This problem set essentially reviews notions of conditional expectation, conditional distribution, and Jointly Gaussian random variables. Not all exercises are to be turned in. Only those with the sign ★ are due on Thursday, September 21 at the beginning of the class. Although the remaining exercises are not graded, you are encouraged to go through them. We will discuss some of the exercises during discussion sections. Please feel free to point out errors and notions that need to be clarified.

We give solutions for only the problems that were required to be turned in. You are encouraged to try the other problems and discuss your solutions during office hours if needed.

**Exercise 1: Linear Algebra**

In this exercise, we will review linear system of equations by studying the conditions under which there exists a solution. Some exercises require using matlab or other math software. The following matlab functions can be useful in this exercise: *inv, det, pinv, eig, svd, range, rank, rref, null* and the ones you might find yourself...Remember: to get help on one command, type *help command*.

Let \( Ax = b \) be a system of \( m \) linear equations and \( n \) unknowns, where \( A \) is the \( m \times n \)-matrix of coefficients, \( b \) the vector of known terms, \( x \) is the vector of unknowns. If the \( \text{rank}(A)=r \), then there exist at least one \( r \times r \) sub matrix of \( A \) that is not singular, we choose one and call it \( M \). We call the equations that correspond to the rows of \( M \) main equations and side equations the others. The unknowns that correspond to columns of \( M \) are called main unknowns and the others are called sides unknowns.

The characteristic matrix, associated with a particular side equation, is the matrix formed by adding to the main matrix:
1. a row at the bottom, with the coefficients of the main unknowns in that side equation
2. a column at the right, with the known terms of the main equations and of the known term of that side equation.

The characteristic determinant is the determinant of the characteristic matrix. So there are as much characteristic determinants as side equations.

In the following examples, in addition to specific questions (for each example), you are asked to:
1. find the main and side equations, the main and side unknowns
2. determine whether the matrix $A$ is full rank or not
3. give a guess on how many solutions exist
4. compute all the characteristic determinants
5. compute the row echelon form of $(A|b)$
6. confirm or infirm your guess in 3)
7. write down the solution(s) if there is any.

**Example 1:** Cramer system

$$A = \begin{pmatrix} 2 & 1 & -1 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \quad b = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$$

Give answers to the questions 1-7.
What can you conclude for such kind of linear system?

**Example 2:** Fat matrix

$$A = \begin{pmatrix} 1 & -3 & 1 \\ 1 & 1 & 2 \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

Give answers to the questions 1-7.
Now suppose that $b$ is in the form
$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

For what value(s) of $b_1$ and $b_2$ a solution exits?
What can you conclude for such kind of linear system?

**Example 3:** Tall matrix

$$A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & -2 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix} \quad b_1 = \begin{pmatrix} 20 \\ 0 \\ 5 \\ 10 \end{pmatrix} \quad b_2 = \begin{pmatrix} 6 \\ 2 \\ 4 \\ 2 \end{pmatrix}$$

Give answers to the questions 1-7 for both $b_1$ and $b_2$.
What can you conclude for such kind of linear system?
We will continue this exercise in example 5.
Example 4: Rank deficient matrix

\[
A = \begin{pmatrix}
1 & 0 & 1 \\
1 & -1 & 2 \\
4 & -1 & 5
\end{pmatrix}
\quad b_1 = \begin{pmatrix}
4 \\
5 \\
17
\end{pmatrix}
\quad b_1 = \begin{pmatrix}
4 \\
5 \\
10
\end{pmatrix}
\]

Give answers to the questions 1-7 for both \(b_1\) and \(b_2\).

What can you conclude for such kind of linear system?

Example 5: Least-Square Solution

Re-consider example 3.

For which \(b_i\), \(i = 1, 2\) there was a solution?

Can you write \(b_i\), \(i = 1, 2\) as a linear combination of the columns of \(A\)?

Conclude!

Consider the \(b_i\) for which there is no solution call it \(b_o\). We will compute an approximative solution. To be more explicit, we will compute the Least-Square solution. In fact since \(Ax - b_o \neq 0\), \(\|Ax - b_o\| > 0\). We would like to find \(y \in \mathbb{R}^n\) such that

\[
\|Ay - b_o\| \leq \|Au - b_o\| \quad \forall u \in \mathbb{R}^n
\]

Method 1:

Compute the projection of \(b_o\) into the space of the rows of \(A\), call it \(proj(b_o)\). Now answer the 7-questions for the system \(Ax = proj(b_o)\).

Is there a solution? If yes write down the solution(s).

