Midterm Exam

Last name	First name	SID

Rules.

- You have 80 mins (5:10pm - 6:30pm) to complete this exam.
- The exam is not open book, but you are allowed half a sheet of handwritten notes; calculators will be allowed.
- No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.

Please read the following remarks carefully.

- Show all work to get any partial credit.
- Take into account the points that may be earned for each problem when splitting your time between the problems.

Problem	Points earned	out of
Problem 1		40
Problem 2		25
Problem 3		25
Problem 4		10
Extra Credit		10
Total		100+10

Problem 1 [40] Answer the following problems briefly but clearly.

(a) [10] Let X be a continuous random non-negative variable (i.e. it has a density function, $f_X(x)$) with strictly increasing cdf F_X. Find the density functions of $Y = \sqrt{X}$ and $Z = F_X(X)$.

Solution:

Let $F_Y(y)$ be the cdf of Y. We have

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(\sqrt{X} \leq y) = \mathbf{P}(X \leq y^2) = F_X(y^2).$$

Differentiating with respect to y, we obtain $f_Y(y) = 2y f_X(y^2)$.

Note that Z can only take values in the interval $[0,1]$. Using a similar idea as above, we have

$$F_Z(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(F_X(X) \leq z) = \mathbf{P}(X \leq F_X^{-1}(z)) = F_X(F_X^{-1}(z)) = z.$$

Differentiating, we obtain $f_Z(z) = 1$ for $z \in [0,1]$. Thus, Z is a uniform random variable over $[0,1]$.

(b) [10] X, Y, Z are iid uniform r.v. over $[0,1]$.

1. Determine the density function for $U = X + Y + Z$
2. Let $V = XY$ and find the joint pdf $f_{VZ}(v,z)$.

Solution:

1. Let $r(t)$ be the distribution of X. The distribution of $X+Y+Z$ is given by the convolution $r \star r \star r$. A direct calculation gives

$$r(t) = \begin{cases} \frac{t^2}{2} & \text{if } 0 \leq t \leq 1, \\ \frac{-2t^2+6t-3}{2} & \text{if } 1 \leq t \leq 2, \\ \frac{(3-t)^2}{2} & \text{if } 2 \leq t \leq 3. \end{cases}$$

2. We first find the cdf of $F_V(v)$.

$$\mathbf{P}(V \leq v) = \mathbf{P}(XY \leq v) = \mathbf{P}(X \leq v) + \int_{x=v}^{1} \mathbf{P}(Y \leq \frac{v}{x})dx = v + \int_{x=v}^{1} \frac{v}{x}dx = v - v\log v.$$

Differentiating, we arrive at $f_V(v) = -\log v$ for $v \in [0,1]$.

Since V and Z are independent, their joint distribution is given by

$$f_{VZ}(v,z) = \begin{cases} -\log v & \text{if } 0 \leq v, z \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) [10] An expensive blood test must be performed on n individuals. The probability that any individual is infected is p. Each person can be tested separately but to save money, the individuals are grouped groups of k (assume that $\frac{n}{k}$ is an integer), and the test applied to each group. If the test is negative, all the individuals in the pool are negative. If it tests positive at least one of the individuals is positive and k more tests have to be performed to determine the status of each of the members of the group.

Find the expected number of tests necessary to test the n individuals under this strategy.
Solution:
For $1 \leq i \leq n/k$, let X_i be the number of tests required for group i. We are asked to find $\mathbb{E}[X_1 + \cdots + X_{n/k}]$. Since all groups are of the same size, using linearity of expectation we conclude that this is simply $\frac{n}{k}E[X_1]$. We have

$$X_1 = \begin{cases} 1 & \text{if test is negative} \\ 1+k & \text{if test is positive.} \end{cases}$$

The test is negative is no one is infected, which happens with probability $(1-p)^k$. The test is positive with probability $1-(1-p)^k$. Thus we have

$$E[X_1] = (1-p)^k + (1+k)(1-(1-p)^k),$$

and the expected number of tests is given by

$$\frac{n}{k}\left[(1-p)^k + (1+k)(1-(1-p)^k)\right] = \frac{n}{k}\left[1+k-k(1-p)^k\right].$$

(d) [10] Let X and Y be random variables and define $U = X+Y$ and $V = X-Y$.

1. Find $cov(U,V)$.
2. If X,Y are exponentially distributed with parameters λ_x and λ_y respectively, can U and V be independent?

