Index of /~abhisheksaha/sta4321

Name	Last modified	Size	Description
Parent Directory		-	
beta.jpg	21-Oct-2012 16:59	923K	
cdf1.jpeg	09-Sep-2012 20:42	50K	
density.jpg	04-Oct-2012 16:14	110K	
density2.jpg	04-Oct-2012 16:53	99K	
exam-4discussion.pdf	12-Dec-2012 01:19	2.0M	
exam2-solution.pdf	01-Nov-2012 00:35	2.5M	
exam2scores.pdf	02-Nov-2012 10:11	53K	
exam3_tips.pdf	09-Nov-2012 15:20	85K	
gamma_less.jpg	14-Oct-2012 23:34	98K	
gamma_more.jpg	13-Oct-2012 09:12	99K	
gradecutoff.pdf	26-Sep-2012 11:01	71K	
lect1.pdf	28-Aug-2012 02:13	256K	
lect1.tex	28-Aug-2012 02:13	2.9K	
lect2.pdf	25-Aug-2012 09:28	151K	
lect2.tex	25-Aug-2012 09:28	6.8K	
lect3.pdf	26-Aug-2012 14:20	119K	

lect3.tex	26-Aug-2012 14:19	5.1K
lect4.pdf	13-Sep-2012 20:43	173K
lect4.tex	13-Sep-2012 20:43	7.7K
lect5.pdf	30-Aug-2012 22:38	164K
lect5.tex	30-Aug-2012 22:38	5.8K
lect6.pdf	04-Sep-2012 20:29	105K
lect6.tex	04-Sep-2012 20:29	3.8K
lect7.pdf	10-Sep-2012 10:56	151K
lect7.tex	08-Sep-2012 21:27	8.5K
lect8.pdf	10-Sep-2012 00:25	190K
lect8.tex	10-Sep-2012 00:25	6.3K
lect9.pdf	12-Sep-2012 10:54	174K
lect9.tex	12-Sep-2012 10:54	5.6K
lect10.pdf	14-Sep-2012 15:38	101K
lect10.tex	14-Sep-2012 15:38	3.7K
lect11.pdf	04-Oct-2012 12:41	150K
lect11.tex	04-Oct-2012 12:41	5.7K
lect12.pdf	21-Sep-2012 00:02	139K
lect12.tex	21-Sep-2012 00:02	4.6K

lect13.pdf	23-Sep-2012 23:05	127K
lect13.tex	22-Sep-2012 18:12	5.2K
lect14.pdf	04-Oct-2012 12:12	138K
lect14.tex	04-Oct-2012 12:12	5.9K
lect15.pdf	28-Sep-2012 11:26	125K
lect15.tex	28-Sep-2012 11:26	6.4K
lect16.pdf	03-Oct-2012 12:27	142K
lect16.tex	03-Oct-2012 12:27	6.2K
lect17.pdf	03-Oct-2012 16:01	177K
lect17.tex	03-Oct-2012 16:01	3.8K
lect18.pdf	04-Oct-2012 23:26	210K
lect18.tex	04-Oct-2012 23:26	4.3K
lect19.pdf	08-Oct-2012 10:04	118K
lect19.tex	08-Oct-2012 10:04	5.3K
lect20.pdf	09-Oct-2012 23:09	133K
lect20.tex	10-Oct-2012 07:13	5.1K
lect21.pdf	12-Oct-2012 09:07	134K
lect21.tex	13-Oct-2012 07:45	5.9K
lect22.pdf	15-Oct-2012 18:35	520K
lect22.tex	14-Oct-2012 23:35	7.3K
lect23.pdf	14-Nov-2012 10:53	146K
lect23.tex	19-Oct-2012 07:12	9.4K
lect24.pdf	22-Oct-2012 08:54	1.0M

lect24.tex	22-Oct-2012 08:54	6.0K
lect25.pdf	24-Oct-2012 10:21	152K
lect25.tex	24-Oct-2012 10:21	5.6K
lect26.pdf	29-Oct-2012 15:06	135K
lect26.tex	29-Oct-2012 15:06	8.3K
lect27.pdf	31-Oct-2012 17:37	121K
lect27.tex	31-Oct-2012 17:37	5.6K
lect28.pdf	04-Nov-2012 00:47	130K
lect28.tex	04-Nov-2012 00:47	6.4K
lect29.pdf	05-Nov-2012 16:33	148K
lect29.tex	05-Nov-2012 16:33	5.9K
lect30.pdf	08-Nov-2012 02:01	130K
lect30.tex	08-Nov-2012 02:01	5.8K
lect31.pdf	18-Nov-2012 22:30	212K
lect31.tex	18-Nov-2012 22:30	5.7K
lect32.pdf	24-Nov-2012 15:10	325K
lect33.pdf	26-Nov-2012 13:16	112K
lect33.tex	26-Nov-2012 13:16	4.2K
lect34.pdf	28-Nov-2012 13:23	213K
lect35.pdf	30-Nov-2012 18:52	2.0M
negbin-var.pdf	17-Oct-2012 13:29	194K

norm_mean.jpg	13-Oct-2012 13:35	79K	
norm_var.jpg	13-Oct-2012 13:15	100K	
partition.jpg	29-Aug-2012 16:38	34K	
rainfall.jpg	03-Oct-2012 00:38	51K	
sol-1-2-3.pdf	29-Aug-2012 11:28	403K	
sol-4-5.pdf	31-Aug-2012 13:20	575K	
sol-7-8.pdf	11-Sep-2012 11:45	500K	
sol-10-11-12.pdf	24-Sep-2012 12:13	527K	
sol-13-14-16.pdf	03-Oct-2012 15:56	350K	
sol-18-19-20.pdf	12-Oct-2012 21:02	503K	
sol-22-23-24-25-27.pdf	01-Nov-2012 00:39	5.5M	
sol-28.pdf	23-Nov-2012 15:01	3.2M	
sol-32-33.pdf	01-Dec-2012 22:55	3.0M	
syllabus.pdf	21-Aug-2012 22:17	53K	
venn1.jpg	21-Aug-2012 09:52	69K	
venn2.jpg	21-Aug-2012 09:58	45K	
venn3.jpg	21-Aug-2012 10:29	53K	

Apache/2.2.15 Server at www.stat.ufl.edu Port 80

Lecture 1

Agenda

1. Basic Set Theory

Basic Set Theory

Definition 1. *A set is a well defined collection of distinct objects, which we call the elements or points of the set.*

Sets are generally denoted by capital letters A, B, C, ... and their elements by small letters like a, b, c, ...

Notation ::

1. $a \in A$ means the element a belongs to the set A. $a \notin A$ means the element a does not belong to the set A.
2. $A \subset B$ means the set A is a subset of the set B; i.e. every element of A is also an element of B.
3. $A \cup B$ pronounced as A union B, is the collection the points which belong to either A or B or both.
4. $A \cap B$ pronounced as A intersection B, is the set of points which are common to both A and B.
5. $\emptyset$ is the null set or the empty set which contains no points.
6. S is the universal set i.e. the collection of all points of interest for the present situation
7. $\bar{A}$ pronounced as A complement, denotes set of all points in S which are not in A
8. A and B are called *disjoint* or *mutually exclusive* if $A \cap B = \emptyset$ i.e. A and B have nothing in common

