MATH 323 Class Notes

Owen Lewis

Summer 2018

Contents

1 May 1, 2018

1.1 Definitions

Let Ω be the set of all possible outcomes. We call Ω the *Sample Space*.

Ex 1: Flipping a coin. The possible outcomes are H and T
$\Rightarrow \Omega = \{H, T\}$.
Ex 2: Tossing a die. We list all the outcomes as ω_i where i is the face of the die that we land on. We'll assume a normal 6-sided die
$\Rightarrow \Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$.
Ex 3: Flipping a coin until an H appears. The possible outcomes are H, TH, $TTH, \ldots, TT\ldots TH$ (with $n-1$ Ts), $\ldots$ ad infinitum.

$\Rightarrow \Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\}$.
Note that in this case, Ω is a (countably) infinite set!

Let Ω be a sample space. Any subset A of Ω is called an *Event*.

- If $A = \emptyset$, then we call A the *Null Event*
- If $A = \Omega$, then we call A the *Certain Event*
- If $|A| = 1$, then we call A an *Elementary Event*

Ex: Tossing a die. Let $\Omega := \{\omega_1, \omega_2, \ldots, \omega_6\}$, $A := \{\omega_1, \omega_2\}$. Then A is an event but is not an elementary event.

If A is an event, then A^c is also an event called the *complement event* of A.
If A, B are two disjoint events then we call A, and B *mutually exclusive*, or *disjoint*.

Let Ω be a sample space, $\mathcal{P}$ be the power set of Ω. A *Probability* $\mathbb{P}$ on Ω is a function $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$, such that:

1. $\forall\ A \subseteq \Omega$, $0 \leq \mathbb{P}(A) \leq 1$
2. $\mathbb{P}(\Omega) = 1$
3. If $A_1, A_2, \ldots, A_n, \ldots$ is a sequence of pairwise disjoint events then

$$\mathbb{P}(\bigcup_{i=1}^{\infty} p_i) = \sum_{i=0}^{\infty} \mathbb{P}(A_i)$$

1.2 How do we Apply this?

Let Ω be a discrete set, $E_i = \{\omega_i\}$ be an elemental event, with $E_i \subseteq \Omega$. A probability on Ω is given by a sequence $\mathbb{P}_1$, $\mathbb{P}_2$,..., $\mathbb{P}_n$,... of positive numbers such that

$$\mathbb{P}(E_i) = p_i, \ and \ \sum_i \mathbb{P}(p_i) = 1$$

If $A \subseteq \Omega$, then

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(p_i)$$

Ex 1: Toss a die.
a) Given that $\mathbb{P}(\omega_2) = \mathbb{P}(\omega_4) = \mathbb{P}(\omega_5) = \mathbb{P}(\omega_6) = \frac{1}{6}$, $\mathbb{P}(\omega_1) = \frac{1}{4}$, find $\mathbb{P}(\omega_3)$.
b) Find the probability that the die will land on an odd face.
Solution:
a) $\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$. $\mathbb{P}(\omega_3)$ is a singleton, and as we know $\sum_i \mathbb{P}(p_i) = 1$ then $\sum_{i=1}^{6} \mathbb{P}(\omega_i) = 1 \implies \mathbb{P}(\omega_2) + \mathbb{P}(\omega_4) + \mathbb{P}(\omega_5) + \mathbb{P}(\omega_6) + \mathbb{P}(\omega_1) + \mathbb{P}(\omega_3)$ $= 1 \implies \frac{4}{6} + \frac{1}{4} + \mathbb{P}(\omega_3) = 1 \implies \mathbb{P}(\omega_3) = 1$ - $\frac{11}{12}$
$\implies \mathbb{P}(\omega_3) = \frac{1}{12}$.
b) Let $A \subseteq \Omega$ be the subset of Ω containing all the odd faces. The total probablilty of A is therefore the sum of all the probabilities of the elements of $A \Rightarrow$ $\mathbb{P}(A_i) = \frac{1}{4} + \frac{1}{12} + \frac{1}{6} = \frac{1}{2}$.

Ex 2: Given a countably infinite sample space, find a constant c such that $\mathbb{P}(\{\omega_n\}) = c(\frac{1}{5})^n$ for some n.
Solution: $\sum_{n=1}^{\infty} \mathbb{P}(\{\omega_n\}) = 1 \Rightarrow \sum_{n=1}^{\infty} c(\frac{1}{5})^n = 1 \Rightarrow \sum_{n=1}^{\infty} (\frac{1}{5})^n = \frac{1}{c}$. Notice how $\sum_{n=1}^{\infty} (\frac{1}{5})^n$ is a geometric series that converges to $\frac{1}{4} \Rightarrow$ therefore $\frac{1}{c} = \frac{1}{4}$
$\Rightarrow c = 4$.

2 May 2, 2018

2.1 Properties of $\mathbb{P}$

Let Ω be a sample space and let $\mathbb{P}$ be a probability on Ω. Then:

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(A^c) = 1$ - $\mathbb{P}(A)$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ - $\mathbb{P}(A \cap B)$

Proof:

1. Set $A_1 = \Omega$, $A_2 = \emptyset$. Then $A_1 \cup A_2 = \Omega$, and $A_1 \cap A_2 = \emptyset$. Therefore $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) \Rightarrow \mathbb{P}(\Omega) = \mathbb{P}(\emptyset) + \mathbb{P}(\Omega) \Rightarrow 1 = \mathbb{P}(\emptyset) + 1$ $\Rightarrow 0 = \mathbb{P}(\emptyset)$. $\square$
2. $A^c \cup A = \Omega$, and $A^c \cap A = \emptyset$. Then $= \mathbb{P}(\Omega) \Rightarrow 1 = \mathbb{P}(A^c \cup A)$ $\Rightarrow 1 = \mathbb{P}(A) + \mathbb{P}(A^c) \Rightarrow \mathbb{P}(A^c) = 1$ - $\mathbb{P}(A)$. $\square$
3. It's easy to show that $A = (A \setminus B) \cup (A \cap B)$, and $\emptyset = (A \setminus B) \cap (A \cap B)$. Similarily, $\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)$. From these, it follows that $A \cup B = A \cup (B \setminus A) \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ $\Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ - $\mathbb{P}(A \cap B)$. $\square$

2.2 Equiprobability

Let Ω be a finite sample space. Set $N := |\Omega|$. Equiprobability means that all outcomes have the same probability $\mathbb{P} = \frac{1}{N}$. Let $A \subseteq \Omega$ be an event. Then we have $\mathbb{P}(A) = |A| \cdot \frac{1}{N}$, or

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

This is great because it means that in **equiprobability** problems we just need to count the cardinality of A, count the cardinality of Ω, and divide them, and we're done. Too bad counting isn't really all that easy.

2.3 Counting Tools

We have here three tools to help us calculate the cardinalities of huge (finite) subsets of huge sample spaces, each with their own specific situations that require its use

1. The Cartesian Product
2. Permutations
3. Combinations

2.3.1 The Cartesian Product

We all know what the cartesian product *is*. In probability we use it in our experiment when we have more than one """input""" each with its own possible outcome, for example, we roll three dice, or flip two coins.

Let A, B, be sets such that $|A| = a$, and $|B| = b$. Then the cardinality of the cartesian product $|(A \times B)|$ is $|(A \times B)| = a \cdot b$.

Ex 1: Suppose we roll a die twice. What is the cardinality of the sample space Ω?
Solution: If we roll a die once we have $\Omega = \{\omega_1, \ldots, \omega_6\}$. Therefore, the sample space for rolling a die twice is $\Omega \times \Omega$. The cardinality of our sample space $\Omega \times \Omega$ is $|\Omega \times \Omega| = 6 \cdot 6 = 36$.

Ex 2: Suppose we roll a fair die twice. What is the probability that the outcome is even?
Solution: If the sum of the two outcomes is even then both must either be even or both must be odd. Let A be the event where the sum of the rolls is even, and let A_1 be event where both individual rolls are even, A_2 be the event where both individual rolls are odd. For example,

$$A_1 = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$$

Where the 1^{st} element in each ordered pair is the outcome of the 1^{st} roll and the 2^{nd} element in each ordered pair is the outcome of the 2^{nd} roll. Therefore, we have that $|A_1| = |A_2| = 9$.
Since naturally $A = A_1 \cup A_2$ then $\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$. We also know that the die was fair, so we can use our equiprobability formula here.

$$\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = \frac{|A_1|}{|\Omega|} + \frac{|A_2|}{|\Omega|} = \frac{9}{36} + \frac{9}{36} = \frac{18}{36} = \frac{1}{2}.$$

2.3.2 Permutations

A permutation of r integer elements chosen from n (possible elements) is equivalent to a successive draw, without replacement, of r elements from a list of n elements. We denote the number of possibilities by P_r^n. The general formula for P_r^n is

$$P_r^n = n \cdot (n-1) \cdot \cdots \cdot (n-r+1) = \frac{n!}{(n-r)!}$$

Ex:
a) A thick black bag contains 4 balls: 1 green, 1 blue, 1 red, 1 yellow. The bag is made of lead, or something, and also light cannot exist in this bag. You couldn't see into this bag if your life depended on it. Draw successively two balls from the bag without putting them back in. What is the probability that

the second ball drawn is green?
b) What is the probability that one of the two balls drawn will be green?
Solution:
a) Let Ω be the set of permutations of 2 balls chosen from the bag containing 4 balls. Then, $|P_2^4| = \frac{4!}{2!} = \frac{24}{2} = 12$. Then let A be the event where the second ball is green. The cardinality of A is 3, as if we take for granted that the second ball is green, then there are 3 other different-coloured balls in the bag to accompany it. Since the bag is so dark that it's physically impossible to see inside it, all the balls in the bag have an equal probability of being drawn. Therefore,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{12} = \frac{1}{4}$$

b) Let B be the event where one of the balls drawn is green. Suppose B_1 is the event where the first ball drawn is green, and B_2 is the event where the second ball drawn is green. In part a) we found that $|B_2| = 3$, and a parallel argument shows that $|B_2| = 3$ as well. Thus,

$$\mathbb{P}(B) = \frac{|B_1|}{|\Omega|} + \frac{|B_2|}{|\Omega|} = \frac{3}{12} + \frac{3}{12} = \frac{6}{12} = \frac{1}{2}$$

2.3.3 Combinations

Consider a set Ω with n elements. Let r be an integer such that $0 \leq r \leq n$. A combination C_r^n, also denoted $\binom{n}{r}$ (pronounced n choose r), is the number of subsets of Ω containing r elements. The general formula is

$$C_r^n = \frac{P_n^r}{r!} = \frac{n!}{r! \cdot (n-r)!}$$

Ex:
a) Out of a deck of 52 cards how many distinct 5-card hands are possible?
b) What is the probability that a given hand contains at least one ace?
Solution:
a) $\binom{52}{5} = \frac{52!}{5! \cdot (47!)} = 2{,}598{,}960$
b) In this case it is easier to calculate the probability where the hand contains no aces and then subtract that from 1 to find the probability that we have an ace. If A is the event where the hand contains an ace, then A^c is the event where a hand contains no ace. Then

$$\mathbb{P}(A^c) = \frac{|A^c|}{|\Omega|} = \frac{C_5^{48}}{C_5^{52}} = \frac{1,712,304}{2,598,960} = \frac{35,673}{54,145}$$

Then we need to subtract this from 1 and we're done

$$\mathbb{P}(A) = 1 - \frac{35,673}{54,145} \approx 0.3412$$

2.4 Properties of Combinations

Here are some properties of combinations

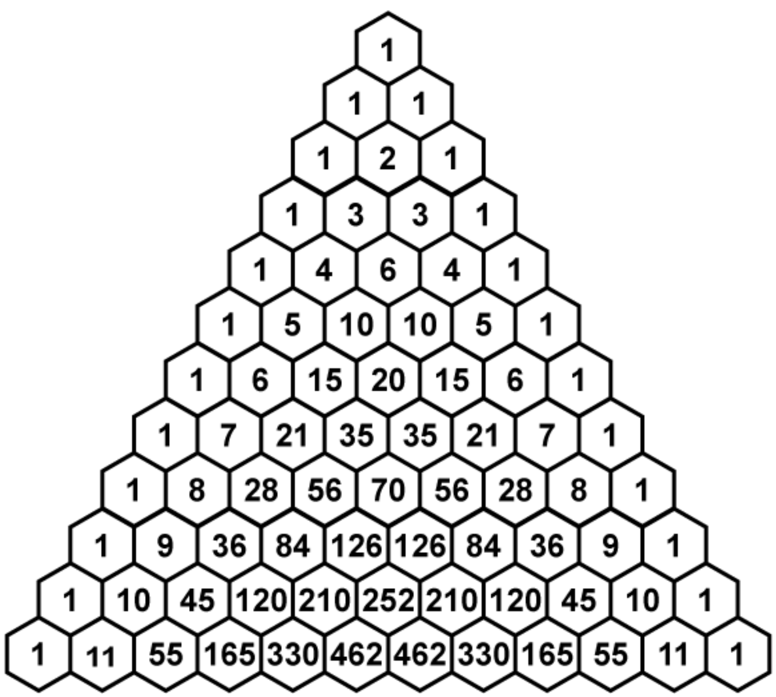
1. $C^n_0 = C^n_n = 1$
2. $C^n_r = C^n_{n-r}$ $0 \leq r \leq n$
3. $C^n_r = C^{n-1}_r + C^{n-1}_{r-1}$ $0 \leq r \leq n$

The proofs are really easy and I dont wan't to bother typing them out but here is the idea behind each of them:

1. Trivial
2. Induction on n
3. Also induction on n

2.4.1 Pascal's Triangle

Pascal's triangle is a table of combinations C^n_r. The rows of the triangle represent n, starting at 0 at the tip and working down, and the r^{th} element from the left (starting at 0) represents r. Each number in the triangle is determined by summing the two numbers directly above it.