Solution: The covariance of U and V is given by

$$\begin{aligned} cov(U,V) = E[UV] - E[U]E[V] &= E[(X+Y)(X-Y)] - (E[X]+E[Y])(E[X]-E[Y]) \\ &= E[X^2] - E[Y^2] - E[X]^2 + E[Y]^2 \\ &= var(X) - var(Y). \end{aligned}$$

In order for U,V to be independent X and Y must be uncorrelated, so they would have to be identically distributed exponentials with parameter λ. But that isn't enough!

If they are independent,

$$P(U \leq u, V \leq v) = P(U \leq u)P(V \leq v)$$

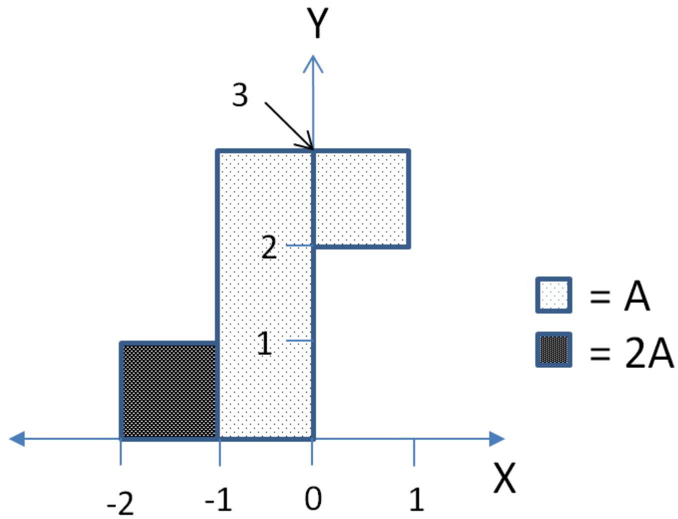
for all u, v so let's pick $v = u$ and see what happens:

$$P(X+Y \leq u, X-Y \leq u) = P(U \leq u)P(V \leq u)$$

Then since X,Y are exponentials they are non-negative and LHS simplifies:

$$P(U \leq u) = P(U \leq u)P(V \leq u)$$

But this isn't true for any finite value of u so U and V can;t be independent.



Problem 2 [25] X and Y are two random variables with joint distribution $f_{XY}(x,y)$ as shown in the figure above. Find the value of A, $cov(X,Y)$ and determine $f_{X|Y}(x|y)$ for all values of y.

Solution: From the pdf, we see that $2A+3A+A=1$, and therefore $A=\frac{1}{6}$.

To quickly solve for the covariance, we condition on the events E_1, E_2 and E_3 as shown in

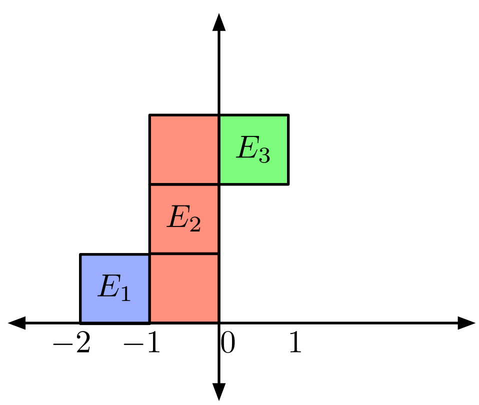


Figure 1: Conditioning on events E_1, E_2 and E_3

Figure 1. Note that conditioned on E_i, the random variables are uniformly distributed on the corresponding rectangle, and are thus independent. This way, calculating expectation becomes easy:

$$\begin{aligned}
E[XY] &= \mathbf{P}(E_1)E[XY|E_1] + P(E_2)E[XY|E_2] + P(E_3)E[XY|E_3] \\
&= \frac{2}{6} \times (-1.5 \times 0.5) + \frac{3}{6} \times (-0.5 \times 1.5) + \frac{1}{6} \times (0.5 \times 2.5) \\
&= \frac{-1.5 - 2.25 + 1.25}{6} = \frac{-2.5}{6} = \frac{-5}{12}.
\end{aligned}$$

Similarly, we may calculate

$$
\begin{aligned}
E[X] &= \frac{2}{6} \times -1.5 + \frac{3}{6} \times -0.5 + \frac{1}{6} \times 0.5 \\
&= \frac{-4}{6} = \frac{-2}{3}
\end{aligned}
$$

and

$$
\begin{aligned}
E[Y] &= \frac{2}{6} \times 0.5 + \frac{3}{6} \times 1.5 + \frac{1}{6} \times 2.5 \\
&= \frac{8}{6} = \frac{4}{3}.
\end{aligned}
$$