Example

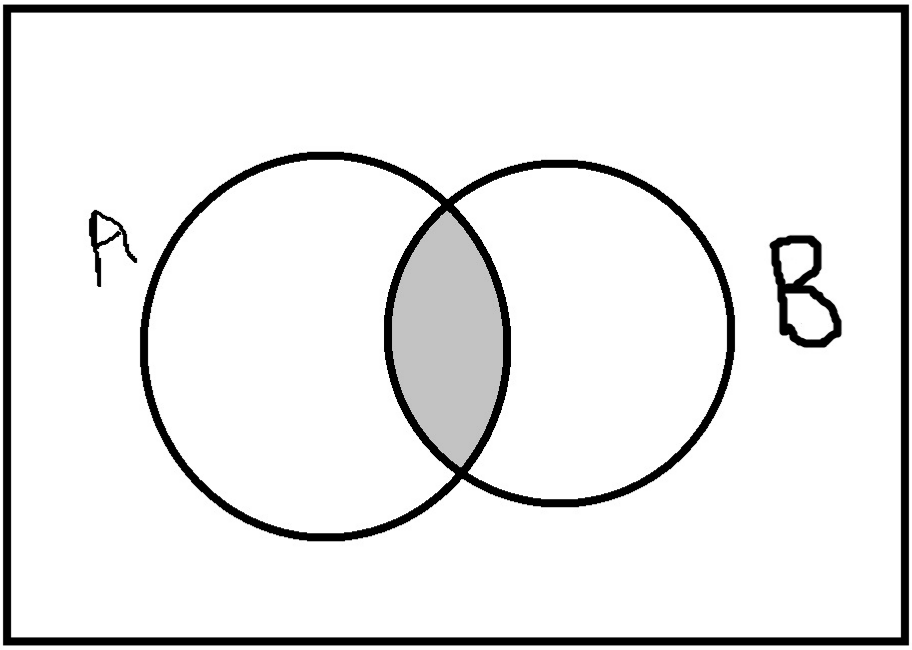
$S = \{1, 2, 3, 4, 5\}$
$A = \{1, 2, 3\}$ $B = \{2, 3, 4\}$ and $C = \{4\}$
then $A \cup B = \{1, 2, 3, 4\}$
$A \cap B = \{2, 3\}$
$A \cap C = \emptyset$
$C \subset B$ thus $B \cup C = B$ and $B \cap C = C$
$\bar{A} = \{4, 5\}$

Example practiced in class

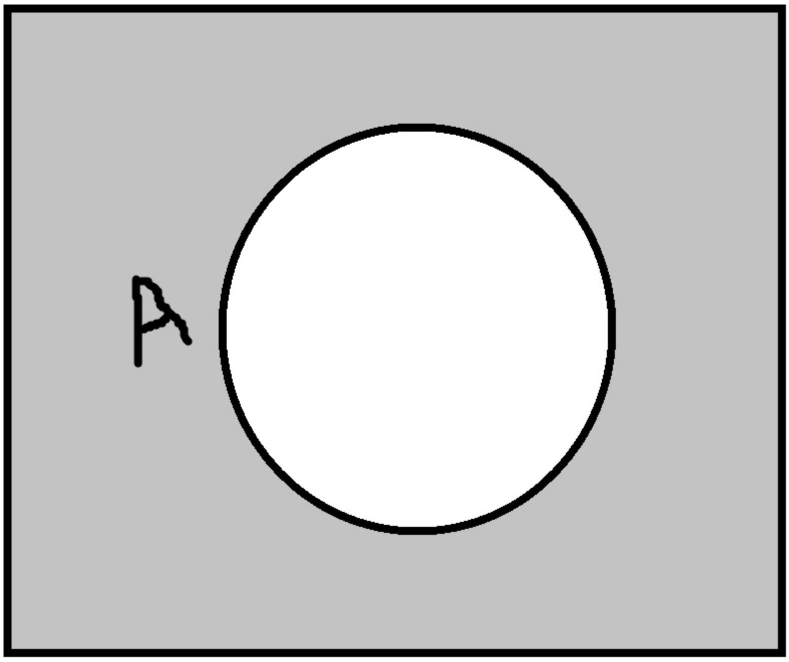
$S = \{Red, Blue, Green, Yellow, Pink\}$
$A = \{Red, Blue, Green\}$ $B = \{Red, Green, Yellow\}$ and $C = \{Blue, Pink\}$
then $A \cup B =?$
$A \cap B =?$
$A \cap C =?$
$\bar{A} =?$

Venn Diagrams

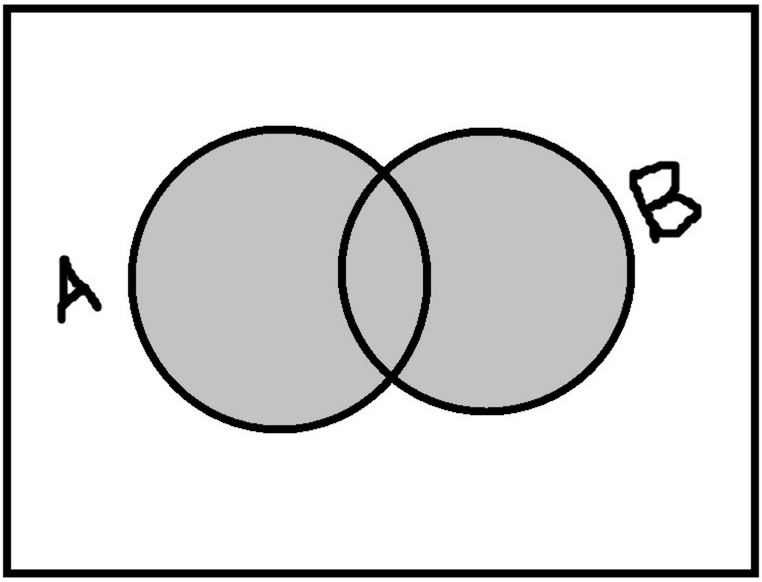
Venn Diagrams are graphical ways of representing sets.



The shaded area is $A \cap B$



The shaded area is $\bar{A}$



The shaded area is $A \cup B$

Distributive Laws

1. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
2. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

De Morgan's Laws

1. $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$
2. $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$

`Homework :: Verify the Distributive and De Morgan's Laws for the examples given above. From the textbook do 2.2 and 2.8`

Lecture 2

Agenda

1. Sample Spaces and Events

2. Probability

Sample Spaces and Events

Whenever we perform any experiment it can result in several different outcomes. But before we perform the experiment we can't exactly say which outcome will occur.

Definition 1. *The sample space S of any random experiment is the set of all possible outcomes of the experiment.*

Example

Experiment :: Toss a coin 3 times
Sample Space :: $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$
There are $2^3 = 8$ possible outcomes. This is an example of a discrete or countable sample space.

Example

Experiment :: Observe the height in ft of a randomly chosen UF student.
Sample Space :: $S = [4, 7]$ i.e. all real numbers between 4 to 7
This is an example of a continuous or uncountable sample space.

Definition 2. *An EVENT is any collection of sample points. In other words any subset of the sample space S is called an event.*

Example

Experiment :: Toss a coin 3 times
Sample Space :: $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$
A = Event that there is atleast one heads
$A = \{HHH, HHT, HTH, THH, HTT, THT, TTH\}$
B = event that there is atmost one heads
$B = \{HTT, THT, TTH, TTT\}$

Probability

Intuitively the probability of any event is a number between 0 and 1 which shows us how likely the event is to occur in a single performance of the experiment. If the probability is 1 that means the event will surely occur, if it's 0 then the event won't occur. And likewise if the probability moves up from 0 towards 1 that means it becomes more likely to occur.

All this talk gives us the feeling of probability. But if we want to define probability rigorously what do we do ?

Old fashioned way

If I give you a coin and tell you that the probability of Heads for this coin is 0.4, that means if I toss the coin a very large number of times then approximately 40% of the times Heads will occur.
Thus previously mathematicians used to define probability of an event like this

Definition 3. ***Probability*** *of an event A is a number p between 0 and 1, such that if you run the experiment a very large number of times, then for p proportion of times A will occur and for $(1-p)$ proportion of times $\bar{A}$ will occur.*

This is a very nice and good definition of probability. But after some time mathematicians encountered some problems with it. So now we will learn the new definition which we will use in this course.

The new definition

Let
$S =$ Sample space of a random experiment
$\mathcal{A} =$ Collection of all possible events

Definition 4. *A probability assisgnment* ***P()*** *for a random experiment is a numerically valued function that assigns a value $P(A)$ to every event A such that the following axioms are satisfied*

- $P(A) \geq 0$ *for any event A*
- $P(S) = 1$

- *If $A_1, A_2, A_3, \ldots$ is a sequence of mutually exclusive events i.e. $A_i \cap A_j = \emptyset$ for all pairs $i \neq j$ then*

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Notice two things, firstly we are not defining the probability of an single event, rather we are defining the probabilities for all the events together. Secondly we are not saying how to actually calculate the probabilities, rather we are giving 3 axioms such that a function $P()$ will be called a probability if it satisfies those axioms; as to how to provide such a function without disturbing the 3 axioms is the headache of the mathematician involved not of this definiton.

