CS 70 Discrete Mathematics and Probability Theory
Spring 2018 Ayazifar and Rao

# Final Solutions

## 1. Discrete Math: True/False (12 parts: 3 points each.)

1. $\forall x, \forall y, \neg P(x,y) \equiv \neg \exists y, \exists x, P(x,y)$

   **Answer:** True. If for every $x$ and $y$ $P(x,y)$ is not true, then there doesn't exist an $x$ and $y$ where $P(x,y)$ is true.
2. $(P \implies Q) \equiv (Q \implies P)$.

   **Answer:** False. The converse of a statement is not logically equivalent to the statement.
3. Any simple graph with $n$ vertices can be colored with $n-1$ colors.

   **Answer:** False. The complete graph requires $n$ colors to properly color it.
4. The set of all finite, undirected graphs is countable.

   **Answer:** True. Can enumerate by considering all graphs with just one vertex, then two vertices, three, etc., all of which have a finite number of configurations of edges.
5. The function $f(x) = ax \pmod N$ is a bijection from and to $\{0,\ldots,N-1\}$ if and only if $\gcd(a,N) = 1$.

   **Answer:** True. $a$ has a multiplicative inverse mod $N$.
6. For a prime $p$, the function $f(x) = x^d \pmod p$ is a bijection from and to $\{0,\ldots,p-1\}$ when $\gcd(d,p-1) = 1$.

   **Answer:** True. The inverse function is $g(x) = x^e \pmod p$ with $e = d^{-1} \pmod{p-1}$.
7. A male optimal pairing cannot be female optimal.

   **Answer:** False. There could just be a single stable pairing.
8. For any undirected graph, the number of odd-degree vertices is odd.

   **Answer:** False. The sum of the degrees is even, since it is twice the number of edges, and thus the number of odd degree vertices is even.
9. For every real number $x$, there is a program that given $k$, will print out the $k$th digit of $x$.

   **Answer:** False. The number of programs is countable but the number of real numbers is not.
10. There is a program that, given another program $P$, will determine if $P$ halts when given no input.

    **Answer:** False. One can reduce from the halting problem as follows. Given a program $P$ and $x$, produce a program that has a constant string with value $x$ and runs $P$ on $x$. This program does not take any input, so determining if it halts also determines whether $P$ halts on $x$.
11. Any connected simple graph with $n$ vertices and *exactly* $n$ edges is planar.

    **Answer:** True. It is a tree plus an edge. Since a tree has one face, that edge can be drawn in that face.
12. Given two numbers, $x$ and $y$, that are relatively prime to $N$, the product $xy$ is relatively prime to $N$.

    **Answer:** True. Neither have a prime factor in common with $N$ and neither does their product which only has prime factors from $x$ and $y$.

## 2. Discrete Math:Short Answer (10 parts: 4 points each)

1. If $\gcd(x,y) = d$, what is the least common multiple of $x$ and $y$ (smallest natural number $n$ where both $x|n$ and $y|n$)? [Leave your answer in terms of $x,y,d$]

   **Answer:** $\frac{xy}{d}$. We have $x = kd$ and $y = \ell d$ where $gcd(k,\ell) = 1$. $\frac{xy}{d} = k\ell d$, and thus is divisible by $x$ and $y$. Moreover, $n$ must contain all the prime factors of $k$ and $\ell$ and $d$ and since they have none in common, $n$ must be as large as this product.

2. Consider the graph with vertices $\{0,\dots,N-1\}$ and edges $(i,i+a) \pmod N$ for some $a \not\equiv 0 \pmod N$. Let $d = gcd(a,N)$. What is the length of the longest cycle in this graph in terms of some subset of $N, a$, and $d$?

   **Answer:** $N/d$. If one takes $j = N/d$ steps along edges from $i$ one gets to $i + ja = i \pmod N$ since $ja = N/d(kd) = Nk = 0 \pmod N$.

3. What is the minimum number of women who get their favorite partner (first in their preference list) in a female optimal stable pairing? (Note that the minimum is over any instance.)

   **Answer:** 0. Consider a 3-men, 3-women example where women $A$ and $B$ have man 1 as their favorite, and man 1 rejects $B$, who then proposes to woman $C$'s current partner. She then asks man 1, who likes her best and rejects woman $A$. Now, no woman has her favorite partner.

4. What is the number of ways to split 7 dollars among Alice, Bob and Eve? (Each person should get an whole number of dollars.)

   **Answer:** $\binom{9}{2}$. We think of assigning a number to Alice, one to Bob and one to Eve, where the numbers sum to 5. We map this situation to one of having 5 stars and 2 bars and then choosing where the bars go out of the seven positions.

5. What is $6^{24} \pmod{35}$?

   **Answer:** 1. This is from the fact that $a^{(p-1)(q-1)} \equiv 1 \pmod{pq}$ when $gcd(a,pq) = 1$. Here $a = 6$ and $pq = (7)(5) = 35$ and $(p-1)(q-1) = 24$.

   Alternative Answer: Notice that $6^2 = 1 \pmod{35}$, which tells us $6^{24} = 1 \pmod{35}$.

6. If one has three distinct degree at most $d$ polynomials, $P(x), Q(x), R(x)$, what is the maximum number of intersections across all **pairs** of polynomials?

   Recall that we define intersections to be two polynomials having the same value at a point. (That is if $P(1) = Q(1)$, and $P(2) = R(2)$ and $R(3) = Q(3)$, that is three intersections. If they all meet at a point $P(1) = Q(1) = R(1)$, that is three intersections.)

   **Answer:** $3d$. Consider if there were $3d+1$. Any intersection point can be assigned to a pair of polynomials that intersect at that point. The number of points assigned to at least one pair of polynomials must be $d+1$ which means that pair of polynomials is the same.

   Have all three polynomials intersect on the same $d$ points and differ on a $d+1$ fixed point yields $3d$ intersections.

7. Working modulo a prime $p > d$, given a degree exactly $d$ polynomial $P(x)$, how many polynomials $Q(x)$ of degree at most $d$ are there such that $P(x)$ and $Q(x)$ intersect at exactly $d$ points?