This gives

$$cov(X,Y) = \frac{-5}{12} - \frac{-2}{3} \times \frac{4}{3} = \frac{17}{36}.$$

The conditional distribution $f_{X|Y}(x|y)$ is given by

$$
f_{X|Y}(x|y) = \begin{cases}
\begin{cases} \frac{2}{3} & \text{if } \; -2 \le x \le -1 \\ \frac{1}{3} & \text{if } \; -1 \le x \le 0 \\ 0 & \text{otherwise} \end{cases} & \text{if } \; 0 \le y \le 1, \\
\begin{cases} 1 & \text{if } \; -1 \le x \le 0 \\ 0 & \text{otherwise} \end{cases} & \text{if } \; 1 \le y \le 2, \\
\begin{cases} \frac{1}{2} & \text{if } \; -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} & \text{if } \; 2 \le y \le 3.
\end{cases}
$$

Problem 3: Coin Sequences[25]

(a) [6] Bob flips a fair coin coin. Let X be the number of flips until he gets the sequence HH (gets two consecutive heads). Let Y be the number of flips until he gets the sequence TH. Find $E[X]$ and $E[Y]$.

Solution: If Bob gets a tail on his first flip he is back to square 1, but if he gets a head, things depend on the next toss. This suggests that we break the sample space in 3 disjoint events based as follows: E_1 is the even that the first coin toss is a T, E_2 is the event that the first 2 coin tosses are HT, E_3 is the even that the first two coin tosses are HH. We have

$$\begin{aligned} E[X] &= \mathbf{P}(E_1)E[X|E_1] + P(E_2)E[X|E_2] + P(E_3)E[X|E_3] \\ &= \frac{1}{2} \times (1 + E[X]) + \frac{1}{4} \times (2 + E[X]) + \frac{1}{4} \times 2. \end{aligned}$$

Solving the resulting linear equation gives $E[X] = 6$.

To find $E[Y]$ observe that if Bob gets a Tails on his first flip, then he is done after he flips the next Heads. If he gets a Heads then he is back to square 1. So:

$$E[Y] = .5(1+2) + .5(1 + E]Y])$$

$$.5E[Y] = 1.5 + .5 \Rightarrow E[Y] = 4$$

A longer way to say basically the same thing is the following: To find $E[Y]$, we partition the sample space into the following sequence of events: E_0 is the event that the first coin toss is H. For $i \geq 1$, E_i is the event the first i tosses are T, followed by H on the $i+1$-th toss.

$$\begin{aligned} E[Y] &= \sum_{i=0}^{n} \mathbf{P}(E_i)E[Y|E_i] \\ &= \frac{1}{2}(1 + E[Y]) + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \dots \\ &= \frac{1}{2}(1 + E[Y]) + \frac{3}{2}. \end{aligned}$$

Solving the linear equation gives $E[Y] = 4$.

(b) [6] Bob plays the following game with Alice who is watching him flip a fair coin. Bob wins if the sequence HT comes before HH, and Alice wins otherwise. What is the probability that Bob will win this game?

Solution: For $i \geq 1$, let E_i be the event that the first time H appears is on the i-th toss. Using the total probability rule,

$$\begin{aligned} \mathbf{P}(\text{Bob wins}) &= \sum_{i=1}^{\infty} \mathbf{P}(E_i)\mathbf{P}(\text{Bob wins}|E_i) \\ &= \sum_{i=1}^{\infty} \frac{1}{2^i} \times \frac{1}{2} \\ &= \frac{1}{2}. \end{aligned}$$

(c) [6] Bob flips a coin which lands heads with probability p. What is the probability that he will observe r Heads before he observes s Tails?

Solution: The key point to note is that by the time the coin has been tossed $r+s-1$ times, either r heads have occurred or s tails have occurred, but not both. From this, it is easy to conclude that the required probability is

$$(*) : \sum_{i=r}^{r+s-1} p^i (1-p)^{r+s-1-i} \binom{r+s-1}{i}.$$

Another way to solve this is as follows: Let E_i be the event that the r-th Heads comes on the i-th flip. Note that this event is possible only when $i \geq r$. The probability of the event of r Heads before s Tails occurs is simply

$$(**) : \sum_{i=r}^{r+s-1} \mathbf{P}(E_i) = \sum_{t=0}^{s-1} \binom{r-1+t}{r-1} p^r (1-p)^t.$$

It is surprising but true that the answers $(*)$ and $(**)$ evaluate to the same number.

