

## Problem Set 12

Due: May 8

### Reading:

- *Independence*: Chapter 17.7, 17.8;
- *Random variables*: Chapter 18 through 18.3;
- *Expectation*: Chapter 18.4, 18.5 in the course textbook.

### Problem 1.

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

(a) Suppose  $R$  is uniform—that is,

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

for all  $b \in V$ —and  $R$  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

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

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 (1).

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

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

(b) Let  $S \times T$  be the random variable giving the values of  $S$  and  $T$ .<sup>1</sup> 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  $[R = S]$  is independent of  $[S = T]$ .

 2015, Eric Lehman, F Tom Leighton, [Albert R Meyer](#). This work is available under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 license](#).

<sup>1</sup>That is,  $S \times T : S \rightarrow V \times V$  where

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

for every outcome  $\omega \in S$ .

(c) Let  $V = \{1, 2, 3\}$  and  $(R, S, T)$  take the 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.

### 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 approximately 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 to 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.

(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  $k$ ?

(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  $\ln(1 - p) \approx -p$ .

(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?

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

### 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,

$$P \text{ OR } \bar{Q} \text{ OR } \bar{R} \text{ OR } V,$$

is a 4-clause, but

$$\bar{V} \text{ OR } \bar{Q} \text{ OR } \bar{X} \text{ OR } V,$$

is not, since  $V$  appears twice.

Let  $\mathcal{S}$  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.

(a) What is the probability that the last  $k$ -clause in  $\mathcal{S}$  is true under the random assignment?

(b) What is the expected number of true  $k$ -clauses in  $\mathcal{S}$ ?

(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  $\mathcal{S}$  is satisfiable.

MIT OpenCourseWare  
<https://ocw.mit.edu>

6.042J / 18.062J Mathematics for Computer Science  
Spring 2015

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.