Lecture 1: Introduction

January 5, 2023

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Probability theory is the branch of mathematics that is concerned with random experiment. So it's natural to ask what is the meaning of a random experiment? Many phenomena have the property that their repeated observation under a specified set of conditions invariably leads to the same outcome. For example, if we deposit a certain amount in a fixed deposit for one year then we know beforehand how much amount will receive at the end of the year. FD is an example of deterministic process. There are other processes whose outcome can not be certain, for example if we invest a certain amount in mutual funds or share markets then we can not predict the maturity amount. These examples gives us some idea about deterministic and random processes. Another example you may take from physics, the classical physics is a deterministic model of studying the motion of particles (if you specify the initial velocity and force etc. then we may predict the position, velocity etc at future times) and quantam physics is a stochastic model or probabilistic model.

Another familiar example of random experiments is the tossing of a coin. If a coin is tossed 1000 times the occurrences of heads and tails alternate in a seemingly erratic and unpredictable manner. It is such phenomena that we think of as being random and which are the object of our investigation.

At first glance it might seem impossible to make any worthwhile statements about such random phenomena, but this is not so. Experience has shown that many nondeterministic phenomena exhibit a "statistical regularity" that makes them subject to study. For example if we toss a coin large number of times, the proportion of heads seems to fluctuate around $\frac{1}{2}$ unless the coin is severely unbalanced.

The eighteenth century French naturalist Comte de Buffon tossed a coin 4040 times and got 2048 heads. The proportion (or relative frequency) of heads in this case is 0.507. J.E. Kerrich from Britain, recorded 5067 heads in 10000 tosses. Proportion of heads in this case is 0.5067. Statistician Karl Pearson spent some more time, making 24000 tosses of a coin. He got 12012 heads, and thus, proportion of heads in this case is 0.5005.

Of course, this proportion of heads depends on number of tosses and will fluctuate, even wildly, as number of tosses increases. But if we let number of tosses go to infinity, will the sequence of relative proportion "settle down to a steady value"? Such a question can never be answered empirically, since by the very nature of a limit we cannot put an end to the trials. So it is a mathematical idealization (or belief) to assume that such a limit does exist and is equal to 0.5, and then write $P(\text{Head}) = 0.5$. We think of this limiting proportion 0.5 as the "probability" that the coin will land heads up in a single toss.

So there is a certain level of belief/assumption when we say that probability of head is 0.5.

More generally the statement that a certain experimental outcome has probability p can be interpreted as meaning that if the experiment is repeated a large number of times, that outcome would be observed "about" (approximately) $100p$ percent of the time. This interpretation of probabilities is called the relative frequency interpretation. It is very natural in many applications of probability theory to real world problems, especially to those involving the physical sciences.

Let us summarize what we have done "When we say that probability of head is $\frac{1}{4}$ while tossing a coin, we mean if we toss the coin large number of times then approximately $\frac{1}{4}$th times we should get head". Same interpretation can applied when we say that probability of an even number is $\frac{1}{2}$ when we through a die.

Question: What interpretation you make of the following statement " Modi has 70% chance of winning 2024 Loksabha elections"?

It is obvious that we can not have relative frequency interpretation for the above statement.

For the mathematical theory of probability the interpretation of probabilities is irrelevant. In general, mathematics is concerned solely with relations among undefined things. For example, geometry does not care what a point and a straight line "really are." They remain undefined notions, and the axioms of geometry specify the relations among them : two points determine a line, etc . Different forms of geometry are based on different sets of axioms, and the logical structure of non-Euclidean geometries is independent of their relation to reality.

We shall no more attempt to explain the "true meaning" of probability than the modern physicist dwells on the "real meaning" of mass and energy. Instead, we shall prove theorems and see how they are applied. We will use the relative frequency interpretation of probabilities only as an intuitive motivation for the definitions and theorems we will be developing throughout the course.

1.1 Sample Space, Countable Sets

Definition 1.1 *By a random experiment (or chance experiment), we mean an experiment (a imaginary thought experiment which may never actually be performed but can be conceived of as being performed or a real physical process) which has multiple outcomes (at least two) and one don't know in advance which outcome is going to occur, unless one perform the experiment. Also, when experiment is performed exactly one out of several possible outcomes will be produced.*

Example 1.2 *Tossing a coin, throwing a die are random experiment. Unless you throw the coin or die you don't know what is coming up. Though we know all possible outcomes of the random experiment.*

Remark 1.3 *There is no restriction on what constitutes a random experiment. For example, it could be a single toss of a coin, or three tosses, or an infinite sequence of tosses. However, it is important to note that in our mathematical model of randomness, there is only one random experiment. So, three tosses of a coin constitute a single experiment, rather than three experiments.*

We begin with a model for a random experiment whose performance, results in an idealized outcome from a family of possible outcomes. The first element of the model is the specification of an abstract sample space (Ω) representing the collection of idealized outcomes of the random experiment. Next comes the identification of a family of events ($\mathcal{F}$) of interest, each event represented by an aggregate of elements of the sample space. The final element of the model is the specification of a consistent scheme of assignation of probabilities (P) to events. We consider these elements in turn.

1.2 The Sample Space

Definition 1.4 *Sample space of a random experiment is the set of all idealized outcomes of the random experiment. We denote it by the uppercase Greek letter* Ω.

Example 1.5 *In random experiment of tossing a coin, the sample space is* $\Omega = \{head, tail\}$ *or* $\{H, T\}$.

Remark 1.6 *In the sample space for the random experiment in Example 1.5, one might also include as possible outcomes "coin stands on the edge" or "coin disappears" (Shaktimaan threw the coin and it crossed the gravity of the earth!!). It is not too serious if we admit more things into our consideration than can really occur, but we want to make sure that we do not exclude things that might occur. As a guiding principle, the sample space should have enough detail to distinguish between all outcomes of interest to the modeler, while avoiding irrelevant details. In this spirit we have made use of idealized outcomes.*

Example 1.7 *Toss two coins. Here sample space is* $\Omega = \{HH, HT, TH, TT\}$

Example 1.8 *In the random experiment of tossing a coin till you get a head, the sample space is*

$$\Omega = \{H, TH, TTH, TTTH, \cdots\}$$

Example 1.9 *Consider throwing a dart on a square target and viewing the point of impact as the outcome. The sample space is the set of all the points on the square.*

Example 1.10 *The physical variables such as electric and magnetic fields, currents, and voltages are represented by complex numbers. Also complex numbers provide the most economical descriptions of quantum states of matter. So if our random experiment if observing a physical process with these quantities then the sample space is the set of complex numbers.*

Natural models for a variety of problems lead to even richer sample spaces, for instance, involving functions as the raw objects.

Example 1.11 *The space of continuous functions whose sample points f are continuous real-valued functions of a real variable. In a variety of problems, one can model random phenomena as producing a function, say $f(\cdot)$, as the outcome of an experiment. This is the natural model for noise processes in electrical communications, for instance.*

The sample space of a random experiment may consist of a finite or an infinite number of possible outcomes. Infinite sets are further divided into two categories: countably infinite and uncountable. The sample space in Example 1.8 is countably infinite and the sample space in Example 1.9 is uncountable.

Dart square.

$\Omega = \{H, TH, TTH, TTTH, \cdots\}$

1.2.1 Countable Sets

We want to measure number of elements in a set. When set S is a finite set, it is very clear to us that we may list the elements as $\{s_1, s_2, \cdots s_n\}$ for some positive integer n. When we deal with infinite sets, things are not that straight. For example, set of natural numbers $\mathbb{N}$ "appears" to contain as much as "double" of elements in set $S = \{2n | n \in \mathbb{N}\}$ (set of all even positive integers). Can we say that number of elements in $\mathbb{N}$ is 2 times the number of elements in S?

Now here is the idea due to Cantor.

Definition 1.12 *We say that two sets A and B have the same cardinality (or number of elements) if there exists a one-one and onto function (or bijection) from A to B (or equivalently from B to A).*

⇒ one-one and Onto.

Let us recast the notion of finiteness as follows.

Definition 1.13 *A set S is finite if there exists a bijection between S and the set $\{1, 2, \cdots, n\}$ for some positive integer n. In this case we say S has n elements or the number of elements in S is equal to n.*

So finiteness has been recasted as a one-to-one correspondence between the set and a subset of natural number. This idea paves the way to define the following notion.

Definition 1.14 *A set S is said to be countably infinite if there exists a function $f : \mathbb{N} \to S$ such that f one-one and onto.*

In other words, a set is countably infinite when it can be put into 1-to-1 correspondence with the set of positive integers. In this case we say that the set S is as infinite as $\mathbb{N}$.

Example 1.15 *Let S be the set of all even positive integers. Then it is countable.*

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \vdots \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & \cdots \end{array}$$

Consider $f : \mathbb{N} \to S$ defined as $f(n) = 2n$. Then f is a bijection. For one-one, $f(n) = f(m) \implies 2n = 2m \implies m = n$. To see f is onto, let $x \in S$, then $x = 2n_0$ for some $n_0 \in \mathbb{N}$. Hence $f(n_0) = x$.

It is highly counter intuitive that both the sets have "same order of infinity", though empirically it appears that set S have half of the elements compare to $\mathbb{N}$.

Example 1.16 *Set of all integers $\mathbb{Z}$ is a countable set.*

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \vdots \\ 0 & 1 & -1 & 2 & -2 & 3 & -3 & 4 & -4 & \cdots \end{array}$$

Define $f : \mathbb{N} \to \mathbb{Z}$ as

$$f(n) = \begin{cases} \dfrac{n}{2} & \text{if } \; n \textit{ is even} \\ -\dfrac{n-1}{2} & \text{if } \; n \textit{ is odd} \end{cases}$$

That is $f(1) = 0, f(2) = 1, f(3) = -1, f(4) = 2, f(5) = -2, f(6) = 3, f(7) = -3 \cdots$. To see that f is one-one, $m \neq n \implies \dfrac{m}{2} \neq \dfrac{n}{2}$ and $-\dfrac{n-1}{2} \neq -\dfrac{m-1}{2} \implies f(m) \neq f(n)$. Now let $m \in \mathbb{Z}$. Consider three cases:

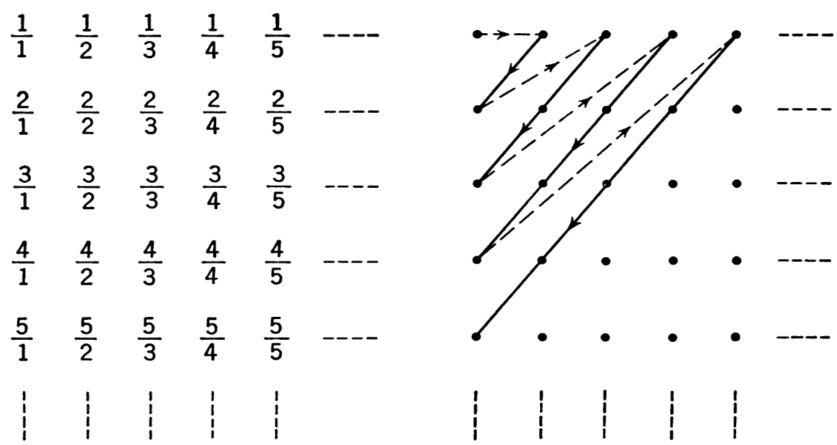
1. *If $m = 0$, then $f(1) = m$.*
2. *If $m < 0$, then $f(1-2m) = -\dfrac{1-2m-1}{2} = m$.*
3. *If $m > 0$, then $f(2m) = m$.*

Hence f is onto.

This shows that $\mathbb{Z}$ and $\mathbb{N}$ have same of number of elements.

Example 1.17 *The set of all positive rational numbers is countably infinite.*

The following figure gives a schematic representation of listing the positive rationals. In this figure the first row contains all positive rationals with numerator 1, the second all with numerator 2, etc.; and the first column contains all with denominator 1, the second all with denominator 2, and so on. Our listing amounts to traversing this array of numbers as the arrows indicate, where of course all those numbers already encountered are left out.



We see that every natural number is associated with the unique positive rational number and for every positive rational number, there exists a natural number. Hence set of all positive rational numbers is countable.

I don't know whether one can write a formula for the bijection just described as we did in Example 1.15 and Example 1.16. But there are other bijection with a formula are defined.

Lecture 3: Countable & Uncountable Sets

January 6, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

3.1 Countability of Rational Numbers

Example 3.1 *Show that set of all rational numbers is countably infinite.*

Solution: In proof of countability of positive rationals, we counted as follows:

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \vdots \\ 1 & \frac{1}{2} & 2 & \frac{1}{3} & 3 & \frac{1}{4} & \frac{2}{3} & \frac{3}{2} & 4 & \cdots \end{array}$$

Now what we did for counting integers, we do the same thing to include non-positive rationals as follows:

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \vdots \\ 0 & 1 & -1 & \frac{1}{2} & -\frac{1}{2} & 2 & -2 & \frac{1}{3} & -\frac{1}{3} & \cdots \end{array}$$

■

Remark 3.2 *Since every countably infinite set is the range of a 1-1 function defined on $\mathbb{N}$, we may regard every countably infinite set as the range of a sequence of distinct terms (Note that in general the terms $x_1, x_2, x_3, \cdots$ of a sequence need not be distinct). Speaking more loosely, we may say that the elements of any countably infinite set can be "arranged" in a sequence of distinct terms.*

3.2 Uncountable Set

Definition 3.3 *An infinite set which is not countably infinite is called uncountable.*

Theorem 3.4 *Each infinite set S contains a denumerable (or countably infinite) subset.*

Proof: Since S is infinite, it is non-empty and contains an element s_1. Since S is infinite the set $S \setminus \{s_1\}$ is non-empty and there exists $s_2 \in S \setminus \{s_1\}$. Since S is an infinite set, $S \setminus \{s_1, s_2\}$ is non-empty and there exists $s_3 \in S \setminus \{s_1, s_2\}$. Continuing this way gives a list (s_n) of distinct elements of S. The set of these elements forms a denumerable subset of S. ■

Proposition 3.5 *Every infinite subset of a countable set A is countable.*

Proof: Since A is countable so we can arrange all the elements of set A as a sequence of distinct terms say $(a_n)_{n\geq 1}$. Suppose B is an infinite subset of A then B can seen as subsequence of $(a_n)_{n\geq 1}$ say $(a_{n_k})_{k\geq 1}$. A subsequence is a sequence. Also all the terms of the subsequence $(a_{n_k})_{k\geq 1}$ are distinct. Therefore B is countably infinite. ■

Definition 3.6 *Let A and B be sets. If there is a one-to-one function from A to B but no bijection between A and B then we will write $card(A) < card(B)$. This notation is read "A has smaller cardinality than B." We use the notation $card(A) \leq card(B)$ to mean that either $card(A) < card(B)$ or $card(A) = card(B)$.*

Hence if $A \subset B$, then the inclusion map $i(a) = a$ is a one-to-one function of A into B; therefore $card(A) \leq card(B)$.

Remark 3.7 *If A is an infinite set, there are two possibilities, it might be countable or uncountable. The Proposition 3.5 tell us that an uncountable set can not be a subset of a countable set, whereas Theorem 3.4 tells us that an uncountable set have a countable subset. This shows that, the cardinality of a countable infinite set is strictly less than the cardinality of an uncountable set. In other words, although countable and uncountable are both infinite, uncountable is more infinite than countable.*

Example 3.8 *Consider the random experiment of tossing a coin indefinitely. Then the same space Ω consists infinite sequences of Hs and Ts, i.e.,*

$$\Omega = \{\omega = (\omega_1, \omega_2, \cdots) : \omega_i \in \{H, T\} \text{ for each } i = 1, 2, \cdots\}$$

We could just view heads and tails in a coin toss as 1 and 0, respectively. Then the set Ω could be rewritten as $\Omega = \{(a_n)_{n\geq 1} : a_n \in \{0, 1\} \text{ for each } n = 1, 2, \cdots\}$.

Claim 3.9 *The set Ω is uncountable.*

It is clear that Ω is infinite. Now suppose contrary that Ω is countable. Then we can enumerate it's element in a sequence $s_1, s_2, s_3, \cdots$. Now we construct a sequence s as follows: If the nth term in s_n is 1, we let nth term of s be 0, and vice versa. Then the sequence s differs from each of s_n. But by construction $s \in \Omega$, which is a contradiction.

The idea of the above proof is called Cantor's diagonalization argument; for, if the sequences $s_1, s_2, s_3, \cdots$ are placed in an array, it is the elements on the diagonal which are involved in the construction of the sequence s.

Lecture 4: Uncountable Sets

January 8, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

4.1 Binary Expansion of real numbers

Let $x \in [0,1)$ be given. The key geometric idea is to subdivide $[0,1)$ into 2-equal parts, choose the one which contains our element x, then subdivide this subinterval into 2-equal parts choose the one in which x lies and so on.

Example 4.1 *Let us compute the binary expansion of* $\frac{1}{4}$. *First, we split* $[0,1)$ *into the* 2 *closed-open intervals*

$$\left[0, \frac{1}{2}\right), \left[\frac{1}{2}, 1\right)$$

and we use the 2 digits $0, 1$ *to number them in order. Our number* $\frac{1}{4}$ *belongs to exactly one of these intervals, namely, to* $\left[0, \frac{1}{2}\right)$. *We have labeled this interval with the digit 0, so 0 is the first digit after the decimal point in the binary expansion of* $\frac{1}{4}$.

$$\frac{1}{4} = 0.0\cdots$$

Next, we split the interval $\left[0, \frac{1}{2}\right)$ *into the 2 closed-open intervals*

$$\left[0, \frac{1}{4}\right), \left[\frac{1}{4}, \frac{1}{2}\right)$$

and we use the 2 *digits to number these in order. Our number* $\frac{1}{4}$ *belongs to* $\left[\frac{1}{4}, \frac{1}{2}\right)$ *which is labeled with the digit* 1, *so* 1 *is the second number after the decimal point in the binary expansion of* $\frac{1}{4}$.

$$\frac{1}{4} = 0.01\cdots$$

If we continue this process exactly as we started it, we can obtain the binary expansion of $\frac{1}{4}$ *to as many places as we wish. As a matter of fact, if we do continue, we get* 0 *at each stage from this point on:*

$$\frac{1}{4} = 0.0100000\cdots$$

Remark 4.2 *Notice that contrary to customary usage,* $0.0011111\cdots$ *is not to be regarded as another binary expansion of* $\frac{1}{4}$ *which is "equivalent" to* $0.010000\cdots$. *In our system, each real number* x *in* $[0,1)$ *has one and only one binary expansion which cannot end in an infinite chain of* 1*'s.* ??

4.2 Uncountablity of real numbers

Proposition 4.3 *Let* X *be uncountable and* A *a countable subset of* X*. Then* $X \setminus A$ *is uncountable. Furthermore,* X *and* $X \setminus A$ *have the same cardinality.*

Proposition 4.4 *Let* $B_1 \subsetneq B_2 \subsetneq \cdots$ *be an strictly increasing sequence of finite sets. Then* $\bigcup_{n=1}^{\infty} B_n$ *is countably infinite.*

Example 4.5 *Let* $\Omega = \{s = (a_1, a_2, \cdots) : a_i = 1 \text{ or } 0 \ \forall\, i \in \mathbb{N}\}$. *Now* B *be the subset of* Ω *consisting all the sequences which are eventually constant equal to* 1*. Then* $B = \bigcup_{n=1}^{\infty} B_n$ *where*

$$B_n := \{(a_1, a_2, \cdots) : a_m = 1 \text{ for all } m \geq n\}$$

Clearly each B_n *is a finite set with cardinality* 2^{n-1} *and* $B_1 \subsetneq B_2 \subsetneq B_3, \cdots$. *Hence by Proposition 4.4,* B *is countable.*

Example 4.6 *Show that the unit interval* $[0,1)$ *is uncountable.*

Solution: We have seen that every real number in $[0,1)$ have a unique binary expansion (which cannot end in an infinite chain of 1's). Recall, that the set $A = \{s = (a_1, a_2, \cdots) : a_i = 1 \text{ or } 0 \ \forall\, i \in J\}$ is uncountable. Then in view of Proposition 4.3 and Example 4.5, $A \setminus B$ (where B as defined in Example 4.5) is uncountable and have the same cardinality as A.

Now there is a natural bijection between the set $A \setminus B$ and points in $[0,1)$, it follows that $[0,1)$ is numerically equivalent to Ω. ∎

Remark 4.7 *Sample space which has finitely many outcomes or countably infinite outcomes are called discrete sample space. Whereas the sample space of the form $[0, 1)$ are called continuous space. If the outcomes of a random experiment take continuum of values on the real line we shall refer to the space a continuous.*

What we have discovered that the sample space of an unending sequence of coin tosses is continuous!

Theorem 4.8 (Schroeder-Bernstein) *If X and Y are two sets and $f : X \to Y, g : Y \to X$ are injections then there exists a bijection $h : X \to Y$.*

Example 4.9 *Show that $[0, 1)$ and $(0, 1)$ are numerically equivalent.*

Solution: Take identity map from $(0, 1)$ to $[0, 1)$. It is clearly one-one. Consider function $f : [0, 1) \to (0, 1)$ defined by

$$f(x) = \frac{x}{2} + \frac{1}{4}$$

It is trivial to show that f is one-one. So by Schroeder-Bernstein Theorem, we are done. ∎

Example 4.10 *Let a and b be any two real numbers with $a < b$. Show that $(a, b) \sim (0, 1)$.*

Solution: Define $f : (0, 1) \to (a, b)$ by $f(x) = a + (b - a)x$. Show that f is a bijective map.

Example 4.11 *Show that $\mathbb{R} \sim (a, b)$, where a and b are any two real numbers with $a < b$.*

Solution: Let $f : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$ by $f(x) = \tan x$. Then f is a bijection. By Example 4.10, $(a, b) \sim \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Since the composition of two bijection is again a bijection, we have $\mathbb{R} \sim (a, b)$. ∎

It follows from the Proposition 4.3, that $\mathbb{R} \setminus \mathbb{Q}$ is uncountable and $\#\mathbb{R} = \#\mathbb{R} \setminus \mathbb{Q}$.

Remark 4.12 *Note that rationals and irrationals both are densely spread over the real line but rationals are countably infinite and irrational numbers uncountable.*

Remark 4.13 *Let $a, b \in \mathbb{R}$ such that $a < b$. Then we have just proved that all the intervals $(a, \infty), (-\infty, a), (-\infty, \infty), (a, b), [a, b], [a, b), (a, b]$ have the same number of elements, though $(a, b), [a, b], [a, b), (a, b]$ have length $b - a$ and $(a, \infty), (-\infty, a), (-\infty, \infty)$ a have length infinite.*

So the concept of length is independent of the concept of cardinality. You will be surprise to know that Cantor constructed a subset of $[0, 1]$ whose cardinality is same as of $\mathbb{R}$ but have the length equal to zero.

Lecture 05: Probability Measure

January 11, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

5.1 Class of Events

Definition 5.1 *The elements ω of Ω will be called sample points, each sample point ω identified with an idealised outcome of the underlying random experiment experiment.*

In Example 2.5, Head and Tail are the sample points.

Definition 5.2 *Any subset of a sample space is said to be an event. An event A is said to be occur if the observed outcome ω of the random experiment is an element of the set A.*

Example 5.3 $\{H\}, \{H, T\}$ *are events corresponding to the sample space in Example 2.5.*

Since empty-set is also considered as subset of every set hence this is also a valid event. It is called impossible event. When there is no outcome corresponding some event, that is referred as null event.

Remark 5.4 *H is a sample point and $\{H\}$ is an event.*

It might be desirable that we take the largest possible class of events of a sample space, i.e., power set of Ω. Indeed we are going to take the power set $\mathcal{P}(\Omega)$ if the sample space is finite or countably infinite. But if the sample space is uncountable, the power set is too large to handle. Therefore we need to settle down for a smaller class of events, which is rich enough to contain all the events of our interest. We shall denote in general this class by $\mathcal{F}$.

5.2 The probability measure

So we now begin with an abstract sample space Ω equipped with a class of events $\mathcal{F}$ containing the events of interest to us. The last element in a probability model is to determine a consistent scheme of assigning probabilities to events.

Definition 5.5 *A function P from $\mathcal{F}$ to the set of real-numbers is called a probability measure if it satisfies the following properties*

1. **(Nonnegativity)** $P(A) \geq 0$ *for all* $A \in \mathcal{F}$.
2. **(Normalization)** $P(\Omega) = 1$.
3. **(Countable Additivity)** *If* $A_1, A_2, \cdots$ *is a sequence of* ~~*mutually exclusive events*~~ *in* $\mathcal{F}$ *(i.e.,* $A_i \cap A_j = \emptyset$ *if* $i \neq j$*), then the probability of their union satisfies*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

How do these axioms gel with our experience? Our intuitive assignment of probabilities to results of random experiments is based on an implicit mathematical idealisation of the notion of limiting relative frequency. Suppose A is an event. If, in n independent trials (we use the word "independent" here in the sense that we attach to it in ordinary language) A occurs m times then it is natural to think of the relative frequency m/n of the occurrence of A as a measure of its probability. Indeed, we anticipate that in a long run of trials the relative frequency of occurrence becomes a better and better fit to the "true" underlying probability of A.

As $0 \leq m/n \leq 1$, the positivity and normalisation axioms are natural if our intuition for odds in games of chance is to mean anything at all. The selection of 1 as normalisation constant is a matter of convention.

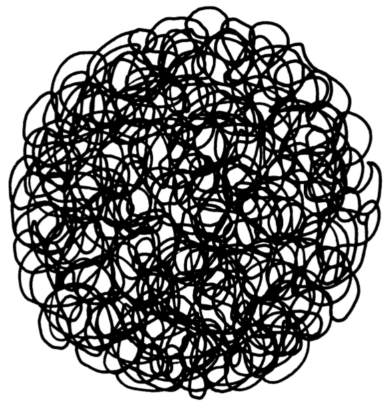
Likewise, from point of relative frequencies, if A and B are two mutually exclusive events and if in n independent trials A occurs m_1 times and B occurs m_2 times, then the relative frequency of occurrence of either A or B in n trials is $(m_1 + m_2)/n = m_1/n + m_2/n$. Probability measure is now forced to be additive if it is to be consistent with experience. $P(A_1)$

When we deal with infinite sample space, countable additivity is a natural property to assume in order to compute probability of certain events which is disjoint countable union. $+P(A_2)$

Probability measure is additive. To visualize countable additivity: Think of probability measure as mass. So it assigns mass 1 unit to the sample space Ω (think of some physical object). Every subset (or piece) of Ω is going have mass ≥ 0. And if we have subspecies $A_1, A_2, \cdots$ which are disjoint, then mass of union of these is sum of all individuals masses.

Probability behaves like length, area, volume. These are natural examples of probability measure. And all the axioms of probability are so natural in these examples.

Example 5.6 *Consider an experiment of throwing a dart on the unit square target. The "area" of a set is a natural candidate of probability measure which satisfies all the axioms. So a smaller subclass $\mathcal{F}$ of events in probability theory is illustrated by the fact that only "nice"* *subset of unite square are measurable in terms of area. Look the following subset which is too jagged, rough, or inaccessible.*



So from purely mathematical point of view, we can not work with power set (the largest class of events) in this case. Hence we have to restrict ourselves to the class of subsets for which we can define area. The collection of those nice sets forms our class $\mathcal{F}$.

Remark 5.7 *Temprature is not an example of probability measure, as additivity fails.*

5.3 Deductions from the Axioms

There are many natural properties of a probability measure, which have not been included in the definition for the simple reason that they can be derived using, the axioms. In this respect the axioms of a mathematical theory are like the constitution of a government. Unless and until it is changed or amended, every law must be made to follow from it. In mathematics we have the added assurance that there are no divergent views as to how the constitution should be construed.

Theorem 5.8 (Properties of Probability Measure) *Suppose $(\Omega, \mathcal{F}, P)$ be a probability space. Then we have the following:*

1. $P(\emptyset) = 0$.
2. **(Finite Additivity)** *If A and B are two mutually exclusive events then show that $P(A \cup B) = P(A) + P(B)$.*
3. *For any event A, $P(A^c) = 1 - P(A)$.*
4. *For any two events such that $A \subset B$, we have $P(B \setminus A) = P(B) - P(A)$.*

5. **(monotonicity)** *For any two events such that $A \subset B$, we have $P(A) \leq P(B)$.*
6. *$P(A) \leq 1$ for any event A.*
7. **(finite sub-additivity)** *For any two events A and B, we have $P(A \cup B) \leq P(A) + P(B)$.*

Proof:

1. Let $A_n = \emptyset$ for all n, then (A_n)'s are disjoint. Therefore by countable additivity

$$P(\emptyset) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P(\emptyset)$$

This is possible iff $P(\emptyset) = 0$.

2. Set $A_1 = A, A_2 = B, A_k = \emptyset$, for $k = 3, 4, \cdots$. Then (A_n) is sequence of mutually exclusive events and $\bigcup\limits_{n=1}^{\infty} A_n = A \cup B$. Therefore by countable additivity we get

$$P(A \cup B) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) = P(A) + P(B) + \sum_{n=3}^{\infty} P(\emptyset) = P(A) + P(B)$$

3. $A \cup A^c = \Omega$. Also A and A^c are mutually exclusive events, hence using normalization and finite additivity of probability measure we are done.

4. Note that $B = A \cup (B \setminus A)$. Now A and $B \setminus A$ are mutually exclusive. Using finite additivity

5. Using previous one and non-negativity we are done.

6. Since $A \subset \Omega$, hence by normlization axiom and monotonicity we are done.

7. Note that $A \cup B = A \cup (B \setminus (A \cap B))$. Now A and $B \setminus (A \cap B)$ are mutually exclusive. Also from 2 above $P(B \setminus (A \cap B)) = P(B) - P(A \cap B)$. So we obtain

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Now by non-negativity gives the desired result.

∎

Lecture 06: Probability Measure in finite Sample Space

January 12, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Theorem 6.1 (Continuity of Probability Measure) *Let $A_n, n \geq 1$ be events in $\mathcal{F}$.*

(a) If $A_1 \subset A_2 \subset \cdots$ Then $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} P(A_k)$.

(b) If $A_1 \supset A_2 \supset \cdots$ Then $P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} P(A_k)$.

Proof:

1. Suppose $A_1 \subset A_2 \subset \cdots$ and $A := \bigcup_{k=1}^{\infty} A_k$. Set $B_1 = A_1$, and for each $n \geq 2$, let B_n denote those points which are in A_n but not in A_{n-1} , i.e., $B_n = A_n \setminus A_{n-1}$. By definition, the sets B_n are disjoint. Also $A_n = \bigcup_{k=1}^{n} B_k$ and $A = \bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k$. Hence

$$P(A_n) = \sum_{k=1}^{n} P(B_k)$$

Since the left side above cannot exceed 1 for all n, $P(B_k) \geq 0$ for all k, so sequence of partial sums is increasing and bounded above hence the series on the right side must converge. Hence we obtain

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) =: \sum_{k=1}^{\infty} P(B_k) = P(A). \tag{6.1}$$

2. Now if $A_1 \supset A_2 \supset \cdots$ Then $A_1^c \subset A_2^c \subset \cdots$. Hence by part (a),

$$P\left(\bigcup_{k=1}^{\infty} A_k^c\right) = \lim_{k\to\infty} P(A_k^c)$$

$$1 - P\left[\left(\bigcup_{k=1}^{\infty} A_k^c\right)^c\right] = \lim_{k\to\infty} [1 - P(A_k)]$$

$$1 - P\left(\bigcap_{k=1}^{\infty} A_k\right) = 1 - \lim_{k\to\infty} P(A_k)$$

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} P(A_k)$$

■

6.1 Examples of Probability Space

A probability space has three elements Ω the sample space, $\mathcal{F}$-collection of events, P the probability measure. Choice of sample space is relatively easy and clear from underlying random experiment. Choosing appropriate class $\mathcal{F}$ which contains events of interest is also a doable task. Defining a probability function is not at all an easy task and definition makes no attempt to tell what particular function P to choose; it merely requires P to satisfy the axiom. As we shall see that for a sample space with same class of events we may define many different probability measure.

