Homework 9 Written Solutions

Written Part

3. Let $X : \Omega \to \mathbb{R}$ be a random variable, taking values in the set $\{x_1, x_2, \cdots x_r\}$.

   (a) Show that if $a \neq b$ then the events $(X = a)$ and $(X = b)$ are disjoint.

   Proof. We prove the contrapositive. If $\omega \in (X = a) \cap (X = b)$, this says that $X(\omega) = a$ and $X(\omega) = b$, so that $a = b$. □

   (b) Show that

$$\Omega = \bigcup_i (X = x_i)$$

   Proof. Since the right side is a union of subsets of the left, the right is clearly contained in the left. Conversely, suppose that $\omega \in \Omega$. Then $X(\omega) = x_i$ for some i, so that $\omega \in (X = x_i)$. Since ω was arbitrary, this show the left is contained in the right. □

   (c) Show that

$$\sum_{i=1}^{r} f_X(x_i) = 1.$$

   Proof. Applying HW8 Problem 4e we observe that

$$\begin{aligned}
\sum_{i=1}^{r} f_X(x_i) &= \sum_{i=1}^{r} Pr(X = x_i) && \text{Definition of } f_X \\
&= Pr\left(\bigcup_{x=1}^{r} (X = x_i)\right) && \text{HW8 4e and part (a)} \\
&= Pr(\Omega) && \text{Part (b)} \\
&= 1,
\end{aligned}$$

   as desired. □

   (d) Recall that the expected value of X was defined to be:

$$E(X) = \sum_{i=1}^{r} x_i f_X(i).$$

   Prove that it this is equal to the folloing value:

$$\sum_{\omega \in \Omega} X(\omega) Pr(\omega).$$

Proof. By definition of event probability, we see that:

$$f_X(x_i) = Pr(X = x_i) = \sum_{\omega \in (X = x_i)} Pr(\omega).$$

Since for all such ω we have $X(\omega) = x_i$, this means:

$$x_i f_X(x_i) = \sum_{\omega \in (X = x_i)} X(\omega) Pr(\omega).$$

Add this up over all i and applying part (b) gives the desired formula. □

4. In the following cases compute the expected value of the variable X

(a) X is uniformly distributed on $\{0, 1, \cdots, N-1\}$.

Proof. You may freely use the formula for triangle numbers: $\sum_{k=1}^{M} k = \frac{M(M+1)}{2}$. Since we have a uniform distribution, we showed in class that for each $i = 0, ..., N-1$ we have that $Pr(X = i) = 1/N$. Therefore:

$$E(X) = \sum_{i=0}^{N-1} i \cdot Pr(X = i) = \sum_{i=0}^{N-1} \frac{i}{N} = \frac{1}{N} \sum_{i=1}^{N-1} i = \left(\frac{1}{N}\right)\left(\frac{(N-1)N}{2}\right) = \frac{N-1}{2}$$

□

(b) X is uniformly distributed on $\{1, 2, \cdots, N\}$.

Proof. As in part (a), we know for $i = 1, ..., N$ that $Pr(X = i) = 1/N$. Therefore:

$$E(X) = \sum_{i=1}^{N} \frac{i}{N} = \frac{1}{N} \sum_{i=1}^{N} = \left(\frac{1}{N}\right)\left(\frac{N(N+1)}{2}\right) = \frac{N+1}{2}$$

□

(c) X is uniformly distributed on the first 7 prime numbers.

Proof. The output set is $S = \{2, 3, 5, 7, 11, 13, 17\}$, and for each $i \in S$ we know $f_X(i) = 1/7$. Thefore:

$$\begin{aligned} E(X) &= \sum_{i \in S} i f_X(i) \\ &= \sum_{i \in S} i/7 \\ &= \left(\frac{1}{7}\right)(2 + 3 + 5 + 7 + 11 + 13 + 17) \\ &= \frac{58}{7} \approx 8.286 \end{aligned}$$

□

(d) X is a random variable with a binomial density function. (Hint: use the binomial theorem and a differentiation to get a closed form for the sum).