Properties

1. $P(\emptyset) = 0$
   Take $A_1 = S$ and $A_2 = A_3 = \ldots = \emptyset$.We can check that the events are mutually exclusive. Thus $P(S) = P(S) + P(\emptyset) + P(\emptyset) + P(\emptyset) + \ldots$
   i.e. $1 = 1 + P(\emptyset) + P(\emptyset) + P(\emptyset) + \ldots$
   Now if $P(\emptyset)$ is even slightly positive then the LHS will be 1 and RHS will be ∞.Hence $P(\emptyset) = 0$.

2. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$
   Simply take $A_1 = A$, $A_2 = B$ and the rest of the A_i's to be $\emptyset$.

3. If $A \subset B$ then $P(A) \leq P(B)$

Defining and calculating the probability of an event by sample point method

1. Define the experiment

2. Construct the sample space

3. Assign probabilites to each of the sample points, making sure that they sum up to 1.

4. Express the event of interest as a collection of sample points

5. Find $P(A)$ by summing probabilites of sample points in A.

Example

Random Experiment Choose a person from 4 persons, with no preference to any person.

- What is the the sample space ?
  Since there are only 4 possible outcomes the sample space is $S = \{1, 2, 3, 4\}$.
- What is the set of all possible events? Recall that $\mathcal{A}$ which is the set of all possible events is essentially the set of all possible subsets of S.Hence

$$\begin{aligned}\mathcal{A} = \{&\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\},\\ &\{2, 4\}, \{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\}\end{aligned}$$

- How to assign a probability assignment $\mathbf{P()}$ which express our belief in how the experiment is conducted and **also satisfies the 3 axioms** ?

Solution: Based on your belief in how the experiment is conducted assign probability for each of the individual outcomes, i.e. *for each point in the sample space* making sure that they add up to 1. For this experiment since there is no preference to any person,

$$P(\{1\}) = \frac{1}{4}, P(\{2\}) = \frac{1}{4}, P(\{3\}) = \frac{1}{4}, P(\{4\}) = \frac{1}{4}$$

NOW DEFINE THE PROBABILITY OF ANY EVENT AS SUM OF PROBABILITIES OF POINTS BELONGING TO THAT EVENT.
Foe example, if A is the event that person 1 or 2 is chosen,

$$P(A) = P(\{1, 2\}) = P(\{1\}) + P(\{2\}) = \frac{1}{4} + \frac{1}{4} = 0.5$$

This procedure will guarantee that the probability assignment satisfies those 3 axioms, for discrete sample spaces.

`Homework :: From the book try 2.20,2.21,2.22,2.29,2.31. If you can't do 2.29 or 2.31 after lecture 2 don't be discouraged we will do do these more in lecture 3`

Lecture 3

Agenda

1. One more example to understand the formal definition of probability
2. Fundamental principles of counting
3. Evaluating probabilities using permutations

Example

Experiment: There are 3 mail boxes. Three people come one after another, and choose one mailbox at random (with no preference to any one) and put mail in it.
Find the probability that no mailbox remain empty.

- Each outcome is a sequence of 3 numbers, each number representing the mailbox chosen by the corresponding person.

  $$S = \{123, 122, 131, \ldots\}$$

  For example 122 means the outcome where the 1st person puts in mailbox 1, the 2nd and 3rd person chooses mailbox 2.
  Total number of such outcomes $=3 \times 3 \times 3 = 27$

- Since no person has any preference over any mailbox we assign same probability for all the outcomes. $P(\{s\}) = \frac{1}{27}\ \forall s \in S$

- A = event that all mailboxes are chosen once

  $$\Rightarrow A = \{123, 321, 132, 231, 312, 213\}$$

  Hence

  $$\begin{aligned} P(A) &= P(\{123\}) + P(\{321\}) + P(\{132\}) + P(\{231\}) + P(\{312\}) + P(\{213\}) \\ &= \frac{1}{27} + \frac{1}{27} + \frac{1}{27} + \frac{1}{27} + \frac{1}{27} + \frac{1}{27} \\ &= \frac{6}{27} \end{aligned}$$

Do we always need to go through this procedure for calculating probabilites of events ? No. We can often use counting rules to get around the situation.

Fundamental Principles of Counting

Counting Rule 1

Suppose we are performing an experiment where all outcomes are equally likely, hence we assign the same probability to every sample point. If there are N sample points and for any event of interest A let there be n_A sample points in A, then

$$P(A) = \frac{n_A}{N}$$

The Fundamental Principle of Counting

Suppose that an experiment consists of doing two succesive tasks. The first task can result in n_1 outcomes and then for each of those outcomes, the second task can result in n_2 outcomes. Then the total number of outcomes of the experiment is $n_1 n_2$.

Example

Experiment: Choose 10 people randomly.
Find the probability that there is atleast one match in the 10 birthdays ?

- Sample space consists of collection of all 10-tuples like

  $$(X_1, X_2, X_3, \ldots, X_{10})$$

  where each X_i is a number between 1 and 365.
  So total number of points in the sample space $= 365^{10}$.

- Lets first find the number of points in the sample space where there is no match in the 10 birthdays; i.e. for any $i \neq j$, $X_i \neq X_j$. This we do by the counting principle. First we choose X_1 in 365 ways. For each of those choices we can choose X_2 in 364 ways because X_2 can't equal X_1. For each of these choices of X_1 and X_2 we can choose X_3 in 363 different ways. Continuing like this we can deduce that the total number of

sample points where bithdays don't match $= 365 \times 364 \times 363 \times \ldots \times 356$
$= \frac{365!}{(365-10)!}$.
Hence total no sample points where atleast one birthday match $= 365^{10} - \frac{365!}{(365-10)!}$

- Thus the probability that there is atleast one birthday match $= \frac{365^{10} - \frac{365!}{(365-10)!}}{365^{10}}$
$= 1 - \frac{365!}{(365-10)! \times 365^{10}}$

Permutations

- Total number of ways of selecing r objects from n objects where
  - *order of selection is important*
  &
  - *the same object can be selected more than once (we call this selection with replacement)*
  $= n \times n \times \ldots \times n = n^r$

- Total number of ways of selecing r objects from n objects where
  - *order of selection is important*
  &
  - *the same object cannot be selected more than once (we call this selection without replacement)*
  $= n \times (n-1) \times \ldots \times (n-(r-1)) = \frac{n!}{(n-r)!}$
  This happens because the first object can be chosen in n ways, and then the second object in $(n-1)$ ways and then the third object in $(n-2)$ ways; and so on ....
  Notice that $r \leq n$. By the way, the quantity $\frac{n!}{(n-r)!}$ is written as ${}^nP_r$

Homework ::
I am choosing a 4 digit number at random.
(a) What's the probability that all the digits are different?
(b) If we represent the 4-digit number by abcd what's the probability $|a-d| = 2$ *?*

You can visit http://math.uncc.edu/ hbreiter/problems/Hawaii/Combo.pdf for more problems. You may not be able to do all the problems now, because we haven't learnt everything yet. But I will be using this problem list in the future also.

Lecture 4

Agenda

1. Some more counting rules

2. And the related examples

Remember the following two results from previous lecture.

Lemma 1. *Total number of ways of selecing r objects from n objects where*
- order of selection is important
&
- the same object can be selected more than once (we call this selection with replacement)
$= n \times n \times \ldots \times n = n^r$

Lemma 2. *Total number of ways of selecing r objects from n objects where*
- order of selection is important
&
- the same object cannot be selected more than once (we call this selection without replacement)
$= n \times (n-1) \times \ldots \times (n-(r-1)) = \frac{n!}{(n-r)!}$

Please note that when order of selection is important we call it a **permutation**. When order is not important we call it a **combination** and the following two rules are for combinations.

