ECE 361 Probability for Engineers (Fall, 2016)
Homework Solutions 4

Please answer the following questions:

1. (3 points) An undirected simple graph G of *order* n and *size* m has n vertices $V = [n]$ and m edges $E = (e_1, \ldots, e_m)$, with each e_i a distinct unordered pair of vertices. A graph is often denoted as $G = (V, E)$. A graph on n vertices may have anywhere from $m = 0$ to $m = M_n \equiv \binom{n}{2}$ edges. A graph with $m = 0$ is called an empty graph, while a graph with $m = M_n$ edges is a complete graph. Given a graph, we define the *neighbors* of a vertex $v \in [n]$ as those vertices $N(v)$ connected by an edge to v. Formally, $N(v) = \{u \in [n] : \{u, v\} \in E\}$ is the set of neighbors of node v, and every node $v \in [n]$ has a neighbor set, possibly empty, meaning there are no edges incident at that vertex. Finally, the last graph-theoretic notion we require is *degree*: the degree of a vertex is the number of neighbors, i.e., $d(v) = |N(v)|$, and again, every node $v \in [n]$ has a degree. Note $d(v) \in \{0, \ldots, n-1\}$, where $d(v) = 0$ means vertex v is *isolated* (has no neighbors) and $d(v) = n-1$ means vertex v is connected to every possible other vertex.

   A *random* graph with parameters (n, p), with $n \in \mathbb{N}$ and $p \in (0, 1)$, is defined in terms of a collection of independent and identically distributed Bernoulli random variables $X = (X_1, \ldots, X_{M_n})$, with $X_e \sim \mathrm{Ber}(p)$ for $e \in M_n$. We construct the random graph from the vector of RVs X as follows:

$$X_e = \begin{cases} 1, & \text{graph contains edge } e \\ 0, & \text{graph does NOT contain edge } e \end{cases} \quad (1)$$

   Thus, we effectively flip a (biased) coin once for each *possible* edge in the graph, i.e., for each unordered pair of vertices, and add an edge connecting those two vertices when that coin flip shows a head, which happens with probability p. Example, for $n = 4$ there are $M_4 = \binom{4}{2} = 6$ possible edges, and each such edge is added independently with probability p.

   Consider a random graph with order $n = 100$ nodes constructed as above with edge probability $p = 1/3$. Please answer the following questions:

   - Consider vertex 1. Let Y_1 be the RV giving the *degree* of vertex 1. What type of RV is Y_1?
   - Compute $\mathbb{E}[Y_1]$ and $\mathrm{Var}(Y_1)$.
   - Find $\mathbb{P}(Y_1 > 40)$. *Hint: use a computer or a table.*

   Solution. Observe there are $n-1$ possible edges connecting to vertex 1, and the presence or absence of each such edge is determined by the value of the corresponding Bernoulli RV. It follows that $Y_1 = X_2 + \cdots X_n$, where X_e corresponds to edge $\{1, e\}$, for $e \in \{2, \ldots, n\}$. Thus Y_1 is the sum of $n-1$ independent and identically distributed Bernoulli RVs, and as such Y_1 is a binomial RV with parameters $(n-1)$ and p: $Y_1 \sim \mathrm{bin}(n-1, p)$. It follows that $\mathbb{E}[Y_1] = (n-1)p$ and $\mathrm{Var}(Y_1) = (n-1)p(1-p)$. Finally, $\mathbb{P}(Y_1 > 40)$ is computed to equal

$$\mathbb{P}(Y_1 > 40) = \frac{3243836749930359736221192989952949691966379913}{57264168970223481226273458862846808078011946889} \approx 0.0566469. \quad (2)$$

   In Mathematica, for example, this is obtained from the command `1 - CDF[BinomialDistribution[99, 1/3], 40]`.

2. (3 points) Recall the previous problem. Consider a random graph with $n \geq 2 \in \mathbb{N}$ nodes and edge probability $p \in (0, 1)$, and let Y_1 be the RV giving the degree of vertex 1. Let A be the event that $\{1, 2\}$ is an edge in the graph. Give the conditional distribution of Y_1 given A.

