Midterm 2 Solutions

Note: These solutions are not necessarily model answers. 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 number of points is 70. Comments in italics following the solutions highlight some common errors or give explanations.

1. True-False/Multiple Choice [26 points]

(a) **[True or false?]** *1 point for each correct answer; -1 point for each incorrect answer; 0 points for no answer* *15pts*

- In an undirected graph, the sum of the degrees of the vertices is equal to the number of edges in the graph: **False.** [*The sum of the degrees is equal to* **twice** *the number of edges.*]

- In a directed graph, the sum of the in-degrees of the vertices is equal to the sum of the out-degrees: **True.**

- The number of edges in an n-dimensional hypercube is $n2^{n-1}$: **True.** [*The hypercube has* 2^n *vertices and each of them has degree* n*, so the number of edges is* $\frac{1}{2} \times 2^n \times n = n2^{n-1}$.]

- The length of a de Bruijn sequence is always a power of two: **True.**

- The following is a de Bruijn sequence for $n = 3$: 10010110: **False.** [*For example, the 3-bit string 101 appears twice as a subsequence, and the string 111 doesn't appear at all.*]

- In any probability space, there is an event E such that E and A are independent for all events A: **True.** [*We can take* E *to be either the entire probability space* Ω*, or the empty set.*]

- If the events A, B are independent, and the events B, C are independent, then the events A, C must be independent: **False.**

- For any events A and B such that $\Pr[B] > 0$, we have $\Pr[A|B] \geq \Pr[A \cap B]$: **True.** [*By definition of conditional probability,* $\Pr[A|B] = \Pr[A \cap B]/\Pr[B]$*, and this is* $\geq \Pr[A \cap B]$ *since* $\Pr[B] \leq 1$.]

- Let A and B be events such that $\Pr[A|B] = \Pr[A|\overline{B}]$, where $\overline{B}$ denotes the complement of B. Then A and B are independent: **True.** [$\Pr[A] = \Pr[A \cap B] + \Pr[A \cap \overline{B}] = \Pr[A|B]\Pr[B] + \Pr[A|\overline{B}]\Pr[\overline{B}] = \Pr[A|B](\Pr[B] + \Pr[\overline{B}]) = \Pr[A|B]$. *(In the third step here we used the assumption* $\Pr[A|B] = \Pr[A|\overline{B}]$*.) Hence* A, B *are independent.*]

- For events A, B, the probability that neither of the events happens is $1 - \Pr[A] - \Pr[B]$: **False.** [*The correct expression is* $1 - \Pr[A] - \Pr[B] + \Pr[A \cap B]$.]

- For three events A, B, C, the probability that *exactly one* of the events happens is $\Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[B \cap C] - \Pr[A \cap C] + 2\Pr[A \cap B \cap C]$: **False.** [*The correct expression is* $\Pr[A] + \Pr[B] + \Pr[C] - 2\Pr[A \cap B] - 2\Pr[B \cap C] - 2\Pr[A \cap C] + 3\Pr[A \cap B \cap C]$.]

- For any indicator (i.e., 0/1-valued) random variable X, $\mathrm{E}(X^2) = \mathrm{E}(X)^2$: **False.** [*If* $\Pr[X = 1] = p$ *then we have* $\mathrm{E}(X^2) = \mathrm{E}(X) = p$ *but* $\mathrm{E}(X)^2 = p^2$. *What* is *true, however, is that* $\mathrm{E}(X^2) = \mathrm{E}(X)$*, simply because* $X^2 = X$ *for any 0/1-valued r.v.*.]