Method 2:

Answer the questions 1), 2), and 3) for the system \(A^*Ax = A^*b_o\).

The matrix \(A^*A\) on the left-hand side is a square matrix, which is invertible if \(A\) has full column rank (that is, if the rank of \(A\) is \(n\)). In that case, the solution of the system of linear equations is unique and given by

\[
x = (A^*A)^{-1}A^*b_o
\]

Compute the solution if it exists.

The matrix \((A^*A)^{-1}A^*\) is called the pseudo inverse of \(A\). It is returned by the matlab function \(pinv\) applied to \(A\). Use this function to verify the previous result.

Method 3:

Compute a svd of \(A = U\Sigma V^*\). Considering \(\Sigma^+\) (the transpose of \(\Sigma\) with every nonzero entry replaced by its reciprocal) as the inverse of \(\Sigma\), compute the ”solution” of the linear system \(U\Sigma V^*x = b_o\).

Did the results of all methods agree?

Which method is less/more computationally expensive?
Probability Space and Random Variable

Measurable Space
A measurable space \((\Omega, \mathcal{B})\) is a pair consisting of a sample space \(\Omega\) together with a \(\sigma\)-field \(\mathcal{B}\) of subsets of \(\Omega\) (also called the event space). A \(\sigma\)-field or \(\sigma\)-algebra \(\mathcal{B}\) is a collection of subsets of \(\Omega\) with the following properties:

1. \(\Omega \in \mathcal{B}\) \hfill (1.1)
2. If \(F \in \mathcal{B}\), then \(F^c = \{\omega : \omega \notin F\} \in \mathcal{B}\) \hfill (1.2)
3. If \(F_i \in \mathcal{B}, i = 1, 2, \ldots\), then \(\bigcup_i F_i \in \mathcal{B}\) \hfill (1.3)

Exercise 1.1. Use de Morgan’s law, show that

If \(F_i \in \mathcal{B}, i = 1, 2, \ldots\), then \(\bigcap_i F_i \in \mathcal{B}\)

An event space is a collection of subsets of a sample space (called events by virtue of belonging to the event space) such that any countable sequence of set theoretic operations (union, intersection, complement) on events produces other events.

Exercise 1.2. What is the largest possible \(\sigma\)-field of \(\Omega\)? What is the smallest possible \(\sigma\)-field of \(\Omega\)?

Probability Spaces
A probability space \((\Omega, \mathcal{B}, P)\) is a triple consisting of a sample space \(\Omega\), a \(\sigma\)-field \(\mathcal{B}\) of subsets of \(\Omega\), and a probability measure \(P(F)\) defined on the \(\sigma\)-field; that is, \(P(F)\) assigns a real number to every member \(F\) of \(\mathcal{B}\) so that the following conditions are satisfied:

Nonnegativity

\[ P(F) \geq 0 \quad \forall F \in \mathcal{B}, \] \hfill (1.4)

Normalization

\[ P(\Omega) = 1, \] \hfill (1.5)

Countable Additivity

If \(F_i \in \mathcal{B}, i = 1, 2, \ldots\) are disjoints, then

\[ P(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} P(F_i) \] \hfill (1.6)

A set function \(P\) satisfying only 1.4 and 1.6 but not necessarily 1.5 is called a measure, and the triple \((\Omega, \mathcal{B}, P)\) is called a measurable space.

Exercise 1.3. ★
Let \(\Omega\) be an arbitrary space. Suppose that \(\mathcal{F}_i, i = 1, 2, \ldots\) are all \(\sigma\)-fields of subsets of \(\Omega\). Define the collection \(\mathcal{F} = \bigcap_i \mathcal{F}_i\); that is, the collection of all sets that are in all of the \(\mathcal{F}_i\). Show that \(\mathcal{F}\) is a \(\sigma\)-field.
Solution:

\[ \Omega \in \mathcal{F}_i \forall i \Rightarrow \Omega \in \cap_i \mathcal{F}_i = \mathcal{F} \]

\[ A \in \cap_i \mathcal{F}_i \iff A \in \mathcal{F}_i \forall i \iff A^{c} \in \cap_i \mathcal{F}_i\]

\[ A_j \in \mathcal{F}_i, j = 1, 2, \ldots \iff A_j \in \mathcal{F}_i \forall i \iff \cup_j A_j \in \mathcal{F}_i \]

Thus the set \( \mathcal{F} \) is a \( \sigma \)-field.