Pascal's Triangle

3 May 3, 2018

3.1 Binomial Theorem

$(a+b)^n = \sum_{k=0}^{n} C_k^n \cdot a^k \cdot b^{n-k}$

Ex 1: $(a+b)^3 = C_0^3 \cdot a^6 b^3 + C_1^3 \cdot ab^2 + C_2^3 \cdot a^2 b + C_3^3 \cdot a^3 b = b^3 + 3ab^2 + 3ba^3 + a^3$

Ex 2: Find the coefficient of x^6 in the expansion of $(x^2+2)^7$.
Solution: If we let $x^2 := a$ and $2 := b$, then from the binomial theorem:

$$(a+b)^n = \sum_{k=0}^{n} C_k^n \cdot a^k \cdot b^{n-k} = \sum_{k=0}^{n} C_k^7 \cdot a^k \cdot b^{7-k} = C_3^7 \cdot 2^4 = 560$$

3.2 Conditional Probability

Let Ω be a sample space and let $\mathbb{P}$ be a probability on Ω. Let $A \subseteq \Omega$ be an event such that $\mathbb{P}(A) > 0$, and let $B \subseteq \Omega$ be another event. The *conditional probability of B given A*, is defined as

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

Ex 1: We have 2 urns. Urn 1 contains 7 red balls and 4 blue balls, and urn 2 contains 5 red balls and 6 blue balls. First, we choose a ball from urn 1 and place it into urn 2, then we remove a ball from urn 2. What is the probability that the ball that we remove from urn 2 will be blue?
Solution: Let B be the event where the ball drawn from urn 2 is blue, and let A_r, A_b be the events where we draw a red or a blue ball from urn 1 respectively. Then $\mathbb{P}(A_r) = \frac{7}{11}$ and $\mathbb{P}(A_b) = \frac{4}{11}$. Also,

$$B = B \cap \Omega \iff B = B \cap (A_r \cup A_b) \iff B = (B \cap A_r) \cup (B \cap A_b)$$

Which then implies

$$\mathbb{P}(B) = \mathbb{P}(B \cap A_r) + \mathbb{P}(B \cap A_b) \iff \mathbb{P}(B) = \mathbb{P}(B|A_r) \cdot \mathbb{P}(A_r) + \mathbb{P}(B|A_b) \cdot \mathbb{P}(A_b)$$

So all that is left is to compute $\mathbb{P}(B|A_r)$, and $\mathbb{P}(B|A_b)$.
From the definition, we can see that $\mathbb{P}(B|A_r) = \frac{1}{2}$, and $\mathbb{P}(B|A_b) = \frac{7}{12}$, so finally

$$\mathbb{P}(B) = \frac{6}{12} \cdot \frac{7}{11} + \frac{7}{12} \cdot \frac{4}{11} = \frac{70}{132}$$

Ex: 2 Roll a fair die. Let $A \subseteq \Omega$ be the event where the outcome is even, $B \subseteq \Omega$ be the event where the outcome is odd, $C \subseteq \Omega$ be the event where the outcome is either a 1 or a 2. Compute all the conditional probabilities.
Solution First $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{2}$, and $\mathbb{P}(C) = \frac{1}{3}$ Then:

$\mathbb{P}(A|B) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$, and $\mathbb{P}(B|A) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$
$\mathbb{P}(A|C) = \frac{1}{2}$, and $\mathbb{P}(C|A) = \frac{1}{3}$
$\mathbb{P}(B|C) = 1$, and $\mathbb{P}(C|B) = \frac{2}{3}$

3.3 Independent Events

Let A, and B be two events over some sample space Ω. A, and B are said to be *independent* if and only if:

1. $\mathbb{P}(A|B) = \mathbb{P}(A)$
2. $\mathbb{P}(B|A) = \mathbb{P}(B)$
3. $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Ex:
a) From the previous example, are A and B independent?
b) What about A and C?
c) B and C?
Answers:
a) No.
b) Yes.
c) No.

3.4 Baye's Rule

Let $A, B \subseteq \Omega$ be events on a sample space Ω. Then the equation

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

is calles Baye's Rule, or Baye's Theorem.

Theorem: Total Probability Rule
Let Ω be a sample space, and let $A \subseteq \Omega$ be an event. Suppose we partition Ω like $\Omega = \{B_1, B_2, \ldots, B_n\}$, where the B_is are pairwise disjoint events such that

$$\bigcup_{i=1}^{n} B_i = \Omega$$

Then:

1. $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$, and
2. If $k \in \mathbb{N}$ is fixed, then $\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k) \cdot \mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}$

Proof:

1. We know that $A = A \cap \Omega \iff A = A \cap (\bigcup_{i=1}^{n} B_i) \iff A = \bigcup_{i=1}^{n} A \cap B_i$. Therefore, $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) \iff \mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$ $\square$

2. $\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k) \cdot \mathbb{P}(B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k) \cdot \mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}$ $\square$

Ex: A (very simple) forecast model
Suppose that on a given day the weather is one of two states: sunny or rainy. Let $R \subseteq \Omega$ be the event where it's rainy and $S \subseteq \Omega$ be the event where it's sunny. If today is rainy then the probability that tomorrow will also be rainy is 60%. On the other hand, if today is sunny then the probability that tomorrow will be sunny is 70%.
a) If Monday is sunny then what is the probability that Wednesday will also be sunny?
b) If Wednesday is sunny then what is the probability that Tuesday was rainy?
Solutions:
a) Let A be the event where it is sunny on Wednesday. Then

$$\mathbb{P}(A) = \mathbb{P}(A|A_1) \cdot \mathbb{P}(A_1) + \mathbb{P}(A|A_2) \cdot \mathbb{P}(A_2) = 0.7 \cdot 0.7 + 0.4 \cdot 0.3 = 0.61$$

where A_1 is the event where it's sunny on Tuesday, and A_2 is the event where it's rainy on Tuesday.
b) $\mathbb{P}(A_2|A) = \frac{\mathbb{P}(A|A_2) \cdot \mathbb{P}(A_2)}{\mathbb{P}(A)} = \frac{0.12}{0.61} = \frac{12}{61}$

4 May 7, 2018

4.1 When to use Permutations vs Combinations

To determine whether we need to use a permutation or a combination we first need to consider two things. First, the sampling method:

1. Without Replacement
2. With Replacement

and whether we care about the order of the sample points:

3. Order doesn't matter
4. Order matters

We use permutations when we have 1&4, and
we use combinations when we have 1&3 *or* 2&4.

4.2 Multinomial Coefficients

The number of ways we can partition n objects into k distinct groups containing n_1, n_n, ..., n_k objects, respectively, where each object appears in exactly one group and $\sum_{i=1}^{k} n_i = n$, is

$$N := \binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}$$

where we call

$$\binom{n}{n_1, n_2, \ldots, n_k}$$

the *multinomial coefficient*.

Ex 1: We wish to expand $(x + y + z)^{17}$. What is the coefficient on the $x^2 y^5 z^{10}$ term?
Solution: The coefficient is $\binom{17}{2,5,10} = \frac{17!}{2!5!10!}$

Ex 2: #62 from the first assignment.
Solution: Total number $= \binom{9}{3,3,3}$, Desired number $= \binom{7}{1,3,3}$.
Therefore, $\mathbb{P}(A) = \frac{\binom{7}{1,3,3}}{\binom{9}{3,3,3}}$

4.3 Discrete Random Variables

Random Variables are variables that take on random values based on the outcome of the experiment. Each random variable is associated with a *probability distribution* that specifies the possible random variable values and the probability each value will occur. A random variable is said to be *discrete* if it can get only a finite or countably infinite number of possible distinct values.

The probability that the random variable Y will take on the value y, denoted by $\mathbb{P}(Y = y)$, is defined as the sum of the probabilities of all the sample points $\omega \in \Omega$ that are assigned the value of y.

4.4 Probability Distribution Representations

The probability distribution for a discrete random variable Y can be represented as a rule (formula), a table, or a graph.
Ex:

y	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(Y = y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

is a probability distribution of y.

Theorem: For any discrete probability distribution, the following hold:

1. $\mathbb{P}(Y = y) \geq 0, \forall y$
2. $\sum_y \mathbb{P}(Y = y) = 1$

4.5 Expected Value

Let Y be a discrete random variable with a probability function $\mathbb{P}(Y = y)$. Then the expected value of Y, $E(Y)$, is defined to be

$$E(Y) = \sum_y y \cdot \mathbb{P}(Y = y).$$

The expected value exists only if the above summation is *absolutely convergent*. We often denote $E(Y)$ by μ.

Ex: Consider a random variable Y with the following probability distribution.

y	0	1	2
$\mathbb{P}(Y = y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Then

$$\mu = \sum_y y \cdot \mathbb{P}(Y = y) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = \frac{1}{2} + \frac{1}{2} = 1$$

4.6 Expected Value of a Function of a Random Variable

Theorem: Let Y be the discrete random variable with probability function $\mathbb{P}(Y = y)$, and let $g(Y)$ be a real-valued function of Y. Then, the expected value of $g(Y)$ is

$$E(g(Y)) = \sum_y g(Y) \cdot \mathbb{P}(Y = y).$$

4.7 Varience of a Random Variable

For a random variable Y with mean $E(Y) \equiv \mu$, the *varience* of Y, $V(Y)$, is defined as the expected value of $(Y-\mu)^2$. That is

$$V(Y) = E((y-\mu)^2).$$

The varience represents the "average squared deviation of Y from its mean". Intuitively, it's a measure of how variable a random variable is. The bigger the value of the varience is, the more spread out the values that the random variable can take on are.

4.8 Standard Deviation of a Random Variable

The *standard deviation*, σ, of a random variable Y, is defined as the principle square root of $V(Y)$. That is

$$\sigma^2 = V(Y) \iff \sigma = |\sqrt{V(Y)}|$$

This is a little easier to visualize than varience.

4.9 Some Theorems (without proof)

<u>Theorem:</u> Let Y be a discrete random variable with the probability function $\mathbb{P}(Y=y)$, and let $c \in \mathbb{R}$ be a constant. Then

$$E(c) = c$$

<u>Theorem:</u> Let Y be a discrete random variable with the probability function $\mathbb{P}(Y=y)$, let $g(Y)$ be a function of Y, and let c be a constant. Then

$$E(c \cdot g(Y)) = c \cdot E(g(Y))$$

<u>Theorem:</u> Let Y be a discrete random variable with the probability function $\mathbb{P}(Y=y)$, and let $g_1(Y), g_2(Y), \ldots, g_k(Y)$ be k functions of Y. Then

$$E(g_1(Y) + g_2(Y) + \cdots + g_k(Y)) = E(g_1(Y)) + E(g_2(Y)) + \cdots + E(g_k(Y))$$

<u>Theorem:</u> Let Y be a discrete random variable with the probability function $\mathbb{P}(Y=y)$ and mean μ. Then

$$V(Y) \equiv \sigma^2 = E((Y-\mu)^2) = E(Y^2) - \mu^2$$

4.10 Binomial Probability Distribution

The *binomial probability distribution* is a probability distribution that comes up pretty commonly. It applies where

- There's a sequence of independent or identical trials

- Each trial can result in one of two outcomes (flipping a coin, for example).

More precisely, a binomial experiment is such that

- Consists of a fixed number of n trials
- Each trial results in one of two possible outcomes: S(uccess), or F(ailure).
- The probability of S on any trial is p, and the probability of F on any trial is $q = 1 - p$.
- All trials are independent
- The random variable Y is the number of successes out of n trials.

A random variable Y is said to have binomial distribution based on n trials iff

$$\mathbb{P}(Y = y) = \binom{n}{y} p^n q^{n-y}$$

5 May 8, 2018

5.1 Bernoulli Distribution

Let X be a random variable. Toss a coin. The outcomes are H, and T. Set $X(H) = 1$, and $X(T) = 0$. We have $X(\Omega) = \{0, 1\}$.
Assume $\mathbb{P}(H) = p$, for $0 < p < 1$. Then the probability function $\mathbb{P}(X = x)$ of X is

x	0	1
$\mathbb{P}(X = x)$	1-p	p

A random variable X that has the above probability function is called a *Bernoulli random variable* on X. Also, X is said to have the *Bernoulli Distribution* with parameter p. We write $X \sim \text{Ber}(p)$.

