STAB52 Lecture Notes

Joshua Concon

University of Toronto Scarborough – Summer 2017

Pre-reqs are MATA37, which is Calculus for Mathematical Sciences II. Instructor is Dr. Mahinda Samarakoon. I highly recommend sitting at the front since he likes to teach with the board and can have a bit of trouble projecting his voice in large lecture halls. If you find any problems in these notes, feel free to contact me at conconjoshua@gmail.com.

Contents

1   Wednesday, May 3, 2017

1.1   What is probability?

He wanted us to think about what it really was. i.e. What does it mean for something to have a 50% probability? If a coin landing on heads has a 50% probability, it doesn't guarantee that such an event would occur if you flipped a coin twice.

After a few guesses, he gave us a definition.

1.2   Relative Frequency Definition of Probability

> **Relative Frequency**
>
> Consider the case where an experiment is performed n times.
>
> Let $|A|$ be the number of trials resulting in 'event' A.
>
> The **Relative Frequency** of $A = \frac{|A|}{n} = \gamma_n$

This is the Probability of A when n is large γ_n by itself is not an accurate definition of the probability of A occurring, so we take the limit as n goes to infinity and then we have such:

$$\lim_{n \to \infty} \gamma_n$$

This definition is difficult to use in most cases but relatively easy to understand. So here is a definition that is easier to do calculations with:

1.3   Formal Definition of Probability

He doesn't actually get into the definition of probability this lecture, instead he goes through a few terms that we need to define before we get to this definition.

> ### Sample Space (S)
>
> Sample Space (S): the set of all possible outcomes in an experiment.
> Size of S is denoted by $n(S)$, $\#S$ or $|S|$

ex. Tossing a coin once: $S = \{H, T\}$, $n(S) = 2$

ex. Rolling a 6-sided die: $S = \{1, 2, 3, 4, 5, 6\}$, $n(S) = 6$

ex. Tossing 2 different coins: $S = \{HH, HT, TH, TT\}$, $n(S) = 4$

> ### Events
>
> Subsets of a sample space

ex. Experiment rolling a die, $S = \{1, 2, 3, 4, 5, 6\}$
$A = \{1, 2\}$, $A \subseteq S$, so A is an event of S.
$B = \{5, 7\}$, $B \nsubseteq S$, so B is not an event of S.

You can also describe events with words.

$C =$ the result is an odd number $= \{1, 3, 5\}$, $C \subseteq S$, so C is an event of S.

An event can be the entire sample space and the null set.

$D = S \subseteq S$, so D is an event of S.
$E = \varnothing \subseteq S$, so E is an event of S.

Operations on Events (Unless specified, valid for all events A, B)

1. Where $S = \{1, 2, 3, 4, 5, 6\}$, $A = \{1, 2\}$, $B = \{2, 4, 5\}$
   $A \cup B$: elements in A or B
   ex. $A \cup B = \{1, 2, 4, 5\}$

2. Where $S = \{1, 2, 3, 4, 5, 6\}$, $A = \{1, 2\}$, $B = \{2, 4, 5\}$
   $A \cap B$: elements in A and B
   ex. $A \cap B = \{2\}$
   And for $C = \{5, 6\}$
   $A \cap C = \varnothing$

> Recall: So essentially all of the logic laws from CSCA67 hold for these sets as well.
> i.e.
> $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
> $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

3. $(A \cap B)^c = (A^c \cup B^c)$ (where A^c is the complement of A, which is the elements in S that is not in A)

   ex.
   Consider $A = \{1, 2\}$ and $S = \{1, 2, 3, 4, 5, 6\}$
   $A^c = S - A = \{3, 4, 5, 6\}$

4. $(A \cup B)^c = (A^c \cap B^c)$

5. $A \cup \varnothing = A$

6. $A \cap \varnothing = \varnothing$

Event Space

Is a set of all the possible events of S (i.e All the possible subsets of the set S). Denoted P

ex. The Event Space of the set $\{1, 2, 3\}$ is
$\{\{ 1 \},\{ 2 \},\{ 3 \},\{ 1,2 \},\{ 2,3 \},\{ 1,3 \},\{ 1,2,3 \}\}$

2   Friday, May 5, 2017

2.1   Formal Definition of Probability

> **Probability Measure Function (PMF)**
>
> A function $P : S \longmapsto [0,1]$ defined on sample space S is called a **Probability Measure Function**. It satisfies the following axioms:
>
> 1. $0 \geq P(A) \geq 1$
>
> 2. $P(S) = 1$ where S is a sample space
>
> 3. For disjoint events $A_1, ..., A_n$ (disjoint means where $A_i \cap A_k = \varnothing$ $\forall i, k \in \mathbb{N} \; 1 \geq i < k \geq n$)
>   then $P(A_1) \cup A_2 \cup ... \cup A_n = P(A_1) + P(A_2) + ... + P(A_n)$
>
> 4. $P(\varnothing) = 0$

2.2   Properties of Probability Measure Functions

Dr. Mahinda ends up adding **Result 1** to the definition of the Probability Measure Function (4) to stay consistent with the book, but also adds it here as well. It will be left again as a property to stay consistent with the lecture and the numbering.

Result 1: $P(\varnothing) = 0$

Result 2:   If $A_1, ..., A_n$ is a finite collection of disjoint events, then

$$P(\bigcup_{i=1}^{n} A_i) = P(A_1 \cup A_2 \cup ... \cup A_n)$$

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

Proof. (Proof of Result 2)

Define $A_i = \varnothing$, $i \geq n + 1$

So since there is a finite amount of events, A_i where $i \geq n + 1$ are just the null set

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$$

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

By Axiom 3 of the PMF

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

$$= \sum_{i=n+1}^{\infty} P(\varnothing) + \sum_{i=1}^{n} P(A_i)$$

$$= 0 + \sum_{i=1}^{n} P(A_i)$$

By Axiom 4 of the PMF

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

■

Result 3:   If $A_1, A_2, A_3, \ldots$ is a partition of S and B is any event then

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

Partitions

A parition is a collection of events satisfying the following axioms:

1. $A_1, A_2, A_3, \ldots$ are disjoint events

2. $\displaystyle\bigcup_{i=1}^{\infty} A_i = S$

Proof. (Proof of Result 3)

For an arbitrary event B, Consider the partition $A_1, A_2, A_3, ...$ of S
The events $B \cap A_1, B \cap A_2, B \cap A_3, ...$ are, by the definition of partitions...

1. disjoint

2. $\bigcup\limits_{i=1}^{\infty} B \cap A_i = B \cap S = B$

Therefore

$$P(B) = P(\bigcup\limits_{i=1}^{\infty} B \cap A_i) = \sum\limits_{i=1}^{\infty} P(A_i \cap B)$$

By Axiom 3 of the PMF ∎

Set Difference

For arbitrary sets A, B
$A \setminus B$ or $A \cap B^c$ refers to everything in A but not in B.