Now we illustrate how to construct probability spaces starting from some common sense assumptions about the random experiment.

6.2 Finite Sample Space

If the sample space consists of a finite number of possible outcomes, then we can take $\mathcal{F}$ to be the largest class, the power set of the sample space. The probability measure is specified by the probabilities of each single outcome. Let $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ for some $n \in \mathbb{N}$. Then choose numbers $p_1, p_2, \cdots, p_n$ such that

$$p_i \geq 0, \quad \text{and} \quad \sum_{i=1}^{n} p_i = 1.$$

Define $P(\{\omega_i\}) = P(\omega_i) = p_i$ for each $i = 1, 2, \cdots, n$. If A is any event in $\mathcal{F}$, then $P(A) = \sum_{\omega \in A} P(\omega)$. Then P is a probability measure. Indeed,

1. $P(A)$ is sum of $P(\omega_i)$'s such that $\omega_i \in A$ and $P(\omega_i) = p_i \geq 0$ hence, there sum is also non-negative.

2. $P(\Omega) = \sum_{i=1}^{n} P(\omega_i) = \sum_{i=1}^{n} p_i = 1.$

3. If $A_1, A_2, \cdots$ is a sequence of pairwise disjoint events, then

$$P(\cup_k A_k) = \sum_{\omega \in \cup_k A_k} P(\omega) = P(A_1) + P(A_2) + \cdots$$

Example 6.2 *In a coin tossing experiment, let us assign the probability p of coming head and $1-p$ of coming tail such that $0 \leq p \leq 1$. Then this assignment satisfy all the properties of probability measure. Hence, we see that there are infinitely many ways to define a probability law on a sample space.*

6.3 Equally likely outcomes or Discrete Uniform Probability law

There is a special probability law when the sample space is finite. It is based on symmetry. Also in absence any prior knowledge it is natural to assume all the outcomes have same probability. If we assume that all the outcomes $\{\omega_1, \omega_2, \cdots, \omega_n\}$ are "equally likely" or they have same chance of occurring then $P(\omega_i) = \frac{1}{n}$ for each $i = 1, 2, \cdots, n$. This probability assignment is called discrete uniform probability law. You must realize by now that whatever probability theory you people have done in your earlier classes, it falls into this setup. Recall your definition of the probability of an event A,

$$P(A) = \frac{\text{Number of outcomes favourable to } A}{\text{Number of all possible outcomes of the experiment.}}$$

How is one going to decide whether the otcomes are equally likely or not?

A celebrated example will illustrate this.

Example 6.3 *Let two coins be tossed. D'Alembert (mathematician, philosopher, and encyclopedist, 1717–83) argued that there are three possible cases, namely:*
(i) both heads, (ii) both tails, (iii) a head and a tail.
So he went on to conclude that the probability of "a head and a tail" is equal to 1/3. If he had figured that this probability should have something to do with the relative frequency of the occurrence of the event, he might have changed his mind after tossing two coins more than a few times.

The three cases he considered are not equally likely. Case (iii) should be split into two:
(iiia) first coin shows head and second coin shows tail.
(iiib) first coin shows tail and second coin shows head.
It is the four cases (i), (ii), (iiia) and (iiib) that are equally likely by symmetry and on empirical evidence. This should be obvious if we toss the two coins one after the other rather than simultaneously. However, there is an important point to be made clear here. The two coins may be physically indistinguishable so that in so far as actual observation is concerned, D'Alembert's three cases are the only distinct patterns to be recognized. In the model of two coins they happen not to be equally likely on the basis of common sense and experimental evidence.

But in an analogous model for certain microcosmic particles, called Bose–Einstein statistics (see below) they are indeed assumed to be equally likely in order to explain some types of physical phenomena.

6.3.1 An Example from Statistical Mechanics

Given n cells and r particle ($n > r$), we place at random each particle in one of cell. We wish to find the probability p that in r preselected cell, one and only one particle will be found. Suppose $n = 6$ and $r = 2$. This problem can be stated in terms of a pair of dice: Then $n = 6$ faces corresponds to n cells and $r = 2$ dice to the r partciles. We assme that the preselected faces (cells) are 3 and 4. The solution of this problem depends on the choice of possible and favorable outcomes. We shall consider the following three cases:

1. **particles are distingushible:** In this case we distinguish between $(3, 4)$ and $(4, 3)$ as different outcomes. There are total 36 outcomes. Therefore probability of getting 3, 4 in the game of two dice is $\frac{2}{36}$.

2. **particles are not distingushible:** In this case we count $(3, 4)$ and $(4, 3)$ as same outcome. Theare are 21 possible outcomes:

   $$(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,2),(2,3),(2,4),(2,5),(2,6),(3,3),(3,4),(3,5),$$
   $$(3,6),(4,4),(4,5),(4,6),(5,5),(5,6),(6,6)$$

   Therefore probability of getting 3, 4 in the game of two dice is $\frac{1}{21}$.

3. **partcile are not distingushible and any cell can not have more than one particle** We need to throw those outcomes from previous case where both dice shows the same outcome. There are 15 possible outcomes:

   $$(1,2),(1,3),(1,4),(1,5),(1,6),(2,3),(2,4),(2,5),(2,6),(3,4),(3,5),$$
   $$(3,6),(4,5),(4,6),(5,6),$$

   Therefore probability of getting 3, 4 in the game of two dice is $\frac{1}{15}$.

Statistical Mechanics: Consider a mechanical system of r particles. In statistical mechanics it is usual to subdivide the phase space into a large number, n, of small regions or cells so that each particle is assigned to one cell. In this way the state of the entire system is described in terms of a random distribution of the r particles in n cells. Then we wishe to compute the probability p that in r preselected cell, one and only one particle will be found.

1. **Maxwell-Boltzmann Statistics:** If we accept as outcomes all possible ways of placing r particles in n cells distinguishing the identity of each particle, then

   $$p = \frac{r!}{n^r}$$

   Modern theory has shown beyond doubt that this statistics does not apply to any known particles (atomic level).

2. **Bose-Einstein Statistics:** If we assume that the particles are not distinguishable, then

   $$p = \frac{(n-1)!r!}{(n+r-1)!}$$

   It is shown in statistical mechanics that photons, nuclei, and atoms containing an even number of elementary particles, follows this model.

3. **Fermi-Dirac Statistics:** If we assume that the particles are not distinguishable and in each cell, we allowed to place at most one particle, then

   $$p = \frac{r!(n-r)!}{n!}$$

   This model successfully explains the behavior of electrons, neutrons, and protons.

Let $n = 5, r = 3$. The arrangement $(*| - | * | * |-)$ has probability $\frac{6}{125}, \frac{1}{35}$, or $\frac{1}{10}$, according to whether Maxwell-Boltzmann, Bose-Einstein, or Fermi-Dirac statistics is used .

Lecture 07: Probability Measure in Countable and Uncountable Sample Space

January 13, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

7.1 Countable Sample Space

We shall show how easy it is to construct probability measures for any countable space $\Omega = \{\omega_1, \omega_2, \cdots\}$. Once again we can work with any σ-field $\mathcal{F}$ (including the largest one, the power set of the sample space). The probability measure is specified by the probabilities of each single outcome. Each sample point ω_n has probability p_n subject to the conditions

$$\forall\ n : p_n \geq 0,\ \sum_{n=1}^{\infty} p_n = 1. \tag{7.1}$$

In symbols, we write $P(\{\omega_n\}) = P(\omega_n) = p_n$ for all n. Now for any subset A of Ω, we define its probability to be

$$P(A) := \sum_{\omega \in A} P(\omega).$$

Once again it is very easy to verify all the axioms of probability function.

Example 7.1 *In the random experiment of tossing a coin till we get a head, the sample space is*

$$\Omega = \{H, TH, TTH, TTTH, \cdots\}.$$

It clear that sample space is countably infinite.

Choose $p \in (0,1)$. We assign probability

$$P(H) = p$$
$$P(TH) = (1-p)p$$
$$P(TTH) = (1-p)^2 p$$

Then $\sum_{n=1}^{\infty}(1-p)^{n-1}p = p\sum_{n=1}^{\infty}(1-p)^{n-1} = p\dfrac{1}{1-(1-p)} = 1$, hence it defines a valid candidate for probability measure. This is the geometric distribution with parameter p.

In fact, if $(a_n)_{n\geq 1}$ is sequence of non-negative terms such that $\sum_{n=1}^{\infty} a_n < \infty$. Then define $S = \sum_{n=1}^{\infty} a_n$, which is a positive real number. Now define probability of each outcome as follows:

$$P(\omega_n) = \frac{a_n}{S}, \quad \text{for each } n = 1, 2, \cdots$$

Then we get a probability measure. As a particular example, recall that $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges for every real number $p > 1$. Let us denote the sum by S_p. Then for every $p > 1$, we define a probability measure as follows:

$$P(\omega_n) = \frac{1}{S_p n^p}, \quad \text{for each } n = 1, 2, \cdots$$

So we have defined uncountably many probability measures on the same sample space and same class of events.

Example 7.2 *Suppose $\Omega = \{0, 1, 2, \cdots, \}$. Suppose $\lambda > 0$. Then a probability measure μ given by*

$$\mu(\{k\}) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \cdots,$$

is called a Poisson distribution with parameter λ.

Exercise 7.3 *Show that all outcomes of a countable sample space can not be equally likely.*

Solution: Suppose the contrary, i.e., there exists $p \in [0, 1]$ such that $P(\omega_n) = p$ for all $n = 1, 2, \cdots$. Now by countable additivity

$$P(\Omega) = \sum_{n=1}^{\infty} P(\omega_n) = \sum_{n=1}^{\infty} p = 1.$$

This not possible for any choice of p, which is a contradiction. ■

7.2 Uncountable Sample Space

The situation is more complicated when we deal with uncountable sample space. It is not possible to define probability for every subset of sample space, and at the same time be consistent with the axioms of probability function and it's consequences. We have already seen an example where we can not talk about area of every subset of a unit square.

We know that every uncountable set has a countable subset, hence one may jump to the conclusion from Exercise 7.3, that in an uncountable sample space, all outcomes can not be equally likely. But this is not the case, for example, consider a random experiment of picking a number 'at random' (there is no bias towards any number or all numbers have equal chance of getting picked up) from interval $[0, 1]$. Then each individual number ω has zero probability to get picked (if $P(\omega)$ is positive for each $\omega \in [0, 1]$ then probability of every countable subset of $[0, 1]$ will be $+\infty$). But this does not give any contradiction like $P([0, 1]) = 0$ because uncountable sums are not well defined (even if we are adding zero's only!!). Also, the information that $P(\omega) = 0$ for each $\omega \in [0, 1]$ is not suffices to define probability of any interval $I \subset [0, 1]$. So the idea is we directly define the probabilities of certain fundamental events like intervals and then appeal to the property of probability measure to compute the probability of other events of our interest.

Example 7.4 *We construct a mathematical model for choosing a number at random from the unit interval $[0, 1]$ so that the probability is distributed uniformly over the interval. We define the probability of closed intervals $[a, b]$ by the formula*

$$P([a, b]) = b - a, 0 \leq a \leq b \leq 1 \tag{7.2}$$

(i.e., the probability that the number chosen is between a and b is $b - a$). If $b = a$, then $[a, b]$ is the set containing only the number a, and (7.2) says that the probability of this set is zero (i.e. , the probability is zero that the number we choose is exactly equal to a) . Because single points have zero probability, the probability of an open interval (a, b) is the same the probability of the closed interval $[a, b)$; we have

$$P\{(a, b)\} = b - a, 0 \leq a \leq b \leq 1$$

. *There are many other subsets of $[0, 1]$ whose probability is determined by the formula (7.2) and the properties of probability measures. For example, the set $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ is not an interval, but we know from finite additivity that its probability is $\frac{2}{3}$.*

Also let $A = \mathbb{Q} \cap [0, 1]$, then A can be written as countable union of disjoint events of all rationals in $[0, 1]$. Hence $P(A) = 0$. This implies that event $B = \mathbb{Q}^c \cap [0, 1]$ which is the set of irrational numbers in $[0, 1]$, probability 1, because

$$P(B) = 1 - P(A) = 1.$$

This shows that the set of irrational in interval $[0, 1]$ has length 1!! As mentioned earlier that Cantor set (which is subset of $[0, 1]$) has zero length (you have to take it on belief!!). This tell us that uncountable infinite sets can have any probability between $[0, 1]$ including zero and one. Uncountability does not help in determining its probability.

Remark 7.5 *In Example 7.4, simple events $\{\omega\}$ have probability zero, but when we choose a number from $[0, 1]$ we do get a particular outcome as ω. Hence events with zero probability do occur.*

Remark 7.6 *Area, volume are examples of uniform probability distributions on unit square and unit cube respectively.*

We know very well, there are mass distributions that are typically nonuniform. For example distribution of snow in region.

Example 7.7 *Let $\Omega = [0, \infty)$. For subinterval $[a, b)$ define the probability as*

$$P\{[a,b)\} = e^{-a} - e^{-b}, \quad 0 \leq a < b \leq \infty.$$

Once again it is easy to verify the axioms of probability function for class of intervals of the form $[a, b)$.

1. *If $a < b$ then $-a > -b$ and e^x is an strictly increasing function, hence $e^{-a} > e^{-b}$. Hence we have non-negativity.*
2. *$P(\Omega) = 1$.*

Now we use the continuity property of the probability measure to compute the probability of a single outcome.

$$P\{a\} = P\left\{\bigcap_{n=1}^{\infty}\left[a, a+\frac{1}{n}\right)\right\} = \lim_{n\to\infty} P\left(\left[a, a+\frac{1}{n}\right)\right) = \lim_{n\to\infty} e^{-a} - e^{-a-1/n} = 0, \forall a \in \Omega$$

Hence it follows that for $0 \leq a < b < \infty$,

$$P\{[a,b]\} = P\{(a,b)\} = P\{(a,b]\} = P\{[a,b)\}.$$

Note that $P([1,2]) = e^{-1} - e^{-2} = 0.232544158$ $P([3,4]) = e^{-3} - e^{-4} = 0.031471429$. This shows that the distribution of the mass is not uniform over the interval $[0, \infty)$.

Lecture 08: Probability Density function & Conditional Probability

January 15, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

8.1 Probability Density function

If the outcomes of a conceptual probability experiment take a continuum of values on the real line we shall refer to the space as continuous and identify the sample space Ω with the real line $\mathbb{R} = (-\infty, \infty)$. In this setting our focus is naturally on intervals (rather individual points) which now become the basic objects of interest, events which special subsets of the line.

Definition 8.1 *Let f be any positive and integrable function defined on the real line $\mathbb{R}$ and normalised so that $\int_{-\infty}^{\infty} f(x)dx = 1$. We call any such function a probability density function.*

With each such function we may associate the probability map $A \to P(A)$ which, to each subset A of the line of interest, assigns the value

$$P(A) = \int_A f(x)dx. \tag{8.1}$$

We shall interpret integrals of the above form as Riemann integrals or improper integrals. There is surely no difficulty in the interpretation of the integral if f is piecewise continuous and A is simply an interval or a finite (or even countable) union of disjoint intervals. This covers most applications and requires little mathematical abstraction.

Thus, if I is any of the four types of intervals $(a, b), [a, b), [a, b]$, or $(a, b]$, we obtain $(I) = \displaystyle\int_I f(x)dx = \int_a^b f(x)dx$, the integral being insensitive to the addition or deletion of points on the boundary, while, if $\{I_n, n \geq 1\}$ is a countable collection of pairwise disjoint intervals then $P\left(\bigcup I_n\right) = \displaystyle\sum_{n=1}^{\infty} \int_{I_n} f(x)dx$ as integration is additive. It follows that the constructed set function P is indeed non-negative, and countably additive at least when restricted to intervals and finite or countable unions of them.

The value $f(x)$ assumed by f at a given point x of the line is the probability mass per unit length that is attached at that point. The function f is hence called a probability density (or simply density in short). The intuition very much parallels what one sees in physics when one progresses from point masses to a continuum of mass.

Example 8.2 *Consider the following function*

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{if} \quad 0 < x \leq 1 \\ 0 & \quad \textit{otherwise} \end{cases}$$

Is it a valid pdf? Even though f becomes arbitrarily large as x approaches zero, this is still a valid pdf, because Clearly $f \geq 0$ and

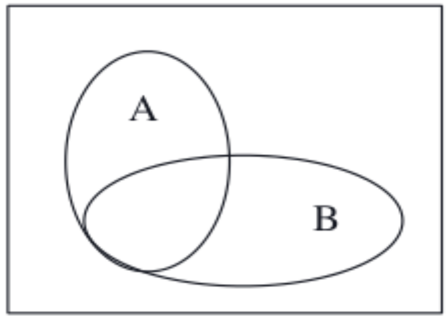
$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_0^1 \frac{1}{2\sqrt{x}}dx \\ &= \sqrt{x}\Big|_0^1 = 1 \end{aligned}$$

It is important to realize that even though the pdf is used to calculate probabilities of events but $f(x)$ is not the probability of point x, it is the rate at which probability mass accumulates near point x.

8.2 Conditional Probability

We know something about the world and based on what we know when we set up a probability model. Then something happens, and somebody tells us a little more about the world, gives us some new information. This new information, in general, should change our beliefs about what happened or what may happen. So whenever we're given new information, some partial information about the outcome of the experiment, we should revise our beliefs. And conditional probabilities are just the probabilities that apply after the revision of our beliefs, when we're given some information.

In more precise terms, given an experiment, a corresponding sample space, and a probability law, suppose that we know that the outcome is within some given event A. We wish to quantify the likelihood that the outcome also belongs to some other given event B. We thus seek to construct a new probability law that takes into account the available knowledge: a probability law that for any event B, specifies the conditional probability of B given A, denoted by $P(B|A)$. We would like the conditional probabilities $P(B|A)$ of different events B to constitute a legitimate probability law, which satisfies the probability axioms.



Saying that A has occurred means outcome already lies in A. So A is our new universe. Now we ask what is the probability that outcome is in set B given it is already in A. Think of probability measure as area or mass. It is clear that it should area or mass of $A \cap B$. Now we want this new assignment of probability to follow our axioms of probability, one of them is probability of sample sapce is 1, since our new universe or sample space is set A, in order to have $P(A|A) = 1$ we divide by $P(A)$. This motivate the definition.

Definition 8.3 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A \in \mathcal{F}$ be such that $P(A) > 0$. The conditional probability of an event $B \in \mathcal{F}$ given A is denoted by $P(B|A)$ is defined as*

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

Define

$$\mathcal{F}_A := \{B \cap A | B \in \mathcal{F}\}$$

So $\mathcal{F}_A$ is class (or family or collection or set) of subsets of A.

Define P_A on $\mathcal{F}_A$ as follows:

$$P_A(B) = P(B|A), B \in \mathcal{F}_A.$$

Then $(A, \mathcal{F}_A, P_A)$ is a probability space. We are already assuming $P(A) > 0$ which implies A is non-empty. Now only thing left is to show P_A is a probability measure.

1. $P(B|A) \geq 0$ by definition.
2. $P_A(A) = P(A|A) = 1$.
3. Countable additivity of P_A follows from countable additivity of P. Let $C_1, C_2, \cdots \in \mathcal{F}_A$ be sequence of pairwise disjoint events. Then there exists $B_1, B_2, \cdots \in \mathcal{F}$ a sequence

of pairwise disjoint events such that $C_n = A \cap B_n$ for all $n \in \mathbb{N}$. Note that $\bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$ and

$$\begin{aligned}
P_A\left(\bigcup_{n=1}^{\infty} C_n\right) &= P\left(\bigcup_{n=1}^{\infty} C_n \;\Big|\; A\right) = \frac{P\left\{\left(\bigcup_{n=1}^{\infty} C_n\right) \bigcap A\right\}}{P(A)} = \frac{P\left\{\bigcup_{n=1}^{\infty} (C_n \bigcap A)\right\}}{P(A)} \\
&= \frac{P\left\{\bigcup_{n=1}^{\infty} (B_n \bigcap A)\right\}}{P(A)} = \frac{\sum_{n=1}^{\infty} P\left\{B_n \bigcap A\right\}}{P(A)} = \sum_{n=1}^{\infty} \frac{P\left\{B_n \bigcap A\right\}}{P(A)} \\
&= \sum_{n=1}^{\infty} P_A(C_n)
\end{aligned}$$

Remark 8.4 *Let us also note that since we have $P(A|A) = 1$, all of the conditional probability is concentrated on A. Thus, we might as well discard all possible outcomes outside A and treat the conditional probabilities as a probability law defined on the new universe A.*

Since conditional probabilities constitute a legitimate probability law, all general properties of probability measure remain valid.

Example 8.5 *Consider an experiment involving two successive rolls of a fair die. If the sum of the two rolls is 9, how likely is it that the first roll was a 6?*

Solution: Let B be event that the sum of the two rolls is 9. Then $B = \{(3,6),(6,3),(4,5),(5,4)\}$. Let A be the event that the first roll is 6, then A has 6 elements. Now

$$P(A|B) = \frac{P\{(6,3)\}}{P(B)} = \frac{\frac{1}{36}}{\frac{4}{36}} = \frac{1}{4}$$

■

In Example 8.5 we considered, the probability space was specified, and we computed conditional probability. In many problems however, we actually proceed in the opposite direction. We are given in advance what we want some conditional probabilities to be, and we use this information and the rules of probabilities to compute the requested probabilities. A typical example of this situation is the following.

Example 8.6 *Suppose that the population of a certain city is 40% male and 60% female. Suppose also that 50% of the males and 30% of the females smoke. Find the probability that a smoker is male.*

Solution: Let M denote the event that a person selected is a male and let F denote the event that the person selected is a female. Also let S denote the event that the person selected smokes. The given information can be expressed in the form $P(S|M) = .5, P(S|F) = .3, P(M) = .4$, and $P(F) = .6$. The problem is to compute $P(M|S)$. By definition

$$P(M|S) = \frac{P(M \cap S)}{P(S)}.$$

Now $P(M \cap S) = P(M)P(S|M) = (.4)(.5) = 0.20$, so the numerator can be computed in terms of the given probabilities. Now how to compute $P(S)$? Here is a very important technique in probability known as total probability theorem. This is known as "divide and rule", divide or partition the set S so that it is possible for us to compute the probability of those partitioned events. Note that S is the union of the two disjoint sets $S \cap M$ and $S \cap F$. it follows that

$$P(S) = P(S \cap M) + P(S \cap F).$$

Since $P(S \cap F) = P(F)P(S \cap F) = (.6)(.3) = .18$. Therefore, $P(S) = 0.38$. Hence

$$P(M|S) = \frac{0.20}{0.38} = 0.526315789$$

■

15.4 Double Integrals in Polar Form

Integrals are sometimes easier to evaluate if we change to polar coordinates. This section shows how to accomplish the change and how to evaluate integrals over regions whose boundaries are given by polar equations.

Integrals in Polar Coordinates

When we defined the double integral of a function over a region R in the xy-plane, we began by cutting R into rectangles whose sides were parallel to the coordinate axes. These were the natural shapes to use because their sides have either constant x-values or constant y-values. In polar coordinates, the natural shape is a "polar rectangle" whose sides have constant r- and θ-values.

Suppose that a function $f(r, \theta)$ is defined over a region R that is bounded by the rays $\theta = \alpha$ and $\theta = \beta$ and by the continuous curves $r = g_1(\theta)$ and $r = g_2(\theta)$. Suppose also that $0 \le g_1(\theta) \le g_2(\theta) \le a$ for every value of θ between α and β. Then R lies in a fan-shaped region Q defined by the inequalities $0 \le r \le a$ and $\alpha \le \theta \le \beta$. See Figure 15.21.

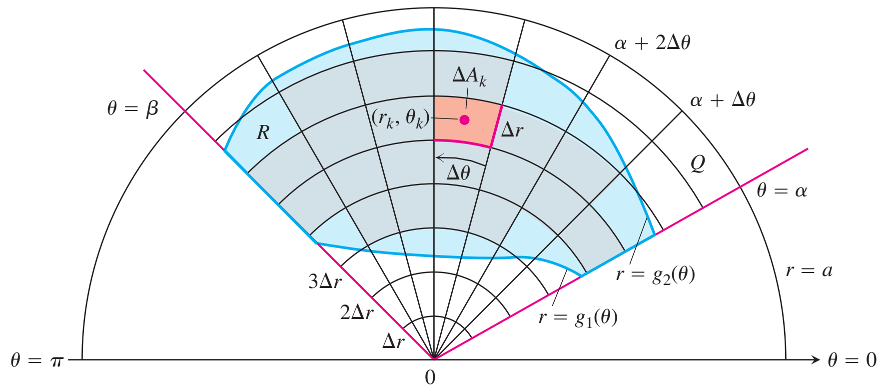


FIGURE 15.21 The region R: $g_1(\theta) \le r \le g_2(\theta)$, $\alpha \le \theta \le \beta$, is contained in the fan-shaped region Q: $0 \le r \le a$, $\alpha \le \theta \le \beta$. The partition of Q by circular arcs and rays induces a partition of R.

We cover Q by a grid of circular arcs and rays. The arcs are cut from circles centered at the origin, with radii $\Delta r, 2\Delta r, \ldots, m\Delta r$, where $\Delta r = a/m$. The rays are given by

$$\theta = \alpha, \qquad \theta = \alpha + \Delta\theta, \qquad \theta = \alpha + 2\Delta\theta, \qquad \ldots, \qquad \theta = \alpha + m'\Delta\theta = \beta,$$

where $\Delta\theta = (\beta - \alpha)/m'$. The arcs and rays partition Q into small patches called "polar rectangles."

We number the polar rectangles that lie inside R (the order does not matter), calling their areas $\Delta A_1, \Delta A_2, \ldots, \Delta A_n$. We let (r_k, θ_k) be any point in the polar rectangle whose area is ΔA_k. We then form the sum

$$S_n = \sum_{k=1}^{n} f(r_k, \theta_k)\, \Delta A_k.$$

If f is continuous throughout R, this sum will approach a limit as we refine the grid to make Δr and $\Delta\theta$ go to zero. The limit is called the double integral of f over R. In symbols,

$$\lim_{n\to\infty} S_n = \iint_R f(r, \theta)\, dA.$$

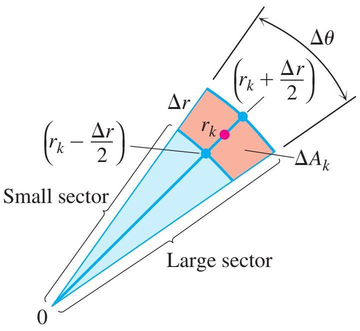


FIGURE 15.22 The observation that

$$\Delta A_k = \begin{pmatrix}\text{area of}\\ \text{large sector}\end{pmatrix} - \begin{pmatrix}\text{area of}\\ \text{small sector}\end{pmatrix}$$

leads to the formula $\Delta A_k = r_k\,\Delta r\,\Delta\theta$.

To evaluate this limit, we first have to write the sum S_n in a way that expresses ΔA_k in terms of Δr and $\Delta\theta$. For convenience we choose r_k to be the average of the radii of the inner and outer arcs bounding the kth polar rectangle ΔA_k. The radius of the inner arc bounding ΔA_k is then $r_k - (\Delta r/2)$ (Figure 15.22). The radius of the outer arc is $r_k + (\Delta r/2)$.

The area of a wedge-shaped sector of a circle having radius r and angle θ is

$$A = \frac{1}{2}\theta \cdot r^2,$$

as can be seen by multiplying πr^2, the area of the circle, by $\theta/2\pi$, the fraction of the circle's area contained in the wedge. So the areas of the circular sectors subtended by these arcs at the origin are

$$\text{Inner radius:}\quad \frac{1}{2}\left(r_k - \frac{\Delta r}{2}\right)^2 \Delta\theta$$

$$\text{Outer radius:}\quad \frac{1}{2}\left(r_k + \frac{\Delta r}{2}\right)^2 \Delta\theta.$$

Therefore,

$$\begin{aligned}\Delta A_k &= \text{area of large sector} - \text{area of small sector}\\ &= \frac{\Delta\theta}{2}\left[\left(r_k + \frac{\Delta r}{2}\right)^2 - \left(r_k - \frac{\Delta r}{2}\right)^2\right] = \frac{\Delta\theta}{2}(2r_k\,\Delta r) = r_k\,\Delta r\,\Delta\theta.\end{aligned}$$

Combining this result with the sum defining S_n gives

$$S_n = \sum_{k=1}^{n} f(r_k, \theta_k) r_k\,\Delta r\,\Delta\theta.$$

As $n \to \infty$ and the values of Δr and $\Delta\theta$ approach zero, these sums converge to the double integral

$$\lim_{n\to\infty} S_n = \iint_R f(r, \theta)\, r\,dr\,d\theta.$$

A version of Fubini's Theorem says that the limit approached by these sums can be evaluated by repeated single integrations with respect to r and θ as

$$\iint_R f(r, \theta)\,dA = \int_{\theta=\alpha}^{\theta=\beta}\int_{r=g_1(\theta)}^{r=g_2(\theta)} f(r, \theta)\, r\,dr\,d\theta.$$

Finding Limits of Integration

The procedure for finding limits of integration in rectangular coordinates also works for polar coordinates. To evaluate $\iint_R f(r, \theta)\,dA$ over a region R in polar coordinates, integrating first with respect to r and then with respect to θ, take the following steps.

1. *Sketch*. Sketch the region and label the bounding curves (Figure 15.23a).
2. *Find the r-limits of integration*. Imagine a ray L from the origin cutting through R in the direction of increasing r. Mark the r-values where L enters and leaves R. These are the r-limits of integration. They usually depend on the angle θ that L makes with the positive x-axis (Figure 15.23b).
3. *Find the θ-limits of integration*. Find the smallest and largest θ-values that bound R. These are the θ-limits of integration (Figure 15.23c). The polar iterated integral is

$$\iint_R f(r, \theta)\,dA = \int_{\theta=\pi/4}^{\theta=\pi/2}\int_{r=\sqrt{2}\csc\theta}^{r=2} f(r, \theta)\, r\,dr\,d\theta.$$

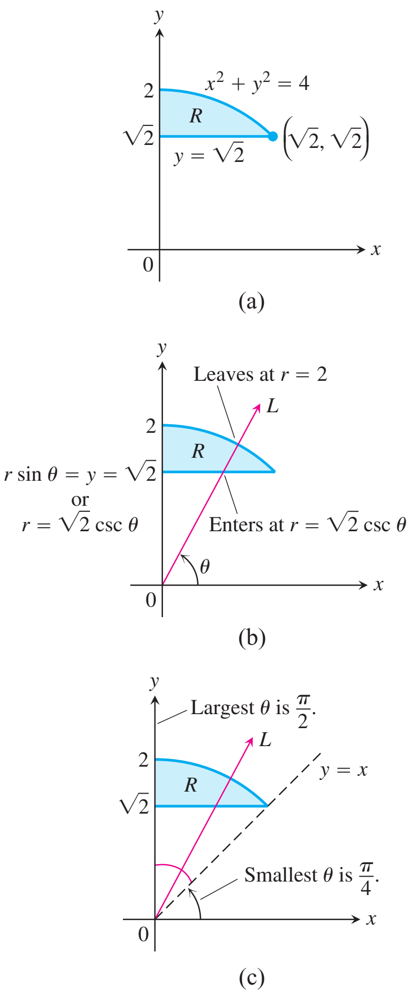


FIGURE 15.23 Finding the limits of integration in polar coordinates.

