**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

16.09 Spring 2023
Recitation 1 1/7/2023

---

**COMBINATORICS: A REVIEW**

We will use the following notation

- $n! = n \times (n-1) \times (n-2) \times ... \times 1$
- $\binom{n}{r} = {}_nC_r = \frac{n!}{r!(n-r)!}$
- $\binom{n}{r_1\ r_2\ ...\ r_k} = \frac{n!}{r_1!r_2!...r_k!}$
- ${}_nP_r = \frac{n!}{(n-r)!}$

## 1 PERMUTATIONS

**Q:** How many ways can you order (or permute) $n$ distinct items?

**A:** An ordering or **permutation** of $n$ items assigns a unique position in the set $\{1, ..., n\}$ to each of the $n$ items. Arranging the items from left to right based on these positions results in a particular permutation.

To count the number of permutations, note that for the first item, we have $n$ possible positions that it can be assigned. For the second item, we have $n-1$ possible positions that it can be assigned, since we have already assigned one position to the first item. Repeating this argument $k$ times, the $k$th item can be assigned $(n-k+1)$ positions, since the $k-1$ items ahead of it have already been assigned unique positions which cannot be used again. Continuing this line of reasoning, we conclude that the total number of assignments possible, which is also the number of permutations, is simply $n \times (n-1) \times (n-2) \times ... \times 1 = n!$.

We can generalize this result to count the number of ways to choose and uniquely order $r$ items from a set of $n$ distinct objects.

**Lemma 1.** *The number of ways to choose and uniquely order $r$ items from a set of $n$ distinct objects is given by*

$${}_nP_r \triangleq \frac{n!}{(n-r)!} = n \times (n-1) \times ... \times (n-r+1).$$

In the discussion above, we have implicitly assumed that once an object is chosen from the set, it cannot be chosen again, since the $r$ chosen items are distinct. This is usually referred to as picking *without replacement*. Another way to chose or sample items from a set of $n$ items is to allow items to be picked repeatedly. This is referred to as picking *with replacement*.

> **Lemma 2.** *The number of ways to permute $r$ items with replacement from a set of $n$ distinct items is given by $n^r$.*

Permutations with replacement are rather straigthforward, since for every pick, you always have $n$ available choices to be made.

**Q:** What if the $n$ items are not distinct? How many ways can one order a set of $n$ items consisting of $k$ groups, where group $i$ has $r_i$ identical and mutually interchangeable items?
**A:** If all the items were distinct, we know that there are $n!$ possible orderings. Clearly, $n!$ is an overcount when some of the items are not distinct. Note that for each such ordering, if we just consider the items belonging to group 1 , they can be freely exchanged with each other and the new resulting ordering will look the same, since the objects are identical. Since we can permute the items in group 1 in $r_1!$ ways, we are overcounting the number of orderings by a factor of $r_1!$. Repeating the argument for every group of identical objects, we conclude that the number of orderings is given by

$$\binom{n}{r_1\ r_2\ \dots\ r_k} \triangleq \frac{n!}{r_1!r_2!\dots r_k!}.$$

## 2 COMBINATIONS

Now, we are interested the number of ways to choose $r$ items from a set of $n$ distinct items, *without* any regard for their relative ordering. We already know that the number of ways to choose these $r$ items while taking their relative order into account is ${}_nP_r$. Further, the total number of permutations for a given choice of $r$ items is simply $r!$. So, when we take ordering into acccount, we are counting $r!$ permutations for each unique set of $r$ items chosen from the original $n$. This implies the following result.

**Lemma 3.** *The number of ways to choose $r$ distinct items from a set of $n$ distinct items is given by*

$$ {}_nC_r \triangleq \frac{n!}{r!(n-r)!} = \frac{{}_nP_r}{r!}. $$

As with the discussion on permutations, we have implicitly assumed that once an object is chosen from the set, it cannot be chosen again, since the $r$ chosen items are distinct. The following result considers combinations *with replacement*.

**Lemma 4.** *The number of ways to choose $r$ items with replacement from a set of $n$ distinct items is given by $\binom{n+r-1}{r}$.*

The proof for this result is a bit more involved. It is essentially equivalent to Exercise 2.2 below, so we will discuss it once we solve that exercise.

## 3 EXERCISES

**Exercise 1 (Binomial Theorem).** Provide combinatorial arguments for the following.

1. Show that the coefficient of $x^r$ in $(1+x)^n$ is given by $\binom{n}{r}$.
2. Show the identity
$$ \sum_{r=0}^{n} \binom{n}{r} = 2^n. $$

**Solution:**

1. Write $(1+x)^n$ as the product of $n$ monomials $(1+x)(1+x)...(1+x)$. In the expansion, the term $x^r$ will appear if the $x$ term from $r$ of the $(1+x)$ is chosen and the 1 term from $n-r$ of the $(1+x)$ is chosen. The total number of such $x^r$ terms is equal to the number of ways to choose $r$ out of $n$ objects, which in turn is $\binom{n}{r}$.
2. One way to prove this result is to simply set $x = 1$ and use the result from part 1. A combinatorial proof, however, involves interpreting the

summation on the LHS as the total number of subsets that can be created from a set of $n$ objects. This is because we are simply counting the number of subsets containing exactly $r$ objects and then summing over all possible values of $r$. Given a subset and any particular element in the set, there are only two possibilities - it could either belong to the subset or not. Thus, an alternative way to calculate the total number of subsets is to simply multiply the possibilities for each element of the set. This leads to $2 \times 2 \times ... \times 2 = 2^n$. This completes the proof.

**Exercise 2 (Sums of Integers).**

1. Count the number of **positive** integer solutions to the following equation

$$x_1 + x_2.... + x_r = n.$$

In other words, find all possible solutions of the form $x_k \geq 1, x_k \in \mathbb{Z}, \forall k \in \{1, ..., r\}$ such that they sum up to $n$.