   **Answer:** $\binom{p}{d}(p-1)$.

   For any polynomial $Q(x)$ that intersects at $d$ points, we have $R(x) = P(x) - Q(x) = 0$ at these $d$ points. This polynomial can be factored into the form $R(x) = a_0(x-r_1)\cdots(x-r_d)$. We need to choose $a_0$ and $r_1,\dots,r_d$ to specify $Q(x)$. The number of ways to choose $a_0$ is $p-1$ and the number of ways to choose $r_1,\dots,r_d$ without repetition as the intersection points are distinct.

8. Recall that the vertices in a $d$-dimensional hypercube correspond to $0-1$ strings of length $d$. We call the number of 1's in this representation the **weight** of a vertex.

   (a) How many vertices in a $d$-dimensional hypercube have weight $k$?
   **Answer:** $\binom{d}{k}$. Need to choose where the 1's are.

   (b) How many edges are between vertices with weight at most $k$ and vertices with weight greater than $k$?
   **Answer:** $\binom{d}{k}(d-k)$. Each vertex of weight $k$ has $d-k$ neighbors of weight $k+1$.

9. For a prime $p$, how many elements of $\{0,\ldots,p^k-1\}$ are relatively prime to $p$?
   **Answer:** $p^{k-1}(p-1)$. One can eliminate all the numbers that are divisible by $p$, which gives $p^k - p^{k-1}$ and factor out $p$.
   Or one can also recall Euler's totient formula which we used on the homework.

**3. Some proofs. (3 parts. 5/5/8 points.)**

1. Recall for $x,y$, with $gcd(x,y)=d$, that there are $a,b \in Z$ where $ax+by=d$. Prove that $gcd(a,b)=1$.
   **Answer:** Notice $x=dt$ and $y=ds$ for integers $s$ and $t$, plugging in we get $asd+btd=d$. Dividing, we get $as+bt=1$. Now, this suggests $gcd(a,b)=1$ by Extended Euclid.
2. You have $n$ coins. The probability of the $i$th coin being heads is $1/(i+1)$ (i.e., the biases of the coins are $\frac{1}{2},\frac{1}{3},\ldots,\frac{1}{n+1}$). You flip all the coins. What is the probability that you see an even number of heads? Prove it. (Hint: the answer is quite simple.)
   **Answer:** $\frac{1}{2}$. For one coin we are asking for the probability that there are zero heads, or $1/2$. Let $A_i$ be the event that the number of heads is even in the first $i$ coins, and $H_i$ be the event that the $i$th coin is heads. We have
   $P(A_{i+1}) = P(A_i)P(\overline{H_i}) + P(\overline{A_i})P(H_i) = (1/2)(1/i) + (1/2)(1-1/i) = 1/2.$
   Alternative Solution: Note that, because of the 1st coin with bias 1/2, every odd-headed arrangement has another even-headed arrangement of equal probability, corresponding to flipping the first coin.
3. Consider a game with two players alternating turns. The game begins with $N>0$ flags. On each turn, each player can remove $1,2,3$, or $4$ flags. A player wins if they remove the last flag (even if they removed several in that turn).
   Show that if both players play optimally, player 2 wins if $N$ is a multiple of 5, and player 1 wins otherwise. (Note player 1 goes first.)
   **Answer:** The base case is that Player 1 wins if $N \in \{1,2,3,4\}$. We assume the statement by induction, and note that if $N$ is not a multiple of 5, that Player 1 can make it a multiple of 5 by removing one of $\{1,2,3,4\}$ flags. Then player 2 must remove some number of flags in $\{1,2,3,4\}$ and the resulting number of flags that player 1 faces is not a multiple of 5 and it is smaller. Thus, by induction player 1 has a winning strategy.

**4. Probability:True/False. (7 parts, 3 points each.)**

1. For a random variable $X$, the event "$X=3$" is independent of the event "$X=4$".
   **Answer:** False. They are mutually exclusive not independent.
2. Let $X,Y$ be Normal with mean $\mu$ and variance $\sigma^2$, independent of each other. Let $Z=2X+3Y$. Then, $LLSE[Z \mid X] = MMSE[Z \mid X]$.
   **Answer:** True.
   Notice, $LLSE[Z|X] = 5\mu + \frac{cov(X,2X+3Y)}{var(X)}(X-\mu) = 2X+3\mu$ as $X$ and $Y$ are independent.
   $MMSE[Z|X] = \mathbb{E}[2X+3Y \mid X] = 2X + \mathbb{E}[3Y \mid X] = 2X+3\mu$ as X and Y are independent.
3. Any irreducible Markov chain where one state has a self loop is aperiodic.
   **Answer:** True. Since one can reach this state from any state and vice versa, there is a cycle of every length, which means that the period which is the gcd of all cycle lengths is at most 1.
4. Given a Markov Chain, let the random variables $X_1,X_2,X_3,\ldots$, where $X_t =$ the state visited at time $t$ in the Markov Chain. Then $E[X_t|X_{t-1}=x] = E[X_t|X_{t-1}=x \cap X_{t-2}=x']$.
   **Answer:** True. The value of $X_t$ conditioned on $X_{t-1}$ is independent of all previous times.

5. Given an expected value $\mu$, a variance $\sigma^2 \geq 0$, and a probability $p \in (0,1)$, it is always possible to choose $a$ and $b$ such that a discrete random variable $X$ which is $a$ with probability $p$ and $b$ with probability $1-p$ will have the specified expected value and variance.

   **Answer:** True. Can think of it as a system of two equations for variance and expected value, with two unknowns $a$ and $b$.

6. Consider two random variables, $X$ and $Y$, with joint density function $f(x,y) = 4xy$ when $x,y \in [0,1]$ and 0 elsewhere. $X$ and $Y$ are independent.

   **Answer:** True. $f(x) = \int_0^1 f(x,y)dy = 2x$. $f(x|Y=y) = f(x,y)/\int_0^1 f(x,y)dx = 4xy/(4y(1/2)) = 2x$

7. Suppose every state in a Markov chain has exactly one outgoing transition. There is one state, $s$, whose outgoing transition is a self-loop. All other states' outgoing transitions are not self-loops. If a unique stationary distribution exists, it must have probability 1 on s and 0 everywhere else.

   **Answer:** True. You can always leave every other state, but $s$ is the hotel california; you may check out, but you can never leave.

**5. Probability: Short Answer. (17 parts, 4 points each.)**

1. Consider $X \sim G(p)$, a geometric random variable $X$ with parameter $p$. What is $Pr[X > i | X > j]$ for $i \geq j$?

