Algorithms M2–IF TD 3

October 6, 2021

1 Blood tests

[MU 2.25] We want to perform a blood test to detect a virus. We have a test with perfect accuracy: if we perform the test on a blood sample that contains the virus (even in small amounts), the test always detects the virus; on the other hand the test never gives false positives (i.e. if the virus is not there, it says that it's not there).

The only problem now is that the test is expensive. We have n people that we wish to test, but running the test on each one costs too much. So we consider the following scheme: we take a group of k people and construct a blood sample by mixing the samples of these k people; then we test this mixed sample. If the test does not detect the disease, all of the k people are healthy. If it does, we must check each of them individually.

Assume that each person has probability p of having the virus, independently of the others.

1. For a specific group of k people, what is the probability that the test on their mixed sample will be positive?

2. What is the expected number of tests we will perform?

3. For which values of p is it better to forget about this scheme and simply test everyone from the beginning?

Solution:

1. Let us calculate the probability that the test is negative, that is, all k people are healthy. This is equal to $Pr[k \text{ people healthy}] = (1-p)^k$.

2. For a specific group of k people the expected number of test we will run is $1 \cdot (1-p)^k + (k+1)(1-(1-p)^k)$, because if one person is sick, we will test everyone individually (so we will run $k+1$ tests in total). This expectation is $1 + k(1-(1-p)^k)$. Since we have n/k groups the total expectation is $n/k + n(1-(1-p)^k)$.

3. From the calculation above we expect to perform $n+n/k-n(1-p)^k$ tests. Since we could simply have performed n tests to begin with, this scheme is more efficient when $\frac{1}{k} \leq (1-p)^k$. So we must have $p \leq 1 - \frac{1}{k^{1/k}}$. For example, for $k = 2$, we must have $p \leq 1 - \frac{1}{\sqrt{2}}$.

2 More random coins

[MU 3.22] We flip n random coins. Observe that there are $m = \binom{n}{2}$ pairs of coin flips in this experiment. For each pair of coin flips we define a random variable Y_i which is equal to 1 if the two flips of this pair gave different outcomes (so, Y_i is the exclusive or of the two random bits of the pair). Let $Y = \sum_{i=1}^{m} Y_i$.

1. Show that $Pr[Y_i = 0] = Pr[Y_i = 1] = 1/2$.
2. Show that the Y_i are not mutually independent.
3. Show that the Y_i are pair-wise independent and satisfy the property $E[Y_iY_j] = E[Y_i]E[Y_j]$.
4. Find $Var[Y]$.
5. Use Chebyshev's inequality to upper bound $Pr[|Y - E[Y]| \geq n]$.

Solution:

1. Let (a, b) be the i-th pair. Then $Pr[Y_i = 0] = Pr[X_a = 0 \wedge X_b = 0] + Pr[X_a = 1 \wedge X_b = 1] = 1/4 + 1/4 = 1/2$.
2. Consider the pairs $(a, b), (a, c), (b, c)$. (See also TD1)
3. It suffices to show that $Pr[Y_i = 1 \wedge Y_j = 1] = 1/4$. This is clear if the i-th and j-th pair do not share coin flips. If $i = (a, b)$ and $j = (a, c)$ then $Pr[Y_i = 1 \wedge Y_j = 1] = \frac{1}{2}Pr[Y_i = 1 \wedge Y_j = 1 \mid X_a = 1] + \frac{1}{2}Pr[Y_i = 1 \wedge Y_j = 1 \mid X_a = 0] = \frac{1}{2}Pr[X_b = 1 \wedge X_c = 1] + \frac{1}{2}Pr[X_b = 0 \wedge X_c = 0] = 1/4$.
4. We have $E[Y] = mE[Y_i] = m/2$. Now,

$$
\begin{aligned}
E[Y^2] &= E[(\sum_{i=1}^{m} Y_i)^2] = \\
&= \sum_{i=1}^{m} E[Y_i^2] + \sum_{i=1}^{m} \sum_{j \neq i} E[Y_iY_j] = \\
&= m/2 + m(m-1)/4 = m^2/4 + m/4
\end{aligned}
$$

Therefore, $Var[Y] = E[Y^2] - (E[Y])^2 = m/4$. Note that $Var[Y] = \sum_{i=1}^{m} Var[Y_i]$, which is logical since we proved that the Y_i are pairwise independent.

5. With Chebyshev's inequality, this probability is at most $\frac{Var[Y]}{n^2} = \frac{m/4}{n^2} = \frac{n(n-1)}{8n^2} \leq 1/8$.