Solution. As there is one known edge we have $Y_1|A \sim 1 + \mathrm{Bin}(n-2,p)$. That is, the degree at vertex 1 equals one plus a binomially-distributed number with parameters $n-2$ and p, where $n-2$ reflects the number of possible edges from vertex 1 to vertices $3,\ldots,n$. Thus

$$\mathbb{P}(Y_1 = k) = \mathbb{P}(\mathrm{Bin}(n-2,p) = k-1),\ k \in \{1,\ldots,n-1\}. \tag{3}$$

3. (3 points) Let (X_1, X_2) be two independent and identically distributed geometric random variables, each with parameter $p \in (0,1)$. That is, $X_1 \sim \mathrm{Geo}(p)$ and $X_2 \sim \mathrm{Geo}(p)$. Let $A_n = \{X_1 + X_2 = n\}$ for $n \in \mathbb{N}$ be their total. You are told A_n is true and asked to guess the most likely value for X_1, i.e., given (n,p) find the value i^* as a function of n,p such that $\mathbb{P}(X_1 = i^*|X_1+X_2=n) > \mathbb{P}(X_1 = i|X_1+X_2=n)$ for all $i \neq i^*$. *Hint: first express $\mathbb{P}(X_1+X_2=n)$ as a simple expression of n and p. Verify that adding up this expression over all $n \in \{2,3,4\ldots\}$ yields one. Next, use Bayes's rule to express $\mathbb{P}(X_1 = i|X_1+X_2=n)$ in terms of $\mathbb{P}(X_1+X_2=n|X_1=i)$, $\mathbb{P}(X_1=i)$, and $\mathbb{P}(X_1+X_2=n)$.*

Solution. Fix $\bar{p} \equiv 1-p$ and let us first compute for $n \in \{2,3,\ldots,\}$:

$$\begin{aligned}
\mathbb{P}(X_1+X_2=n) &= \sum_{k=1}^{n-1} \mathbb{P}(X_1+X_2=n|X_1=k)\mathbb{P}(X_1=k) \\
&= \sum_{k=1}^{n-1} \mathbb{P}(X_2=n-k|X_1=k)\mathbb{P}(X_1=k) \\
&= \sum_{k=1}^{n-1} \mathbb{P}(X_2=n-k)\mathbb{P}(X_1=k) \\
&= \sum_{k=1}^{n-1} \bar{p}^{n-k-1}p \times \bar{p}^{k-1}p \\
&= (n-1)\bar{p}^{n-2}p^2
\end{aligned} \tag{4}$$

Next, although it is not required for the solution, we verify this is in fact a valid distribution, i.e., if $p_n = \mathbb{P}(X_1+X_2=n) = (n-1)\bar{p}^{n-2}p^2$ then we wish to show that $p_2 + p_3 + \cdots = 1$:

$$\begin{aligned}
\sum_{n=2}^{\infty} p_n &= \sum_{n=2}^{\infty}(n-1)\bar{p}^{n-2}p^2 \\
&= p^2 \sum_{n=2}^{\infty}(n-1)\bar{p}^{n-2} \\
&= p^2 \sum_{n=2}^{\infty} \frac{\mathrm{d}}{\mathrm{d}\bar{p}}\left(\bar{p}^{n-1}\right) \\
&= p^2 \frac{\mathrm{d}}{\mathrm{d}\bar{p}} \sum_{n=2}^{\infty} \bar{p}^{n-1} \\
&= p^2 \frac{\mathrm{d}}{\mathrm{d}\bar{p}} \left(\bar{p}\sum_{n=0}^{\infty} \bar{p}^{n}\right) \\
&= p^2 \frac{\mathrm{d}}{\mathrm{d}\bar{p}} \left(\frac{\bar{p}}{1-\bar{p}}\right) \\
&= p^2 \frac{(1-\bar{p})(1) - \bar{p}(-1)}{(1-\bar{p})^2} = 1
\end{aligned} \tag{5}$$