- For any random variable X, $\mathrm{E}(X) = \sum_{k=0}^{\infty} \Pr[X \geq k]$: **False.** [*This equality holds only for positive, integer-valued random variables (see proof in Lecture Note 15).*]

- For any two random variables X, Y on the same probability space, $\mathrm{E}((2X+1)(Y+1)) = 2\mathrm{E}(XY) + 2\mathrm{E}(X) + \mathrm{E}(Y) + 1$: **True.** [*This is an immediate consequence of linearity of expectation.*]

- For any two random variables X, Y on the same probability space, if $\mathrm{E}(X) > \mathrm{E}(Y)$ then $\Pr[X > Y] > 0$: **True.** [*Proof by contradiction: if* $\Pr[X > Y] = 0$ *then we always have* $X \leq Y$*, so clearly* $\mathrm{E}(X) \leq \mathrm{E}(Y)$.]

Multiple Choice *1 point for each correct answer; 0 points for incorrect answers* *11pts*

(b) You are dealt a hand of five cards from a randomly shuffled standard deck of 52 cards.

(i) The probability that your hand contains no pair of cards with the same numerical value (there are 13 numerical values: ace, 2, ..., king) is

$$\frac{48 \cdot 44 \cdot 40 \cdot 36}{51 \cdot 50 \cdot 49 \cdot 48}$$

The number of hands of this form is $(52 \cdot 48 \cdot 44 \cdot 40 \cdot 36)/5!$: *52 choices for the first card, 48 for the second, 44 for the third, 40 for the fourth, 36 for the fifth. The total number of hands is* $(52 \cdot 51 \cdot 50 \cdot 49 \cdot 48)/5!$. *The probability is the ratio of these two numbers.*

(ii) The probability that the hand contains *at least three* cards of the same value is

$$\frac{13\left(4 \cdot \binom{48}{2} + 48\right)}{\binom{52}{5}}$$

This can happen in two ways: three or four cards of the same value. The number of hands with three cards of the same value is $13 \cdot 4 \cdot \binom{48}{2}$ *(13 choices for the value,* $\binom{4}{3} = 4$ *choices for the 3 suits,* $\binom{48}{2}$ *choices for the remaining 2 cards); the number of hands with four cards of the same value is* $13 \cdot 48$ *(13 choices for the value, 48 choices for the one remaining card).*

(c) A fair six-sided die is thrown repeatedly.

(i) The expected number of throws until the first 6 appears is : **6**.
The number of throws until 6 appears is a geometric r.v. with parameter $p = \frac{1}{6}$. *The expectation is* $\frac{1}{p} = 6$.

(ii) The expected number of throws until two 6's have appeared is **12**.
This is the sum of two r.v.'s as in part (i). By linearity of expectation, the expectation is just twice that in part (i).

(iii) The expected number of throws until *two different* numbers have appeared is **11/5**.
The first number appears for sure on the first throw. After this, the number of throws until a new number appears is a geometric r.v. with parameter $p = \frac{5}{6}$*, which has expectation* $\frac{1}{p} = \frac{6}{5}$.

(d) A book contains 10^6 characters. Each character is mis-typed with probability 10^{-5}, independently of all other characters.

(i) The expected number of mis-typed characters in the book is **10**.
By linearity of expectation, the expectation is $10^6 \times 10^{-5} = 10$.

(ii) The probability that the book contains exactly seven mis-typed characters is approximately

$$e^{-10}\frac{10^7}{7!}$$

The number of mistyped characters is approximately Poisson with parameter $\lambda = 10$.

(e) The n volumes of an encyclopedia are arranged in a random sequence on a shelf, so that all $n!$ orderings are equally likely. Let $\pi(i)$ denote the position of volume i in the sequence.

(i) The probability that $\pi(1) < \pi(2) < \pi(3)$ is **1/6**.
There are a total of $3! = 6$ *possible orderings of volumes* $1, 2, 3$*, each of which is equally likely. So the ordering* $\pi(1) < \pi(2) < \pi(3)$ *has probability 1/6.*

(ii) *Conditional on* the event $\pi(1) < \pi(3)$, the probability that $\pi(1) < \pi(2) < \pi(3)$ is **1/3**.
We have $\Pr[\pi(1) < \pi(2) < \pi(3) | \pi(1) < \pi(3)] = \Pr[(\pi(1) < \pi(2) < \pi(3)) \cap (\pi(1) < \pi(3)] / \Pr[\pi(1) < \pi(3)] = (1/6)/(1/2) = 1/3$.