Lecture 09: Total Probability Theorem and Baye's Theorem

January 18, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Generally speaking, most problems of probability have to do with several events and it is their mutual relation or joint action that must be investigated. The following result is useful for such situation.

Proposition 9.1 *For arbitrary events $A_1, A_2, \cdots, A_n$, we have*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cdots \cap A_{n-1}) \quad (9.1)$$

provided $P(A_1 \cap A_2 \cdots A_{n-1}) > 0$.

Proof: Since

$$P(A_1) \geq P(A_1 \cap A_2) \cdots \geq P(A_1 \cap A_2 \cdots A_{n-1}) > 0,$$

therefore, all the conditional probabilities in (9.1) are well-defined. Now the right side of (9.1) is

$$P(A_1) \times \frac{P(A_2 \cap A_1)}{P(A_1)} \times \frac{P(A_3 \cap A_2 \cap A_1)}{P(A_2 \cap A_1)} \times \cdots \times \frac{P(A_1 \cap A_2 \cdots \cap A_{n-1} \cap A_n)}{P(A_1 \cap A_2 \cdots \cap A_{n-1})}$$

■

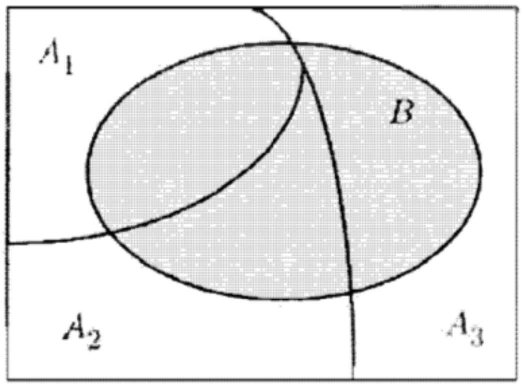
Definition 9.2 *A collection $\{A_1, A_2, \cdots, A_N\}$ of events is said to be a partition of Ω if*

1. *A_i's are pairwise disjoint.*
2. $\displaystyle\bigcup_{i=1}^{N} A_i = \Omega$

If $N < \infty$ then partition is said to be finite partition and if $N = \infty$, then it is called a countable partition.

Example 9.3 *If $\Omega = \mathbb{N}$ then $\{E, O\}$ where E is collection even numbers and O is set of odd numbers. This is a finite partition. If take $\{\{1\}, \{2\}, \cdots\}$ as partition then it is a countable partition.*

Now we explore some further applications of conditional probability. The following theorem, which is often useful for computing the probabilities of various events, using a "divide-and-conquer" approach.



Theorem 9.4 (Total Probability Theorem) *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_1, A_2, \cdots, A_N\}$ be a partition of Ω such that $P(A_i) > 0$ for all i. Then for any event $B \in \mathcal{F}$,*

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

Proof: Event B is decomposed into the disjoint union

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots (A_N \cap B)$$

Now using additivity of probability measure we have

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

By definition of conditional probability $P(B \cap A_i) = P(B|A_i)P(A_i)$. Hence we get the theorem. ■

One of the uses of the theorem is to compute the probability of various events B for which the conditional probabilities $P(B|A_i)$ are known or easy to derive. The key is to choose appropriately the partition $\{A_1, A_2, \cdots, A_N\}$ and this choice is often suggested by the structure of the problem.

Example 9.5 *You enter a chess tournament where your probability of winning a game is* 0.3 *against half the players (call them type 1),* 0.4 *against a quarter of the players (call them type 2), and* 0.5 *against the remaining quarter of the players (call them type 3). You play a game against a randomly chosen opponent. What is the probability of winning?*

Solution: Let A_i be the event of playing with an opponent of type i. We have

$$P(A_1) = 0.5, P(A_2) = 0.25, P(A_3) = 0.25$$

Also, let B be the event of winning. We have

$$P(B|A_1) = 0.3, P(B|A_2) = 0.4, P(B|A_3) = 0.5$$

Thus, by the total probability theorem, the probability of winning is

$$\begin{aligned} P(B) &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) \\ &= 0.5 \times 0.3 + 0.25 \times 0.4 + 0.25 \times 0.5 \\ &= 0.375. \end{aligned}$$

■

The total probability theorem is often used in conjunction with the following celebrated theorem, which relates conditional probabilities of the form $P(A|B)$ with conditional probabilities of the form $P(B|A)$, in which the order of the conditioning is reversed.

Theorem 9.6 (Bayes Theorem) *Let $\{A_1, A_2, \cdots, A_N\}$, be a partition of the sample space, and assume that $P(A_i) > 0$, for all $i = 1, 2, \cdots, N$. Then, for any event B such that $P(B) > 0$, we have*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{N} P(A_k)P(B|A_k)}, \quad \textit{for each } i = 1, 2, \cdots N.$$

Proof: For fixed i,

$$P(A_i|B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{N} P(A_k)P(B|A_k)}$$

where the last equality follows from Total probability theorem. ■

Bayes rule is often used for inference. There are a number of causes that may result in certain effect. We observe the effect and wish to infer the cause. The events $A_1, A_2, \cdots, A_N$ is associated with the causes and event B represents the effect. The probability $P(B|A_i)$ that the effect will be observed when the cause A_i is present amounts to a probabilistic model of the cause-effect relation. Given that effect B has been observed, we wish to evaluate the probability $P(A_i|B)$ that cause A_i is present. We refer to $P(A_i|B)$ as the posterior probability of event A_i given the information, to be distinguished from $P(A_i)$ which we call the prior probability.

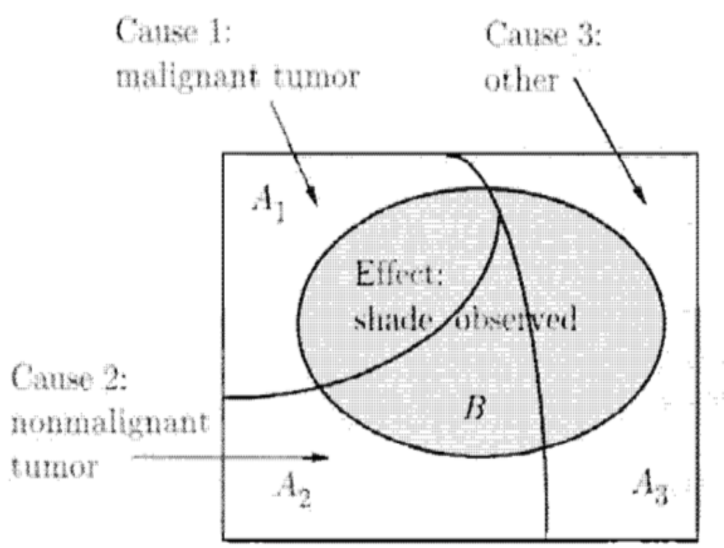
Numerous applications were made in all areas of natural phenomena and human behavior. Let us look at few examples of the "inference".

Example 9.7 *1. If B is a "body" and the A_n's are the several suspects of the murder, then the Baye's theorem will help the jury or court to decide the whodunit.*

2. *If B is an earthquake and the A_n's are the different physical theories to explain it, then the theorem will help the scientists to choose between them.*

3. *We observe a shade in a person's X-ray (this is event B the "effect") and we want to estimate the likelihood of three mutually exclusive and collectively exhaustive potential causes : cause 1 (event A_1) is that there is a malignant tumor, cause 2 (event A_2) is that there is a nonmalignant tumor, and cause 3 (event A_3) corresponds to the reasons other than a tumor. We assume that we know the probabilities $P(A_i)$ and $P(B|A_i), i = 1, 2, 3$. Given that we see a shade (event B occurs), Baye's theorem gives the posterior probabilities of the various causes as:*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{3} P(A_k)P(B|A_k)}, \quad \textit{for each } i = 1, 2, 3.$$



Example 9.8 *A test for a certain rare disease is assumed to be correct* 95% *of the time: if a person has the disease, the test results are positive with probability* 0.95, *and if the person does not have the disease, the test results are negative with probability* 0.95. *A random person drawn from a certain population has probability* 0.001 *of having the disease. Given that the person just tested positive, what is the probability of having the disease?*

Solution: If A is the event that the person has the disease, and B is the event that the test results are positive, the desired probability, $P(A|B)$, is

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\ &= \frac{0.001 \times 0.95}{0.001 \times 0.95 + 0.999 \times 0.05} \\ &= 0.0187 \end{aligned}$$

Since $P(A^c) = 0.999$, $P(\cdot|A^c)$ is a probability measure. Since test is 95% correct, that is $P(B^c|A^c) = 0.95$. Hence $P(B|A^c) = 0.05$.

Note that even though the test was assumed to be fairly accurate, a person who has tested positive is still very unlikely (less than 2%) to have the disease. According to The Economist (February 20th, 1999), 80% of those questioned at a leading American hospital substantially missed the correct answer to a question of this type; most qf them thought that the probability that the person has the disease is 0.95. ■

Remark 9.9 *The practical utility of Baye's formula is limited by our usual lack of knowledge on the various a priori probabilities.*

Lecture 10: Independence

January 19, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

We have introduced the conditional probability $P(A|B)$ to capture the partial information that event B provides about event A. An interesting and important special case arises when the occurrence of B provides no such information and does not alter the probability that A has occurred, i.e. ,

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

When the above equality holds, we say that A is independent of B. Note that by the definition $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, this is equivalent to

$$P(A \cap B) = P(A)P(B)$$

We adopt this latter relation as the definition of independence because it can be used even when $P(B) = 0$, in which case $P(A|B)$ is undefined. The symmetry of this relation also implies that independence is a symmetric property; that is, if A is independent of B, then B is independent of A, and we can unambiguously say that A and B are independent events.

Definition 10.1 *Events A and B are said to be independent if*

$$P(A \cap B) = P(A)P(B)$$

Independence is often easy to grasp intuitively. For example, if the occurrence of two events is governed by distinct and noninteracting physical processes, such events will turn out to be independent .

Example 10.2 *We toss a fair coin two times. Then $\Omega = \{HH, HT, TT, TH\}$ and probability measure is*

$$P(\omega) = \frac{1}{4}, \forall \omega \in \Omega.$$

Consider the events $A = \{HH, HT\}, B = \{HH, TH\}$ and $C = \{HT, TT\}$. Clearly $P(A) = P(B) = P(C) = \frac{1}{2}$. Event A is first toss is head and event B is second toss is head. Physically there is no connection what happens in first toss and second toss, intuitively it is clear that A and B should be independent. In fact that is the case.

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

Once again A and C should be independent. Indeed we have In fact that is the case.

$$P(A \cap C) = P(HT) = \frac{1}{4} = P(A)P(C).$$

Note that independence is not easily visualized in terms of the sample space. A common first thought is that two events are independent if they are disjoint, but in fact the opposite is true: two disjoint events A and B with $P(A) > 0$ and $P(B) > 0$ are never independent, since their intersection $A \cap B$ is empty and has probability 0. For example, an event A and its complement A^c are not independent [unless $P(A) = 0$ or $P(A) = 1$] , since knowledge that A has occurred provides precise information about whether A^c has occurred. In Example 10.2, the event C is second toss is tail. Now B and C determine each other in the sense, if B occurs then C can not occur and vice-versa, so they are not independent. In fact that is the case.

$$P(C \cap B) = P(\emptyset) = 0 \neq P(C)P(B)$$

Intuition, however, is not a reliable guide and one needs to turn to actual computation to verify whether events are indeed, formally, independent.

The following example tell us that sometimes the notion of independence does not appear intuitive but the mathematical coincidence of equality happens and we have to declare the two events independent.

Example 10.3 *Let us consider a random experiment of choosing a real number from $(0, 1]$. Then $\Omega = (0, 1]$ and $\mathcal{F}$ is the σ-field which contains all the subintervals of Ω, and probability measure is defined as "length" of the set. Consider events $A = \left(0, \frac{1}{2}\right]$, $B = \left[\frac{1}{4}, \frac{3}{4}\right]$, $C = \left[\frac{1}{4}, 1\right]$. Then $P(A) = 0.5, P(B) = 0.5, P(C) = 0.75$. Event A is that the chosen number belong to the interval $(0, 0.5]$. Event B is that the chosen number belongs the interval $[0.25, 0.75]$ and event C is that the chosen number belongs the interval $[0.25, 1]$.*

In this experiment the notion of independence/dependence in not at all intuitive. But Then A, B are independent, A, C are dependent and B, C are dependent.

$$P(A \cap B) = P\{[0.25, 0.5]\} = \frac{1}{4} = P(A)P(B)$$
$$P(A \cap C) = P\{[0.25, 0.5]\} = \frac{1}{4} \neq P(A)P(C)$$
$$P(B \cap C) = P\{B\} = \frac{1}{2} \neq P(B)P(C)$$

Example 10.4 *Suppose a family has $n \geq 2$ children. All arrangements of the sexes of the children are assumed to be equally likely. For example listing the genders of the two children,* *elder first, the sample space is* $\Omega = \{bb, bg, gb, gg\}$*. Hence for* n *children the sample space would contain* 2^n *arrangements of the sexes of the children. Let* A *be the event that the family has at most one girl. Let* B *be the event that the family has children of both sexes. Are* A *and* B *independent?*

How many elements in events A*? There is exactly one girl and no girl. There is exactly one sample point for no girl and there are* n *places where a girl can be put, hence* $P(A) = \dfrac{(n+1)}{2^n}$*. Now Event* B *is the compliment of the event that family either have only boys or only girls. This tells us that* $P(B) = 1 - \frac{2}{2^n}$*. Event* $A \cap B$ *has* n *common points, so* $P(A \cap B) = \frac{n}{2^n}$

Now A *and* B *would be independent if*

$$P(A \cap B) = P(A)P(B) \iff \frac{n}{2^n} = \left(1 - \frac{1}{2^{n-1}}\right)\frac{(n+1)}{2^n}$$
$$\iff n = n + 1 - \frac{n+1}{2^{n-1}} \iff n + 1 = 2^{n-1} \iff n = 3$$

Consequently, A *and* B *are independent if* $n = 3$ *but dependent if* $n \neq 3$*.*

As mentioned earlier, if A and B are independent, the occurrence of B does not provide any new information on the probability of A occurring. It is then intuitive that the non-occurrence of B should also provide no information on the probability of A. Indeed, we have the following proposition.

Proposition 10.5 *If* A *and* B *are independent events, then the following pairs are also independent:*

(a) A *and* B^c*,*

(b) A^c *and* B*,*

(c) A^c *and* B^c*.*

Proof:

(a) We must show that $P(A \cap B^c) = P(A)P(B^c)$.

$$\begin{aligned} P(A \cap B^c) &= P(A \setminus (A \cap B)) = P(A) - P(A \cap B) \quad (\because A \cap B \subset A) \\ &= P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(B^c) \end{aligned}$$

(b) Let us relabel the event A as event C and B as event D. So events D and C are independent, hence by part (a) it follows that D and C^c are independent. That is to say B and A^c are independent.

(c) If A and B are independent then A^c and B are independent by part (b). Now let us relabel A^c as event C and relabel B as event D. Applying part (a) on the pair C and D, we get independence of C and D^c. But $C = A^c$ and $D^c = B^c$. This completes the proof.

$\blacksquare$

Lecture 11: Independence of Three Events

January 20, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

The following example illustrate that given a sample space Ω, two events A and B might be independent according to one probability measure and may be dependent according to another probability measure.

Example 11.1 *Let $\Omega = \{(1,1),(1,0),(0,1),(0,0)\}$. Define two probability measure.*

$$P\{\omega\} = \frac{1}{4}, \text{ for all } \omega \in \Omega.$$
$$Q\{(1,1)\} = Q\{(0,0)\} = \frac{1}{3}, \quad Q\{(1,0)\} = Q\{(0,1)\} = \frac{1}{6}$$

Now consider event $A = \{(1,1),(1,0)\}$ and event $B = \{(1,0),(0,0)\}$ Then A and B are independent under P. But $Q(A) = \frac{1}{2} = Q(B), Q(A \cap B) = P\{(1,0)\} = \frac{1}{6} \neq Q(A)Q(B)$

As you know that on a given sample space we may define uncountably many probability measures, hence this example reinforce the point that the notion of independence of two events is w.r.t. a given probability measure.

11.1 Independence of more than two events

We might think that we could say A, B and C are independent if $P(A \cap B \cap C) = P(A)P(B)P(C)$. However, this is not the correct condition.

Example 11.2 *Consider two rolls of a fair six-sided die, and the following events:*

$$A = \{\text{1st roll is 1 , 2, or 3}\},$$
$$B = \{\text{1st roll is 3 , 4, or 5}\},$$
$$C = \{\text{sum of the two rolls is 9}\}$$

Then $A \cap B = \{(3, i)|i = 1, 2, 3, 4, 5, 6\}, A \cap C = \{(3, 6)\}, B \cap C = \{(3, 6), (4, 5), (5, 4)\}$

$$P(A \cap B) = \frac{6}{36} \neq \frac{18}{36} \cdot \frac{18}{36} = P(A)P(B)$$
$$P(A \cap C) = \frac{1}{36} \neq \frac{18}{36} \cdot \frac{4}{36} = P(A)P(C)$$
$$P(C \cap B) = \frac{3}{36} \neq \frac{18}{36} \cdot \frac{4}{36} = P(B)P(C)$$

On the other hand,

$$P(A \cap B \cap C) = \frac{1}{36} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{9} = P(A)P(B)P(C)$$

So it would be grossly embarrassing if we say that three events A, B, C *are independent but they are not pairwise (any two at a time) independent .*

The preceding Example show that mutual (or total) independence of three events requires an extremely strong condition. The following definition works.

Definition 11.3 *We say three events,* A_1, A_2 *and* A_3 *are independent, if it satisfying the four conditions*

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$
$$P(A_2 \cap A_3) = P(A_3)P(A_2)$$
$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

The first three conditions simply assert that any two events are independent, a property known as pairwise independence. But the fourth condition is also important and does not follow from the first three.

Example 11.4 *Let* $\Omega = \{1, 2, 3, 4\}$. *Define*

$$P(i) = \frac{1}{4} \quad \textit{for } i = 1, 2, 3, 4$$

Consider the events $A = \{1, 2\}, B = \{1, 3\}$ *and* $C = \{1, 4\}$. *Then* $P(A) = P(B) = P(C) = \frac{1}{2}$. *Note that* A, B, C *are pairwise independent*

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

but $P(A \cap B \cap C) = P\{1\} = \frac{1}{4} \neq P(A)P(B)P(C)$.

Example 11.5 *Consider the experiment of tossing a fair coin three times. Then* $\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$. *Each outcome has equal chance of occurrence. Let* A_i *be the event that the* i*th toss is a head and* $i = 1, 2, 3$. *Intuitively,* A_1, A_2, A_3 *seems to be independent. Let us verify the same as per the definition* **??**. *Note that*

$$\begin{aligned} A_1 &= \{HHH, HHT, HTH, HTT\} \\ A_2 &= \{HHH, HHT, THH, THT\}, \\ A_3 &= \{HHH, HTH, THH, TTH, \}. \end{aligned}$$

Hence $P(A_i) = \frac{1}{2}$ *for each* $i = 1, 2, 3$.

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) = P(HHH) = \frac{1}{8} = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = P(A_1)P(A_2)P(A_3) \\ P(A_1 \cap A_2) = P(HHT, HHH) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A_1)P(A_2) \\ P(A_2 \cap A_3) = P(HHH, THH) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A_3)P(A_2) \end{aligned}$$

Exercise 11.6 *For any two events* A *and* B*, show that* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Solution: Write $A \cup B = A \cup (B \setminus (A \cap B))$. Then by additivity of probability measure $P(A \cup B) = P(A) + P(B \setminus (A \cap B)) = P(A) + P(B) - P(A \cap B)$. ■ Recall if events A and B are independent then any event determines by A is independent of any other event determined by B. Similarly we expect that if if A, B and C are independent then any event determined by events A and B, will be independent from event C.

Example 11.7 *If events* A, B*, and* C *are independent, show that* $A \cup B$ *is also independent of* C *and* $A^c \cap B$ *is independent of* C.

Solution:

$$\begin{aligned} P((A \cup B) \cap C) &= P\{(A \cap C) \cup (B \cap C)\} \\ &= P(A \cap C) + P(B \cap C) - P\{(A \cap C) \cap (B \cap C)\} \\ &= P(A)P(C) + P(B)P(C) - P(A \cap B \cap C) \\ &= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C) \\ &= P(C)\left[P(A) + P(B) - P(A)P(B)\right] \\ &= P(C)\left[P(A) + P(B) - P(A \cap B)\right] = P(C)P(A \cup B) \end{aligned}$$

and

$$\begin{aligned}
P((A^c \cap B) \cap C) &= P(A^c \cap B \cap C) = P(B \cap C) - P(B \cap C \cap A) \\
&= P(B)P(C) - P(A)P(B)P(C) \\
&= P(B)P(C)\,[1 - P(A)] = P(B \cap C)P(A^c)
\end{aligned}$$

■

Lecture 12: Independence of multiple events & Random Variables

January 22, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

The definition of independence can be extended to multiple events (more than three).

Definition 12.1 *We say that the events $A_1, A_2, \cdots, A_n$ are independent if*

$$P\left(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_m}\right) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_m}).$$

for every $2 \leq m \leq n$, and for every choice of indices $1 \leq i_1 < i_2 \cdots < i_m \leq n$

For $m = 2$, we have $\binom{n}{2}$ conditions to be checked for pairs. For $m = 3$, we have $\binom{n}{3}$ conditions to be checked for triples and so on. Hence checking independence of n events require to check

$$\binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = (1+1)^n - \binom{n}{0} - \binom{n}{1} = 2^n - n - 1$$

non-trivial conditions. Independence places very strong requirements on the interrelationships between events.

Example 12.2 *Let $\Omega = \{(a_1, a_2, \cdots, a_n) | a_i$ is either 0 or 1 for each i $\}$. For $i = 1, 2, \cdots, n$ set $A_i = \{(a_1, a_2, \cdots, a_n) | a_i = 1\}$. If all the outcomes are equally likely, then show that $A_1, A_2, \cdots, A_n$ are independent.*

Solution: We have to count the number of elements in A_i. Since only *ith* place is fixed and for each of rest of $n-1$ places we have two choices either 1 or 0 hence each A_i has 2^{n-1} elements. Hence

$$P(A_i) = \frac{2^{n-1}}{2^n} = \frac{1}{2}, \ \forall i = 1, 2, \cdots, n.$$

Note that

$$A_{i_1} \cap A_{i_2} \cap \cdots A_{i_m} = \{(a_1, a_2, \cdots, a_n) | a_{i_1} = a_{i_2} = \cdots = a_{i_m} = 1\}$$

Hence

$$P(A_{i_1} \cap A_{i_2} \cap \cdots A_{i_m}) = \frac{2^{n-m}}{2^n} = \frac{1}{2^m} = P(A_{i_1})P(A_{i_2} \cdots P(A_{i_m})$$

■

Notion of independence can be extended easily to countably infinite collection of events.

Definition 12.3 *We say that a countably infinite collection of events $\{A_i | i \geq 1\}$ is independent if any finite sub-collection of events is independent.*

12.1 Random Variables

In many random experiments, the outcomes are numerical, e.g., throwing a dice, when outcome corresponds to instrument reading or stock prices. In other random experiments, the outcomes may not be numerical, e.g., tossing a coin, choosing a student randomly from the class. In general, the points of a sample space may be very concrete objects such as apples, molecules, and people. Sometimes we may be interested in certain real-valued (sometimes we may require to consider complex-valued) function of the sample space. For example, if the experiment is the selection of students (for placement!!) from a given population, we may wish consider their CGPA.

But all functions defined on the sample space are not useful, in the sense that we may not be able to assign probabilities to all basic events associated with the function. So one need to restrict to certain class of functions of the sample space. This motivates us to define random variables.

Definition 12.4 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, A function $X : \Omega \to \mathbb{R}$ is said to be a random variable if for each $x \in \mathbb{R}$,*

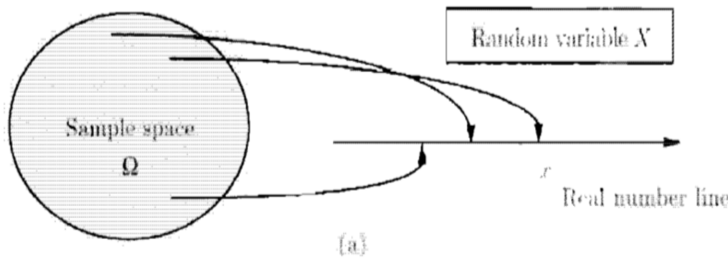
$$\{\omega \in \Omega | X(\omega) \leq x\} =: \{X \leq x\} \in \mathcal{F}$$

Remark 12.5

1. *The adjective "random" is just to remind us that we are dealing with a sample space, which is related to something called random phenomena or random experiment. Once ω is picked, $X(\omega)$ is thereby determined and there is nothing vague, or random about it anymore.*
2. *Observe that random variables can be defined on a sample space before any probability is mentioned.*
3. *When experimenters are concerned with random variables that describe observations, their main interest is in the probabilities with which the random variables take various values. For example, Let Ω be set of students in A Section for P&S class. These may be labeled as $\Omega = \{\omega_1, \omega_2, \cdots, \omega_{241}\}$. If we are interested in their attendance distribution, let $A(\omega)$ denote the attendance of ω. Now we are interested in determining the probability of events like $\{A \geq 1\} = \{\omega \in \Omega | A(\omega) \geq 1\}, \{A = 0\}$ etc. For that these events should be in the class $\mathcal{F}$. This idea suggest the condition in the Definition 12.4.*

If you look at most of the books on probability theory they give the following definition of random variable.

Definition 12.6 (Engineering or Practical Definition) *A random variable is a function from Ω to $\mathbb{R}$ (or $\mathbb{C}$).*



(a)

Remark 12.7 *When sample space is finite or countably infinite, any function can be regarded as random variable in the sense that we can always take the power set as the event class $\mathcal{F}$. Recall that if sample space is finite or countably infinite, we can define probability of any event, hence Definition 12.6 is perfect in this case.*

But there are certain difficulties arise when sample space is uncountable or continuous and random variable with a continuous range of possible values. Then it may not be possible for us to define probability to every event. So we can not work with power set, hence we need a class of events which is rich enough to include all the events of our interest. Precisely, this is the reason we need the Definition 12.4. It is comforting to know that mathematical subtleties of this type do not arise in most of the physical applications.

Example 12.8 *In an experiment involving two rolls of a die, the following are examples of random variables:*

1. *The sum of the two rolls.* $\boxed{X((i,j)) = i + j}$ ✓

2. *The number of sixes in the two rolls.* $X((i,j)) = \begin{cases} 0 & \text{if} \quad i \neq 6, j \neq 6 \\ 1 & \text{if} \quad \textit{exactly of one of } i \textit{ or } j \textit{ is } 6 \\ 2 & \text{if} \quad i = j = 6 \end{cases}$

3. *The second roll raised to the fifth power.* $X((i,j)) = j^5$.

So this example tell us that with a single sample space we may have several random variables sitting on it.

α **Example 12.9** *Let* $\Omega = [0,1]$, $\mathcal{F} =$ *class of events containing all the subintervals of the* $[0,1]$. *Define* $X : \Omega \to \mathbb{R}$ *by*

$$X(\omega) = 3\omega + 1.$$

Claim 12.10 *X is a random variable.*
Note that the range of X is $[1,4]$. *Hence for given* $x \in \mathbb{R}$,

1. *If* $x < 1$, *the* $\{X \leq x\} = \emptyset$.
2. *If* $1 \leq x \leq 4$, *the* $\{X \leq x\} = \left[0, \dfrac{x-1}{3}\right]$.
3. *If* $4 < x$, *the* $\{X \leq x\} = \Omega$.

Hence for each $x \in \mathbb{R}$, $\{X \leq x\} \in \mathcal{F}$.

Starting with some random variables, we can at once make new ones by operating on them in various ways.

Proposition 12.11 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space.*

1. *If* X *and* Y *are random variables, then so are*
$$\boxed{X + Y, X - Y, XY, \frac{X}{Y}\ (Y \neq 0),}$$
and $aX + bY$ *where* a *and* b *are two real numbers.*
2. *If* X *is a random variable and* $f : D \subset \mathbb{R} \to \mathbb{R}$ *is a function of one variable then* $f(X)$ *is also a random variable.*
3. *If* X *and* Y *are random variables and* $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ *is a function of two variables then* $f(X, Y)$ *is also random variable.*

Lecture 13: Discrete Random Variables

January 25, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

13.1 Discrete Random Variables → finite sample space pas defined.

Definition 13.1 *Let $(\Omega, \mathcal{F})$ be a measurable space and $X : \Omega \to \mathbb{R}$ be a random variable. Random variable X is called discrete if its range is either finite or countably infinite subset of $\mathbb{R}$.*

$\Omega \to$ uncountable

The random variables in Example 12.8 can take at finite number of numerical values, and are therefore discrete, whereas the random variable in Example 12.9 have the range $[1, 4]$, which is an uncountable set and therefore it is not a discrete random variable.

If sample space is finite then by definition of a function it follows that, any random variable defined on the sample space necessarily have finite range. Therefore any random variable on finite sample space is a discrete random variable.

If sample space is countably infinite then any random variable defined on the sample space have range which is either finite or countably infinite. Therefore any random variable on countably infinite sample space is a discrete random variable.

Example 13.2 *In the random experiment of tossing a coin till you get a head, the sample space is*

$$\Omega = \{H, TH, TTH, TTTH, \cdots\}$$

which is a countably infinite set. Define a random variable X as the number of toss required to get the head, then

$$X(\omega) = \begin{cases} 1 & \text{if} \quad \omega = H \\ 2 & \text{if} \quad \omega = TH \\ 3 & \text{if} \quad \omega = TTH \\ \vdots & \vdots \quad \vdots \end{cases}$$

$S \to \mathbb{N}$

We may define a random variable Y as follows:

$$Y(\omega) = \begin{cases} 1 & \text{if} \quad \omega = H \\ 0 & \quad \text{otherwise} \end{cases}$$

Note that we may have a sample space which is uncountable but we may define discrete random variables on it.

For an example, consider the experiment of choosing a point a from the interval $[-1, 1]$. So now our $\Omega = [-1, 1]$. The random variable that associates with x the numerical value

$$X = sgn(x) = \begin{cases} -1 & \text{if} \quad x < 0 \\ 0 & \text{if} \quad x = 0 \\ 1 & \text{if} \quad x > 0 \end{cases} \qquad \{-1, 0, 1\}$$

is discrete.