Now use Bayes' rule:

$$\begin{aligned}
\mathbb{P}(X_1 = i | X_1 + X_2 = n) &= \frac{\mathbb{P}(X_1 + X_2 = n | X_1 = i)\mathbb{P}(X_1 = i)}{\mathbb{P}(X_1 + X_2 = n)} \\
&= \frac{\mathbb{P}(X_2 = n - i | X_1 = i)\mathbb{P}(X_1 = i)}{\mathbb{P}(X_1 + X_2 = n)} \\
&= \frac{\mathbb{P}(X_2 = n - i)\mathbb{P}(X_1 = i)}{\mathbb{P}(X_1 + X_2 = n)} \\
&= \frac{\bar{p}^{n-i-1} p \times \bar{p}^{i-1} p}{(n-1)\bar{p}^{n-2} p^2} = \frac{1}{n-1} \qquad (6)
\end{aligned}$$

Thus, given knowledge that $X_1 + X_2 = n$, it follows that X_1 is uniformly distributed over $\{1, \ldots, n-1\}$. As such, all values i are equally likely.

4. (3 points) Consider a collection of N fair M-sided dice, for $N \in \mathbb{N}$ and $M \geq 2$. We roll all the dice simultaneously, and put aside all those that show face M. We then re-roll all the remaining dice, and again put aside all those that show face M. We repeat this procedure until all dice show face M. Define the RV X_N with support $\mathbb{N}$ as the number of times we rolled dice (not the number of dice rolled) starting with N dice. Find $\mathbb{E}[X_5]$ for $M = 6$. In words, find the average number of rolls required until five fair six-sided dice each show six, under the above procedure. You may give either an analytical expression or simulate this process on a computer and report simulation results. If you use a computer, use at least 1000 trials and average the results.

Solution. Define $m_N = \mathbb{E}[X_N]$. Observe $m_0 = 0$ and $m_1 = M$ since the random number of times the one die is rolled until it shows an M is a geometric RV with parameter $1/M$ and mean M. It is very easy to simulate this process, although this is not required for the solution. Here is Mathematica code:

```
Clear[sim4];
sim4[Nn_, M_] := Module[{X},
   X = RandomVariate[BinomialDistribution[Nn, 1/M]];
   If[X == Nn, 1, 1 + sim[Nn - X, M]]
   ];
```

This simply generates a binomial RV $X \sim \mathrm{Bin}(N, 1/M)$ and returns 1 if $X = N$ or else returns $1 + \mathsf{sim4}(N - X, M)$, meaning it recurses, as is natural in this problem.

Condition on the number of dice $k \in \{0, \ldots, N\}$ removed after the first toss:

$$m_N = 1 + \sum_{k=0}^{N} m_{N-k} \binom{N}{k} \left(\frac{1}{M}\right)^k \left(1 - \frac{1}{M}\right)^{N-k}. \qquad (7)$$

This expresses m_N in terms of $m_N, \ldots, m_0$. We separate the $k = 0$ term:

$$m_N = m_N \left(1 - \frac{1}{M}\right)^N + \left(1 + \sum_{k=1}^{N} m_{N-k} \binom{N}{k} \left(\frac{1}{M}\right)^k \left(1 - \frac{1}{M}\right)^{N-k}\right). \qquad (8)$$

and solve for m_N to get a recursion for m_N in terms of $m_{N-1}, \ldots, m_0$

$$m_N = \frac{1 + \sum_{k=1}^{N} m_{N-k} \binom{N}{k} \left(\frac{1}{M}\right)^k \left(1 - \frac{1}{M}\right)^{N-k}}{1 - \left(1 - \frac{1}{M}\right)^N} \qquad (9)$$

Solving this for $N \in \{2,3,4,5\}$ gives:

$$
\begin{aligned}
m_0 &= 1 \\
m_1 &= M \\
m_2 &= \frac{3M^2-2M+1}{2M-1} \\
m_3 &= \frac{11M^4-19M^3+12M^2-3M}{6M^3-9M^2+5M-1} \\
m_4 &= \frac{25M^6-69M^5+85M^4-58M^3+22M^2-4M}{12M^5-30M^4+34M^3-21M^2+7M-1} \\
m_5 &= \frac{137M^{10}-655M^9+1509M^8-2171M^7+2132M^6-1476M^5+720M^4-240M^3+50M^2-5M}{60M^9-270M^8+590M^7-805M^6+747M^5-485M^4+219M^3-66M^2+12M-1}
\end{aligned}
\tag{10}
$$