Some more counting rules – Combinations

Lemma 3. *No of ways to selecting r objects from n distinct objects*
without replacement (i.e. without repetition)
&
where order is not important
is ${}^nC_r = \frac{n!}{r!(n-r)!}$.

Proof. Note that now we are only interested in, which r objects are being chosen and not the order in which they are chosen. So lets call the total number of ways of choosing r objects as ${}^nC_r$. We have to calculate ${}^nC_r$, but instead of directly calculating that, lets try a round about way.
We already know from the previous lecture that the total number of ways

of choosing r objects without replacement from n objects where **order is important** is ${}^nP_r$. But this can be done also in another way. First we choose r objects from n without replacement so that **order is not important** in ${}^nC_r$ ways. Then for each such group of r objects, we can arrange them in $r!$ ways, so as to make the order important. So we can do the same job in two different ways. Hence

$${}^nP_r = {}^n C_r \times r!$$

Thus ${}^nC_r = \frac{{}^nP_r}{r!} = \frac{n!}{r!(n-r)!}$ ☐

Notation :: ${}^nC_r$ is also denoted as $\binom{n}{r}$

Lemma 4. *Number of ways of dividing n things into k groups where the 1st group should contain n_1 elements, 2nd group n_2 elements, ..., k-th group n_k elements, where obviously $n = n_1 + n_2 + \ldots + n_k$ holds is*

$$\frac{n!}{n_1!n_2!\ldots n_k!}$$

Proof. The n_1 elements for the 1st group can be chosen in $\binom{n}{n_1}$. Then the n_2 elements for the second group can be chosen in $\binom{n-n_1}{n_2}$ ways. Continuing like this finally for the k-th group the n_k elements can be chosen in $\binom{n-n_1-n_2-\ldots-n_{k-1}}{n_k}$. Hence the total number of ways of partitioning the n elements in k groups

$$= \binom{n}{n_1}\binom{n-n_1}{n_2}\ldots\binom{n-n_1-n_2-\ldots-n_{k-1}}{n_k} = \frac{n!}{n_1!n_2!\ldots n_k!}$$

☐

Lemma 5. *No of ways to selecting r objects from n distinct objects*
with replacement (i.e. with repetition)
&
where order is not important
is $\binom{n+r-1}{r}$.

Proof. Since order is not important and objects can be repeated, a sample point looks like $(l_1, l_2, \ldots, l_n)$ where l_i represents the number of times i-th object is selected. Please note that l_i may take the value 0. The condition that l_i's should satisfy is that

$$l_1 + l_2 + \ldots + l_n = r$$

where all the l_i's are non-negative integers. So we have to find the number of $(l_1, l_2, \ldots, l_n)$ which satisy the above equation and $l_i \geq 0$ for all i.
Let us represent $(l_1, l_2, \ldots, l_n)$ as follows
$\underbrace{0, 0, \ldots, 0}_{l_1 \text{ times}} 1 \underbrace{0, 0, \ldots, 0}_{l_2 \text{ times}} 1..........1 \underbrace{0, 0, \ldots, 0}_{l_n \text{ times}}$
Note that if some l_i is 0 then there will be no string of 0's in the corresponding place. Thus if we reflect for a moment we will observe that for any $(l_1, l_2, \ldots, l_n)$ we have a string of ones-and-zeroes consisting of r zero's and $n-1$ ones's. And for any string of ones-and-zeroes consisting of r zero's and $n-1$ ones's we have a n-tuple $(l_1, l_2, \ldots, l_n)$. So the number of n-tuples equals the number of string of ones-and-zeroes consisting of r zero's and $n-1$ ones's.
How to calculate the number of such strings?
Imagine $n+r-1$ blank spaces; choose r places from it in $\binom{n+r-1}{r}$ ways and fill them with 0's and fill the rest with 1's. And you are done.

So the answer is $\binom{n+r-1}{r}$. □

Example 1

`Suppose there are 50 watches in a shop but 4 of them are defective. If I buy 3 watches at random what's the probability that all of them are defective?`

All the sample points looks like a 3 tuple (s_1, s_2, s_3) where s_i belong to $\{1, 2, 3, \ldots, 50\}$ and the s_i's are different.
Total number of sample points $= 50 \times 49 \times 48$.
Total number of sample points where all watches are defective $= 4 \times 3 \times 2$.
P(all watches are defective) $= \frac{4*3*2}{50*49*48}$.

Example 2

There are m mailboxes and n people. Each person comes and chooses one mailbox at random and puts his mail in it. What's the probability that the first mailbox has exactly k letters.

Total no. of sample points $= m^n$.
No of ways first mailbox gets exactly k letters $= \binom{n}{k} \times (m-1)^{(n-k)}$, because we choose k people from n people in $\binom{n}{k}$ ways for the first mailbox. Now we have $n-k$ people left who can put their mails in any of the remaining $m-1$ mailboxes. That they can do in $(m-1)^{(n-k)}$ ways.
Hence probability that the first mailbox has exactly k letters $= \frac{\binom{n}{k} \times (m-1)^{(n-k)}}{m^n}$

Example 3

`A company hired 20 people. They will be placed in 4 cities; 6 in 1st, 4 in 2nd, 5 in 3rd & 5 in 4th.If among the 20 there are 4 friends, what's the probability they will be together. People are placed in the cities randomly.`

Total No. of sample points = Number of ways of partitioning 20 people in groups of $6, 4, 5, 5$

$$= \frac{20!}{6! \times 4! \times 5! \times 5!}$$

Number of ways 4 friends remain together in City 1
= Number of ways of dividing $20 - 4$ people into groups of $6 - 4, 4, 5, 5$

$$= \frac{16!}{2! \times 4! \times 5! \times 5!}$$

Number of ways 4 friends remain together in City 2
= Number of ways of dividing $20 - 4$ people into groups of $6, 4 - 4, 5, 5$

$$= \frac{16!}{6! \times 0! \times 5! \times 5!}$$

By the way, $0! = 1$.

Number of ways 4 friends remain together in City 3
= Number of ways of dividing $20 - 4$ people into groups of $6, 4, 5 - 4, 5$

$$= \frac{16!}{6! \times 4! \times 1! \times 5!}$$

Number of ways 4 friends remain together in City 4
= Number of ways of dividing $20 - 4$ people into groups of $6, 4, 5, 5 - 4$

$$= \frac{16!}{6! \times 4! \times 5! \times 1!}$$

Hence the required probability

$$= \frac{\frac{16!}{2! \times 4! \times 5! \times 5!} + \frac{16!}{6! \times 0! \times 5! \times 5!} + \frac{16!}{6! \times 4! \times 1! \times 5!} + \frac{16!}{6! \times 4! \times 5! \times 1!}}{\frac{20!}{6! \times 4! \times 5! \times 5!}}$$

`Homework :: From the book try 2.38,2.42,2.45,2.47,2.48,2.53, 2.54,2.56,2.57,2.69`

Lecture 5

Agenda

1. Conditional Probability

2. Theorem of Total probability

3. Bayes Rule

Conditional Probability

Suppose you are going to have a surgery, but before that you want to talk to a Doctor about how likely it is that the surgery will be a success. The Doctor informs you that about 60% cases have success and you find this number too low and depressing. But if the Doctor looked at the proportion of success among young patients he would have observed that 85% of young patients have success (and that number is not so bad!).
So often, while calculating probabilty of an event (e.g. probability of success of operation) if you already have any piece of information(e.g. the patient is young), it's a good idea to use it.
But how to use that extra information ?

Definition 1. *If A and B are two events in an experiment with sample space S and suppose $P(B) > 0$, (i.e. B is not an impossible event) then the probability of happening A given that B has happened is defined as*

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

$P(A|B)$ is read as "Probability of A given B"

Example

Toss a fair die. Let A denote the event that the outcome is 2, 4 or 6. If someone asks you to play the following game will you play ? "If A occurs you pay me \$10, otherwise I will pay you \$10"

Answer $S = \{1, 2, 3, 4, 5, 6\}$. All outcomes are equally likely, as it is a fair die. $A = \{2, 4, 6\}$

$P(A) = \frac{3}{6} = \frac{1}{2}$
So it seems like a fair game and \$10 is not a big amount, so I will bet.