If X is a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function, then $g(X)$ is also a discrete random variable.

True/ False: If X is not a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function, then $g(X)$ is also not discrete.

The statement is False. Let $\Omega = (-1, 1), X(\omega) = \omega$ and $g(x) = \begin{cases} -1 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad x \geq 0 \end{cases}$ Then $g(X)$ is discrete since it's range is $\{-1, 1\}$.

13.2 Probability Mass Function

The most important way to characterize a discrete random variable is through the probabilities of the values that it can take. For a discrete random variable X, these are captured by the probability mass function (pmf for short) of X, denoted $f_X(\cdot)$. In particular, if x is any possible value of X, the probability mass of x, denoted $f_X(x)$, is the probability of the event $\{X = x\}$ consisting of all outcomes that give rise to a value of X equal to x:

$$f_X(x) = P\{X = x\} \qquad f_x(x) = P\{X = x\}$$

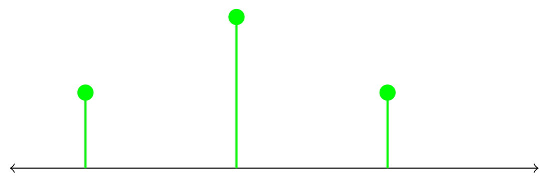
Definition 13.3 *Let X be a discrete random variable on a probability space $(\Omega, \mathcal{F}, P)$. A real-valued function f defined on $\mathbb{R}$ by $f(x) = P(X = x)$ is called the pmf of X.*

Example 13.4 *Consider the experiment consist of two tosses of a fair coin, and let X be the number of heads obtained. Then the pmf of X is*

$$f_X(x) = \begin{cases} \frac{1}{4} & \text{if} \quad x = 0, 2 \\ \frac{1}{2} & \text{if} \quad x = 1 \\ 0 & \quad \textit{otherwise} \end{cases}$$

Pmf captures the "point probabilities or point mass", in the sense that $P(X = x)$ is the mass attached at point x. We may draw pmf of X as follows:

Aryan Baghla



Proposition 13.5 (Properties of pmf) *Let f be the pmf of a discrete random variable X. Then it has the following properties*

1. $f(x) \geq 0$ *for all* $x \in \mathbb{R}$.
2. $\{x \in \mathbb{R} : f(x) > 0\}$ *is a finite or countably infinite subset of* $\mathbb{R}$.
3. $\sum_x f(x) = 1$, *where summation is over all x belongs to the range of X.*

Proof:

1. Since $P\{X = x\} \geq 0$ for any x.
2. Since range of X is either finite or countably infinite.
3. As x ranges over all possible values of X the events $\{X = x\}$ are disjoint and form a partition of the sample space, that is

$$\Omega = \bigcup_{x \in R(X)} \{X = x\}$$

Now the result follows from the additivity and normalization axioms of probability measure.

■

The pmf contains all the information about its discrete random variable, i.e., for any set $S \subset \mathbb{R}$, we can compute the probability of the event $\{X \in S\}$ we have

$$P(X \in S) = \sum_{x \in S \cap R(X)} f_X(x) = \sum_{x \in S \cap R(X)} P\{X = x\}$$

For example, if X is the number of heads obtained in two tosses of a fair coin, as above, the probability of at least one head is

$$P(X \geq 1) = \sum_{x=1}^{2} f_X(x) = \frac{3}{4},$$

since $S = [1, \infty)$ and $R(X) = \{0, 1, 2\}$.

Calculating the PMF of X is conceptually straightforward. For each possible value x of X:

1. Collect all the possible outcomes that give rise to the event $\{X = x\}$
2. Add their probabilities to obtain $f_X(x)$.

Example 13.6 *Let* $\Omega = \{\omega_1, \omega_2, \omega_3\}, \mathcal{F} = \mathcal{P}(\Omega), P(\omega_i) = \frac{1}{3}$ *for* $i = 1, 2, 3,$ *and define random variables* X, Y *and* Z *as follows:*

$$X(\omega_1) = 1, X(\omega_2) = 2, X(\omega_3) = 3;$$
$$Y(\omega_1) = 2, Y(\omega_2) = 3, Y(\omega_3) = 1;$$
$$Z(\omega_1) = 3, Z(\omega_2) = 1, Z(\omega_3) = 2.$$

Show that X, Y *and* Z *have the same pmf.*

Solution: Note that $R(X) = R(Y) = R(Z) = \{1, 2, 3\}$.

$$P\{X = 1\} = P(\omega_1) = \frac{1}{3}$$
$$P\{X = 2\} = P(\omega_2) = \frac{1}{3}$$
$$P\{X = 3\} = P(\omega_3) = \frac{1}{3}$$
$$P\{Y = 1\} = P(\omega_3) = \frac{1}{3}$$
$$P\{Y = 2\} = P(\omega_1) = \frac{1}{3}$$
$$P\{Y = 3\} = P(\omega_2) = \frac{1}{3}$$
$$P\{Z = 1\} = P(\omega_2) = \frac{1}{3}$$
$$P\{Z = 2\} = P(\omega_3) = \frac{1}{3}$$
$$P\{Z = 3\} = P(\omega_1) = \frac{1}{3}$$

■

Moral of the Example 13.6 is, different random variables may have same pmf.

Lecture 14: Standard Discrete Random Variables

27 January, 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

The three properties in the Proposition 13.5 characterize the pmf.

Theorem 14.1 *Let $f : \mathbb{R} \to \mathbb{R}$ be a function which satisfies the following three properties:*

1. $f(x) \geq 0$.
2. *The set $S := \{x \in \mathbb{R} : f(x) > 0\}$ is a finite or countably infinite subset of $\mathbb{R}$.*
3. $\sum_{x \in S} f(x) = 1$.

Then there exist a probability space $(\Omega, \mathcal{F}, P)$ and discrete random variable X on it such that f is the pmf of X.

Proof: Take $\Omega = S$. Since S is at most countable, hence take $\mathcal{F}$ be the power set of Ω. Now it is enough to define probability of each singleton and we define P the probability measure by $P(x) = f(x)$ for $x \in \Omega$. The random variable X defined by $X(x) = x$. Of course the range of X is S which is either finite or countably infinite. Note that the pmf of X is

$$f_X(x) = P(X = x) = P(x) = f(x) \quad \forall x \in S$$

For $x \in S^c$ both f and f_X are zero, hence they agree everywhere on real line. So f is the pmf of X. ∎

Remark 14.2 *The above result assures us that statements like "Let X be a random variable with the pmf f" always make sense, even if we do not specify directly a probability space upon which X is defined.*

The notion of a discrete random variable forms a convenient way of describing a random experiment that has a finite or countably infinite number of possible outcomes. We need not bother to set up a probability space for the experiment. Instead we can simply introduce a random variable X taking values $x_1, x_2, \cdots$ such that $X = x_i$, if and only if the experiment results in the ith outcome. Thus, for example, consider the experiment of picking a point at random from the finite subset S of $\mathbb{R}$ consisting of the distinct points $x_1, \cdots, x_n$. We say that a discrete random variable X is taking value x_i if the selected point is x_i. Then the function f defined by $f(x_i) = \frac{1}{n}$ for $i = 1, 2, \cdots, n$, and $f(x) = 0$ elsewhere is clearly a pmf. A random variable X having this pmf is said to be uniformly distributed on S. Observing a value of X corresponds to our intuitive notion of choosing a point at random from S.

In general, conducting an experiment which has a finite or countably infinite number of possible outcomes can be described as observing the value of a discrete random variable X. Many times, in fact, this is how the experiment already appears to us, and often it is easier to think of the experiment in these terms rather than in terms of a probability space.

14.1 Standard Discrete Random Variables

Now we look at some standard random variables.

14.1.1 Discrete Uniform Random Variable:

If the random variable X takes finitely many values (say N) with the probability of taking each value is same, then we call random variable X discrete uniform random variable.

To standardize, we assume that X takes values $\{1, 2, \cdots, N\}$. Then pmf of X is

$$f(i) = P\{X = i\} = \frac{1}{N} \ \forall\ i = 1, 2, \cdots, N.$$

The positive integer N is called the parameter of the discrete uniform distribution. That is if we say X has the pmf

$$P(X = i) = \frac{1}{10}, \quad i = 1, 2, \cdots, 10$$

and Then X is discrete uniform random variables with the parameter $N = 10$. Suppose Z is random variable with the pmf

$$P(Z = i) = \frac{1}{11}, \quad i = 1, 2, \cdots, 11$$

Then Z is also a discrete uniform random variable but with parameter $N = 11$.

It is like we have small, medium, XL, XXL size of the same PETER ENGLAND shirt, where the basic structure of all shirts is same with different values of the parameters(length, size).

In statistics, we usually deal with a family of distribution rather than a single distribution. This family is indexed by one or more parameters, which allow us to vary certain characteristics of the distribution while staying with one functional form. For example, we may specify the discrete uniform distribution as reasonable choice to model a particular population, but we can not precisely specify the number N. Then we deal with a parametric family of discrete uniform distributions with unspecified parameter N, where $N \in \mathbb{N}$.

14.1.2 The Bernoulli Random Variable

→ Involves two possible results.

Consider the toss of a coin, which comes up a head with probability p and a tail with probability $1-p$. The Bernoulli random variable takes the two values 1 and 0 depending on wether the outcome is head or tail. Let $\Omega = \{H, T\}, P(H) = p, P(T) = 1-p$ where $0 \le p \le 1$. Define $X : \Omega \to \mathbb{R}$ by $X(H) = 1, X(T) = 0$.
Question: What is the pmf of Bernoulli random variable ?

$$f_X(x) = \begin{cases} p & \text{if} \quad x = 1 \\ 1-p & \text{if} \quad x = 0 \\ 0 & \quad\text{otherwise} \end{cases} \tag{14.1}$$

Any random variable having pmf (14.1) is called a Bernoulli (p) random variable, where p is the parameter. If an experiment involves only two possible results, we say that we have a Bernoulli trial. The two possible results can be anything, e.g., "it rains" or "it doesn't rain," but we will often think in terms of coin toss and refer to the two results as "head" (H) and "tail" (T). For all its simplicity Bernoulli random variable is very important. In practice, it is used to model generic probabilistic situations with just two outcomes such as:

1. The state of a telephone at a given time that can either be free or busy.

2. A person is either healthy or sick with a given disease.

3. The preference of a person who can be either for or against Narendra Modi.

14.1.3 Binomial Random Variable

Consider an experiment that consists of n independent tosses of a coin, in which probability of head (in each trail) is p (where $0 < p < 1$). In this context, Independence means that the events $A_1, A_2, \cdots A_n$ are independent where $A_i = \{i\text{th toss is a head}\}$. For example, take $n = 3$ then we see that any particular outcome (3-long sequence of heads and tails) that involves k heads and $3-k$ tails has probability $p^k(1-p)^{3-k}$, that is $P(HHT) = P(HTH) = p^2(1-p)$. This formula extends to the case a general number n of tosses. We obtain that any particular outcome (n-long sequence of heads and tails) that involves k heads and $n-k$ tails has probability $p^k(1-p)^{n-k}$ for all k from 0 to n. We define a random variable X on Ω such that $X(\omega)$ is the number of heads in n independent tosses. Then X takes values

$0, 1, 2, \cdots, n$. Let us compute it's pmf.

$$\begin{aligned} P(X = k) &= P(\text{exactly } k \text{ heads come up in an } n\text{-toss sequence }) \\ &= \text{number of distinct } n\text{-toss sequences that} \\ &\quad\ \text{contain exactly } k \text{ heads} \times p^k(1-p)^{n-k} \\ &= \binom{n}{k} p^k(1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k} \end{aligned}$$

Note that the pmf must add to 1, thus we have

$$\sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} = 1$$

So Binomial random variable is determined by two parameters n and p, where $n \in \mathbb{N}$ is number of independent trials and $p \in (0, 1)$ is probability of success in each trail.

14.1.4 Geometric Distribution

Let us consider a random experiment, we toss a coin till we get first head. In each toss the probability of getting head is p. Also we assume each coin toss is independent. If X denotes the number of coin tosses required to get first head, then X is a random variable which takes values $1, 2, \cdots$. So pmf of X would be

no. of coin tosses req. to get 1st head.

$$P(X = k) = p(1-p)^{k-1}, \text{ for } k = 1, 2, \cdots$$

A random variable with above pmf is called a geometric random with parameter p. The name geometric is motivated from the fact that pmf resemble the geometric sequence $a, ar, ar^2, \cdots$. So a geometric random variable is determined by one parameters p, where $0 < p < 1$ is probability of success or head in each trail.

Proposition 14.3 (Memoryless Property) *Let X be a geometric(p) random variable. Then for any nonnegative integers m and n,*

$$P(X > n + m | X > m) = P(X > n). \tag{14.2}$$

Proof: Note that for a non-negative integer n

$$P(X > n) = \sum_{k=n+1}^{\infty} P(X = k) = \sum_{k=n+1}^{\infty} p(1-p)^{k-1} = p(1-p)^n[1+(1-p)+(1-p)^2+\cdots] = (1-p)^n$$

$$P(X > n+m|X > m) = \frac{P(X > n+m, X > m)}{P(X > m)} = \frac{P(X > n+m)}{P(X > m)}$$
$$= \frac{(1-p)^{n+m}}{(1-p)^m} = (1-p)^n = P(X > n)$$

■

Think of X as the waiting time till success. Given $X > m$ means we have already waited for m unit of times, and $X > n+m$ means we have to wait $m+n$ unit times. So if waiting time till success X follows geometric distribution, property (14.2) says that probability of waiting additional n times (having already waited for m times) is same as the probability of waiting n times at the start of the sequence.

In other words, the probability of getting the first success depends only on the length of the time interval not on its position.

Remark 14.4 *Note that memoryless property for geometric distribution holds only for non-negative integers. For e.g., if $m = n = \frac{1}{2}$ then (14.2) does not hold.*

$$P(X > 1|X > \frac{1}{2}) = P(X > 1|X \geq 1) = \frac{P(X > 1)}{P(X \geq 1)} = (1-p) \neq P(X > \frac{1}{2}) = 1$$

Theorem 14.5 *If a discrete random variable X taking values on $\mathbb{N}$ satisfies (14.2), then X is a geometric random variable.*

Lecture 15: Poisson Distribution & Random Variable with pdf

29 January 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

15.1 Poisson random variable

A random variable X with pmf

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \cdots,$$

is called a Poisson random variable with parameter $\lambda > 0$.

Many counting type random phenomena are known from experience to be approximately Poisson distributed. Some examples of such phenomena are the number of atoms of a radioactive substance that disintegrate in a unit time interval, the number of calls that come into a telephone exchange in a unit time interval, and the number of misprints on a page of a book.

Proposition 15.1 *Define,*

$$B_k(n; p) := \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2 \cdots, n.$$

Then for $\lambda > 0$

$$\lim_{n\to\infty} B_k\left(n; \frac{\lambda}{n}\right) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2 \cdots \tag{15.1}$$

The limit in equation (15.1) is binomial approximation to the Poisson in the sense, we let n grows to infinity simultaneously decreasing p, in a manner that keeps the product np at a constant value say λ, then, it turns out that binomial pmf converges to Poisson pmf.

$B_k(n,p)$

$p = \lambda/n$

$\boxed{\lambda = np}$

Proof: For given k, we have

$$\begin{aligned} B_k\left(n; \frac{\lambda}{n}\right) &= \frac{n!}{(n-k)!k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-(k-1))}{k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n}{n}\times\frac{n-1}{n}\cdots\times\frac{n-k+1}{n}\times\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k} \end{aligned}$$

here k is fixed. So as $n \to \infty$ each of the ratios: $\frac{n}{n}, \frac{n-1}{n}, \cdots \frac{n-k+1}{n}$ converge to 1. Furthermore,

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1, \qquad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$$

Hence we conclude, that for each fixed k,

$$\lim_{n\to\infty} B_k\left(n; \frac{\lambda}{n}\right) = \frac{e^{-\lambda}\lambda^k}{k!}$$

■

Remark 15.2 *In general, the Poisson distribution is a good approximation to the binomial as long as $\lambda = np$, n is very large, and p is very small. As a thumb rule,*

$$\binom{n}{k}p^k(1-p)^{n-k} \approx \frac{e^{-\lambda}\lambda^k}{k!}, \qquad \textit{for } k = 0, 1, 2\cdots,$$

is valid to several decimal places if $n \geq 100, p < 0.01$, and $\lambda = np$. For example take $n = 100$ and $p = .01$. Then we have $\lambda = np = 1$ and following is comparison

k	*Binomial probabilities*	*Poisson probability*
0	*0.366*	*0.368*
2	*0.185*	*0.184*
5	*0.00290*	*0.00306*
10	7.006×10^{-8}	1.001×10^{-8}

The limit (15.1) represents the simplest from of law of rare events. Given an indefinitely large number of independent trials, where success on each trial occurs with the same arbitrarily small probability, then the total number of successes will follow, approximately, a Poisson distribution.

15.2 Random variable with density or absolutely continuous random variable

Definition 15.3 *A random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$ is called absolutely continuous if there is a nonnegative function f_X defined on $\mathbb{R}$, called the probability density function of X (sometimes just density of X), or pdf for short, such that*

$$P(X \in S) = \int_S f_X(x)dx,$$

for every Borel subset S of $\mathbb{R}$.

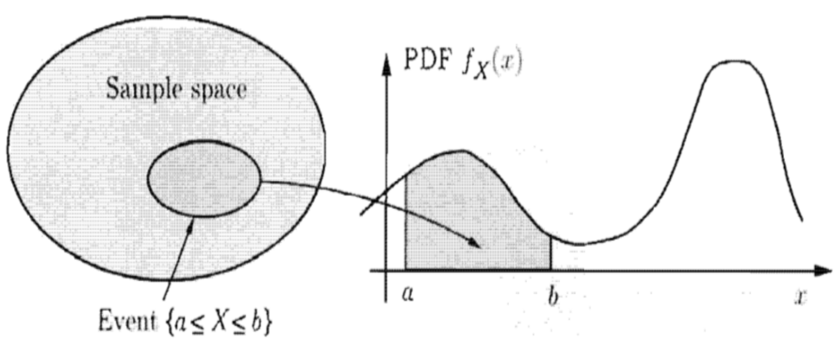
Example of Borel sets: Every singleton, every countable set, all types of intervals, open set, closed set, their (finite or counatble) union or intersections etc.

In particular, the probability that the value of X falls within an interval is

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

pdf

and can be interpreted as the area under the graph of the pdf.



For any single value a, we have

$$P(X = a) = \int_a^a f_X(x)dx = 0.$$

For this reason for an absolutely continuous random variable, including or excluding the endpoints of an interval has no effect on its probability

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b).$$

Example 15.4 *Show that the range of an absolutely continuous random variable is uncountable.*

Solution: Assume contrary that the range of X is countable, i.e., $R(X) = \{a_1, a_2, \cdots\}$. Note that $(\{X = a_n\})_{n=1}^{\infty}$ is a countable partition of sample space Ω.

$$\Omega = \bigcup_{n=1}^{\infty} \{X = a_n\} \implies P(\Omega) = 1 = \sum_{n=1}^{\infty} P(X = a_n) = 0,$$

where last equality follows from the fact that X is absolutely continuous. ∎

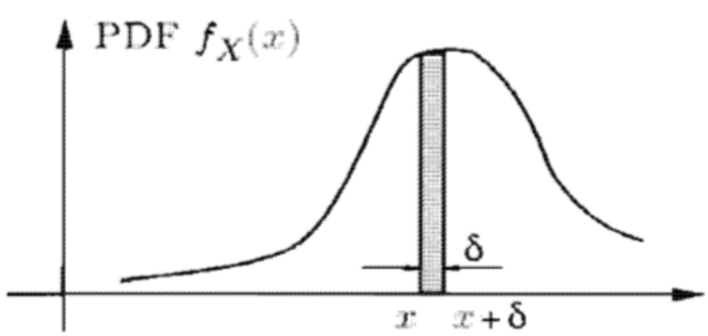
Remark 15.5 *Discrete random variables were defined in term of the range of the random variable. So you might be tempted to think that a absolutely continuous (or continuous by many authors) random variable should be one which has range as uncountable set or a continuous subset of $\mathbb{R}$, i.e., intervals. Actually this is not the way we define continuous random variable, it is the probability distribution via pdf. We shall see random variables with range as interval but does not have the pdf.*

15.2.1 PDF

Interpretation of the pdf For an interval $[x, x+\delta]$ with very small length δ, we have

$$P(x \leq X \leq x+\delta) = \int_x^{x+\delta} f_X(t)dt \approx f_X(x) \cdot \delta \implies \frac{P(x \leq X \leq x+\delta)}{\delta} \approx f_X(x)$$

So we can view $f_X(x)$ as the "probability mass per unit lenght" near x or in other words it the rate at which probability accumulates near point x.



Lecture 16: Standard Random Variables with PDF

01 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

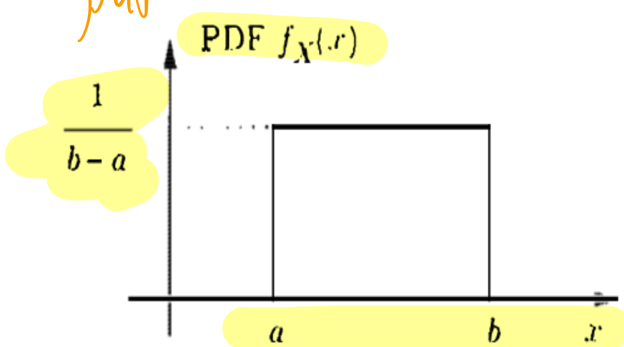
discrete =

16.1 Continuous Uniform Random variable

A random variable X that takes values in an interval $[a, b)$ and any two sub-interval of equal length have the same probability is called uniform random variable and is denoted by $U([a, b))$. It's pdf is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \leq x < b \\ 0, & \text{otherwise.} \end{cases}$$

pdf



Limiting Uniform Experiments: Suppose $X \sim U([0, 1))$. Then it can be seen as limit (in sense of probability distribution) of discrete uniform random variables.

For each $n \geq 1$, suppose X_n taking values in the discrete set of points $\left\{0, \frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}\right\}$ with $P\left(X_n = \frac{k}{n}\right) = \frac{1}{n}$ for each $0 \leq k \leq n-1$. The points $\frac{1}{n}$ pepper the unit interval increasingly densely as n increases and in the limit we pass to a continuum of values. The number of the points $\dfrac{k}{n}$ encountered in any given subinterval (a, b) of $[0, 1)$ differs from $\frac{b-a}{\frac{1}{n}}$ only in the addition or deletion of a point at either boundary. For e.g., Take $n = 10$ and $(a, b) = (0.5, 0.75)$ Then $\frac{6}{10}, \frac{7}{10} \in (a, b)$, where as $n(b-a) = 2.5$. As $n \uparrow \infty$, this difference $|n(b-a) -$ number of points $\frac{k}{n}$ in (a,b) $| \downarrow 0$. Now $P(a < X_n < b) = \frac{1}{n} \times$ number of points $\frac{k}{n}$ in (a,b) $\approx b - a$

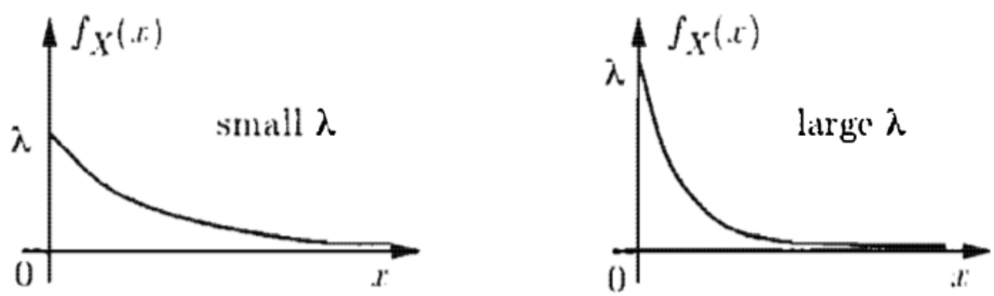
Hence we may conclude $P(a < X_n < b) \to b - a$ as $n \to \infty$. If we set $u(x) := 1$ for x in the unit interval $[0, 1)$ then we may write $P\{a < X_n < b\} \to \int_a^b u(x)dx$ where $u(x)$ stands in the role of a limiting probability density on the unit interval. The function u is the uniform density over interval $[0, 1)$ and serves as a model for the generation of a random point in the unit interval.

16.2 Exponential Random variable

A random variable X with the pdf

$$f(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \lambda e^{-\lambda x} & \text{if} \quad x \geq 0 \end{cases}.$$

is called a exponential random variable with parameter $\lambda > 0$.



Remark 16.1 *A geometric random variable X takes values $1, 2, \cdots$, and we interpret X as the number of trials until the first success. Sometimes the term "geometric distribution" is used in referring with the following pmf.* → discrete

$$f(x) = \begin{cases} (1-p)^x p & \text{if} \qquad x = 0, 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}.$$

This is merely the pmf of random variable $X' = X - 1$, has the interpretation as the number of failures prior to the first success. $P(X = k) = P(X' = k - 1)$. The X' connotes the waiting time to the first success.

Limiting geometric probabilities Now we show that an exponential random variable can be seen as limit (in some sense) of geometric random variable.

Consider a sequence of coin tosses where the success probability $p = p_n$ decreases to zero as n increases in such a way that $np_n \to \alpha$ for some fixed α, i.e., $p_n \approx \frac{\alpha}{n} = \frac{\alpha x}{nx}$ for large values of n and for each fixed $x > 0$.

For each positive integer k, suppose X_n takes value $\dfrac{k}{n}$ with probability $(1-p_n)^k p_n$. Then X_n represents a suitably scaled (by n) waiting time till the first success in a succession of tosses. As the approximation $(1-\frac{x}{n})^n \approx e^{-x}$ gets increasingly good as $n \to \infty$, in the limit of large n we obtain

$$P(X_n \leq x) = 1 - \sum_{k:k>nx} (1-p_n)^k p_n = 1 - (1-p_n)^{[nx+1]} \to 1 - e^{-\alpha x}$$

and it follows that $P(a < X_n < b) \to e^{-\alpha a} - e^{-\alpha b}$. If we write $w(x) := \alpha e^{-\alpha x}$ for $x \geq 0$, we hence obtain $P(a < X_n < b) \to \displaystyle\int_a^b w(x)dx$ and, again, the function $w(x)$ may be seen in the light of a limiting probability density for a scaled geometric random variable with support now on the entire half-line. This is the exponential density with parameter α.

Recall that a geometric distribution have a memoryless property for non-negative integers. Since exponential distribution can be seen as a limit of scaled geometric random variables, it is natural to expect the memoryless property for non-negative real numbers for exponential distribution. Indeed, the answer is yes.

Proposition 16.2 *Let* $X \sim \exp(\lambda)$. *Then*

$$P(X > x + y | X > x) = P(X > y), \forall\, x, y \in [0, \infty). \tag{16.1}$$

Proof: Note that for $x, y \geq 0$, $\{X > x + y\} \subseteq \{X > x\}$ and $\{X > x + y\} \subseteq \{X > y\}$. Hence

$$\begin{aligned} P(X > x + y | X > x) &= \frac{P(\{X > x + y\} \cap \{X > x\})}{P(X > x)} \\ &= \frac{P(X > x + y)}{P(X > x)} \\ &= \frac{1 - P(X \leq x + y)}{1 - P(X \leq x)} \\ &= \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} \\ &= e^{-\lambda y} \\ &= P(X > y) \end{aligned}$$

■

Theorem 16.3 *If a random variable X with pdf taking values on $[0, \infty)$ satisfies (16.1), then X is an exponential random variable.*

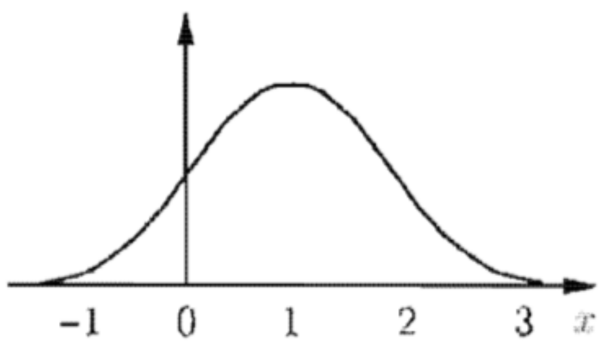
16.3 Normal Random Variable

A random variable with the following pdf

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

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

is called a normal random variable with parameters μ, σ^2, where σ is assumed to be positive. Below is the graph of normal pdf with $\mu = 1$ and $\sigma = 1$.



A normal random variable with $\mu = 0$ and $\sigma^2 = 1$ is said to be a standard normal random variable.

The de Moivre–Laplace limit law Let S_n be the number of successes in n tosses of a coin with (fixed) success probability p. Then S_n has the binomial distribution with paramter n and p. The shifted and scaled variable $S_n^* = \dfrac{S_n - np}{\sqrt{np(1-p)}}$ takes values in the set $\left\{ \dfrac{k - np}{\sqrt{np(1-p)}}, 0 \leq k \leq n \right\}$ (note that $np < n$ so S_n^* take negative values also) and as n increases these values pepper the real line increasingly finely. The de Moivre–Laplace theorem asserts that

$$P(a < S_n^* < b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Lecture 17: CDF or Distribution Functions

02 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

$F_x(x) = P(X \leq x)$
$\forall x \in \mathbb{R}$

Definition 17.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X be a random variable on it. The distribution function (or cumulative distribution function CDF) of X, denoted by F_X, is defined as*

$$F_X(x) = P(X \leq x), \forall x \in \mathbb{R}.$$

Note that the notion of distribution function is well-defined for every random variable (discrete, absolutely continuous or mixed).

In particular, for every x we have

$$F_X(x) = P(X \leq x) = \begin{cases} \displaystyle\sum_{t \leq x : t \in R(X)} \overset{\to \text{pmf}}{f_X(t)} & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{x} \overset{\to \text{pdf}}{f_X(t)}\, dt & \text{if } X \text{ is absolutley continuous} \end{cases}$$

Loosely speaking, the CDF $F_X(x)$ "accumulates" probability "up to" the point x.

Example 17.2 *Let X be discrete uniform random variable with parameter N. Determine it's CDF.*

$1/N$

Solution: Recall $R(X) = \{1, 2, \cdots, N\}$. Therefore

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 1 \\ \frac{1}{N} & \text{if} \quad 1 \leq x < 2 \\ \frac{2}{N} & \text{if} \quad 2 \leq x < 3 \\ \vdots & \vdots \quad \vdots \\ \frac{N-1}{N} & \text{if} \quad N-1 \leq x < N \\ 1 & \text{if} \quad x \geq N \end{cases}$$

■

Exercise 17.3 *Write down the distribution function of a Bernoulli random variable with probability of success is $\frac{3}{4}$.*

Solution: Recall the pmf of Bernoulli random variable with probability of success $\frac{3}{4}$ is

$$f_X(x) = \begin{cases} \frac{3}{4} & \text{if} \quad x = 1 \\ \frac{1}{4} & \text{if} \quad x = 0 \\ 0 & \quad \text{otherwise} \end{cases} \tag{17.1}$$

Hence the distribution function is

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{4} & \text{if} \quad 0 \le x < 1 \\ 1 & \text{if} \quad x \ge 1 \end{cases} \tag{17.2}$$

■

Exercise 17.4 *Let $X \sim U[-1,1]$. Determine its cdf.*

Solution: The pdf of X is

$$f(x) = \begin{cases} \frac{1}{2}, & \text{if} -1 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Hence the cdf is

$$F(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ \displaystyle\int_{-1}^{x} \frac{1}{2} dt = \frac{x+1}{2} & \text{if} \quad -1 \le x \le 1 \\ 1 & \text{if} \quad x > 1 \end{cases} \tag{17.3}$$

■

Example 17.5 *Let $X \sim \exp(\lambda)$. Find it's distribution function.*

Solution: The pdf of X is

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

Hence the cdf is

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x} & \text{if} \quad x \ge 0 \end{cases} \tag{17.4}$$

■

We observe the following properties of the distribution function.