Proposition: If $X \sim \text{Ber}(p)$, then

1. $E(X) = p$
2. $V(X) = p(1 - p)$

Proof of (1):
$E(X) = \sum_x x \cdot \mathbb{P}(X = x) = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = p.$ □

Proof of (2):
$V(X) = E(X - E(X)^2) = E(X^2) - (E(X))^2.$
Also, $E(X^2) = \sum_x x^2 \cdot \mathbb{P}(X = x) = p.$
Therefore, $V(X) = p - p^2 = p(1 - p).$ □

5.2 Some Properties of E, and V

Let X be a random variable, and let $c, \alpha \in \mathbb{R}$ be constants.

1. $E(X + c) = E(X) + c$
2. $V(X + c) = V(X)$
3. $E(\alpha X) = \alpha E(X)$
4. $V(\alpha X) = \alpha^2 V(X)$

5.3 A Proof that $V(X) = E(X^2) - (E(X))^2$

Property: If f, g are functions, then $E(f(x) + g(x)) = E(f(x)) + E(g(x))$.

Set $\mu := E(X)$. Then, by definition,

$$\begin{aligned} V(X) &= E((X - E(X))^2) \\ &= E(x^2 - 2\mu x + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

□

5.4 More on Binomial Distribution

Proposition: If $X \sim \mathrm{Bin}(n, p)$, then

1. $E(X) = np$
2. $V(X) = np(1-p)$

To prove there we first have to note that

$$k \cdot \binom{n}{k} = k \cdot \frac{n!}{k!(n-k)!} = n \cdot \frac{(n-1)!}{(k-1)!(n-k)!} = n \cdot \binom{n-1}{k-1}$$

Proof of (1): By definition,

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k \cdot \mathbb{P}(X = k) \\ &= \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} n \cdot \binom{n-1}{k-1} \cdot p^k (1-p)^{n-k} \end{aligned}$$

Let $\ell = k - 1$

$$\begin{aligned} &= n \cdot \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} \cdot p^{\ell+1} (1-p)^{(n-1)-\ell} \\ &= np \cdot \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} \cdot p^{\ell} (1-p)^{(n-1)-\ell} \\ &= np(p + (1-p))^{n-1} \\ &= np \end{aligned}$$

$\square$

<u>Proof</u> of (2): By definition,

$$\begin{aligned}
V(X) &= E(X^2) - n^2p^2 \\
&= E(X(X-1) + X) - n^2p^2 \\
&= E(X(X-1)) + np - n^2p^2
\end{aligned}$$

Now we need to reduce $E(X(X-1))$ down.

$$\begin{aligned}
E(X(X-1)) &= \sum_{k=0}^{n} k \cdot (k-1) \cdot \binom{n}{k} \cdot p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} k \cdot (k-1) \cdot \binom{n}{k} \cdot p^k (1-p)^{n-k}
\end{aligned}$$

Now we need to change the variables in $k(k-1) \cdot \binom{n}{k}$.

$$k \cdot (k-1) \cdot \binom{n}{k} = n \cdot (n-1) \cdot \binom{n-2}{k-2}$$

so now plugging that back into $E(X)$ we get

$$\begin{aligned}
E(X(X-1)) &= \sum_{k=2}^{n} k \cdot (k-1) \cdot \binom{n}{k} \cdot p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} n \cdot (n-1) \cdot \binom{n-2}{k-2} \cdot p^k (1-p)^{n-k}
\end{aligned}$$

Let $\ell = k - 2$

$$\begin{aligned}
&= n \cdot (n-1) \cdot \sum_{\ell=0}^{n-2} \binom{n-2}{\ell} \cdot p^\ell (1-p)^{(n-2)-\ell} \\
&= n \cdot (n-1) \cdot p^2 \cdot (p + (1-p))^{n-2} \\
&= n \cdot (n-1) \cdot p^2
\end{aligned}$$

Plugging this back into $V(X)$, we get

$$\begin{aligned}
V(X) &= E(X(X-1)) + np - n^2p^2 \\
&= n \cdot (n-1) \cdot p^2 + np - n^2p^2 \\
&= np - np^2 \\
&= np(1-p)
\end{aligned}$$

$\square$

Ex 1: 3.60 from the textbook

Solution Fish die with $\mathbb{P} = 0.2$. Therefore the prob of a success (they survive), $\mathbb{P}(S) = 0.8$. Let the random variable X be the number of fish that survive. There are 20 fish. Thus $X \sim \text{Bin}(n = 20,\, p = 0.8)$. The probability that 14 fish survive is

$$\mathbb{P}(X = 14) = \binom{20}{14} \cdot (0.8)^{14} \cdot (0.2)^{6} \approx 0.1091 = 10.91\%$$

5.5 Geometric Distribution

FIrst recall that

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x},\ |x| < 1.$$

Suppose an experiment leads to either an $S(uccess)$ or an $F(ailure)$. Suppose we want to repeat the experiment until an S occurs
Let $\mathbb{P}(S) = p$, when $0 < p < 1$, and of course, $\mathbb{P}(F) = 1 - p$. Then

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$$

where

$$\omega_n = FFF \ldots FS$$

with $n - 1$ Fs.
Let the random variable X reperesent the number of trials needed for as S to occur. Then

$$\Omega(X) = \mathbb{N}$$

The probability function of X is then

$$\mathbb{P}(X = k) = p \cdot (1 - p)^{k-1}$$

If X satisfies this then we write $X \sim \text{Geometric}(p)$.

<u>Proposition:</u>

1. $E(X) = \frac{1}{p}$
2. $V(X) = \frac{1}{p} \cdot (\frac{1}{p} - 1)$

Proof of (1): By definition

$$\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) \\
&= \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} \\
&= p \cdot \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \\
&= p \cdot \frac{1}{(1-(1-p))^2} \\
&= \frac{p}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

$\square$

Proof of (2): By definition

$$\begin{aligned}
V(X) &= E(X^2) - (E(X))^2 \\
&= E(X^2) - \frac{1}{p^2} \\
&= E(X(X-1)) + E(X) - \frac{1}{p^2} \\
&= E(X(X-1)) + \frac{1}{p} - \frac{1}{p^2}
\end{aligned}$$

Now we need to reduce down $E(X(X-1))$

$$\begin{aligned}
E(X(X-1)) &= \sum_{k=1}^{\infty} k \cdot p \cdot (k-1) \cdot (1-p)^{k-1} \\
&= p \cdot (1-p) \cdot \sum_{k=2}^{\infty} k \cdot (k-1)(1-p)^{k-2} \\
&= p \cdot (1-p) \cdot \frac{2}{p^3} \\
&= \frac{2(1-p)}{p^2}
\end{aligned}$$

Plugging this back into $V(X)$, we get

$$\begin{aligned} V(X) &= E(X(X-1)) + \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{2-2p+1-1}{p^2} \\ &= \frac{1-p}{p^2} \\ &= \frac{1}{p} \cdot \left(\frac{1}{p} - 1\right) \end{aligned}$$

$\square$

6 May 9, 2018

6.1 Hypergeometric Distribution

Suppose we have a population of size N, and a subpopulation of size r. Each member of the subpopulation has a certain characteristic that the remaining $N - r$ members do not have. Take a sample from the population of size n. Then, let X be the random variable containing the number of elements from the subpopulation in the sample.

Ex: Suppose we draw a 5-card hand out of a deck of 52 cards . Let X be the number of aces in the hand. What are the possible values that X can take on?
Solution: We have $N = 52$, $n = 5$, and $r = 4$, and we want to find $X(\Omega)$.
First, $X \le n$, and $X \le r \Rightarrow X \le \min\{n, r\}$.
Also $n - X \le n$, and $n - X \le N - r \Rightarrow X \ge 0$, and $X \ge n + r - N$, and therefore $X \ge \max\{0, n + r - N\}$.
So, $X \in \mathbb{Z}$ such that

$$\begin{gathered} \max\{0, n + r - N\} \le X \le \min\{n, r\} \\ \Rightarrow \max\{0, 9 - 52\} \le X \le \min\{4, 5\} \\ \Rightarrow 0 \le X \le 4 \end{gathered}$$

And thus $X \in$ {0, 1, 2, 3, 4}, which makes sense.

<u>Theorem:</u> Suppose X is a hypergeometric distribution with parameters N, n, and r. Then

1. $E(X) = n \cdot \frac{r}{N}$
2. $V(X) = (n \cdot \frac{r}{N})(1 - \frac{r}{n})(\frac{N-n}{N-1})$

We will prove this later when we're on chapter 5.

6.1.1 The Probability Function of X

If $k \in \mathbb{N}_0$ such that $\max\{0, n + r - N\} \le k \le$ min{n, r}, then

$$p_X(k) = \mathbb{P}(X = k) = \frac{C_k^r \cdot C_{n-k}^{N-r}}{C_n^N}$$

A random variable X with the above probability function is calles a *hypergeometric distribution* with parameters N, n, and r, and we write

$$X \sim \text{Hypergeometric}(N,\ n,\ r).$$

Ex: Let X be the number of aces in a 5-card hand drawn from a 52-card deck. Then

$$
\begin{aligned}
p_X(0) &= \frac{C_0^4 \cdot C_5^{48}}{C_5^{52}} \\
p_X(1) &= \frac{C_1^4 \cdot C_4^{48}}{C_5^{52}} \\
p_X(2) &= \frac{C_2^4 \cdot C_3^{48}}{C_5^{52}} \\
&\vdots
\end{aligned}
$$

6.2 Poisson Distribution

Recall if $\lambda \in \mathbb{R}$ then

$$e^\lambda = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}$$

Let $\lambda \geq 0$ be fixed. The function

$$p(n) = e^{-\lambda} \cdot \frac{\lambda^n}{n!}$$

for $n \in \mathbb{N}$ is a probability function. This isn't immediately obvious and requires proof.
Proof:

$$
\begin{aligned}
\sum_{n=0}^{\infty} p(n) &= e^{-\lambda} \cdot \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\
&= e^{-\lambda} e^{\lambda} \\
&= 1
\end{aligned}
$$

$\square$

A random variable X with the above probability function is said to have *poisson distribution* with parameter λ. We write $X \sim \text{Poisson}(\lambda)$.

Proposition: Assume $X \sim \text{Poisson}(\lambda)$. Then

1. $E(X) = \lambda$
2. $V(X) = \lambda$

Proof of (1): By definition

$$\begin{aligned}
E(X) &= \sum_{n=0}^{\infty} n \cdot p_x(n) \\
&= \sum_{n=1}^{\infty} n \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!} \\
&= e^{-\lambda} \cdot \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\
&= \lambda \cdot e^{-\lambda} \cdot \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \\
&= \lambda \cdot e^{-\lambda} \cdot \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\
&= \lambda \cdot e^{-\lambda} \cdot e^{\lambda} \\
&= \lambda
\end{aligned}$$

$\square$

Proof of (2): By definition

$$\begin{aligned}
V(X) &= E(X^2) - (E(X))^2 \\
&= E(X(X-1)) + E(X) - (E(X))^2 \\
&= E(X(X-1)) + \lambda - \lambda^2 \\
&= \sum_{n=2}^{\infty} [n(n-1) \cdot p_X(n)] + \lambda - \lambda^2 \\
&= \lambda^2 e^{-\lambda} \cdot \sum_{n=2}^{\infty} \left[\frac{\lambda^{n-2}}{(n-2)!}\right] + \lambda - \lambda^2 \\
&= \lambda^2 e^{-\lambda} \cdot \sum_{n=0}^{\infty} \left[\frac{\lambda^n}{n!}\right] + \lambda - \lambda^2 \\
&= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}$$

$\square$

Ex 1: 3.122 from the book.
Solution: Let X be the number of customers at a store. 7 is the mean $\Rightarrow$ $X \sim \text{Poisson}(\lambda = 7)$.
a) $\mathbb{P}(X \leq 3) = \sum_{n=0}^{3} p_X(n) = \sum_{n=0}^{3} e^{-7} \cdot \frac{7^n}{n!}$
b) $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(x < 2) = 1 - \mathbb{P}(x \leq 1) \ldots$ plug into probability function.

Ex 2: Assume that $X_1 \sim \text{Poisson}(\lambda_1)$, $X_2 \sim \text{Poisson}(\lambda_2)$. Suppose X_1, X_2 are independent events, and $\{X_1 = n_1\}$, $\{X_2 = n_2\}$ are independent events for every n_1, n_2, $\in \mathbb{N}$. Prove that the distribution given by $X = X_1 + X_2$ is also poisson.
Solution: Proceeding by induction on n, we have
Base Case: $n = 1$

$$\begin{aligned}
\mathbb{P}(X = 1) &= \mathbb{P}((X_1 = 1 \wedge X_2 = 0) \vee (X_1 = 0 \wedge X_2 = 1)) \\
&= \mathbb{P}(X_1 = 1 \wedge X_2 = 0) + \mathbb{P}(X_1 = 0 \wedge X_2 = 1) \\
&= \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 0) + \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 1) \\
&= \lambda_1 e^{-\lambda_1} e^{-\lambda_2} + \lambda_2 e^{-\lambda_1} e^{-\lambda_2} \\
&= (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda 2)}
\end{aligned}$$

Q.E.D

Inductive Step: $n = k$

$$\begin{aligned}
\mathbb{P}(X = k) &= \sum_{i=0}^{k} \mathbb{P}(X_1 = i \;\wedge\; X_2 = k - i) \\
&= \sum_{i=0}^{k} \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = k - i) \\
&= e^{-(\lambda_1 + \lambda_2)} \cdot \sum_{i=0}^{k} \frac{\lambda_1^i}{i!} \cdot \frac{\lambda_2^{k-i}}{(k-i)!} \\
&= \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \cdot \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i} \\
&= \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \cdot \sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \\
&= \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} (\lambda_1 + \lambda_2)^k
\end{aligned}$$

Q.E.D

And thus by induction, the distribution of the random variable, formed by summing two random variables with poisson distribution, is itself poisson.