Result 4: $P(A \setminus B) = P(A) - P(A \cap B)$

Proof. (Proof of Result 4)

Note that $A \setminus B$ and $A \cap B$ are disjoint
$P((A \setminus B) \cup (A \cap B)) = P(A \setminus B) + P(A \cap B)$
Since $A \setminus B$ is disjoint from $A \cap B$
So for arbitrary events A, B
$P(A) = P(A \setminus B) + P(A \cap B) \implies P(A \setminus B) = P(A) - P(A \cap B)$ ∎

Note: if $B \subseteq A$ then $P(A \setminus B) = P(A) - P(B)$

Proof. (Proof of Note)

if $B \subseteq A$ then $A \cap B = B$
therefore $P(A \setminus B) = P(A) - P(A \cap B) = P(A) - P(B)$ ∎

Result 5:   If $A \subseteq B$ then $P(A) \leq P(B)$

Proof. (Proof of Result 5)

$$A \subseteq B \Longrightarrow P(A \setminus B)$$
$$= P(B) - P(A) \geq 0$$

By Axiom 1 of the PMF

$$\Longrightarrow P(A) \leq P(B)$$

∎

Result 6: $P(A^c) = 1 - P(A)$

Proof. (Proof of Result 6)
Consider an arbitrary event $A \subseteq S$
Since we know that if $A \subseteq S$ then $P(A \setminus B) = P(B) - P(A)$
and that $A^c = S \setminus A$

$$P(A^c) = P(S \setminus A) = P(S) - P(A)$$
$$= 1 - P(A)$$

∎

Result 7:   we know $P(A \cup B) = P(A) + P(B)$ if A, B are disjoint and $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any 2 events A, B.

Proof. (Proof for Result 7)
since we know that $A \cup B = A \cup (B \setminus A)$ and that events A and $(B \setminus A)$ are disjoint.

$$P(A \cup B) = P(A \cup (B \setminus A))$$
$$= P(A) + P(B \setminus A)$$
$$= P(A) + P(B) - P(A \cup B)$$

By Result 4

∎

Note: This result can be extended to more than 2 events e.g.

$$P(A{\cup}B{\cup}C) = P(A){+}P(B){+}P(C){-}P(A{\cap}B){-}P(A{\cap}C){-}P(B{\cap}C){+}P(A{\cap}B{\cap}C)$$

Example 1: Consider a bag with 3 red marbles and 7 green marbles inside the bag. What is the probability that you would select a red marble from the bag at random? How about the probability of selecting a green marble?

Solution 1: Simply consider all possible events, and then for each probability x, take all the events where x happens and divide it by the sample space. Let R, G represent the events where a red and green marble respectively is selected.

$$P(R) = \frac{R}{R + G} = 0.3$$

$$P(G) = \frac{G}{R + G} = 0.7$$

Example 2: Consider a class of 33 students, 17 of them aced the midterm, 14 of them aced the final and 11 did not have As on either. How many of them aced both the midterm and the final?

Solution 2: Let A, B refer to the students who have aced the midterm and final respectively. We are trying to solve for $P(A \cap B)$.

$$P(A) = 17/33$$

$$P(B) = 14/33$$

$$P(A \cup B) = 22/33$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$22/33 = 17/33 + 14/33 - P(A \cap B)$$
$$22/33 = 31/33 - P(A \cap B)$$
$$22/33 + P(A \cap B) = 31/33$$
$$P(A \cap B) = 31/33 - 22/33$$
$$P(A \cap B) = 9/33$$

2.3 Uniform Probability Measure on Finite Sample Spaces

> **Uniform Probability Measure on Finite Sample Spaces (UPMFSS)**
>
> If S is a finite sample space (i.e. $|S| < \infty$) of equally likely outcomes, and A is any event, then $P(A) = \frac{|A|}{|S|}$

But is this sample space a probability measure?

Proof. (Proof of UPMFSS being a probability measure)

1. $\underline{0 \le P(A) \le 1 \text{ for an arbitrary event } A}$

$$\varnothing \subseteq A \subseteq S$$
$$|\varnothing| \le |A| \le |S|$$
$$0 \le \frac{|A|}{|S|} \le 1$$

2. $\underline{P(S) = 1}$

$$P(S) = \frac{|S|}{|S|} = 1$$

3. $\underline{P(A_1) \cup A_2 \cup ... \cup A_n = P(A_1) + P(A_2) + ... + P(A_n)}$
   Let $A_1, A_2, ...$ be disjoint events
   Consider $|\bigcup\limits_{i=1}^{\infty} A_i|$

$$|\bigcup_{i=1}^{\infty} A_i| = |A_1| + |A_2| + ...$$

Since the A_i's are disjoint

$$\frac{|\bigcup\limits_{i=1}^{\infty} A_i|}{|S|} = \frac{|A_1|}{|S|} + \frac{|A_2|}{|S|} + ...$$

$$P(|\bigcup_{i=1}^{\infty} A_i|) = P(A_1) + P(A_2) + ...$$

∎

2.4   Counting (Combinatorics)

Example 1: How many 4 digit numbers can be formed from digits $1, 2, ..., 9$ if repeats are not allowed?

Solution 1: $P(9,4) = \frac{9!}{4!} = 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5$ since there are 9 numbers to choose from and there are 4 digits that these numbers can "occupy".

Example 2: Find the probability that the 4-digit number formed from the question above is divisible by 5.

Solution 2: $\frac{P(8,3)}{P(9,4)} = \frac{8 \cdot 7 \cdot 6}{9 \cdot 8 \cdot 7 \cdot 6} = \frac{1}{9}$ since there are $P(9,4)$ combinations in total and the numbers divisible by 5 end with a 5 (since it is not possible for the number in the question above to end in a 0). So there are 3 digits left to be assigned a number with only 8 numbers since 5 is already used, so the numerator is $P(8,3)$.

Counting Subsets: The number of subsets of size k that can be formed from n objects is

$$\frac{n!}{k!(n-k)!} = C(n,k) = \binom{n}{k}$$

3 Wednesday, May 10, 2017

3.1 Counting (continued)

Example 1: A professor will select 5 out of 10 questions for an exam, and a student studies the answers to 6 of these 10 questions. What is the probability that all the questions that the professor will select questions are questions that the student has studied for?

Solution 1: The sample space $|S| = \binom{10}{5}$ and the cases that all the questions that the professor will select questions that the student has studied for is $\binom{6}{5}$ so the probability is:

$$\frac{\binom{6}{5}}{\binom{10}{5}}$$

Example 2: There are 20 members in a group. We want to divide this group into 3 sub-groups (say, A, B, C) of sizes $8, 8, 4$ respectively. In how many ways can we do this?

Solution 2: We choose 8 for A, and then 8 from the remaining 12 for B where the order does not matter and the rest go to C, so we have.

$$\binom{20}{8}\binom{12}{8}\binom{4}{4} = \frac{20!}{8!12!} \cdot \frac{12!}{8!4!} \cdot \frac{4!}{4!} = \frac{12!}{8!8!4!} = \binom{12}{8\ 8\ 4}$$

Example 3: A deck of 52 cards has 12 picture cards. We want to divide this among 4 players (say, A, B, C, D). What is the probability that each player will get exactly 3 picture cards?

Solution 3: The sample space $|S| = \binom{52}{13\ 13\ 13\ 13}$ The cases for distributing all of the 12 picture cards between the 4 players is $\binom{12}{3\ 3\ 3\ 3}$ and the cases for distributing all of the 40 regular cards between the players is $\binom{40}{10\ 10\ 10\ 10}$ so the total probability is:

$$\frac{\binom{12}{3\ 3\ 3\ 3} \cdot \binom{40}{10\ 10\ 10\ 10}}{\binom{52}{13\ 13\ 13\ 13}}$$

3.2　Conditional Probability

Conditional Probability

Let A, B be two events such that $P(B) > 0$, then the conditional probability of A given B is:

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

Proof. (Proof that $P(\cdot|B)$ is a probability measure)
Axiom 1

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

Therefore $P(\cdot|B)$ satisfies axiom 1.