Exercise 1.4. ★

Let \( \Omega = (0, 1) \), \( \mathcal{B} \) all countable unions of intervals of \( (0, 1) \), and \( P : \Omega \rightarrow \Omega \) with \( P((a, b)) = b - a \) for \( a < b \in (0, 1) \).

Show that \( P \) is a probability measure. Do you recognize it?

Prove the following elementary properties of probability measure:

1. \( P(F \cup G) = P(F) + P(G) - P(F \cap G) \)
2. \( P(F^{c}) = 1 - P(F) \)
3. For all events \( F \), \( P(F) \leq 1 \)
4. If \( G \subset F \), then \( P(F - G) = P(F) - P(G) \)

Solution:

Note that the proof given here is not rigorous, a more rigorous proof is the construction of the Lebesgue measure which needs more background on measure theory.

- \( \mu(A) \geq 0 \) by definition
- \( \mu(\emptyset) = 0 \)
- The additivity property is not easy to show. However if we have disjoint intervals \( (a_i, b_i], i = 1, 2, \ldots, n \) such that \( a_1 < b_1 = a_2 < b_2 = a_3 \cdots < b_{n-1} = a_n < b_n \), we can easily verify that \( \mu(\biguplus_i (a_i, b_i)) = \sum_i (b_i - a_i) = \sum_i \mu(a_i, b_i) \), where \( \biguplus \) means disjoint union.
- The other cases are left out for measure theory course!.

- This is the definition of the uniform random variable with the Lebesgue Measure.

We will show (4) first and then we use the result to show (1).

(4)-If \( G \in F \), then we can write \( F = G \cup (F \cap G^{c}) = G \cup (F - G) \) and \( G \cap (F - G) = \emptyset \). Thus \( P(F) = P(G \cup (F - G)) = P(G) + P(F - G) \) by the additivity property and the result follows.

(1)-We have that \( F \cup G = (F - (F \cap G)) \cup G \) and \( (F - (F \cap G)) \cap G = \emptyset \). Using the additivity property, we have \( P(F \cup G) = P(F - (F \cap G)) + P(G) \). Now we use (4) and the fact that \( (F \cap G) \in F \) to rewrite \( P(F \cup G) = P(F) - P(F \cap G) + P(G) \).
(2)- Notice that $\Omega = F \cup F^c$, thus $1 = P(\Omega) = P(F) + P(F^c)$.

(3)- Since $P(F^c) \geq 0$, we have $P(F) \leq 1$.

**Random Variables**

Given a measurable space $(\Omega, \mathcal{B})$, let $(\mathcal{A}, \mathcal{B}_A)$ denote another measurable space. A *random variable* or *measurable function* defined on $(\Omega, \mathcal{B})$ and taking values in $(\mathcal{A}, \mathcal{B}_A)$ is a mapping or function $f : \Omega \to \mathcal{A}$ with the property that:

$$\text{If } F \in \mathcal{B}_A, \text{ then } f^{-1}(F) = \{\omega : f(\omega) \in F\} \in \mathcal{B}$$

The name *random variable* is commonly associated with the case where $\mathcal{A}$ is the real line and $\mathcal{B}_A$ the Borel $\sigma$-field (the minimum set that contains all the open intervals of the real line).

A random variable is just a function or mapping with the property that inverse images of input events determined by the random variable are events in the original measurable space. This simple property ensures that the output of the random variable will inherit its own probability measure. For example, with the probability measure $P_f$ defined by:

$$P_f(B) = P(f^{-1}(B)) = P(\{\omega : f(\omega) \in B\}), \text{ for } B \in \mathcal{B}_A$$

$(\mathcal{A}, \mathcal{B}_A, P_f)$ becomes a probability space since measurability of $f$ and elementary set theory ensure that $P_f$ is indeed a probability measure. The induced probability measure $P_f$ is called the distribution of the random variable $f$.

**Exercise 1.5.** If $F \in \mathcal{B}$, show that the indicator function $1_F$ defined by $1_F(x) = 1$ if $x \in F$ and 0 otherwise is a random variable.

**Probability and Common Random Variables**

**Exercise 1.6.** Star Probability Space

1. Pick 3 balls without replacement from an urn with 15 identical balls (10 red and 5 blue balls). Specify the probability space.