It is helpful to use a computer to obtain the solution of the recurrence in (9). Here is Mathematica code:

```
Clear[mr];
mr[0,m_]:=0;
mr[n_,m_]:=(1+Sum[mr[n-k,m] Binomial[n,k] (1/m)^k (1-1/m)^(n-k), {k,1,n}])/(1-(1-1/m)^n);
Assuming[m in Integers && m>1,FullSimplify[mr[1,m]]]
Assuming[m in Integers && m>1,FullSimplify[mr[2,m]]]
Assuming[m in Integers && m>1,FullSimplify[mr[3,m]]]
Assuming[m in Integers && m>1,FullSimplify[mr[4,m]]]
Assuming[m in Integers && m>1,FullSimplify[mr[5,m]]]
```

Evaluating these quantities at $M = 6$ gives values

N	1	2	3	4	5
m_N	6	$\frac{96}{11}$	$\frac{10{,}566}{1{,}001}$	$\frac{728{,}256}{61{,}061}$	$\frac{3{,}698{,}650{,}986}{283{,}994{,}711}$
m_N	6	8.727	10.555	11.927	13.024

Thus, the answer to the original question is that just over 13 rolls are required on average for 5 six-sided dice to all show a six.

5. (3 points) A soft drink manufacturer runs a promotion where each bottle cap shows a random number between 1 and $N = 10$, and a prize is offered to any consumer holding a collection of N bottle caps with each of the N possible numbers included in the set. Let T_N be the RV denoting the number of soft drinks purchased until the complete set is obtained. Give $\mathbb{E}[T_N]$. *Hint: think of opening one can each time unit and thus T_N is the random time at which you first have a complete set. Define N stages, labeled $1,\ldots,N$, where in stage n we have obtained $n-1$ distinct bottle caps and are looking for the nth distinct bottle cap. Let $(X_1,\ldots,X_N)$ be RVs where X_n is the number of bottles opened in stage n. Observe $T = X_1 + \cdots + X_N$. Use linearity of expectation. Read about the Nth harmonic number and the Euler-Mascheroni constant.*

Solution. Define T_N as the random time to collect all N caps, where in each time slot we open one new bottle. The key to the solution is to decompose the T_N bottles into stages. In particular, we define N stages, labeled $1,\ldots,N$, where in stage n we have obtained $n-1$ distinct bottle caps and are looking for the nth distinct bottle cap. Let $(X_1,\ldots,X_N)$ be RVs where X_n is the number of bottles opened in stage n. In stage n we have $n-1$ caps and are missing $N-(n-1)$ caps, and so the probability of a cap being new is $p_n = (N-(n-1))/N = 1-(n-1)/N$. It follows that $X_n \sim \mathrm{Geo}(p_n)$ and thus $\mathbb{E}[X_n] = 1/p_n = \frac{N}{N-(n-1)}$.

Finally, we observe $T_N = X_1 + \cdots + X_N$ and, by linearity of expectation,

$$\begin{aligned}
\mathbb{E}[T_N] &= \sum_{n=1}^{N} \mathbb{E}[X_n] \\
&= \sum_{n=1}^{N} \frac{N}{N-(n-1)} \\
&= N\left(\frac{1}{N} + \frac{1}{N-1} + \cdots + \frac{1}{2} + 1\right) \\
&= N \sum_{n=1}^{N} \frac{1}{n} \\
&= N H_N
\end{aligned} \tag{11}$$

Here $H_N = 1 + 1/2 + \cdots + 1/N$ is the Nth harmonic number, and $H_N \approx \log N + \gamma$, for $\gamma \approx 0.577$ the Euler-Mascheroni constant. Thus, in particular for $N = 10$ we have $\mathbb{E}[T_{10}] = \frac{7381}{252} \approx 29.2897$, and using the H_N approximation we have $\mathbb{E}[T_{10}] \approx 10(\log 10 + \gamma) \approx 28.798$. We must purchase on average approximately 28 or 29 bottles in order to collect all ten bottle caps.