Now suppose the die is cast in a secret chamber and you have a helpful informer who tells you that the results were 4,5 or 6. Would you still play if you had the option to move out?
So let $B = \{4, 5, 6\}$, and we know B has occured. So

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{4, 6\})}{P(\{4, 5, 6\})} = \frac{2}{3}$$

So the game is fair no more and I wouldn't play.

Conditional probability satisfies the 3 axioms of probability.

Lemma 1. *Let B be an event with $P(B) > 0$. Then*

1. *For any event A, $0 \leq P(A|B) \leq 1$.*
2. *If S is the sample space, $P(S|B) = 1$*
3. *If $A_1, A_2, A_3, \ldots$ are disjoint events then*

$$P(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$$

Proof.

1. $A \cap B \subset B$. Hence $0 \leq P(A \cap B) \leq P(B)$.
   Lets divide all of them by $P(B)$ to get,

$$0 \leq \frac{P(A \cap B)}{P(B)} \leq 1$$

2. $$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

3. If $A_1, A_2, A_3, \ldots$ are disjoint events then so are $A_1 \cap B, A_2 \cap B, A_3 \cap B, \ldots$. Hence by the 3rd axiom of probability

$$P(\cup_{i=1}^{\infty}(A_i \cap B)) = \sum_{i=1}^{\infty} P(A_i \cap B)$$

But $\cup_{i=1}^{\infty}(A_i \cap B) = (\cup_{i=1}^{\infty} A_i) \cap B$. Hence

$$P((\cup_{i=1}^{\infty} A_i) \cap B) = \sum_{i=1}^{\infty} P(A_i \cap B)$$

We divide both sides by $P(B)$ to get,

$$P(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$$

$\square$

`Question :: If P(A) = 0, does that mean P(A|B) = 0 ?`
`If P(A|B) = 0, does that mean P(A) = 0 ?`
`For any event A, P(A|S) = P(A) TRUE or FALSE ?`
We know, that a sequence of events $B_1, B_2, \ldots, B_k$ is called mutually exclusive (also called disjoint), if for any $i \neq j$, $B_i \cap B_j = \emptyset$.
A sequence of events $B_1, B_2, \ldots, B_k$ is called **mutually exhaustive** if

$$B_1 \cup B_2 \cup \ldots \cup B_k = S$$

. If a sequence of events $B_1, B_2, \ldots, B_k$ is mutually exclusive and mutually exhaustive, we call it a **partition of the sample space**. It is as if we are dividing the sample space into several disjoint pieces.

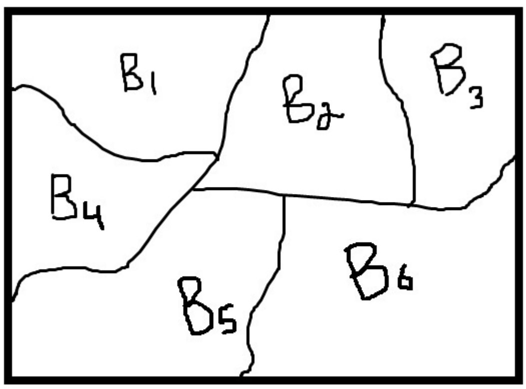


Figure 1: This sample space has been partition into $B_1, B_2, B_3, B_4, B_5, B_6$

Theorem of Total probability

Theorem 1. *If $B_1, B_2, \ldots, B_k$ is a partition of the sample space S, such that $P(B_i) > 0$ for all i; then for any event A*

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

Proof. Since $B_1, B_2, \ldots, B_k$ are mutually exhaustive,

$$S = B_1 \cup B_2 \cup \ldots \cup B_k$$

Hence

$$\begin{aligned} A \cap S &= A \cap (B_1 \cup B_2 \cup \ldots \cup B_k) \\ \Rightarrow A &= (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k) \text{ [By Distributive Law]} \end{aligned}$$

Now $B_1, B_2, \ldots, B_k$ are mutually exclusive, hence so are $A \cap B_1,\ A \cap B_2,\ \ldots,\ A \cap B_k$. Thus

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k)$$

Now for each $A \cap B_i$ we can write $P(A|B_i) = \frac{P(A \cap B_i)}{P(B_i)}$ which implies

$$P(A \cap B_i) = P(A|B_i)P(B_i)$$

. Thus

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_k)P(B_k)$$

□

Bayes Rule

Theorem 2. *If $B_1, B_2, \ldots, B_k$ is a partition of the sample space S, then for any event A*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

Proof.

$$\begin{aligned}
P(B_i|A) &= \frac{P(B_i \cap A)}{P(A)} \\
&= \frac{P(A \cap B_i)}{P(A)} \\
&= \frac{P(A|B_i)P(B_i)}{P(A)} \\
&= \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} \text{ [Apply Theorem of Total Probability]}
\end{aligned}$$

□

`Homework ::  2.124,2.125,2.129,2.134,2.135,2.137`

Lecture 6

Agenda

1. Independence

Independence

In our everyday lives when we see two things which are not influenced by one another, we call them independent of each other. For example how many sandwitches you will eat today and whether the Gators are gonna win this season. One thing doesn't influence the other.So we call them independent. If we want to translate this concept in terms of probability it should look something like this.

Definition 1. *Two events A and B are said to be independent if*

$$P(A|B) = P(A)$$

.

Properties

If A and B are independent then

- $P(A \cap B) = P(A)P(B)$
- $P(B|A) = P(B)$
- A and $\overline{B}$ are independent.

Proof. 1.

$$\begin{aligned} P(A|B) &= P(A) \\ \Rightarrow \frac{P(A \cap B)}{P(B)} &= P(A) \\ \Rightarrow P(A \cap B) &= P(A)P(B) \end{aligned}$$

2.

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

3.

$$\begin{aligned} P(A|\bar{B}) &= \frac{P(A\cap\bar{B})}{P(\bar{B})} \\ &= \frac{P(A) - P(A\cap B)}{P(\bar{B})} \\ &= \frac{P(A) - P(A)P(B)}{P(\bar{B})} \\ &= \frac{P(A)(1 - P(B))}{P(\bar{B})} \\ &= \frac{P(A)P(\bar{B})}{P(\bar{B})} \\ &= P(A) \end{aligned}$$

Hence A and $\bar{B}$ are independent.

□

Please note that $P(A \cap B) = P(A)P(B)$ can be used as an alternate definition of independence.

Example

Suppose I draw a card out of a standard deck of 52. A denotes the event that the card is a diamond and B denotes the event that the it's a face card. Then A and B are independent.

Proof. $P(A) = 13/52 = 1/4$
$P(B) = 12/52 = 3/13$
$P(A \cap B) = 3/52 = (3/13) * (1/4) = P(A)P(B)$ Hence A and B are independent. □

A note

Independence is not same as disjoint, infact quite different.
If A and B are two independent events which are not impossible events i.e. $P(A) > 0$ and $P(B) > 0$, then they can't be disjoint. $P(A \cap B) = P(A) \times P(B) > 0$.Hence not disjoint.

Independence of a 3 or more events

Let $\{A_1, A_2, \ldots, A_n\}$ be a collection of n events. We call this collection of events independent if for any subcollection $\{A_{i_1}, A_{i_2}, \ldots, A_{i_m}\}$,$2 \leq m \leq n$

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}) = P(A_{i_1}) \times P(A_{i_2}) \times \ldots \times P(A_{i_m})$$

`Notice that for independence of` n `events we have to check many conditions, can you calculate exactly how many ?`

Example

Suppose I toss a fair coin independently twice. Let's define the following events
A_1 = we get heads in the first toss
A_2 = we get heads in the second toss
A_3 = the two tosses give same result

A_1 and A_2 are independent by design.

$$P(A_1) = P(A_2) = \frac{1}{2}$$

$$P(A_3) = P(A_1 \cap A_2) + P(\bar{A}_1 \cap \bar{A}_2) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$$

For A_1 and A_3

$$P(A_1 \cap A_3) = P(A_1 \cap A_2) = \frac{1}{4} = P(A_1) \times P(A_3)$$

Hence A_1 and A_3 are independent.
Similarly we can show A_2 and A_3 are independent.
So A_1, A_2, A_3 are pairwise independent. Does this imply independence ?
NO.