2. You flip a coin until you get 3 consecutive "heads". Specify the probability space.

**Solution:**

(1)- One possible choice is to specify the color of the three balls in the order they are picked. Then

$$\Omega = \{R, B\}^3, \mathcal{F} = 2^\Omega, P(\{RRR\}) = \frac{10 \cdot 9 \cdot 8}{15 \cdot 14 \cdot 13}, \ldots, P(\{BBB\}) = \frac{5 \cdot 4 \cdot 3}{15 \cdot 14 \cdot 13}$$

(2)- Again here one possible choice is $\Omega = H, T^*$, where we consider all finite sequences of $H$ and $T$. Since this set is countable, we can consider $\mathcal{F} = 2^\Omega$. Observing that all the sequences of the same length are equally likely, we have:

$$P(\omega) = 2^n$$

where $n$ is the length of the sequence $\omega$. 

1-6
Exercise 1.7. Birthday Paradox
In a class of 23 students, pick two students at random. What is the probability that they
have the same birthday? What is the probability that at least two students have the same
birthday?
Hints: \( P(A) = 1 - P(A^c) \)

Exercise 1.8. Independence and conditioning

1. Draw a ball from an urn containing 4 balls, numbered 1, 2, 3, 4. Let \( E = \{1, 2\} \), \( F = \{1, 3\} \), \( G = \{1, 4\} \). Assume that all the outcomes are equally likely.
   Are the events \( E \), \( F \), and \( G \) pairwise independent? Are they mutually independent?
   Give an example of events that are pairwise independent but not mutually independent.

2. Let \( A \) and \( B \) be two independent events. Show that \( A \) and \( B^c \) are independent

3. If \( P(A) = 1 \) or \( 0 \), show that \( A \) is independent to all other events.

4. ★
   For the following statements, give a proof of correctness or a counterexample to show
   incorrectness (\( \perp \perp \) = independent)
   - If \( X_1 \perp \perp X_2 \), then \( (X_1 \perp \perp X_2)|X_3 \)
   - If \( (X_1 \perp \perp X_2)|X_3 \) for some \( X_3 \), then \( X_1 \perp \perp X_2 \)
   - If \( (X_1 \perp \perp X_2)|X_4 \) and \( (X_1 \perp \perp X_3)|X_4 \) then If \( (X_1 \perp \perp (X_2, X_3))|X_4 \)
   - If \( (X_1 \perp \perp (X_2, X_3))|X_4 \) then \( (X_1 \perp \perp X_2)|X_4 \)

5. Let \( X_1, \ldots, X_n \) be independent, identically distributed random variables with common
   mean and variance. Find the values of \( c \) and \( d \) that will make the following formula
   true:
   \[
   E \left[ (X_1 + \cdots + X_n)^2 \right] = cE[X_1]^2 + d(E[X_1])^2
   \]
   Solution:
   - If \( X_1 \perp \perp X_2 \), then \( (X_1 \perp \perp X_2)|X_3 \). False

\[
X_1 = \begin{cases} 
1 & wp \\ 0 & 1/2 
\end{cases} \quad X_2 = \begin{cases} 
1 & wp \\ 0 & 1/2 
\end{cases} \quad X_3 = X_1 \oplus X_2
\]

\( X_1 \) and \( X_2 \) are by definition independent. However given \( X_3 = 0 \), we know that \( X_1 = X_2 = 1 \) with probability 1.
• If \((X_1 \perp \perp X_2)\mid X_3\) for some \(X_3\), then \(X_1 \perp \perp X_2\) \textbf{False}

Let \(X_1, X_2, \text{ and } X_3\) be defined as follow:

\[
X_1 = \begin{cases} 
  P(X_1 = 0, X_2 = 3) = \frac{2}{3}, & X_3 = X_1 + X_2 = 3 \\
  P(X_1 = 0, X_2 = 5) = \frac{2}{9}, & \text{if } X_1 = 0, X_2 = 3 \\
  P(X_1 = 1, X_2 = 3) = \frac{1}{9}, & \text{if } X_1 = 1, X_2 = 3 \\
  P(X_1 = 0, X_2 = 5) = \frac{1}{9}, & \text{if } X_1 = 0, X_2 = 5 \\
\end{cases}
\]

In this example we see that given \(X_3\), \(X_1\) and \(X_2\) become deterministic and are thus independent (e.g. \(X_3 = 0\) implies that \(X_1 = 0\) and \(X_3\)). However, \(X_1\) and \(X_2\) are not independent.

• If \((X_1 \perp \perp X_2)\mid X_4\) and \((X_1 \perp \perp X_3)\mid X_4\) then If \((X_1 \perp \perp (X_2, X_3))\mid X_4\) \textbf{False}

For a counterexample, consider the one used in the first case and let \(X_4\) be the ”event that it will rain tomorrow”. We see that \(X_4\) has nothing to do with the other variables and we can drop the conditioning on \(X_4\). We have \((X_1 \perp \perp X_2)\) and \((X_1 \perp \perp X_3)\) (Verify that!). However, as was shown \((X_1 \not\perp \perp (X_2, X_3))\).

