Math 4981 Spring 2021
Cryptography HW 4 Solutions
Due Thursday, February 11

1. Let Ω be a (finite) probability space, and let E, F, G be events.

   (a) Prove that $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

   (b) Prove that $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$.

   Solution:

   (a) We can look at this two ways. One is to decompose each event into its corresponding observations. By disjointness, we know that

$$\begin{aligned} P(E) &= \sum_{\omega \in E} P(\omega) \\ P(F) &= \sum_{\omega \in F} P(\omega) = \sum_{\omega \in F \setminus E} P(\omega) + \sum_{\omega \in F \cap E} P(\omega) \\ P(E) + P(F) &= \sum_{\omega \in E} P(\omega) + \sum_{\omega \in F \setminus E} P(\omega) + \sum_{\omega \in F \cap E} P(\omega) \\ &= \sum_{\omega \in E \cup F} P(\omega) + \sum_{\omega \in E \cap F} P(\omega) \\ P(E) + P(F) &= P(E \cup F) + P(E \cap F) \end{aligned}$$

   which gives us what we wanted.

   Alternatively, we can observe that $E = (E \cap F) \cup (E \cap F^C)$ is a disjoint union. So we have

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap F^C) \\ P(F) &= P(E \cap F) + P(E^C \cap F) \\ P(E \cup F) &= P(E \cap F) + P(E^C \cap F) + P(E \cap F^C) \\ &= P(E \cap F) + (P(F) - P(E \cap F)) + (P(E) - P(E \cap F)) \\ &= P(E) + P(F) - P(E \cap F). \end{aligned}$$

   (b) We could prove this by an argument of decomposition into observations as in (a),

but it's much easier to use (a) to prove this.

$$\begin{aligned}
P(E \cup F \cup G) &= P\big((E \cup F) \cup G\big) = P(E \cup F) + P(G) - P\big((E \cup F) \cap G\big) \\
P\big((E \cup F) \cap G\big) &= P\big((E \cap G) \cup (F \cap G)\big) \\
&= P(E \cap G) + P(F \cap G) - P\big((E \cap G) \cap (F \cap G)\big) \\
&= P(E \cap G) + P(F \cap G) - P(E \cap F \cap G) \\
P(E \cup F) &= P(E) + P(F) - P(E \cap F) \\
P(E \cup F \cup G) &= P(E) + P(F) - P(E \cap F) + P(G) \\
&\quad - \big(P(E \cap G) + P(F \cap G) - P(E \cap F \cap G)\big)
\end{aligned}$$

which is what we wanted.

2. Suppose we have two boxes. Box A contains three blue balls and four red balls; box B contains five blue balls and two red balls.

   (a) We choose a box at random (50% chance of each box), and then pick a ball out of it at random. What is the probability that the ball is blue?

   (b) Suppose we choose a box at random, draw out a random ball, and the ball is blue. What is the probability that the box we chose is box A?

   Solution:

   (a) The probability is $\frac{1}{2} \cdot \frac{3}{7} + \frac{1}{2}\frac{5}{7} = \frac{4}{7}$.

   (b) Here we can use Bayes's Rule. We compute

$$P(A|\text{blue}) = \frac{P(\text{blue}|A)P(A)}{P(\text{blue})} = \frac{3/7 \cdot 1/2}{4/7} = \frac{3}{8}.$$

3. Suppose we have two boxes. Box A contains three blue balls and three red balls; box B contains six blue balls and zero red balls.

   (a) We choose a box at random and then choose a ball at random. What is the probability the ball is red?

   (b) Suppose we choose a box at random, draw out a random ball, and the ball is red. What is the probability that the box we chose is box A?

   (c) Suppose we choose a box at random, draw out a random ball, and the ball is blue. What is the probability that the box we chose is box A?

   (d) How does this relate to the slogan "absence of evidence is not evidence of absence?"

   Solution:

   (a) $\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2}\frac{0}{6} = \frac{1}{4}$.