(f) Each cereal box contains one coupon, chosen independently and uniformly at random from a set of n different coupons.

(i) The expected number of boxes that need to be bought before at least one copy of *all* n coupons is obtained is on the order of $\mathbf{n \ln n}$.
This is just coupon collecting.

(ii) The number of boxes that need to be bought before the probability that some coupon appears twice reaches 1/2 is on the order of $\sqrt{\mathbf{n}}$.
The event that some coupon appears twice after a total of m *boxes have been bought is the same as the event that two people have the same birthday when the number of people is* m *and the number of possible birthdays is* n*. From Lecture Note 13 we know that the probability of this event reaches 1/2 when* m *is on the order of* $\sqrt{n}$.

2. Euler and Hamilton [8 points]

(a) The graph in figure (a) has an Eulerian cycle but no Hamiltonian cycle.
One possible Eulerian cycle is: *2pts*

$$[A, E, B, F, C, G, D, F, A]$$

We can see that there is no Hamiltonian cycle by observing that any path from A to (for instance) C must pass through F; however, to complete the cycle, it must return to A passing through F a second time. Alternatively and more formally, observe that the graph is a bipartite graph with parts $\{A, B, C, D\}$ and $\{E, F, G\}$. We know that any cycle in a bipartite graph must have an even number of vertices. Since our graph contains 7 vertices, at least one vertex must be visited more than once in a cycle that visits all vertices at least once. *2pts*

(b) The graph in figure (b) has no Eulerian cycle but has a Hamiltonian cycle.
By Euler's theorem, a graph (with no isolated vertices) has an Eulerian cycle iff it is connected and every vertex has even degree. Since vertices such as B have odd degree, the graph has no Eulerian cycle. *2pts*
A common mistake was to state that a graph has an Eulerian tour iff it is connected and has no more than 2 odd vertices. This is the condition for Eulerian paths, not tours. This mistake did not usually affect the correctness of the final answer.

One possible Hamiltonian cycle is: *2pts*

$$[A, B, C, D, H, G, F, J, K, L, P, O, N, M, I, E, A]$$

A number of students gave an answer that was a Hamiltonian path (it visited all vertices once) but not a Hamiltonian cycle (ending at the starting node)

Final Note: Several students mentioned hypercubes in their answers. However, neither of these graphs is a hypercube!

3. Artistic Counting [14 points]

(a) The answer is *2pts*

$$\frac{50!}{40!}$$

since this is the balls and bins variant where the balls are the 10 patrons (which are labeled) and the bins are the 50 paintings (which are without replacement).

Most students answered this question correctly. The most common mistake was to treat the patrons as unlabeled, yielding $\binom{50}{40}$ possible sets of paintings that could be given away (but not accounting for who gets which painting).

(b) The answer is *2pts*

$$\binom{50}{40}$$

since this is the unlabeled/without replacement balls and bins variant where the balls represent which 40 paintings are not selected and the bins are the 50 paintings.

Many students didn't realize that sets are inherently unordered (e.g. the set {1,2,3} is the same as {2,3,1}) and so gave the same answer to this part as the previous part.