• If \((X_1 \perp \perp (X_2, X_3))\mid X_4\) then \((X_1 \perp \perp X_2)\mid X_4\) \textbf{True}

\[
(X_1 \perp \perp (X_2, X_3))\mid X_4 \Rightarrow P(X_1, (X_2, X_3)\mid X_4) = P(X_1\mid X_4)P(X_2, X_3\mid X_4)
\]

Now

\[
P(X_1, X_2\mid X_4) = \sum_{X_3} P(X_1, (X_2, X_3)\mid X_4)
\]

\[
= \sum_{X_3} P(X_1\mid X_4)P(X_2, X_3\mid X_4)
\]

\[
= P(X_1\mid X_4)\sum_{X_3} P(X_2, X_3\mid X_4)
\]

\[
= P(X_1\mid X_4)P(X_2\mid X_4)
\]

Thus \((X_1 \perp \perp X_2)\mid X_4\).

\begin{center}
\textbf{Distribution, Expectation, and Conditional Expectation}
\end{center}

Part 1: Discrete case

\textbf{Exercise 1.9.} A random variable is said to a Bernoulli random variable if its p.m.f is \(p(0) = P(X = 0) = 1 - p\), \(p(1) = P(X = 1) = p\). It models an experiment that has a probability of \(p\) to succeed.

If such an experiment is repeated independently \(n\) times and \(N\) counts the number of success, argue that \(N\) is a random variable, specify the corresponding probability space and compute the distribution of \(N\).
Now suppose that the experiment is repeated until the time $T$ of the first success. Argue that $T$ is a random variable, specify the corresponding probability space and compute its distribution.

Compute the expectation of $T$ by applying the expectation formula. Now compute it by conditioning on the outcome of the first experiment.

Let $Y = \sum_{i=1}^{N} T_i$ where the $T_i$’s are independent and identically distributed random variables and independent to $N$. $N$ and $T_1$ have respectively the distributions computed in the questions above. Compute the mean of $Y$. Compute its variance.

Exercise 1.10. ⋆

Suppose that $X$ and $Y$ are independent discrete random variables with the same geometric PMF:

$$p_X(k) = p_Y(k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

where $p$ is a scalar with $0 < p < 1$. Show that for any integer $n > 2$, the conditional PMF

$$P(X = k | X + Y = n)$$

is uniform.

Solution:

$$P(X = k | X + Y = n) = \frac{P(X = k, X + Y = n)}{P(X + Y = n)} = \frac{P(X = k, Y = n - k)}{P(X + Y = n)} = \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)}$$

where the last equality is obtained because $X$ and $Y$ are independent.

Now let us compute $P(X + Y = n)$

$$P(X + Y = n) = \sum_{m=1}^{n} P(X + Y = n | Y = m)P(Y = m)$$

$$= \sum_{m=1}^{n} P(X = n - m | Y = m)P(Y = m)$$

$$= \sum_{m=1}^{n} P(X = n - m)P(Y = m)$$

$$= \sum_{m=1}^{n} p(1-p)^{n-m-1}p(1-p)^{m-1}$$

$$= p^2(1-p)^{n-2} \sum_{m=1}^{n} 1$$

$$= p^2(1-p)^{n-2}n$$

1-9
Now going back to \( P(X = k | X + Y = n) \) we have:

\[
P(X = k | X + Y = n) = \frac{p(1 - p)^{k-1}p(1 - p)^{n-k-1}}{p^2(1 - p)^{n-2n}}
\]

\[
= \frac{1}{n}
\]

Thus \( X|X + Y = n \) is uniformly distributed in \( \{1, \ldots, n\} \).

Exercise 1.11. ★
A student shops for probability books for \( K \) hours, where \( K \) is a random variable that is equally likely to be 1, 2, 3, or 4. The number of books \( N \) that she buys is random and depends on how long she shops according to the conditional PMF

\[
p_{N|K}(n|k) = \frac{1}{k}, \text{ for } n = 1, \ldots, k.
\]

(a) Find the joint PMF of \( K \) and \( N \).
(b) Find the marginal PMF of \( N \).
(c) Find the conditional PMF of \( K \) given that \( N = 2 \).
(d) Find the conditional mean and variance of \( K \), given that she bought at least 2 but no more than 3 books.
(e) The cost of each book is a random variable with mean $30. What is the expected value of his total expenditure?