Axiom 2
Let S be the sample space, and let B be any event

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

Therefore $P(\cdot|B)$ satisfies axiom 2.

Axiom 3
Let $A_1, A_2, \dots$ be a collection of disjoint sets

Note that $\forall$ events B that $A_1 \cap B, A_2 \cap B, A_3 \cap B, ...$ are also disjoint.

$$P((A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \cup ...) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + ...$$
$$= P((A_1 \cup A_2 \cup A_3 \cup ...) \cap B)$$
$$\frac{P((A_1 \cup A_2 \cup A_3 \cup ...) \cap B)}{P(B)} = \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} + \frac{P(A_3 \cap B)}{P(B)} + ...$$
$$P(A_1 \cup A_2 \cup A_3 \cup ...|B) = P(A_1|B) + P(A_2|B) + P(A_3|B) + ...$$

Therefore $P(\cdot|B)$ satisfies axiom 3.
Therefore $P(\cdot|B)$ is a probability measure. ■

Example 4: Consider the two events A, B such that
$P(A) = 0.5, P(B) = 0.3, P(AB) = 0.1$. Find...

1. $P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.1}{0.3} = \frac{1}{3}$

2. $P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.1}{0.5} = \frac{1}{5}$

3. $P(A|A \cup B) = \frac{P(A \cap (A \cup B))}{P(A \cup B)} = \frac{P(A)}{P(A) + P(B) - P(AB)} = \frac{0.5}{0.5 + 0.3 - 0.1} = \frac{5}{7}$

4. $P(A|A \cap B) = \frac{P(A \cap A \cap B)}{P(A \cap B)} = \frac{P(A \cap B)}{P(A \cap B)} = 1$

5. $P(AB|A \cup B) = \frac{P(AB \cap (A \cup B))}{P(A \cup B)} = \frac{P(AB)}{P(A \cup B)} = \frac{0.1}{0.5 + 0.3 - 0.1} = \frac{1}{7}$

4   Friday, May 12, 2017

4.1   Conditional Probability (continued)

Example 1: An unbiased coin is flipped twice. What is the probability that both flips result in Heads given that the first flip does?

Solution 1: $S = \{HH, TH, HT, HH\}$

$$P(HH|\{HT, HH\}) = \frac{P(HH \cap \{HT, HH\})}{P(\{HT, HH\})} = \frac{P(HH)}{P(\{HT, HH\})} = \frac{1}{2}$$

> **Theorem: Law of Total Conditional Probability**
>
> Let $A_1, A_2, ...$ be a partition of S such that $P(A_i) > 0 \forall i$.
> Then $\forall$events B, $P(B) = \sum\limits_{i=1}^{\infty} P(B|A_i) \cdot P(A_i)$

Proof. (Proof of the Law of Total Conditional Probability)

$$P(B) = P(B \cap A_1) + P(B \cap A_1) + ...$$

since we know that $P(B|A_i) = \frac{P(B \cap A_i)}{P(A_i)} \iff P(B \cap A_i) = P(B|A_i) \cdot P(A_i)$
so now we can conclude that...

$$P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + ... = \sum\limits_{i=1}^{\infty} P(B|A_i) \cdot P(A_i)$$

∎

> **Theorem: Bayes' Theorem**
>
> If A, B are events such that $P(A), P(B) > 0$
> then $P(A|B) = \frac{P(A)}{P(B)} \cdot P(B|A)$

Proof. (Proof of Bayes' Theorem)

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

∎

Example 2: In a population of voters, 40% are Republican and 60% are Democrats. 30% of Republicans and 50% of Democrats support an election issue. A person is selected at random from this population and he supports the election issue, find the probability that he is a democrat.

Solution 2: Let Su be the probability that the person supports the issue.

$$P(R) = 0.4, P(D) = 0.6, P(Su|R) = 0.3, P(Su|D) = 0.5$$

$$\begin{aligned} P(Su) &= P(Su|R) \cdot P(R) + P(Su|D) \cdot P(D) \\ &= (0.3)(0.4) + (0.5)(0.6) \\ &= 0.42 \end{aligned}$$

$$P(D|Su) = P(Su|D) \cdot \frac{P(D)}{P(Su)} = 0.5 \cdot \frac{0.6}{0.42} = \frac{0.3}{0.42}$$

Definition: Independent Events

Two events A, B are independent $(A \perp B)$ if $P(A \cap B) = P(A) \cdot P(B)$
Weaker Version: If $P(B) > 0$ then $A \perp B$ if $P(A|B) = P(A)$

Example 3: Consider the experiment where you roll a 6-sided die:
Let A be the events where the result is an odd number on the die. Let B be the events where the result is an even number on the die. Let $C = \{1, 2\}$

1. Are A and B independent?
   $P(A) = P(B) = \frac{1}{2}$
   $AB = \varnothing$
   $P(AB) = 0 \neq P(A) \cdot P(B)$
   No.

2. Are A and C independent?
   $AC = \{1\}$
   $P(AC) = \frac{1}{6} = P(A) \cdot P(C)$
   $Yes.$

Result 1: $A \perp B$ then $A \perp B^c$

Proof. (Proof of Result 1)
Assume $A \perp B$.
Consider $P(A \cap B^c)$

$$
\begin{aligned}
P(A \cap B^c) &= P(A \setminus B) \\
&= P(A) - P(AB) \\
&= P(A) - P(A) \cdot P(B) \\
&= P(A) \cdot (1 - P(B)) \\
&= P(A) \cdot P(B^c)
\end{aligned}
$$

Therefore $A \perp B \implies A \perp B^c$ ∎

Definition: General Definition of Independent Events

The events $A_1, A_2, \dots$ are independent if for every subcollection of events $A_i1, A_i2, \dots A_ik$ where a subcollection is a subset of a collection of events.

Example 4: Consider the experiment where a fair 4 sided die is rolled, and
Let $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$
Are A,B,C pairwise independent? Are they independent?
Solution 4: We know $P(A) = P(B) = P(C) = 1/2$
So are A and B independent?

$$
P(AB) = P(\{1\}) = 1/4 = P(A) \cdot P(B)
$$

Yes, so are A and C independent?

$$
P(AC) = P(\{1\}) = 1/4 = P(A) \cdot P(C)
$$

Yes, so are B and C independent?

$$
P(BC) = P(\{1\}) = 1/4 = P(B) \cdot P(C)
$$

Yes, so A, B, C are pairwise independent.
Are A, B, C independent? $P(A \cap B \cap C) = 1 \neq P(A) \cdot P(B) \cdot P(C) = 1/8$
So no, they are not independent.