$\square$

6.3 Moment Generating Functions

First we will define *moments of a random variable.* Let X be a random variable. The n^{th} moment of X, μ_n, is defined as

$$\mu_n = E(X^n)$$

for $n \in \mathbb{N}$.

<u>Remark:</u> If X is a discrere random variable, then

$$\mu_n = E(X^n) = \sum_x x^n \cdot p_X(x)$$

Now we'll define *moment generating functions (MGFs).* Let X be a random variable. The moment generate function of X is defined as

$$m_X(t) = E(e^{tX})$$

<u>Remark:</u>

1. $t = 0$ is always in the domain of m_X
2. The moment generating function, m_X, of X is a unique identification of the distribution of X, not unlike Laplace transforms serve as a unique identification for differential equation solutions!

Ex 1: Consider the probability distribution given by

x	0	-1	3
$\mathbb{P}(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

find the moment generating function.
Solution: Straight from the definition of the moment generating function, we have

$$\begin{aligned} m_x(t) = E(e^{tX}) &= \sum_x x^n \cdot p_X(x) \\ &= e^{0t} \cdot p_X(0) + e^{-t} \cdot p_X(-1) + e^{3t} \cdot p_X(3) \\ &= \frac{1}{4}(1 + e^{3t}) + \frac{1}{2}e^{-t}, \ t \in \mathbb{R} \end{aligned}$$

Ex 2: Let X be a random variable. Consider the moment generating function of X, $m_X(t)$ given by

$$m_X(t) = \frac{1}{2}e^{t} + \frac{1}{6}e^{5t} + \frac{1}{3}e^{-6t}$$

write the probability distribution.

Solution:

x	1	5	6
$\mathbb{P}(X = x)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$

It's pretty easy to see the relationship between the equatoin and the table.

Remark:

$$\begin{aligned} m_X(t) &= E(e^{tX}) \\ &= E\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) \\ &= \sum_{n=0}^{\infty} E\left(\frac{(tX)^n}{n!}\right) \end{aligned}$$

(Of course one needs to be careful of doing the last step!)

$$m_X(t) = \sum_{n=0}^{\infty} \frac{E(X^x)}{n!} t^n$$

If that holds on an interval of the form $]-\varepsilon, \varepsilon[$, then

$$\frac{d^n}{dt^n}(m_X(t))\bigg|_{t=0} = E(X^n)$$

We can write this more formally as a theorem.

Theorem: Let $\varepsilon > 0$, and let X be a random variable such that the moment generating function contains an interval of the form $]-\varepsilon, \varepsilon[$. Then

$$E(X^n) = \frac{d^n}{dt^n}(m_X(t))\bigg|_{t=0}$$

We will not prove this right now. ..Maybe later.

6.4 The MGF of the Binomial Distribution

Proposition: Let X be a random variable and suppose that $X \sim \text{Bin}(n, p)$. Then

$$m_X(t) = (pe^t + (1-p))^n$$

for $t \in \mathbb{R}$.
Proof: By definition, we have

$$m_X(t) = \sum_{k=0}^{n} e^{tk} \cdot p_X(n)$$

Substituting in the probability function for the binomial distribution, we get

$$\begin{aligned} &= \sum_{k=0}^{n} \binom{n}{k} e^{tk} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= (pe^t + (1-p))^n, \ t \in \mathbb{R} \end{aligned}$$

□

Proposition: Let X be a random variable and suppose that $X \sim \text{Bin}(n, p)$. Then

$$E(X) = np$$

Proof:

$$E(X^n) = \frac{d^n}{dt^n}(m_X(t))\Big|_{t=0}$$

Setting $n = 1$

$$\begin{aligned} E(X) &= \frac{d}{dt}(m_X(t))\Big|_{t=0} \\ &= \frac{d}{dt}\left[(pe^t + (1-p))^n\right]\Big|_{t=0} \\ &= np(pe^t + (1-p))^{n-1} \cdot e^t\Big|_{t=0} \\ &= np \end{aligned}$$

□

Note that finding $V(X)$ is just a matter of finding $\frac{d^2}{dt^2}$.

7 May 10, 2018

7.1 The MGF of the Geometric Distribution

Proposition: Let X be a random variable and suppose that $X \sim \text{Geometric}(p)$. Then

$$m_X(t) = \frac{pe^t}{1-(1-p)e^t},\ t < \ln\left(\frac{1}{1-p}\right)$$

Proof:

$$\begin{aligned}
m_X(t) = E(e^{tX}) &= \sum_{n=1}^{\infty} e^{nt} p_X(n) \\
&= \sum_{n=1}^{\infty} e^{nt}(1-p)^{n-1}p \\
&= pe^t \cdot \sum_{n=1}^{\infty}((1-p)e^t)^{n-1} \\
&= pe^t \cdot \sum_{n=0}^{\infty}((1-p)e^t)^{n} \\
&= \frac{pe^t}{1-(1-p)e^t},\ t < \ln\left(\frac{1}{1-p}\right)
\end{aligned}$$

□

7.2 The MGF of the Poisson Distribution

Proposition: Let X be a random variable, let $\lambda > 0$, and suppose that $X \sim \text{Poisson}(\lambda)$. Then

$$m_X(t) = e^{\lambda(e^t-1)}$$

Proof:

$$\begin{aligned}
m_X(t) = E(e^{tX}) &= \sum_{n=0}^{\infty} e^{nt} p_X(n) \\
&= \sum_{n=0}^{\infty} e^{nt} e^{-\lambda} \cdot \frac{\lambda^n}{n!} \\
&= e^{-\lambda} \cdot \sum_{n-0}^{\infty} \frac{(\lambda e)^n}{n!} \\
&= e^{-\lambda} e^{\lambda e^t},\ \forall\ t \in \mathbb{R} \\
m_X(t) &= e^{\lambda(e^t-1)},\ t \in \mathbb{R}
\end{aligned}$$

□

Ex: Let X be the random variable containing the number of defects in a fabric. Suppose $X \sim \text{Poisson}(\lambda = 4)$. The company pays a cost C of $C := 3^X$ for each defect. Find the expected cost.
Solution: We want to find $E(3^X)$.

$$\begin{aligned} E(3^X) &= m_X(t = \ln(3)) \\ &= e^{4(e^{\ln 3} - 1)} \\ &= e^{4 \cdot 2} \\ &= e^8 \end{aligned}$$

7.3 Chebyshev's Inequality

Theorem: Let X be a random variable, let $E(X) := \mu$, and let $V(X) := \sigma^2$. Then

$$\mathbb{P}(|x - \mu| \;\geq\; k\sigma) \;\leq\; \frac{1}{k^2}$$

Lemma: (*) If X_1, X_2 are random variables with $X_1 \;\geq\; X_2$, then

$$E(X_1) \geq E(X_2)$$

Proof of (*): We will prove this in chapter 5.

Proof: Let Y be an arbitrary random variable such that $E(Y) = 0$, and such that $V(Y) = 0$. Let $k \in \mathbb{R}$ be fixed. We want to show that

$$\mathbb{P}(|Y| \;\geq\; k\sigma) \;\leq\; \frac{1}{k^2}$$

Let Z be a random variable such that:

$$Z := \begin{cases} 1, & \text{if } |Y| \;\geq\; k\sigma \\ 0, & \text{otherwise} \end{cases}$$

Therefore

$$\begin{aligned} & |Y| \;\geq\; (k\sigma) \cdot Z \\ \Rightarrow\; & |Y|^2 \;\geq\; (k^2\sigma^2) \cdot Z^2 \\ \Rightarrow\; & |Y|^2 \;\geq\; (k^2\sigma^2) \cdot Z \\ \Rightarrow\; & E(Y^2) \;\geq\; E((k^2\sigma^2) \cdot Z) \\ \Rightarrow\; & E(Y^2) \;\geq\; k^2\sigma^2 \cdot E(Z) \\ \Rightarrow\; & E(Y^2) \;\geq\; k^2\sigma^2 \cdot \mathbb{P}(|Y| \;\geq\; k\sigma) \\ \Rightarrow\; & \frac{E(Y^2)}{k^2\sigma^2} \;\geq\; \mathbb{P}(|Y| \;\geq\; k\sigma) \\ \Rightarrow\; & \frac{1}{k^2} \;\geq\; \mathbb{P}(|Y| \;\geq\; k\sigma) \end{aligned}$$

Since Y was arbitrary, we conclude that if we set $Y := X - \mu$, then we have

$$\mathbb{P}(|X - \mu| \ \geq \ k\sigma) \ \leq \frac{1}{k^2}$$

$\square$

7.4 The Cumulative Distribution Function

The *Cumulative Distribution Function*. Let X be a random variable. The cuumulative distribution function (cdf) is defined as

$$F_X(x) = \mathbb{P}(X \leq x)$$

Ex: Let X be a discrete random variable whose probability function is given by the table

x	-1	0	2
$\mathbb{P}(X = x)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

This gives us the cumulative distribution function $F_X(x)$

$$F_X(x) = \begin{cases} 0, & \text{if } x < \ -1 \\ \frac{1}{2}, & \text{if } -1 \ \leq \ x < 0 \\ \frac{5}{6}, & \text{if } 0 \ \leq \ x < 2 \\ \frac{1}{2}, & \text{if } x \ \geq \ 2 \end{cases}$$

Note that in this case $F_X(x)$ is a right-continuous nondecreasing step-function, such that $\lim_{x \to -\infty} F_X(x) = 0$, and $\lim_{x \to \infty} F_X(x) = 1$.

7.5 Continuous Random Variables

A random variable X is said to be *continuous* if its cumulative distribution function $F_X(x)$ is a continuously nondecreasing function such that $\lim_{x \to -\infty} F_X(x) = 0$, and $\lim_{x \to \infty} F_X(x) = 1$.

7.6 Computing Probability with the Cumulative Distribution Function

$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X{<}a)$.
Note that $\mathbb{P}(X = a) = 0 \Rightarrow \mathbb{P}(X{<}a) = \mathbb{P}(X \leq a) = F_X(a)$, where F_X is the cumulative distribution function of a random variable X, so

$$\begin{aligned} \mathbb{P}(a \leq X \leq b) &= F_X(b) - F_X(a) \\ &= \mathbb{P}(a \leq X{<}b) \\ &= \mathbb{P}(a{<}X \leq b) \\ &= \mathbb{P}(a{<}X{<}b) \end{aligned}$$

Note that $\mathbb{P}(X \geq a) = \mathbb{P}(X > a) = 1 - F_X(a)$

Ex 1: Let X be a random variable, and let

$$F_X(x) = \frac{2}{\pi}\Big(\tan^{-1}(x) + \frac{\pi}{2}\Big)$$

In this case, $F_X(x)$ is not a cumulative distribution function because $\lim x \to \infty = 2$. However

$$F_X(x) = \frac{1}{\pi}\Big(\tan^{-1}(x) + \frac{\pi}{2}\Big)$$

indeed is a cumulative distribution function.

In the above example, the random variable X for which F_X is a cumulative distribution function, is said to have the *Cauchy distribution*.

Ex 2: Let X be a random variable with the Cauchy distribution.
Find $\mathbb{P}(X \geq 1)$, and $\mathbb{P}(-1 \leq X \leq 1)$.
Solution: To find $\mathbb{P}(X \geq 1)$ we do

$$\mathbb{P}(X \geq 1) = 1 - F_X(1) = 1 - \frac{3}{4} = \frac{1}{4}$$

and to find $\mathbb{P}(-1 \leq X \leq 1)$ we do

$$\mathbb{P}(-1 \leq X \leq 1) = F_X(1) - F_X(-1) = \frac{1}{2}$$

Theorem: If F is a cumulative distribution function, then F is differentiable almost everywhere.
Proof: Take honours analysis 4 lol.

7.7 Probability Density Function

Let F be a cumulative distribution function, let X be a random variable, and let $f : \mathbb{R} \to \mathbb{R}$ be a function such that $f(x) = F'(x)$. Then if, $f(x) \geq 0$, and $\int_{-\infty}^{\infty} f(x)\ dx = 1$, then f is called a *probability density function* of X. The cumulative distribution function of X is given by $F_X(x) = \int_{-\infty}^{x} f(t)\ dt$.