   **Answer:** $(1-p)^{i-j}$. This is the event that the first $i-j$ trials fail after the $j$th trial.

2. Suppose we have a random variable, $X$, with pdf

$$f(x) = \begin{cases} cx^2, \text{if } 0 \leq x \leq 1 \\ 0, \text{ otherwise} \end{cases}$$

   What is $c$?

   **Answer:** 3. $\int_0^1 cx^2 = cx^3/3|_0^1 = c/3 = 1$, which implies that $c = 3$.

3. Given a binomial random variable $X$ with parameters $n$ and $p$, ($X \sim B(n,p)$) what is $Pr[X = E[X]]$? (You should assume $pn$ is an integer.)

   **Answer:** $\binom{n}{pn} p^{pn}(1-p)^{(1-p)n}$

4. $Pr[A|B] = 1/2$, and $Pr[B] = 1/2$, and $A$ and $B$ are independent events. What is $Pr[A]$?

   **Answer:** 1/2. $Pr[A|B] = Pr[A]$ for independent events. For $A$ and $B$ to be independent, $Pr[A] \times Pr[B] = Pr[A|B]Pr[B]$.

5. Aaron is teaching section and has 6 problems on the worksheet. The time it takes for him to finish covering each question are i.i.d. random variables that follow the exponential distribution with parameter $\lambda = 1/20$. Additionally, for each question, Aaron may choose to skip it entirely with probability $p = 1/3$. What is the expected time of section?

   **Answer:** 80 minutes. Let $X_i$ be the random variable corresponding to the time for the $i$th problem where it is 0, if Aaron doesn't do it. $E[X_i] = E[X_i|skip]Pr[skip] + E[X_i|notskip]Pr[notskip] = 0*1/3 + 20*2/3 = 40/3$. The total time is $6E[X_1] = 80$.

6. Let $X$ be a uniformly distributed variable on the interval $[3,6]$. What is Var(X)?

   **Answer:** $9/12$. The variance of $X$ being uniform over $[0,1]$ is $1/12$. The width of this interval is 3 so we get $9/12$.

7. Label $N$ teams as team 1 through team $N$. They play a tournament and get ranked from rank 1 to rank $N$ (with no ties). All rankings are equally likely.

(a) What is the total number of rankings where team 1 is ranked higher than team 2?
**Answer:** $\frac{N!}{2}$. The number of orders with 1 before 2 is the same a the number where 1 is after 2; a bijection is to switch the 1 and 2.

(b) What is the expected number of teams with a strictly lower rank number than their team number? For example, if team 3 was rank 1, their rank number (1) is lower than their team number (3). Simplify your answer (i.e. no summations).
**Answer:** $\frac{N-1}{2}$. For team numbered $i$, let $X_i$ be the indicator variable that $i$ ends up at a lower rank. $E[X_i] = Pr[X_i = 1] = (i-1)/N$ since each position is equally likely and there are $i-1$ positions before $i$. Summing up over $i$, we obtain $\frac{1}{N}(N-1)(N)/2 = (N-1)/2$.

8. Let $X$ be a random variable that is never smaller than $-1$ and has expectation 5. Give a non-trivial upper bound on the probability that $X$ is at least 12.

**Answer:** $6/13$. Let $Y = X+1$, which is a non-negative random variable with expectation 6. By Markov's inequality, $Pr[X \geq 12] = Pr[Y \geq 13] \leq \frac{6}{13}$.

9. Let $X$ be a random variable with mean $E[X] = 5$ with $E[X^2] = 29$. Give a non-trivial upper bound on the probability that $X$ is larger than 12.

**Answer:** $4/49$. Chebyshev's inequality suggests that $Pr[|X - E[X]| > t] \leq \frac{Var(X)}{t^2} = \frac{4}{49}$. Moreover, $Var(X) = E[X^2] - E[X]^2 = 4$.

10. Let $T$ be the event that an individual gets a positive result on a medical test for a disease and $D$ be the event that an individual has the disease. The test has the property that $Pr[T|D] = .9$ and $Pr[T|\overline{D}] = .01$. Morever, $Pr[D] = .01$. Given a positive result, what the probability that the individual has a disease? (No need to simplify your answer, though it should be a complete expression with numbers.)

**Answer:** $Pr[D|T] = \frac{Pr[T|D]\times Pr[D]}{Pr[T|D]\times Pr[D]+Pr[T|D^c]Pr[D^c]} = \frac{.009}{.009+.0099}$.

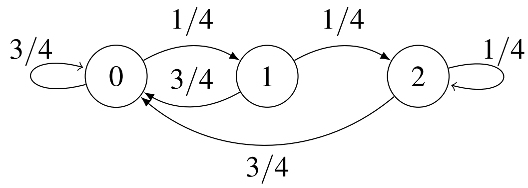
11. Let $R$ be a continuous random variable corresponding to a reading on a medical test for an individual and $D$ be the event that the individual has a disease. The probability of an individual having the disease is $p$. Further, let $f_{R|D}(r)$ (and $f_{R|\overline{D}}(r)$) be the conditional probability density for $R$ conditioned on $D$ (respectively conditioned on $\overline{D}$). Given a reading of $r$, give an expression for the probability the individual has the disease in terms of $f_{R|D}(r), f_{R|\overline{D}}(r)$, and $p$.

**Answer:** $\frac{f_{R|D}(r)\times p}{f_{R|D}(r)\times p + f_{R|\overline{D}}(r)\times(1-p)}$.

This can be seen as the limit of $\frac{Pr[(r\leq R\leq r+dr)\cap D]}{Pr[r\leq R\leq r+dr]}$.
The answer computes this with factors of $dr$ in the numerator and denominator which then divide out.

12. For continuous random variables, $X$ and $Y$ where $Y = g(X)$ for some differentiable, bijective function $g : \mathbb{R} \to \mathbb{R}$. What is $f_Y(y)$ in terms of $f_X(\cdot)$, $g(\cdot)$, $g^{-1}(\cdot)$ and $g'(\cdot)$? (Possibly useful to remember that $f_Y(y)dy = Pr[y \leq Y \leq y+dy]$.)
**Answer:** $f_X(g^{-1}(y))/g'(g^{-1}(y))$.
Recall that $f_Y(y)dy = Pr[y \leq Y \leq y+dy]$. We have $f_X(x)dx = Pr[x \leq X \leq x+dx]$.
Now, we have $y = g(x)$, so $dy = g'(x)dx$ which also states that $dx = dy/g'(y)$.
For a fixed $y$ and $dy$, $x = g^{-1}(y)$, and $x+dx = g^{-1}(y) + dy/g'(x)$.
Thus, we have $f_Y(y)dy = Pr[g^{-1}(y) \leq X \leq g^{-1}(y) + dy/g'(x)] = f_X(x)dy/g'(x)$.
Dividing by $dy$ yields the expression.