2. Now, count the number of **non-negative** integer solutions to the equation

$$x_1 + x_2.... + x_r = n.$$

In other words, find all possible solutions of the form $x_k \geq 0, x_k \in \mathbb{Z}, \forall k \in \{1, ..., r\}$ such that they sum up to $n$.

**Solution:**

1. To solve this problem, first write down $n$ 1s.

$$1 \quad 1 \quad ... \quad 1$$

Any valid solution $(x_1, ..., x_r)$ can be represented by adding $(r-1)$ markers in the sequence above

$$1 \quad 1 \Big\uparrow 1 \quad 1 \quad 1 \Big\uparrow 1....1 \Big\uparrow 1 \quad 1$$

The number of ones up to the first marker represent $x_1$, the number of ones between the first and second marker represent $x_2$ and so on. The number of ways in which $r-1$ markers can be inserted into $n-1$ spaces between the $n$ ones is given by $\binom{n-1}{r-1}$.

2. The solution to this problem is essentially the same as the previous part. Define $y_i = x_i + 1, \forall i$. Then, we are back to solving a problem of the form solved in part 1, however, the equaion we want to find solutions for becomes

$$y_1 + y_2 .... + y_r = n + r.$$

The number of solutions for this system with the constraint $y_i \geq 1, \forall i \geq 1$ is given by $\binom{n+r-1}{r-1}$.

Now, going back to combinations with replacement, let's represent the number of times item $i$ gets picked by the variable $x_i$. Then

$$x_1 + x_2 + ... + x_n = r$$

and $x_i \geq 0, \forall i$. Using the result above, the number of solutions to this is simply $\binom{n+r-1}{n-1}$. This is equal to $\binom{n+r-1}{r}$ since $\binom{m}{k} = \binom{m}{m-k}$.

**Exercise 3 (Texas Hold'em).** A *hand* in poker is a set of five distinct cards pulled from a standard deck of 52 cards.

1. A *straight flush* involves 5 consecutive cards of the same suit. Count the total number of straight flushes possible. Remember to subtract *royal flushes*, i.e. the case when the 5 consecutive cards are A, K, J, Q, 10.

2. A *three of a kind* involves 3 cards with the same rank and 2 cards with other different ranks. Count the total number of three of a kind hands possible. Remember to avoid counting *full houses* which are three of a kind and one pair.

**Solution:**

1. To have a straight flush the hand must consist of all five cards being of the same suit and in numerical order. There are 10 possible sequences: A – 5, 2 – 6, … , 9 – K, and 10 – A. However, we remove the last one, since it is a royal flush. Since there are 4 suits, then the number of straight flushes possible is just 9 * 4 = 36, plus the highest four (each a straight flush 10 – A of one of the four suits) being royal flushes.

2. There are 13 ranks in total. Given a rank, there are 4 cards, out of which we need to choose 3. Discarding the cards with this rank, we are left with 48 cards. The first card can be chosen in 48 ways, the second can be chosen in 44 ways, since they must be of different ranks. Further, since ordering is irrelevant, we are overcounting by a factor of 2. This gives us the final count $\binom{13}{1}\binom{4}{3}\frac{48\times 44}{2} = 54912$.

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

16.09 Spring 2023
Recitation 2 02/14/2023

## 1 Review

**Experiment:** activity whose outcome is subject to uncertainty.
**Sample space:** denoted by $\mathcal{S}$, collection of all outcomes of an experiment.
**Events:** A collection of outcomes, i.e. a subset of $\mathcal{S}$.
**Mutually Exclusive Events:** Two events $A$ and $B$ are *disjoint or mutually exclusive* if $A \cap B = \phi$.

**Axioms:** Probabilities map events to numbers in the set $[0, 1]$. More formally, they must satisfy the following axioms.

1. $\mathbb{P}(A) \geq 0$, for any event $A$.
2. $\mathbb{P}(\mathcal{S}) = 1$.
3. $\mathbb{P}(A_1 \cup A_2 \cup ....) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ for any collection of disjoint (mutually exclusive) events $A_1, A_2, ....$

**Conditional Probability**: The conditional probability that an event $A$ occured, given that another event $B$ occured is defined as

$$\mathbb{P}(A|B) \triangleq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

assuming $\mathbb{P}(B) > 0$.
**Independence**: Two events $A$ and $B$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Further, if $\mathbb{P}(B) > 0$, then we also get that $\mathbb{P}(A|B) = \mathbb{P}(A)$.

For a collection of events $A_1, ..., A_n$ to be independent, the following must hold for every subset of events $\mathbb{P}(A_{i_1}, A_{i_2}, ..., A_{i_k}) = \mathbb{P}(A_{i_1}) \times ... \times \mathbb{P}(A_{i_k})$.

## 2 Exercises

**Exercise 1 (On Sets).** Consider a sample space $\mathcal{S}$ with three events $A, B$ and $C$.

1. Write a set theoretic expression for the event that contains all outcomes only in $A$ but not in $B$ or $C$, i.e. $A$ occurs but $B$ and $C$ do not occur.

2. Suppose $B \subseteq A$. Using *only* the axioms of probability, show that $\mathbb{P}(B) \leq \mathbb{P}(A)$.

3. Is it possible that $A \cap B \neq \phi$, $B \cap C \neq \phi$ and $C \cap A \neq \phi$ and yet $A \cap B \cap C = \phi$? Can you give an example that involves picking a card from a deck?