1. If X is a discrete random variable, then F_X is discontinuous with jumps occurring at the values of positive probability.
2. If X is an absolutely continuous random variable, then F_X is a continuous function of x. This follows from fundamental theorem of calculus.

Theorem 17.6 (Properties of Distribution function) *Let X be a random variable and F be the distribution function of X. Then F has the following properties.*

(a) *F is non-decreasing.*

(b) *F is right-continuous.*

(c) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$.

Proof:

(a) For $x_1 \le x_2, \{X \le x_1\} \subseteq \{X \le x_2\}$. Hence by monotonicity of probability measure, we have $P\{X \le x_1\} = F(x_1) \le F(x_2) = P\{X \le x_2\}$.

(b) Let us recall the meaning of right-continuity of real-valued function in term of sequences. Let $f : \mathbb{R} \to \mathbb{R}$ and $c \in \mathbb{R}$, we say f is right-continuous at c if for any sequence (x_n) in $[c, \infty)$ such that $x_n \downarrow c$, we have $f(x_n) \to f(c)$.

Let $x_n \downarrow c$. Set $A_n = \{X \le x_n\}$. Then $A_1 \supset A_2 \supset \cdots$ and $\{X \le c\} = \bigcap_{n=1}^{\infty} A_n$.

Therefore continuity property of probability measure gives us

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P\{X \le x_n\} = \lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P\{X \le c\} = F(c).$$

(c) First we show that $\lim_{x\to+\infty} F(x) = 1$.

Let us recall the meaning of $\lim_{x\to\infty} f(x) = l$ in term of sequences. Let $f : \mathbb{R} \to \mathbb{R}$, we say $\lim_{x\to\infty} f(x) = l$ if for any sequence (x_n) in $\mathbb{R}$ such that $x_n \uparrow +\infty$, we have $f(x_n) \to l$.

Let $x_n \uparrow +\infty$, set

$$A_n = \{X \le x_n\}.$$

Then $A_1 \subseteq A_2 \subseteq \cdots$ (since $x_n \le x_{n+1}$ for all n). Also

$$\bigcup_{n=1}^{\infty} A_n = \Omega.$$

By continuity property of probability measure, we have

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P\{X \le x_n\} = \lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\Omega) = 1.$$

Now we show that $\lim_{x\to-\infty} F(x) = 0$.

Let us recall the meaning of $\lim_{x\to-\infty} f(x) = l$ in term of sequences. Let $f : \mathbb{R} \to \mathbb{R}$, we say $\lim_{x\to-\infty} f(x) = l$ if for any sequence (x_n) in $\mathbb{R}$ such that $x_n \downarrow -\infty$, we have $f(x_n) \to l$.
Let $x_n \downarrow -\infty$, set

$$A_n = \{X \le x_n\}.$$

Then $A_1 \supseteq A_2 \supseteq \cdots$ (since $x_n \ge x_{n+1}$ for all n). Also

$$\bigcap_{n=1}^{\infty} A_n = \emptyset.$$

By continuity property of probability measure, we have

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P\{X \le x_n\} = \lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(\emptyset) = 0.$$

■

Theorem 17.7 (Characterization of distribution function) *Let $F : \mathbb{R} \to [0,1]$ be a function, which has the following properties.*

(a) $\lim_{x\to-\infty} F(x) = 0, \ \lim_{x\to+\infty} F(x) = 1.$

(b) F *is non-decreasing.*

(c) F *is right-continuous.*

Then there exist a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X defined on it such that F is the distribution function of random variable X.

Example 17.8 *Consider the following function*

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \le x < 1 \\ \frac{x}{2} & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

Clearly F satisfies all three conditions (a), (b), (c) of Theorem 17.7. Hence F is a CDF of some random variable X.

What type of random variable X is?

Note that X is not absolutely continuous since F is discontinuous at 0.

X is not discrete since X has point mass only at one point and $P(X = 0) = \frac{1}{2} < 1$, means rest $\frac{1}{2}$ unit mass is distributed over the interval $[1, 2]$ according to the uniform mass density $\dfrac{1}{2}$, i.e., $P(1 \leq X \leq 2) = \displaystyle\int_1^2 \frac{1}{2} dx = \frac{1}{2}$. This is an example of a random variable with the range $\{0\} \cup [1, 2]$ and it does not have the pdf. Note that X does take values over a continuous space $[1, 2]$.

Lecture 18: CDF to PMF and PDF, Functions of a Random Variable

03 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Proposition 18.1 *Let X be a random variable and F be the distribution function of X. Then*

$$F(x-) = P\{X < x\},$$

where $F(x-)$ denotes left hand limit at the point x.

Proof: Let us recall the meaning of left-hand limit of real-valued function in term of sequences. Let $f : \mathbb{R} \to \mathbb{R}$ and $c \in \mathbb{R}$, we say left hand limit of f at c exists if there exist some $l \in \mathbb{R}$ such that for any sequence (x_n) in $(-\infty, c)$ such that $x_n \uparrow c$, we have $f(x_n) \to l$.

Let $x_n \uparrow x$. Set $A_n = \{X \leq x_n\}$. Then $A_1 \subseteq A_2 \subseteq \cdots$ and $\{X < x\} = \bigcup_{n=1}^{\infty} A_n$. By continuity property of probability measure, we have

$$F(x-) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P\{X \leq x_n\} = \lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\{X < x\}.$$

■

18.1 From CDF to PMF

Theorem 18.2 *Let $f : I \to \mathbb{R}$ be a monotonic (increasing or decreasing) function, where $I \subset \mathbb{R}$ be an interval. Then the set of discontinuity points of function f is at most countably infinite.*

Corollary 18.3 *Let $F : \mathbb{R} \to [0, 1]$ be a distribution function. Then set of points where F has jump is at most countably infinite.*

Proof: Since any distribution function is increasing, by Theorem 18.2 result follows. ■

Example 18.4 *If F is continuous on $\mathbb{R}$, then set of discontinuity is an empty set, which a finite set of cardinality 0. Recall CDF of Continuous Uniform, Exponential and Normal random variables is continuous.*

Bernoulli, Binomial CDFs have discontinuity at finitely many points.

Geometric, Poisson CDFs have discontinuity at countably infinite points.

From the Proposition 18.1, it follows that for $x \in \mathbb{R}$

$$\boxed{P\{X = x\} = F(x) - F(x-)}$$

In view of what we have obtained just now, Let X be a discrete random variable with distribution function F. Define $f : \mathbb{R} \to [0, 1]$ by

$$f(x) = F(x) - F(x-)$$

Claim 18.5 *f is the probability mass function (pmf) of X.*

In view of Theorem 14.1, we verify the following three conditions.

1. Sicne F is increasing so $F(x) \geq F(x-)$ for all $x \in \mathbb{R}$
2. By corollary 18.3, $\{x : f(x) > 0\}$ is at most countable.
3. Since F can have only left discontinuity,and F is increasing so $F(x) > F(x-)$ if F is left discontinuous at x. So $f(x) > 0$ iff $P(X = x) > 0$. $1 = \sum_{x \in R(X)} P(X = x) = \sum_{x \in R(X)} f(x)$.

Example 18.6 *Suppose X is a random variable with the following distribution function.*

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{4} & \text{if} \quad 0 \leq x < 1 \\ \frac{3}{4} & \text{if} \quad 1 \leq x < 2 \\ 1 & \text{if} \quad x \geq 2. \end{cases}$$

Determine wether X is a discrete random variable? If yes, find it's pmf.

Solution: Recall that the distribution function of a discrete random variable has jump precisely at the points of positive probability.

For given distribution function, the points of jumps are $x = 0, 1, 2$

$$\begin{aligned} P(X = 0) &= F(0) - F(0-) = 1/4 \\ P(X = 1) &= F(1) - F(1-) = 3/4 - 1/4 = 1/2 \\ P(X = 2) &= F(2) - F(2-) = 1 - 3/4 = 1/4 \end{aligned}$$

Also note that

$$P(X=0)+P(X=1)+P(X=2)=1.$$

So there is no other point which has positive probability, hence X is a discrete random variable. ■

Example 18.7 *Suppose X is a random variable with the following cdf*

$$F(x)=\begin{cases} 0 & \text{if} \quad x<1 \\ 1-(1-p)^k & \text{if} \quad k\le x<k+1 \end{cases}$$

where $0<p<1$ *and* $k=1,2,3,\cdots$. *Determine it's pmf.*

Solution: Note that F has jump exactly at $k=1,2,\cdots$. Also

$$\begin{aligned} P(X=k) &= F(k)-F(k-)=[1-(1-p)^k]-[1-(1-p)^{k-1}]=(1-p)^{k-1}-(1-p)^k \\ &= (1-p)^{k-1}p, \quad \text{for } k=1,2,\cdots \end{aligned}$$

■

18.2 From CDF to PDF

Recall that we say that a random variable X has a pdf (or density) if there exist a non-negative, integrable function f such that

$$F_X(x)=P(X\le x)=\int_{-\infty}^{x} f(t)dt, \ \forall x\in\mathbb{R} \tag{18.1}$$

We note that (18.1) does not define f uniquely since we can always change the value of a function at a finite number of points without changing the integral of the function over intervals. Also by Fundamental theorem of calculus, if pdf f is continuous at point $x=a$ then CDF F is differential at $x=a$. So given the CDF of an absolutely continuous random variable X we may obtain the pdf of X, by differentiating F

$$f(x)=F'(x)$$

where equality is valid for those x at which the pdf f is continuous.

Final Method: If CDF F is continuous everywhere on $\mathbb{R}$ then the corresponding random variable may have pdf. One typical way to comput pdf from cdf F by setting $f(x)=F'(x)$ whenever $F'(x)$ exists and $f(x)=0$ otherwise. This defines a pdf of F **provided that F' exists everywhere (except finitely many points) and F' is continuous everywhere but a finite number of points (this will ensure f is integrable)).**

Example 18.8 *Let X be a random variable with the distribution function*

$$F(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ x & \text{if} \quad 0 < x \leq 1 \\ 1 & \text{if} \quad x > 1 \end{cases}$$

Determine wether X is a discrete random variable or an absolutely continuous random variable? Accordingly find it's pmf or pdf.

Solution: Since F is continuous everywhere on $\mathbb{R}$, hence X can not be a discrete random variable. Is it enough to say that X has pdf? Observe that F is not differentiable at two points $x = 0$ and $x = 1$.

$$\lim_{h\uparrow 0} \frac{F(0+h) - F(0)}{h} = \lim_{h\uparrow 0} \frac{0-0}{h} = 0$$

and

$$\lim_{h\downarrow 0} \frac{F(0+h) - F(0)}{h} = \lim_{h\uparrow 0} \frac{h-0}{h} = 1$$

Also

$$\lim_{h\uparrow 0} \frac{F(1+h) - F(1)}{h} = \lim_{h\uparrow 0} \frac{1+h-1}{h} = 1$$

and

$$\lim_{h\downarrow 0} \frac{F(1+h) - F(1)}{h} = \lim_{h\uparrow 0} \frac{1-1}{h} = 0$$

So we can say that

$$F'(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad 0 < x < 1 \\ 0 & \text{if} \quad x > 1 \end{cases}$$

As F' exists and is continuous everywhere except at two points $x = 0, 1$, hence the following is our candidate for the pdf

$$f(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ 1 & \text{if} \quad 0 < x < 1 \\ 0 & \text{if} \quad x \geq 1 \end{cases}$$

Recall that this is the pdf of continuous uniform random variable on the interval $(0, 1)$. ■

Example 18.9 *Let X be the random variable with the distribution function*

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

Determine it's pdf.

Solution: Note that the integrand $\dfrac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}$ is continuous everywhere on $\mathbb{R}$, hence F is a differentiable function on $\mathbb{R}$. Hence pdf is $f(t) = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}$, which is pdf of a standard normal random variable. ■

Exercise 18.10 *Let X be a random variable with the following cumulative distribution function:*

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x < \dfrac{1}{2} \\ \dfrac{3}{4} & \dfrac{1}{2} \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

What type of random variable X is? Accordingly determine it's pmf or pdf. Also compute $P\left(\dfrac{1}{4} \le X < 1\right)$.

Solution:

X is not absolutely continuous since F is discontinuous at $\dfrac{1}{2}, 1$.

X is not discrete since

$$P(X = 1/2) = F(1/2) - F(1/2-) = 3/4 - 1/4 = 1/2$$
$$P(X = 1) = F(1) - F(1-) = 1 - 3/4 = 1/4$$

X has point mass only at these two points and $P(X = 1) + P(X = 1/2) = 3/4 < 1$, means rest $1/4$ unit mass is distributed over the interval $[0, 1/2)$ according to the density $2x$, i.e., $P(0 \le X < 1/2) = \int_0^{1/2} 2x dx = 1/4$.

This is an example of a mixed type of random variable. Now

$$\begin{aligned} P\left(\frac{1}{4} \le X < 1\right) &= P(X < 1) - P\left(X < \frac{1}{4}\right) \\ &= F(1-) - F\left(\frac{1}{4}-\right) \\ &= \frac{3}{4} - \left(\frac{1}{4}\right)^2 = \frac{11}{16} \end{aligned}$$

■

18.3 Functions of a random Variable

Theorem 18.11 *If X is a random variable on a probability space $(\Omega, \mathcal{F}, P)$ and $f : D \subset \mathbb{R} \to \mathbb{R}$ be a Borel measurable function such that $R(X) \subset D$ ($R(X)$ denotes the range of X). Then $f(X)$ is a random variable on $(\Omega, \mathcal{F}, P)$.*

Often, if we are able to model a phenomenon in terms of a random variable X, we will also be concerned with behavior of functions of X. Now we study techniques that allow us to gain complete information (i.e. probability distributions of these functions) about function of X that may be of interest.

Also we have seen that if X is a discrete random variable then $f(X)$ is necessarily discrete. So one can ask given pmf of X how to compute pmf of $f(X)$?

Example 18.12 *Let X be a random variable with pmf*

$$f_X(k) = P(X = k) = \frac{1}{9}, \quad \textit{for } k = -4, -3, -2, \cdots, 3, 4$$

Find the pmf of $|X|$.

Solution: Let $Y = |X|$. So the range of Y is $\{0, 1, \cdots, 4\}$. We want to find pmf of Y,i.e., $P(Y = k)$ for $k = 0, 1, 2, 3, 4$.

$$\{Y = 0\} = \{X = 0\} \implies P(Y = 0) = \frac{1}{9} = \sum_{k:|k|=0} P(X = k)$$

$$\{Y = k\} = \{X = k\} \bigcup \{X = -k\} \implies P(Y = k) = \frac{2}{9} = \sum_{i:|i|=k} P(X = i) \quad \text{for } k = 1, 2, 3$$

■

Lecture 19: PMF & PDF of function of a Random Variable

05 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

19.1 PMF of function of a random variable

Proposition 19.1 *If X is discrete random variable then we can obtain pmf of $Y = g(X)$ as follows:*

$$P\{Y = y\} = \sum_{x:g(x)=y} P\{X = x\}$$

Example 19.2 *Let $X \sim$ B(n,p). Consider random variable $n - X$. Find it's pmf.*

Solution: Since X takes values from $\{0, 1, \cdots, n\}$, hence $Y = n - X$ will also take values $\{0, 1, 2, \cdots, n\}$. Also for given $k \in \{0, 1, 2 \cdots, n\}$, $Y = k \iff X = n - k$. Hence

$$P(Y = k) = P(X = n - k) = \binom{n}{n-k} p^{n-k}(1-p)^{n-(n-k)} = \binom{n}{k} p^{n-k}(1-p)^{k}$$

So $Y \sim B(n, 1-p)$. ∎

Example 19.3 *Suppose $X \sim N(0,1)$ and $g(x) = \begin{cases} -1 & \text{if} \quad x < 0 \\ 0 & \text{if} \quad x = 0 \\ 1 & \text{if} \quad x > 0 \end{cases}$. Then determine pmf of $g(X)$.*

Solution:

$$\begin{aligned} P(Y = -1) &= P(X < 0) = 1/2. \\ P(Y = 0) &= P(X = 0) = 0. \\ P(Y = 1) &= P(X > 0) = 1/2. \end{aligned}$$

∎

19.2 PDF of function of random variable

If X is random variable with density, and g is real-valued function of a real variable, then $g(X)$ may not have pdf (see Example 19.3). Even if we take g to be a continuous function, still $g(X)$ may not have pdf (see Example 19.4).

Let $f, g : I \to \mathbb{R}$ be two functions, $I \subset \mathbb{R}$ is an interval. Then

$$\max\{f(x), g(x)\} = \frac{f(x) + g(x) + |f(x) - g(x)|}{2}, \ \min\{f(x), g(x)\} = \frac{f(x) + g(x) - |f(x) - g(x)|}{2}$$

So if f and g are continuous functions on I then so is $f + g$ and $|f - g|$. Therefore $\max\{f(x), g(x)\}$ and $\min\{f(x), g(x)\}$ are continuous on I.

For example, if we take $f(x) = x$ and $g(x) \equiv 10$ then $h(x) := \min\{x, 10\}$ is continuous function on $\mathbb{R}$.

Example 19.4 *Let $X \sim \exp(5)$. Find the pdf of the random variable $\min\{X, 10\}$ (if it exists).*

Solution: First we determine the distribution function of random variable $Y := \min\{X, 10\}$. Note that the range of the random variable Y is $[0, 10]$ as the range of X is $[0, \infty)$. For $x \in \mathbb{R}$,

$$\{Y \leq x\} = \{X \leq x\} \cup \{10 \leq x\}$$

Note that if $10 > x$ then $\{10 \leq x\} = \emptyset$ and if $10 \leq x$ then $\{10 \leq x\} = \Omega$. Hence

$$\{Y \leq x\} = \begin{cases} \{X \leq x\} & \text{if} \quad x < 10 \\ \Omega & \text{if} \quad x \geq 10 \end{cases}$$

Hence distribution function of Y denoted by F_Y is

$$F_Y(x) = \begin{cases} F_X(x) & \text{if} \quad x < 10 \\ 1 & \text{if} \quad x \geq 10 \end{cases} = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 - e^{-5x} & \text{if} \quad 0 \leq x < 10 \\ 1 & \text{if} \quad x \geq 10 \end{cases}$$

Since F_Y is discontinuous at $x = 10$ hence Y can not have pdf. ∎

Example 19.5 *Let $X \sim U([0, 1])$. Find the pdf of random variable X^2 (if it exists).*

Solution: First we determine the distribution function of X^2.

$$F_{X^2}(x) = P(X^2 \leq x) = \begin{cases} P(\emptyset) = 0 & \text{if} \quad x < 0 \\ P(-\sqrt{x} \leq X \leq \sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x}) = F_X(\sqrt{x}) & \text{if} \quad x \geq 0 \end{cases}$$

we know that

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \leq x \leq 1 \\ 1 & \text{if} \quad x > 1 \end{cases} \tag{19.1}$$

Hence

$$F_{X^2}(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \sqrt{x} & \text{if} \quad 0 \leq x \leq 1 \\ 1 & \text{if} \quad x > 1 \end{cases}$$

The CDF of X^2 is continuous on $\mathbb{R}$ and differentiable everywhere except possibly at $x = 0, 1$. Also the derivative of CDF is continuous everywhere except at $x = 0, 1$. Hence Therefore the density of X^2 would be

$$f_{X^2}(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ \frac{1}{2\sqrt{x}} & \text{if} \quad 0 < x < 1 \\ 0 & \text{if} \quad x \geq 1 \end{cases}$$

■

Lecture 20: PDF of function of a Random Variable

08 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

20.1 Shortcut for computing PDF of function of a random variable

One may use the following formula to determine density of $Y = g(X)$ (cf. Page 130 of book by Papoulis & Pillai). Suppose X is random variable with pdf f_X and $g : I \to \mathbb{R}$ be a function such that $Y := g(X)$ admits a pdf. We wish to compute f_Y in terms of f_X. Compute range of Y, call it $R(Y)$. For $y \in R(Y)$, find all $x \in \mathbb{R}(X)$ such that $y = g(x)$. Suppose $(x_n)_{1 \leq n \leq N}$ (it might be the case $N = +\infty$ as well. This method will not work if uncountable distinct points are mapped to y) are the points such that

$$y = g(x_n) \text{ and } |g'(x_n)| \neq 0 \text{ for all } n = 1, 2, \cdots N.$$

Then

$$f_Y(y) = \sum_{n=1}^{N} \frac{f_X(x_n)}{|g'(x_n)|}, \text{ here } x_n \text{ is written in terms of } y. \tag{20.1}$$

Example 20.1 *Let X be a random variable with pdf f. Then Determine pdf of $aX + b$, where $a \neq 0, b \in \mathbb{R}$.*

Solution: The range of Random variable $aX + b$ is the set $\{ax + b : x \in R(X)$. Now if $y \in R(aX + b)$ then $x = y - b/a$. Note that $g(x) = ax + b$ is differentiable and $g'(x) = a$ So

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

■

When shortcut may lead to wrong conclusion! The best thing about the shortcut formula (20.1) is that we are not suppose to compute the CDF of $g(X)$. But it assume that suppose $g(X)$ has pdf. This is the dangerous point. Look at the Example 19.4 again. The function $g(x) = \min\{x, 10\}$ and $X \sim \exp(5)$. The range of random variable $g(X) = [0, 10]$. For each $y \in [0, 10)$ there is exactly one $x \in [0, 10)$ such that $g(x) = y$. But for $y = 10$, all $x \geq 10$ are solutions of $g(x) = 10$. If one argue that it does not matter if pdf of Y is not defined at a single point $y = 10$ (so we set it equal to zero) and apply the formula (20.1) then we obtain

$$g(x) = \begin{cases} x & \text{if } 0 \leq x < 10 \\ 10 & \text{if } x \geq 10 \end{cases}$$

hence

$$g'(x) = \begin{cases} 1 & \text{if } 0 \leq x < 10 \\ 0 & \text{if } x > 10 \end{cases}$$

$$f_Y(y) = \frac{f_X(x)}{|g'(x)|} = \begin{cases} 5e^{-5y} & \text{if } 0 \leq y < 10 \\ 0 & \text{if } y > 10 \end{cases} \tag{20.2}$$

But is it legitimate pdf? No. it does not integrate to 1.

20.2 Normal Random Variable

A normal random variable has several special properties. The following one is particularly important. That is Normality is Preserved under Linear Transformations.

Proposition 20.2 *If $X \sim N(\mu, \sigma^2)$, and if $a \neq 0, b$ are real numbers, and consider the random variable*

$$\boxed{Y = aX + b}$$

Then $Y \sim N(a\mu + b, a^2\sigma^2)$

Proof: By Example 20.1, the pdf of Y is given by

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

where

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

Therefore

$$f_Y(x) = \frac{1}{|a|}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-b-a\mu)^2}{2a^2\sigma^2}}, \qquad x \in \mathbb{R}$$

So Y is a normal random variable with mean $b + a\mu$ and variance $a^2\sigma^2$. ∎

A normal random variable X with zero mean and unit variance is said to be a standard normal. It's distribution function is denoted by $N(\cdot)$:

$$N(x) = P(X \le x) = P(X < x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$$

We can not get closed form for $N(x)$. One use numerical technique to find the approximate values of $N(x)$. It is recorded in a table and is a very useful tool for calculating various probabilities involving normal random variables. Note that the table only provides the values of $N(x)$ for $x > 0$, because the omitted values can be found using the symmetry of the pdf. For example, if X is a standard normal random variable, for $x > 0$ we have

$$N(-x) = P(X \le -x) = P(X \ge x) = 1 - P(X < x) = 1 - N(x)$$

Let X be a normal random variable with mean μ and variance σ^2. We define a new random variable $Z := \dfrac{X-\mu}{\sigma}$. Then Z is a linear function of X hence normal. Also Z mean zero ($\because \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$) and variance is one($\because \frac{1}{\sigma^2}\sigma^2 = 1$). Thus Z is a standard normal random variable. This fact allows us to calculate the probability of any event defined in term of X: we redefine the event in terms of Z, then use the standard normal normal table. This is how it is done:

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = N\left(\frac{x-\mu}{\sigma}\right)$$

Example 20.3 *Let X be a normal random variable with parameters μ and σ^2. Find $P(\mu - \sigma \le X \le \mu + \sigma)$.*

Solution:

$$\begin{aligned}
P(\mu-\sigma \le X \le \mu+\sigma) = P(X \le \mu+\sigma) - P(X < \mu-\sigma) &= N\left(\frac{\mu+\sigma-\mu}{\sigma}\right) - N\left(\frac{\mu-\sigma-\mu}{\sigma}\right)\\
&= N(1) - N(-1) = N(1) - [1 - N(1)] = 2N(1) - 1\\
&= 2 \times 0.8413 - 1 = 1.6826 - 1\\
&= 0.6826
\end{aligned}$$

∎

Remark 20.4 *The values in the table for standard normal CDF are correct upto 4 decimal places. In the table you see that* $N(3.09) = 1.0000$ *but strictly speaking* $N(3.09) = P(X \leq 3.09) < 1$*. But if we are ready to accept the correct values upto 4 decimal places then* $0.99999...$ *may be written as* 1*.*

TABLE 1 Values of the standard normal distribution function

x	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5363	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9648	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9874	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.	.9987	.9990	.9993	.9995	.9997	.9998	.9998	.9999	.9999	1.0000

Lecture 21: PDF of function of a Random Variable

10 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Example 21.1 *Let X be a random variable with the pdf*

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \ x \in \mathbb{R}.$$

Random variable X is called a Cauchy random variable. Find the pdf of $\dfrac{1}{X}$ if it exists.

Solution: Range of random variable X is $(-\infty, \infty)$. Hence $Y := \dfrac{1}{X}$ is not defined when $X = 0$. But $P(X = 0) = 0$, as X is absolutely continuous. Range of the random variable Y is $(-\infty, \infty) \setminus \{0\}$. For any $y \in R(Y)$ we have unique $x \in R(X)$, such that $x = \frac{1}{y}$. The function $g(x) = \frac{1}{x}$ has derivative $g'(x) = -\frac{1}{x^2}$ Hence

$$f_Y(y) = \frac{f_X(\frac{1}{y})}{|g'(\frac{1}{y})|} = \frac{\frac{y^2}{\pi(1+y^2)}}{y^2} = \frac{1}{\pi(1+y^2)}, \ \text{if } y \neq 0$$

Note that a pdf is not affected if we change it's value at a point. Therefore if we define $f_Y(0) = \frac{1}{\pi}$ then we see that Y is a Cauchy random variable. ■

Example 21.2 *Let $X \sim \exp(\lambda)$. Find the pdf of $\sqrt{X}$, if it exists.*

Solution: Range of random variable X is $[0, \infty)$. Hence range of the random variable $Y := \sqrt{X}$ is $[0, \infty)$. For any $y \in R(Y)$ we have unique $x \in R(X)$, such that $x = y^2$. The function $g(x) = \sqrt{x}$ has derivative $g'(x) = \frac{1}{2\sqrt{x}}$. For $y = 0$, $g'(y^2) = +\infty$. Hence

$$f_Y(y) = \frac{f_X(y^2)}{|g'(y^2)|} = \frac{\lambda e^{-\lambda y^2}}{\frac{1}{2y}} = 2\lambda y e^{-\lambda y^2}, \ \text{if } y > 0$$

Note that a pdf is not affected if we change it's value at a point. Therefore if we define $f_Y(0) = 0$. For $y < 0$, we anyway have $f_Y(y) = 0$. ■

Example 21.3 *Let $X \sim N(0,1)$. Find the pdf of e^X, if it exists.*

Solution: Range of random variable X is $\mathbb{R}$. Hence range of the random variable $Y := e^X$ is $(0, \infty)$. For any $y \in R(Y)$ we have unique $x \in R(X)$, such that $x = \ln y$. The function $g(x) = e^x$ has derivative $g'(x) = e^x$. Hence

$$f_Y(y) = \frac{f_X(\ln y)}{|g'(\ln y)|} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{(\ln y)^2}{2}}}{y} = \frac{1}{y\sqrt{2\pi}} e^{-\frac{(\ln y)^2}{2}}, \text{ if } y > 0$$

The random variable Y is called lognormal random variable. ∎

Example 21.4 *Let $X \sim U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Find the pdf of $\tan X$, if it exists.*

Solution: Range of random variable X is $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Hence range of the random variable $Y := \tan X$ is $(-\infty, \infty)$. For any $y \in R(Y)$ we have unique $x \in R(X)$, such that $x = \tan^{-1} y$. The function $g(x) = \tan x$ has derivative $g'(x) = \sec^2 x \neq 0$ on $R(X)$. Note that $\sec^2 x = 1 + \tan^2 x$. Hence

$$f_Y(y) = \frac{f_X(\tan^{-1} y)}{|g'(\tan^{-1} y)|} = \frac{\frac{1}{\pi}}{1 + y^2} = \frac{1}{\pi(1 + y^2)}, \text{ for all } y \in \mathbb{R}$$

So Y is a Cauchy random variable. ∎

Example 21.5 *Consider $X \sim U[-1, 1]$. Another random variable Y is formed by using the transformation $Y = X^2 + X$. Find density function of the new transformed random variable Y.*

Solution: Note that $R(X) = [-1, 1]$. Consider the function $g : [-1, 1] \to \mathbb{R}$, defined as $g(x) = x^2 + x$. Since $g'(x) = 2x + 1$ so g is strictly increasing on $[-\frac{1}{2}, 1]$ and strictly decreasing on interval $[-1, \frac{1}{2}]$. $g(-1) = 0, g(-\frac{1}{2}) = -\frac{1}{4}, g(1) = 2$. Also g is continuous, hence intermediate value theorem we conclude that the range of g is $[-\frac{1}{4}, 2]$. Now for given $y \in R(Y)$, we solve $x^2 + x - y = 0$ for $x \in [-1, 1]$.

$$x = \frac{-1 \pm \sqrt{1 + 4y}}{2}$$

are the two solutions. Note that if $y \in [-\frac{1}{4}, 0]$ then we have two solutions

$$x = \frac{-1 \pm \sqrt{1 + 4y}}{2} \in R(X)$$

But when $y \in [0,2]$ then only one solution $x = \frac{-1+\sqrt{1+4y}}{2} \in R(X)$ another solution $x = \frac{-1-\sqrt{1+4y}}{2} < -1$ for $y > 0$.

Also note that $g'(x) = 0$ only if $x = -\frac{1}{2}$, i.e., only for $y = 0 \in R(Y)$. As pdf is not affected by value at a single point.

