Massachusetts Institute of Technology
6.042J/18.062J, Fall '02: Mathematics for Computer Science
Prof. Albert Meyer and Dr. Radhika Nagpal

Solutions to Quiz 2

Problem 1 (15 points). Chuck Vest is planning to set aside \$100M of MIT endowment funds in the year 2002 for student scholarships. His plan is to offer the same annual amount for student scholarships, starting as soon as the money is available in 2002 and continuing through 2102, at which point the \$100M will be exhausted.

(a) (6 points) Assuming Chuck can count on earning a 3% annual return on the funds he sets aside, how much scholarship money can he offer each year? Let $r ::= 1/1.03$, and find a closed form formula in terms of r for the annual number of dollars; you do not have to evaluate the formula. (Full credit for a formula that evaluates to the right answer. Partial credit for an incorrect formula will only be considered when it is very clearly explained.)

Solution. To pay out \$1 next year, Chuck needs $\$r$ now. To pay out \$1 in two years, he need $\$r^2$ now. To pay out \$1 in i years from now, he needs $\$r^i$. So to pay out $\$a$ annually from now (2002) until 2102 he needs

$$\sum_{i=0}^{100} ar^i = a\frac{1-r^{101}}{1-r}$$

dollars now. So for \$100M today, Chuck can offer

$$a = \frac{10^8(1-r)}{1-r^{101}}$$

dollars annually. This turns out to be slightly less than \$3.07M annually—barely more than the \$3M he could give out and still have the whole \$100M left at the end of a century. ■

(b) (9 points) Chuck would like to grant \$4M annually until 2102, but he realizes that the \$100M available is not enough to do this. So he decides to hold off on awarding scholarships for d years until the \$100M, earning 3% annually, has grown enough to award

\$4M in annual scholarships for the remaining years until 2102. How many years must Chuck wait before he can start awarding these scholarships? Find a closed form formula in terms of r for the number of years, d, that Chuck must wait; you do not have to evaluate the formula. (Full credit for a formula that evaluates to the right answer. Partial credit for an incorrect formula will only be considered when it is very clearly explained.)

Solution. Paying \$4M/year starting in d years until 2102 requires setting aside

$$\sum_{i=d}^{100} 4Mr^i$$

dollars now. This sum is

$$4M[\sum_{i=0}^{100} r^i - \sum_{i=0}^{d-1} r^i] = 4M[\frac{1-r^{101}}{1-r} - \frac{1-r^d}{1-r}] = \frac{4M}{1-r}(r^d - r^{101}).$$

So with \$100M, it is necessary and sufficient that

$$\begin{aligned}
100M \geq& \frac{4M}{1-r}(r^d - r^{101}) \\
25 \geq& \frac{r^d - r^{101}}{1-r} \\
25(1-r) \geq& r^d - r^{101} \\
25(1-r) + r^{101} \geq& r^d \\
\ln(25(1-r) + r^{101}) \geq& d \ln r, \\
\frac{\ln(25(1-r) + r^{101})}{\ln r} \leq& d. \qquad\qquad (\ln r \text{ is negative})
\end{aligned}$$

Using a calculator (not required for this Quiz), it turns out that this lower bound on d is ≈ 8.46, so Vest must wait 9 years. ■

Problem 2 (20 points). Recall that for functions f, g on the natural numbers, $\mathbb{N}$, $f = O(g)$ iff

$$\exists c \in \mathbb{N}\, \exists n_0 \in \mathbb{N}\, \forall n \geq n_0 \quad c \cdot g(n) \geq |f(n)|\,. \tag{1}$$

For each pair of functions below, determine whether $f = O(g)$ and whether $g = O(f)$. In cases where one function is O() of the other, indicate the *smallest natural number*, c, and for that smallest c, the *smallest corresponding natural number* n_0 ensuring that condition (1) applies.