4. If $A, B, C$ is a collection of independent events then $A$ is independent of $B \cap C$.

5. If $A$ and $B$ are independent, show that $A^c$ and $B^c$ are independent as well.

**Solution:**

1. The set $A \cap (B \cup C)^c$ or alternatively, $A \cap (B^c \cap C^c)$.

2. Since $B \subseteq A$, we can write $A$ as the union of two disjoint sets $A = B \cup (A \cap B^c)$. Using the third axiom, we then know that $\mathbb{P}(A) = \mathbb{P}(B) + \mathbb{P}(A \cap B^c)$. However, $\mathbb{P}(A \cap B^c) \geq 0$ from the first axiom. Thus, $\mathbb{P}(A) \geq \mathbb{P}(B)$.

3. Yes, consider the experiment to be that you pick one card at random out of a standard 52 card deck. Define the event $A$ to be that the card is red, define the event $B$ to be that the card is a number, and defined $C$ to be the event that the card is either a black 4 or a red King. Clearly, $A, B$ and $C$ satisfy the required conditions.

4. For three events to be independent together, we need $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, $\mathbb{P}(A \cap C) = \mathbb{P}(C)\mathbb{P}(A)$, and $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

   Using these, we can write $\mathbb{P}(A \cap (B \cap C)) = \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \mathbb{P}(A)\mathbb{P}(B \cap C)$. This proves the required statement.

5. Since $A$ and $B$ are independent, we know that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Using this, we will prove a similar statement for the complements.

$$
\begin{aligned}
\mathbb{P}(A^c \cap B^c) &= \mathbb{P}((A \cup B)^c) \\
&= 1 - \mathbb{P}(A \cup B) \\
&= 1 - \Big(\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\Big) \\
&= 1 - \Big(\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)\Big) \\
&= (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) = \mathbb{P}(A^c)\mathbb{P}(B^c).
\end{aligned}
$$

**Exercise 2 (Counting Strikes Again).**

1. Consider the $n$ integers $1, 2, ..., n$. Alice picks $k$ integers from this set, one after the other and uniformly at random without replacement. What is the probability that the sequence of integers picked by Alice is in increasing order?

2. We toss a biased coin $n$ times. The probability of it turning up heads is $p$, and each coin toss is independent of other coin tosses.

   (a) What's the probability that there were k heads in total?

   (b) Suppose you were told that $k$ tosses ended up heads. Find the coin bias $p^* \in [0, 1]$ that maximizes the probability that there were $k$ heads out of $n$ tosses.

   This estimate of $p$ is more formally known as the maximum likelihood estimate.

**Solution:**

1. Note that Alice picks each $k$ sized permutation from the set $\{1, 2, ..., n\}$ with equal probability. The total number of these permutations is ${}_nP_k$. We are only interested in permutations that are in increasing order. For every $k$ sized subset , there is only one permutation that is in creasing order. There are a total of $\binom{n}{k}$ subsets of size $k$. Thus, there are $\binom{n}{k}$ permutations that are in increasing order. The probability of interest is then simply $\frac{\binom{n}{k}}{{}_nP_k} = \frac{1}{k!}$.

   An alternative way to come to the same conclusion is to realize that we are only interested in picking one out of the $k!$ permutations for any choice of $k$ numbers, so the probability that we pick the numbers in increasing order is simply $\frac{1}{k!}$.

2. (a) To get $k$ heads in total, we need to choose which $k$ coin tosses turn up heads. We can do this in $\binom{n}{k}$ ways. Once we have chosen the specific tosses that are heads, the probability of this particular sequence occurring is simply $p^k(1-p)^{n-k}$. Summing across all possible choices of $k$ sized subsets of $1, ..., n$, we get the total probability to be $\binom{n}{k}p^k(1-p)^{n-k}$

   (b) We want to solve the following optimization -

$$p^* = \arg\max_{p\in[0,1]} \left(\binom{n}{k}p^k(1-p)^{n-k}\right)$$

.

Differentiating the objective with respect to $p$ and setting it to zero we get

$$
\begin{aligned}
k\binom{n}{k}p^{k-1}(1-p)^{n-k} - (n-k)\binom{n}{k}p^{k}(1-p)^{n-k-1} &= 0 \\
\Rightarrow k(1-p) = (n-k)p & \\
\Rightarrow p = \frac{k}{n}. &
\end{aligned}
$$

Thus, the maximum occurs at $p^* = \frac{k}{n}$, which, at least intuitively would also have been our best guess of $p$ if someone told us that a coin tossed $n$ times resulted in $k$ heads.

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

16.09 Spring 2023
Recitation 3 02/21/2023

## 1 Review

**Random Variables:** A random variable $X$ maps events (which are a collection of outcomes in the sample space) to real numbers. A *discrete* random variable takes at most a countably infinite number of values.

**Example 1:** Suppose $X$ is a random variable that takes a value $n$ with probability $2^{-n}$ for all positive integers $n$. While the range of this random variable is infinite, it is *countable*, so $X$ is a discrete random variable.

**Example 2:** Suppose you generate a random variable $X$ by picking real numbers uniformly at random between 0 and 1. Clearly, the range of $X$ is the entire interval $[0, 1]$ so it is not a discrete random variable.

**PMF:** The probability mass function or pmf of a discrete random variable is simply the probability that it takes a particular value, i.e. $\mathbb{P}(X = x)$. It is also denoted by $p(x)$ or $p_X(x)$. The pmf satisfies $\sum_x p(x) = 1$.

**Support:** The support of a discrete random variable is the set of values $x$ for which there is a positive probability mass, i.e. $p(x) > 0$.

**CDF:** The cumulative distribution function of a random variable $X$ is defined as the probability that it is less than or equal to a specified value $x$, i.e. $\mathbb{P}(X \leq x)$. It is denoted by $F_X(x)$.