Ex 1: The Cauchy distribution

$$F_X(x) = \frac{1}{\pi}\Big(\tan^{-1}(x) + \frac{\pi}{2}\Big)$$

has a probability density function given by

$$f(x) = F_x'(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2},\ \forall\ x \in \mathbb{R}$$

Ex 2: Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} cx^2 e^{-2x}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find c such that f is a probability density function.
Solution: If f is a probability density function, then c is such that

$$c \int_0^\infty x^2 e^{-2x} \ dx = 1$$

which implies that

$$c = \frac{1}{\int_0^\infty x^2 e^{-2x} \ dx}$$

8 May 14, 2018

8.1 More on CDFs and PDFs

Suppose X is a continuous random variable. The distribution of X is defined by either its cumulative distribution function F_X or its probability density function f_X, where of course the relationship between the two is

$$F'_X = f_X$$

We define the *support of the distribution* to be the set

$$\{x \in \mathbb{R} : f_X(x) > 0\}$$

Ex: Consider the CDF

$$f_X(x) = \begin{cases} 1, & \text{if } x \in]0,1[\\ 0, & \text{otherwise} \end{cases}$$

Then $f_X(x) \geq 0$, the support of the distribution is the interval $]0, 1[$, and

$$\int_{-\infty}^{\infty} f_X(x)\ dx = \int_0^1 f_X(x)\ dx = 1$$

Proposition: Let X be a continuous random variable and f_X be its PDF. Set $Y := aX + b$, $a \neq 0$. Then, the PDF of Y is

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Proof: Let F_X be the CDF of X. The CDF of Y is then

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(aX + b \leq y) = \mathbb{P}\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

The PDY of Y is thus

$$f_Y(y) = F'_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

$\square$

8.2 Expected Value of a Continuous Random Variable

Let X be a continuous random variable and $f_X(x)$ be its PDF. If $g : \mathbb{R} \to \mathbb{R}$ is a function, then

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f_X(x)\ dx$$

if and only if the integral converges. In particular, as in the discrete case

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\ dx$$

8.3 Variance of a Continuous Random Variable

Let $\mu := E(X)$. Much like with discrete random varibales, we define the variance of a function, $V(X)$, as

$$V(X) = E((X - \mu)^2)$$

in particular

$$V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)\ dx$$

8.4 Properties of $E(X)$ and $V(X)$

The proofs to all these properties are analagous to the proofs to the same properties in the discrete case. If X is a continuous random variable then

1. $E(aX + b) = aE(X) + b$
2. If g, and h are functions then
   - $E(g(x) + h(x)) = E(g(x)) + E(h(x))$, and
   - $E(\alpha g(x)) = \alpha E(g(x))$
3. $V(X) = E(X^2) - (E(X))^2$
4. $V(aX + b) = a^2 V(X)$

Ex 1: Consider the following function

$$F_X(x) = \begin{cases} cx^2 & \text{if } x \in]0, 1[\\ 0 & \text{otherwise} \end{cases}$$

Find c so that f_X is a PDF and find $E(X)$, and $V(X)$.

Solution If f_X is a PDF then

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_X(x)\ dx \\ &= \int_{-\infty}^{\infty} cx^2\ dx \\ 1 &= \frac{c}{3} \\ c &= 3 \end{aligned}$$

To find $E(X)$ we do

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x)\ dx \\
&= \int_0^1 x3x^2\ dx \\
&= 3\int_0^1 x^3\ dx \\
&= \frac{3}{4}
\end{aligned}$$

To find $V(X)$ we first must find $E(X^2)$

$$\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)\ dx \\
&= \int_0^1 x^2 3x^2\ dx \\
&= 3\int_0^1 x^4\ dx \\
&= \frac{3}{5}
\end{aligned}$$

and so finally

$$\begin{aligned}
V(X) &= E(X^2) - (E(X))^2 \\
&= \frac{3}{5} - \frac{9}{16} \\
&= \frac{3}{80}
\end{aligned}$$

Ex 2: Find $E(X)$ of the Cauchy distribution.

Solution: The Cauchy distribution is given by

$$f_X(x) = \frac{1}{\pi} \cdot \frac{x}{1+x^2}$$

for $x \in \mathbb{R}$. Therefore E(X) is given by

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\pi} \cdot \frac{x}{1+x^2} \cdot dx$$

but that diverges, and thus $E(X)$ does not exist.

8.5 Uniform Distribution Over an Open Interval

Let $a, b \in \mathbb{R}$ such that $b > a$. A continuous random variable X is said to have a uniform distribution over the interval $]a, b[$ if the cumulative distribution function of X is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x \geq b \end{cases}$$

If we take the derivative of that we see the PDF of a uniformly distributed random variable is

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

If X has a uniform distribution then we write $X \sim \text{Uniform}(a, b)$.

Remark 1: If $a \subseteq a_1 \leq b_1 \subseteq b$ then $\mathbb{P}(a_1 \leq X \leq b_1) = F_X(b_1) - F_X(a_1)$

Remark 2: $F_X(b_1) - F_X(a_1) = \frac{b_1 - a_1}{b-a} = \frac{\text{length}(a_1, b_1)}{\text{length}(a, b)}$

Remark 3: If $I \subset \mathbb{R}$ is an interval then $\mathbb{P}(X \in I) = \frac{\text{length}(I \cap (a,b))}{b-a}$.

Ex: Suppose $X \sim \text{Uniform}(a, b)$. Find $\mathbb{P}(X \geq 6)$.
Solution: $\mathbb{P}(X \geq 6) = \frac{10-6}{10-2} = \frac{4}{8} = \frac{1}{2}$

Proposition: If $X \sim \text{Uniform}(a, b)$, then $X = (b-1)U + a$, where we define U as $U \sim \text{Uniform}(0,\ 1)$.
Proof: Suppose $X \sim \text{Uniform}(a, b)$. Then we set $U = \frac{1}{b-a}(x-a)$. We want to show that $U \sim \text{Uniform}(0, 1)$. Note that the PDF of U is given by

$$\begin{aligned} f_u(u) &= \frac{1}{\frac{1}{b-a}} f_X((b-a)u + a) \\ &= (b-a) f_X((b-a)u + a) \end{aligned}$$

Recall that

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

so applying this to U we get

$$\begin{aligned} f_U(u) &= \begin{cases} \frac{b-a}{b-a} & \text{if } a < (b-a)u + a < b \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 & \text{if } a < u < b \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

And thus $U \sim$ Uniform(0, 1). $\square$

Proposition: If a continiuous random variable $X \sim \text{Uniform}(a, b)$, then

1. $E(X) = \frac{a+b}{2}$
2. $V(X) = \frac{(b-a)^2}{12}$

Proof: The proof follows directly from the definitions. I don't wanna type them.

8.6 Normal Distribution

A continuous random variable X is said to that a *normal distribution* with parameters μ, σ^2 if the PDF of X is given by

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

for $x \in \mathbb{R}$. We write $X \sim \text{N}(\mu, \sigma^2)$

If a continuous random variable Z is given by $Z \sim \text{N}(\mu = 0,\ \sigma^2 = 1)$ we say that Z has the *standard normal distribution.* The PDF of Z is then given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

for $z \in \mathbb{R}$.

Proposition: If $X \sim \text{N}(\mu, \sigma^2)$ then $X = \sigma Z + \mu$ where $Z \sim \text{N}(\mu = 0,\ \sigma^2 = 1)$.
Proof: Like all proofs for this class, it's pretty straightforward.

Remark: If $X \sim \text{N}(\mu, \sigma^2)$ then

$$\begin{aligned}
F_X(x) &= \mathbb{P}(X \leq x) \\
&= \mathbb{P}(\sigma Z + \mu \leq x) \\
&= \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) \\
&= F_Z\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}$$

The above implies that if we know F_Z then we can easily figure out F_X for all $X \sim \text{N}(\mu, \sigma^2)$! Since F_Z is pretty special we give it its own symbol. We denote F_Z by Φ where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\ dt$$

Properties of Φ

1. $\Phi(0) = \frac{1}{2}$
2. Φ(x) + Φ(-x)= 1

Proposition: If $X \sim$ N(μ, σ^2)

1. $E(X) = \mu$
2. $V(X) = \sigma^2$

Proof: Recall that $X = \sigma Z + \mu$, where $Z \sim$ N$(\mu = 0,\ \sigma^2 = 1)$. The remainder of the proof follows immediately from the fact that $E(Z) = 0$, and $V(Z) = 1$. I'll show that $E(Z) = 0$ but the remainder is just integration by parts.

$$E(Z) = \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\ dz$$

Note that this is an odd function on a symmetric domain and so the integral evaluates to 0. That's an easy enough thing to prove but that's not the point of this class so I won't do it. $\square$

9 May 15, 2018

9.1 The Gamma Function

The Gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$$

Note that for our purposes, Γ is defined if and only if $\alpha > 0$ as

$$\int_0^\infty t^{\alpha-1}e^{-t}dt = \int_0^1 t^{\alpha-1}e^{-t}dt + \int_1^\infty t^{\alpha-1}e^{-t}dt$$

where

$$\int_0^1 t^{\alpha-1}e^{-t}dt$$

converges iff $\alpha > 0$ and

$$\int_1^\infty t^{\alpha-1}e^{-t}dt$$

converges $\forall \alpha \in \mathbb{R}$, and thus Γ converges for only positive α.

9.2 Properties of the Gamma function

Properties

1. $\Gamma(1) = 1$
2. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
3. If $n \in \mathbb{N}$ then $\Gamma(n) = (n-1)!$
4. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Proof of 1: Just evaluate the integral lol $\square$

Proof of 2:

$$\Gamma(\alpha + 1) = \int_0^\infty t^{\alpha}e^{-t}dt$$

Do one iteration of integration by parts

$$\begin{aligned} &= \alpha \int_0^\infty t^{\alpha-1}e^{-t}dt \\ &= \alpha\Gamma(\alpha) \end{aligned}$$

$\square$

Proof of 3: The proof of this is pretty easy and I don't want to bother with typing it up. It's just induction on n. $\square$

Proof of 4: We'll prove this in class later, but it probably has something to do with integrating in polar coordinates. $\square$

Ex: Find $\Gamma\left(\frac{7}{2}\right)$
Solution:

$$\begin{aligned}\Gamma\left(\frac{7}{2}\right) &= \frac{5}{2}\cdot\Gamma\left(\frac{5}{2}\right)\\ &= \frac{5}{2}\cdot\frac{3}{2}\cdot\Gamma\left(\frac{3}{2}\right)\\ &= \frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\cdot\Gamma\left(\frac{1}{2}\right)\\ &= \frac{15\sqrt{\pi}}{8}\end{aligned}$$

9.3 Some More General Applications of Γ

First Application:

$$\Gamma\left(\frac{2n+1}{2}\right) = \frac{\sqrt{\pi}}{n^n}\cdot\frac{(2n-1)!}{2^{n-1}(n-1)!}$$

Proof:

$$\begin{aligned}\Gamma\left(\frac{2n+1}{2}\right) &= \Gamma\left(\frac{1}{2}\right)\cdot\frac{1}{2}\cdot\frac{3}{2}\cdot\ldots\cdot\frac{2n-1}{2}\\ &= \frac{\sqrt{\pi}}{2^n}\cdot 1\cdot 3\cdot 5\cdot\ldots\cdot(2(n-1)-1)\\ &= \frac{\sqrt{\pi}}{2^n}\cdot\frac{(2n-1)!}{2\cdot 4\cdot\ldots\cdot 2(n-1)}\\ &= \frac{\sqrt{\pi}}{n^n}\cdot\frac{(2n-1)!}{2^{n-1}(n-1)!}\end{aligned}$$

$\square$

Second Application:

$$\int_0^\infty t^a e^{-bt}dt = \frac{\Gamma(a+1)}{b^{a+1}}$$

Ex: Solve the improper integral

$$\int_0^\infty t^{10} e^{-4t} dt$$

Solution: Using the above equation we have

$$\begin{aligned}
\int_0^\infty t^{10} e^{-4t} dt &= \frac{\Gamma(11)}{4^{11}} \\
&= \frac{\Gamma(10+1)}{4^{10+1}} \\
&= \frac{\Gamma(11)}{4^{11}} \\
&= \frac{10!}{4^{11}}
\end{aligned}$$

10 May 16, 2018

10.1 The Gamma Family of Distributions

10.1.1 The Gamma Distribution with One Parameter

Let X be a continiuous random variable, and consider the function

$$f_X(x) = \begin{cases} \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, & \text{if } x{>}0 \\ 0, & \text{if } x \leq 0 \end{cases}$$

if f_X is the PDF of X, then X has the one-parameter gamma distribution. We write $X \sim \text{Gamma}(\alpha)$.

Properties: If $X \sim \text{Gamma}(\alpha)$ then

1. $E(X) = \alpha$
2. $V(X) = \alpha$

Proof of 1: By definition,

$$\begin{aligned} E(X) &= \int_0^\infty x\frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}\ dx \\ &= \int_0^\infty \frac{x^{\alpha}e^{-x}}{\Gamma(\alpha)}\ dx \\ &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \\ &= \alpha \end{aligned}$$

$\square$

Proof of 2: First we need to calculate $E(X^2)$. This is done in evaluating the same kind of integral as in the proof of part 1 (but with x^2 instead of x).

$$E(X^2) = \int_0^\infty x^2\frac{x^{\alpha-1}e^{-1}}{\Gamma(\alpha)}\ dx = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} = \alpha(\alpha+1)$$

Then of course,

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= \alpha(\alpha+1) - \alpha^2 \\ &= \alpha \end{aligned}$$

$\square$

10.1.2 The Gamma Distribution with Two Parameters

Let X be a continuous random variable, and let $\alpha > 0$, and $\beta > 0$. Suppose that if $X \sim \text{Gamma}(\alpha, \beta)$, then the PDF of X is given by

$$f_X(x) = \begin{cases} \frac{1}{\beta^2 \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, \text{if } x > 0 \\ 0, \text{otherwise} \end{cases}$$

Remark: If $X \sim \text{Gamma}(\alpha, \beta)$, then $X = \beta Y$, where $Y \sim \text{Gamma}(\alpha)$.