(b) We *could* use Bayes's Rule here:

$$P(A|\text{red}) = \frac{P(\text{red}|A)P(A)}{P(\text{red})} = \frac{1/2 \cdot 1/2}{1/4} = 1.$$

But this is overkill. If we pull out a red ball the box can't possibly be B, so it must have been A.

(c) Now we really do have to use Bayes's Rule. We can compute that the probability of drawing a blue ball is 1/4 (among other things, because it's $1 - P(\text{red})$. Then

$$P(A|\text{blue}) = \frac{P(\text{blue}|A)P(A)}{P(\text{blue})} = \frac{1/2 \cdot 1/2}{3/4} = \frac{1}{3}.$$

(d) In part (b) we had evidence that we pulled from a box that contained red balls, and everything was straightforward.

In part (c), we did not have evidence that the box contained red balls. And this provides evidence that the box does not contain red balls! In general, the aphorism is false; Bayes's Theorem says that if seeing A makes B more likely, then seeing not-B makes not-A more likely.

A better aphorism might be "absence of proof is not proof of absence"; in part (c) we have no proof that the box contains red balls, but it very well might.

4. For cryptographic purposes, we often need to generate large prime numbers. In practice, we generate a random number, test whether it's prime, and repeat until one of the numbers passes our tests. But it's also difficult to determine for sure whether or not a number is prime, so we use a much faster test that has a chance of giving the wrong answer.

One specific prime test is the Miller test. Every prime number will pass the Miller test; a composite number has a $\frac{1}{4}$ chance of passing the Miller test. (It's actually less than that, but $\frac{1}{4}$ is a good upper bound.)

(a) If we pick a random 100-digit number, it has about a $\frac{1}{230}$ chance of being prime. Suppose we choose a random 100 digit number, and then use the Miller test and it passes. What are the odds that our number is prime?

(b) We can repeat the Miller test with different "bases" to get independent chances of success. Thus if we try the Miller test k times, a prime number will pass all of them, while a composite number will pass all of them with probability $\frac{1}{4^k}$. Suppose our random 100-digit number passes five Miller tests. What are the odds that it's prime?

(c) How many Miller test would our number have to pass to reach a 99% chance of being prime?

Solution:

(a) We use Bayes's Theorem:

$$\begin{aligned} P(\text{prime}|\text{pass}) &= \frac{P(\text{pass}|\text{prime})P(\text{prime})}{P(\text{pass}|\text{prime})P(\text{prime}) + P(\text{pass}|\text{composite})P(\text{composite})} \\ &= \frac{1 \cdot 1/230}{1 \cdot 1/230 + 1/4 \cdot 229/230} = \frac{4}{233} \approx .017 \end{aligned}$$

So a number that passes the Miller test has approximately a 1.7% chance of being prime.

(b)

$$P(\text{prime}|\text{pass}) = \frac{P(\text{pass}|\text{prime})P(\text{prime})}{P(\text{pass}|\text{prime})P(\text{prime}) + P(\text{pass}|\text{composite})P(\text{composite})}$$
$$= \frac{1 \cdot 1/230}{1 \cdot 1/230 + 1/4^5 \cdot 229/230} = \frac{1024}{1253} \approx .82$$

so the number has an 82% chance of being prime.

(c) Guess and check shows that we need to run 7 Miller tests to get a 98.6% chance, and 8 Miller tests gets us to 99.6%.

5. Compute the expected value of the random variable X if:

(a) The values of X are uniformly distributed on the set $\{0, 1, 2, \ldots, n-1\}$

(b) The values of X are uniformly distributed on the set $\{3, 6, 11, 15, 22\}$

(c) The values of X follow a binomial distribution with $n = 5$ trials and probability of success $p = .6$.

(d) The values of X follow a binomial distribution with n trials and probability of success p.