1. If $y \in [-1/4, 0)$, then

$$f_Y(y) = \frac{f_X(\frac{-1+\sqrt{1+4y}}{2})}{\left|g'\left(\frac{-1+\sqrt{1+4y}}{2}\right)\right|} + \frac{f_X(\frac{-1-\sqrt{1+4y}}{2})}{\left|g'\left(\frac{-1-\sqrt{1+4y}}{2}\right)\right|} = \frac{\frac{1}{2}}{\sqrt{1+4y}} + \frac{\frac{1}{2}}{\sqrt{1+4y}} = \frac{1}{\sqrt{1+4y}}$$

2. If $y \in (0, 2]$

$$F_Y(y) = \frac{f_X(\frac{-1+\sqrt{1+4y}}{2})}{\left|g'\left(\frac{-1+\sqrt{1+4y}}{2}\right)\right|} = \frac{1}{2\sqrt{1+4y}}$$

Hence we get the pdf of Y.

$$f_Y(y) = \begin{cases} 0 & \text{if} \quad y \leq -\frac{1}{4} \\ \frac{1}{\sqrt{4y+1}} & \text{if} \quad -\frac{1}{4} < y \leq 0 \\ \frac{1}{2\sqrt{4y+1}} & \text{if} \quad 0 \leq y < 2 \\ 0 & \text{if} \quad y \geq 2 \end{cases}$$

You may check $\int_{-\infty}^{\infty} f_Y(y)dy = 1$.

$$\int_{-1/4}^{0} \frac{1}{2\sqrt{y+\frac{1}{4}}}dy = \left[\sqrt{y+1/4}\right]_{-1/4}^{0} = 1/2 - 0 = 1/2$$

$$\int_{0}^{2} \frac{1}{4\sqrt{y+\frac{1}{4}}}dy = \frac{1}{2}\left[\sqrt{y+1/4}\right]_{0}^{2} = 1/2[3/2 - 1/2] = 1/2$$

■

Lecture 22: Expectation

12 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

22.1 Introduction

The distribution function of a random variable gives us complete information (statistically) about it. In many settings of theoretical and practical interest, however, one is satisfied with cruder information about the random variable. (And if the distribution in question is unknown or only partly known, one may have to resort to these cruder information measures in any case to make a virtue out of necessity.) While a variety of measures can be developed along these lines, principal among them are measures of central tendency in a random variable. The simplest descriptors of central tendency in a random variable include the mean, the median, and the mode. Each has utility but, while no single measure of central tendency is completely satisfactory in all applications, the mean, also called the expectation, is by far the most important and the most prevalent in theory and application.

22.2 Motivation

Suppose you play a certain game and your rewards are 1 with probability $\frac{1}{6}$, 2 with probability $\frac{1}{2}$ and 4 with probability $\frac{1}{3}$. So basically X is random variable which denotes rewards it takes three values $\{1, 2, 4\}$ and we are given it's pmf. So we ask this, how much you expect to get on the average if you play the game a zillion times? So you can think as follows: probability of getting 1 is $\frac{1}{6}$ means $\frac{1}{6}$ times (out of a zillion games) we get 1 (relative frequency interpretation of probability) similarly $\frac{1}{3}$ times we get 2 etc. so

$$\frac{1}{6} \times 1 + \frac{1}{2} \times 2 + \frac{1}{3} \times 4 = 2.5$$

Basically what we have done here is:

$$P\{X = 1\} \cdot 1 + P\{X = 2\} \cdot 2 + P\{X = 4\} \cdot 4$$

Expected value or Expectation of a discrete random variable is a weighted average, where we interpret probabilities as weights associated with the value.

22.3 Expectation of Discrete Random Variable

Definition 22.1 *Let X be a discrete random variable with the pmf $f_X(x)$. Then expectation of X denoted by $E(X)$ is defined as*

$$E(X) = \sum_{x \in R(X)} x f_X(x)$$

provided the right hand side series converges absolutely, i.e., $\sum_{x \in R(X)} |x| f_X(x) < \infty$.

Remark 22.2

1. *If X has finite range then the sum $\sum_{x \in R_X} x f_X(x)$ is a finite sum and there is no need to check the condition $\sum_{x \in R_X} |x| f_X(x) < \infty$ separately (as finite sum of real numbers is a real number).*
2. *If range of discrete random variables X is infinite then the sum $\sum_{x \in R_X} x f_X(x)$ is an infinite sum and the condition $\sum_{x \in R_X} |x| f_X(x) < \infty$ implies that the infinite sum $\sum_{x \in R_X} x f_X(x)$ converges to a finite value that is independent of the order in which the various terms are summed (see Example 22.7). Note that the series $\sum_{x \in R_X} x f_X(x)$ may converge but the series $\sum_{x \in R_X} |x| f_X(x)$ may not. In that case we say that $E[X]$ does not exist (see Example 22.6 below).*
3. *If $R_X \subset [0, \infty)$ then convergence of the infinite series $\sum_{x \in R_X} x f_X(x)$ is equivalent to absolute convergence.*

Example 22.3 *Let X be a Bernoulli(p) random variable where $0 < p < 1$. Determine EX (if it exists).*

Solution: Recall that X has pmf $P(X = 0) = 1 - p, P(X = 1) = p$. Since X takes finitely many values hence we don't have to bother about absolute convergence, hence EX exists. Further $E(X) = p$. ■

Remark 22.4 *Though the number $E(X)$ is called expected value of random variable X, but do not expect the value $E(X)$ when X is observed. In section 22.2, $E(X) = 2.5$ where as*

X assumes values $1, 2, 4$. So expected value 2.5 is never going to be observed. In Example 22.3, X takes values $0, 1$, where as expected value p is between 0 and 1. So expected value p is never going to be observed.

Expectation of X, tell us about the average behavior of random variable X or 'typical value' taken by the random variable. Sometimes this estimate could be very vague.

Example 22.5 *Consider a random variable X that takes the value 2^k with probability 2^{-k} for each $k \in \mathbb{N}$. Find it's mean (if it exists).*

Solution: Since X take values over positive integers $\{2, 2^2, 2^3, \cdots\}$, hence absolute convergence is same convergence. Now

$$EX = \sum_{k=1}^{\infty} 2^k P(X = 2^k) = \sum_{k=1}^{\infty} 1 = +\infty$$

So EX does not exists. ■

Example 22.6 *Let X have the PMF given by*

$$P\left(X = \frac{(-1)^{n+1}3^n}{n}\right) = \frac{2}{3^n}, \quad n = 1, 2, \cdots$$

$\frac{1}{n^p}$ diverges

Then

$$\sum_{x \in R_X} |x| P(X = x) = \sum_{n=1}^{\infty} \frac{3^n}{n} \times \frac{2}{3^n} = \sum_{n=1}^{\infty} \frac{2}{n} = \infty.$$

$p > 1$ $p \leq 1$

Therefore $E[X]$ does not exist, although the series

converges

$$\sum_{x \in R_X} x P(X = x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}3^n}{n} \times \frac{2}{3^n} = 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} < \infty.$$

Question: One may ask why we need the condition of absolute convergence, if expectation is defined as the sum $\sum_{x \in R(X)} x P(X = x)$?

Answer: The absolute convergence implies that the infinite sum $\sum_{x \in R(X)} x f_X(x)$ converges to a finite value that is independent of the order in which the various terms are summed. In order to appreciate this point let us look at the following example.

Example 22.7 *Consider the convergent series* $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots$ *and one of its rearrangements*

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \cdots \tag{22.1}$$

in which two positive terms are always followed by one negative. If s is the sum of $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$, *then one can show that the rearrangement series (22.1) converges to a number strictly bigger then s.*

If we rearrange the series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots$ *in such a way that two negative terms follow a positive term:*

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \cdots \tag{22.2}$$

then one can show that the rearranged series (22.2) converges to $\frac{s}{2}$.

In fact, the theorem below says something very dramatic and startling.

Theorem 22.8 (Riemann) *A conditionally convergent series can be made to converge to any arbitrary real number or even made to diverge by a suitable rearrangement of its terms.*

In view of Riemann's Theorem, the following theorem highlights the importance of absolute convergence.

Theorem 22.9 *If the series* $\sum_n a_n$ *converges absolutely, then every rearrangement of the series* $\sum_n a_n$ *converges, and they all converge to the same sum.*

Example 22.10 *Let $X \sim B(n, p)$. Find EX.*

Solution: We know that a Binomial random variable X has pmf

$$P(X = k) = \binom{n}{k} p^k (1-p)^k, \quad k = 0, 1, 2, \cdots, n$$

So

$$\begin{aligned}
EX &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} n\frac{(n-1)!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} = n\sum_{k=1}^{n} \binom{n-1}{k-1}p^k(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} \binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} = np\underbrace{\sum_{i=0}^{n-1} \binom{n-1}{i}p^i(1-p)^{(n-1)-i}}_{=1} \qquad (\text{put } k = i+1) \\
&= np
\end{aligned}$$

■

Lecture 23: Expectation

15

February 2021

Sunil

Kumar Gauttam

De-

partment of Mathematics, LNMIIT

Example 23.1 *Let $X \sim$geometric(p). Then X takes values over set of positive integers. Hence*

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\sum_{k=1}^{\infty} -\frac{d}{dp}((1-p)^k) = -p\frac{d}{dp}\left(\sum_{k=1}^{\infty}(1-p)^k\right)$$
$$= -p\frac{d}{dp}\left(\frac{(1-p)}{1-(1-p)}\right) = -p\frac{d}{dp}\left(\frac{1}{p}-1\right) = -p\frac{-1}{p^2} = \frac{1}{p}$$

Example 23.2 *Let $X \sim$Poisson(λ). Then X takes values over set of non-negative integers. Hence*

$\frac{\lambda^n}{n!} = e^{\lambda}$

$$E(X) = \sum_{n=0}^{\infty} n\frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda}\sum_{n=1}^{\infty}\frac{\lambda^n}{n-1!} = \lambda e^{-\lambda}\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{n-1!} = \lambda e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} = \lambda$$

23.1 Expectation of an Absolutely Continuous Random Variable

Definition 23.3 *Let X be a random variable with pdf f. Then expectation of X denoted by $E(X)$ is defined as*

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

provided the improper integral in the right hand side converges absolutely, i.e., $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$.

This is similar to the discrete case except that the PMF is replaced by the PDF, and summation is replaced by integration.

Remark 23.4 *We emphasize that the condition* $\int_{-\infty}^{\infty} |x|f_X(x)dx < \infty$ *must be checked before it can be concluded that* $E[X]$ *exists and* EX *equals* $\int_{-\infty}^{\infty} xf_X(x)dx$.

23.2 Improper integral

Definition 23.5 *let $f : \mathbb{R} \to \mathbb{R}$ be a function that is integrable on $[a, b]$ for all $a, b \in \mathbb{R}$ with $a \leq b$. We say that $\int_{-\infty}^{\infty} f(t)dt$ is convergent if both $\int_{0}^{\infty} f(t)dt$ and $\int_{-\infty}^{0} f(t)dt$ are convergent.*

23.2.1 Cauchy Principle value

Definition 23.6 *If the limit $\lim\limits_{x\to\infty} \int_{-x}^{x} f(t)dt$ exists, then this limit is called the Cauchy principal value of the improper integral $\int_{-\infty}^{\infty} f(t)dt$.*

If improper integral $\int_{-\infty}^{\infty} f(t)dt$ converges (as per Definition 23.5), then since

$$\int_{-x}^{x} f(t)dt = \int_{-x}^{0} f(t)dt + \int_{0}^{x} f(t)dt, \text{ for all } x \geq 0.$$

Therefore Cauchy principal value of $\int_{-\infty}^{\infty} f(t)dt$ exists and is equal to $\int_{-\infty}^{\infty} f(t)dt$. But Cauchy principal value may exit even when integral $\int_{-\infty}^{\infty} f(t)dt$ is divergent. For example consider the improper integral $\int_{-\infty}^{\infty} \sin t dt$. It diverges because $\int_0^x \sin t \ dt = 1 - \cos x$,and $\lim\limits_{x\to\infty} \cos x$ does not exist. But

$$\int_{-x}^{x} \sin t dt = 0, \text{ for all } x \geq 0$$

So Cauchy principle value exists and it is equal to zero.

Theorem 23.7 *If $f : \mathbb{R} \to \mathbb{R}$ be a nonnegative function and the Cauchy principal value $\lim\limits_{x\to\infty} \int_{-x}^{x} f(t)dt$ exists then improper integral $\int_{-\infty}^{\infty} f(t)dt$ is convergent.*

23.3 Solved Examples

Example 23.8 *Let $X \sim U[a, b]$. Find mean of X.*

Solution: The pdf of X is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \le x \le b \\ 0, & \text{otherwise.} \end{cases}$$

Hence

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \frac{1}{b-a} \int_a^b x dx \\ &= \frac{b+a}{2} \end{aligned}$$

Since pdf is non-zero on a finite interval, hence absolute convergence of improper integral is not needed to be checked. ■

Example 23.9 *Consider a random variable X with pdf $f(x) = \dfrac{1}{\pi(1+x^2)}, -\infty < x < \infty$. Find it's mean (if it exists).*

Solution: Note that Cauchy principle value exists and is equal to zero. For each $a > 0$ we have

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

But the improper integral $\int_{-\infty}^{\infty} x f(x) dx$ diverges, since

$$\begin{aligned} \int_0^{\infty} x f(x) dx &= \int_0^{\infty} \frac{2x}{\pi(1+x^2)} dx \\ &= \int_0^{\infty} \frac{2x}{\pi(1+x^2)} dx = \frac{1}{\pi} \times \lim_{x\to\infty} \log(1+x^2) = \infty \end{aligned}$$

Therefore expectation does not exist. ■

Example 23.10 *Let $X \sim \exp(\lambda)$. Find its means (if it exists).*

Solution:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \lambda \left(\frac{1}{-\lambda} x e^{-\lambda x} \Big|_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x} dx \right) = \frac{1}{-\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda}$$

■

By Parts $\int u \cdot v \, dx = u \int v dx - \int \frac{du}{dx} \times (\int v dx) dx$

[check karlena lmao]

Example 23.11 *Let $X \sim N(0,1)$. Find it's mean (if it exists).*

Solution: Recall the pdf of a $N(0,1)$ random variable is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

Note that density is positive on entire negative real axis therefore we need to check the absolute convergence of the improper integral $\int_{-\infty}^{\infty} xf(x)dx$. Note that the integrand of improper integral $\int_{-\infty}^{\infty} |x|f(x)dx$ is nonnegative, therefore by Theorem 23.7 it is enough to observe that

$$\lim_{a\to+\infty} \int_{-a}^{a} |x|f(x)dx = 2\frac{1}{\sqrt{2\pi}} \int_0^{\infty} xe^{-\frac{x^2}{2}} dx = 2\frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-u} du = 2\frac{1}{\sqrt{2\pi}}.$$

Hence one conclude that $EX < \infty$. Now we compute it as follows:

$$\int_0^{\infty} xf(x)dx = \frac{1}{\sqrt{2\pi}}$$

$$\int_{-\infty}^{0} xf(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} xe^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}}$$

Hence $EX = 0$. ∎

Lecture 24: Expectation of Function of a Random Variable

16 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

24.1 Expectation of Function of Random Variable

Theorem 24.1 1. *Let X be a discrete random variable with pmf f_X, and let $g : \mathbb{R} \to \mathbb{R}$ be any function. Then,*

$$E[g(X)] = \sum_{x \in R_X} g(x) f_X(x) \tag{24.1}$$

provided $\displaystyle\sum_{x \in R_X} |g(x)| f_X(x) < \infty$.

2. *Let X be a random variable with pdf f and $g : \mathbb{R} \to \mathbb{R}$ be a Borel function (e.g., piecewise continuous, continuous) such that the integral $\displaystyle\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$. Then,*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \tag{24.2}$$

Remark 24.2 1. *If X is discrete, for any function g, random variable $Z := g(X)$ admits pmf. So one can use the formula $E[Z] = \sum_{z \in R_Z} z f_Z(z)$ to compute expectation of Z. But if one is just interested in $E[g(X)]$, then formula (24.1) paves the way without calculating pmf of $g(X)$.*

2. *If X has density then for some functions g it is possible that $g(X)$ does not have pdf (see Example 24.3), but the formula (24.2) paves the way to define the expectation of $g(X)$.*

   Suppose $g(X)$ has pdf then the formula (24.2) tells us how to compute expectation, without calculating the pdf of $g(X)$.

Example 24.3 *Let $Z \sim N(0, 1)$ be a random variable. Then find the $E[\max\{Z, 0\}]$ (if it exists).*

Solution: Random variable Z has pdf

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

Hence

$$\begin{aligned}
E\left[\max\{Z,0\}\right] &= \int_{-\infty}^{\infty} \max\{x,0\} f(x) dx \\
&= \int_{-\infty}^{0} \max\{x,0\} f(x) dx + \int_{0}^{\infty} \max\{x,0\} f(x) dx \\
&= \int_{-\infty}^{0} 0 \cdot f(x) dx + \int_{0}^{\infty} x f(x) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} x e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-t} dt \quad \text{(substituting } \tfrac{x^2}{2} = t\text{)} \\
&= \frac{1}{\sqrt{2\pi}}
\end{aligned}$$

■

24.2 Expectation of a Mixed Random Variable

In the Example 24.3, it was mere a coincidence that the a continuous function of a random variable with pdf turn out to be a mixed type random variable and using the formula 24.2, we were able to compute the expectation of the mixed random variable. But given a mixed type of random variable it may not be possible to express this a function of some random variable with density. For example suppose random variable X has following CDF

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \le x < 1 \\ \frac{x}{2} & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

point par expection

+ $\int$ x differention of the function dx

and Suppose random variable Y has the following CDF

$$F_Y(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x < \dfrac{1}{2} \\ \dfrac{3}{4} & \dfrac{1}{2} \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

But one can ask whether X and Y has finite mean or not? The idea as follows: In general, the CDF of a mixed random variable Y can be written as the sum of a continuous function and a staircase function:

$$F_Y(y) = C(y) + D(y).$$

We differentiate the continuous part of the CDF. In particular, let's define

$$c(y) = \frac{dC(y)}{dy}, \text{ wherever } C(y) \text{ is differentiable.}$$

Note that this is not a valid PDF as it does not integrate to one. Also, let $\{y_1, y_2, y_3, \cdots\}$ be the set of jump points of $D(y)$, i.e., the points for which $P(Y = y_k) > 0$. We then have

$$\int_{-\infty}^{\infty} c(y)dy + \sum_{y_k} P(Y = y_k) = 1.$$

The expected value of Y can be obtained as

$$EY = \int_{-\infty}^{\infty} yc(y)dy + \sum_{y_k} y_k P(Y = y_k).$$

Example 24.4 *Let us revisit the Example 24.3, with this new method. We compute the CDF of $Y := \max\{Z, 0\}$ as*

$$F_Y(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt & \text{if} \quad x \geq 0 \end{cases}$$

Hence $C(y) = N(y)$ and $D(y)$ has jump at $y = 0$. Hence

$$EY = \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + 0 \times P(Y = 0) = \frac{1}{\sqrt{2\pi}}$$

Example 24.5

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ \frac{x}{2} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

Hence

$$EX = \int_1^2 x\frac{1}{2}dx + 0P(X = 0) = \left[\frac{x^2}{4}\right]_1^2 = \frac{3}{4}$$

Example 24.6

$$F_Y(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x < \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

Then

$$EY = \int_0^{\frac{1}{2}} x2xdx + \frac{1}{2}P(X = 0.5) + 1P(X = 1) = \frac{2}{3}\left[x^3\right]_0^{\frac{1}{2}} + \frac{1}{2} \times \frac{1}{2} + \frac{1}{4} = \frac{2}{3} \times \frac{1}{8} + \frac{1}{2} = \frac{7}{12}$$

Lecture 25: Variance

17 February 2021

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

25.1 Higher Order Moments

Definition 25.1 *Let X be a random variable. Then $E[X^n]$ is called the nth moment of X and $E[(X - E(X))^n]$ is called the nth central moment of X. The second central moment is called the variance.*

The most important quantity associated with a random variable X, other than the mean, is its variance, it is denoted by $var(X)$. The variance provides the measure of dispersion around its mean. Another measure of dispersion is the standard deviation, which is defined as positive square root variance and is denoted by σ_X. So $\sigma_X = \sqrt{var(X)}$.

If X is a discrete random variable then by Theorem 24.1,

$$Var(X) = E[(X - \mu)^2] = \sum_{x \in R(X)} (x - \mu)^2 f_X(x),$$

where $\mu = EX$.

Example 25.2 *Let X be a Bernoulli(p) random variable where $0 < p < 1$. Then find it's variance.*

Solution: Recall $E(X) = p$.

$$\begin{aligned} Var(X) = E[(X-p)^2] &= (0-p)^2 P(X=0) + (1-p)^2 P(X=1) = p^2(1-p) + (1-p)^2 p \\ &= p(1-p)[p + (1-p)] = p(1-p) \end{aligned}$$

■

Example 25.3 *Let X be a Geometric(p) random variable where $0 < p < 1$. Then find it's variance.*

Solution: Recall $E(X) = \frac{1}{p}$.

$$
\begin{aligned}
Var(X) = E\left[\left(X - \frac{1}{p}\right)^2\right] &= \sum_{k=1}^{\infty}\left(k - \frac{1}{p}\right)^2 P(X = k) \\
&= \sum_{k=1}^{\infty}\left(k^2 + \frac{1}{p^2} - \frac{2}{p}k\right) p(1-p)^{k-1} \\
&= p\sum_{k=1}^{\infty} k^2(1-p)^{k-1} + \frac{1}{p}\sum_{k=1}^{\infty}(1-p)^{k-1} - 2\sum_{k=1}^{\infty} k(1-p)^{k-1} \\
&= \frac{2}{p^2} - \frac{1}{p} + \frac{1}{p^2} - \frac{2}{p^2} = \frac{1-p}{p^2}
\end{aligned}
$$

Recall for $|x| < 1$

$$
\begin{aligned}
&\sum_{k=1}^{\infty} x^{k-1} = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \\
&\implies \sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2} (\text{ term by term differentiation of power series }) \\
&\sum_{k=1}^{\infty} kx^{k-1} = \sum_{k=0}^{\infty}(k+1)x^k \implies \\
&\frac{2}{(1-x)^3} = \sum_{k=1}^{\infty}(k+1)kx^{k-1} = \sum_{k=1}^{\infty} k^2x^{k-1} + \sum_{k=1}^{\infty} kx^{k-1} \\
&\implies \sum_{k=1}^{\infty} k^2x^{k-1} = \frac{2}{(1-x)^3} - \frac{1}{(1-x)^2}
\end{aligned}
$$

■

If X is a random variable with density f_X then by Theorem 24.1,

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx,$$

where $\mu = EX$.

Example 25.4 *Let X be a Uniform$[a, b]$ random variable. Then find it's variance.*

Solution: Recall $E(X) = \frac{a+b}{2}$.

$$\begin{aligned}
Var(X) &= E\left[\left(X - \frac{a+b}{2}\right)^2\right] = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}dx \\
&= \frac{1}{3(b-a)}\left[\left(x - \frac{a+b}{2}\right)^3\right]_{x=a}^{x=b} = \frac{1}{3(b-a)}\left[\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3\right] = \frac{(b-a)^3}{12(b-a)} \\
&= \frac{(b-a)^2}{12}
\end{aligned}$$

■

Theorem 25.5 (Properties of Expectation) *Let X be a random variable and let a, b and c be real numbers. Suppose g_1 and g_2 are real-valued function of one real variable such that $E[g_1(X)] < \infty$ and $E[g_2(X)] < \infty$. Then*

(a) $E[ag_1(X) + bg_2(X) + c] = aE[g_1(X)] + bE[g_2(X)] + c$.

(b) *If $g_1(x) \geq g_2(x)$ for all x, then $E[g_1(X)] \geq E[g_2(X)]$.*

Proof: We will prove only for X which are absolutely continuous, the discrete case being similar.

(a) By (24.2)

$$\begin{aligned}
E[ag_1(X) + bg_2(X) + c] &= \int_{-\infty}^{\infty} [ag_1(x) + bg_2(x) + c]\, f_X(x)dx \\
&= \int_{-\infty}^{\infty} ag_1(x)f_X(x)dx + \int_{-\infty}^{\infty} bg_2(x)f_X(x)dx + \int_{-\infty}^{\infty} cf_X(x)dx \\
&= a\int_{-\infty}^{\infty} g_1(x)f_X(x)dx + b\int_{-\infty}^{\infty} g_2(x)f_X(x)dx + c\int_{-\infty}^{\infty} f_X(x)dx \\
&= aE[g_1(X)] + bE[g_2(X)] + c
\end{aligned}$$

(b)

$$\because g_1(x) \geq g_2(x) \implies g_1(x)f_X(x) \geq g_2(x)f_X(x) \text{ for all } x \ (\because f_X(x) \geq 0)$$

$$\int_{-\infty}^{\infty} g_1(x)f_X(x)dx \geq \int_{-\infty}^{\infty} g_2(x)f_X(x)dx$$

$$\text{Hence } E[g_1(X)] \geq E[g_2(X)] \ \text{(By 24.2)}$$

■

Alternate Expression for Variance

$$\begin{aligned}\text{Var}(X) = E[(X-\mu)^2] &= E[X^2+\mu^2-2\mu X]\\ &= E[X^2]+\mu^2-2\mu E[X]\\ &\text{(By Theorem 25.5 (a) with } g_1(x)=x^2, g_2(x)=x, a=1, b=-2\mu, c=\mu^2)\\ &= E[X^2]+\mu^2-2\mu^2 = E[X^2]-\mu^2,\end{aligned}$$

where $\mu = EX$.

Hence

$$\boxed{\text{Var}(X) = E[X^2] - (E[X])^2.}$$

Second Moment implies first moment Note that function $g(x) = (x-\mu)^2 \geq 0$ for any real number μ. Let $\mu = EX$, If a random variable X has finite variance then it follows from the Theorem 25.5 (b), that $\text{Var}(X) \geq 0$. Hence by alternate form we get

$$\infty > E[X^2] - (EX)^2 \geq 0 \implies E[X^2] \geq (EX)^2$$

Hence if $E[X^2] < \infty$ then $E[X] < \infty$. In other words, if a random variable has second moment finite, then it's mean is finite. Also, from alternate expression it follows that if a random variable has finite variance then it's first and second moment must be finite.

Proposition 25.6 *Let X be a random variable with finite variance and $a, b \in \mathbb{R}$. Then $Var(aX+b) = a^2 var(X)$.*

Proof: By alternate definition of variance

$$\begin{aligned}\text{Var}(aX+b) &= E[aX+b]^2 - [E(aX+b)]^2\\ &= E[a^2X^2+b^2+2abX] - [aEX+b]^2\\ &= a^2E[X^2]+b^2+2abE[X]-a^2[EX]^2-b^2-2abEX\\ &= a^2 var(X)\end{aligned}$$

■

Example 25.7 *If X is a random variable taking nonnegative integer values and has finite mean then show that*

$$EX = \sum_{n=1}^{\infty} P\{X \geq n\}.$$

Solution: Note that for any given $n = 0, 1, 2, \cdots$

$$\{X \geq n\} = \{X = n\} \cup \{X = n+1\} \cup \{X = n+3\} \cup \cdots$$

$$\begin{aligned}
\text{Hence } EX &= \sum_{n=0}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} n[P(X \geq n) - P(X \geq n+1)] \\
&= \sum_{n=1}^{\infty} nP(X \geq n) - \sum_{n=1}^{\infty} nP(X \geq n+1) \\
&= \sum_{n=1}^{\infty} nP(X \geq n) - \sum_{n=2}^{\infty} (n-1)P(X \geq n) \\
&= \sum_{n=1}^{\infty} nP(X \geq n) - \sum_{n=2}^{\infty} nP(X \geq n) + \sum_{n=2}^{\infty} P(X \geq n) \\
&= \sum_{n=1}^{\infty} P\{X \geq n\}.
\end{aligned}$$

■

Example 25.8 *Suppose that $Y = -2X + 3$. If we know $EY = 1$ and $EY^2 = 9$, find EX and Var(X).*

Solution:

$$\begin{aligned}
EY = -2EX + 3 &\implies EX = 1 \\
Var(Y) = E[Y^2] - [E(Y)]^2 &= 9 - 1 = 8. \\
Var(Y) = (-2)^2 var(X) &\implies Var(X) = 2.
\end{aligned}$$

■

Lecture 26: Multiple Random Variables

11 March, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Probabilistic models often involve several random variables. For example, in a medical diagnosis context, the results of several tests may be significant (These different tests would be observations on different random variables.), or in a networking context, the workloads of several routers may be of interest (These different workloads would be observations on different random variables, one for each router measured). Thus, we need to know how to describe and use probability models that deals with more than one random variable at a time.

All of these random variables are associated with the same experiment, sample space, and probability law, and their values may relate in interesting ways. This motivates us to consider simultaneously several random variables, i.e., the notion of random vector.

Definition 26.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A map $X : \Omega \to \mathbb{R}^n, X(\omega) = (X_1(\omega), X_2(\omega), \cdots, X_n(\omega))$ is called a n-dimensional random vector on $(\Omega, \mathcal{F}, P)$ if each X_i is a random variable on $(\Omega, \mathcal{F}, P)$.*

We will discuss mainly bivariate random vectors, i.e., two random variables. The extension to higher dimension is analogue.

Example 26.2 *Consider the experiment of tossing two fair dice. Then we know sample space would be $\Omega = \{(i, j) : i, j = 1, \cdots, 6\}$ with all outcomes be equally likely and σ-algebra is power set. Let us consider the following random variable*

$$X((i, j)) = i + j, Y((i, j)) = |i - j|$$

Then (X, Y) is a random vector.

Having define a random vector (X, Y), we can now discuss probabilities of events that are defined in terms of (X, Y). For example we can ask what is $P(X = 5 \text{ and } Y = 3)$?, Henceforth, we will write $P(X = 5, Y = 3)$ for $P(X = 5 \text{ and } Y = 3)$. Read the comma as "and". It is easy to see that there are only two sample points $(1, 4)$ and $(4, 1)$ that yields $X = 5$ and $Y = 3$. Please note that we are interested in those sample points (i, j) such that $X((i, j)) = 5, Y((i, j)) = 3$ simultaneously. Hence

$$P(X = 5, Y = 3) = \frac{2}{36} = \frac{1}{18}$$

Definition 26.3 (Discrete random vector) *We say that a random vector $X = (X_1, X_2)$ is a discrete random vector, if X_1 and X_2 both are discrete random variables.*

It follows from the definition that range of a discrete random vector is either finite or countable infinite.

1. If range of X_1 and X_2 both are finite, then X has finite range.
2. If range of X_1 is finite and X_2 is countable infinite, then X has countable infinite range. Similarly, If range of X_2 is finite and X_1 is countable infinite, then X has countable infinite range.
3. If range of X_1 and X_2 both are countable infinite, then X has countable infinite range.

Definition 26.4 (joint pmf) *Let $X = (X_1, X_2)$ be a discrete random vector. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by*

$$f(x_1, x_2) = P\{X_1 = x_1, X_2 = x_2\}, \forall\, x_1, x_2 \in \mathbb{R}$$

Then f is called joint pmf of X.