(a) (6 points) $f(n) = n^2, g(n) = 3n.$

$f = O(g)$ YES NO If YES, $c =$, n_0 =

Solution. NO. ■

$g = O(f)$ YES NO If YES, $c =$, n_0 =

Solution. YES, with $c = 1$, $n_0 = 3$, which works because $3^2 = 9$, $3 \cdot 3 = 9$. ■

(b) (7 points) $f(n) = (3n - 7)/(n + 4), g(n) = 4$

$f = O(g)$ YES NO If YES, $c =$, n_0 =

Solution. YES, with $c = 1, n_0 = 0$ (because $|f(n)| < 3$). ■

$g = O(f)$ YES NO If YES, $c =$, n_0 =

Solution. YES, with $c = 2, n_0 = 15.$

Since $\lim_{n\to\infty} f(n) = 3$, the smallest possible c is 2. For $c = 2$, the smallest possible $n_0 = 15$ which follows from the requirement that $2f(n_0) \geq 4$. ■

(c) (7 points) $f(n) = 1 + (n \sin(n\pi/2))^2, g(n) = 3n$

$f = O(g)$ YES NO If yes, $c =$ n_0 =

Solution. NO, because $f(2n) = 1$, which rules out $g = O(f)$ since $g = \Theta(n)$. ■

$g = O(f)$ YES NO If yes, $c =$ n_0 =

Solution. NO, because $f(2n + 1) = n^2 + 1 \neq O(n)$ which rules out $f = O(g)$. ■

Problem 3 (20 points). Theory Hippo received a mysterious bag in the mail. When he opened the bag he found that it contained 100 6-sided dice and a note that said that 99 of the dice were fair, but one always throws a 6. Theory Hippo decides to experiment with the contents of the bag by pulling a die out of the bag at random, throwing it k times and then putting it back. Call this experiment $X(k)$.

(a) (5 points) Theory Hippo decides perform $X(10)$.

- Let a be the probability that Theory Hippo gets 10 sixes given that he pulled out a fair die. What is the value of a?

Solution.

$$a = \left(\frac{1}{6}\right)^{10} = 6^{-10}.$$

■

- Let b be the probability that Theory Hippo gets 10 sixes when he performs X(10). What is the value of b? You may (but are not required to) express b in terms of a above.

  Solution.

  $$b = \left(\frac{1}{100}\right) + \left(\frac{99}{100}\right) a$$

  ■

(b) (7 points) Theory Hippo has just performed $X(10)$ and ended up getting 10 sixes! Let c be the probability that he pulled out the unfair die, given that he got 10 sixes. What is the value of c? You may (but are not required to) express c in terms of a and/or b.

Solution. Let A be the event that Theory Hippo pulled out the unfair die. Let B be the event that Theory Hippo got 10 sixes. Then:

$$\begin{aligned}
c &= \Pr\{A \mid B\} \\
&::= \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \\
&= \frac{\Pr\{A\}\Pr\{B \mid A\}}{\Pr\{B\}} && \text{(Product Rule)} \\
&= \frac{(1/100)\cdot 1}{b} = \frac{1}{100b}.
\end{aligned}$$

■

(c) (8 points) Theory Hippo's assistant performs $X(20)$ and reports that the *last* 10 of the 20 throws were sixes. What is the probability, f, that the *first* 10 throws were also sixes? You may (but are not required to) express f in terms of a, b, and/or c.

Solution.

$$f = c + \frac{1-c}{6^{10}}.$$

The previous reasoning about the first 10 throws applies equally to the second 10 of 20, since successive throws are assumed to be independent. So the assistant has pulled out the unfair die with probability c, and a fair die with probability $1 - c$. Therefore, the probability of getting another 10 heads is

$$c + (1 - c) \cdot a = c + \frac{1 - c}{6^{10}}.$$

Another way to approach this problem is by defining the following events:

$$\begin{aligned}
B &::= \text{the first 10 throws were sixes,} \\
L &::= \text{the last 10 throws were sixes,} \\
F &::= \text{the fair dice was chosen.}
\end{aligned}$$