Proposition: If $X \sim \text{Gamma}(\alpha, \beta)$ then

1. $E(X) = \alpha\beta$
2. $V(X) = \alpha\beta^2$

Proof: The proof of these properties immediately follows from the fact that $X = \beta Y$ where $Y \sim \text{Gamma}(\alpha)$.

10.1.3 The Exponential Distribution

Suppose $X \sim \text{Gamma}(\alpha = 1,\ \beta)$. Then we say that X has the exponential distribution and we write $X \sim \text{Exp}(\beta)$.

Proposition: If $X \sim \text{Exp}(\beta)$ then

1. $E(X) = \beta$
2. $V(X) = \beta^2$

Proof: The exponential distribution is a special case of the two-parameter gamma distribution, with $\alpha = 1$. The result thus follows from setting $\alpha := 1$ in the expected value and variance formulae.

The Memoryless Property If $X \sim \text{Exp}(\beta)$ then

$$\mathbb{P}(X > a + b \mid X > a) = \mathbb{P}(X > b)$$

Proof of the Memoryless Property: Recall the *survival function* $S(t)$ is defined as the compliment of the CDF. i.e.

$$S(t) = \mathbb{P}(x > t) = 1 - F_X(t) = e^{\frac{t}{\beta}}$$

for $t > 0$. Intuitively, it's the probability that something will survive past a point. Next,

$$\begin{aligned}
\mathbb{P}(X > a+b \mid X > a) &= \frac{\mathbb{P}(X > a+b \ \cap \ X > a)}{\mathbb{P}(X > a)} \\
&= \frac{\mathbb{P}(X > a+b)}{\mathbb{P}(X > a)} \\
&= \frac{S(a+b)}{S(a)} \\
&= \frac{e^{-\frac{a+b}{\beta}}}{e^{-\frac{a}{\beta}}} \\
&= e^{-\frac{b}{\beta}} \\
&= S(b) \\
&= \mathbb{P}(X > b)
\end{aligned}$$

$\square$

Theorem: Let X be a continuous random variable. If X is such that the memoryless property is satisfied, then X must have the exponential distribution.

Proof: The proof was left as an exercise "for math majors only" and I intend to work it out later. Who knows, maybe I'll even type it up here.

10.1.4 The Chi-Square Distribution

Suppose X is a continuous random variable. X is said to have the chi-square distribution with p degrees of freedom ($p \in \mathbb{N}$), iff $X \sim \text{Gamma}(\alpha = \frac{p}{2}, \beta = 2)$. We write $X \sim \chi^2(p)$.

To find the CDF and the PDF we let Z be a standard normal, and set $X := Z^2$. Then

$$\begin{aligned}
F_X(x) &= \mathbb{P}(X \leq x) \\
&= \mathbb{P}(Z^2 \leq x) \\
&= \begin{cases} 0, & \text{if } x < 0 \\ \mathbb{P}(-\sqrt{x} \leq Z \leq \sqrt{x}), & \text{otherwise} \end{cases} \\
&= \begin{cases} 0, & \text{if } x < 0 \\ \Phi(\sqrt{x}) - \Phi(-\sqrt{x}), & \text{otherwise} \end{cases} \\
F_X(x) &= \begin{cases} 0, & \text{if } x < 0 \\ 2\Phi(\sqrt{x}) - 1, & \text{otherwise} \end{cases}
\end{aligned}$$

is the CDF, and to find the PDF we just differentiate that.

$$F'_X(x) = f_X(x) = \begin{cases} 0, & \text{if } x \ lt \ 0 \\ \frac{x^{-\frac{1}{2}} e^{\frac{-x}{2}}}{\sqrt{2\pi}}, & \text{otherwise} \end{cases}$$

Preposition: If $X \sim \chi^2(p)$ then

1. $E(X) = 2 \cdot \frac{p}{2} = p$
2. $V(X) = 2^2 \cdot \frac{p}{2} = 2p$

10.1.5 A One-Line "Proof" that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$\sqrt{2\pi} = 2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right) \Rightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ □

10.2 The Beta Function

First we define the *beta function* as

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \, dx$$

for $\alpha > 0$, $\beta > 0$.

Proposition:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

10.3 The Beta Distribution

Suppose X is a continuous random variable. If the function

$$f_X(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

is the PDF of X, then X has the beta distribution with parameters α, β. We write $X \sim \text{Beta}(\alpha, \beta)$.

Proposition: If $X \sim \text{Beta}(\alpha, \beta)$ the

1. $E(X) = \frac{\alpha}{\alpha+\beta}$
2. $V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Ex 1: Let $X \sim \text{Beta}(\alpha = 1,\ \beta = 1)$. Then,

$$B(\alpha = 1,\ \beta = 1) = \frac{\Gamma(1)\Gamma(1)}{\Gamma(1)} = 1$$

and thus the PDF is given by

$$f_X(x) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Remark: $X \sim \text{Uniform}(0,\, 1) \equiv X \sim \text{Beta}(\alpha = 1,\ \beta = 1)$

Ex 2: Let $Y \sim \text{Beta}(\alpha = 2,\ \beta = 1)$. Then,

$$B(\alpha = 2,\ \beta = 1) = \frac{\Gamma(2)\Gamma(1)}{\Gamma(3)} = \frac{1}{2}$$

so the PDF is given by

$$f_X(x) = \begin{cases} 2y, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

11 May 17, 2018

11.1 Moment Generating Functions

If X is a random variable, and if f_X is its PDF, then the *moment generating function* of X, m_X, is defined by

$$m_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\ dx$$

Theorem: Let $\varepsilon > 0$. If the domain of $m_X(t)$ contains an interval of the form $]-\varepsilon, \varepsilon[$, then

$$\frac{d^n}{dt^n}(m_X(t))\Big|_{t=0} = E(X^n)$$

Property: If $Y = aX + b$, for $a > 0$, then $m_Y(t) = e^{tb} m_X(at)$
Proof: $m_Y(t) = E(e^{ty}) = E(e^{t(aX+b)}) = e^{tb}E(e^{atX}) = e^{tb}m_X(at)$ $\square$

11.2 The MGF of the Uniform Distribution

Proposition: Suppose $X \sim \text{Uniform}(a, b)$. Then, the moment generating function of X is given by

$$m_X(t) = \begin{cases} \frac{e^{bt}-e^{at}}{(b-a)t}, & \text{if } t \neq 0 \\ 1, & \text{otherwise} \end{cases}$$

Proof: Since $X \sim \text{Uniform}(a, b)$, $X = (b-a)U + a$, where $U \sim \text{Uniform}(0, 1)$. Therefore,

$$m_X(t) = e^{at} m_U((b-a)t)$$

Now we need to find the MGF of U.

$$\begin{aligned} m_U(s) &= E(e^{sU}) \\ &= \int_0^1 e^{su}\ du \\ &= \frac{1}{s} e^{su}\Big|_0^1, s \neq 0 \\ m_U(s) &= \begin{cases} \frac{e^s - 1}{s}, & \text{if } s \neq 0 \\ 1, & \text{if } s = 0 \end{cases} \\ m_U((b-a)t) &= \begin{cases} \frac{e^{bt}-e^{at}}{(b-a)t}, & \text{if } s \neq 0 \\ 1, & \text{if } s = 0 \end{cases} \end{aligned}$$

$\square$

11.3 The MGF of the Normal Distribution

Proposition: Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$. Then the MGF of X is given by

$$m_X(t) = e^{\frac{1}{2}\sigma^2 t^2 + \mu t}$$

Proof: If $X \sim \mathrm{N}(\mu, \sigma^2)$ then $X = \sigma Z + \mu$, where $Z \sim \mathrm{N}(0,\ 1)$. Therefore $m_X(t) = e^{\mu t} m_Z(\sigma t)$. To find m_Z we do

$$\begin{aligned}
m_Z(s) &= E(e^{sZ}) \\
&= \int_{-\infty}^{\infty} e^{sz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\ dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2zs)}\ dz \\
&\vdots
\end{aligned}$$

(We tackle the integral a bit then do a U-substituiton and end up here)

$$\begin{aligned}
&= e^{\frac{1}{2}s^2} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2}\ du \right) \\
m_Z(s) &= e^{\frac{1}{2}s^2}, s \in \mathbb{R} \\
m_Z(t) &= e^{\frac{1}{2}\sigma^2 t^2} e^{\mu t} \\
&= e^{\frac{1}{2}\sigma^2 t^2 + \mu t}
\end{aligned}$$

□

11.4 The MGF of the Gamma Distribution

Proposition: If $X \sim \mathrm{Gamma}(\alpha, \beta)$ then the MGF of X is given by

$$m_x(t) = \frac{1}{(1 - \beta t)^\alpha}$$

Proof: If $X \sim \text{Gamma}(\alpha, \beta)$ then $X = \beta Y$ where $Y \sim \text{Gamma}(\alpha)$. Therefore, $m_X(t) = mY(\beta t)$. We need to find m_Y.

$$\begin{aligned}
m_Y(s) &= E(e^{sY}) \\
&= \int_0^\infty e^{sy} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} \; dy \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-y(1-s)} y^{\alpha-1} \; dy \\
&= \frac{1}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{(1-s)^\alpha} \\
m_Y(s) &= \frac{1}{(1-s)^\alpha}, s < 1 \\
m_Y(t) &= \frac{1}{(1-\beta t)^\alpha}
\end{aligned}$$

□

11.5 The Jointly Discrete Bivariate Distribution

If X, and Y are two discrete random variables, the *joint probability function of X and Y*, $p(x, y)$ is defined as

$$p(x, y) = \mathbb{P}(X = x \ \cap \ Y = y)$$

Ex: Suppose a die has 6 faces numbered 1, 2, 3, 3, 4, 4. Roll the die twice. Let X represent the number obtained the first time, and let Y represent the number obtained the second time. Then the probability distribuion for X is

x	1	2	3	4
$p_X(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$

and the probability distribution for Y is

y	1	2	3	4
$p_Y(y)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$

And thus the joint probability distribution for X and Y is

$y \backslash x$	1	2	3	4
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$
3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{9}$
4	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{9}$

Note that in this example, rolling a die twice is two independent events and so

$$\begin{aligned}
p(x_1, y_1) &= \mathbb{P}(X = x_1 \ \cap \ Y = y_1) \\
&= \mathbb{P}(X = x_1)\mathbb{P}(Y = y_1)
\end{aligned}$$

11.6 Jointly Continuous Random Variables

11.6.1 The Joint Cumulative Distribution Function

Suppose X, and Y are continuous random variables. We define the *joint cumulative distribution function* (JCDF) of X, and Y, F, as

$$F(X,Y) = \mathbb{P}(X \leq x \ \cap \ Y \leq y)$$

Properties of F:

1. $\lim_{x\to-\infty} F(X,Y) = 0 = \lim_{y\to-\infty} F(X,Y)$
2. $\lim_{x\to\infty} F(X,Y) = \mathbb{P}(Y \leq y)$= $F_Y(y)$ $F(X,Y)$
3. $\lim_{y\to\infty} F(X,Y) = \mathbb{P}(X \leq x)$= $F_X(x)$ $F(X,Y)$
4. $\lim_{(x,y)\to\infty} F(X,Y) = 1$
5. Inclusion/Exclusion Property: If $x_2 \geq x_1$, and $y_2 \geq y_1$, then $F(x_1,x_2) + F(x_2,y_2) \geq 0$, and $F(x_2,y_2) + F(x_1,y_1) \geq F(x_2,y_1) + F(x_1,y_2)$.
6. $0 \leq F(X,Y) \leq 1$

11.6.2 The Joint Probability Density Function

Suppose X, and Y are two continuous random variables. Then the *joint probability density function* (JPDF) of X, and Y, f, is defined as

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(X,Y)$$

Properties of f: If f is the JPDF of two continuous random variables, X, and Y, then

1. $f(x,y) \geq 0$
2. $\iint_A f(x,y)\ dA = 1,\ A = \mathbb{R}^2$

Ex 1: Let X, and Y be two continuous random variables and and consider the function f.