4.2   Continuity of P

Definition: Continuous

A function f [Domain$(f) \subseteq \mathbb{R}$] is continuous at x_0 if for every sequence $\{x_n\}$ such that if

$$\lim_{n \to \infty} x_n = x_0$$

then

$$\lim_{n \to \infty} f(x_n) = f(x_0)$$

---

so

$$\lim_{n \to \infty} f(x_n) = f(\lim_{n \to \infty} x_n)$$

Definition: Increase\Decrease to

We say that $\{A_n\}$ increases to A (write $\{A_n\} \uparrow A$) if

$A_1 \subseteq A_2 \subseteq ... \subseteq A$; $\bigcup\limits_{n=1}^{\infty} A_n = A$; $\lim_{n \to \infty} A_n = A$

We say that $\{A_n\}$ decreases to A (write $\{A_n\} \downarrow A$) if

$A_1 \supseteq A_2 \supseteq ... \supseteq A$; $\bigcap\limits_{n=1}^{\infty} A_n = A$; $\lim_{n \to \infty} A_n = A$

Theorem: Continuous Function

P is a continuous function if $\{A_n\} \uparrow A$ or $\{A_n\} \downarrow A$ then $P(A) = \lim_{n \to \infty} P(A_n)$

Proof. (Proof of the Continuous Function Theorem)
Case 1: $\{A_n\} \uparrow A$
Define $A_2, A_3, ..., A$ as

$$A_2 = A_1 \cup (A_2 \setminus A_1)$$

$$A_3 = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2)$$

$$...$$

$$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup ...$$

Consider $P(A)$

$$
\begin{aligned}
P(A) &= P(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup ...) \\
&= P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + ... \\
&= P(A_1) + \sum_{i=2}^{\infty} P(A_i \setminus A_{i-1}) \\
&= P(A_1) + \lim_{n\to\infty} \sum_{i=2}^{n} P(A_i \setminus A_{i-1}) \\
&= P(A_1) + \lim_{n\to\infty} \sum_{i=2}^{n} [P(A_i) - P(A_{i-1})]
\end{aligned}
$$

Notice that the series is telescoping

$$
\begin{aligned}
&= P(A_1) - P(A_1) + \lim_{n\to\infty} P(A_n) \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}
$$

Case 2: $\{A_n\} \downarrow A$
if $\{A_n\} \downarrow A$ then $\{A_n^c\} \uparrow A^c$
<u>i.e.</u> $P(A^c) = \lim_{n\to\infty} P(A_n^c)$
Consider $(1 - P(A))$

$$
\begin{aligned}
(1 - P(A)) &= 1 - \left[\lim_{n\to\infty} P(A_n^c) \right] \\
-P(A) &= - \lim_{n\to\infty} P(A_n^c) \\
P(A) &= \lim_{n\to\infty} P(A_n^c)
\end{aligned}
$$

∎

Example 5: Suppose $P([0, \frac{8}{4+n})) = \frac{2+e^{-n}}{6}$ find $P(\{0\})$

Solution 5: $\{A_n\} \uparrow 0$

$$
\begin{aligned}
P(\{0\}) &= \lim_{n\to\infty} P(A_n) \\
&= \frac{2 + e^{-n}}{6} \\
&= \frac{1}{3}
\end{aligned}
$$

5   Wednesday, May 17, 2017

5.1   Random Variables

Definition: Random Variables

A random variable X is a function from S to the set of real numbers ($\mathbb{R}$)

Example 1: Consider the experiment where you roll a die ($S = \{1, 2, 3, 4, 5, 6\}$).

Define $\forall s \in S$ that X as $X(s) = s$. So $X(5) = 5$, $X(2) = 2$, and etc.

Example 2: The Indicator Function

$$X(s) = \begin{cases} 1, & s \in A \\ 0, & \text{otherwise} \end{cases}$$

5.2   Distribution of a Random Variable

For random variable X and $B \subseteq S$, $X \in B = \{s \in S | X(S) \in B\}$. Note that $X \in B \subseteq S$ is an event.

Definition: The Distribution of a Random Variable X

The collection of probabilities $P(x \in B)$, $\forall$ subset of $\mathbb{R}$ is called the **Distribution of X**.

Example 3: Consider $S = \{\text{Clear, Sunny, Rain}\}$ and $P(\{\text{Clear}\}) = 0.5$, $P(\{\text{Sunny}\}) = 0.2$, $P(\{\text{Rain}\}) = 0.3$.

And define $X(\text{Clear}) = 200$, $X(\text{Sunny}) = 100$, $X(\text{Rain}) = -50$.

Find the distribution of X.
$P(x \in B) = (0.5) \cdot I_B(200) + (0.2) \cdot I_B(100) + (0.5) \cdot I_B(-50)$ for any $B \in \mathbb{R}$
$P(X \in \{-50, 100, 200\}) = 0.5 + 0.2 + 0.3 = 1.0$

5.3   Discrete Random Variables

Definition: Discrete Random Variables

A random variable X is discrete of there is a finite or countable set of real numbers $x_1, x_2, ...$ and a corresponding sequence of non-negative real numbers $p_1, p_2, ...$ such that

$$P(X = x_i) = p_i, \forall i, and \sum_i p_i = 1$$

Probability Mass Function

The Probability Mass Function of a discrete random variable X is the function $P_X : \mathbb{R} \rightarrow [0, 1]$ defined by $P_X(x) = P(X = x)$. Note:

$$P(X \in A) = \sum_{x \in A} P(X = x) = \sum_{x \in A} P_X(x)$$

5.4   Important Discrete Distributions

1. **Degenerate (Point Mass) Distribution**
   A random variable X is S and has a degenerate distribution if

$$P_X(x) = \begin{cases} 1, & x = c \\ 0, & x \neq x \end{cases}$$

2. **Bernouelli Distribution**
   A random variable X is said to have a Bernoulli Distribution ($X \sim \text{Ber}(\theta)$) if

   $$P_X(1) = P(X = 1) = \theta \text{ while } P_X(0) = P(X = 0) = 1 - \theta$$

3. **Binomial Distribution**
   A random variable X is said to have a binomial distribution with parameters n and θ if

$$P_X(x) = \begin{cases} \binom{n}{x} \theta^x (1 - \theta)^{n-x} & \text{for } x = 0, 1, 2, ..., n \\ 0 & \text{otherwise} \end{cases}$$

Example 4: Toss a coin n times $(P(H) = \theta)$
Let X be the number of heads

$$P_X(x) = P(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

Binomial Experiment must satisfy the following conditions

(a) A fixed number of trials (n)

(b) Trials must be independent

(c) Outcomes must fall into 2 categories (either success or failure)

(d) Probability of a success should be the same for all trials

> **Note**
>
> (a) $P_X(x) \geq 0$
>
> (b) $\sum_{x=0}^{n} P_X(x) = 1$

Proof. (Proof of (2.) in Note)

$$\sum_{x=0}^{n} P_X(x) = \sum_{x=0}^{n} \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$
$$= (\theta + 1 - \theta)^n = 1^n = 1$$

Recall that

$$(a + b)^n = \sum_{x=0}^{n} a^x b^{n-x}$$

∎

4. **Geometric Distribution**
   A random variable X is said to have a geometric distribution with parameter $(\theta \in (0,1))$ if $P_X(x) = (1 - \theta)^x \cdot \theta$ for $x = 0, 1, 2, ...$

   Note: $P_X(x) \geq 0 \; \forall x$, but is it a Probability Measure Function?