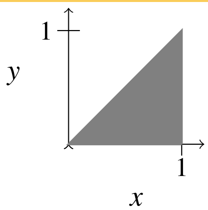
13. What is the stationary distribution, $\pi$, for the following three state Markov chain? (Hint: $\pi(0) = 3/4$)




$\pi(1)$ $\pi(2)$

**Answer:** $\pi(0) = 3/4, \pi(1) = (3/4) * (1/4) = 3/16, \pi(2) = 1/16$. $\pi(0)$ is by noting that no matter where you are, the next state is 0 with probability $3/4$, $\pi(1) = (1/4)\pi(0)$, and finally $\pi(2)$ is whatever it needs to be to add up to 1.

14. Consider continuous random variables, $X$ and $Y$, with joint density that is $f(x,y) = 2$ for $x, y \in [0,1]$ and where $y < x$. That is, the distribution is uniform over the shaded region in the figure below.




Say someone takes a sample of $X$ or $Y$ with equal probability, and then announces that the value is $2/3$. What is the probability that the sample is from $X$?

**Answer:** 2/3.

We wish to compute $Pr[\text{ from } X \mid \text{ see } 2/3]$ which is just $\frac{Pr[\text{ from } X \text{ and } 2/3]}{Pr[2/3]}$.

Now $Pr[\text{ from } X \text{ and } 2/3] \propto (1/2) \times f_X(2/3) = (2/3)(2) = 4/3$

The $Pr[\text{from } Y \text{ and } 2/3] \propto (1/2) \times f_Y(2/3) = ((1/3)(2)) = 2/3$

Thus, the ratio is $\frac{4/3}{2/3+4/3}$ or $2/3$.

15. Given a random variable $X \sim \text{Expo}(\lambda)$, consider the integer valued random variable $K = \lceil X \rceil$.

    (a) What is $Pr[K = k]$?

    **Answer:** $(1 - e^{-\lambda})(e^{-\lambda})^{k-1}$

    A proof is as follows:

$$\begin{aligned} Pr[K = k] &= \int_{k-1}^{k} \lambda e^{-\lambda t} dt \\ &= (-e^{-\lambda(k)} + e^{-\lambda(k-1)}) \\ &= e^{-\lambda(k-1)}(1 - e^{-\lambda}) \end{aligned}$$

    (b) What standard distribution with associated parameter(s) does this correspond to?

    **Answer:** It is geometric with parameter $p = 1 - e^{-\lambda}$.

    This can be seen as the process where the probability of success in an integer interval is the probability that the exponential variable is in an interval of length 1, which is $1 - e^{-\lambda}$.

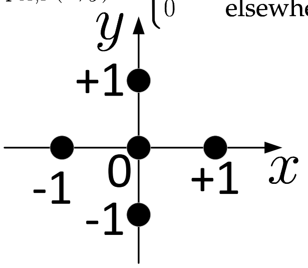
**6. Longer Probability Questions.**

~~1. [I iterated my~~ **expectations, and you can, too!] (4 parts. 5 points each.)**

Consider two discrete random variables $X$ and $Y$. For notational purposes, $X$ has probability mass function (or distribution), $p_X(x) = Pr[X = x]$, mean $\mu_X$, and variance $\sigma_X^2$. Similiarly, random variable $Y$ has PMF $p_Y(y) = Pr[Y = y]$, mean $\mu_Y$ and variance $\sigma_Y^2$.

For each of True/False parts in this problem, either prove the corresponding statement is True in general or use exactly one of the counterexamples provided below to show the statement is False.

$$p_{X,Y}(x,y) = \begin{cases} 1/5 & \text{At each } \bullet \\ 0 & \text{elsewhere.} \end{cases}$$




(a) Potential Counterexample I

The PMF for random variable $X$ is

$$p_X(x) = \Pr(X = x) = \begin{cases} 1/3 & x = -1, 0, +1 \\ 0 & \text{elsewhere.} \end{cases}$$

Random Variable Y is

$$Y = X^2 \quad \text{for all X.}$$

(b) Potential Counterexample II

~~(a) Suppose $E[Y|X$~~$] = c$, where $c$ is a fixed constant. This means that the conditional mean $E[Y|X]$ does *not* depend on $X$.

i. Show that $c = \mu_Y$, the mean of $Y$.

**Answer:** Apply the Law of Iterated Expectations:

$$E[Y] = E[E[Y|X]] = E[c] = c.$$

So, $c = \mu_Y$. To show this in more detail,

$$E[Y] = E[E[Y|X]] = \sum_x \underbrace{E[Y|X = x]}_{=c} p_X(x) = c \underbrace{\sum_x p_X(x)}_{=1} = c.$$

ii. True or False?

The random variables $X$ and $Y$ are independent.

**Answer:** False. In Potential Counterexample I, $E[Y|X = 1] = E[Y|X = -1] = E[Y|X = 0] = 0$. However, $X$ and $Y$ are not independent, as the $p_{Y|X}(1|0) = 1/3 \neq 1/5 = p_Y(1)$. Alternatively, note that $p_{Y,X}(1,1) = 0$, whereas neither $p_Y(1)$ nor $p_X(1)$ is zero. So, $X$ and $Y$ are dependent, even though the conditional mean $E[Y|X]$ does *not* depend on $X$.

iii. True or False?

The random variables $X$ and $Y$ are *uncorrelated*, meaning that $cov(X,Y) = 0$.