Proof. Recall that this is an event where n experiments are conducted, and each one either success or fails, with the probability of success being p. We showed in class that for each $i = 0, \cdots, n$ the probability of having exactly i successful outcomes was:

$$f_X(i) = Pr(X = i) = \binom{n}{i} p^i (1-p)^{n-i}.$$

Therefore we can compute the expected value to be:

$$E(X) = \sum_{i=0}^{n} i f_X(i) = \sum_{i=0}^{n} i \cdot \binom{n}{i} p^i (1-p)^{n-1}.$$

This is not very enlightening, but it looks close to the derivative of a binomial expansion (at least if we consider p and $1-p$ separately). Let's formalize this observation by considering the binomial function in two variables:

$$G(x, y) = (x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}.$$

Then we can take the partial derivative with respect to x, and multiply through by x:

$$x \cdot \frac{\partial G}{\partial x} = x \cdot \sum_{i=0}^{n} i \cdot \binom{n}{i} x^{i-1} y^{n-i} = \sum_{i=0}^{n} i \cdot \binom{n}{i} x^i y^{n-i}.$$

Now observe that when we evaluate this at $x = p$ and $y = 1-p$ we get $E(X)$. That is:

$$E(X) = \left(x \cdot \frac{\partial G}{\partial x} \right) (p, 1-p).$$

Of course, we can also compute the partial derivative of G using it's expression as a binomial:

$$x \cdot \frac{\partial G}{\partial x} = x \cdot n(x+y)^{n-1}.$$

Evaluating this at $(p, 1-p)$ gives:

$$E(X) = pn(p+1-p)^{n-1} = np.$$

That's much nicer! □

5. In this problem we will use probability and expected values to study why the `findPrime` algorithm from problem 1 was so successful.

(a) Let $L < U$ be a positive integers. Use the prime number theorem to estimate

$\rho = \rho(L, U) =$ (the probability that a randomly chosen number n with $L < n \leq U$ is prime)

in terms of L and U.

Proof. We compute:

$$\rho(L,U) = \frac{\pi(U) - \pi(L)}{U - L} \approx \frac{U/\ln U - L/\ln L}{U - L}.$$

□

(b) Let Ω the set of outcomes consisting of infinite sequences of numbers between L and U:

$$\Omega = \{n_1, n_2, n_3, \cdots | L < n_i \leq U \text{ for all } i\}.$$

Let $X : \Omega \to \mathbb{Z}$ be the random variable whose value is number of guesses until the first prime. That is:

$$X(n_1 n_2 n_3 ...) = i \iff n_i \text{ is prime and } n_j \text{ is not prime for any } j < i.$$

Let a be a positive integer. Compute the probability density $f_X(i)$ in terms of i and the probability ρ from part (a). (That is, what is the probability that the ith number is the first prime?)

Proof. This is a perscribed event, the chances that n_j is not prime is $(1 - \rho)$ for each $j < i$, so we multiply these together to get the probability that none of them are prime:

$$Pr(n_0, n_1, ..., n_{i-1} \text{ are not prime}) = (1 - \rho)^{i-1}$$

This is independent from the probability that n_i is prime (since we are just choosing numbers at random, so we multiply this by:

$$Pr(n_i \text{ is prime}) = \rho.$$

In total we see that $f_X(i) = \rho(1 - \rho)^{i-1}$. □

(c) Compute the expected value $E(X)$. Interpret in words what this number means. (This computation should look a lot like the expected value of the coin flipping example in the 11/5 lecture).