Proof. (Proof of axiom 3 and 2 of a PMF)

$$\sum_{n=0}^{\infty} P_X(x) = \sum_{n=0}^{\infty} (1 - \theta)^x \theta$$
$$= \theta \sum_{n=0}^{\infty} \gamma^x (\text{ where } \gamma = 1 - \theta)$$
$$= \theta \frac{1}{1 - \gamma}$$
$$= \theta \frac{1}{\theta}$$
$$= 1$$

∎

5. **Negative Binomial Distribution**
   We need r successes and where x is the failures before the rth success

$$P(X = x) = \binom{x + r - 1}{x} (1 - \theta)^x \theta^{r-1} \theta$$

A random variable X is said to have a negative binomial distribution with parameters r and θ where $\theta \in (0, 1)$ if

$$P_X(x) = \begin{cases} \binom{x+r-1}{x}(1 - \theta)^x \theta^r & x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

6   Friday, May 19, 2017

6.1   Important Distributions (continued again)

Some Notes on Binomial Distributions

- Experiments Involving the Binomial Distributions have the following properties:

    1. Fixed Number of trials

    2. Each trial should result in 1 of 2 possible outcomes

    3. The trials must all be independent

    4. The probability of success should be the same in every trial

- Is this distribution (Negative Binomial Distribution) a probability measure function?

i.e. is $\sum_{x=0}^{\infty} P_X(x) = 1$?

$$\sum_{x=0}^{\infty} P_X(x) = \sum_{x=0}^{\infty} \binom{x+r-1}{x} (1-\theta)^x \theta^r$$

$$= \theta^r \sum_{x=0}^{\infty} \binom{x+r-1}{x} (1-\theta)^x$$

$$= \theta^r \sum_{x=0}^{\infty} \frac{(r+x-1)(r+x-2)...(r+x-1-(x-1))}{1 \cdot 2 \cdot ... \cdot x} (1-\theta)^x$$

$$= \theta^r \sum_{x=0}^{\infty} \frac{(r+x-1)(r+x-2)...(r)}{1 \cdot 2 \cdot ... \cdot x} (1-\theta)^x$$

$$= \theta^r \sum_{x=0}^{\infty} \frac{r(r+1)(r+2)...(r+x-1)}{1 \cdot 2 \cdot ... \cdot x} (1-\theta)^x$$

$$= \theta^r \sum_{x=0}^{\infty} \frac{(-1)^x[(-r)(-r-1)(-r-2)...(-r-(x-1))]}{1 \cdot 2 \cdot ... \cdot x} (1-\theta)^x$$

$$= \theta^r \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} (1-\theta)^x$$

$$= \theta^r \sum_{x=0}^{\infty} \binom{-r}{x} (1-\theta)^x$$

$$= \theta^r (1+\theta-1)^{-r}$$

$$= \theta^r \theta^{-r} = 1$$

Therefore it is a probability measure function.

6.2   Hyper Geometric Distribution

Definition: Hyper Geometric Distribution

A random variable X is said to have a Hyper Geometric Distribution
if

$$P_X(x) = \frac{\binom{M}{x}\binom{N-M}{m-x}}{\binom{N}{m}}$$

where $x \geq 0, m - x \leq N - M, x \geq m + M - N$
Denoted as $X \sim$ Hyper-Geometric(N, M, m)

Note: $P_X(x) \geq 0 \ \forall x \in \mathbb{R}$

Is $\sum_x P_X(x) = 1$?
The number of ways of choosing n balls with exactly x white balls is

$$\binom{M}{x}\binom{N-M}{m-x}$$

$$\sum_x \binom{M}{x}\binom{N-M}{m-x} = \text{Number of ways of choosing m balls} = \binom{N}{m}$$

$$\sum_x \frac{\binom{M}{x}\binom{N-M}{m-x}}{\binom{N}{m}} = 1, \text{ so yes, it is.}$$

7   Wednesday, May 24, 2017

7.1   Important Distributions (continued again)

Example 1: 5% of products produced are defectives. Take a random sample of $n = 20$ items.
$X = \sim$ Binomial$(m = 20, \theta = 0.05)$ (# of defectives)
What is $P(X = 2)$?

Solution 1:

$$P(X = 2) = P_X(x) = \binom{20}{2}(0.05)^2(0.95)^{18}$$

Example 2: A machine is producing 100 items. 5% are defective. Let $X \sim$ Hypergeometric$(N = 100, M = 5, n = 20$

Solution 2:

$$P(X = 2) = \frac{\binom{5}{2}\binom{95}{18}}{\binom{100}{20}}$$

7.2   Poisson Distribution

> **Definition: Poisson Distribution**
>
> A random variable X is said to have a Poisson Distribution with parameters $\lambda > 0$, if
>
> $$P_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} \text{ for } x = 0, 1, 2, ...$$

Example 3: Customers arrive during a particular period according to a Poisson Distribution with the parameters $\lambda = 7$. Find the probability:

1. No more than 3 customers arrive

2. At least 2 customers arrive

Solution 3:

1. Let $X = $ # of costumers that arrive $\sim \text{Poisson}(X = 7)$

$$P(X \leq 3) = P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0)$$
$$= e^{-7}\frac{7^3}{3!} + e^{-7}\frac{7^2}{2!} + e^{-7}\frac{7}{1} + e^{-7}$$
$$= e^{-7}(\frac{7^3}{3!} + \frac{7^2}{2!} + 8)$$

2.

$$P(X \geq 2) = 1 - P(X < 2)$$
$$= 1 - (P(X = 1) + P(X = 0))$$
$$= 1 - (e^{-7}7 + e^{-7})$$
$$= 1 - e^{-7}(8)$$

7.3 Continuous Distributions

Definition: Continuous Random Variables

A random variable X is continous if $P(X = x) = 0, \forall x \in \mathbb{R}$

Definition: Density Function

A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a density function if:

- $f(x) \geq 0, \forall x \in \mathbb{R}$

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

Definition: Absolutely Continuous Random Variables

A random variable X is absolutely continuous if there exists a density function f such that

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx \text{ whenever } a \leq b$$

we call this density function the probability density function of X (PDF of X).

Example 4: An absolutely continuous random variable has a PDF

$$f(x) = \begin{cases} kx, & 1 \le x \le 5 \\ 0, & \text{otherwise} \end{cases}$$

1. Find k

2. Find $P(2 \le X \le 3)$

Solution 4:

1.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{1} f(x)dx + \int_{1}^{5} f(x)dx + \int_{5}^{\infty} f(x)dx$$

$$= 0 + \int_{1}^{5} f(x)dx + 0$$

$$= \int_{1}^{5} f(x)dx$$

$$= k[\frac{x^2}{2}]_{1}^{5}$$

$$k = 1/12$$

2.

$$P(2 \le X \le 3) = \int_{2}^{3} \frac{1}{12}xdx$$

$$= \frac{1}{12}[\frac{x^2}{2}]_{2}^{3}$$

$$= \frac{1}{12}((9-4)/2)$$

$$= \frac{5}{24}$$

Example 5: A random variable X has the PDF

$$f_X(x) = \begin{cases} 0.2, & -1 \le x \le 0 \\ 0.2 + cx, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

1. Find c

2. Find $P(-0.5 \leq X \leq 0.05)$

Solution 5:

1.