Solution:

(a)

$$\sum_{i=0}^{n-1} \frac{1}{n} i = \frac{1}{n} \frac{n(n-1)}{2} = \frac{n-1}{2}.$$

(b)

$$\frac{1}{5}(3 + 6 + 11 + 15 + 22) = \frac{57}{5} = 11.4.$$

(c) One (over-complicated) approach is just to brute force this. We have the formula

$$f_X(k) = \binom{5}{k} .6^k (.4)^{5-k} = \frac{5!}{k!(5-k)!} .6^k (.4)^{5-k}$$

and this gives

$$f_X(0) = \frac{5!}{0!(5)!}.6^0(.4)^5 = .4^5\frac{32}{3125} = .01024$$
$$f_X(1) = \frac{5!}{1!(4)!}.6^1(.4)^4 = 5 \cdot .6 \cdot .4^4 = .0768$$
$$f_X(2) = \frac{5!}{2!(3)!}.6^2(.4)^3 = 10 \cdot .6^2 \cdot .4^3 = .2304$$
$$f_X(3) = \frac{5!}{3!(2)!}.6^3(.4)^2 = 10 \cdot .6^3 \cdot .4^2 = .3456$$
$$f_X(4) = \frac{5!}{4!(1)!}.6^4(.4)^1 = 5 \cdot .6^4 \cdot .4 = .2592$$
$$f_X(5) = \frac{5!}{5!(0)!}.6^5(.4)^0 = .6^5\frac{243}{3125} = .07776$$

and we can compute a weighted sum

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=0}^{n} i \cdot f_X(i) \\ &= 0 \cdot .01024 + 1 \cdot .0768 + 2 \cdot .2304 + 3 \cdot .3456 + 4 \cdot .2592 + 5 \cdot .07776 \\ &= 3.\end{aligned}$$

This number works out much, much nicer than we have any right to expect, and should maybe make us think about whether this problem is secretly easier than it looks, and whether we could solve it more easily.

And once we ask that, maybe we realize: We know that each trial has a 3/5 chance of success, and there are 5 trials. So the expected value of the number of successes is $^3/_5 \cdot 5 = 3$.

(d) By the same argument, the expected value of the number of successes is pn. I'm sure there's a long and clever algebraic argument that gets that out of the binomial formula too, but this is really the right way to think of it.

6. Which of the follinwg random variables should be given a uniform distribution, and why?

   (a) X is the roll of one die.

   (b) X is the number of heads obtained in ten coin tosses.

   (c) X is the number of coin tosses it takes to get one head.

   (d) X is the birthday of a randomly chosen person.

   Solution:

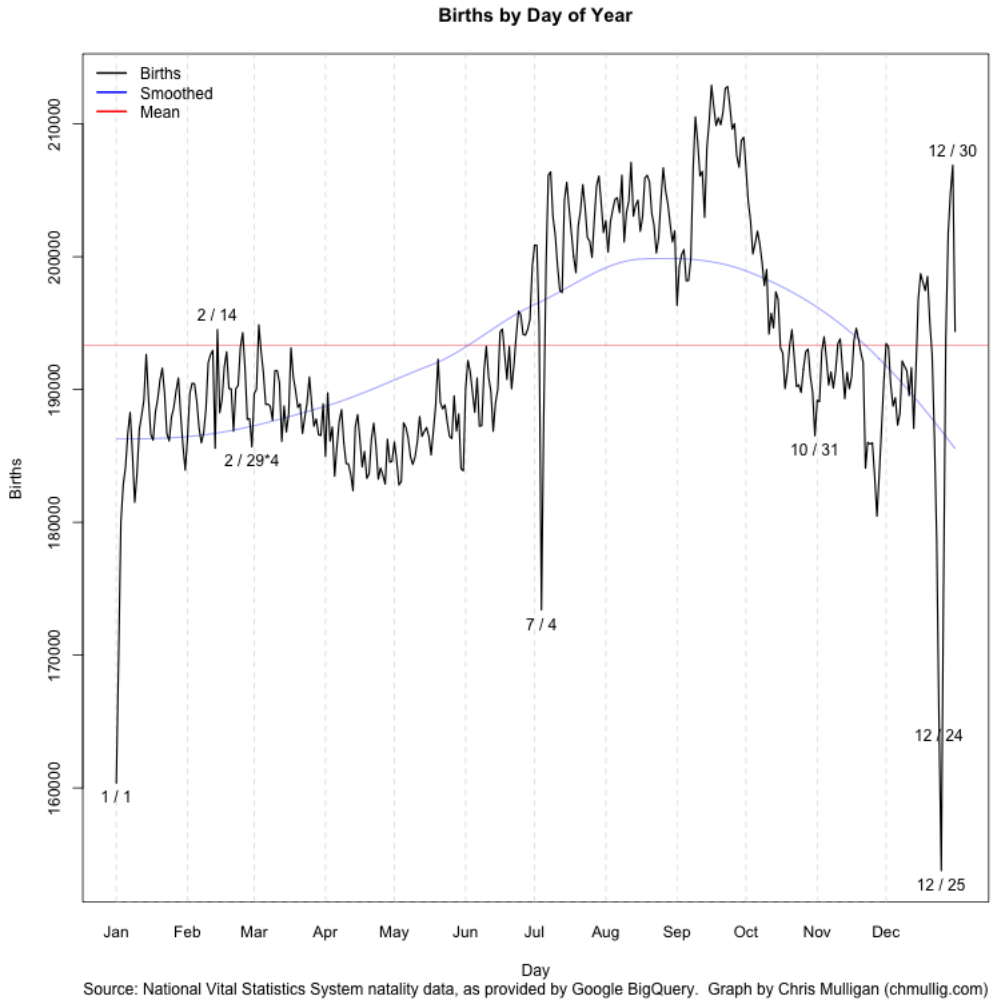
   (a) Yes: each outcome is equally likely.

   (b) No: it's more likely to get five heads than to get one.

(c) No: the mostly likely number is one, followed by two, etc.

(d) I would say no, but it depends on how you define the question! The most obvious problem is Feb 29, which is less likely than other days, but you could maybe ignore that one. More subtlely, some parts of the year generate more birthdays than others (birthdays are most common in August through October, probably corresponding to winter conceptions) and they are dramatically unlikely on major holidays (Christmas, New Year's, July 4).

Here's a picture of (roughly) the pdf of birthdays:



7. Suppose X is the random variable associated to rolling two (six-sided) dice and adding them together. Give an explicit description of the probability density function $f_X(x)$, and graph the probability density function f_X and cumulative density function F_X associated to X.

Solution: The probability density function has the formula

$$\begin{array}{lll} f_X(2) = 1/36 & f_X(3) = 1/18 & f_X(4) = 1/12 \\ f_X(5) = 1/9 & f_X(6) = 5/36 & f_X(7) = 1/6 \\ f_X(8) = 5/36 & f_X(9) = 1/9 & f_X(10) = 1/12 \\ f_X(11) = 1/18 & f_X(12) = 1/46 & \end{array}$$