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

Find a constant $c \in \mathbb{R}$ such that f is the JPDF of X, and Y.
Solution:

$$\begin{aligned}
1 &= \iint_A cx^2y \; dA, A = \mathbb{R}^2 \\
&= c\int_0^1 \int_0^y x^2y \; dx \; dy \\
&= c\int_0^1 \frac{y^4}{3} \; dy \\
\frac{3}{c} &= \frac{1}{5} \\
c &= 15
\end{aligned}$$

Ex 2: Let X, and Y be two continuous random variables and and consider the function f.

$$f(x,y) = \begin{cases} ce^{-2y}, & \text{if } y \geq x^2, \text{ and } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find a constant $c \in \mathbb{R}$ such that f is the JPDF of X, and Y.

$$\begin{aligned}
1 &= c\int_0^\infty \int_0^{\sqrt{y}} e^{-2y} \; dx \; dy \\
&= c\int_0^\infty \sqrt{y}e^{-2y} \; dy \\
&= c\frac{\Gamma\left(\frac{3}{2}\right)}{2^{\frac{3}{2}}} \\
c &= \frac{2\sqrt{2}}{\frac{1}{2}\sqrt{\pi}} \\
c &= 4\sqrt{\frac{2}{\pi}}
\end{aligned}$$

12 May 22, 2018

12.1 An Example That We Didn't Get To on Friday

Ex: Consider the following JPDF of two continuous random variables X, and Y.

$$f(x,y) = \begin{cases} 6x, & \text{if } 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Find $\mathbb{P}\left(\frac{1}{4} \leq X \leq 1 \wedge Y \leq \frac{1}{2}\right)$

Solution: $\mathbb{P}\left(\frac{1}{4} \leq X \leq 1 \wedge Y \leq \frac{1}{2}\right)$ is simply given by

$$\mathbb{P}\left(\frac{1}{4} \leq X \leq 1 \wedge Y \leq \frac{1}{2}\right) = \iint_A f(x,y)\ dA$$

By sketching out the support its can easily be seen that $A = \left[\frac{1}{4}, 1\right] \times \left[x, \frac{1}{2}\right]$, and thus

$$\mathbb{P}\left(\frac{1}{4} \leq X \leq 1 \wedge Y \leq \frac{1}{2}\right) = \int_{\frac{1}{4}}^{\frac{1}{2}} \int_{x}^{\frac{1}{2}} 6x\ dy\ dx = \cdots = \frac{1}{16}$$

12.2 Marginal Distributions

Suppose we have two random variables X, and Y where we're given the JPF (if discrete) or the JPDF (if continuous) and we want to find the distribution of one of those variables. That's where the idea of the marginal distribution comes in. We define it as such

1. If X, and Y are discrete and $p(x,y)$ is their JPF then
   - The PF of X is $p_X(x) = \sum_y p(x,y)$
   - The PF of Y is $p_Y(y) = \sum_x p(x,y)$
2. If X, and Y are continuous and $f(x,y)$ is their JPDF then
   - The PDF of X is $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\ dy$
   - The PDF of Y is $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\ dx$

Ex 1: Given the following table find $p_X(x)$, $p_Y(y)$, and the distributions of X^2, and Y.

$y \backslash x$	-1	0	1
0	1/4	1/6	1/12
1	1/4	0	1/4

Solution: $p_X(x)$ is given by

$$p_X(x) = \sum_y p(x,y) \Rightarrow p_X(-1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$
$$\Rightarrow p_X(0) = \frac{1}{6} + 0 = \frac{1}{6}$$
$$\Rightarrow p_X(1) = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}$$

Note that $X^2 \sim \text{Ber}\left(\frac{5}{6}\right)$

Similarily, $p_Y(y)$ is given by

$$p_X(x) = \sum_y p(x,y) \Rightarrow p_Y(0) = \frac{1}{4} + \frac{1}{6} + \frac{1}{12} = \frac{1}{2}$$
$$\Rightarrow p_Y(1) = \frac{1}{4} + 0 + \frac{1}{4} = \frac{1}{2}$$

Note that $Y \sim \text{Ber}\left(\frac{1}{2}\right)$

Ex 2: Consider the following JPDF

$$f(x,y) = \begin{cases} 6x, & \text{if } 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Then

$$f_X(x) = \begin{cases} \int_x^1 6x\ dy, & \text{if } 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$
$$= \begin{cases} 3y^2, & \text{if } 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

And thus $X \sim \text{Beta}(\alpha = 3,\ \beta = 1)$

12.3 The Conditional Probability Function

Let $p(x,y)$ be the JPF of (X,Y). Ley X be such that $p_X(x) \neq 0$, then the *conditional probability probability function* of Y given $\{X = x\}$ is $p_Y(y|x)$ where

$$p_Y(y|x) = \frac{p(x,y)}{p_X(x)}$$

Similarily if $p_Y(y) \neq 0$ then the conditional probability of X given $\{Y = y\}$ is $p_X(x|y)$ where

$$p_X(x|y) = \frac{p(y,x)}{p_Y(y)}$$

Ex: Refer to the table a couple examples ago. Find the PF of Y given $\{X = 1\}$ and the PF of X given $\{Y = 1\}$.

Solution: Fixing $X = 1$ then we get the probability fistribution given by the table

y	0	1
$p_Y(y\|x = 1)$	1/4	3/4

Note that $Y \sim \text{Ber}\big(p = \frac{3}{4}\big)$. Similarily by fixing $Y = 1$ then we get the probability fistribution given by the table

x	-1	0	1
$p_X(x\|y = 1)$	1/2	0	1/2

12.4 Conditional Density

Let $f(x, y)$ be the JPDF of (X, Y). By definition

1. The conditional density function of X given $\{Y = y\}$ is given by

$$f_X(x|y) = \frac{f(x, y)}{f_Y(y)}$$

2. The conditional density function of Y given $\{X = x\}$ is given by

$$f_Y(y|x) = \frac{f(x, y)}{f_X(x)}$$

Ex: Consider the JPDF given by

$$f(x, y) = \begin{cases} 6x, & \text{if } 0 \le x \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Let $y \in]0, 1[$ be fixed. Then

$$f_X(x|y) = \begin{cases} \frac{6x}{3y^2}, & \text{if } 0 \le x \le y \\ 0, & \text{otherwise} \end{cases}$$

And therefore, given Y fixed, $X \sim Y\text{Beta}(2, 1)$.
Now suppose we let $x \in]0, 1[$ be fixed. Then

$$\begin{aligned} f_Y(y|x) &= \begin{cases} \frac{6x}{6x(1-x)}, & \text{if } x \le y \le 1 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{1}{1-x}, & \text{if } x \le y \le 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

And thus $Y \sim \text{Uniform}(x, 1)$

12.5 Independence

Suppose X, and Y are two random variables.

1. If X, and Y are discrete then they are said to be independent if

$$p(x,y) = p_X(x)p_Y(y)$$

where $p(x,y)$ is the JPF of X, and Y, and $p_X(x)$, and $p_Y(y)$ are the PFs of X and Y respectively.

2. If X, and Y are continuous then they are said to be independent if

$$f(x,y) = f_X(x)f_Y(y)$$

where $f(x,y)$ is the JPDF of X, and Y, and $f_X(x)$, and $f_Y(y)$ are the PDFs of X and Y respectively.

Ex 1: Given the table deduce whether X and Y are independent or not

$y \backslash x$	-1	0	1	$p_Y(y)$
0	1/4	1/6	1/12	1/2
1	1/4	0	1/4	1/2
$p_X(x)$	1/2	1/6	1/3	

$p(X=0, Y=1) = 0 \neq p_X(0)p_Y(1)$ and so X and Y aren't independent.

Ex 2: Consider the JPDF of X, and Y given by

$$f(x,y) = \begin{cases} 6x, & \text{if } 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

X and Y are not independent because the sketch of the support forms a triangle.

Ex 3: Consider the JPDF given by

$$f(x,y) = \begin{cases} x+y, & \text{if } x \in \,]0,1[\text{, and } y \in \,]0,1[\\ 0, & \text{otherwise} \end{cases}$$

then

$$f_X(x) = \begin{cases} x+\frac{1}{2}, & \text{if } x \in \,]0,1[\\ 0, & \text{otherwise} \end{cases}$$

and

$$f_Y(y) = \begin{cases} y+\frac{1}{2}, & \text{if } y \in \,]0,1[\\ 0, & \text{otherwise} \end{cases}$$

but

$$f(x,y) \neq f_X(x)f_Y(y), \ \forall x,y$$

as

$$f\Big(\frac{1}{4},\frac{1}{4}\Big) = \frac{1}{2} \neq \frac{9}{16} = f_X\Big(\frac{1}{4}\Big)f_Y\Big(\frac{1}{4}\Big)$$

thus X, and Y aren't independent.

Proposition: Let (X,Y) be 2 continuous random variables. Then X, and Y are independent if and only if the following conditions hold

1. The support of the JPDF of (X,Y) is a rectangle
2. The JPDF f of (X,Y) can be written as $f(x,y) = h(x)g(y)$

Ex: Consider the following JPDF

$$f(x,y) = \begin{cases} x^2e^{-x-2y}, & \text{if } 0 < x, \text{and } 0 < y \\ 0, & \text{otherwise} \end{cases}$$

Then the support obviously forms a rectangle, and

$$f(x,y) = (x^2e^{-x})(e^{-2y})$$

and thus X and Y are independent.

12.6 Expected Value

Let X, Y be two random variables. Then

1. Suppose X, and Y are discrete and $p(x,y)$ is their JPF. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be a function such that
$$E(F(X,Y)) = \sum_x \sum_y F(X,Y)p(x,y)$$
2. Suppose X, and Y are continuous and $f(x,y)$ is theit JPDF. Let the function $F : \mathbb{R}^2 \to \mathbb{R}$ be such that
$$E(F(X,Y)) = \iint_{\mathbb{R}^2} F(X,Y)f(x,y)\ dx\ dy$$

Ex: Suppose X, Y are continuous. Find $E(X)$.

Solution: Since we're looking for $E(X)$ then we want to find the expected value where $F(X,Y) = x$. Therefore

$$E(X) = \iint_{\mathbb{R}^2} xf(x,y)\ dy\ dx = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)\ dy\ dx\ = \int_{-\infty}^{\infty} xf_X(x)\ dx$$

13 May 23, 2018

13.1 Examples of Expected Value

Ex 1: Consider the JPDF of X, and Y given by

$$f(x,y) = \begin{cases} 6x, & \text{if } 0 \leq X \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Then $E(X^2Y) = \iint_D x^2 y f(x,y)\ dA = \int_0^1 \int_0^4 6x^3 y\ dx\ dy = \dfrac{1}{4}$

Ex 2: Consider the JPDF of X, and Y given by

$$f(x,y) = \begin{cases} x^2 e^{-x-2y}, & \text{if } 0 \leq y, 0 \leq x \\ 0, & \text{otherwise} \end{cases}$$

Then $E(X^3Y^2) = \iint_D x^3 y^2 f(x,y)\ dA$. Note that X and Y are independent so $E(X^2Y^2) = E(X^3)E(Y^2) = \left(\int_0^\infty y^2 e^{-2y}\ dy\right)\left(\int_0^\infty x^5 e^{-x}\ dx\right) = \Gamma(6)\dfrac{\Gamma(3)}{2^3} =$ 30

13.2 Properties of Expected Value

Let X, and Y be two random variables. Then

1. If $C \in \mathbb{R}$ is constant then $E(C) = C$
2. If $F : \mathbb{R}^2 \to \mathbb{R}$, and $G : \mathbb{R}^2 \to \mathbb{R}$ are functions, and $\alpha, \beta \in \mathbb{R}$ are constants, then
   - $E(\alpha F(X,Y) + \beta G(X,Y)) = E(\alpha F(X,Y)) + E(\beta G(X,Y))$
   - In particular $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$

Proposition: If X, and Y are independent random variables and $F : \mathbb{R}^2 \to \mathbb{R}$, $G : \mathbb{R}^2 \to \mathbb{R}$ are functions, then

$$E(F(X)G(Y)) = E(F(X))E(G(Y))$$

Proof: Follows immediately from the definitions of expected value and independence.

13.3 Moment Generating Functions

If X, and Y are two independent random variables then the *moment generating function* of X and Y, $m_{X+Y}(t)$, is defined as

$$m_{X+Y}(t) = m_X(t)m_Y(t)$$

Ex 1: Suppose $X_1 \sim \text{Poisson}(\lambda_1)$, and $X_2 \sim \text{Poisson}(\lambda_2)$ are independent. Then

$$\begin{aligned} m_{X_1+X_2}(t) &= m_{X_1}(t) m_{X_2}(t) \\ &= e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)} \end{aligned}$$

and therefore $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$.

Ex 2: Suppose $X_1 \sim \text{N}(\mu_1,\ \sigma_1^2)$, and $X_2 \sim \text{N}(\mu_2,\ \sigma_2^2)$ are independent. Then

$$\begin{aligned} m_{X_1+X_2}(t) &= m_{X_1}(t) m_{X_2}(t) \\ &= \left(e^{\mu_1 t + \sigma_1^2 t^2/2}\right)\left(e^{\mu_2 t + \sigma_2^2 t^2/2}\right) \\ &= e^{\mu_1 t + \sigma_1^2 t^2/2 + \mu_2 t + \sigma_2^2 t^2/2} \end{aligned}$$

and thus $X_1 + X_2 \sim \text{N}(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$

Ex 3: Suppose $X_1 \sim \text{Gamma}(\alpha_1,\ \beta)$ and $X_2 \sim \text{Gamma}(\alpha_2,\ \beta)$. Then

$$m_{X_1+X_2(t)} = m_{X_1}(t) m_{X_2}(t) = \frac{1}{(1-\beta t)^{\alpha_1}} + \frac{1}{(1-\beta t)^{\alpha_2}} = \frac{1}{(1-\beta t)^{\alpha_1+\alpha_2}}$$

and thus $X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2,\ \beta)$

The above example holds for all Gamma random variables X_1, $X2$ given their βs are equal.