(c) There are several ways to count this. One correct answer is *2pts*

$$\frac{50!}{40!} - 10 \cdot 9 \cdot \frac{48!}{40!}$$

since there are $\frac{50!}{40!}$ total ways to give away the paintings, and we must subtract off the number of ways such that both special paintings are given away: there are 10 choices for who gets the first special painting, then 9 choices for who gets the other special painting, and then $\frac{48!}{40!}$ ways for the remaining 8 patrons to choose among the 48 non-special paintings. Another correct answer is

$$\left(\binom{50}{10} - \binom{48}{8}\right) \cdot 10!$$

since $\binom{50}{10}$ is the total number of sets of paintings and $\binom{48}{8}$ is the number of sets of paintings that include both special ones, and then after an acceptable set of 10 paintings is chosen there are $10!$ ways to distribute them to the patrons. Another correct answer is

$$\frac{48!}{38!} + 2 \cdot 10 \cdot \frac{48!}{39!}$$

since there are $\frac{48!}{38!}$ ways such that neither special painting is given away, and $10 \cdot \frac{48!}{39!}$ ways such that the first special painting is given away but not the second, and similarly if the second special painting is given away but not the first. Another correct answer is

$$2 \cdot \frac{49!}{39!} - \frac{48!}{38!}$$

since there are $\frac{49!}{39!}$ ways such that the first special painting is not given away (and similarly for the second), but then we must subtract off the $\frac{48!}{38!}$ ways such that neither is given away.

Of course, all of the above answers evaluate to the same quantity!

Relatively few students had correct answers to this question, and there was an extremely wide range of incorrect answers. Partial credit was very rarely given. A common wrong answer was $2 \cdot \frac{49!}{39!}$, which double-counts the cases in which neither special painting is given away.

(d) The answer is *2pts*

$$50^{10}$$

since this is the balls and bins variant where the balls are the 10 patrons (which are labeled) and the bins are the 50 paintings (which are now with replacement).

The most common wrong answer was $\binom{59}{10}$, which corresponds to treating the patrons as unlabeled.

(e) The answer is *2pts*

$$\binom{59}{10}$$

since this is the balls and bins variant where the balls are unlabeled and the bins are with replacement. The crucial difference from part (d) is that the printer only needs to know how many copies of each painting to print; she doesn't care which patron will receive each reproduction.

Several students confused this with other balls/bins variants, answering 50^{10} or $\binom{50}{10}$. A few people applied the formula incorrectly and got answers such as $\binom{61}{10}$ or $\binom{59}{50}$.

(f) The answer is *2pts*

$$\binom{54}{5}^{10}$$

since each patron has $\binom{50+5-1}{5}$ choices for the copies they receive (by unlabeled / with replacement balls and bins), and the 10 patrons may choose independently.

A common wrong answer was 50^{50}, presumably because students thought that each patron has 50^5 ways to choose 5 reproductions (which is not accurate since a patron only cares about how many of each painting he got, and not about the order in which he receives his 5 copies). But 50^{50} all by itself received no credit since it was not clear whether the 50 in the exponent was referring to the total number of paintings (which would be a grievous misunderstanding).

(g) The answer is *2pts*

$$\frac{50!}{40!} \cdot 49^{10}$$

since there are $\frac{50!}{40!}$ ways for the patrons to choose their originals (by part a), and then each patron independently has 49 choices for which reproduction to get (any one except the original he got).

Partial credit was only awarded for solutions quite close to the correct answer. Some students wrote $50^{10} \cdot \frac{49!}{39!}$, presumably trying to first count the ways to give away reproductions and then count the ways to give away originals. This does not work since, for example, if the reproduction chosen by the 2nd patron is the same as the original chosen by the 1st patron, then the 2nd patron has 49 choices for his reproduction (but would have only 48 choices if his reproduction is different from the original chosen by the 1st patron).