**Answer:** True. The covariance is 0, if $E[XY] = E[X]E[Y]$. In this case,

$$\begin{aligned} E[XY] &= \sum_x x \sum_y y Pr[X=x, Y=y] \\ &= \sum_x Pr[X=x] x \sum_y y Pr[Y=y|X=x] \\ &= \sum_x Pr[X=x] x E[Y|X=x] \\ &= c \sum_x Pr[X=x] \cdot = cE[X] = E[Y]E[X]. \end{aligned}$$

An easier proof is that if the MMSE is a constant (and thus a linear function), the LLSE must be the same as the MMSE, and so is also a constant. From our formula for the LLSE, that implies that $\mathrm{cov}(X,Y)=0$.
Another Alternative: $E[E[XY \mid X]] = E[XE[Y \mid X]] = E[cX] = cE[X] = E[Y]E[X]$.

~~(b) Suppose $X$ and~~ $Y$ are *uncorrelated*, meaning that $cov(X,Y)=0$.
True or False?
The conditional mean is $E[Y|X]=c$, where $c$ is a fixed constant, meaning that $E[Y|X]$ does *not* depend on $X$.

**Answer:** False.
We use Counterexample II to show this. Notice that $\mu_X = E(X) = 0$. Therefore,

$$\begin{aligned} cov(X,Y) &= E(XY) - \underbrace{\mu_X}_{=0} \mu_Y \\ &= E(XY) = E(X^3) \\ &= \sum_{x=-1}^{+1} x^3 p_X(x) = \frac{1}{3} \sum_{x=-1}^{+1} x^3 \\ &= 0. \end{aligned}$$

So far we've shown that $X$ and $Y$ are uncorrelated. However, note that

$$E(Y|X) = E(X^2|X) = X^2.$$

In particular,

$$E(Y|X) = \begin{cases} +1 & \text{if } X=-1 \\ 0 & \text{if } X=0 \\ +1 & \text{if } X=+1. \end{cases}$$

Clearly, $E(Y|X)$ varies with $X$; it is *not* a constant.

2. **[Estimations of a random variable with noise.] (6 parts. 2/4/2/2/4/8 points.)**
Let random variable $Y$ denote the blood pressure of a patient, and suppose we model it as a Gaussian random variable having mean $\mu_Y$ and variance $\sigma_Y^2$.
Our blood pressure monitor (measuring device) is faulty. It yields a measurement

$$X = Y + W$$

where the noise $W$ is a zero mean Gaussian random variable ($\mu_W = 0$) with variance $\sigma_W^2$. Assume that the noise $W$ is *uncorrelated* with $Y$. Note, that the actual blood pressure $Y$ is inaccessible to us, due to the additive noise $W$.

(a) Show that $\sigma_X^2 = \sigma_Y^2 + \sigma_W^2$.

**Answer:** The variance of the sum of the two uncorrelated random variables is the sum of the variances.

~~(b) Show that $L(Y$~~$|X)$, the Linear Least-Square Error Estimate for the blood pressure $Y$, based on the measured quantity $X$, is given by

$$L(Y|X) = a + bX, \quad \text{where} \quad a = \frac{\sigma_W^2}{\sigma_Y^2 + \sigma_W^2}\mu_Y \quad \text{and} \quad b = \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_W^2}.$$

**Answer:** The formula is $L(Y|X) = \frac{cov(X,Y)}{var(X)}(X - E(X)) + E(Y)$ or $b = \frac{covX,Y}{var(X)}$ and $a = \mu_Y - b\mu_X$. We have all the terms but $cov(X,Y)$.

$$\begin{aligned}
cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= E[(Y+W-\mu_Y-\mu_W)(Y-\mu_Y)] \\
&= E[(Y-\mu_Y)(Y-\mu_Y)+(W-\mu_W)(Y-\mu_Y)] \\
&= E[(Y-\mu_Y)(Y-\mu_Y)]+E[(W-\mu_W)(Y-\mu_Y)] \\
&= \sigma_Y^2 + 0
\end{aligned}$$

(Alternatively, $cov(X,Y) = E[XY] - E[X]E[Y] = E[Y^2 + YW] - (E[Y]^2 + E[Y]E[W]) = E[Y^2] - E[Y]^2 + E[YW] - E[Y]E[W] = \sigma_Y^2 + Cov(Y,W) = \sigma_Y^2$.)

Thus we have $b = \frac{\sigma_Y^2}{\sigma_W^2 + \sigma_Y^2}$ and $a = \mu_y - b\mu_x$. Since $\mu_Y = \mu_X$, we can rewrite $a$ as follows:

$$\begin{aligned}
a &= \mu_Y - b\mu_X = \mu_Y - b\mu_Y \\
&= (1-b)\mu_Y \\
&= \left(1 - \frac{\sigma_Y^2}{\sigma_W^2 + \sigma_Y^2}\right)\mu_Y \\
&= \frac{\sigma_W^2}{\sigma_W^2 + \sigma_Y^2}\mu_Y
\end{aligned}$$

~~(c) We now consid~~er two extreme cases.

i. Suppose the blood pressure monitor has been repaired —that is, it introduces no noise. Determine a simple expression for $L(Y|X)$ in this case.

**Answer:** $L(Y|X) = X$. In this case, $Y = X$.

ii. Suppose the blood pressure monitor's performance has deteriorated, so it now introduces noise whose variance $\sigma_W^2 \gg \sigma_Y^2$. In the limit $\sigma_W^2 \to \infty$, what does your best linear estimator converge to? Explain briefly, in plain English words, why your answer makes sense.

**Answer:** $L(Y|X) \to \mu_Y$. In the limit the measurement, $X = Y + W$, is just noise, so one should just predict the mean of $Y$ as $X$ gives no information.

~~(d) Recall $L[Y|X]$~~ is a function of $X$ and is a random variable. Let $\hat{Y} = L[Y|X] = a + bX$. Determine the distribution of $\widehat{Y}$ and the appropriate parameters.