**Expectation**: Suppose that a discrete random variable $X$ has support over the set $\mathcal{X}$. Then, the expectation of $X$ is defined as

$$\mathbb{E}[X] \triangleq \sum_{x \in \mathcal{X}} xp(x).$$

**Moments**: The $r$th moment of $X$ is defined as

$$\mathbb{E}[X^r] \triangleq \sum_{x \in \mathcal{X}} x^r p(x).$$

**Variance**: The variance of $X$ is defined as

$$var(X) \triangleq \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

**MGF**: The moment generating function (mgf) of $X$ is defined as

$$M_X(t) \triangleq \mathbb{E}\big[e^{tX}\big] = \sum_{x \in \mathcal{X}} e^{tx} p(x).$$

## 2 Exercises

**Exercise 1 (Properties of CDFs).** We consider a random variable $X$ with the pmf $p(x)$ distributed over the support $\mathcal{X}$.

1. Argue that $\lim_{x \to -\infty} F_X(x) = 0$, your proof does not need to be rigorous.
2. Argue that $\lim_{x \to \infty} F_X(x) = 1$, your proof does not need to be rigorous.
3. Show that $F_X(x)$ is non-decreasing in $x$.

**Solution:**

1. We are interested in the following limit

$$\lim_{x \to -\infty} F_X(x) = \lim_{x \to -\infty} \mathbb{P}(X \leq x).$$

   If we could push the limit from outside the probability sign $\mathbb{P}$ to inside, we would get

$$\lim_{x \to -\infty} \mathbb{P}(X \leq x) = \mathbb{P}\left( \lim_{x \to -\infty} X \leq x \right).$$

   Now, note that the set $\lim_{x \to -\infty} \{X \leq x\} = \{X \leq -\infty\}$ is the null set, since $X$ is a random variable that takes real values and so it must *always* be greater than negative infinity.

   Thus,

$$\lim_{x \to -\infty} \mathbb{P}(X \leq x) = \mathbb{P}\left( \lim_{x \to -\infty} X \leq x \right) = \mathbb{P}(\phi) = 0.$$

   The exchange of limits and probability can be justified by a property called continuity of probability. However, we leave the details of this to a more advanced class.

2. Along similar lines as part 1, we are interested in the following limit

$$\lim_{x \to \infty} F_X(x) = \lim_{x \to \infty} \mathbb{P}(X \leq x).$$

   If we could push the limit from outside the probability sign $\mathbb{P}$ to inside, we would get

$$\lim_{x \to \infty} \mathbb{P}(X \leq x) = \mathbb{P}\left( \lim_{x \to \infty} X \leq x \right).$$

Now, note that the set $\lim_{x\to\infty} \{X \leq x\} = \{X \leq \infty\}$ is the entire sample space $\mathcal{S}$, since $X$ is a random variable that takes real values and so it must *always* be less than infinity.

Thus,

$$\lim_{x\to\infty} \mathbb{P}(X \leq x) = \mathbb{P}\left(\lim_{x\to\infty} X \leq x\right) = \mathbb{P}(\mathcal{S}) = 1.$$

3. Consider two numbers $x_1$ and $x_2$ such that $x_1 \leq x_2$. Consider the following sets $A = \{X \leq x_1\}$ and $B = \{X \leq x_2\}$. Since $x_1 \leq x_2$, we know that if event $A$ occurs then event $B$ must also occur. This is because event $A$ indicates that $X$ is less than or equal to $x_1$ but $x_1$ is smaller than $x_2$ so $X$ must also be smaller than $x_2$, which is precisely event $B$. Thus, $A$ is a subset of $B$, or $A \subseteq B$. Now, we had shown in the previous recitation that if $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$. Thus, given any values $x_1$ and $x_2$ such that $x_1 \leq x_2$, we have shown that $\mathbb{P}(X \leq x_1) \leq \mathbb{P}(X \leq x_2)$, or alternatively $F_X(x_1) \leq F_X(x_2)$. This shows that $F_X(\cdot)$ must be a non-decreasing function.

**Exercise 2 (Computing Expectations).**

1. Consider a geometric random variable with parameter $p$. Show that its expectation is $1/p$.

2. Compute the expectation and variance of a Poisson random variable with parameter $\lambda$.

   *Note: We derive these results for completeness. In exams or homeworks, you can directly use the expectation/variance formulas for any distributions we discuss in class.*

**Solution:**

1. The pmf of a geometric random variable is given by

$$\mathbb{P}(X = k) = \begin{cases} (1-p)^{k-1}p, \forall n \in \mathbb{N}, \\ 0, \text{ otherwise.} \end{cases}$$

So, its expectation is given by

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} kp(1-p)^{k-1} = -p\sum_{k=1}^{\infty}\frac{d}{dp}(1-p)^k \\
&= -p\frac{d}{dp}\left(\sum_{k=1}^{\infty}(1-p)^k\right) = -p\frac{d}{dp}\left(\frac{1-p}{1-(1-p)}\right) \\
&= -p\frac{d}{dp}\left(\frac{1}{p}-1\right) = -p\frac{-1}{p^2} = \frac{1}{p}.
\end{aligned}
$$

We used the following identity above $\sum\limits_{k=1}^{\infty} r^k = \frac{r}{1-r}, \forall |r| < 1$.

2. The Poisson pmf is given by

$$
\mathbb{P}(X = k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!}, \forall n \in \{0, 1, 2, ...\}, \\ 0, \text{ otherwise.} \end{cases}
$$

So, its expectation is given by

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty}\frac{\lambda^k e^{-\lambda}}{(k-1)!} \\
&= \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!} \\
&= \lambda\sum_{j=0}^{\infty}\frac{\lambda^j e^{-\lambda}}{j!} = \lambda.
\end{aligned}
$$

The last inequality follows by substituting $j = k - 1$ in the summation and observing that we are summing over the Poisson pmf again, so the summation becomes $\sum_{j=0}^{\infty}\mathbb{P}(X = j) = 1$.