If it is necessary to stress the fact that f is the joint pmf of (X_1, X_2) rather than some other random vector, the notation $f_{X_1,X_2}(x_1, x_2)$ will be used.

The joint pmf of (X, Y) completely determines the probability distribution of the discrete random vector (X, Y), just as the pmf of a discrete random variable completely defines its distribution.

For the discrete random vector defined in Example 26.2, there are 21 possible values of (X, Y).

Y \ X	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$
1	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0
2	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0
3	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0
4	0	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0	0
5	0	0	0	0	0	$\frac{1}{18}$	0	0	0	0	0

As in the one-dimensional case, this function f has the following three properties:

1. $f(x, y) \geq 0$, for all $(x, y) \in \mathbb{R}^2$.
2. The set $\{(x, y) : f(x, y) \neq 0\}$ is at most countably infinite subset of $\mathbb{R}^2$.

3. $\displaystyle\sum_{(x,y)\in R(X,Y)} f(x,y) = 1.$

Any real-valued function f defined on $\mathbb{R}^2$ having these three properties will be called a two dimensional joint pmf.

One can easily verify all three properties for the joint pmf of Example 26.2

The joint PMF determines the probability of any event that can be specified in terms of the discrete random variables X_1 and X_2.

Theorem 26.5 *Let $X = (X_1, X_2)$ be a discrete random vector, with the joint pmf f. Then for any $A \subseteq \mathbb{R}^2$,*

$$P\{(X_1, X_2) \in A\} = \sum_{(x,y)\in A} f(x,y)$$

Exercise 26.6 *Suppose X be a random variable taking three values $-2, 1$ and 3, and let Y be a random variable that assume four values $-1, 0, 4, 6$. Their joint probabilities are given by the following table.*

X \ Y	-1	0	4	6
-2	$\frac{1}{9}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{9}$
1	$\frac{2}{9}$	0	$\frac{1}{9}$	$\frac{1}{9}$
3	0	0	$\frac{1}{9}$	$\frac{4}{27}$

Compute the probability of the event that XY is odd.

Solution: $\{XY \text{ is odd}\} = \{X = 1, Y = -1\} \bigcup \{X = 3, Y = -1\}$. Therefore

$$P(XY \text{ is odd}) = f(1,-1) + f(3,-1) = \frac{2}{9}$$

■

26.1 Marginal PMF

Even if we are considering a random vector (X, Y) there may be probabilities of interest that involve only one of the random variables in the random vector. We may wish to know $P(X = 2)$, for instance in Example 26.2. The random variable X is discrete and its probability distribution if described by it's pmf f_X (we now use the subscript to distinguish

$f_X(x)$ from $f_{X,Y}(x,y)$). We now call $f_X(x)$ the marginal pmf of X to emphasize the fact that it is the pmf of X but in the context of the joint pmf of the random vector (X,Y).

The marginal pmf of X or Y is easily calculated from the joint pmf of (X,Y).

Proposition 26.7 *If f is the joint pmf of X and Y, then*

$$f_X(x) = \sum_{y \in R(Y)} f(x,y), \quad f_Y(y) = \sum_{x \in R(X)} f(x,y),$$

Proof: Note that $\Omega = \bigcup_{y \in R(Y)} \{Y = y\}$. For each $x \in \mathbb{R}$, the events $\{X = x, Y = y\}, y \in R(Y)$ are disjoint and their union is the event $\{X = x\}$. Thus

$$\begin{aligned} f_X(x) &= P(X = x) \\ &= P\left(\bigcup_{y \in R(Y)} \{X = x, Y = y\} \right) \\ &= \sum_{y \in R(Y)} P\left(X = x, Y = y\right) \\ &= \sum_{y \in R(Y)} f(x,y) \end{aligned}$$

∎

Lecture 27: Marginal PMFs and Random Vector with density

12 March, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Example 27.1 *Suppose X be a random variable taking two values 1 and 2, and let Y be a random variable that assume four values $1, 2, 3, 4$. Their joint probabilities are given by the following table.*

X \ Y	1	2	3	4
1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$

Determine the marginal pmf of random variables X and Y.

Solution:

$$f_X(1) = \sum_{y=1}^{4} f(1, y) = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{1}{2}$$

$$f_X(2) = \sum_{y=1}^{4} f(2, y) = \frac{1}{16} + \frac{1}{16} + \frac{1}{4} + \frac{1}{8} = \frac{1}{2}$$

$$f_Y(1) = \sum_{x=1}^{2} f(x, 1) = \frac{1}{4} + \frac{1}{16} = \frac{5}{16} = f_Y(3)$$

$$f_Y(2) = \sum_{x=1}^{2} f(x, 2) = \frac{1}{8} + \frac{1}{16} = \frac{3}{16} = f_Y(4)$$

Clearly X is uniformly distributed but Y is not. ■

The marginal pmfs f_X and f_Y do not completely determines the joint pmf of X and Y. Indeed there are many different joint pmfs that have same marginal pmfs.

Example 27.2 *Define a joint pmf by*

$$f(0, 0) = \frac{1}{12}, f(1, 0) = \frac{5}{12}, f(0, 1) = f(1, 1) = \frac{3}{12}$$

The marginal pmf of Y *is* $f_Y(0) = f_Y(1) = \frac{1}{2}$. *The marginal pmf of* X *is* $f_X(0) = \frac{1}{3}, f_X(1) = \frac{2}{3}$.

Define a joint pmf by

$$f(0,0) = f(0,1) = \frac{1}{6}, f(1,0) = f(1,1) = \frac{1}{3}$$

The marginal pmf of Y *is* $f_Y(0) = f_Y(1) = \frac{1}{2}$. *The marginal pmf of* X *is* $f_X(0) = \frac{1}{3}, f_X(1) = \frac{2}{3}$.

Thus, it is hopeless to try to determine the joint pmf from the knowledge of only the marginal pmfs. The marginals does not capture the information how X and Y are interrelated.

27.1 Random Vectors with density

Definition 27.3 *A random vector* (X, Y) *defined on a probability space* $(\Omega, \mathcal{F}, P)$ *is called absolutely continuous if there is a nonnegative function* $f(x, y)$ *defined on* $\mathbb{R}^2$*, called the joint pdf of* (X, Y) *(sometimes just joint density of* (X, Y)*), such that*

$$P((X, Y) \in S) = \iint_S f(x, y)dxdy,$$

for every Borel subset S *of* $\mathbb{R}^2$.

Example of Borel subsets of $\mathbb{R}^2$: polygons, disks, ellipses, and finite or countably unions of such shapes. Open set, closed set, their (finite or counatble) union or intersections etc.

In particular, the probability that the value of (X, Y) falls within an rectangle $[a, b] \times [c, d]$ is

$$P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f(x, y)dxdy.$$

and can be interpreted as the volume of region lying below the surface $z = f(x, y)$ and above the rectangle $[a, b] \times [c, d]$.

Example 27.4 *The joint probability density function of* X *and* Y *is*

$$f(x, y) = \begin{cases} 2 & \text{if} \quad x > 0, y > 0, 0 < x + y < 1 \\ 0 & \quad \textit{elsewhere} \end{cases}$$

Then find $P\left(X + Y < \frac{1}{2}\right)$.

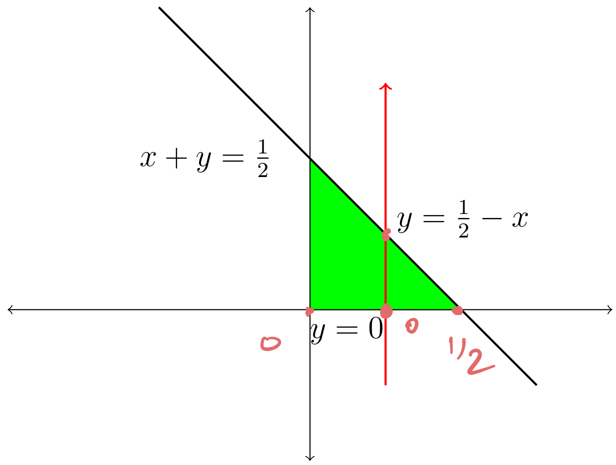
↓

① write in the given below form

② find Limits of double integral.

Solution:

Define $A := \left\{(x,y) \in \mathbb{R}^2 : x + y < \frac{1}{2}\right\}$. Then



$$P\left(X+Y<\frac{1}{2}\right) = \iint_A f(x,y)dxdy$$
$$= \int_0^{\frac{1}{2}} \left(\int_0^{\frac{1}{2}-x} 2dy\right) dx = \int_0^{\frac{1}{2}} [2y]\Big|_0^{\frac{1}{2}-x} dx$$
$$= \int_0^{\frac{1}{2}} (1-2x)dx = \left[x - x^2\right]\Big|_0^{\frac{1}{2}} = \frac{1}{4}$$

■

27.2 Properties of Joint Density

Let f be the joint pdf of random variable X and Y. Then

$$P(-\infty < X < \infty, -\infty < Y < \infty) = P(\Omega \cap \Omega) = 1. \tag{27.1}$$

But by definition we have

$$P(-\infty < X < \infty, -\infty < Y < \infty) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy \tag{27.2}$$

Hence joint pdf integrate to 1 on the entire plane.

Theorem 27.5 (Characterization of joint pdf) *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function satisfying*

(a) $f(x,y) \geq 0$ *for all* $(x,y) \in \mathbb{R}^2$.

(b) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1.$

Then there exists a probability space $(\Omega, \mathcal{F}, P)$ *and a random vector* (X,Y) *defined on it such that* f *is the joint pdf of* (X,Y).

Example 27.6 *Let* $f(x,y) = ce^{-\frac{x^2-xy+4y^2}{2}}, x,y \in \mathbb{R}$. *Find the value of* c *such that* f *is a joint pdf.*

Solution:

If f is a joint pdf then $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t,s)dtds = 1.$

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy &= c\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2-xy+4y^2}{2}}dxdy \\
&= c\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2-2\cdot x\cdot\frac{y}{2}+\frac{y^2}{4}-\frac{y^2}{4}+4y^2}{2}}dxdy \\
&= c\int_{-\infty}^{\infty} e^{-\frac{15y^2}{8}}\left(\int_{-\infty}^{\infty} e^{-\frac{(x-\frac{y}{2})^2}{2}}dx\right)dy \\
&= c\int_{-\infty}^{\infty} e^{-\frac{15y^2}{8}}\left(\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}du\right)dy \quad (\text{put } x = u+\tfrac{y}{2}) \\
&= c\sqrt{2\pi}\int_{-\infty}^{\infty} e^{-\frac{15y^2}{8}}dy \quad (\because \int_{-\infty}^{\infty}\frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}du = 1) \\
&= c\sqrt{2\pi}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}\frac{2du}{\sqrt{15}} \quad (\text{put } y = \tfrac{2u}{\sqrt{15}}) \\
&= c\sqrt{2\pi}\frac{2}{\sqrt{15}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}du \\
&= c\sqrt{2\pi}\frac{2}{\sqrt{15}}\sqrt{2\pi}
\end{aligned}$$

Hence $c = \dfrac{\sqrt{15}}{4\pi}$. ■

Lecture 28: Marginal PDFs and Joint CDF

13 March, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Proposition 28.1 *If f is the joint pdf of X_1 and X_2, then*

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2, \quad f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_1.$$

Proof:

$$\begin{aligned} P\{X_1 \le x_1\} &= P\{X_1 \le x_1, X_2 < \infty\} \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f(x, x_2)dx_2 dx \\ &= \int_{-\infty}^{x_1} g(x)dx, \text{ where } g(x) = \int_{-\infty}^{\infty} f(x, x_2)dx_2 \end{aligned}$$

Hence $f_{X_1}(x_1) = g(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2$. ■

Example 28.2 *The joint probability density function of (X, Y) is given as*

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

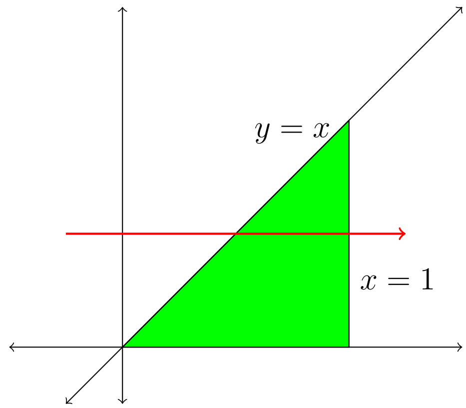
Determine the marginal density of random variables X and Y.

Solution:

Density of Y: We have

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx, \ \forall y \in \mathbb{R}$$

Since $f(x, y)$ is zero for $y \ge 1$ and $y \le 0$, therefore $f_Y(y) = 0$ if $y \ge 1$ or $y \le 0$.



Now for $y \in (0, 1)$ we have

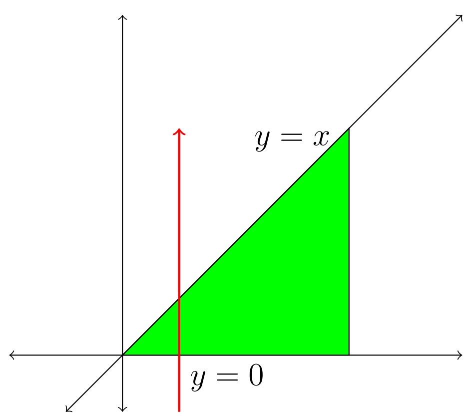
$$f_Y(y) = \int_y^1 6(1-x)dx = 6\left(x - \frac{x^2}{2}\right)\Big|_y^1 = 3(y-1)^2.$$

Hence

$$f_Y(y) = \begin{cases} 3(y-1)^2 & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Density of X: We have

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy.$$



Again $f(x, y)$ is zero for $x \geq 1$ and $x \leq 0$, therefore $f_X(x) = 0$ if $x \geq 1$ or $x \leq 0$. For $x \in (0, 1)$ we have

$$f_X(x) = \int_0^x 6(1-x)dy = 6\left(y - xy\right)\Big|_0^x = 6(x - x^2)$$

Therefore

$$f_X(x) = \begin{cases} 6(x - x^2), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

■

28.1 Joint Distribution Function

Definition 28.3 (joint distribution function) *Let $X = (X_1, X_2)$ be a random vector on $(\Omega, \mathcal{F}, P)$. Then the function $F : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$F(x_1, x_2) = P\{X_1 \leq x_1, X_2 \leq x_2\}$$

is called the joint distribution function of (X_1, X_2).

Example 28.4 *Suppose the joint pmf of X and Y is given as*

$$f(0,0) = f(0,1) = \frac{1}{6}, f(1,0) = f(1,1) = \frac{1}{3}$$

Then determine the joint CDF of X and Y.

Solution: We have

$$F(x,y) = \sum_{(i,j): i \leq x, j \leq y} P(X = i, Y = j)$$

Hence

$$F(x,y) = \begin{cases} 0, & x < 0 \text{ or } y < 0 \\ \frac{1}{6}, & 0 \leq x < 1, 0 \leq y < 1 \\ \frac{2}{6}, & 0 \leq x < 1, y \geq 1 \\ \frac{1}{2}, & x \geq 1, 0 \leq y < 1 \\ 1, & x \geq 1, y \geq 1 \end{cases}$$

Look at the following figures to get an idea how to compute joint CDF $F(x, y)$. In all the figure below we denote the infinite rectangle $(-\infty, x] \times (-\infty, y]$ by the R, which is the green lined region in the figures.

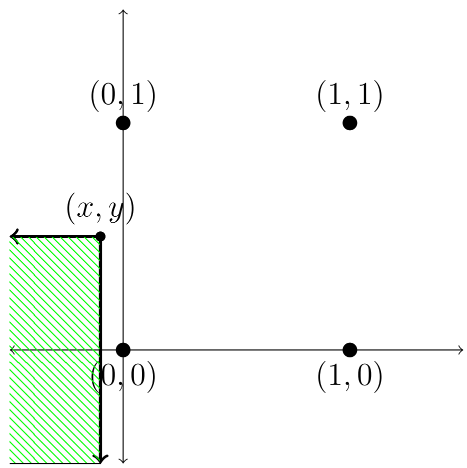


Fig. 1. Region R : if $x < 0$

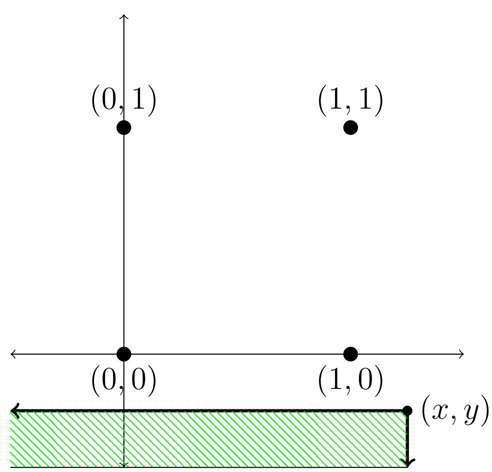


Fig. 2. Region R : if $y < 0$

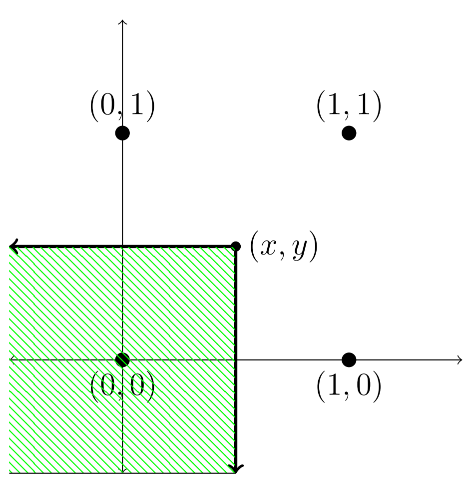


Fig. 3. R : if $0 \leq x < 1, 0 \leq y < 1$

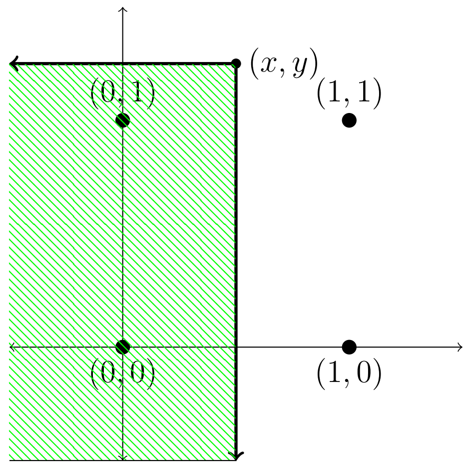


Fig. 4. R : if $0 \le x < 1, y \ge 1$

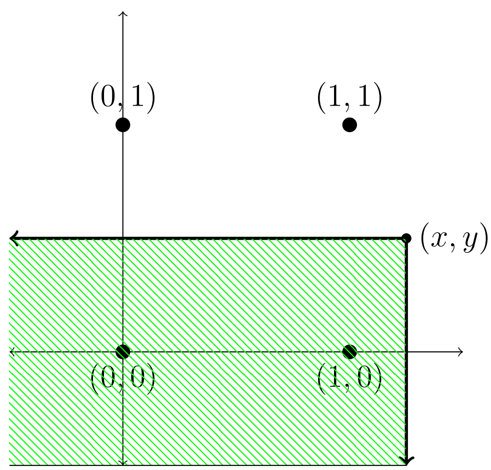


Fig. 5. R : if $x \ge 1, 0 \le y < 1$

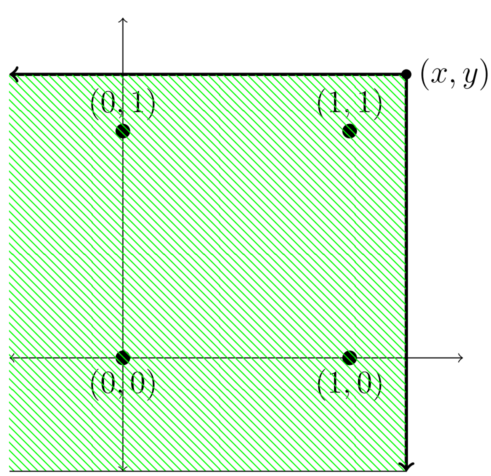


Fig. 6. R : if $x \ge 1, y \ge 1$

■

Lecture 29: Properties of Joint CDF

16

March, 2019

Sunil

Kumar Gauttam

De-

partment of Mathematics, LNMIIT

Theorem 29.1 (Properties of Joint CDF) *Let F be the joint distribution function of a random vector $X = (X_1, X_2)$. Then F satisfies the following.*

1. (a) *$F(x_1, x_2) \to 0$ as $x_i \to -\infty$ for at least one i where $i \in \{1, 2\}$. That is*
      i. $\lim_{x \to -\infty} F(x, y) = 0, \quad \forall y \in \mathbb{R}$.
      ii. $\lim_{y \to -\infty} F(x, y) = 0, \quad \forall x \in \mathbb{R}$.
      iii. $\lim_{(x,y) \to (-\infty,-\infty)} F(x, y) = 0$.

   (b) $\lim_{(x,y) \to (\infty,\infty)} F(x, y) = 1$.

2. *F is right continuous in each argument. That is*

$$\lim_{h \to 0^+} F(x + h, y) = \lim_{h \to 0^+} F(x, y + h) = F(x, y), \text{ for all } (x, y) \in \mathbb{R}^2.$$

3. *F is nondecreasing in each argument. That is*

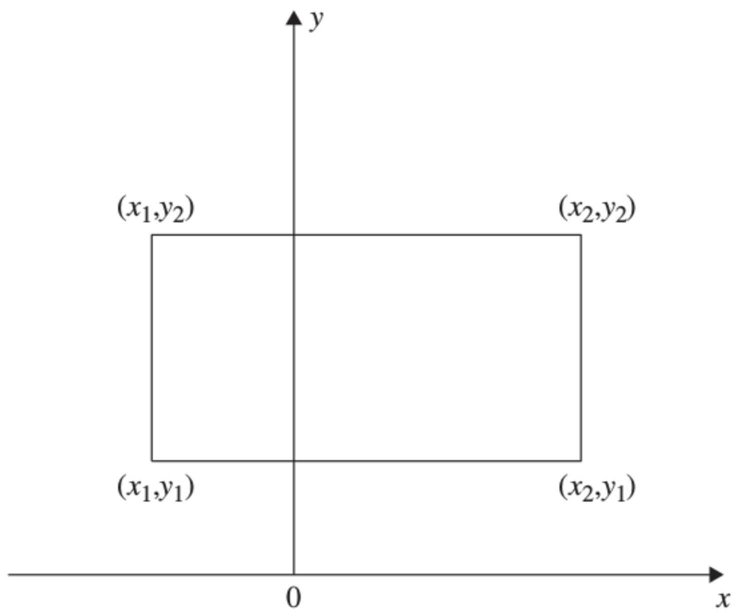
$$F(x, y) \le F(x + h, y), \ \forall h > 0.$$
$$F(x, y) \le F(x, y + k), \ \forall k > 0.$$

4. *For every $(x_1, y_1), (x_2, y_2)$ with $x_1 < x_2$ and $y_1 < y_2$ the following inequality holds:*

$$F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) \ge 0. \tag{29.1}$$

Proof: Proof of the inequality (29.1). By figure we have

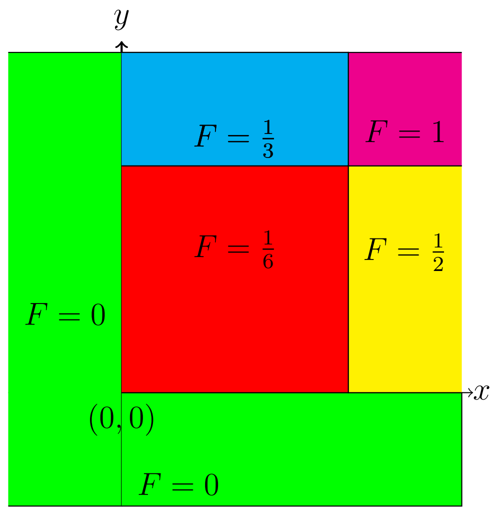
$$\begin{aligned} P\{x_1 < X \le x_2, y_1 < Y \le y_2\} &= P\{X \le x_2, Y \le y_2\} + P\{X \le x_1, Y \le y_1\} \\ &\quad - P\{X \le x_1, Y \le y_2\} - P\{X \le x_2, Y \le y_1\} \\ &= F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) \ge 0 \end{aligned}$$



■

Example 29.2 *Let recall the joint cdf obtained in Example 28.4.*

$$F(x,y) = \begin{cases} 0, & x < 0 \ or \ y < 0 \\ \frac{1}{6}, & 0 \le x < 1, 0 \le y < 1 \\ \frac{2}{6}, & 0 \le x < 1, y \ge 1 \\ \frac{1}{2}, & x \ge 1, 0 \le y < 1 \\ 1, & x \ge 1, y \ge 1 \end{cases}$$



Let us verify the properties 1-4 in Theorem 29.1.

1. (a) *For any given* $y \in \mathbb{R}$, $F(x, y) = 0$ *for all* $x < 0$. *Hence* $\lim_{x\to-\infty} F(x, y) = 0, \quad \forall y \in \mathbb{R}$.
   (b) *Similarly, For any given* $x \in \mathbb{R}$, $F(x, y) = 0$ *for all* $y < 0$. *Hence* $\lim_{y\to-\infty} F(x, y) = 0, \quad \forall x \in \mathbb{R}$.
   (c) *Also since* $F(x, y) = 0$ *in third quadrant, hence* $\lim_{(x,y)\to(-\infty,-\infty)} F(x, y) = 0$.

2. (a) **(Right continuity w.r.t. x)**
   i. *If* $c < 0$ *then along line* $y = c$, *joint cdf* $F = 0$ *which is continuous for all* $x \in \mathbb{R}$.
   ii. *If* $0 \le c < 1$, *then along line* $y = c$,

$$F(x, c) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{6}, & 0 \le x < 1, \\ \dfrac{1}{2}, & x \ge 1 \end{cases}$$

   which is right continuous everywhere.
   iii. *If* $c \ge 1$, *then along line* $y = c$,

$$F(x, c) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{3}, & 0 \le x < 1, \\ 1, & x \ge 1 \end{cases}$$

   which is right continuous everywhere.

(b) **(Right continuity w.r.t. y)**

i. *If $c < 0$ then along line $x = c$, joint cdf $F = 0$ which is continuous for all $y \in \mathbb{R}$.*

ii. *If $0 \le c < 1$, then along line $x = c$,*

$$F(c, y) = \begin{cases} 0, & y < 0 \\ \frac{1}{6}, & 0 \le y < 1, \\ \frac{1}{3}, & y \ge 1 \end{cases}$$

which is right continuous everywhere.

iii. *If $c \ge 1$, then along line $x = c$,*

$$F(c, y) = \begin{cases} 0, & y < 0 \\ \frac{1}{2}, & 0 \le y < 1, \\ 1, & y \ge 1 \end{cases}$$

which is right continuous everywhere.

3. *Non-decreasing in each argument is also clear from the discussion about right continuity in each coordinate.*

4. *For every $(x_1, y_1), (x_2, y_2)$ with $x_1 < x_2$ and $y_1 < y_2$, we have*

$$F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) = P\{x_1 < X \le x_2, y_1 < Y \le y_2\},$$

So non-negativity of probability measure the inequality hold true.

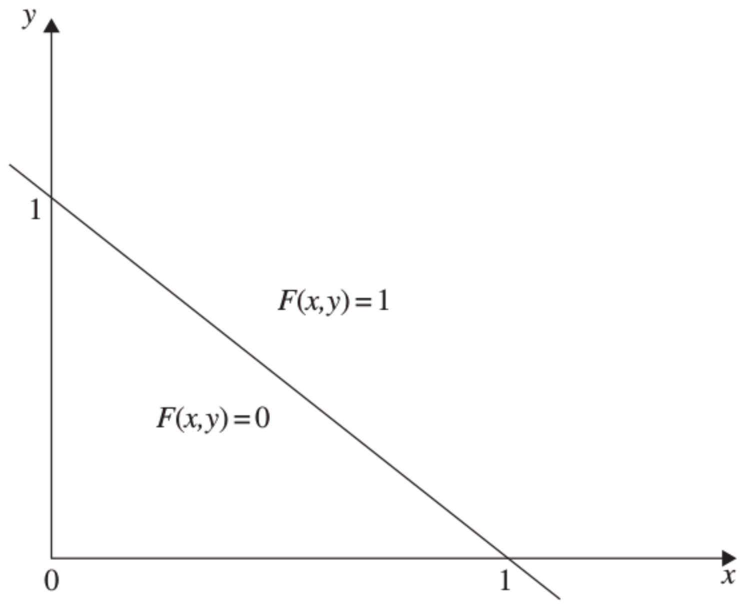
Theorem 29.3 (Characterization of Joint CDF) *Any function F defined on $\mathbb{R}^2$ and satisfying conditions 1-4 in Theorem 29.1 can be identified as a joint distribution function of some 2-dimensional random vector.*

Example 29.4 *Let F be a function of two variables defined by*

$$F(x, y) = \begin{cases} 0, & x < 0 \text{ or } x + y < 1 \text{ or } y < 0 \\ 1, & \text{otherwise} \end{cases}$$

Determine whether F is a joint CDF?

Solution:



Let us verify properties 1-4 in Theorem 29.1.

1. (a) For any given $y \in \mathbb{R}$, $F(x, y) = 0$ for all $x < 0$. Hence $\lim_{x \to -\infty} F(x, y) = 0, \quad \forall y \in \mathbb{R}$.
   (b) Similarly, For any given $x \in \mathbb{R}$, $F(x, y) = 0$ for all $y < 0$. Hence $\lim_{y \to -\infty} F(x, y) = 0, \quad \forall x \in \mathbb{R}$.
   (c) Also since $F(x, y) = 0$ in third quadrant, hence $\lim_{(x,y) \to (-\infty, -\infty)} F(x, y) = 0$.

2. (a) **(Right continuity w.r.t. x)**
      i. If $c < 0$ then along line $y = c$, joint cdf $F = 0$ which is continuous for all $x \in \mathbb{R}$.
      ii. If $0 \leq c < 1$, then along line $y = c$,

$$F(x, c) = \begin{cases} 0, & x < 1 - c \\ 1, & x \geq 1 - c \end{cases}$$

      which is right continuous everywhere.
      iii. If $c \geq 1$, then along line $y = c$,

$$F(x, c) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

      which is right continuous everywhere.
   (b) **(Right continuity w.r.t. y)**

i. If $c < 0$ then along line $x = c$, joint cdf $F = 0$ which is continuous for all $y \in \mathbb{R}$.

ii. If $0 \leq c < 1$, then along line $x = c$,

$$F(c, y) = \begin{cases} 0, & y < 1 - c \\ 1, & y \geq 1 - c \end{cases}$$

which is right continuous everywhere.

iii. If $c \geq 1$, then along line $x = c$,

$$F(c, y) = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases}$$

which is right continuous everywhere.

3. Non-decreasing in each argument is also clear from the discussion about right continuity in each coordinate.

4. Take $(x_1, y_1) = (\frac{1}{3}, \frac{1}{3}), (x_2, y_2) = (1, 1)$. Then

$$F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) = 1 - 1 + 0 - 1 = -1.$$

Hence given F is not a CDF.