Proof. Viewing ρ as a formal variable, we compute:

$$\begin{aligned}
E(X) &= \sum_{i=1}^{\infty} i \cdot f_X(i) \\
&= \sum_{i=1}^{\infty} i \cdot r\rho(1-\rho)^{i-1} \\
&= \rho \cdot \sum_{i=1}^{\infty} i(1-\rho)^{i-1} \\
&= \rho \cdot \sum_{i=1}^{\infty} -\frac{d}{d\rho}\left((1-\rho)^i\right) \\
&= -\rho \cdot \frac{d}{d\rho}\left(\sum_{i=1}^{\infty}(1-\rho)^i\right) \\
&= -\rho \cdot \frac{d}{d\rho}\left(-1+\sum_{i=0}^{\infty}(1-\rho)^i\right) \\
&= -\rho \cdot \frac{d}{d\rho}\left(-1+\frac{1}{1-(1-\rho)}\right) \\
&= -\rho \cdot \frac{d}{d\rho}\left(\frac{1}{\rho}-1\right) \\
&= -\rho \cdot \left(-\frac{1}{\rho^2}\right) \\
&= 1/\rho.
\end{aligned}$$

□

(d) Use part (c) to estimate the following:

i. If I randomly guess 2 digit numbers how many guesses will it take to find a prime?

Proof.

$$\rho(99,9) \approx \frac{99/\ln 99 - 9/\ln 9}{90} \approx .19387.$$

Therefore:

$$E(X) = 1/\rho(99,9) \approx 5.158$$

□

ii. If I randomly guess 100 digit numbers how many guesses will it take to find a prime?

Proof.

$$\rho(10^{100}-1, 10^{99}) \approx .00434,$$

so that

$$E(X) = 1/\rho \approx= 230.516.$$

□

iii. If I randomly guess 500 digit numbers how many guesses will it take to find a prime?

Proof.
$$E(X) = 1/\rho(10^{500} - 1, 10^{499}) \approx 1151.55.$$
□

(e) Use the evidence you've gathered to explain why `findPrime` from the first project was successful.

Proof. The Millar-Rabin test for primality ran very quickly, detecting if a number was prime in around logarithmic time. Find prime just guessed random numbers and fed it to the Miller-Rabin test, and we just observed that you should expect to find a prime in a little over 1000 guesses on average, which is essentially no time at all. □

6. Suppose 23 random people are in a room. Compute the probability that at least 2 of them share a birthday. (This is the most well known statement of the *birthday paradox*).

Proof. Let E be the probability that 2 people share a birthday. We will begin by computing E^c: the probability that no 2 people do. Order the 23 people as person 1,2,3,...,23. Let F_i be the even that person i doesn't share a birthday with persons $1, 2, ..., i-1$. Then, and using the definitions of conditional probability we have

$$Pr(E^c) = \cap Pr(F_i) = Pr(F_1)Pr(F_2|F_1)Pr(F_3|F_2 \cap F_1) \cdots Pr(F_{23}|F_{22} \cap F_{21} \cap \cdots \cap F_1).$$

For the i'th person to not share a birthday with persons $1, 2, ..., i-1$, assuming none of them share a birthday, means that the i'th person must not have their birthday on $i-1$ particular days. That is:
$$Pr(F_i|F_{i-1} \cap \cdots \cap F_1) = \frac{365 - (i-1)}{365}.$$
Therefore, we see that:
$$Pr(E^c) = \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{343}{365}\right) \approx .4927.$$
Therefore,
$$Pr(E) = 1 - Pr(E^c) \approx .5073.$$
□

The next problem concerns the following theorem from class, for which we did prove part (i).

Theorem 1. *Suppose there is an urn with N balls, of which n are red and $N-n$ are blue. Suppose further that you randomly choose m balls, replacing after each selection.*

(i) $Pr(\text{at least one red}) = 1 - \left(1 - \frac{n}{N}\right)^m$

(ii) $Pr(\text{at least one red}) \geq 1 - e^{\frac{-mn}{N}}$

(iii) If N is large and m, n are not much larger than $\sqrt{N}$ then the estimate from (ii) is quite accurate

7. Lets prove parts (ii) and (iii) of Theorem 1. We may use part (i) in our proofs since it was established in class.

(a)
$$e^{-x} \geq 1 - x \text{ for all } x.$$
(Hint: use calculus to find the global minimum of $e^{-x} - (1 - x)$)).