Now, for the variance, we first compute the second moment

$$
\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{(k-1)!} \\
&= \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \\
&= \lambda \sum_{j=0}^{\infty} (1+j) \frac{\lambda^j e^{-\lambda}}{j!} = \lambda \sum_{j=0}^{\infty} (1+j) \mathbb{P}(X=j) \\
&= \lambda \sum_{j=0}^{\infty} \mathbb{P}(X=j) + \lambda \sum_{j=0}^{\infty} j \mathbb{P}(X=j) = \lambda \mathbb{P}(\mathcal{S}) + \lambda \mathbb{E}[X] \\
&= \lambda + \lambda^2.
\end{aligned}
$$

Now, we can use the formula for variance in terms of the first and second moment $var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (\lambda + \lambda^2) - \lambda^2 = \lambda$.

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

16.09 Spring 2023
Recitation 4 02/28/2023

## 1 Review

**Continuous Random Variables:** A random variable $X$ is continuous if it takes values over intervals of the real line, and has a function associated with it called the *probability density function*, denoted by $f_X(\cdot)$ that satisfies

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx, \forall a, b.$$

The cdf of a continuous random variable can then be obtained as follows

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(x)dx.$$

From the above, it is also easy to see that the pdf is the derivative of the cdf, i.e.

$$\frac{d}{dx}F_X(x) = f_X(x).$$

This relationship holds at all values of $x$ where $F_X(\cdot)$ is differentiable.

**PDFs:** For any function $f$ to be a probability density function, it must satisfy the two conditions below:

1. $f(x) \geq 0, \forall x$
2. $\int\limits_{-\infty}^{\infty} f(x)dx = 1.$

**Support:** The support of a continuous random variable, denoted by $\mathcal{X}$, is the set of values for which there is a positive density, i.e. $f_X(x) > 0$.
**Expectation**: The expectation of a continuous random variable $X$ is defined as

$$\mathbb{E}[X] \triangleq \int_{x\in\mathcal{X}} xf_X(x)dx.$$

Higher moments, the variance and the moment generating function can be defined similarly using the definitions for discrete random variables, all we need to do is replace summations with integrals and the pmf with the pdf.
**Common distributions:** We list the densities for some common continuous random variables below

1. Uniform on the interval $[a, b]$, $X \sim Unif([a, b])$.

$$f_X(x) = \begin{cases} \frac{1}{b-a}, \text{ if } x \in [a, b] \\ 0, \text{ otherwise.} \end{cases}$$

2. Exponential with parameter $\lambda$, $X \sim Exp(\lambda)$.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, \text{ if } x > 0 \\ 0, \text{ otherwise.} \end{cases}$$

3. Normal with mean $\mu$ and variance $\sigma^2$, $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}.$$

## 2 Exercises

**Exercise 1 (Transforming Random Variables).** Consider a random variable $U$ distributed uniformly over the interval $[0, 1]$. Using this, we create a new random variable

$$Y = \frac{1}{\sqrt{1-U}}.$$

What is the pdf and the cdf of $Y$?

**Solution:** Note that the support of $Y$ is the set $[1, \infty)$ and the cdf of $U$ is given by

$$F_U(x) = \begin{cases} 0, \text{ if } x < 0 \\ x, \forall x \in [0, 1] \\ 1, \text{ if } x > 1. \end{cases}$$

We will directly compute the cdf of $Y$.

$$\begin{aligned} F_Y(x) = \mathbb{P}(Y \leq x) &= \mathbb{P}\left(\frac{1}{\sqrt{1-U}} \leq x\right) \\ &= \mathbb{P}\left(\frac{1}{x} \leq \sqrt{1-U}\right) \\ &= \mathbb{P}\left(\frac{1}{x^2} \leq 1-U\right) \\ &= \mathbb{P}\left(U \leq 1 - \frac{1}{x^2}\right) \\ &= 1 - \frac{1}{x^2}, \forall x \geq 1. \end{aligned}$$

Using this, we can now compute the density of $Y$ as well

$$f_Y(x) = \begin{cases} \frac{2}{x^3}, \forall x \geq 1. \\ 0, \text{ otherwise.} \end{cases}$$

**Exercise 2 (Bayes' Rule).** Suppose that the number of turnovers in a Celtics win is uniformly distributed over the set $\{7, ..., 15\}$ and the number of turnovers in a Celtics loss is uniformly distributed over $\{10, ..., 18\}$. The Celtics win probability this season is approximately $2/3$. What is the probability that they win a game given that they had 15 or more turnovers?

*The numbers in this problem are made up but the Celtics do have a turnover problem.*

**Solution:** We are interested in the probability $\mathbb{P}(\text{Win}|\text{15 or more turnovers})$. To compute this, we can use Bayes' rule

$$\mathbb{P}(\text{Win}|\text{15 or more turnovers}) = \frac{\mathbb{P}(\text{15 or more turnovers}|\text{Win})\mathbb{P}(\text{Win})}{\mathbb{P}(\text{15 or more turnovers})}.$$

To compute $\mathbb{P}(\text{15 or more turnovers})$ we can use the law of total probability.

$$\begin{aligned} \mathbb{P}(\text{15 or more turnovers}) &= \mathbb{P}(\text{15 or more turnovers}|\text{Win})\mathbb{P}(\text{Win}) \\ &+ \mathbb{P}(\text{15 or more turnovers}|\text{Loss})\mathbb{P}(\text{Loss}) \\ &= \frac{1}{9} \times \frac{2}{3} + \frac{4}{9} \times \frac{1}{3} \\ &= \frac{6}{27}. \end{aligned}$$

Putting these values back into the Bayes' rule expression above, we get

$$\mathbb{P}(\text{Win}|\text{15 or more turnovers}) = \frac{\frac{1}{9} \times \frac{2}{3}}{\frac{6}{27}} = \frac{1}{3}.$$

We see that the win probability drops from $2/3$ to $1/3$ if we know that the number of turnovers is 15 or more.

