Solution

Week 16 (12/30/02)

Letters in envelopes

First Solution: (This solution is due to Aravi Samuel.) We will use induction on N. Let B_N denote the number of "bad" configurations where none of the N letters end up in the correct envelope. We claim that $B_{N+1} = N(B_N + B_{N-1})$. This can be seen as follows. In proceeding inductively from N to $N+1$ letters, there are two possible ways we can generate bad configurations:

- Given a bad configuration with N letters, we can create a bad configuration with $N+1$ letters by simply placing down the $(N+1)$st letter in its envelope, and then trading that letter with any of the other N letters. This provides us with NB_N bad configurations.
- We can also create a bad configuration with $N+1$ letters by taking a configuration of N letters where exactly one letter is in the correct envelope (there are NB_{N-1} such configurations) and then trading that letter with the $(N+1)$st letter. This provides us with NB_{N-1} bad configurations.

We therefore see that $B_{N+1} = N(B_N + B_{N-1})$. The probability of obtaining a bad configuration with N letters is $P_N = B_N/N!$. Hence, $B_N = N!\,P_N$, and so

$$\begin{aligned} (N+1)!\,P_{N+1} &= N\Big(N!\,P_N + (N-1)!\,P_{N-1}\Big) \\ \Longrightarrow \qquad (N+1)P_{N+1} &= NP_N + P_{N-1}. \end{aligned} \tag{1}$$

To solve this recursion relation, we can write it in the more suggestive form,

$$P_{N+1} - P_N = -\frac{1}{N+1}\Big(P_N - P_{N-1}\Big). \tag{2}$$

Since $P_1 = 0$ and $P_2 = 1/2$, we have $P_2 - P_1 = 1/2$. We then find inductively that $P_k - P_{k-1} = (-1)^k/k!\,$. Therefore,

$$\begin{aligned} P_N &= P_1 + \sum_{k=2}^{N} (P_k - P_{k-1}) \\ &= 1 - 1 + \sum_{k=2}^{N} \frac{(-1)^k}{k!} \\ &= \sum_{k=0}^{N} \frac{(-1)^k}{k!}\,. \end{aligned} \tag{3}$$

This is simply the partial series expansion for e^{-1}. So for large N, it approaches $1/e \approx 37\%$. This series expansion for $1/e$ converges very rapidly, so N does not have to be very large for the approximation $P_N \approx 1/e$ to be valid. For example, if $N = 5$ we have $P_5 - 1/e \approx 0.001$.

Remark: This $1/e$ result in the large-N limit can also be seen in the following way. The probability that a given letter does not end up in its corresponding envelope is $1 - 1/N$.

Therefore, if we ignore the fact that the placements of the letters are related (they are related because two letters cannot end up in the same envelope), then the probability that no letter ends up in the correct envelope is

$$\left(1 - \frac{1}{N}\right)^N \approx \frac{1}{e}\,. \tag{4}$$

It is not obvious that the correlations between the letters can be neglected here, but in view of the above result, this is apparently the case.

Second Solution: Let P_N be the probability that none of the N letters end up in the correct envelope. Let L_i and E_i denote the ith letter and corresponding envelope, respectively.

Consider a given letter, L_{a_1}, and assume that no letter ends up in the correct envelope. Then L_{a_1} must end up in some E_{a_2}, with $a_2 \neq a_1$. L_{a_2} will then end up in some E_{a_3}. L_{a_3} will then end up in some E_{a_4}, and so on. Eventually, one of the envelopes in this chain must be E_{a_1}. Let it be $E_{a_{n+1}}$. We may describe this situation by saying that L_{a_1} belongs to a "loop" of length n. If no letter ends up in the correct envelope, then n can be any number from 2 to N.