4. Light Bulb Factory [12 points]

(a) Let A, B and C be the events that the light bulb comes from machines A, B, and C respectively. If we define D as the event that the light bulb is defective, then applying the law of total probability (as *4pts*

A, B, and C are disjoint events that cover the whole sample space):

$$\begin{aligned}
\Pr[D] &= \Pr[D \cap A] + \Pr[D \cap B] + \Pr[D \cap C] \\
&= \Pr[D|A]\Pr[A] + \Pr[D|B]\Pr[B] + \Pr[D|C]\Pr[C] \\
&= \frac{4}{100}\frac{50}{100} + \frac{2}{100}\frac{30}{100} + \frac{7}{100}\frac{20}{100} \\
&= 0.04
\end{aligned}$$

Some students confused $\Pr[D|A]$ *(given in the problem) with* $\Pr[D \cap A]$*. If students left their result as a single fraction they were awarded full points, but we deducted points if the result was left as a sum of several terms.*

(b) Bayes' rule lets us compute the conditional probability $\Pr[A|D]$ with $\Pr[D|A]$ (which is given in the problem). Applying Bayes' rule we obtain: *4pts*

$$\begin{aligned}
\Pr[A|D] &= \frac{\Pr[D|A]\Pr[A]}{\Pr[D]} = 0.5 \\
\Pr[B|D] &= \frac{\Pr[D|B]\Pr[B]}{\Pr[D]} = 0.15 \\
\Pr[C|D] &= \frac{\Pr[D|C]\Pr[C]}{\Pr[D]} = 0.35
\end{aligned}$$

using $\Pr[D]$ computed in part (a).

If students had part (a) wrong, they were still awarded full credit as long as their calculation in part (b) was correct.

(c) Let D_2 be the event that a second light bulb coming from the same machine as the first one is defective. In this part we are asked about $\Pr[D_2|D]$, the probability that this second light bulb (coming from the same machine as the first one) will be defective given that the first bulb was defective. By the law of total probability: *4pts*

$$\Pr[D_2|D] = \Pr[D_2|A, D]\Pr[A|D] + \Pr[D_2|B, D]\Pr[B|D] + \Pr[D_2|C, D]\Pr[C|D]$$

Note that $\Pr[A|D]$, $\Pr[B|D]$ and $\Pr[C|D]$ were computed in part (b). Also, $\Pr[D_2|A, D] = \Pr[D_2|A] = \Pr[D|A]$ as, conditioning on the event that the first light bulb was produced by machine A gives us that the probability of the second light bulb being defective is only dependent on the machine that produced it (similarly for B and C). Plugging in all the terms, we obtain that $\Pr[D_2|D] = 0.0475$. [Note that the probability of being defective has increased (compared with $\Pr[D] = 0.04$) as a result of having already observed a defective light bulb that came from the same machine.]

Many people said that D_2 *and* D *were independent, but this is not true. If the second light bulb were produced by a random machine (rather than by the same as the first one), then the events would be independent, but the information that a previous bulb from the same machine was defective gives us information about the second bulb.*

5. Random Networking [10 points]

(a) The random variables $X_{\{i,j\}}$ are **not** mutually independent. The r.v.'s are mutually independent if and only if conditioning on the values of any subset of the $X_{\{i,j\}}$ does not affect the distribution of any other $X_{\{i,j\}}$. But here we know that if people 1 and 2 get the same color, and also 1 and 3 get the same color, then 2 and 3 for sure also get the same color. In other words the conditional probability *3pts*

$$\Pr[X_{\{2,3\}} = 1 \mid (X_{\{1,2\}} = 1) \cap (X_{\{1,3\}} = 1)] = 1,$$

while the unconditional probability of the same event is $\Pr[X_{\{2,3\}} = 1] = \frac{1}{m}$, which is not the same.

Most people got this part wrong. The most common error was to claim that the r.v.'s are independent because the hat colors are assigned independently. It is true that the hat colors themselves are mutually independent, but the r.v.'s $X_{\{i,j\}}$ are **not** *the same as the hat colors, so they do not automatically inherit their independence properties. Another very common error was to say that the r.v.'s $X_{\{i,j\}}$ and $X_{\{k,l\}}$ are independent for any two distinct pairs $\{i, j\}$ and $\{k, l\}$, because if we know whether i, j get the same color then this does not affect the probability that some other pair k, l gets the same color (even if $i = k$ or $j = l$). This is true, but it is not enough for mutual independence: for mutual independence, we have to be able to condition on any subset, not just on one other event. Another way of saying this is that the r.v.'s $X_{\{i,j\}}$ are* pairwise independent *(i.e., any two of them are independent) but not mutually independent. For more on this point, see Q1 of HW10. Finally, some people simply stated that the $X_{\{i,j\}}$ are not mutually independent but gave either no justification or an incorrect justification. Such solutions did not receive any credit since it was impossible to distinguish them from random guessing.*