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

$$\int_{-\infty}^{-1} f(x)dx + \int_{-1}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{\infty} f(x)dx = 1$$

$$0 + \int_{-1}^{0} f(x)dx + \int_{0}^{1} f(x)dx + 0 = 1$$

$$0.2 \int_{-1}^{0} dx + \int_{0}^{1} (0.2 + cx)dx = 1$$

$$0.2 + [0.2x + \frac{cx^2}{2}]_0^1 = 1$$

$$0.2 + 0.2 + c/2 = 1$$

$$0.4 + c/2 = 1$$

$$c = 1.2$$

2.

$$P(-0.5 \leq X \leq 0.05) = \int_{-0.5}^{0} f(x)dx + \int_{0}^{0.05} f(x)dx$$

$$= 0.2 \int_{-0.5}^{0} dx + \int_{0}^{0.05} (0.2 + 1.2x)dx$$

$$= 0.1 + [0.2x + 1.2x^2]_0^{0.05}$$

$$= 0.1 + 0.01 + 0.003$$

$$= 0.113$$

8   Friday, May 26, 2017

8.1   Continuous Distributions (continued)

Clarification: What is the difference between discrete and continuous random variables?

Consider $X(s) = s$ where $s = \{1, 2, 3, 4, 5, 6\}$.

This is discrete because the output can only be these 6 values.

Now, consider the different heights of people in the range of $[20, 50]$. Every height in this range is technically possible to have, so this is continuous.

8.2   Important Continuous Distributions

1. **Uniform Distribution over some interval** $[L, R]$

> Defintion: Uniform Distribution
>
> A random variable X has a Uniform $[L, R]$ Distribution if
>
> $$f_X(x) = \begin{cases} \frac{1}{R-L}, & L \leq x \leq R \\ 0, & \text{otherwise} \end{cases}$$

Example 1: $X \sim$Uniform$[5, 10]$

(a) Find $P(6 \leq x \leq 8)$

(b) Find $P(8 \leq x \leq 12)$

Solution 1:

(a)

$$P(6 \leq X \leq 8) = \int_6^8 \frac{1}{10 - 5} dx$$
$$= (1/5) \int_6^8 dx$$
$$= (2/5)$$

(b)

$$P(8 \leq X \leq 12) = P(8 \leq X \leq 10) + P(10 \leq X \leq 12)$$
$$= \int_8^{10} \frac{1}{10 - 5} dx + 0$$
$$= (2/5)$$

2. **Exponential Distribution**

> **Defintion: Exponential Distribution**
>
> A random variable X has a Exponential Distribution with the parameter $(\lambda > 0)$ if
>
> $$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Example 2: $X \sim$Exponential$(\lambda = 0.2)$, Find $P(x > 10)$

Solution 2:

$$P(X > 10) = \int_{10}^{\infty} 0.2 e^{-0.2x} dx$$
$$= [0.2 \frac{e^{-0.2x}}{-0.2}]_{10}^{\infty}$$
$$= [-e^{-0.2x}]_{10}^{\infty}$$
$$= e^{-2}$$

Note:

- For Uniform Distribution, $f_X(x) \geq 0$
- For Exponential Distribution,

$$\int_0^{\infty} f_X(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \lambda [\frac{e^{-\lambda x}}{-\lambda}]_0^{\infty} = [e^{-\lambda x}]_0^{\infty} = 1$$

Therefore $f_X(x)$ is a density function.

3. **Normal Distribution**

> Defintion: Normal Distribution
>
> A random variable X has a Normal Distribution with the parameter ($\mu \in \mathbb{R}$) and ($\sigma \in \mathbb{R}^+$) if
>
> $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(1/2)(\frac{x-\mu}{\sigma})^2} \text{ where } x \in \mathbb{R}$$
>
> We write $X \sim$Exponential(μ, σ^2)

Note: $N(\mu = 0, \sigma^2 = 1)$ is called the Standard Normal Distribution ($f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-(1/2)x^2}$)

Multivariable Calculus and Polar Coordinates Crash Course

Dr. Mahinda sort of speeds through these parts, so feel free to refer to other resources to understand what is happening here, since you need to understand these concepts to understand what is happening in the following material.

So in single variable calculus, you're probably used to integrals which find the area under an curve. Well, double integration is finding the area in-between two curves at different boundaries. Kind of like a circle, centred around the origin, you can either represent it as two integrals of the bottom and top part of the circle, or as a double integral, with the x going from the leftmost side of the circle to the rightmost side, and the y going from the top of the circle to the bottom. As an example, consider the circle with radius 1.

$$\int_{-1}^{1} \sqrt{1-x^2}dx + \int_{-1}^{1} -\sqrt{1-x^2}dx = \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1\,dy\,dx$$

with the general case for double integrals being

$$\int \int f(x,y)dA \text{ where } dA = dxdy$$

We also have polar coordinates that make it a lot easier to find the area of these types of circles, the definition being that $x = rcos\theta$, $y = rsin\theta$, $dxdy = rdrd\theta$ and

$$\int \int f(x,y)dA = \int \int f(rcos\theta, rsin\theta)rdrd\theta$$

So we can represent a circle with radius 1 centred around the origin as

$$\int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1\,dy\,dx = \int_{0}^{1} \int_{0}^{2\pi} 1\,d\theta\,dr$$

So, is

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2} = 1 \ ?$$

Proof.

$$\sqrt{2\pi} = I = \int_{-\infty}^{\infty} e^{-(1/2)x^2} dx \qquad \text{(so we need to prove that } I = \sqrt{2\pi})$$

$$I = \int_{-\infty}^{\infty} e^{-(1/2)y^2} dy \qquad \text{(Variable change)}$$

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(1/2)y^2} e^{-(1/2)x^2} dydx \qquad \text{("just assume this is true" - Mahinda}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(1/2)(y^2+x^2)} dxdy$$

$$= \int_{0}^{2\pi} \int_{0}^{\infty} e^{-(1/2)(r^2 cos^2\theta + r^2 sin^2\theta)} r dr \int_{0}^{2\pi} \quad \text{(change to polar coordinates)}$$

$$= \int_{0}^{2\pi} [-e^{-(1/2)r^2}]_0^{\infty} d\theta$$

$$= \int_{0}^{2\pi} 1 d\theta$$

$$I^2 = 2\pi$$

$$I = \sqrt{2\pi}$$

∎

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

$$\int_{\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)(\frac{x-\mu}{\sigma})^2} dx$$

sub $z = \frac{x-\mu}{\sigma}$ we also get:

$$dz = \frac{1}{\sigma} dx$$

so now we have

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

And this is the standard normal distribution

4. **Gamma Distribution**

> Definition: Gamma Function
>
> For $\alpha > 0$,
> $$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Properties of the Gamma Function

(a) $\Gamma(1) = 1$

(b) If $\alpha > 1$ then $\Gamma(\alpha) = (\alpha - 1)(\Gamma(\alpha - 1))$

(c) If α is a positive integer, $\Gamma(\alpha) = (\alpha - 1)!$

(d) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Proof. (Proof of (d))

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

$$\text{sub } x = \frac{1}{2}t^2$$
$$dx = tdt$$

$$= \int_0^\infty (\frac{1}{2}t^2)^{\frac{1}{2}} e^{-(\frac{1}{2}t^2)} tdt$$

$$= \sqrt{2} \int_0^\infty t^{-1} e^{-(\frac{1}{2}t^2)} tdt$$

$$= \sqrt{2} \int_0^\infty e^{-(\frac{1}{2}t^2)} dt$$

$$= \sqrt{2} \frac{\sqrt{2\pi}}{2}$$

$$= \sqrt{\pi}$$

■

Now Consider $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$$

for $\lambda > 0$, sub $x = \lambda t$
$$dx = \lambda dt$$

$$= \int_0^\infty \lambda^{\alpha-1}t^{\alpha-1}e^{-x}\lambda dt$$

$$= \int_0^\infty \lambda^{\alpha}t^{\alpha-1}e^{-x}dt$$

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

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

Definition: Gamma Distribution

A random variable has a Gamma Distribution with parameters $(\alpha > 0, \lambda > 0)$ if

$$f_X(x) = \begin{cases} \frac{\lambda^{\alpha}t^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Note: Gamma$(\alpha = 1, \lambda) = $ Exponential(λ)

5. **Beta Distribution**

Definition: Beta Function

A Beta (a, b) function for $a > 0, b > 0$ is defined as

$$B(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1}dx$$

Result: It can be shown that $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

$$\int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}dx = 1$$

Beta Distribution

A random variable X has a Beta Distribution with $a > 0, b > 0$ if

$$f_X(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Example 3:

$$\int_0^1 x^9(1-x)^{19}dx = \frac{\Gamma(10)\Gamma(20)}{\Gamma(30)}$$

9   Wednesday, May 31, 2017

Note: For the Normal Distribution, a larger sigma results in a larger horizontal spread.