**Answer:** Since, $\widehat{Y} = a + bX$ is an affine function of $X$, $\widehat{Y}$ has the same type of PDF as $X$—that is, $\widehat{Y}$ is a Gaussian random variable. All we must do is determine its mean and variance. Since the noise is zero-mean, we know that $\mu_X = \mu_Y$. Therefore, $E(\widehat{Y}) = a + b\mu_X = a + b\mu_Y$. Furthermore $var(\widehat{Y}) = b^2\sigma_X^2$. Therefore,

$$\widehat{Y} \sim \mathcal{N}(a + b\,\mu_Y, b^2\sigma_X^2)$$
$$f_{\widehat{Y}}(\widehat{y}) = \frac{1}{b\,\sigma_X\sqrt{2\pi}} e^{-(\widehat{y}-a-b\mu_Y)^2/2b^2\sigma_X^2}.$$

(e) We estimate $\hat{\mu}_Y$ of the true mean $\mu_Y$ as

$$\widehat{\mu}_Y = \frac{X_1 + \cdots + X_n}{n},$$

where $X_i$ are independent measurements of the random variable $X = Y + W$.
We want to be *at least* 95% confident that the absolute error $|\widehat{\mu}_Y - \mu_Y|$ is within 4% of $\mu_Y$. Your task is to determine the *minimum* number of measurements $n$ needed so that

$$Pr\big[\,|\widehat{\mu}_Y - \mu_Y| \le 0.04\,\mu_Y\big] \ge 0.95.$$

You may assume that $\sigma_Y^2 = 12$ and $\sigma_W^2 = 4$ and that the true mean $\mu_Y \in [60, 90]$.
(Remember that in this course, you may assume that a Gaussian random variable lies within $2\sigma$ of its mean with 95% probability.)
**Answer:** We know that

$$var(\widehat{\mu}_Y) = \frac{n\,var(X_i)}{n^2} = \frac{\sigma_Y^2}{n} = \frac{\sigma_Y^2 + \sigma_W^2}{n} = \frac{16}{n}, \quad \text{so} \quad StdDev(\widehat{\mu}_Y) = \frac{\sigma_Y}{\sqrt{n}} = \frac{4}{\sqrt{n}}.$$

$$Pr\big(\,|\widehat{\mu}_Y - \mu_Y| \le 0.04\,\mu_Y\big) \ge 0.95.$$
$$Pr\left(\left|\frac{\widehat{\mu}_Y - \mu_Y}{StdDev(\widehat{\mu}_Y)}\right| \le \frac{0.04\,\mu_Y}{4/\sqrt{n}}\right) \ge 0.95$$
$$Pr\left(\left|\frac{\widehat{\mu}_Y - \mu_Y}{4/\sqrt{n}}\right| \le \frac{0.04\,\mu_Y}{4/\sqrt{n}}\right) \ge 0.95$$
$$Pr\left(|Z_n| \le \frac{0.04\,\mu_Y\sqrt{n}}{4}\right) \ge 0.95,$$

where $Z_n = \frac{\widehat{\mu}_Y - \mu_Y}{4/\sqrt{n}}$ is a standardized Gaussian random variable. It must be that

$$\frac{0.04\,\mu_Y\sqrt{n}}{4} \ge 2 \qquad \text{(A more accurate right side is 1.96.)}.$$

Taking the minimum value $\mu_Y = 60$, so we obtain a conservative lower bound on $n$,

$$\frac{60\sqrt{n}}{100} \ge 2$$
$$\sqrt{n} \ge \frac{200}{60} = \frac{10}{3}$$
$$n \ge \frac{100}{9} = 11.11, \quad \text{which leads to } n \ge 12.$$

3. **[Derive the Unexpected from a Uniform PDF] (2 parts. 3/2 points.)**

You wish to use $X \sim U[0,1)$ to produce a different *nonnegative* random variable $Y = -\frac{1}{\lambda}\ln(1-X)$, for $0 \le X < 1$, where $\lambda$ is a positive constant, and ln is the natural logarithm function.
(Note that the pdf for $X \sim U[0,1)$ is the same as for $X \sim U[0,1]$.)

(a) Determine the CDF $F_Y(y) = Pr[Y \le y]$. [It may be useful to recall that $F_x(x) = x$ for $x \in [0,1)$.]
**Answer:** The CDF for $Y$ is

$$F_Y(y) = Pr(Y \le y) = \begin{cases} 0 & y < 0 \\ Pr\left(-\frac{1}{\lambda}\ln(1-X) \le y\right) & y \ge 0. \end{cases}$$

For $y \ge 0$, we have

$$\begin{aligned} F_Y(y) &= Pr\left(-\frac{1}{\lambda}\ln(1-X) \le y\right) = Pr(-\ln(1-X) \le \lambda y) \\ &= Pr(-\lambda y \le \ln(1-X)) = Pr\left(e^{-\lambda y} \le 1-X\right) \\ &= Pr\left(X \le 1-e^{-\lambda y}\right) \\ &= F_X(1-e^{-\lambda y}) \\ &= 1-e^{-\lambda y}, \end{aligned}$$

where the last equality is due to the fact that $F_X(x) = x$ for $0 \le x < 1$.

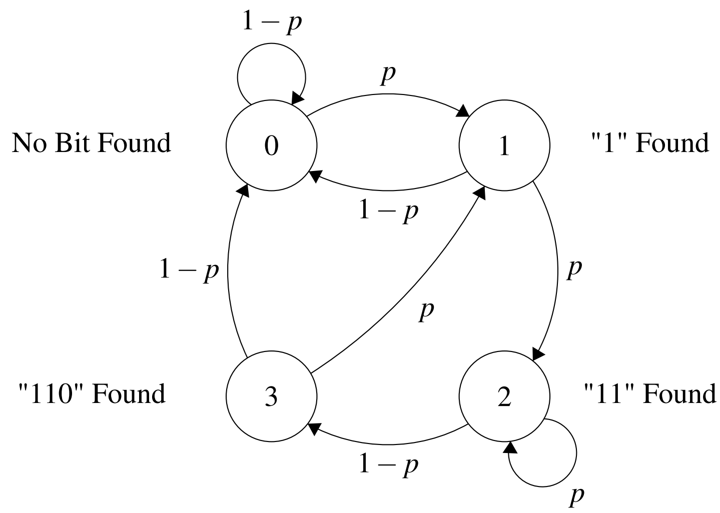
(b) Determine the PDF $f_Y(y)$ and indicate what standard distribution it corresponds to.
**Answer:** Therefore,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} 0 & y < 0 \\ \lambda\, e^{-\lambda y} & y \ge 0. \end{cases}$$

Clearly, $Y \sim Expo(\lambda)$—that is, $Y$ has an exponential PDF with parameter $\lambda$.