We just observed that

$$b ::= \Pr\{B\} = \Pr\{L\}. \tag{2}$$

Note that we cannot assume $\Pr\{B \cap L\} = \Pr\{B\}\Pr\{L\}$, that is, that B and L are independent. They are not: if you know there are 10 sixes already, it increases the likelihood that it was the unfair dice that was selected, and thus increases the chance of another 10 sixes. However, they are *conditionally* independent depending on which die was selected. That is, given that the fair die was picked, then the outcomes of the first 10 rolls are independent of the second 10:

$$\Pr\{B \cap L \mid F\} = \Pr\{B \mid F\} \cdot \Pr\{L \mid F\} \tag{3}$$

Likewise, given that the unfair die was picked, the outcomes of the first 10 and second 10 rolls are *vacuously* independent, since in this case every outcome has probability 0 or 1.

$$\Pr\left\{B \cap L \mid \overline{F}\right\} = \Pr\left\{B \mid \overline{F}\right\} \cdot \Pr\left\{L \mid \overline{F}\right\}. \tag{4}$$

So now we can calculate:

$$\begin{aligned}
f &::= \Pr\{B \mid L\} \\
&::= \frac{\Pr\{B \cap L\}}{\Pr\{L\}} \\
&::= \frac{\Pr\{B \cap L\}}{b} && \text{(by (2))} \\
&= \frac{\Pr\{B \cap L \mid F\} \cdot \Pr\{F\} + \Pr\left\{B \cap L \mid \overline{F}\right\} \cdot \Pr\left\{\overline{F}\right\}}{b} && \text{(Total Probability)} \\
&= \frac{\Pr\{B \mid F\} \cdot \Pr\{L \mid F\} \cdot \Pr\{F\} + \Pr\left\{B \mid \overline{F}\right\} \cdot \Pr\left\{L \mid \overline{F}\right\} \cdot \Pr\left\{\overline{F}\right\}}{b} && \text{(by (3) \& (4))} \\
&= \frac{a^2 \cdot \frac{99}{100} + 1^2 \cdot \frac{1}{100}}{b} \\
&= \frac{99a^2}{100} + \frac{1}{100b}.
\end{aligned}$$

Common errors on this problem included assuming that the first 10 rolls and last 10 rolls *are* independent without conditioning on which die was chosen (fair or unfair). Another common error was trying to apply the law of total probability directly to "$B \mid L$," leading to the

False Claim.

$$\Pr\{B \mid L\} = \Pr\{B \mid L \cap F\} \cdot \Pr\{F\} + \Pr\left\{B \mid L \cap \overline{F}\right\} \cdot \Pr\left\{\overline{F}\right\}$$

The correct claim requires that the events F and $\overline{F}$ also be conditioned on the event L:

$$\Pr\{B \mid L\} = \Pr\{B \mid L \cap F\} \cdot \Pr\{F \mid L\} + \Pr\left\{B \mid L \cap \overline{F}\right\} \cdot \Pr\left\{\overline{F} \mid L\right\}.$$

■

Problem 4 (25 points). We are going to classify different counting problems by figuring out which ones have the same formula. Here is a set of six formulas.

$$\begin{aligned} n^m&: \\ m^n&: \\ P(n,m)&: \\ C(n-1+m,m)&: \\ C(n-1+m,n)&: \\ 2^{mn}&: \end{aligned}$$

For each problem part below, write its label—(a),(b), … —on the line next to the corresponding formula above.

Solution.

$$\begin{aligned} n^m &: bd \\ m^n &: fh \\ P(n,m) &: cj \\ C(n-1+m,m) &: a \\ C(n-1+m,n) &: e \\ 2^{mn} &: gi \end{aligned}$$

■

(a) (2 points) Number of arrangements of m indistinguishable balls in n distinguishable urns.

(b) (2 points) Number of arrangements of m distinguishable balls in n distinguishable urns.

(c) (2 points) Number of m letter words from an alphabet of size n, with no letter used more than once ($m \leq n$).

(d) (2 points) Number of m letter words from an alphabet of size n, where any letter can be repeated any number of times.