Proof. We consider the function $F(x) = e^{-x} - (1-x)$. To optimize we take the derivative: $F'(X) = -e^{-x} + 1$. This has one root at $x = 0$. Evaluating this at F gives $F(0) = 0$, so that $(0, 0)$ is the unique critical point of F. To see this is a global minium, we observe that $F''(X) = e^{-x} > 0$ for all x, so the function is concave up. □

(b) Use part (a) and Theorem 1(i) to prove Theorem 1(ii).

Proof. Letting $x = n/N$, we see that $1 - n/N \leq e^{-n/N}$. Raising to the m'th power gives $(1 - n/N)^m \leq e^{-nm/N}$, so that applying part (i) we see that:
$$Pr(\text{at least one red}) = 1 - \left(1 - \frac{n}{N}\right)^m \geq 1 - e^{\frac{-mn}{N}},$$
as desired. □

(c) Prove that for all $a > 1$ and $0 \leq x \leq 1$ the following inequality holds.
$$e^{-ax} \leq (1 - x)^a + \frac{1}{2}ax^2.$$

Proof. Let $F(x) = (1-x)^a + \frac{1}{2}ax^2 - e^{-ax}$. We hope to show that $F(x) \geq 0$ for $0 \leq x \leq 1$. Since $F(0) = 0$, it suffices to show that $F'(x) \geq 0$ for $0 \leq x \leq 1$, since the function would be increasing from 0 on that whole interval and therefore could never fall below 0.
We compute $F'(x) = -a(1 - x)^{a-1} + ax + ae^{-ax}$. Notice that $F'(x) \geq 0$ if and only if $F'(x)/a \geq 0$, so we can cancel the a from each term and consider $-(1-x)^{a-1} + x + e^{-ax}$. By part (a), since $e^{-x} \geq 1 - x$ we have $e^{-ax} \geq (1 - x)^a$. Therefore:
$$F'(x)/a = -(1 - x)^{a-1} + x + e^{-ax} \geq -(1 - x)^{a-1} + x + (1 - x)^a =: G(x),$$
where we define $G(x)$ as the right hand side. It therefore suffices to show that $G(x) \geq 0$ for $0 \leq x \leq 1$. Factoring gives:
$$G(x) = (1 - x)^{a-1}(-1 + 1 - x) + x = -x(1 - x)^{a-1} + x.$$
Since $a - 1 \geq 0$, and $0 \leq x \leq 1$, the $0 \leq (1 - x)^{a-1} \leq 1$. Multipyling through by x gives:
$$x(1 - x)^{a-1} \leq x,$$
so that we conclude that $G(x) = x - x(1 - x)^{a-1} \geq 0$, as desired. □

(d) Use part (c) and Theorem 1(i) to prove the following identity:

$$Pr(\text{at least one red}) \leq 1 - e^{\frac{-mn}{N}} + \frac{mn^2}{2N^2}.$$

Use this to deduce Theorem 1(iii).

Proof. Using part (c) with $x = n/N$ and $a = m$, with $m > 1$ and $n \leq N$ we compute directly that $e^{-mn/N} \leq (1 - n/N)^m + \frac{1}{2}m(n/N)^2$. In particular, we have

$$\left(1 - \frac{n}{N}\right)^m \geq e^{\frac{-mn}{N}} + \frac{mn^2}{2N^2}.$$

Plugging this into part (i) of the theorem gives:

$$Pr(\text{at least one red}) = 1 - \left(1 - \frac{n}{m}\right)^m \leq 1 - e^{\frac{-mn}{N}} + \frac{mn^2}{2N^2},$$

as desired. Putting this together with part (ii) of the theorem gives the following chain of inequalities:

$$1 - e^{\frac{-mn}{N}} \leq Pr(\text{at least one red}) \leq 1 - e^{\frac{-mn}{N}} + \frac{mn^2}{2N^2}.$$

In particular, the error of the estimate from part (ii) is controlled by the value $\frac{mn^2}{2N^2}$. If $n, m \approx \sqrt{N}$ then this value is approximately (on the order of) $\frac{\sqrt{N}}{N}$ which is very small if N is very large. □