Solution:
(a)- Using

\[
P(N = n, K = k) = P(N = n|K = k)P(K = k) = P(N = n|K = k)\frac{1}{4}
\]

we obtain the following table for the joint PMF of \( N \) and \( K \).

\[
\begin{array}{c|cccc}
N \backslash K & 1 & 2 & 3 & 4 \\
\hline
1 & 1/4 & 1/4 & 1/4 & 1/4 \\
2 & 0 & 1/4 & 1/4 & 1/4 \\
3 & 0 & 0 & 1/4 & 1/4 \\
4 & 0 & 0 & 0 & 1/4 \\
\end{array}
\]

(b)- The marginal PMF of \( N \) is given by:

\[
\begin{array}{c|cccc}
P(N) & 1 & 2 & 3 & 4 \\
\hline
\frac{1}{48} & \frac{13}{48} & \frac{7}{48} & \frac{3}{48} \\
\end{array}
\]

(c)-

\[
P(K = k|N = 2) = \frac{P(K = k, N = 2)}{P(N = 2)} = \begin{cases} 0 & \text{if } k = 1 \\ \frac{1}{1} \frac{48}{13} & \text{if } k = 2, 3, 4 \end{cases}
\]
(d)- Let us first compute the conditional PMF of $K$ given that $2 \leq N \leq 3$.

| $k$ | $P(K = k | 2 \leq N \leq 3)$ |
|-----|---------------------|
| 1   | $\frac{1}{10}$     |
| 2   | $\frac{2}{5}$      |
| 3   | $\frac{3}{10}$     |

Hence $E[K | 2 \leq N \leq 3] = 3$ and $E[K | 2 \leq N \leq 3] = 0.6$

(e)- Let $Z$ be the expenditure of the student. We have:

$$Z = \sum_{i=1}^{N} Y_i$$

where $N$ is the number of book purchased and $Y_i$ is the price of book $i$. Now let us compute $E[Z]$.

$$E[Z] = E\left[\sum_{i=1}^{N} Y_i\right] = E\left[E\left[\sum_{i=1}^{N} Y_i | N\right]\right]$$

But

$$E\left[\sum_{i=1}^{N} Y_i | N = n\right] = E\left[\sum_{i=1}^{n} Y_i\right] = nE[Y]$$

This yields to

$$E\left[\sum_{i=1}^{N} Y_i | N\right] = NE[Y]$$

And thus

$$E[Z] = E\left[E\left[\sum_{i=1}^{N} Y_i | N\right]\right] = E\left[NE[Y]\right] = E[N]E[Y] = \frac{85}{48} \times 30$$

Part 2: Continuous case

Review the common random variables (uniform, exponential, normal, Gamma).

**Exercise 1.12.** Let $X$ be a random variable with PDF $f_X$. Find the PDF of the random variable $|X|$ in the following three cases.
1. $X$ is exponentially distributed with parameter $\lambda$.

2. $X$ is uniformly distributed in the interval $[-1, 2]$.

3. $f_X$ is a general PDF.

**Exercise 1.13. Markov's inequality**
Suppose $X$ is a continuous random variable that takes nonnegative values with pdf $f_X$.

1. Show that for any value $a > 0$
   \[ P(X \geq a) \leq \frac{E[X]}{a} \]

2. Show that if $f_X(x) \leq c$ for all $x$, then
   \[ P(X > a) \geq 1 - ac \]

**Exercise 1.14. ★**
Let $X$ be exponential with mean $\frac{1}{\lambda}$, find
\[ E[X|X > 1] \]

**Solution:**
This exercise illustrates the memoryless property of the exponential distribution (that we will see later in the course!)

Note that:
\[ F_{X|X>1}(x|X > 1) = P(X \leq x|X > 1) = 1 - P(X > x|X > 1) = \begin{cases} 0 & \text{if } x \leq 1 \\ 1 - \exp\{-\lambda(x - 1)\} & \text{else} \end{cases} \]

Thus we can derive the pdf of $X|X > 1$
\[ f_{X|X>1}(x|X > 1) = \begin{cases} 0 & \text{if } x < 1 \\ \lambda \exp\{-\lambda(x - 1)\} & \text{else} \end{cases} \]

and then compute the mean.
\[ E[X|X > 1] = 1 + \frac{1}{\lambda} \]

**Exercise 1.15. Matlab: Inversion Formula**
For random variables where there exist a closed form of the inverse distribution function $F^{-1}(U)$, one can generate samples of it by first generating samples of uniform random variables and applying the inversion formula $X = F^{-1}(U)$.