Ex 4: Suppose $X_1 \sim \mathcal{X}^2(p_1)$, are $X_2 \sim \mathcal{X}^2(p_2)$ are independent random variables, then $X_1 + X_2 \sim \mathcal{X}^2(p_1 + p_2)$

13.4 Covariance

Let X, and Y be two random variables. *The covariance* of X, and Y, $\text{cov}(X, Y)$ is defined as

$$\text{cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

where μ_X, and μ_Y are defined as $E(X)$, and $E(Y)$ respectively.

Remark: Note that $\text{cov}(X, X) = V(X)$.
Proof: The proof of this immediately mirrors the proof of the following proposition.

Proposition: $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$

Proof: Let $\mu_X = E(X)$, $\mu_Y = E(Y)$.

$$\begin{aligned}
\text{cov}(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) \\
&= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y) \\
&= E(XY) - E(X)\mu_Y - E(Y)\mu_Y + \mu_X\mu_Y \\
&= E(XY) - \mu_X\mu_Y \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

□

Ex: Consider the following JPDF

$$f(x,y) = \begin{cases} 6x, & \text{if } 0 \le x \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Note that $X \sim \text{Beta}(2,2)$, $Y \sim \text{Beta}(3,1)$, therefore $E(X) = \frac{1}{2}$, and $E(Y) = \frac{3}{4}$. Therefore, to find $\text{cov}(X,Y)$ we simply do

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y) = E(XY) - \frac{3}{8}$$

Now we need to find E(XY). By definition

$$E(XY) = \int_0^1 \int_0^4 xy6x \; dx \; dy = \int_0^1 2y^3 \; dy = \frac{2}{5}$$

13.5 Properties of Covariance

Let X_1, X_2, Y_1, and Y_2 be random variables, and let α_1, $\alpha_2 \in \mathbb{R}$ be constants. Then

1. $\text{cov}(X,Y) = \text{cov}(Y,X)$
2. $\text{cov}(\alpha_1 X_1 + \alpha_2 X_2, Y) = \alpha_1 \text{cov}(X_1, Y) + \alpha_2 \text{cov}(X_2, Y)$
3. $\text{cov}(X, \alpha_1 Y_1 + \alpha_2 Y_2) = \alpha_1 \text{cov}(X, Y_1) + \alpha_2 \text{cov}(X, Y_2)$
4. $\text{cov}(X,X) = V(X)$
5. $\text{cov}(X,X) = 0 \iff X$ is constant

Ex 1:

$$\begin{aligned}
V(X+Y) &= \text{cov}(X+Y, X+Y) \\
&= \text{cov}(X+Y, X) + \text{cov}(X+Y, Y) \\
&= \text{cov}(X,X) + \text{cov}(Y,X) + \text{cov}(X,Y) + \text{cov}(Y,Y) \\
&= V(X) + 2 \cdot \text{cov}(X,Y) + V(Y)
\end{aligned}$$

6. $V(\alpha X + \beta Y) = \alpha^2 V(X) + 2\alpha\beta \text{cov}(X,Y) + \beta^2 V(Y)$

7. $V\left(\sum_{i=1}^{n} \alpha_i X_i\right) = \sum_{i=1}^{n} \alpha_i^2 V(X_i) + 2 \sum_{i\ <\ j} \alpha_i \alpha_j \mathrm{cov}(X_i, X_j)$

Proposition: If X and Y are independent, then $\mathrm{cov}(X, Y) = 0$.
Proof: Trivial

13.6 Uncorrelated Random Variables

Two random variables X, and Y are said to be *uncorrelated* if $\mathrm{cov}(X, Y) = 0$.

Remark: If X, and Y are independent then they must be uncorrelated, but the converse isn't necessarily true.

Proposition: If X, and Y are random variables then $|\mathrm{cov}(X, Y)| \leq \sqrt{V(X)}\sqrt{V(Y)}$. Moreover, $|\mathrm{cov}(X, Y)| = \sqrt{V(X)}\sqrt{V(Y)}$ when $Y = \alpha X + \beta$.

We define the *correlation coefficient* between X, and Y as

$$p(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Ex: Consider the JPDF

$$f(x, y) = \begin{cases} 6x, & \text{if } 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Then $X \sim \mathrm{Beta}(2,\ 2)$, and $Y \sim \mathrm{Beta}(3,\ 1)$ and therefore $V(X) = \frac{4}{80}$, and $V(Y) = \frac{3}{80}$. Also $\mathrm{cov}(X, Y) = \frac{1}{40}$. Therefore

$$p(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{1/40}{\sqrt{12}/80} = \frac{1}{\sqrt{5}}$$

14 May 24, 2018

14.1 Multinomial Distribution

Suppose an experiment yields $k \geq 2$ outcomes $E_1, E_2, \ldots, E_k$, where $p_1 = \mathbb{P}(E_i)$, with $i = 1, 2 \ldots, k$. Suppose we run the experiment n times and let X_i be the random variable containing the number of times that E_i occured. Note that the distribution of each of the X_is is binomial. The probability function $\mathbb{P}$ for the multinomial distributions is given by

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \binom{n}{X_1, X_2, \ldots, X_k} p_1^{X_1} p_2^{X_2} \ldots p_k^{X_k}$$

Ex: A restaurant offers 3 menus, I, II, and III. Any customer chooses either I, II, or III with probability p_1, p_2, or p_3 respectively. Supporse 10 customers arrive at the restaurant and select menus independently. Then

$$\mathbb{P}(X_{\mathrm{I}} = 1, X_{\mathrm{II}} = 3, X_{\mathrm{III}} = 6) = \frac{10!}{1!3!6!} p_1^1 p_2^3 \ldots p_k^6$$

Proposition: In the multinomial setting

1. $X_i \sim \mathrm{Bin}(n,\ p_i)$
2. If $i, j \in \mathbb{R}$ with $i \neq j$ then $\mathrm{cov}(X_i,\ X_j) = -np_ip_j$

Proof of 2: Let $i, j \in \mathbb{R}$ be fixed and WLOG suppose $i = 1$, $j = 2$. Then define a random variable U_ℓ as

$$U_\ell := \begin{cases} 1, & \text{if the } \ell^{th} \text{ element gives } E_1 \\ 0, & \text{otherwise} \end{cases}$$

Then $U_\ell \sim \mathrm{Ber}(p_1)$, and $\mathrm{cov}(U_\ell,\ U_{\ell'}) = 0$ if $\ell \neq \ell'$. Now define a random variable X_1 as $X_1 := \displaystyle\sum_{\ell=1}^{n} U_\ell$

Similarily, define a random variable, V_k, as

$$V_k := \begin{cases} 1, & \text{if the } k^{th} \text{ element gives } E_2 \\ 0, & \text{otherwise} \end{cases}$$

Then $V_k \sim \mathrm{Ber}(p_2)$, and $\mathrm{cov}(V_k,\ V_{k'}) = 0$ if $k \neq k'$. Now define the random variable X_2 as $X_2 := \displaystyle\sum_{k=1}^{n} V_k$

Note that by just looking at all cases, $U_xV_y = 0$ for any $x, y \in \mathbb{R}$, and so $\mathrm{cov}(U_\ell, V_k) = E(U_\ell V_k) - E(U_\ell)E(V_k) = -E(U_\ell)E(V_k) = -p_1p_2$. Therefore

$$\mathrm{cov}(U_\ell V_k) = \begin{cases} 0 & \text{if } k \neq \ell \\ -p_1p_2 & \text{otherwise} \end{cases}$$

FInally

$$\begin{aligned}
\mathrm{cov}(X_1, X_2) &= \mathrm{cov}\left(\sum_{\ell=1}^{n} U_\ell,\ \sum_{k=1}^{n} V_k\right) \\
&= \sum_{\ell=1}^{n}\sum_{k=1}^{n} \mathrm{cov}(U_\ell,\ V_k) \\
&= \cancelto{0}{\sum_{\ell\neq k}^{n} \mathrm{cov}(U_\ell,\ V_k)} \quad + \sum_{\ell=1}^{n} \mathrm{cov}(U_\ell,\ V_k) \\
&= -np_1p_2
\end{aligned}$$

□

14.2 Conditional Expectation

Suppose X and Y are discrete random variables. Let $p_X(x|y)$ be the conditional probability function of X given $\{Y = y\}$. Set $G(Y) := E(X|Y = y)$. The random variable $G(Y)$ is called the *conditional expectation* of X given Y and is denoted $E(X|Y)$. $E(Y|X)$ is defined in parallelly. Note that

$$G(Y) = E(X|Y) = \sum_x x p_X(x, y)$$

Suppose X and Y are continuous random variables. Let $f(x|y)$ be the conditional PDF of X given $\{Y = y\}$. Set $G(Y) := E(X|Y = y)$. The random variable $G(Y)$ is called the *conditional expectation* of X given Y and is denoted $E(X|Y)$. Again, $E(Y|X)$ is defined in parallelly. Note that

$$G(Y) = E(X|Y) = \int_{-\infty}^{\infty} x f(x|y)\ dx$$

Remark: If $F : \mathbb{R} \to \mathbb{R}$ is a function then we can use the above definitions to define $E(F(X)|Y)$.

Ex: Consider the JPDF

$$f(x, y) = \begin{cases} 6x, & \text{if } 0 \le x \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Then $X \sim$ Beta(2, 2), and $Y \sim$ Beta(3, 1)...
... but given $\{Y = y\}$ $X \sim y\cdot$Beta(2, 1), and given $\{X = x\}$ $Y \sim$ Uniform(x, 1).

Then $G(Y) = E(X|Y = y) = \frac{2}{3}y$ and $E(X|Y) = \frac{2}{3}Y$, and $G(X) = E(Y|X = x) = \frac{1+x}{2}$, and $E(Y|X) = \frac{1+X}{2}$

Properties: If X and Y are random variables then

1. (Tower Property): $E(E(F(Y)|X)) = E(F(Y))$
2. If $H : \mathbb{R} \to \mathbb{R}$ then $E((Y \cdot H(X))|X) = H(X)E(Y|X)$

Proof of the Tower Property in the Continuous Case:
If $G(X) = E(F(Y)|X)$ then

$$G(X) = \int_{-\infty}^{\infty} F(y)f(y|x)\ dy$$

Also

$$E(E(F(Y)|X)) = E(G(X)) = \int_{-\infty}^{\infty} G(X)f_X(x)\ dx$$

Therefore

$$\begin{aligned}
E(E(F(Y)|X)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(y)f(x|y)\ dy\ f_X(x)\ dx \\
&= \int_{-\infty}^{\infty} F(y)\int_{-\infty}^{\infty} f(y|x)f_X(x)\ dx\ dy \\
&= \int_{-\infty}^{\infty} F(y)\int_{-\infty}^{\infty} f(x,y)\ dx\ dy \\
&= \int_{-\infty}^{\infty} F(y)f_Y(y)\ dy \\
&= E(F(Y))
\end{aligned}$$

□

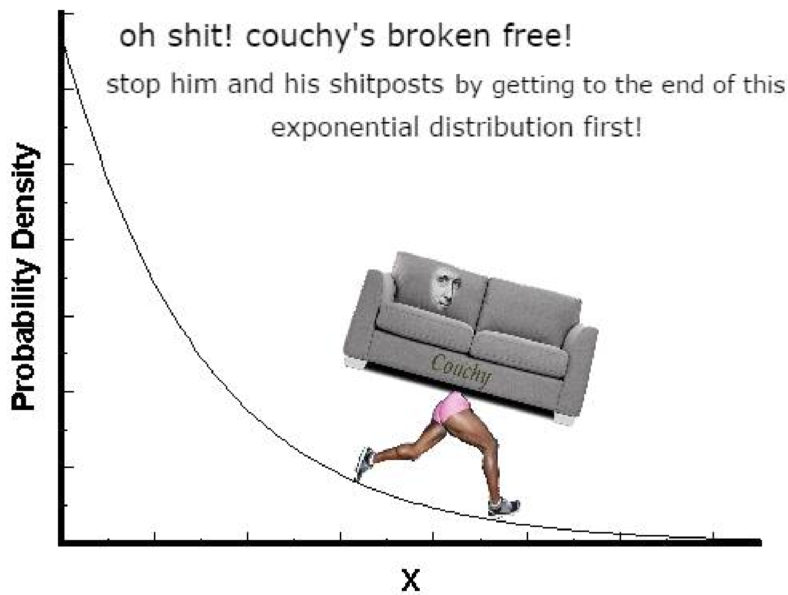
Theorem: (Variance Formula) If X and Y are random variables then

$$V(Y) = E(V(Y|X)) + V(E(Y|X))$$

where $V(Y|X) = E(Y^2|x) - (E(Y|X))^2$

Proof: He did the proof but I didn't write it down so I need to find it.

15 May 29, 2018


oh shit! couchy's broken free!
stop him and his shitposts by getting to the end of this exponential distribution first!
Probability Density
Couchy
X