(d) [7] The bias of Bob's coin (prob coin will land on heads) is uniformly distributed over $[0, 1]$. He tosses the coin n times. What is the expected value and variance of, X, the total number of heads?

Solution: Let X_i be the random variable with equals 1 if the i-th coin toss is H, and 0 if it is T. The total number of H's in n coin tosses is

$$X = X_1 + X_2 + \cdots + X_n.$$

Let Y be the bias of Bob's coin. Note that Y is uniformly distributed over $[0, 1]$. Using iterated conditional expectation, we have

$$\begin{aligned}
E[X] &= E[E[X|Y]] \\
&= E[E[X_1 + X_2 + \cdots + X_n|Y]] \\
&= E[nY] \\
&= nE[Y] = \frac{n}{2}.
\end{aligned}$$

Using the law of total variance,

$$\begin{aligned}
Var(X) &= E[Var(X|Y)] + Var(E[X|Y]) \\
&= E[Var(X_1 + \cdots + X_n|Y)] + Var(nY) \\
&= E[ny(1-y)] + n^2 \frac{1}{12} \\
&= \frac{n}{6} + \frac{n^2}{12} \\
&= \frac{n(n+2)}{12}.
\end{aligned}$$

Problem 4 [15]

A company has created a test to determine if a worker is proficient at a particular task. There are N workers, m of who are proficient. Workers are picked at random and tested, and X is the number of workers needed to be tested in order to find some fixed number, $k(\leq m)$, proficient workers. Assume that once a worker has been tested they are removed from the pool and cannot be selected again.

(a) [8] Find $p_X(x)$. Your answer for the first part should have no summations but can have terms such as $\binom{N+m}{k+r}$. Hint: Find the right quantity to condition on.

Consider the event that the xth person tested is proficient. This event has probability $\frac{m}{N}$. Now conditioned on that event we require that of the remaining $x-1$ workers, there must $k-1$ proficient workers and $x-k$ non-proficient ones. There are $\binom{m-1}{k-1}\binom{N-m}{x-k}$ ways for this to occur. The total number of ways to pick the $x-1$ workers is $\binom{N-1}{x-1}$ This yields

$$P(X=x) = \frac{\binom{m-1}{k-1}\binom{N-m}{x-k}}{\binom{N-1}{x-1}}\frac{m}{N}$$

(b) [7] Find $E[X]$. Hint: Use linearity of expectations.

Let X_j be the number of non-proficient workers tested between the $j-1$st and jth proficient worker tested. Consider $E[X_j]$: Number the non-proficient workers and let Y_i be the indicator r.v. that is 1 when non-proficient worker i is tested after the $j-1$st and before the jth proficient worker. $P(Y_j=1) = \frac{1}{m+1}$ since the relative position of i is fixed with respect to the m proficient workers. Thus $E[X_j] = P(X_j=1) = \frac{N-m}{m+1}$

Now clearly, $X = X_1 + ... + X_k + k$ but from the above and linearity of expectations:

$$E[X] = k(\frac{N-m}{m+1}+1) = k\frac{N+1}{m+1}$$

Extra Credit: [10] No partial credit for this problem.

Acme Inc has the following interview policy for filling a single position. n candidates are interviewed one at a time. Each candidate receives a score (no ties and a higher score is better). A candidate is either accepted or rejected for the job immediately after their interview. The company decides to interview and reject the first m candidates. After the mth candidate, the policy is to hire the first candidate who scores higher than the all the previous candidates interviewed.

Let E be the event that the best candidate out of the n is hired, and let E_i be the event that the ith candidate is best candidate and is hired. Find $P(E_i)$ and $P(E)$. Note, $i > m$. Also, you can leave your answer as a sum over i or try to bound it.

Let F_i be the event that of the first i candidates interviewed, the second best candidate was among first m candidates interviewed. Also, let B_i be the event that i is the best candidate in the first m interviewed. Then

$$P(E_i) = P(B_i \cap F_i) = P(F_i|B_i)P(B_i)$$

$P(B_i) = \frac{1}{n}$, and since the second best candidate has to be in the first m places out of $i-1$ places: $P(F_i|B_i) = \frac{m}{i-1}$. So

$$P(E_i) = \frac{1}{n}\frac{m}{i-1}$$

$$P(E) = \frac{m}{n}\sum_{i=m+1}^{n}\frac{1}{i-1}$$

You can bound this by

$$\frac{m}{n}(\ln n - \ln m) \leq P(E) \leq \frac{m}{n}(\ln(n-1) - \ln(m-1))$$

but this is optional.