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2) = \frac{1}{4}$$

But

$$P(A_1) \times P(A_2) \times P(A_3) = \frac{1}{8}$$

So A_1, A_2, A_3 are not independent.

Lecture 8

Agenda

1. Discrete Random Variable from previous class completed
2. Probability mass function
3. Probability distribution function

Probability mass function

Let X be a discrete random variable. Hence $Range(X)$ is a countable set and thus it can be written as a list like $\{x_1, x_2, x_3, \ldots\}$. Thus the only values the random variable X can take are $x_1, x_2, x_3, \ldots$.
Now the next natural question to ask is what is $P(X = x_i)$ for the various values $x_1, x_2, x_3, \ldots$?

Definition 1. *The* ***probability mass function*** *of a discrete random variable X is given by*

$$p_X(x) = P(X = x)\ \forall x \in Range(X)$$

Note that

$$p_X(x) \geq 0 \qquad \text{and} \qquad \sum_{x \in Range(X)} p_X(x) = 1$$

Example

Experiment :: Toss a fair coin 3 times
Sample Space :: $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$
Random variable X is the number of tosses.
Thus $X : S \to \mathbb{R}$ looks like this

$$\begin{aligned} X(HHH) &= 3 \\ X(HHT) &= X(HTH) = X(THH) = 2 \\ X(HTT) &= X(THT) = X(TTH) = 1 \\ X(TTT) &= 0 \end{aligned}$$

Thus, $Range(X) = \{0, 1, 2, 3\}$ and

$$P(X = 0) = \frac{1}{8}, P(X = 1) = \frac{3}{8}, P(X = 2) = \frac{3}{8}, P(X = 3) = \frac{1}{8}$$

Hence the probability mass function is given by

$$p_X(0) = \frac{1}{8}, p_X(1) = \frac{3}{8}, p_X(2) = \frac{3}{8}, p_X(3) = \frac{1}{8}$$

Example

Experiment :: Suppose I toss a fair die twice independently.
Random Variable :: X = Minimum of the two outcomes

Sample space $S = \{(i, j) | 1 \leq i, j \leq 6\}$.
Please note that the sample space has 36 elements and hence element has probability $\frac{1}{36}$.

The values X can take are $\{1, 2, 3, 4, 5, 6\}$.
$X = i$ corresponds to the event $\{(i, i), (i, i + 1), (i + 1, i), \ldots, (i, 6), (6, i)\}$. Hence

$$P(X = i) = \frac{1 + (6 - i) * 2}{36} = \frac{13 - 2i}{36}$$

Thus we can write the pmf as

$$p_X(x) = \frac{13 - 2x}{36} \text{ for } x \in \{1, 2, \ldots, 6\}$$

Example

Experiment :: There are N UF students, from them I collect a sample of n students. Random Variable :: The average height of n students.

If $X_1, X_2, \ldots, X_n$ are the n heights, let $\bar{X}_n$ denote the average height. The first question is why we are bothering with $\bar{X}_n$ since clearly height of students and hence average height of students should be a continuous variable. But here the population consists of N students and the total number of sample points is $\binom{N}{n}$. Hence $\bar{X}_n$ can take take atmost $\binom{N}{n}$ many different values; the number can be even smaller since two different samples may give the same value of $\bar{X}_n$. Each sample point has probability $\frac{1}{\binom{N}{n}}$ and hence we can calculate the probabilities corresponding to the different values of $\bar{X}_n$ and thus get it's pmf.

Let's look at an example.

$$N = 10 \; n = 3$$

The 10 heights are

4.956	5.1672	5.2746	5.4631	6.2162
6.2486	6.3499	6.552	6.5782	6.8448

Now $\binom{10}{3} = 120$, so lets list down all possible values of $\bar{X}_n$.

5.1326	5.1954	5.2312	5.3016	5.4465
5.4573	5.4823	5.491	5.4931	5.5268
5.5451	5.5527	5.5559	5.5584	5.5635
5.5671	5.5897	5.5942	5.5972	5.6029
5.6155	5.6263	5.6513	5.656	5.6571
5.6601	5.6621	5.6646	5.6658	5.6733
5.6918	5.6959	5.7275	5.7362	5.7546
5.7622	5.7632	5.7719	5.8069	5.825
5.8407	5.8515	5.8608	5.8773	5.9081
5.9111	5.9131	5.9168	5.9189	5.9219
5.9276	5.9469	5.9527	5.9577	5.9614
5.976	5.9785	5.9872	5.9893	5.998
6.0057	6.0097	6.0143	6.0165	6.0205
6.023	6.0231	6.0251	6.0287	6.0318
6.0338	6.0502	6.0588	6.0675	6.0761
6.0771	6.0858	6.0869	6.0879	6.0966
6.0991	6.1119	6.1176	6.1206	6.1217
6.1227	6.1263	6.1304	6.1349	6.1564
6.1747	6.1855	6.188	6.1967	6.1978
6.2193	6.2238	6.2325	6.2716	6.2867
6.2954	6.339	6.3477	6.3727	6.3814
6.3835	6.3922	6.4365	6.4488	6.4596
6.4703	6.4811	6.4934	6.5377	6.5464
6.5485	6.5572	6.5823	6.591	6.6583

So we have successively computed the pmf of $\bar{X}_n$, for every x in the above list, $p_{\bar{X}_n}(x) = \frac{1}{120}$.

But if someone tells me $P(\bar{X}_n = 6.2325) = \frac{1}{120}$ how much does that mean to us ? How much idea does that give us about the distribution of $\bar{X}_n$? **Not much !**

So if some random variable X takes many values within some given range, it is not very helpful if somebody gives a long list of the values that X takes along with the probabilities.

But if someone tells me $P(5.5 < \bar{X}_n \leq 6) = \frac{51}{120} = 0.425$ that gives us some idea about how $\bar{X}_n$ is distributed. So for any random variable X we define it's probability disttribution function.

Probability distribution function

Definition 2. *For any random variable X, we define it's probability distribution function as*

$$F(x) = P(X \leq x) \textit{ for all } x \in \mathbb{R}$$

Example

For example consider the random variable X in Example 1.

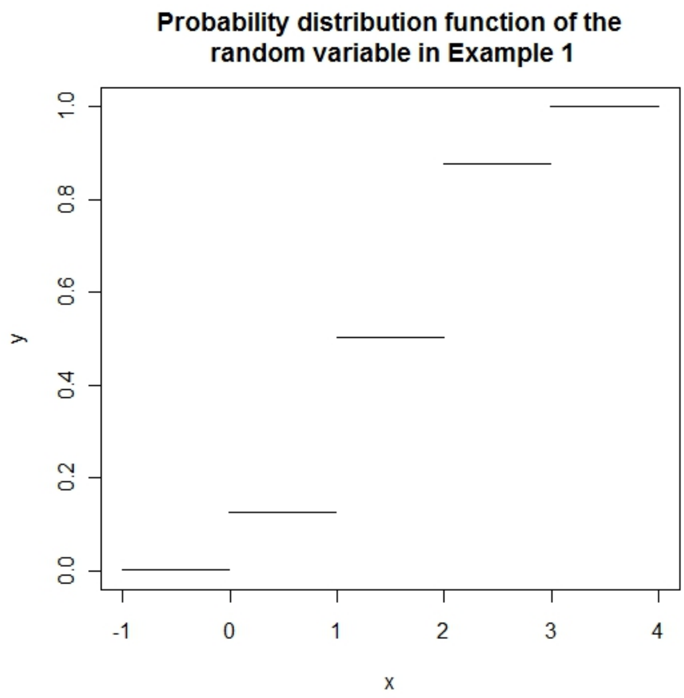
$$P(X=0) = \frac{1}{8}, P(X=1) = \frac{3}{8}, P(X=2) = \frac{3}{8}, P(X=3) = \frac{1}{8}$$

Hence for any $x \in \mathbb{R}$,

$$F(x) = P(X \leq x) = \sum_{y \leq x} p_X(y)$$

We can calculate that

$$\begin{aligned} F(x) &= 0 \text{ if } x < 0 \\ &= \frac{1}{8} \text{ if } 0 \leq x < 1 \\ &= \frac{4}{8} \text{ if } 1 \leq x < 2 \\ &= \frac{7}{8} \text{ if } 2 \leq x < 3 \\ &= 1 \text{ if } 3 \leq x \end{aligned}$$



Homework :: 3.1,3.2,3.5,3.6,3.10

Lecture 9

Agenda

1. Properties of distribution function
2. Expectation

Properties of distribution function

Let X be a random variable. Then it's probability distribution function $F_X(b)$ is defined as

$$F_X(b) = P(X \leq b)$$

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

Properties

1. $$\lim_{x \to -\infty} F_X(x) = 0$$

2. $$\lim_{x \to \infty} F_X(x) = 1$$

3. $F_X()$ is an non-decreasing function.

