Math for CS 2015/2019 Problem Set 12 solutions

https://github.com/spamegg1

June 10, 2022

Contents

1 Problem 1 **1**
1.1 (a) . . . 1
1.2 (b) . . . 3
1.3 (c) . . . 4

2 Problem 2 **6**
2.1 (a) . . . 6
2.2 (b) . . . 7
2.3 (c) . . . 8
2.4 (d) . . . 8

3 Problem 3 **9**
3.1 (a) . . . 9
3.2 (b) . . . 9
3.3 (c) . . . 10

1 Problem 1

Let R, S, and T be random variables with the same codomain, V.

1.1 (a)

Suppose R is uniform, that is

$$Pr[R = b] = \frac{1}{|V|}$$

for all $b \in V$; and suppose R is independent of S. Originally this text had the following argument:

The probability that $R = S$ is the same as the probability that R takes whatever value S happens to have, therefore

$$Pr[R = S] = \frac{1}{|V|}$$

Are you convinced by this argument? We decided to replace it by a reference to this problem. We'd like your advice on whether it should be put back in the text. Before advising us, write out a careful proof of $Pr[R = S] = \frac{1}{|V|}$.

Hint: The event $[R = S]$ is a disjoint union of events

$$[R = S] = \bigcup_{b \in V} [R = b \text{ AND } S = b]$$

Proof. Let's follow the hint.

1. The event $[R = S]$ is a disjoint union of events

$$[R = S] = \bigcup_{b \in V} [R = b \text{ AND } S = b]$$

2. By the Sum rule for disjoint events, from (1) we get

$$Pr[R = S] = \sum_{b \in V} Pr[R = b \text{ AND } S = b]$$

3. Since R and S are independent, by definition of independence of random variables (18.2, page 741) for all $b_1, b_2 \in V$, the events $[R = b_1]$ and $[S = b_2]$ are independent.

4. By (3) and the alternate formulation of independence of events (Theorem 17.7.2), we have, for all $b \in V$:

$$Pr[R = b \text{ AND } S = b] = Pr[R = b] \cdot Pr[S = b]$$

5. By (2) and (4) we get

$$Pr[R = S] = \sum_{b \in V} Pr[R = b] \cdot Pr[S = b]$$

6. Since R is uniform, by (5) we get

$$Pr[R = S] = \sum_{b \in V} \frac{1}{|V|} Pr[S = b]$$

7. Notice that $1/|V|$ is constant, so it can be pulled out in front of the summation. By (6) we have

$$Pr[R = S] = \frac{1}{|V|} \sum_{b \in V} Pr[S = b]$$

8. Notice that $\sum_{b \in V} Pr[S = b] = 1$ (by definition of a probability space). Therefore

$$Pr[R = S] = \frac{1}{|V|}$$

□

1.2 (b)

Let $S \times T$ be the random variable giving the values of S and T, that is,

$$S \times T : \mathcal{S} \to V \times V$$

where

$$(S \times T)(\omega) ::= (S(\omega), T(\omega))$$

for every outcome $\omega \in \mathcal{S}$. Now suppose R has a uniform distribution, and R is independent of $S \times T$.

How about this argument?

The probability that $R = S$ is the same as the probability that R equals the first coordinate of whatever value $S \times T$ happens to have, and this probability remains equal to $1/|V|$ by independence. Therefore the event $[R = S]$ is independent of $[S = T]$.

Write out a careful proof that the event $[R = S]$ is independent of the event $[S = T]$.

Proof. 1. Since R is independent of $S \times T$, by definition of independence of random variables (section 18.2, page 741) for all $b \in V$ and $(b_1, b_2) \in V \times V$ the two events

$$[R = b] \text{ and } [(S, T) = (b_1, b_2)]$$

are independent.

2. Notice that $(S, T) = (b_1, b_2)$ iff $S = b_1$ and $T = b_2$. So we can rewrite (1) as: for all $b \in V$ and $(b_1, b_2) \in V \times V$, the two events

$$[R = b] \text{ and } [S = b_1 \text{ and } T = b_2]$$

are independent.

3. In particular, by (2) it holds that for all $b \in V$, the two events

$$[R = b] \text{ and } [S = b \text{ and } T = b]$$

are independent.

4. By (3) and the alternate formulation of independence of events (Theorem 17.7.2), we have for all $b \in V$:

$$Pr[R = b \text{ and } (S = b \text{ and } T = b)] = Pr[R = b] \cdot Pr[S = b \text{ and } T = b]$$

5. Summing the equation in (4) over all $b \in V$ we get

$$\sum_{b \in V} Pr[R = b \text{ and } (S = b \text{ and } T = b)] = \sum_{b \in V} Pr[R = b] \cdot Pr[S = b \text{ and } T = b]$$

6. Notice that, just like the hint in part (a), because of the disjoint union of events again, by the Sum rule, the LHS is $Pr[R = S \text{ and } S = T]$. Also, since R has uniform distribution, $Pr[R = b] = 1/|V|$ for all $b \in V$.