**Exercise 3 (Queuing and Memorylessness).** People arrive into a queue at a bank. The bank agent spends a random time addressing each person's request. These times are independent and exponentially distributed with a mean of 5 minutes.

Suppose you enter the queue, and observe 3 people waiting ahead of you. In addition, there is also one person currently talking to the agent at his desk. The people ahead of you tells you that this person has been talking to agent for at least 6 minutes. What is the expected time until you will reach the bank agent's desk? Assume that the queue is first-come-first-serve.

**Solution:** Suppose $X_1$ represents the time spent by person in front of you in the queue, $X_2$ for the person ahead of that and $X_3$ for the person ahead of that. Each of these is exponentially distributed. Let $Y$ denote the time remaining for the person who is currently at the agent's desk. Then we are interested in the quantity $\mathbb{E}[X_1 + X_2 + X_3 + Y]$. Using the memorylessness property, we know that $Y$ has the same distribution as if the person just arrived at the desk, i.e. exponential with mean 5 minutes. Each of $X_1, X_2$ and $X_3$ is distributed in the same way. So the total expected time until you reach the bank agent's desk is simply $4 \times 5 = 20$ minutes.

**Exercise 4 (More Memorylessness).** *This problem is optional, the analysis involved is not essential to understanding probability.*

Show that the only probability density that satisfies memorylessness for all positive real numbers is the exponential pdf. In other words, the only memoryless continuous random variables on $(0, \infty)$ are exponentially distributed.

*You can derive a similar result for discrete random variables - the only memoryless discrete random variables are geometric.*

**Solution:** The memorylessness property states that the following must hold

$$\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s), \forall s, t > 0.$$

Suppose that $X$ has a cdf represented by $F_X(x)$. Further, let $G(x) = \mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = 1 - F_X(x)$. Then, we can rewrite the memorylessness condition as follows

$$\begin{aligned}
&\frac{\mathbb{P}(X > s + t \cap X > t)}{\mathbb{P}(X > t)} = \mathbb{P}(X > s), \forall s, t > 0. \\
&\mathbb{P}(X > s + t) = \mathbb{P}(X > t)\mathbb{P}(X > s) \\
&G(s + t) = G(s)G(t).
\end{aligned}$$

Setting $s = t$ in the relationship above, we get $G(2t) = G^2(t), \forall t > 0$. Then, setting $s = 2t$, we get $G(3t) = G^3(t), \forall t > 0$. Repeating this procedure for any integer, we can get $G(mt) = G^m(t), \forall t > 0, m \in \mathbb{N}$. Setting $t' = tm$, we can also get $G(t') = G^m(t'/m), \forall t' > 0, m \in \mathbb{N}$. Alternatively, we get $G^{1/m}(t) = G(t/m), \forall t > 0, m \in \mathbb{N}$. Putting all of these together, for any rational number $m/n$, we get that $G(\frac{m}{n}t) = G^{m/n}(t)$.

Now, note that any real number $x$ can be written as the limit of a sequence of rational numbers, so we get that $G(xt) = G^x(t)$ must hold for all positive $x$ and $t$. Setting $t = 1$, we get $G(x) = G^x(1)$. Let $G(1) = e^{-\lambda}$ since it's just a constant smaller than 1. Then, we get $G(x) = e^{-\lambda x}$ or alternatively that $F_X(x) = 1 - e^{-\lambda x}$. Thus, $X$ must be exponentially distributed for memorylessness to hold.

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

16.09 | Spring 2023
Recitation 5 | 03/07/2023

## 1 Review

Given two random variables $X, Y$, their joint pdf is denoted by $f_{X,Y}(x, y)$ if they are continuous, their joint pmf is denoted by $p_{X,Y}(x, y)$ if they are discrete, and their joint cdf is denoted by $F_{X,Y}(x, y)$.

**Marginals:** To find marginal distributions from the joint, simply integrate or sum over the other variable. For example,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

$$p_X(x) = \sum_{x \in D} p_{X,Y}(x, y).$$

**Evaluating probabilities:** Given some set $A$ on $\mathbb{R}^2$, the probability that the pair of random variables $(X, Y)$ lies in this set is given by

$$\mathbb{P}((X, Y) \in A) = \int_{(x,y) \in A} f_{X,Y}(x, y)dxdy.$$

This gives us an expression for the joint cdf as well

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x, y)dydx.$$

**Independence:** Two random variables $X, Y$ are independent if their distributions factor as a product of their marginals, i.e. $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ or $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ or $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.

**Expectations:** To compute the expectation of a function of two random variables, use the following result

$$\mathbb{E}[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)f_{X,Y}(x, y)dxdy.$$

**Covariance and Correlation:** Given two random variables $X, Y$, their covariance is defined as

$$cov(X, Y) = \mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Further, their correlation coefficient is defined as

$$\rho_{XY} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}.$$

## 2 Exercises

**Exercise 1 (Tossing Coins).** Suppose that we toss a coin 10 times, where the coin ends up heads in each toss independently with probability $p$. Let $X$ be a random variable that denotes the number of heads in tosses 1-6. Let $Y$ denote the number of heads in tosses 4-10. What is the covariance and correlation between $X$ and $Y$?

**Solution:** Let $Z_i$ be a random variable that's 1 if the $i$th coin toss ends up heads. Then $X = Z_1 + ... + Z_6$ and $Y = Z_4 + ... + Z_{10}$.