4. **[Finding a Three-Bit String in a Binary Bitsream] (3 parts. 2/5/5 points.)**

Consider a bitstream $B_1, B_2, \ldots$ consisting of IID Bernoulli random variables obeying the probabilities $Pr[B_n = 1] = p$, and $Pr[B_n = 0] = 1-p$, for every $n = 1, 2, \ldots$.
Here, $0 < p < 1$.
We begin parsing the bitstream from the beginning, in search of a desired binary string represented by the codeword $c = (1,1,0)$. We say that we've encountered the codeword $c$ at time $n$ if $(B_{n-2}, B_{n-1}, B_n) = (1,1,0)$. We model this process using the Markov chain shown below.




**Answer:**

There are four states, labeled 0,1,2, and 3. The state number $i$ represents the number of the leading (leftmost) bits of the codeword $c = 110$ for which we've found a match at time $n$—starting from the leading (leftmost) bit. For example, being in state 2, means you saw a 11 in the two latest bits.

That is, if $X_n$ denote the state of the process at time $n$ and and the bit-stream consists of $B_1, \ldots, B_n$. We have $X_n = 2$ when $(B_{n-1}, B_n) = 11$. We begin with $X_0$ in state 0 by default which corresponds to no prefix of the codeword $c = 110$ has been read.

~~(a) Provide a clear~~, succinct explanation as to why the Markov chain above has a set of unique limiting-state (i.e., stationary) probabilities:

$$\pi_i = \lim_{n \to \infty} Pr[X_n = i], \qquad i = 0, 1, 2, 3.$$

**Answer:** The chain has a single recurrent state (i.e., it is irreducible). Further more, the single recurrent state has at least one self-loop, so it is aperiodic. Therefore, the Markov chain has limiting-state probabilities $\pi_i$, $i = 0, 1, 2, 3$.

(b) Determine a simple expression for the limiting-state probability $\pi_3$ of State 3.
To receive full credit, you must explain your answer.
Depending on how you tackle this part, you may need only a small fraction of the space given to you below.
**Answer:**
**Method I: Brute Force** Write the Balance Equations and the Normalization Equation governing the limiting-state probabilities $\pi_0, \ldots, \pi_3$, and then solve. We can read the Balance Equations from the Markov Chain diagram:

$$\begin{aligned}
\pi_0 &= (1-p)\,\pi_0 + (1-p)\,\pi_1 + (1-p)\,\pi_3 \\
\pi_1 &= p\,\pi_0 + p\,\pi_3 \\
\pi_2 &= p\,\pi_1 + p\,\pi_2 \\
\pi_3 &= (1-p)\,\pi_2
\end{aligned}$$

$$\sum_{i=0}^{3} \pi_i = 1. \quad \text{Normalization Equation}$$

From class we know that the Balance Equations are linearly dependent; we don't need all of them. That's why we need the Normalization Equation. Using standard techniques to solve the linear equations yields $\pi_3 = p^2(1-p)$.

**Method II: Exploit the IID Nature of the Bits** At any time $n \geq 3$, it's possible for the Markov chain to be in State 3. In fact

$$Pr(X_n = 3) = Pr(B_n = 0, B_{n-1} = 1, B_{n-2} = 1).$$

But the bitstream's underlying process is an independent Bernoulli, which is what makes the bits independent. Accordingly, for every $n \geq 3$,

$$Pr(X_n = 3) = Pr(B_n = 0)\,Pr(B_{n-1} = 1)\,Pr(B_{n-2} = 1) = p^2(1-p).$$

In fact, we can say

$$Pr(X_n = 3) = \begin{cases} 0 & n = 1,2 \\ p^2(1-p) & n = 3,4,5,\ldots \end{cases}$$

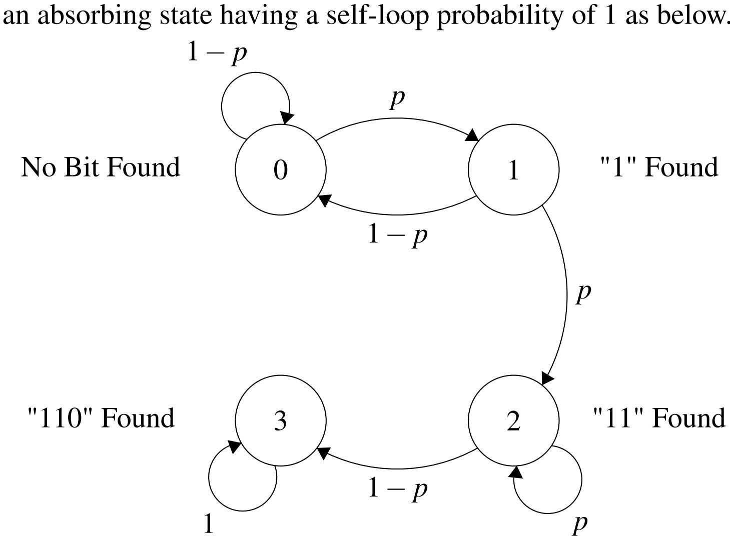
Hence, $\pi_3 = p^2(1-p)$.

**Method III: Long-Term Average View** Another way involves thinking of the stationary distributions as "long-term averages."

According to this view, for a string of length $n$, we can compute the expected number of 110's as $p^2(1-p)(n-2)$ out of $n$.

(c) For the remainder of this problem, we want to find the *expected time* $E(N)$ until the first occurrence of the string $c = 110$ in the bitstream.

Accordingly, we remove all the outgoing edges from State 3 in the original Markov chain, and turn State 3 into an absorbing state having a self-loop probability of 1 as below.




Determine $E(N)$, the expected time at which we first enter State 3—that is, the time at which the string $c = (1,1,0)$ occurs for the first time.

**Hint:** We recommend that you break down $N$ into two parts. Let $N = N_{02} + N_{23}$, where $N_{02}$ denotes the number of steps until first passage into State 2, starting from State 0, and $N_{23}$ denotes the number of steps it takes to transition for the first time from State 2 to State 3. Show that

$$E(N_{02}) = \frac{1}{p} + \frac{1}{p^2},$$

determine $E(N_{23})$, and put your results together to obtain $E(N)$.

**Answer:** Let $a_i$ be the expected number of steps until the Markov chains reaches the absorbing State 3 for the first time, *given* that the initial state is $i$. Since in this process we begin in State 0, we're after $E(N) = a_0$.