Claim: *The probability that the loop containing L_{a_1} has length n is equal to $1/N$, independent of n.*

Proof: L_{a_1} has an $(N-1)/N$ probability of ending up in some E_{a_2}, with $a_2 \neq a_1$. L_{a_2} then has an $(N-2)/(N-1)$ probability of ending up in some E_{a_3}, with $a_3 \neq a_2, a_1$. This continues until $L_{a_{n-1}}$ then has an $(N-(n-1))/(N-(n-2))$ probability of ending up in some E_{a_n}, with $a_n \neq a_{n-1}, \cdots, a_2, a_1$. Finally, L_{a_n} has a $1/(N-(n-1))$ probability of ending up in $E_{a_{n+1}} = E_{a_1}$. The probability that L_{a_1} belongs to a loop of length n is therefore equal to

$$\left(\frac{N-1}{N}\right)\left(\frac{N-2}{N-1}\right)\cdots\left(\frac{N-(n-1)}{N-(n-2)}\right)\left(\frac{1}{N-(n-1)}\right) = \frac{1}{N}\,. \quad \blacksquare \tag{5}$$

Given that a loop of length n is formed, which happens with probability $1/N$, the probability that all the $N-n$ other letters end up in the wrong envelopes is simply P_{N-n}. We therefore arrive at the relation,

$$P_N = \frac{1}{N}\Big(P_{N-2} + P_{N-3} + \cdots + P_1 + P_0\Big). \tag{6}$$

There is no P_{N-1} term in this equation, because a loop of length 1 would mean that L_{a_1} went into E_{a_1}. Note that $P_1 = 0$, and that $P_0 = 1$ here.

Multiplying eq. (6) through by N, and then subtracting the analogous equation for P_{N-1} (after multiplying through by $N-1$), gives

$$\begin{aligned} NP_N - (N-1)P_{N-1} &= P_{N-2} \\ \Longrightarrow \quad P_N - P_{N-1} &= -\frac{1}{N}\left(P_{N-1} - P_{N-2}\right). \end{aligned} \tag{7}$$

This is the same as eq. (2), with $N+1$ replaced by N. The solution proceeds as above.

Third Solution: We will find P_N by counting the number of cases that have no letter in the correct envelope, and then dividing this by the total number of possible arrangements, $N!$.

We may count these cases in the following manner. There are $N!$ total combinations. To count the number that have no letter in the correct envelope, we must subtract from $N!$ the number of combinations with, for example, (at least) L_1 in the correct envelope; there are $(N-1)!$ of these combinations. Likewise for the situations where another letter is in the correct envelope. So there seem to be $N! - N(N-1)!$ combinations with no letter in the correct envelope. However, we have double-counted some of the cases. For example, a combination which has (at least) L_1 and L_2 in the correct envelopes has been subtracted twice; there are $(N-2)!$ of these. Likewise for all the other pairs of letters. So we must add on $\binom{N}{2}(N-2)!$ combinations. But now a combination which has (at least) L_1, L_2, and L_3 in the correct envelopes has not been counted at all (because we have subtracted it off three times, and then added it on three times); there are $(N-3)!$ of these. Likewise for the other triplets. So we must subtract off $\binom{N}{3}(N-3)!$ combinations. Now, however, the combinations with (at least) L_1, L_2, L_3, and L_4 in the correct envelopes have been counted $-\binom{4}{1}+\binom{4}{2}-\binom{4}{3}=-2$ times. Likewise for the other quadruplets. So we must add on $\binom{N}{4}(N-4)!$ combinations.

In general, if we have done this procedure up to $(k-1)$-tuplets, then a combination having (at least) k letters in the correct envelopes has been counted T times, where

$$T = -\binom{k}{1} + \binom{k}{2} - \cdots + (-1)^{k-1}\binom{k}{k-1}. \tag{8}$$

However, the binomial expansion gives

$$\begin{aligned} 0 &= (1-1)^k \\ &= 1 - \binom{k}{1} + \binom{k}{2} + \cdots + (-1)^{k-1}\binom{k}{k-1} + (-1)^k \\ &= 1 + T + (-1)^k. \end{aligned} \tag{9}$$

Therefore, $T = -2$ for even k, and $T = 0$ for odd k. So we have either undercounted by one, or overcounted by one. Hence, the total number of combinations having no letter in the correct envelope is

$$N! - \binom{N}{1}(N-1)! + \binom{N}{2}(N-2)! + \cdots = \sum_{k=0}^{N} \frac{(-1)^k N!}{k!}. \tag{10}$$

To obtain the probability, P_N, that no letter is in the correct envelope, we must divide this result by $N!$. Therefore,

$$P_N = \sum_{k=0}^{N} \frac{(-1)^k}{k!}. \tag{11}$$

Remarks:

1. What is the probability (call it P_N^l) that exactly l out of the N letters end up in the correct envelopes? (With this notation, P_N^0 equals the P_N from above.) We can find P_N^l as follows.