First, we will compute the covariance

$$\begin{aligned}
cov(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[\sum_{i=1}^{6}\sum_{j=4}^{10} Z_i Z_j] - \mathbb{E}[\sum_{i=1}^{6} Z_i]\mathbb{E}[\sum_{j=4}^{10} Z_j] \\
&= \sum_{i=1}^{6}\sum_{j=4}^{10}\Big(\mathbb{E}[Z_i Z_j] - \mathbb{E}[Z_i]\mathbb{E}[Z_j]\Big) \\
&= \sum_{i=1}^{6}\sum_{j=4}^{10} cov(Z_i, Z_j) \\
&= \sum_{i=4}^{6} cov(Z_i, Z_i) \\
&= 3var(Z_4) = 3p(1-p).
\end{aligned}$$

**Exercise 2 (Sampling uniformly from 2D regions).** Suppose we sample points $(X,Y)$ uniformly from the unit circle centered on the origin. What is the joint distribution of $X, Y$?

*Bonus: Find the marginal densities of $X$ and $Y$, and use them to show that $X$ and $Y$ are not independent.*

**Solution:** We postulate that to sample points uniformly at random from the circle, the joint density of $(X,Y)$ should be some constant $c > 0$ over the entire region irrespective of the coordinates $(x, y)$, i.e.

$$f_{X,Y}(x,y) = \begin{cases} c, & \text{if } x^2 + y^2 \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Why do we call this density uniform over the unit disc? Suppose there is a subset within the unit disc $A_1$. Then the probability that we sample something from $A_1$ is given by

$$\mathbb{P}((X,Y) \in A_1) = \int_{(x,y)\in A_1} c\, dxdy = c\,(\text{Area of } A_1).$$

Thus, the probability only depends on the area of the subset, and not its shape or location. Further, two subsets of equal area have equal probability of being sampled from. Thus, this intuitively extends the notion of uniformity to regions in two dimensions.

How do we find the constant $c$? All we need to do is ensure that the integral of $f_{X,Y}$ over the disc equals 1.

$$\int_{x=-1}^{x=1} \int_{y=-\sqrt{1-x^2}}^{y=\sqrt{1-x^2}} c\, dxdy = c\,(\text{Area of unit circle}) = c\pi.$$

Setting this integral to 1, we get $c = \frac{1}{\pi}$.
To compute the marginal pdf of $X$, we need to integrate out $y$

$$\begin{aligned} f_X(x) &= \int_{y=-\sqrt{1-x^2}}^{y=\sqrt{1-x^2}} \frac{1}{\pi} dy \\ &= \frac{2\sqrt{1-x^2}}{\pi}, \text{ if } |x| \leq 1. \end{aligned}$$

By symmetry, we obtain that

$$f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi}, \text{ if } |y| \leq 1.$$

It is easy to verify that $f_X$ and $f_Y$ are indeed valid densities since they integrate to 1 over the interval $[-1, 1]$.

Note that the product of the marginals equals

$$f_X(x) f_Y(y) = \frac{4\sqrt{(1-x^2)(1-y^2)}}{\pi^2}, \forall |x| \leq 1, \forall |y| \leq 1.$$

Clearly, this does not equal the actual joint density

$$f_{X,Y}(x,y) = \frac{1}{\pi}, \text{ if } x^2 + y^2 \leq 1.$$

In fact, they don't even have the same support. So, $X$ and $Y$ are not independent.

**Exercise 3 (Bounding Correlation).** Consider two random variables $X$ and $Y$. Let their variances be $var(X) = \sigma_x^2$ and $var(Y) = \sigma_y^2$. Show that

$$var(X+Y) = var(X) + var(Y) + 2cov(X,Y).$$

Now, define two new random variables

$$Z_1 = \frac{X}{\sigma_x} + \frac{Y}{\sigma_y}$$
$$Z_2 = \frac{X}{\sigma_x} - \frac{Y}{\sigma_y}.$$

Show that $var(Z_1) = 2 + 2\rho_{XY}$ and $var(Z_2) = 2 - 2\rho_{XY}$. Using these, show that $|\rho_{XY}| \leq 1$.

**Solution:** First, we show the variance identity:

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

Now, we can apply the identity to $Z_1$. Using the facts that $var(aX) = a^2 var(X)$ and $cov(aX, bY) = abcov(X,Y)$, we get

$$\begin{aligned} var(Z_1) &= var(X/\sigma_x) + var(Y/\sigma_y) + 2cov(X/\sigma_x, Y/\sigma_y) \\ &= \sigma_x^2/\sigma_x^2 + \sigma_y^2/\sigma_y^2 + 2\frac{cov(X,Y)}{\sigma_x\sigma_y} \\ &= 2 + 2\rho_{XY}. \end{aligned}$$

Repeating the analysis for $Z_2$, we get

$$\begin{aligned} var(Z_2) &= var(X/\sigma_x) + var(-Y/\sigma_y) + 2cov(X/\sigma_x, -Y/\sigma_y) \\ &= \sigma_x^2/\sigma_x^2 + \sigma_y^2/\sigma_y^2 + 2\frac{cov(X,-Y)}{\sigma_x\sigma_y} \\ &= 2 - 2\rho_{XY}. \end{aligned}$$

Note that variances must always be non-negative. Since $var(Z_1) \geq 0$, we get that $-1 \leq \rho_{XY}$. Since $var(Z_2) \geq 0$, we get that $\rho_{XY} \leq 1$. Putting the two together, we get that $-1 \leq \rho_{XY} \leq 1$ or alternatively that $|\rho_{XY}| \leq 1$.