4. $F()$ is right continuous, i.e.

$$\lim_{h \to 0^+} F_X(b+h) = F_X(b)$$

Let $\{A_n : n \geq 1\}$ be a sequence of events, and A be an event. We say that $A_n \to A$ if either
(i) $A_1 \subset A_2 \subset A_3 \subset \ldots$ i.e. A_n is an increasing sequence and $\cup_{i=1}^{\infty} A_i = A$
or
(ii) $A_1 \supset A_2 \supset A_3 \supset \ldots$ i.e. A_n is an decreasing sequence and $\cap_{i=1}^{\infty} A_i = A$

Please note that in your calculus course, for a sequence of real numbers $\{x_n : n \geq 1\}$ when you defined $x_n \to x$, there was no condition like $\{x_1 \leq x_2 \leq x_3 \ldots\}$ or $\{x_1 \geq x_2 \geq x_3 \ldots\}$

Theorem 1. *If $A_n \to A$ then $P(A_n) \to P(A)$.*

Proof of properties

1. Let $x_n \downarrow -\infty$, i.e. $x_n \to -\infty$ and $\{x_n : n \geq 1\}$, is a decreasing sequence. We have to show $F_X(x_n) \to 0$.
   Define, A_n as the event, $A_n = \{X \leq x_n\}$. Since $\{x_n\}$ is decreasing sequence, $A_1 \supset A_2 \supset A_3 \supset \ldots$, i.e. $\{A_n : n \geq 1\}$ is also a decreasing sequence of sets. Secondly since $x_n \downarrow -\infty$, we can say $\cap_{i=1}^{\infty} A_i = \emptyset$. Thus $A_n \to \emptyset$, which gives $P(A_n) \to P(\emptyset)$, i.e. $F_X(x_n) \to 0$.

2. Proof similar to Property 1.

3. Let $a < b$, we have to show $F_X(a) \leq F_X(b)$. Let's define the events, $A = \{X \leq a\}$ and $B = \{X \leq b\}$. Note that $A \subset B$, hence

$$\begin{aligned} P(A) &\leq P(B) \\ \Rightarrow P(X \leq a) &\leq P(X \leq b) \\ \Rightarrow F_X(a) &\leq F_X(b) \end{aligned}$$

   Please note that we said **non-decreasing** and didn't say increasing.

4. Let b be any real number. We have to show, $F_X()$ is right-continuous at b. Let the sequence $\{b_n : n \geq 1\}$ decrease to b, written as $b_n \downarrow b$. That is, $\{b_1 > b_2 > b_3 > \ldots\}$ and $b_n \to b$. Then we have to show,

$$F_X(b_n) \to F_X(b)$$

   Define A_n as the event $\{X \leq b_n\}$, and A as the event $\{X \leq b\}$. Since $b_n \downarrow b$, we can argue $A_n \to A$, hence $P(A_n) \to P(A)$ i.e. $F_X(b_n) \to F_X(b)$.

However it is not true that $F_X()$ has to be left continuous. Actually what happens is, for $b \in \mathbb{R}$,

(i) If $P(X = b) = 0$ then $F_X()$ is both left and right continuous at b.

(ii) If $P(X = b) > 0$ then $F_X()$ is only right continuous at b and not left continuous.

In the last class we drew the distribution funtion of some random variables, if you look at those graphs the above fact will be more clear.

Another Fact

Any function satisfying Properties 1,2,3,4 is a distribution function of some random variable.

Expectation

A random variable X takes many values, with different probabilities. Sometimes we would like to represent those values with a single number.Thus we define the expectation of the random variable, which is a representative of those values.

Definition 1. *Let X be a discrete random variable. Then we define it's expectation as*

$$E(X) = \sum_{x \in Range(X)} x * p_X(x)$$

. But we only define it only when $\sum_{x \in Range(X)} |x| * p_X(x) < \infty$

Thus if $RangeX = \{x_1, x_2, x_3, \ldots\}$ and $p_X()$ represents it's pmf, then we define

$$E(X) = x_1 * p_X(x_1) + x_2 * p_X(x_2) + x_3 * p_X(x_3) + \ldots$$

But we define it only when

$$(|x_1| * p_X(x_1) + |x_2| * p_X(x_2) + |x_3| * p_X(x_3) + \ldots) < \infty$$

Example

Consider the following game. We toss a fair die twice. If the sum of the two values is less than or equal to 3, we have to pay \$10. If the sum is 4,5 or 6 we pay \$4. If the sum is 7,8 or 9 we earn \$4. If the sum is 10,11 or 12 we earn \$10. What's our expected winning ?

$$P(X \leq 3) = \frac{3}{36}$$

$$P(X = 4, 5, 6) = \frac{12}{36}$$

$$P(X = 7, 8, 9) = \frac{15}{36}$$

$$P(X = 10, 11, 12) = \frac{6}{36}$$

Let W be the winnings.

$$P(W = -10) = \frac{1}{12}, P(W = -4) = \frac{1}{3}, P(W = 4) = \frac{5}{12}, P(W = 10) = \frac{1}{6}$$

$$E(W) = -10 \times \frac{1}{12} - 4 \times \frac{1}{3} + 4 \times \frac{5}{12} + 10 \times \frac{1}{6} = \frac{7}{6}$$

Theorem 2. *Let X be a discrete random variable with pmf $p_X()$, and $g : \mathbb{R} \to \mathbb{R}$ is any function. Then $g(X)$ is also another discrete random variable. If it's expectation exists then, that can be calculated the usual way, i.e. using the definition or using the following formula*

$$E(g(X)) = \sum_{x \in Range(X)} g(x) p_X(x)$$

`Homework :: No Homework.`

Lecture 10

Agenda

1. Examples involving expectation of a random variable

2. Properties of Expectation

Examples involving expectation of a random variable

Example 1

Roll a fair die twice. Let X denote the sum of the two outcomes. Let's find the expectation of X.

Sample space ,$S = \{(i,j) | 1 \leq i,j \leq 6\}$
Total number of sample points = 36.
$Range(X) = \{2, 3, 4, \ldots, 12\}$
Number of sample points corresponding to the event $\{X = x\}$
= Number of points in the set $\{i | 1 \leq i \leq 6 \text{ and } 1 \leq x - i \leq 6\}$
= Number of points in the set $\{i | max\{1, x-6\} \leq i \leq min\{6, x-1\}\}$.

Hence

$$P(X = x) = \frac{min\{6, x-1\} - max\{1, x-6\} + 1}{36}$$

So,

$$E(X) = \sum_{x=2}^{12} x * P(X = x) = 7$$

Example 2

Consider a coin with probability of heads = p, and probability of tails = $1 - p$.
I will keep tossing the coin until I get an head. Let X be the number of tosses needed before my first head.
$Range(X) = \{0, 1, 2, 3, \ldots\}$
$P(X = x) = P($we will have tails for the first x tosses and heads in the $x + 1$-th toss$)$
$= (1-p)^x \times p$
Hence, $E(X) = \sum_{x=0}^{\infty} x \times (1-p)^x \times p = \frac{1-p}{p}$
Proof of the last equality is done at the end.