7. By (5) and (6)

$$Pr[R = S \text{ and } S = T] = \sum_{b \in V} \frac{1}{|V|} \cdot Pr[S = b \text{ and } T = b]$$

8. Moving the constant outside the sum, by (7) we have

$$Pr[R = S \text{ and } S = T] = \frac{1}{|V|} \sum_{b \in V} Pr[S = b \text{ and } T = b]$$

9. By part (a), $Pr[R = S] = 1/|V|$. So

$$Pr[R = S \text{ and } S = T] = Pr[R = S] \sum_{b \in V} Pr[S = b \text{ and } T = b]$$

10. Once again by the disjoint union of events, the sum on the RHS of (9) is $Pr[S = T]$. Therefore

$$Pr[R = S \text{ and } S = T] = Pr[R = S] \cdot Pr[S = T]$$

11. By (10) and the alternate formulation of independence of events again, the event $[R = S]$ is independent of the event $[S = T]$. □

1.3 (c)

Let $V = \{1, 2, 3\}$ and (R, S, T) take following triples of values with equal probability

$$(1, 1, 1), (2, 1, 1), (1, 2, 3), (2, 2, 3), (1, 3, 2), (2, 3, 2)$$

Verify that

1. R is independent of $S \times T$,
2. The event $[R = S]$ is not independent of $[S = T]$,
3. S and T have a uniform distribution.

1.

Proof. We need to show that for all $b \in V$ and all $(b_1, b_2) \in V \times V$, the events $[R = b]$ and $[S \times T = (b_1, b_2)]$ are independent.

R only takes the values 1 and 2; whereas $S \times T$ only takes the values $(1, 1), (2, 3), (3, 2)$.

So we need to check the events $[R = 1]$ and $[R = 2]$ against every one of the events $[S \times T = (1, 1)]$, $[S \times T = (2, 3)]$ and $[S \times T = (3, 2)]$.

To prove independence of events, we will use the alternate formulation

$$Pr[A \text{ and } B] = Pr[A] \cdot Pr[B]$$

So we need to verify 6 equations of the form

$$Pr[R = b \text{ and } S \times T = (b_1, b_2)] = Pr[R = b] \cdot Pr[S \times T = (b_1, b_2)]$$

First let's calculate all the relevant probabilities. Go over the info above, and verify the calculations below one by one.

$Pr[R = 1] = 1/2$, $Pr[R = 2] = 1/2$.

$Pr[S \times T = (1, 1)] = 1/3$, $Pr[S \times T = (2, 3)] = 1/3$, $Pr[S \times T = (3, 2)] = 1/3$.

$Pr[R = 1 \text{ and } S \times T = (1, 1)] = 1/6$, $Pr[R = 2 \text{ and } S \times T = (1, 1)] = 1/6$.

$Pr[R = 1 \text{ and } S \times T = (2, 3)] = 1/6$, $Pr[R = 2 \text{ and } S \times T = (2, 3)] = 1/6$.

$Pr[R = 1 \text{ and } S \times T = (3, 2)] = 1/6$, $Pr[R = 2 \text{ and } S \times T = (3, 2)] = 1/6$.

From these calculations we can see that all 6 equations are verified:

$$Pr[R = 1 \text{ and } S \times T = (1, 1)] = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = Pr[R = 1] \cdot Pr[S \times T = (1, 1)]$$

$$Pr[R = 2 \text{ and } S \times T = (1, 1)] = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = Pr[R = 2] \cdot Pr[S \times T = (1, 1)]$$

$$Pr[R = 1 \text{ and } S \times T = (2, 3)] = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = Pr[R = 1] \cdot Pr[S \times T = (2, 3)]$$

$$Pr[R = 2 \text{ and } S \times T = (2, 3)] = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = Pr[R = 2] \cdot Pr[S \times T = (2, 3)]$$

$$Pr[R = 1 \text{ and } S \times T = (3, 2)] = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = Pr[R = 1] \cdot Pr[S \times T = (3, 2)]$$

$$Pr[R = 2 \text{ and } S \times T = (3, 2)] = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = Pr[R = 2] \cdot Pr[S \times T = (3, 2)]$$

This proves that R is independent of $S \times T$. □

2.