6. (3 points) Consider N distinct N-sided dice. The N dice are rolled simultaneously and we let the RV $X_N \in \{1, \ldots, N\}$ denote the number of distinct values shown on the N dice. Find $\mathbb{P}(X_N = 1)$, $\mathbb{P}(X_N = 2)$, $\mathbb{P}(X_N = N-1)$, and $\mathbb{P}(X_N = N)$ for generic $N \geq 2$. *Hint: try $N = 2$, $N = 3$, and $N = 4$ to build intuition for the general case. I advise you number the dice from 1 to N so they are distinguishable, and observe there are N^N possible outcomes when they are all rolled. The event $\{X_N = 1\}$, for example, has a probability given by the fraction of the N^N outcomes with 1 distinct value shown. Use the concepts of combinations, permutations, and multinomial coefficients studied earlier. Verify your proposed formulas for general N with calculations by hand for small N such as $N \in \{2, 3, 4\}$.*

Solution. Number the dice $1, \ldots, N$ so they are distinguishable. There are then N^N different outcomes for the experiment.

- The event $\{X_N = 1\}$ corresponds to the N different outcomes where all N dice show the same face. Thus:
$$\mathbb{P}(X_N = 1) = \frac{N}{N^N}. \tag{12}$$
- The event $\{X_N = 2\}$ means all faces show one of two numbers; there are $\binom{N}{2}$ possible pairs of numbers. Let k be the number of dice that show the first number, so that $N - k$ dice show the second number. For each k there are $\binom{N}{k}$ orderings of the positions of these k numbers. Combining:
$$\mathbb{P}(X_N = 2) = \frac{1}{N^N}\binom{N}{2}\sum_{k=1}^{N-1}\binom{N}{k} = \frac{1}{N^N}\binom{N}{2}\left(\sum_{k=0}^{N}\binom{N}{k}1^k 1^{N-k} - 2\right) = \frac{1}{N^N}\binom{N}{2}\left(2^N - 2\right) \tag{13}$$
Alternately, and more directly, observe that once we have selected the $\binom{N}{2}$ distinct numbers to appear, we then fill in each of the N faces with one of these two numbers. There are 2^N binary strings, each mapping to an assignment of these two numbers to the N dice, and the only inadmissable strings are the 2 constant-value strings.
- The event $\{X_N = N-1\}$ means $N-1$ faces show distinct numbers, and the last face shows one of the previous $N-1$ numbers.
$$\mathbb{P}(X_N = N-1) = \frac{1}{N^N}\binom{N}{N-1}(N-1)\frac{N!}{2} = \frac{N(N-1)N!}{2N^N} \tag{14}$$

To explain.

- The $\binom{N}{N-1} = N$ choices for the $N-1$ distinct numbers to be included, or, equivalently, there are N choices for the number to be omitted.
- The $N-1$ means there are $N-1$ choices for which of these $N-1$ numbers will be chosen by the last die.
- The $N!/2$ corresponds to the multinomial coefficient $\binom{N!}{1!\cdots 1!2!}$ for the number of distinct ways to order a word with N letters where $N-1$ letters occur once and one letter occurs twice.

• The event $\{X_N = N\}$ means N faces show N distinct numbers, and thus

$$\mathbb{P}(X_N = N) = \frac{N!}{N^N}. \tag{15}$$

7. (3 points) Recall the previous problem. Find $\mathbb{P}(X = n)$ for each $n \in [N]$ for $N = 6$. *Hint: you may use a computer for this problem.*

Solution. This is computed using the following Mathematica code:

```
Clear[p6];
p6[n_]:=Table[Length[Select[Tuples[Range[n],n],CountDistinct[#]==k&]]/(n^n),{k,n}];
MatrixForm[p6[6]]
```

This produces

n	1	2	3	4	5	6
p_n	$\frac{6}{46,656}$	$\frac{930}{46,656}$	$\frac{10,800}{46,656}$	$\frac{23,400}{46,656}$	$\frac{10,800}{46,656}$	$\frac{720}{46,656}$
p_n	$\frac{1}{7776}$	$\frac{155}{7776}$	$\frac{25}{108}$	$\frac{325}{648}$	$\frac{25}{108}$	$\frac{5}{324}$
p_n	0.000128	0.019933	0.231481	0.501543	0.231481	0.015432

(16)

Observe there is a 50% chance that 4 faces show, and a 96% chance that the number of faces will be one of $\{3, 4, 5\}$.