(b) Note that $X = \sum_{i<j} X_{\{i,j\}}$. Also, since each $X_{\{i,j\}}$ is an indicator r.v., we have *3pts*

$$\mathrm{E}(X_{\{i,j\}}) = \Pr[X_{\{i,j\}} = 1] = \Pr[i, j \text{ get the same color hat}] = \frac{1}{m}.$$

By linearity of expectation, we have

$$\mathrm{E}(X) = \sum_{i<j} \mathrm{E}(X_{\{i,j\}}) = \binom{n}{2}\frac{1}{m},$$

since there are $\binom{n}{2}$ pairs $\{i, j\}$.

Many people got this right. The most common error was to say that X has a binomial distribution with $\binom{n}{2}$ coin tosses and Heads probability $p = \frac{1}{m}$. While this gives the correct value for the expectation, it is a seriously wrong derivation because the $X_{\{i,j\}}$ are **not** *independent (as would be required for a binomial distribution). As the above derivation shows, however, all we need is linearity of expectation, and this holds regardless of whether the $X_{\{i,j\}}$ are independent. Another common error was to say that $\Pr[X_{\{i,j\}} = 1] = \frac{1}{m^2}$, because (say) the probability that both i and j get the color blue is $\frac{1}{m} \times \frac{1}{m} = \frac{1}{m^2}$; but this does not take account of the fact that there are m possibilities for the color, so we have to multiply this probability by m, getting $\frac{1}{m}$ as in the above argument. Yet another error made by several students was to write $X = \binom{Z_1}{2} + \binom{Z_2}{2} + \ldots + \binom{Z_m}{2}$, where Z_k is the number of people who receive hats with color k; this is OK, but the problem is that it's not so easy to compute $\mathrm{E}(\binom{Z_k}{2})$. (It's easy to compute $\mathrm{E}(Z_k) = \frac{n}{m}$, but it's* **not** *true that $\mathrm{E}(\binom{Z_k}{2}) = \binom{\mathrm{E}(Z_k)}{2} = \binom{n/m}{2}$ as several people claimed.) Partial credit was awarded for mentioning linearity of expectation, for computing $\Pr[X_{\{i,j\}} = 1]$, and for computing the number of pairs, provided the overall framework of the solution was evident.*

(c) The random variable Y can be written as the sum $Y = \sum_{i \neq j} Y_i$, where j is the number of Jason Nerd, and Y_i (for $i \neq j$) is the indicator r.v. of the event that person i gets the same color hat as Jason. Now the Y_i are mutually independent (knowing that some bunch of people do or do not have the same color hat as Jason does not affect that probability that some other person has the same color as Jason), and $\Pr[Y_i = 1] = \frac{1}{m}$. Hence Y is the sum of $n - 1$ independent coin tosses, each with Heads probability $\frac{1}{m}$, so Y has the binomial distribution $\mathrm{Bin}(n - 1, \frac{1}{m})$. *2pts*

Most people got this right. A common error was to write n rather than $n - 1$, forgetting that Jason himself is excluded from the sum over Y_i; we did not penalize this minor error. Students who just wrote "binomial" without specifying the parameters lost a point.

(d) From part (c), $\mathrm{E}(Y)$ is the expectation of a $\mathrm{Bin}(n-1, \frac{1}{m})$ r.v., which we know from class is $(n-1) \times \frac{1}{m}$. *2pts*

Most people got this right, modulo the minor error n vs $n - 1$ carried over from part (c).