Generate random variables of the following pdfs (or pmf):

1. $f_X(x) = \lambda \exp \lambda x$
2. \( f_X(x) = \frac{\alpha}{\beta} \left( \frac{x}{\beta} \right)^{\alpha-1} \exp \left( -\frac{x}{\beta} \right)^\alpha \)

3. \( P(X = 0) = 0.6, \ P(X = 1) = 0.3, \ P(X = 2) = 0.1, \ P(X = i) = 0 \) elsewhere

Show that for a geometric(p) random variable, \( F(x) \) is given by:

\[
F(x) = 1 - (1 - p)^{\lfloor x+1 \rfloor}
\]

Generate a sample of geometric random variables.

For a general discrete random variable over \( x_1 < \cdots < x_k \), one way to generate it is to first build a table of pairs \((x_i, F(X_i))\), and then return \( X = x_I \) where \( I = \min\{i : F(x_i) \geq U\} \) for \( U \sim U(0,1) \).

Generate a sample of geometric random variable and compare the histogram to that of the previous question.

Gaussian random variables

**Exercise 1.16.** In the class notes we mentioned several reasons why Gaussian random variables are important. One of those was that "Gaussian random variables are worst case" i.e it has the maximum entropy among all the random variables having the same variance. In this exercise, we will illustrate this fact.

The differential entropy (or simply entropy) of a continuous random variable \( X \) with pdf \( f_X \) is defined as:

\[
h(X) = - \int_{-\infty}^{\infty} f_X(x) \log(f_X(x)) \, dx
\]

The entropy is a "quantization" of the amount of randomness carried by a random variable. If two random variables \( X \) and \( Y \) have respectively distribution \( f_X \) and \( f_Y \), then the differential entropy (also called Kullback-Leibler divergence) is defined as:

\[
D(f_X \| f_Y) = \int_{-\infty}^{\infty} f_X(x) \log\left( \frac{f_X(x)}{f_Y(x)} \right) \, dx
\]

One property of \( D(f_X \| f_Y) \) is its nonnegativity.

1) Show that \( D(f_X \| f_Y) \geq 0 \).
2) Compute the entropy of \( X \sim \mathcal{N}(\mu, \sigma^2) \).
3) Using the nonnegativity of \( D(f_X \| f_Y) \), show that for a fixed variance of \( X \), the maximum entropy is reached when \( X \) is Gaussian.

Hint: consider the zero-mean case and note that the variance is fixed!

**Exercise 1.17.** Matlab

Suppose you are given a generator of standard normal random variable (e.g. the function \texttt{randn} for matlab) and you are asked to generate a jointly Gaussian random vector with covariance \( K \). How would you proceed?

Use your method to generate a sample of 1000 Gaussian \( \mathcal{N}(1, 2) \) random variables. Generate
another sample using matlab and compare the two histograms.  
Repeat the above question of $X \sim \mathcal{N}(\mu, K)$ where 

$$K = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \quad \mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

**Exercise 1.18. Matlab: Box-Muller**

In exercise 1.15 you have learned how to generate an arbitrary random variable using uniform random variable. In this exercise we will study another method to generate Gaussian random variable. 

Generate two independent uniform random variables $U_1, U_2 \sim \text{Unif}(0, 1)$. Let $R = \sqrt{-2 \log(U_1)}$ and $\theta = 2\pi U_2$; and let $X = R \cos \theta$ and $Y = R \sin \theta$.

Argue that $X$ and $Y$ are two independent standard normal random variables.

Use this method to generate a sample of 1000 Gaussian $\mathcal{N}(1, 2)$ random variables. Generate another sample using the matlab provided function and compare the two histograms.

**Exercise 1.19.** Let $X_1, X_2, \ldots, X_n$ are iid Gaussian random variables. Show that $Y = X_1 + X_2 + \cdots + X_n$ is Gaussian. Specify the mean and variance.

**Exercise 1.20. ⋆**

Assume $X$ and $Y$ are iid Gaussian. Compute $E[(X + Y)^4 | X - Y]$

**Solution:**

Note that $X + Y$ and $X - Y$ are independent because they are jointly Gaussian and uncorrelated. Hence,


(term with odd power vanish, $X \perp \! \! \! \! \perp Y$)

**Exercise 1.21.** Let $X_n, n \geq 0$ be Gaussian $\mathcal{N}(0, 1)$ random variables. Assume that $Y_{n+1} = aY_n + X_n$ for $n \geq 0$ where $Y_0$ is a Gaussian random variable with mean zero and variance $\sigma^2$ independent of the $X_n$‘s and $|a| < 1$.

a) Calculate var($Y_n$) for $n > 0$. Show that var($Y_n$) $\to \delta^2$ as $n \to \infty$ for some $\delta^2$.

b) Find the value of $\sigma^2$ so that the variance of $Y_n$ does not depend on $n \geq 1$.