9.1   Cumulative Distribution Functions (CDF)

Definition: Cumulative Distribution Functions

Given a random variable X, the function $F_X : \mathbb{R} \longrightarrow [0, 1]$ defined by $F_X(x) = P(X \le x)$ is called the Cumulative Distribution Function of X.

Properties:

1. $0 \le F_X(x) \le 1$ since $F_X(x)$ is a probability

2. $x \le y$ implies that $F_X(x) \le F_X(y)$

Proof. (Proof of 2)

$$x_1 \le x_2 \iff (X \le x_1) \subseteq (X \le x_2)$$
$$\iff P(X \le x_1) \le P(X \le x_2)$$
$$\iff F_X(x_1) \le F_X(x_2)$$

<u>i.e.</u> F_X is a non-decreasing function   ∎

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

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

5. If $a < b$, $P(x \in (a, b]) = F_X(b) - F_X(a)$

Proof. (Proof of 5)

$$x \in (a, b] = a < x \leq b$$
$$= (x \leq b) \setminus (x \leq a)$$
$$= P(x \leq b) - P(x \leq a)$$
$$= F_X(b) - F_X(a)$$

∎

6. If $a \leq b$, $P(x \in [a, b]) = F_X(b) - F_X(a)$ where $F_X(a-) = \lim_{n \to \infty} F_X(a - \frac{1}{n})$

Proof. (Proof of 6)
Let $A = x \in [a, b]$, $A_n = x \in (a - \frac{1}{n}]$, $n \in \mathbb{N}$

Note: $a \in A_n \forall n$ therefore $a \in \bigcap_n A_n$

$A_n \downarrow A = [a, b]$
By continuity of probability $P(A) = \lim_{n \to \infty} P(A_n)$

$$P(x \in [a, b]) = \lim_{n \to \infty} P(x \in (a - \frac{1}{n}, b])$$
$$= \lim_{n \to \infty} [F_X(b) - F_X(a - \frac{1}{n})]$$
$$= F_X(b) - F_X(a-)$$

∎

Note: if $a = b$, $P(x \in [a, a]) = P(x = a) = F_X(a) - F_X(a-)$

7. If $a < b$, $P(x \in (a, b)) = F_X(b-) - F_X(a)$

Proof. (Proof of 7)
Let $A = x \in (a, b)$, $A_n = x \in (a, b - \frac{1}{n}]$

$A_n \uparrow A = x \in (a, b)$ since $b \notin A_n \forall n$

By continuity of Probability:

$$P(x \in (a, b)) = \lim_{n \to \infty} P(a, b - \frac{1}{n}]$$
$$= \lim_{n \to \infty} [F_X(b - \frac{1}{n}) - F_X(a)]$$
$$= \lim_{n \to \infty} F_X(b - \frac{1}{n}) - F_X(a)$$
$$= F_X(b-) - F_X(a)$$

∎

9.2   Cumulative Distribution Functions for Discrete Random Variables

Ex. Roll a 4-sided die, $S = \{1, 2, 3, 4\}$. Define $X(s) = s$, this random variable is a discrete random variable.

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ \frac{1}{4}, & 1 \leq x < 2 \\ \frac{2}{4}, & 2 \leq x < 3 \\ \frac{3}{4}, & 3 \leq x < 4 \\ 1, & \text{otherwise} \end{cases}$$

Note: if $F_X(x) = F_X(x+)$, F_X is considered to be right continuous.

Note: In the example above, each of the discontinuity points is right continuous, but not left continuous or just continuous.

10   Friday, June 2, 2017

10.1   Cumulative Distribution Functions of Absolutely Continuous Random Variables

(i.e. random variables that have density functions)

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \text{ where } f(t) \text{ is the density function}$$

Note: $f_X(x) = \frac{dF_X(x)}{dx}$ if $F_X(x)$ is differentiable at x.

Example 1: Find the Probability Density Function of X if

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

Solution 1:
$$f_X(x) = \frac{dF_X(x)}{dx} = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

And since $0 \leq f_X(x) \forall x$ and that

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} 1 \cdot dx = x|_{0}^{1} = 1$$

Therefore, this is a Probability Density Function.

For the Standard Normal Distribution, the probability density function is denoted by ϕ

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, x \in \mathbb{R}$$

The Cumulative Distribution Function of the Normal Distribution

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

Although this integral is integrate, no one actually knows how to actually integrate this, so they made a table of set values of different outputs, called

Normal Tables.

Example 2: $X \sim N(\mu = 0, \sigma^2 = 1)$, Find $P(-0.63 \leq X \leq 2.0)$

Solution 2:

$$
\begin{aligned}
P(-0.63 \leq X \leq 2.0) &= \Phi(2.00) - \Phi(-0.63) \\
&= 1 - \Phi(-2.00) - \Phi(-0.63) \\
&= 1 - 0.0228 - 0.2643 \qquad \text{By Normal Tables}
\end{aligned}
$$

The General Normal Distribution Case: If $X \sim N(\mu, \sigma^2)$

$$
P(a \leq X \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx \qquad\qquad \text{sub } z = \frac{x-\mu}{\sigma}
$$

$$
dx = \sigma dz
$$

$$
\begin{aligned}
&= \int_{((a-\mu)/\sigma)}^{((b-\mu)/\sigma)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dx \\
&= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})
\end{aligned}
$$

Example 3: $Y \sim N(\mu = -8, \sigma^2 = 4)$

1. Find $P(-2 \leq Y \leq 7)$

2. Find $P(Y \geq 7)$

Solution 3:

1.

$$
\begin{aligned}
P(-2 \leq Y \leq 7) &= \Phi(\frac{7-(-8)}{2}) - \Phi(\frac{-2-(-8)}{2}) \\
&= \Phi(7.5) - \Phi(3) \\
&= [1 - \Phi(-7.5)] - [1 - \Phi(-3)] \\
&\qquad\qquad\qquad\qquad \forall x < -3.5, \Phi(x) = \text{basically 0} \\
&= [1 - 0] - [1 - 0.0013]
\end{aligned}
$$

2.

$$P(Y \geq 3) = 1 - P(Y \leq 3)$$
$$= 1 - \Phi((3 - (-8))/2)$$
$$= 1 - \Phi(5.5)$$
$$= 1 - [1 - \Phi(-5.5)]$$
$$= 1 - 1 = 0$$

<u>Note:</u> $P(Y < 3) = P(Y \leq 3) - P(Y = 3)$ and since Y is continuous, $P(Y = 3) = 0$, so $P(Y < 3) = P(Y \leq 3)$.