Write the system of three equations in three unknowns governing $a_0$, $a_1$, and $a_2$ (note that by definition $a_3 = 0$), and solve for $a_0$.

$$\begin{aligned} a_0 &= 1 + (1-p)\,a_0 + p\,a_1 \\ a_1 &= 1 + (1-p)\,a_0 + p\,a_2 \\ a_2 &= 1 + p\,a_2 \end{aligned}$$

**Method I: Brute Force** Using standard techniques, we solve the system of three equations in three unknowns to arrive at

$$a_0 = E(N) = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{1-p}.$$

**Method II: Notice the Salient Feature of State** 2 The last equation in our system of equations yields an immediate expression for $a_2$—namely,

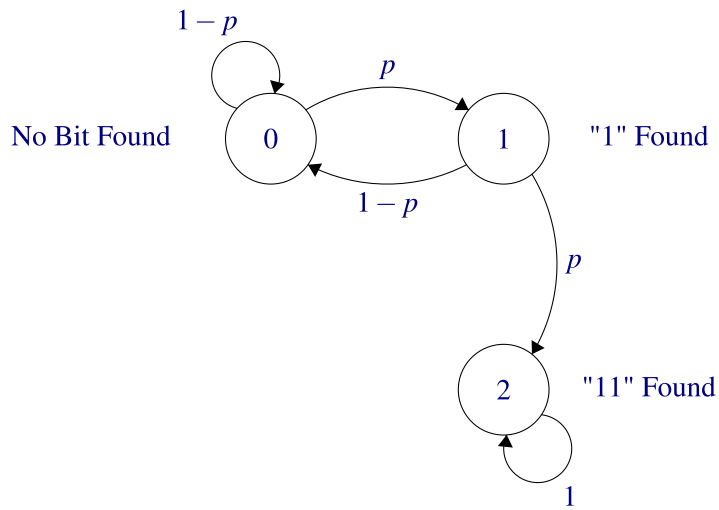
$$a_2 = \frac{1}{1-p}.$$

This should not be surprising, especially in light of the Markov chain diagram where State 3 is absorbing. State 2 is a critical transitional state here. Once we enter State 2, the expected absorption trajectory length turns into a first-order interarrival problem involving a Geometric random variable.

To see this, notice that the first transition from State 2 to State 3 occurs when we encounter a 0 bit for the first time (after arriving in State 2). This happens with probability $1-p$—which we can think of as a probability of "success." Hence, once we're in State 2, the expected number of steps up to, and including, the first "success" (i.e., the first enounter with a 0 bit) is simply the expected length of the first arrival in a Geometric distribution—that is, $\dfrac{1}{1-p}$.

We can now break the original problem into two segments. In particular, we can think of $N = N_{02} + N_{23}$, where $N_{02}$ is the number of steps for first passage into State 2 (starting from State 0), and $N_{23}$ as the number of steps to transition from State 2 to State 3 for the first time—and ultimate absorption. Accordingly, $E(N) = E(N_{02}) + E(N_{23})$. We know $E(N_{23}) = \dfrac{1}{1-p}$. So we must determine $E(N_{02})$. This problem has a reduced size, relative to the original.

**Method II(a)** Since the random variable $N_{02}$ refers to the number of steps, starting from State 0, to enter State 2 for the first time, we redraw the Markov chain by removing State 3 and turning State 2 into an absorbing state. The redrawn Markov chain is shown below.




We now determine the expected absorption time to State 2, which is precisely the $E(N_{02})$ we're after. To this we can then add $E(N_{23}) = \frac{1}{1-p}$ to arrive at the solution to the original problem.
Let $\mu_i$ denote the expected time to absorption in State 2, given that the chain starts in State $i$. Clearly, $\mu_0 = E(N_{02})$, $\mu_2 = 0$, and we have the following system of two equations in two unknowns:

$$\mu_0 = 1 + (1-p)\,\mu_0 + p\,\mu_1$$
$$\mu_1 = 1 + (1-p)\,\mu_0.$$

Solving this system yields

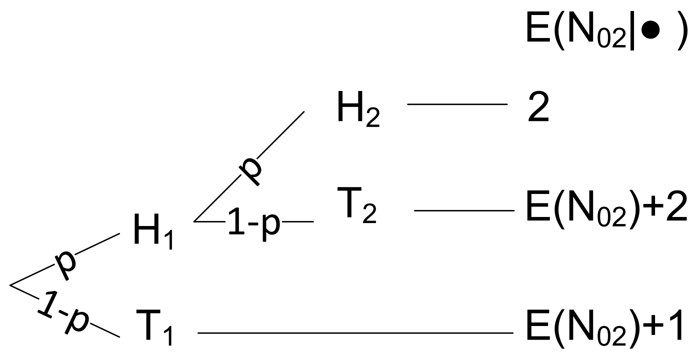
$$\mu_0 = E(N_2) = \frac{1}{p} + \frac{1}{p^2}.$$

Therefore, we arrive at the final solution

$$a_0 = E(N) = E(N_{02}) + E(N_{23}) = \mu_0 + \frac{1}{1-p}$$
$$E(N) = \underbrace{\frac{1}{p} + \frac{1}{p^2}}_{\text{Expected first passage time to State 2}} + \underbrace{\frac{1}{1-p}}_{\text{Expected absorption time into State 3, from State 2}}.$$

**Method II(b)** The expected time for first passage into State 2 (which corresponds to two consecutive 1 bits) can also be thought of as the expected time to the first occurrence of two consecutive Heads in an IID sequence of coin tosses, where the probability of Heads is $p$. So, think of a Heads as encountering a 1 bit. The tree diagram of the sample space, where we condition on the outcome of the first coin toss (i.e., the first bit encountered in the bitstream) allows us to determine the $E(N_{02})$.

Tosses Until Two Consecutive Heads.pdf Tosses Until Two Consecutive Heads.pdf




Clearly,

$$E(N_{02}) = 2p + p(1-p)[E(N_{02})+2] + (1-p)[E(N_{02})+1].$$

Solving for $E(N_{02})$, we obtain:

$$E(N_{02}) = \frac{1+p}{p^2} = \frac{1}{p} + \frac{1}{p^2}.$$

We then add the $\frac{1}{1-p}$ to obtain the expected absorption time in State 3:

$$\begin{aligned} E(N) &= E(N_{02}) + \frac{1}{1-p} \\ &= \frac{1}{p} + \frac{1}{p^2} + \frac{1}{1-p}. \end{aligned}$$