can play the piano and the violin, 2 can play the saxophone and the violin, only 1 can play all three instruments.

   (a) How many people can play **at least** one instrument?

   5 $\qquad$ 6 $\qquad$ 7 $\qquad$ 8 $\qquad$ 9 $\qquad$ 10

   (b) How many can play **exactly** two instruments?

   3 $\qquad$ 4 $\qquad$ 5 $\qquad$ 6 $\qquad$ 7 $\qquad$ 8

8. Alice, Bob and Charlie want to choose one of the numbers 1,2 and 3 with equal probabilities. All they have is a biased coin that comes up heads with probability p, where $0 < p < 1$.

   - Alice suggests the following method. Toss the coin twice. If the two flips are HH, output 1, if they are HT, output 2, if they are TH, output 3. If the flips are TT, repeat the experiment.
   - Bob suggests the following method. Toss the coin twice. Output the number of heads obtained, plus 1.
   - Charlie suggets the following method. Toss the coin three times. If heads is obtained **only** in the i-th toss, where $1 \leq i \leq 3$, output i. Otherwise, repeat the experiment.

   Considering such methods:

   (a) For which values of p, if any, does Alice's method produce the numbers 1,2 and 3 with equal probabilities?

   (b) For which values of p, if any, does Bob's method produce the numbers 1,2 and 3 with equal probabilities?

(c) For which values of p, if any, does Charlie's method produce the numbers 1,2 and 3 with equal probabilities?

(d) What is the **expected** number of tosses used by Alice's method.

(e) What is the **variance** of the number of tosses used by Charlie's method.

(f) Suppose that you have a supply of biased coins, each with an unknown, and possibly different, bias. You are allowed to use each coin only **twice**. Can you use the coins to select one of the numbers 1,2 and 3 with equal probabilities? Show how or explain why this is not possible.

9. Given parameters n and p (where $n \geq 2$ is an integer and $0 < p < 1$ is a real) construct a random graph in the following way:

   (a) start with vertex set $V = \{1, 2, \ldots, n\}$ and edge set $E = \emptyset$

   (b) for each of the $\binom{n}{2}$ possible edges $e = \{i, j\}$ independently, flip a coin with heads probability p; if the coin comes up heads, add e to E

   (c) output $G = (V, E)$

   This distribution over graph is called $G_{n,p}$.

   (a) What is the size of the sample space?

   (b) Let G be a graph in the sample space having exactly m edges. What is the probability of G, as a function of n, p and m?

   (c) Let the random variable X denote the number of edges in a graph G generated according to $G_{n,p}$. What is $\mathbf{E}[X]$, as a function of n and p?

(d) A subset of vertices $S \subseteq V$ is called a **clique** if $\{i, j\} \in E$ for all $i, j \in S$ (i.e., all pairs of vertices in S are connected by an edge). For any given set S of k vertices, what is the probability that S is a clique, as a function of p and k?

(e) What is the expected number of cliques of size k in G, as a function of n, p and k?

10. A certain start-up company announces a piece of software that will compress **any** gigabyte of data by at least 10%. (I.e., given any string of 1Gb = 2^{33} bits, it will return a compressed version of length at most 0.9×2^{33} bits. Of course, the original data must be recoverable from the compressed version. No claim is made about the performance on strings of other lengths.)

    (a) Explain, at a high level, why you should be skeptical about this product. (You need not do any calculations here.)

    (b) How would you construct a gigabyte of data which the software (probably) cannot compress? (Again, you need not do any calculations. Note that you **cannot** assume anything about how the company's software works.)

    (c) Give an upper bound on the probability that the data you constructed in part (b) will be correctly compressed by the software. [Hint: Show that the number of different bit strings of length at most m is $\leq 2^{m+1}$.]