(e) (2 points) Number of sequences of n indistinguishable red balls and $m - 1$ indistinguishable blue balls.

(f) (3 points) Number of matrices of size $\sqrt{n} \times \sqrt{n}$ with m possible values for each entry. (Assume n is a perfect square.)

(g) (3 points) Number of possible subsets of A, where $|A| = nm$.

(h) (3 points) Number of functions from set A to B where $|A| = n$ and $|B| = m$.

(i) (3 points) Number of relations from set A to B where $|A| = n$ and $|B| = m$.

(j) (3 points) Number of injections from set A to B where $|A| = m$ and $|B| = n$, and $m \leq n$.

Problem 5 (20 points). Let $S ::= \{0, 1, \ldots, 9\}^{90}$ be the set of all sequences of length 90 consisting of digits $\{0, 1, \ldots, 9\}$. Two sequences, $s_1, s_2 \in S$ are said to have the *same digit distribution* iff s_1 has the same of number of occurrences of digit j as s_2, for each $j \in \{0, \ldots, 9\}$.

(a) (7 points) Describe a bijection between digits distributions and arrangements of "stars & bars."

Solution. There is a bijection between digit distributions and permutations of 90 stars and 9 bars. The 9 bars divide the 90 stars into 10 blocks, with the number of stars in the ith block equal to the number of occurrences of digit i in the distribution. ■

(b) (7 points) Give a closed formula, possibly also involving factorials and binomial coefficents, for the number, g, of possible digit distributions.

Solution. the number of permutations of 90 stars and 9 bars is

$$g = \binom{99}{9}.$$

■

(c) (6 points) Let n be the smallest number such that any set of n sequences in S is sure to contain four different sequences with the same digit distribution. Give a closed formula, possibly also involving factorials and binomial coefficents, for n. You may write your formula in terms of g in part (b).

Solution. The correct answer, which a few students got and most of the staff did not, is $n = 3g - 19$.

This is a problem about the Generalized Pigeonhole Principle. Here the pigeons are the 10^{90} possible sequences of 90 digits and the holes are the g possible digit distributions. However, for any digit d, there is only one possible sequence with digit distribution specifying 90 d's. That is, each of these 10 *special* holes (digit distributions of 90 d's, for each digit d), can hold a single, unique pigeon (the sequence d^{90}). There are at least 90 pigeons eligible for assignment to each of the other $g - 10$ holes.

So we can put 3 pigeons in each of the other $g - 10$ holes and 1 pigeon in each of these special 10 holes, without putting 4 pigeons in any hole. But one more pigeon will force us to put a 4th pigeon in one of the $g-10$ holes. Hence, once we have $n = 3(g-10)+10+1 = 3g - 19$ pigeons, there will have to be 4 pigeons in some hole, and this is the minimum n with this property.

This would have been a nice problem for class or problem set, but is trickier than we intended for a Quiz. The intended answer was $n = 3g + 1$, which would be correct if we allowed the same sequence to occur several times, *viz.*, if we had asked for the smallest n ensuring that any *multiset* of n sequences had four elements with the same distribution. By the Generalized Pigeonhole Principle, in any multiset of $m \geq 3g + 1$ occurrences of sequences in S, there will be at least $\lceil m/g \rceil \geq \lceil (3g + 1)/g \rceil = 4$ occurrences of sequences with same distribution. Moreover, there is a multiset of $3g$ occurrences in which no distribution occurs 4 times, namely, choose a multiset in which each distribution occurs exactly 3 times.

We decided to leave full credit for the intended answer, since students could easily overlook the trickiness and wind up with the "wrong" $3g+1$ answer—as almost all did—even though they clearly understood the Generalized Pigeonhole Principle.

Actually, for either of these answers to be correct, it's also obviously necessary that $|S| \geq 3g - 19$ or $3g + 1$, respectively. But it's easy to check that $|S| = 10^{90}$ is indeed much larger than $3C(99, 9) + 1$. ■