**Exercise 1.22. ⋆**

Gallager’s note: Exercise 2.1
Solution:
(a) - Let’s compute the cdf of $S = X^2 + Y^2$ first.

$$P(S \leq s) = \int_{-\sqrt{s}}^{\sqrt{s}} dx \int_{-\sqrt{s-x^2}}^{\sqrt{s-x^2}} dy f_{X,Y}(x,y)$$

$$= \int_{-\sqrt{s}}^{\sqrt{s}} dx \int_{-\sqrt{s-x^2}}^{\sqrt{s+x^2}} dy \alpha^2 e^{-\frac{2x^2}{\pi}}$$

$$= \int_0^{\sqrt{s}} r dr \int_{-\pi}^{\pi} d\theta \alpha^2 e^{-\frac{r^2}{\pi}} \text{ by going to polar coordinate}$$

$$= 2\pi \alpha^2 \int_0^{\sqrt{s}} r dr e^{-\frac{r^2}{\pi}}$$

$$= 2\pi \alpha^2 (1 - e^{-\frac{s}{\pi}})$$

Thus $f_S(s) = \alpha^2 \pi e^{-\frac{s}{\pi}}$ for $s \geq 0$.

(b) - In order for $S$ to be a random variable, we must have

$$\int_0^\infty f_S(s) ds = 1$$

so $\alpha^2 \pi (2 - 0) = 1 \Rightarrow \alpha = 1/\sqrt{2\pi}$

(c) - Use the transformation formula for $r = g(s) = \sqrt{s}$

$$f_R(r) = \frac{f_S(r^2)}{|g'(r^2)|} = \begin{cases} 
0 & \text{if } s < 0 \\
re^{-\frac{r^2}{\pi}} & \text{else}
\end{cases}$$

Exercise 1.23. ★
Gallager’s note: Exercise 2.3

Solution:
First, observe that $Y$ is Gaussian:

$$f_Y(y) = f_{Y|X>0}(y)P(X > 0) + f_{Y|X<0}(y)P(X < 0)$$

$$= 2f_Z(y)1_{\{y \geq 0\}} \frac{1}{2} + 2f_Z(y)1_{\{1-y \geq 0\}} \frac{1}{2}$$

$$= f_Z(y)$$

where we use $1_{\{y \geq 0\}}$ to designate the indicator function of the set $\{y : y \geq 0\}$.

To show that $X$ and $Y$ are not jointly Gaussian we will proceed by contradiction. Let’s assume that they are JG. Since $E[(X-Y)(X+Y)] = E[X^2] - E[Z^2] = 0$, $X+Y$ and $X-Y$ must be independent, and so must $|X+Y|$ and $|X-Y|$. But since $X+Y = X + sgn(X)|Z|$ is positive if and only if $X$ is positive (show this!), we have:

$$|X+Y| = |X + sgn(X)|Z|| = |X| + |sgn(X)|Z| \geq |X - sgn(X)|Z| = |X-Y|.$$
Thus $|X + Y|$ and $|X - Y|$ are not independent, which contradicts our assumption. The contours of equal joint pdf are shown in Figure 1.1.

Exercise 1.24. ★
Gallager’s note: Exercise 2.5

Solution:
Note that the MGF of $X$ is given by:

$$g_X(r) = E[\exp(rX)] = E[\exp(rs^TU)] = E[\exp((rs)^TU)] = g_U(rs) = \exp\left(rs^Tm + (rs)^TK(rs)\right)$$

Now computing the mean and variance of $X$, we obtain

$$E[X] = E[s^TU] = s^TE[U] = s^Tm$$

and

$$Var(X) = E[(X - E[X])(X - E[X])^T] = E[(s^TU - s^Tm)(s^TU - s^Tm)^T] = s^TE[(U - m)(U - m)^T]s = s^TK_Us$$
Hence we see that:

\[ g_X(r) = \exp \left( r E[X] + \frac{r^2}{2} Var(X) \right) \]

This is the MGF of a Gaussian random variable and thus \( X \) is Gaussian. In other words, we have shown that \( s^T U \) is Gaussian for any vector \( s \). This tells us that every linear combination of the elements of \( U \) is Gaussian which means that \( U \) is JG.