   The probability that a given set of l letters goes into the correct envelopes is $1/\big(N(N-1)(N-2)\cdots(N-l+1)\big)$. The probability that the remaining $N-l$ letters all go into the wrong envelopes is P_{N-l}^0. This situation can happen in $\binom{N}{l}$ ways. Therefore,

$$\begin{aligned} P_N^l &= \frac{\binom{N}{l}}{N(N-1)\cdots(N-l+1)}\, P_{N-l}^0 \\ &= \frac{1}{l!} P_{N-l}. \end{aligned} \tag{12}$$

   Hence, using eq. (3),

$$P_N^l = \frac{1}{l!}\sum_{k=0}^{N-l}\frac{(-1)^k}{k!}\,. \tag{13}$$

   For large N, we have $P_N^l \approx 1/(l!e)$. The fact that this falls off so rapidly with l means that we are essentially guaranteed of having just a few letters in the correct envelopes. For example, we find (for large N) that the probability of having four or fewer letters in the correct envelopes is about 99.7%.

2. It is interesting to note that the equality, $P_N^l = \frac{1}{l!}P_{N-l}$, may directly yield the large-N result, $P_N \approx 1/e$, without having to go through all the work of the original problem. To see this, note that

$$1 = \sum_{l=0}^{N} P_N^l = \sum_{l=0}^{N}\frac{1}{l!}P_{N-l}. \tag{14}$$

   Since the terms with small l values dominate this sum, we may (for large N) replace the P_{N-l} values with $\lim_{M\to\infty} P_M$. Hence,

$$1 \approx \sum_{l=0}^{N}\frac{1}{l!}\Big(\lim_{M\to\infty} P_M\Big). \tag{15}$$

   Therefore,

$$\begin{aligned} \lim_{M\to\infty} P_M &\approx \left(\sum_{l=0}^{N}\frac{1}{l!}\right)^{-1} \\ &\approx \frac{1}{e}\,. \end{aligned} \tag{16}$$

3. Let's check that the P_N^l given by eq. (13) do indeed satisfy $\sum_{l=0}^{N} P_N^l = 1$. This may be done as follows:

$$\begin{aligned} \sum_{l=0}^{N} P_N^l &= \sum_{l=0}^{N}\sum_{k=0}^{N-l}\frac{1}{l!}\frac{(-1)^k}{k!} \\ &= \sum_{l=0}^{N}\sum_{s=l}^{N}\frac{1}{l!}\frac{(-1)^{s-l}}{(s-l)!} \qquad \text{(with } s = l+k\text{)} \\ &= \sum_{s=0}^{N}\sum_{l=0}^{s}\frac{1}{l!}\frac{(-1)^{s-l}}{(s-l)!} \qquad \text{(rewriting the limits in the } s,l \text{ plane)} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{s=0}^{N} \frac{1}{s!} \sum_{l=0}^{s} \frac{s!}{l!(s-l)!}(-1)^{s-l} \\
&= \sum_{s=0}^{N} \frac{1}{s!}(1-1)^{s} \\
&= 1 \qquad \text{(only } s=0 \text{ contributes).} \qquad (17)
\end{aligned}
$$

4. What is the average number, A, of letters in the correct envelopes? If the setup of the problem is repeated many times, then the average number of times a given letter ends up in the correct envelope is $1/N$. Since there are N letters, the average total number of correct envelopes is $N(1/N) = 1$.

   You can check that the expression for P_N^l in eq. (13) leads to $A = 1$. For finite N, the sum gets a little messy, but the result in eq. (17) will help simplify things a bit if you want to work it out.

   For large N, where we have $P_N^l \approx 1/(l!e)$, the sum is easy, and we obtain

$$
A = \sum_{l=0}^{N} l P_N^l \approx \frac{1}{e} \sum_{l=1}^{N} \frac{1}{(l-1)!} \approx 1. \qquad (18)
$$