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

16.09 Spring 2023
Recitation 6 03/14/2023

## 1 Review

We look at linear combinations and sums of random variables. Let $X_1, X_2, ..., X_n$ be random variables with means $\mu_1, ..., \mu_n$ and variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$. Then the following identities hold

1. $$\mathbb{E}[a_1X_1 + ... + a_nX_n + b] = a_1\mu_1 + ... + a_n\mu_n + b$$

2. $$\begin{aligned} var(a_1X_1 + ... + a_nX_n) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jcov(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2\sigma_i^2 + 2\sum_{i}\sum_{j<i} a_ia_jcov(X_i, X_j). \end{aligned}$$

Next, we look at sums of independent random variables. Let $X_1, ..., X_n$ be independent random variables and let $Z = X_1 + ... + X_n$.

1. If we have just two continuous random variables and $Z = X_1 + X_2$, then the pdf of $Z$ is given by
$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_1}(x)f_{X_2}(z-x)dx.$$

2. The mgf of $Z$ satisfies
$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

3. If $X_i$ are normally distributed, then $Z$ is also normally distributed.

4. If $X_i$ are Poisson, then $Z$ is also Poisson.

## 2 Exercises

**Exercise 1 (Covariance of independent sums).** Suppose $X_1, ..., X_4$ are independent random variables. Let $Y = X_1 + 2X_2 + 3X_3$ and let $Z = X_1 - X_2 + X_3 + X_4$. Compute $cov(Y, Z)$ in terms of the variances of $X_1, ..., X_4$.
*Bonus: Do you notice a pattern? Can you formulate a general result on the covariance of linear combinations based on the analysis in this question?*
**Solution:** Recall the distributive property of covariance

$$cov(aX_1 + bX_2, Y) = a\, cov(X_1, Y) + b\, cov(X_2, Y).$$

We will use this property to compute covariance between $Y$ and $Z$.

$$\begin{aligned}
cov(Y, Z) &= cov(X_1, Z) + 2cov(X_2, Z) + 3cov(X_3, Z) \\
&= cov(X_1, X_1) + 2cov(X_2, -X_2) + 3cov(X_3, X_3) \\
&= var(X_1) - 2var(X_2) + 3var(X_3).
\end{aligned}$$

The second equality follows due to the fact that all terms of the form $cov(X_i, X_j)$ where $i \neq j$ equal zero, since $X_i$ are independent.

**Exercise 2 (Sampling from a circle, again!).** Suppose we sample points $(X, Y)$ uniformly from the unit circle centered on the origin. Find the marginal densities of $X$ and $Y$ and use them to show that $X$ and $Y$ are not independent.
**Solution:** Recall that to sample points uniformly at random from the circle, the joint density of $(X, Y)$ should be

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi}, & \text{if } x^2 + y^2 \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

To compute the marginal pdf of $X$, we need to integrate out $y$

$$\begin{aligned}
f_X(x) &= \int_{y=-\sqrt{1-x^2}}^{y=\sqrt{1-x^2}} \frac{1}{\pi} dy \\
&= \frac{2\sqrt{1-x^2}}{\pi}, \text{ if } |x| \leq 1.
\end{aligned}$$

By symmetry, we obtain that

$$f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi}, \text{ if } |y| \leq 1.$$

It is easy to verify that $f_X$ and $f_Y$ are indeed valid densities since they integrate to 1 over the interval $[-1, 1]$.

Note that the product of the marginals equals

$$f_X(x)f_Y(y) = \frac{4\sqrt{(1-x^2)(1-y^2)}}{\pi^2}, \forall |x| \leq 1, \forall |y| \leq 1.$$

Clearly, this does not equal the actual joint density

$$f_{X,Y}(x,y) = \frac{1}{\pi}, \text{ if } x^2 + y^2 \leq 1.$$

In fact, they don't even have the same support. So, $X$ and $Y$ are not independent.

**Exercise 3 (Markov Inequality).** Suppose $X$ is a continuous random variable that takes only positive values. Show that $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}, \forall a > 0$. We provide a step-by-step procedure to prove this result.

1. First, write down $\mathbb{E}[X]$ as an integral using the pdf of $X$, given by $f_X$.
2. Then, split the integral into the sum of two parts, an integral from $0$ to $a$ and an integral from $a$ to infinity.
3. Lower bound $\mathbb{E}[X]$ by ignoring the first term in the sum.
4. Further lower bound the remaining term by the term $a\mathbb{P}(X \geq a)$. If you can do that, then this completes the proof.

**Solution:** We will go through the recipe provided in the question.

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)dx \\
&= \int_{0}^{\infty} x f_X(x)dx \\
&= \int_{0}^{a} x f_X(x)dx + \int_{a}^{\infty} x f_X(x)dx \\
&\geq \int_{a}^{\infty} x f_X(x)dx \\
&\geq \int_{a}^{\infty} a f_X(x)dx = a\int_{a}^{\infty} f_X(x)dx = a\mathbb{P}(X \geq a).
\end{aligned}$$

We have shown that $\mathbb{E}[X] \geq a\mathbb{P}(X \geq a)$. Dividing by $a$, we get the required result.

## 3 Overall Review

Things you should be comfortable with for the quiz

1. Axioms of Probability
2. Basic set theory and counting (you don't need to worry about the more involved versions of counting we did in recitation 1)
3. Events, sample spaces, notions of mutual exclusion and independence of events
4. Probability identities - conditional probability, Bayes' rule, law of total probability, etc.
5. Discrete random variables - taking expectations/variances, computing probabilities, common distributions (we will provide standard distributions, expectations and variances in the exam)
6. Continuous random variables - taking expectations/variances, computing probabilities, common distributions (we will provide standard distributions, expectations and variances in the exam)
7. Manipulating pmfs, pdfs and cdfs, transforming functions of random variables
8. Basics of moment generating functions (you don't need to worry about evaluating very complicated integrals)
9. Joint distributions, covariances and correlations
10. Properties of sums of many random variables