8. In the Miller-Rabin problem I suggested that we interpret the Prime Number Theorem as saying the probability of a number n being prime is $\frac{\ln n}{n}$, but of course this way off the mark. The prime number theorem says there are $\frac{n}{\ln n}$ primes less than n. Thus the probability of one being prime in particular is approximately

$$\frac{n/\ln n}{n} = \frac{1}{\ln n}.$$

Notice that $\frac{1}{\ln n}$ is MUCH LARGER DENSITY than $\frac{\ln n}{n}$.

(a) To really feel the difference between these two densities, use each to compute the probability that a random number less than 10^{100} is prime. This should illustrate the gravity of the mistake.

Proof. Notice that $1/\ln(10^{100}) \approx .00434$ which says that approximately half a percent of 100 digit (or less) numbers are prime. This is about 1 in every 200, so not too bad. The incorrect estimation would have predicted $\ln(10^{100})/10^100 \approx 2.3 \times 10^{-98}$. This says less than 1 in every 10^{97} numbers is prime, which would say they are extremely rare. □

The beauty of the prime number theorem is it says primes are rather dense, and the value I gave said they are extremely sparse. This likely affected the expected correctness of your Miller-Rabin computation.

(b) Redo the computations from HW8 Problem 8 with this correct probability, so that we have computed the correct values. (Don't worry about re-deriving everything, just show the formula and plug in the correct values.) In particular you should show that:

$$Pr\left(\text{n is prime} \mid \text{Miller-Rabin Fails } N \text{ times}\right) \geq 1 - \frac{\ln n}{4^N}.$$

In particular, how confident are we in are primes when implemented RSA, where $N = 20$ and $2^{511} \leq n < 2^{512}$?

Proof. The formula we use is the following. We let F_N be "the Miller Rabin test returns not composite N times", and E be the event a number is prime. If $Pr(E) = \delta$ we computed in HW 8 the formula

$$Pr(E|F_N) \geq \frac{\delta}{\delta + (1-\delta)(1-P^N)},$$

where P is (a lower bound for) the probability that the Miller-Rabin test returns 'is composite' if n is composite.
For example, a 512 bit number n we approximate by 2^{512} (this is an upper bound), we have $\delta \approx 1/\ln n$ and $P = .75$ Suppose we run the Miller-Rabin test 20 times and it returns not composite each time, then:

$$Pr(n \text{ is prime } | n \text{ is probably prime}) = \frac{1/\ln(2^{512})}{1/\ln(2^{512}) + (1 - \ln(2^{512})(1-.75)^20} \approx 0.99999999967814.$$

Those seem like pretty good odds. I had a lot of pushback and complaints about the bound I suggested, and these were probably warratned, since they won't work for all n, N, but for any that we use they would. Let me explain how they would be derived. First one writes writes $(1-\delta)(1-P)^N = \lambda$ and considers

$$\frac{\delta}{\delta + (1-\delta)(1-P^N)} = \frac{\delta}{\delta + \lambda} = \frac{\delta + \lambda - \lambda}{\delta + \lambda} = 1 - \frac{\lambda}{\delta + \lambda}.$$

Now, we will use the following fact, if λ is very small with respect to δ, it's value in the denominator of a fraction is negligable compared to that of δ, so that we can approximate the above term by $\frac{\lambda}{\delta}$. To apply this approximation we need that 4^N is much larger than $\ln n$, but this is easy to acheive since exponential grows way faster than logrithmic. Indeed, even for $N = 20$ and $n = 2^{512}$ this is certainly true, we have 4^N is a 12 digit number and $\ln n$ is a 3 digit number. Analyzing this more closely we have:

$$\frac{\lambda}{\delta} = \frac{(1 - 1/\ln n)(1/4)^N}{1/\ln n} = \frac{\ln n - 1}{4^N} \approx \frac{\ln n}{4^N},$$

as desired. □