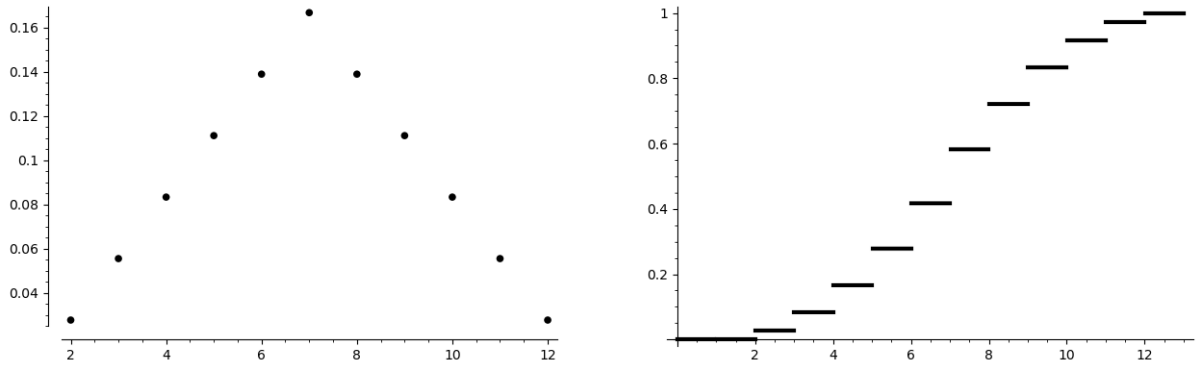


Figure 1: pdf (left) and cdf (right)

8. Consider a cipher with three keys, three plaintexts, and four ciphertexts, given by:

	m_1	m_2	m_3
k_1	c_2	c_4	c_1
k_2	c_1	c_3	c_2
k_3	c_3	c_1	c_4

Suppose all keys are equally likely, and the messages have probability $P(m_1) = 2/5, P(m_2) = 2/5, P(m_3) = 1/5$.

(a) What is the probability of each ciphertext?

(b) Compute $P(c_1|m_1), P(c_1|m_2), P(c_1|m_3)$. Can you tell if the ciphertext has perfect secrecy from this calculation?

(c) Compute $P(c_2|m_1), P(c_3|m_1), P(c_4|m_1)$. Can we combine this with the previous answer to tell if the cipher has perfect secrecy?

(d) Compute $P(k_1|c_3), P(k_2|c_3), P(k_3|c_3)$.

Solution:

(a) $P(c_1) = 1/3, P(c_2) = 1/5, P(c_3) = 4/15, P(c_4) = 1/5$.

(b) $P(c_1|m_1) = 1/3, P(c_1|m_2) = 1/3, P(c_1|m_3) = 1/3$.
This doesn't tell us if we have perfect secrecy. The probability of c_1 overall is $1/3$, and the probability for each message is $1/3$. But it doesn't prove we *don't* have perfect security either, because we'd need to check for all $c \in \mathcal{C}$ to be sure.

(c) $P(c_2|m_1) = 1/3, P(c_3|m_1) = 1/3, P(c_4|m_1) = 0$. This shows that we don't have perfect secrecy, since the ciphertext constrains the message.

(d) $P(k_1|c_3) = 0, P(k_2|c_3) = \frac{2/5 \cdot 1/3}{4/15} = 1/2, P(k_3|c_3) = \frac{2/5 \cdot 1/3}{4/15} = 1/2$.

9. We proved in class that if a cryptosystem has perfect secrecy, then $\#\mathcal{K} \geq \#\mathcal{M}$. Give an example of a cryptosystem with $\#\mathcal{K} = \#\mathcal{M} = \#\mathcal{C} = 3$ that has perfect secrecy, and show explicitly that it has perfect secrecy.

Solution:

There are many correct answers, but the simplest is probably something like this:

	m_1	m_2	m_3
k_1	c_1	c_2	c_3
k_2	c_2	c_3	c_1
k_3	c_3	c_1	c_2

where $P(m_i) = P(k_i) = 1/3$. Then we can check that $P(c_i) = 1/3$ as well, and $P(c_i|m_j) = 1/3$ for each i, j.

10. Suppose $\#\mathcal{M} = \#\mathcal{C}$. Prove that for a fixed key $k \in \mathcal{K}$ and a fixed ciphertext $c \in \mathcal{C}$, there is a unique plaintext $m \in \mathcal{M}$ such that $e(k, m) = c$. (Hint: this is a counting argument using the fact that e_k is 1-1).

    Solution: Fix $k \in \mathcal{K}$. Then the function $e_k : \mathcal{M} \to \mathcal{C}$ is injective, and an injection from one finite set to another is a bijection. Thus e_k is a bijection and thus invertible, and for every $c \in \mathcal{C}$ there is a unique $e_k^{-1}(c)$.