10.2   Mixture Distributions

> **Definition: Mixture Distribution**
>
> Let $F_1, F_2, ..., F_k$ be cumulative distribution functions and $p_1, p_2, ..., p_k$ are positive real numbers such that $\sum_{i=1}^{k} p_i = 1$. Then $p_1 F_1 + p_2 F_2 + ... + p_k F_k$ is a cumulative distribution function. The distribution with cumulative distribution function $p_1 F_1 + p_2 F_2 + ... + p_k F_k$ is called the Mixture Distribution of $F_1, F_2, ..., F_k$.

With this, there exists distributions that are neither continuous nor discrete.

Recall:

- Discrete: $\sum_i P(X = x_i) = 1$

- Continuous: $P(X = x) = 0 \forall x$

Example 4: Let $X_1 \sim P(\lambda = 3)$, $X_2 \sim N(0, 1)$. Let F_1, F_2 be the Cumulative Distribution Functions of X_1 and X_2 respectfully.

Let Y be a random variable with the CDF of $F_Y(y) = 0.2F_1(y) + 0.8F_2(y)$, Let $y \in \mathbb{R}$.

$$P(Y = y) = F_Y(y) - F_Y(y-)$$
$$= [0.2F_1(y) + 0.8F_2(y)] - [0.2F_1(y-) + 0.8F_2(y-)]$$
$$= 0.2[F_1(y) - F_1(y-)] + 0.8[F_2(y) - F_2(y-)]$$
$$= 0.2P(X_1 = y) + 0.8P(X_2 = y)$$

$P(X_2 = y) = 0$ since X_2 is absolutely continuous.

$$= \begin{cases} 0.2(e^{-3}\frac{3^y}{y!}), & y = 0, 1, 2, ... \\ 0, & \text{otherwise} \end{cases}$$

Check if discrete:

$$\sum_{y=0}^{\infty} P(Y = y) = \sum_{y=0}^{\infty} 0.2e^{-3}\frac{3^y}{y!} = 0.2\sum_{y=0}^{\infty} e^{-3}\frac{3^y}{y!} = 0.2(1) = 0.2, \text{therefore, not discrete}$$

Check if Y is continuous:

$$P(Y = 0) = 0.2e^{-3}\frac{3^0}{0!} \neq 0, \text{therefore, not continuous}$$

10.3   One-dimensional change of variables

Idea: Given a distribution of X (P_x or f_X), we have $y = h(x)$, What is the PDF/PMF of Y?

Discrete Case: Example 5: Let X be a random variable taking possible values of $m = \{-3, -2, -1, 0, 1, 2, 3\}$ with equal probability.

$$P_X(x) = \begin{cases} \frac{1}{7}, & x \in m \\ 0, & \text{otherwise} \end{cases}$$

Let $Y = X^2 - X$, find the probability measure function of Y.

Solution 5: Let $h : X \mapsto Y$, so that: $h(-3) = 12$, $h(-2) = 6$, $h(-1) = 2$, $h(0) = 0$, $h(1) = 0$, $h(2) = 2$, $h(3) = 6$.

Consider $P_Y(0) = P(Y = 0) = P(X = 0) + P(X = 1) = 1/7 + 1/7 = 2/7$

Let $h^{-1} : Y \mapsto X$, so that $h^{-1}(y) = \{x : h(x) = y\}$

So $P_Y(y) = \sum_{i \in h^{-1}(y)} P_X(i) = \begin{cases} 2/7, & y \in \{0, 2, 6\} \\ 1/7, & y = 12 \\ 0, & \text{otherwise} \end{cases}$

Absolutely Continuous Case: Example 6: Let X be a random variable

with the PDF of $f_X(x) = \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$. Let $Y = 2X - 1 (h(X) = 2X - 1)$, Find the PDF of Y.

Solution 6: Method 1: (Distribution Function Method)

Find the CDF of Y first and then differentiate.

$$F_Y(y) = P(Y \le y) = P(2X - 1 \le y) = P(X \le \frac{y+1}{2}) = F_X(\frac{y+1}{2})$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\frac{y+1}{2}) = f_X(\frac{y+1}{2}) \frac{1}{2} = \begin{cases} 2(1 - (\frac{y+1}{2})) \frac{1}{2}, & 0 \le \frac{y+1}{2} \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1-y}{2}, & |y| < 1, y \ne 1 \\ 0, & \text{otherwise} \end{cases}$$

11   Wednesday, June 7, 2017

11.1   R syntax and Basic R functions

Mahinda just talks about programming in R as well as the syntax and how to calculate probabilities and solve integrals with it. He simply showed how to use R while referring to these notes:
`http://fisher.utstat.utoronto.ca/~mahinda/stab52/wk6Rcodes.pdf`.

I don't really see the need to type out notes for this section since everything is already provided with the notes of basic R functions linked above.

In terms of this lecture, he stated that students of STAB52 don't have to know how to code in R but need to be able to read the syntax for midterm and final exam questions.

12 Friday, June 9, 2017

12.1 One-dimensional change of variables (continued)

Given a distribution of a random variable X and $Y = h(X)$, we want the probability density function of Y (i.e. f_Y). There are two methods to solve this:

1. Cumulative Distribution Function (shown 2 lectures ago)

2. Transformation Formula

> **Theorem: Transformation Formula**
>
> Let X be an absolute continuous random variable and $Y = h(X)$ where $h : \mathbb{R} \mapsto \mathbb{R}$ is a function that is differentiable and either strictly increasing or strictly decreasing <u>on the support of f(x)</u>, then Y is an absolutely continuous random variable and
>
> $$f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|} \text{ where } h'(x) = \frac{d}{dx}h(x)$$

Definition of "support of distribution/random variable/density function":

$$f_X = \{x | f_X(x) > 0\}$$

Example 1: To continue from our example for the Cumulative Distribution Function. $Y = 1 - 2X = h(X)$.

Solution 1: This implies that

$$h^{-1}(Y) = \frac{1-Y}{2} \text{ and } h'(X) = -2$$

Using the transformation formula

$$f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|} = \begin{cases} \frac{2(1-\frac{1-y}{2})}{|-2|}, & 0 \le \frac{1-y}{2} \le 1 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{1+y}{2}, & -1 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Example 2: considering $Y = X^2 = h(X)$, find f_Y

Solution 2: If the support is $[0, 1]$, we can use method 2, since the function would be increasing and differentiable on the interval. So we're just gonna pretend like it is so we have an actual example to solve.

$$h^{-1}(y) = \sqrt{y}, h'(x) = 2x$$

$$f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|} = \begin{cases} \frac{2(1-\sqrt{y})}{2\sqrt{y}}, & 0 \leq \sqrt{y} \leq 1 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{1-\sqrt{y}}{\sqrt{y}}, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

12.2   Joint Distributions

Definition: Joint Distribution

Let X, Y be 2 random variables, then the collection of probability $P((X, Y) \in B), B \subseteq \mathbb{R}^2$ is called the Joint Distribution of X and Y

Definition: Joint Cumulative Distribution Function of 2 Random Variables X and Y

Let X, Y be 2 random variables, then the function $F_{X,Y} : \mathbb{R}^2 \mapsto [0, 1]$ defined by $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ is called the Joint Cumulative Distribution Function of X and Y.