Proof. To show that the event $[R = S]$ is not independent of the event $[S = T]$, we need to show

$$Pr[R = S \text{ and } S = T] \neq Pr[R = S] \cdot Pr[S = T]$$

If we look at the data again, the only case where $R = S$ and $S = T$ simultaneously, is the triple $(1, 1, 1)$, so

$$Pr[R = S \text{ and } S = T] = 1/6$$

Again looking at the data, we see that $R = S$ for the triples $(1, 1, 1)$ and $(2, 2, 3)$. So

$$Pr[R = S] = 1/3$$

Again looking at the data, we see that $S = T$ for the triples $(1, 1, 1)$ and $(2, 1, 1)$. So

$$Pr[S = T] = 1/3$$

Putting it all together, we get

$$Pr[R = S \text{ and } S = T] = \frac{1}{6} \neq \frac{1}{3} \cdot \frac{1}{3} = Pr[R = S] \cdot Pr[S = T]$$

□

3.

Proof. Looking at the data we see that S takes each of the values 1,2,3 exactly twice, so $Pr[S = 1] = Pr[S = 2] = Pr[S = 3] = 1/3$, which shows S is uniformly distributed.

The same goes for T. □

2 Problem 2

(A true story from World War Two.)

The army needs to test n soldiers for a disease. There is a blood test that accurately determines when a blood sample contains blood from a diseased soldier. The army presumes, based on experience, that the fraction of soldiers with the disease is equal to some small number p.

Approach (1) is to test blood from each soldier individually, this requires n tests. Approach (2) is to randomly group the soldiers into g groups of k soldiers, where $n = gk$. For each group, blend the k blood samples of the people in the group, and test the blended sample. If the group-blend is free of the disease, we are done with that group after one test. If the group-blend tests positive for the disease, then someone in the group has the disease, and we test all the people in the group for a total of $k + 1$ tests on that group.

Since the groups are chosen randomly, each soldier in the group has the disease with probability p, and it is safe to assume that whether one soldier has the disease is independent of whether the others do.

2.1 (a)

What is the expected number of tests in Approach (2) as a function of the number of soldiers n, the disease fraction p, and the group size k?

Proof. 1. For $1 \leq i \leq k$, consider the ith group of k soldiers.

Consider the case when there are no soldiers in group i with the disease.

The probability of this case happening is $(1 - p)^k$ because:

there are k soldiers in the group, and

each soldier has a probability of $1 - p$ of not having the disease, and

whether one soldier has the disease is independent of the others in the group.

In this case, the blended sample test yields negative, so the number of tests done on group i is 1.

2. Consider the case when there is at least 1 soldier in group i with the disease.

The probability of this case happening is $1 - (1 - p)^k$.

In this case the blended sample test yields positive, so in this case the number of tests done on group i is $k+1$.

3. Define R_i to be the random variable that equals the number of tests done on group i. Combining (1) and (2) and using the Law of Total Expectation we have

$$Ex[R_i] = (1-p)^k \cdot 1 + (1-(1-p)^k) \cdot (k+1) = k+1-k(1-p)^k$$

4. Define R to be the random variable that equals the total number of tests on all n soldiers. Summing (3) over all the g groups of k soldiers we get

$$Ex[R] = \sum_{i=1}^{g} Ex[R_i] = \sum_{i=1}^{g} \left(k+1-k(1-p)^k\right) = g(k+1-k(1-p)^k)$$

5. Using $n = gk$ on (4) and simplifying we get

$$Ex[R] = gk + g - gk(1-p)^k = n + \frac{n}{k} - n(1-p)^k = n(1 + \frac{1}{k} - (1-p)^k)$$

□

2.2 (b)

Show how to choose k so that the expected number of tests using Approach (2) is approximately $n\sqrt{p}$.

Hint: Since p is small, you may assume that $(1-p)^k \approx 1$ and $\ln(1-p) \approx -p$.

Proof. 1. Taking derivative with respect to k we get

$$\frac{d}{dk}Ex[R] = n(0 - \frac{1}{k^2} - (1-p)^k \cdot \ln(1-p))$$

2. Set the derivative equal to 0 to find the value of k that minimizes $Ex[R]$. We use the approximations given in the hint:

$$\begin{aligned}
n(0 - \frac{1}{k^2} - (1-p)^k \cdot \ln(1-p)) &= 0 \\
-\frac{1}{k^2} - (1-p)^k \cdot \ln(1-p) &= 0 \\
-\frac{1}{k^2} - 1 \cdot (-p) &= 0 \\
p - \frac{1}{k^2} &= 0 \\
p &= \frac{1}{k^2} \\
k^2 &= \frac{1}{p} \\
k &= \pm\sqrt{\frac{1}{p}}
\end{aligned}$$

3. Since k is at least 1, and at most n, k is positive. So we get $k = \sqrt{1/p}$. This is the only critical value on the interval $[1..n]$.

4. Also notice that as k decreases to 1 (every soldier is an individual group, so every soldier gets tested), the expected number of tests increases to n.