Properties of Expectation

Property 1

Let X and Y be two discrete random variables, such that both $E(X)$ and $E(Y)$ can be defined. Then $Z = X + Y$ is also another discrete random variable and $E(Z)$ can be defined and

$$E(Z) = E(X) + E(Y)$$

Example 1 becomes much easier with Property 1.

Property 2

Let X be a discrete random variable with pmf $p_X()$, and $g : \mathbb{R} \to \mathbb{R}$ is any function. Then $g(X)$ is also another discrete random variable. If it's expectation exists then, that can be calculated using the following formula

$$E(g(X)) = \sum_{x \in Range(X)} g(x) p_X(x)$$

Property 3

If X is a discrete random variable, such that $E(X)$ can be defined and a and b are two constants; $Y = aX + b$ is also a discrete random variable and

$$E(Y) = a * E(X) + b$$

Property 3 can be proved easily from Property 2.

Example for property 2

Let the probability distribution of X be

X	-2	-1	0	1	2
$P(X = x)$	0.25	0.1	0.2	0.2	0.25

Define $Y = |X|$, and suppose we have to calculate $E(Y)$.
Then doing it the traditional way,

$$Range(Y) = \{0, 1, 2\}$$

$$P(Y = 0) = P(X = 0) = 0.2$$

$$P(Y = 1) = P(X = 1) + P(X = -1) = 0.2 + 0.1 = 0.3$$

$$P(Y=2) = P(X=2) + P(X=-2) = 0.25 + 0.25 = 0.5$$

Hence

$$E(Y) = 0 \times 0.2 + 1 \times 0.3 + 2 \times 0.5 = 1.3$$

But if we want to use property 2;

$$\begin{aligned} E(Y) = E(|X|) &= \sum_{x \in Range(X)} |x| \times P(X=x) \\ &= 2 * 0.25 + 1 * 0.1 + 0 * 0.2 + 1 * 0.2 + 2 * 0.25 \\ &= 1.3 \end{aligned}$$

Example 2 proof

Let $S = x + 2 * x^2 + 3 * x^3 + 4 * x^4 + \ldots$, where $0 < x < 1$.

$$\begin{aligned} S * (1-x) &= S - S * x \\ &= \{x + 2 * x^2 + 3 * x^3 + 4 * x^4 + \ldots\} - \{x^2 + 2 * x^3 + 3 * x^4 + 4 * x^5 + \ldots\} \\ &= x + (2 * x^2 - x^2) + (3 * x^3 - 2 * x^3) + (4 * x^4 - 3 * x^4) + \ldots \\ &= x + x^2 + x^3 + x^4 + \ldots \\ &= \frac{x}{1-x} \end{aligned}$$

Put $x = 1 - p$ and we get the result.

`Homework :: 3.12,3.20,3.21,3.23,3.25,3.29`

Lecture 11

Agenda

1. Variance and Standard Deviation
2. Indicator Function
3. Markov Inequality and Chebyshev's inequality

Variance and Standard Deviation

Consider two students and their scores on 4 exams.

Tom	49	51	48	52
Harry	20	80	30	70

Both have mean score 50. Speaking that way the performance of both should be same. But Tom is much more consistent in his performance than Harry. If we want to measure how much the values are spread around the average, we have to define variance.

Theorem 1. *If for any random variable X, $E(X^2) < \infty$ then $E(|X|) < \infty$ that is $E(X)$ can be defined without any doubt.*

Please recall, we defined $E(X)$, only when $E(|X|) < \infty$. The above result says if you can define the 2nd moment $E(X^2)$, then you will have no trouble with the first moment $E(X)$.

Definition 1. *For any random variable X if $E(X^2) < \infty$, then we define it's **variance** as*

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

where $\mu = E(X)$.

Note that

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

Please note that, $V(X) = E(X^2) - E(X)^2$ can also be used as a definition of variance.

Definition 2. *The* ***standard deviation*** *of a random variable X is,*

$$sd(X) = \sqrt{V(X)} = \sqrt{E(X-\mu)^2}$$

Sometimes *standard deviation* is also called the *standard error*; they are exactly the same thing.

Example

Consider the random variable X with the following distribution,

$$P(X=1) = 0.25, P(X=2) = 0.3, P(X=3) = 0.2, P(X=4) = 0.25$$

$$E(X) = 1*0.25 + 2*0.3 + 3*0.2 + 4*0.25 = 2.45$$
$$E(X^2) = 1^2*0.25 + 2^2*0.3 + 3^2*0.2 + 4^2*0.25 = 7.25$$
$$V(X) = 7.25 - 2.45^2 = 1.2475$$
$$sd(X) = \sqrt{V(X)} = \sqrt{1.2475} = 1.1169$$

Property

For any random variable X,

$$V(aX+b) = a^2V(X)$$

Proof.

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

$\square$

Indicator Function

Definition 3. *Let A be any event, we define it's indicator function as, the random variable*

$$\begin{aligned} 1_A &= 1 \textit{ if } A \textit{ happens} \\ &= 0 \textit{ if } \bar{A} \textit{ happens i.e. } A \textit{ does not happen} \end{aligned}$$

You may be thinking why we used the word *function* for 1_A, when it's clearly a random variable. We can define, for any $s \in S =$ the sample space,

$$\begin{aligned} 1_A(s) &= 1 \text{ if } s \in A \\ &= 0 \text{ if } s \in \bar{A} \end{aligned}$$

The above clearly looks like a function from S to $\mathbb{R}$, hence the name indicator function.

Property 1

Lemma 1. *For any events A and B,*

$$1_{A \cap B} = 1_A \times 1_B$$

Property 2

Lemma 2. *For any events A and B, if $A \cap B = \emptyset$*

$$1_{A \cup B} = 1_A + 1_B$$

Property 3

Lemma 3. *For any event A,*

$$1_A + 1_{\bar{A}} = 1$$

Property 4

Lemma 4. *For any event A,*

$$E(1_A) = P(A)$$

The proof of these properties were done in class and they are simple and direct; but not included in the notes.

Markov's Inequality

Theorem 2. *Let X be any random variable and $k > 0$ is any positive real number. Then*

$$P(|X| \geq k) \leq \frac{E(|X|)}{k}$$

Proof.

$$|X| \geq |X| 1_{\{|X| \geq k\}} \geq k 1_{\{|X| \geq k\}}$$

Hence

$$1_{\{|X| \geq k\}} \leq \frac{|X|}{k}$$

Taking expectation on both sides we get

$$P(|X| \geq k) \leq \frac{E(|X|)}{k}$$

□

Chebyshev's Inequality

Theorem 3. *Let X be a random variable with mean μ and standard deviation σ; then for any $k > 0$,*

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

Hence,

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

Proof.

$$P(|X - \mu| \geq k\sigma) = P(\ \ (X - \mu)^2 \geq k^2\sigma^2\ \ )$$

Define $Y = (X - \mu)^2$, then

$$E(|Y|) = E(Y) = E(X - \mu)^2 = \sigma^2$$

Hence by Chebyshev's inequality,

$$\begin{aligned} P(|Y| \geq k^2\sigma^2) &\leq \frac{E|Y|}{k^2\sigma^2} \\ &= \frac{\sigma^2}{k^2\sigma^2} \\ &= \frac{1}{k^2} \end{aligned}$$

Thus,

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

$\square$

Homework ::

1. In lecture 10, you had to calculate the expectation of several random variables; calculate their variances.

2. In this lecture we have used the property $E(X+Y) = E(X)+E(Y)$, which we stated in lecture 10. If you are curious about how to prove this then contact me during office hour.

3. For any X and Y, $E(X+Y) = E(X) + E(Y)$. But this does not happen for variance, i.e. $V(X+Y) = V(X) + V(Y)$ may not be true. Think of a quick example to illustrate that.

4. The Markov's Inequality or the Chebyshev's Inequality can become equality ----- when ?

5. If $A_1, A_2, \ldots A_k$ are any k events and $A = A_1 \cup A_2 \cup \ldots \cup A_k$, then show that $(1_A - 1_{A_1}) \times (1_A - 1_{A_k}) \times \ldots \times (1_A - 1_{A_k}) = 0$. Does this give you any idea about how to prove the IEP ?