5. Also notice that as k increases to n (everyone in one group, all blood samples blended into one) the expected value increases: the derivative is approximately $n(p - 1/k^2)$ and when $k > \sqrt{1/p}$ the derivative is positive.

6. By (3), (4) and (5) the value $k = \sqrt{1/p}$ gives us the minimum expected value. This is approximately

$$Ex[R] \approx n(1 + \frac{1}{\sqrt{1/p}} - 1) = n\sqrt{p}$$

□

2.3 (c)

(c) What fraction of the work does Approach (2) expect to save over Approach (1) in a million-strong army of whom approximately 1% are diseased?

Proof. 1. In this part we are given $n = 1000000$ and $p = 0.01$.

2. Approach 1 requires $n = 1000000$ tests.

3. Approach 2 (by choosing $k = \sqrt{1/p} = 10$ carefully as in part (b)) requires

$$n\sqrt{p} = 1000000 \cdot \sqrt{0.01} = 1000000 \cdot 0.1 = 100000$$

tests.

4. Therefore by (2) and (3) Approach 2 requires 10% of the tests Approach 1 requires. So, 90% of work is saved.

5. If we use the exact formula from part (a), instead of the approximation from part (b), we see that Approach 2 requires (keep in mind $k = 10$):

$$1000000(1 + \frac{1}{10} - (1 - 0.01)^{10}) = 1000000(1 + 0.1 - (0.99)^{10}) = 195618$$

tests. That is, 19.5% of the work required by Approach 1. So 80.5% of the work is saved. □

2.4 (d)

Can you come up with a better scheme by using multiple levels of grouping, that is, groups of groups?

Proof. No. First dividing n soldiers into g_1 groups of k_1 soldiers each (where $n = g_1 k_1$), and then further dividing each group of k_1 soldiers into g_2 groups of k_2 soldiers each (where $k_1 = g_2 k_2$) is the same as dividing the initial n soldiers into $g1 \cdot g2$ groups of k_2 soldiers each (where $n = (g_1 \cdot g_2) \cdot k_2$).

And we already solved the optimization problem of dividing n soldiers into groups of k soldiers for any possible k, and proved that $k = \sqrt{1/p}$ is (approximately) the group size that gives us the absolute minimum. □

3 Problem 3

A literal is a propositional variable or its negation. A k-clause is an OR of k literals, with no variable occurring more than once in the clause. For example,

$$P \text{ OR } \overline{Q} \text{ OR } \overline{R} \text{ OR } V$$

is a 4-clause, but

$$\overline{V} \text{ OR } \overline{Q} \text{ OR } \overline{X} \text{ OR } V$$

is not, since V appears twice.

Let S be a sequence of n distinct k-clauses involving v variables. The variables in different k-clauses may overlap or be completely different, so $k \leq v \leq nk$.

A random assignment of true/false values will be made independently to each of the v variables, with true and false assignments equally likely. Write formulas in n, k, and v in answer to the first two parts below.

3.1 (a)

What's the probability that the last k-clause in S is true under the random assignment?

Proof. A k-clause is true, unless every single literal in it is false. Thus, each k-clause is true with probability

$$1 - \left(\frac{1}{2}\right)^k$$

□

3.2 (b)

What is the expected number of true k-clauses in S?

Proof. 1. For $1 \leq i \leq n$ define T_i to be the indicator variable (so $T_i = 0$ or 1) for the event that the ith k-clause in S is true.

2. Then the number of true k-clauses in S is $T_1 + \ldots + T_n$ and by part (a) the expected number of true k-clauses in S is

$$E[T_1 + \ldots + T_n] = E[T_1] + \ldots + E[T_n] = n \cdot \left(1 - \frac{1}{2^k}\right)$$

□

3.3 (c)

A set of propositions is satisfiable iff there is an assignment to the variables that makes all of the propositions true. Use your answer to part (b) to prove that if $n < 2^k$, then S is satisfiable.

Proof. 1. The random variable $T_1 + \ldots + T_n$ cannot always be less than its expected value. So there exists at least one assignment A of true/false values to the propositional variables such that

$$T_1 + \ldots + T_n \geq n \cdot \left(1 - \frac{1}{2^k}\right)$$

2. Since $n < 2^k$, we have $n/2^k < 1$, so $-n/2^k > -1$, adding n to both sides we have

$$n - \frac{n}{2^k} > n - 1$$

3. By (1) and (2), under the assignment A we have

$$T_1 + \ldots + T_n > n - 1$$

4. Since $T_1 + \ldots + T_n$ only takes integer values, and since its maximum value is n, by (3) under A we have

$$T_1 + \ldots + T_n = n$$

5. So $T_i = 1$ for all $1 \leq i \leq n$, which means all of the k-clauses in S are satisfied under the assignment A.

5. Therefore S is satisfiable. □