■

Lecture 30: Marginal Distributions

18

March, 2019

Sunil

Kumar Gauttam

De-

partment of Mathematics, LNMIIT

30.1 Marginal Distribution Function:

Given a random vector $X = (X_1, X_2, \cdots, X_n)$, the distribution function of X_1, denoted by F_{X_1} is called the marginal distribution of X_1. Similarly the marginal distribution function F_{X_i} of X_i is defined. Given the joint distribution function F of X, one can recover the corresponding marginal distributions as follows.

$$F_{X_1}(x_1) = P\{X_1 \leq x_1\} = P\{X_1 \leq x_1, X_2 \in \mathbb{R}, \cdots, X_n \in \mathbb{R}\} = \lim_{x_k \to +\infty \forall\ k \geq 2} F(x_1, x_2, \cdots, x_n)$$

Similarly,

$$F_{X_i}(x_i) = \lim_{x_k \to +\infty \forall\ k \neq i} F(x_1, x_2, \cdots, x_n)$$

Given the marginal distribution functions of X_1 and X_2, in general it is impossible to construct the joint distribution function of (X_1, X_2). Note that marginal distribution functions doesn't contain information about the dependence of X_1 over X_2 and vice versa.

Example 30.1 *Suppose the joint pmf of X and Y is given as*

$$f(0,0) = f(0,1) = \frac{1}{6}, f(1,0) = f(1,1) = \frac{1}{3}$$

Let recall the joint cdf obtained in Example 28.4.

$$F(x,y) = \begin{cases} 0, & x < 0 \text{ or } y < 0 \\ \frac{1}{6}, & 0 \leq x < 1, 0 \leq y < 1 \\ \frac{2}{6}, & 0 \leq x < 1, y \geq 1 \\ \frac{1}{2}, & x \geq 1, 0 \leq y < 1 \\ 1, & x \geq 1, y \geq 1 \end{cases}$$

$$F_X(x) = \lim_{y \to \infty} F(x, y) = \begin{cases} 0, & x < 0 \\ \frac{1}{3}, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

This corresponds to pmf $P(X = 0) = \frac{1}{3}, P(X = 1) = \frac{2}{3}$.

Similarly we have

$$F_Y(y) = \lim_{x \to \infty} F(x, y) = \begin{cases} 0, & y < 0 \\ \frac{1}{2}, & 0 \le y < 1 \\ 1, & y \ge 1 \end{cases}$$

This corresponds to pmf $P(Y = 0) = P(Y = 1) = \frac{1}{2}$.

30.2 Joint CDF and Joint Density

The joint cdf is usually not very handy to for a discrete random vector. But for a random vector with density we have the important relationship.

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(x, y) dx dy, \forall\, x_1, x_2 \in \mathbb{R} \tag{30.1}$$

1. As in the one-dimensional case, joint density $f(x, y)$ is not uniquely defined by (30.1). We can change f at a finite number of points or even over a finite number of smooth curves in the plane without affecting integrals of f over sets in the plane.

2. Given joint CDF $F(x, y)$, we can determines the joint PDF $f(x, y)$ through the following formula

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y} \tag{30.2}$$

   for every (x, y) at which the joint PDF f is continuous. This relationship is useful when in situations where an expression for $F(x, y)$ can be found. The mixed partial derivative can be computed to find joint pdf.

Example 30.2 *Suppose a joint cdf is given as*

$$F(x, y) = \begin{cases} 6xy + y^3 - 3y^2 - 3x^2 y & ; 0 < y < x, 0 < x < 1 \\ 3x^2 - 2x^3 & ; 0 < x < 1, y \ge x \\ 3y + y^3 - 3y^2 & ; 0 < y < 1, x \ge 1 \\ 1 & ; y \ge 1, x \ge 1 \\ 0, & \textit{otherwise} \end{cases}$$

Find the the joint density (if it exists).

Solution: Instead of checking that the $F(x, y)$ is continuous everywhere on the plane $\mathbb{R}^2$ then computing mixed partials to obtain density, it is usually simpler to first compute the mixed partials first and show that the function f obtained from (30.2) satisfies both the conditions of Theorem 27.5.

We may avoid the boundary points of various regions and compute the mixed partials on in the interior points.

$$\frac{\partial F}{\partial x}(x, y) = \begin{cases} 6y - 6xy & ; 0 < y < x, 0 < x < 1 \\ 6x - 6x^2 & ; 0 < y > x, 0 < x < 1 \\ 0 & ; 0 < y < 1, x > 1 \\ 0 & ; y < 1, x > 1 \\ 0, & \text{otherwise} \end{cases}$$

Further

$$\frac{\partial^2 F}{\partial x \partial y}(x, y) = \begin{cases} 6 - 6x & ; 0 < y < x, 0 < x < 1 \\ 0 & ; 0 < y > x, 0 < x < 1 \\ 0 & ; 0 < y < 1, x > 1 \\ 0 & ; y < 1, x > 1 \\ 0, & \text{otherwise} \end{cases}$$

Therefore, we obtains the following candidate for the joint pdf.

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

Clearly f is non-negative and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_0^1 \left[\int_y^1 6(1 - x) dx \right] dy = \int_0^1 3(y - 1)^2 dy = 1.$$

Hence f is the desired joint pdf. ∎

Example 30.3 *Let (X, Y) be an random vector with joint PDF given by*

$$f(x, y) = \begin{cases} e^{-(x+y)}, & 0 < x < \infty, 0 < y < \infty \\ 0, & \textit{otherwise.} \end{cases}$$

Determine the joint CDF.

Solution: If either $x \leq 0$ or $y \leq 0$, the joint cdf $F \equiv 0$ as the joint pdf is zero in this region.

Let (x, y) be an interior point of the first quadrant. Then

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)dsdt = \int_0^x\int_0^y e^{-(s+t)}dsdt = \left[\int_0^x e^{-s}ds\right]\left[\int_0^y e^{-t}dt\right] = (1-e^{-x})(1-e^{-y})$$

Therefore the joint CDF is

$$F(x,y) = \begin{cases} (1-e^{-x})(1-e^{-y}), & 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise.} \end{cases}$$

■

Lecture 31: Independent Random Variables

19 March, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

We now discuss concepts of independence related to random variables. These are analogous to the concepts of independence between events. They are developed by simply introducing suitable events involving the possible values of various random variables, and by considering the independence of these events.

Definition 31.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and (X, Y) be a random vector defined on it. We say that he random variables X and Y are independent if events $\{X \in A\}$ and $\{Y \in B\}$ are independent for every Borel subsets A and B of $\mathbb{R}$.*

Intuitively, independence means that the value of Y provides no information on the value of X.

Example 31.2 *Consider the experiment of tossing a fair coin and rolling a fair die simultaneously. Intuitively, we feel that whatever the outcome of the coin toss is, it should have no influence on the outcome of the die roll, and vice-versa. Let X be a random variable that is 1 or 0 according as the coin lands heads or tails, i.e., such that the event $\{X = 1\}$ represents the outcome that the coin lands heads and the event $\{X = 0\}$ represents the outcome that the coin lands tails. In a similar way we represent the outcome of the die roll by a random variable Y that takes the value $1, 2, \cdots,$ or 6 according as the die roll results in the face number $1, 2, \cdots,$ or 6. Our intuitive notion that the outcome of the coin toss and die roll have no influence on each other can be stated precisely by saying that if x is one of the numbers 1 or 0 and y is one of the numbers $1, 2, \cdots, 6$, then the events $\{X = x\}$ and $\{Y = y\}$ should be independent.*

How put all this in the frame work of the Definition 31.1. This how is it is done.

$$\Omega = \{(H, i), (T, i) | i = 1, 2, \cdots, 6\}, \mathcal{F} = \mathcal{P}(\Omega), P(\omega) = \frac{1}{12}, \forall \omega \in \Omega.$$

Define random variable $X : \Omega \to \mathbb{R}$ as

$$X(H, i) = 1, X(T, i) = 0, \quad \forall i = 1, 2, \cdots, 6.$$

Define random variable $Y : \Omega \to \mathbb{R}$ as

$$Y(H, i) = i = Y(T, i), \quad \forall i = 1, 2, \cdots, 6.$$

Then

$$\begin{aligned} P(X=1, Y\in\{3,4\}) &= P\left(\{(H,i): i=1,\cdots,6\}\cap\{(H,3),(H,4),(T,3),(T,4)\}\right) \\ &= P\left(\{(H,3),(H,4)\}\right) = \frac{2}{12}. \\ P(X=1) &= P\left(\{(H,i): i=1,\cdots,6\}\right) = \frac{1}{2} \\ P(Y\in\{3,4\}) &= P\left(\{(H,3),(H,4),(T,3),(T,4\}\right) = \frac{1}{3} \end{aligned}$$

We have

$$P(X=1, Y\in\{3,4\}) = P(X=1)P(Y\in\{3,4\}).$$

Similarly we may verify for other events.

One can characterize the independence of X and Y in terms of its joint and marginal distribution functions as in the following theorem.

Theorem 31.3 *Let (X,Y) be a random vector with joint distribution function F, and let F_X and F_Y be the distribution functions of X and X respectively. Then X and Y are independent iff*

$$F(x,y) = F_X(x)F_Y(y), \forall\, (x,y)\in\mathbb{R}^2$$

Remark 31.4 *The Definition 31.1 and the Theorem 31.3 does not assume any special structure on the random variables X or Y. In particular, we may take X as discrete and Y be absolutely continuous or vice-versa or one of them be a general (neither discrete nor absolutely continuous) random variable.*

Theorem 31.3 tells us that if X and Y are independent random variables then marginal distributions of X and Y uniquely determine the joint distribution of X and Y. This suggest us one way to construct the joint CDF.

Example 31.5 *Suppose $X\sim$ Bernoulli$(1/2)$ and $Y\sim$ Discrete Uniform over set $\{1,2,\cdots,6\}$. So recall*

$$F_X(x) = \begin{cases} 0, & x<0 \\ \frac{1}{2}, & 0\le x<1 \\ 1, & x\ge 1. \end{cases} \qquad F_Y(y) = \begin{cases} 0 & \text{if} \quad y<1 \\ \frac{1}{6} & \text{if} \quad 1\le y<2 \\ \frac{2}{6} & \text{if} \quad 2\le y<3 \\ \vdots & \vdots \quad \vdots \\ \frac{5}{6} & \text{if} \quad 5\le y<6 \\ 1 & \text{if} \quad y\ge 6 \end{cases}$$

If we assume independent of X and Y then the joint CDF of X and Y is

$$F(x,y) = \begin{cases} 0, & x < 0 \text{ or } y < 1 \\ \frac{1}{12}, & 0 \leq x < 1, 1 \leq y < 2 \\ \frac{2}{12}, & 0 \leq x < 1, 2 \leq y < 3 \\ \frac{3}{12}, & 0 \leq x < 1, 3 \leq y < 4 \\ \frac{4}{12}, & 0 \leq x < 1, 4 \leq y < 5 \\ \frac{5}{12}, & 0 \leq x < 1, 5 \leq y < 6 \\ \frac{1}{2}, & 0 \leq x < 1, y \geq 6 \\ \frac{1}{6}, & x \geq 1, 1 \leq y < 2 \\ \frac{2}{6}, & x \geq 1, 2 \leq y < 3 \\ \frac{3}{6}, & x \geq 1, 3 \leq y < 4 \\ \frac{4}{6}, & x \geq 1, 4 \leq y < 5 \\ \frac{5}{6}, & x \geq 1, 5 \leq y < 6 \\ 1, & x \geq 1, y \geq 6. \end{cases}$$

Theorem 31.6 *Let X and Y be two discrete random variables. Then X and Y are independent iff joint pmf can be written as product of marginals pmfs, i.e.,*

$$P\{X = x, Y = y\} = P\{X = x\}P(Y = y), \text{ for all } x \in R(X), y \in R(Y).$$

Remark 31.7 *Let us recall that if we are given only marginal distributions of random variables X and Y, in general it is impossible to define the joint distribution of X and Y. But in a very special situation, knowledge about marginal distributions is enough to construct the joint distribution, namely when random variable X and Y are independent. Thanks to the Theorem 31.3.*

Example 31.8 *Let the random vector (X, Y) has joint probabilities given by the following table.*

X \ Y	-1	0	1
-1	0	$\frac{1}{4}$	0
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{4}$	0

Determine if X and Y are independent?

Solution: Note that X and Y are identically distributed and

$$P(X=-1)=P(X=1)=\frac{1}{4} \quad \text{and } P(X=0)=\frac{1}{2}$$

However, X and Y are not independent since

$$P(X=-1,Y=-1)=0\neq\frac{1}{16}=P(X=-1)P(Y=-1)$$

$\blacksquare$

Theorem 31.9 *Let random vector (X,Y) has joint pdf $f(x,y)$, and $f_X(x)$ and $f_Y(y)$ are pdf corresponding to random variables X and Y, respectively. Then X and Y are independent iff joint pdf can be written as product of marginals pdfs, i.e.,*

$$f(x,y)=f_X(x)f_Y(y), \tag{31.1}$$

for each $(x,y)\in\mathbb{R}^2$ where both $f(x,y)$ and $g(x,y):=f_X(x)f_Y(y)$ are continuous.

Example 31.10 *Let X and Y be jointly distributed with the joint density*

$$f(x,y)=\begin{cases}\dfrac{1+xy}{4}, & |x|<1,|y|<1\\ 0, & \textit{otherwise}\end{cases}$$

Determine wether X and Y are independent?

Solution:

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\begin{cases}\displaystyle\int_{-1}^{1}\frac{1+xy}{4}dy=1/2, & |x|<1\\ 0, & \text{otherwise}\end{cases}$$

Since joint pdf is symmetric in x and y, hence

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

But $f(x,y)\neq f_X(x)f_Y(y)$ at many points in the square $(-1,1)\times(-1,1)$, where both, the joint density f and the product f_Xf_Y are continuous. Hence random variables X and Y are not independent. $\blacksquare$

Lecture 32: Independence of Several Random Variables & Real-Valued Functions of Random Vectors

25 March, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Theorem 32.1 *Let X and Y be independent random variables and f and g be Borel-measurable functions. Then $f(X)$ and $g(Y)$ are also independent.*

Example 32.2 *If X and Y are independent then using previous theorem*

1. X^2 *and* Y^2 *are independent.*
2. X *and* $\sin Y$ *are independent*
3. $|X|$ *and* e^Y *are independent.*

32.1 Independence of Several Random Variables

Definition 32.3 *We say $X_1, X_2, \cdots, X_n$ are independent if events $\{X_1 \in A_1\}, \{X_2 \in A_2\}, \cdots, \{X_n \in A_n\}$ are independent for all $A_1, A_2, \cdots, A_n$ Borel subsets of $\mathbb{R}$.*

Theorem 32.4 *A collection of jointly distributed RVs $X_1, X_2, \cdots, X_n$ is said to be mutually or completely independent iff*

$$F(x_1, x_2, \cdots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n) \quad \forall\ (x_1, \cdots, x_n) \in \mathbb{R}^n,$$

where F is the joint CDF of $X_1, X_2, \cdots, X_n$.

Theorem 32.5

1. *Suppose X, Y, Z are discrete random variables. Then they are independent iff*

$$P(X = x, Y = y, Z = z) = P(X = x)P(Y = y)P(Z = z)$$

for all $x \in R(X), y \in R(Y), z \in R(Z)$.

2. *Suppose X, Y, Z are random variables with joint pdf $f(x, y, z)$. Then they are independent iff*

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

for all $(x, y, z) \in \mathbb{R}^3$ where both f and $g(x, y, z) := f_X(x) f_Y(y) f_Z(z)$ are continuous.

When we study law of large numbers and central limit theorem we encounter with the sequence of independent random variables $X_1, X_2, \cdots$. So we need to understand the meaning of independence for countably infinite collection of random variables. In view of Theorem 32.4, we have

Definition 32.6 *We say that a sequence of random variables $(X_n)_{n \in \mathbb{N}}$ is independent if for every $n = 2, 3, \cdots$ the random variables $X_1, X_2, \cdots, X_n$ are independent.*

32.2 Real-Valued Functions of Random Vectors

When there are multiple random variables of interest, it is possible to generate new random variables by considering functions involving several of these random variables . In particular, suppose we have two random variables $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$, and $g : \mathbb{R}^2 \to \mathbb{R}$ be a function. Then $Z := g(X, Y) : \Omega \to \mathbb{R}$ defines another random variable. If both X, Y are discrete (A random variable that can take on at most a countable number of possible values is said to be discrete) then Z is also discrete. Now next natural question is what is the PMF of random variable Z?

PMF of function of two random variables

Now we can employ the same idea to derive PMF of $Z = g(X, Y)$ from the joint PMF $f(x, y)$ according to

$$\begin{aligned} f_Z(z) &= P\{Z = z\} = P\{g(X, Y) = z\} \\ &= P\left\{ \bigcup_{(x,y):g(x,y)=z} \{X = x, Y = y\} \right\} \\ &= \sum_{(x,y):g(x,y)=z} f(x, y). \end{aligned}$$

Example 32.7 *Let X and Y be random variables with the joint pmf given by the following table.*

X \ Y	-1	0	2	6
-2	$\frac{1}{9}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{9}$
1	$\frac{2}{9}$	0	$\frac{1}{9}$	$\frac{1}{9}$
3	0	0	$\frac{1}{9}$	$\frac{4}{27}$

Find the PMF of $|Y - X|$.

Solution: So here $g(x, y) = |y - x|$. First we compute the range of the random variable $Z = g(X, Y) = |Y - X|$.

$$\begin{aligned}
\text{Fix } x = -2, \text{ then } z = |y - (-2)| = |y + 2|. \text{ Now run through the all } y \text{ values, we get } z = 1, 2, 4, 8 \\
\text{Fix } x = 1, \text{ then } z = |y - 1|. \text{ Now run through the all } y \text{ values, we get } z = 2, 1, 1, 5 \\
\text{Fix } x = 3, \text{ then } z = |y - 3|. \text{ Now run through the all } y \text{ values, we get } z = 4, 3, 1, 3
\end{aligned}$$

So range of random variable Z is $\{1, 2, 3, 4, 5, 8\}$. Now

$$P(Z = 1) = \sum_{(x,y):|y-x|=1} f(x, y) = f(-2, -1) + f(1, 0) + f(1, 2) + f(3, 2) = \frac{1}{9} + 0 + \frac{1}{9} + \frac{1}{9} = \frac{1}{3}.$$

$$P(Z = 2) = \sum_{(x,y):|y-x|=2} f(x, y) = f(-2, 0) + f(1, -1) = \frac{1}{27} + \frac{2}{9} = \frac{7}{27}.$$

$$P(Z = 3) = \sum_{(x,y):|y-x|=3} f(x, y) = f(3, 0) + f(3, 6) = 0 + \frac{4}{27} = \frac{4}{27}.$$

$$P(Z = 4) = \sum_{(x,y):|y-x|=4} f(x, y) = f(-2, 2) + f(3, -1) = \frac{1}{27} + 0 = \frac{1}{27}.$$

$$P(Z = 5) = \sum_{(x,y):|y-x|=5} f(x, y) = f(1, 6) = \frac{1}{9}.$$

$$P(Z = 8) = \sum_{(x,y):|y-x|=8} f(x, y) = f(-2, 6) = \frac{1}{9}.$$

■

✠ **Tips For Exam** One check regarding the calculations of pmf in Example 32.7 is, $\sum_{z \in R_Z} f_Z(z) = 1$, where R_Z denotes the range of the random variable Z.

Example 32.8 *Let X and Y be iid (independent & identically distributed) discrete uniform random variables with parameter N. Find the pmf of the random variable* $\min\{X, Y\}$.

Solution: Set $Z := \min\{X, Y\}$. Also it is clear that range of random variable Z would be $\{1, 2, \cdots, N\}$. Now we find its pmf,

$$P\{Z = i\} = \sum_{(x,y):\min\{x,y\}=i} f(x, y)$$

Now both x and y ranges over the set $\{1, 2, \cdots, N\}$. Since X and Y are independent hence the joint pmf is given as follows:

$X \backslash Y$	1	2	$\cdots$	i	$\cdots$	N
1	$\frac{1}{N^2}$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$
2	$\frac{1}{N^2}$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$
$\vdots$	$\frac{1}{N^2}$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$
i	$\frac{1}{N^2}$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$
$\vdots$	$\frac{1}{N^2}$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$
N	$\frac{1}{N^2}$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$	$\cdots$	$\frac{1}{N^2}$

Hence for given $i \in \{1, 2, \cdots, N\}$,

$$\begin{aligned}
\sum_{(x,y):\min\{x,y\}=i} f(x, y) &= \sum_{y=i}^{N} f(i, y) + \sum_{x=i+1}^{N} f(x, i) \\
&= \frac{(N - (i - 1))}{N^2} + \frac{N - i}{N^2} \\
&= \frac{2N - 2i + 1}{N^2}
\end{aligned}$$

■

Lecture 33: Density of Function of two Random Variables

26 March, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

If X and Y have joint pdf and g is a function such that $Z = g(X, Y)$ is absolutely continuous random variable. Then, how to compute the PDF of Z?

Example 33.1 *Let X and Y be independent and exponential random variables with parameters λ and μ, respectively. Find the density of* $\max\{X, Y\}$ *(if it exists).*

Solution: Set $Z := \max\{X, Y\}$. Then for each $z \in \mathbb{R}$, we have

$$\{Z \le z\} = \{X \le z\} \cap \{Y \le z\}.$$

Therefore CDF of Z is

$$F_Z(z) = P(Z \le z) = P(X \le z, Y \le z) = P(X \le z)P(Y \le z) = F_X(z)F_Y(z),$$

where

$$F_X(z) = \begin{cases} 1 - e^{-\lambda z} & \text{if} \quad z \ge 0 \\ 0 & \text{if} \quad z < 0 \end{cases}, \quad F_Y(z) = \begin{cases} 1 - e^{-\mu z} & \text{if} \quad z \ge 0 \\ 0 & \text{if} \quad z < 0 \end{cases}.$$

Therefore

$$F_Z(z) = \begin{cases} [1 - e^{-\lambda z}][1 - e^{-\mu z}] = 1 - e^{-\mu z} - e^{-\lambda z} + e^{-(\lambda+\mu)z} & \text{if} \quad z \ge 0 \\ 0 & \text{if} \quad z < 0 \end{cases}.$$

At $z = 0$, $F_Z(0) = 0$ hence CDF is continuous everywhere. We may differentiate it to get the density

$$F_Z'(z) = \begin{cases} \mu e^{-\mu z} + \lambda e^{-\lambda z} - (\lambda + \mu)e^{-(\lambda+\mu)z} & \text{if} \quad z > 0 \\ 0 & \text{if} \quad z < 0 \end{cases}.$$

In fact $F_Z'(0) = 0$. So the random variable Z has the density

$$f_Z(z) = \begin{cases} \mu e^{-\mu z} + \lambda e^{-\lambda z} - (\lambda + \mu)e^{-(\lambda+\mu)z} & \text{if} \quad z \ge 0 \\ 0 & \text{if} \quad z < 0 \end{cases}.$$

■

Lecture 34: Expectation of function of random vector

27 March, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Example 34.1 *The joint density function of X and Y is given by*

$$f(x,y) = \begin{cases} e^{-(x+y)} & \text{if} \quad 0 < x, y < \infty \\ 0 & \textit{otherwise} \end{cases}.$$

Find the density function of the random variable $\dfrac{X}{Y}$.

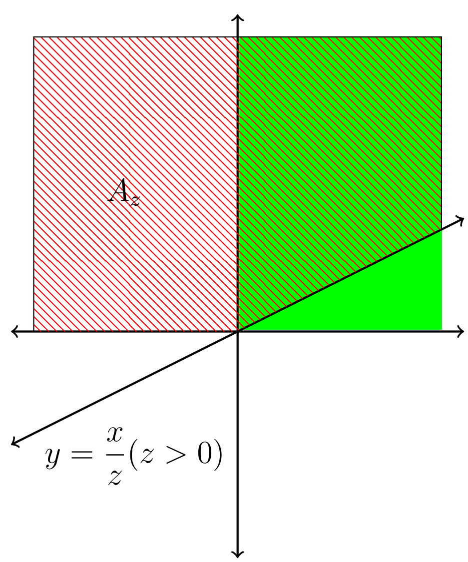
Solution: Let $Z := \dfrac{X}{Y}$. Let $z \in \mathbb{R}$ be given. Then

$$\{Z \leq z\} = \left\{\frac{X}{Y} \leq z\right\} = \{(X,Y) \in A_z\},$$

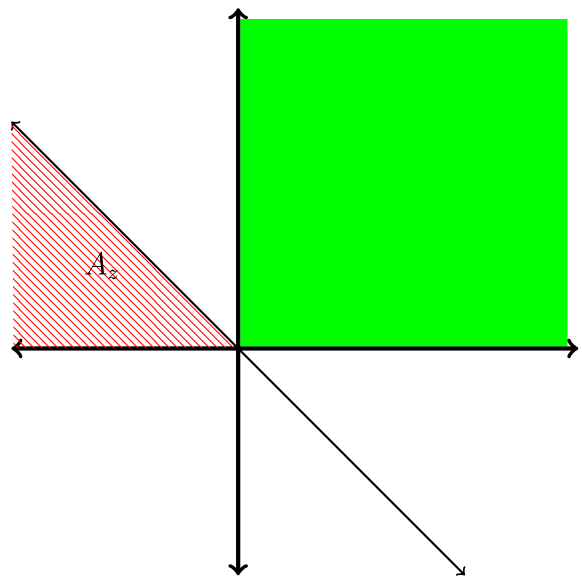
where $A_z = \{(x,y) \in \mathbb{R}^2 : \frac{x}{y} \leq z\}$. If $y > 0$ then $A_z = \{(x,y) \in \mathbb{R}^2 : x \leq yz\}$ and if $y < 0$ then $A_z = \{(x,y) \in \mathbb{R}^2 : x \geq yz\}$. Now we plot the straight line $x = yz$, which we further divide int two cases:

1. **When** $y > 0$**:**

   (a) **When** $z \geq 0$**:** Then the straight line $x = yz$ can be rewritten as $y = \frac{x}{z}$, this the line with positive slope. Hence we are interested in the region $y \geq \frac{x}{z}$. Then the desired region A_z is

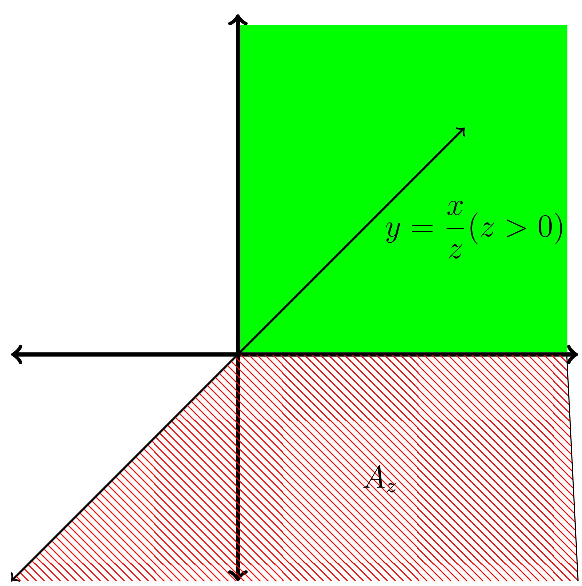


(b) **When** $z < 0$**:** Then the straight line $x = yz$ can be rewritten as $y = \frac{x}{z}$, this the line with negative slope. Hence we are interested in the region $y \leq \frac{x}{z}$. Then the desired region A_z is Then the desired region A_z is

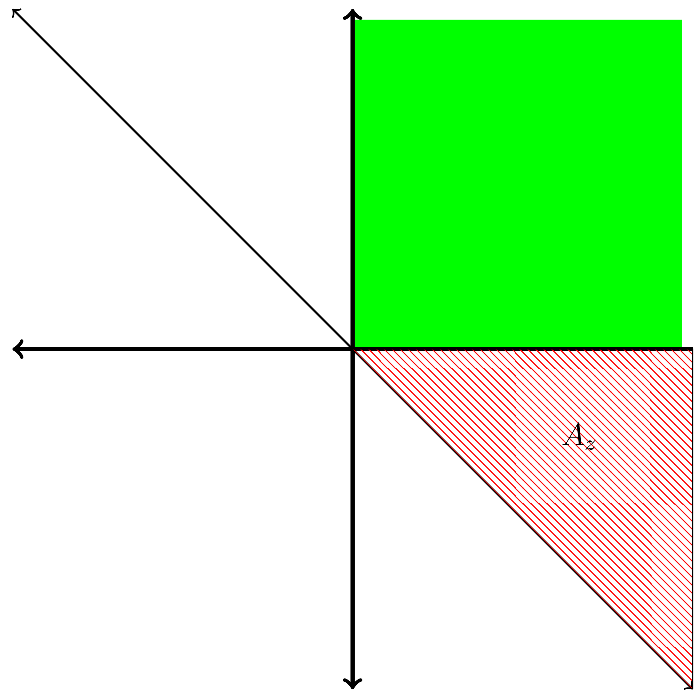


2. **When** $y < 0$**:** These case is futile to analyse as the joint pdf is non-zero only in first quadrant hence we get no contribution from here. In this case $A_z = \{(x, y) \in \mathbb{R}^2 : x \geq yz\}$.

   (a) **When** $z \geq 0$**:** Then the straight line $x = yz$ can be rewritten as $y = \frac{x}{z}$, this the line with positive slope. Hence we are interested in the region $y \leq \frac{x}{z}$. Then the desired region A_z is



(b) **When** $z < 0$**:** Then the straight line $x = yz$ can be rewritten as $y = \frac{x}{z}$, this the line with negative slope. Hence we are interested in the region $y \geq \frac{x}{z}$. Then the desired region A_z is Then the desired region A_z is



Therefore if $z \geq 0$,

$$\begin{aligned}F_Z(z) &:= P\{(X,Y) \in A_z\} = \int_0^\infty e^{-y}\left(\int_0^{zy} e^{-x}dx\right)dy \\ &= \int_0^\infty e^{-y}\left(\int_0^{zy} e^{-x}dx\right)dy = \int_0^\infty e^{-y}\left[-e^{-x}\right]_0^{zy} dy = \int_0^\infty e^{-y}\left[1 - e^{-zy}\right]dy \\ &= \left[-e^{-y} + \frac{1}{z+1}e^{-(z+1)y}\right]_0^\infty = -0 + 0 + 1 - \frac{1}{z+1} = 1 - \frac{1}{z+1}\end{aligned}$$

Therefore if $z < 0$,

$$F_Z(z) := P\{(X,Y) \in A_z\} = 0$$

Hence

$$F_Z(z) = \begin{cases} 1 - \dfrac{1}{z+1} & \text{if} \quad z \geq 0 \\ 0 & \text{if} \quad z < 0 \end{cases}.$$

Since F_Z is a continuous function, we may differentiate it to get the density

$$F'_Z(z) = \begin{cases} \dfrac{1}{(z+1)^2} & \text{if} \quad z > 0 \\ 0 & \text{if} \quad z < 0 \end{cases}.$$

F_Z is not differentiable at $z = 0$, so we set density to be equal to zero at this point. Hence pdf of Z is

$$f_Z(z) = \begin{cases} \dfrac{1}{(z+1)^2} & \text{if} \quad z > 0 \\ 0 & \text{if} \quad z \leq 0 \end{cases}.$$

■

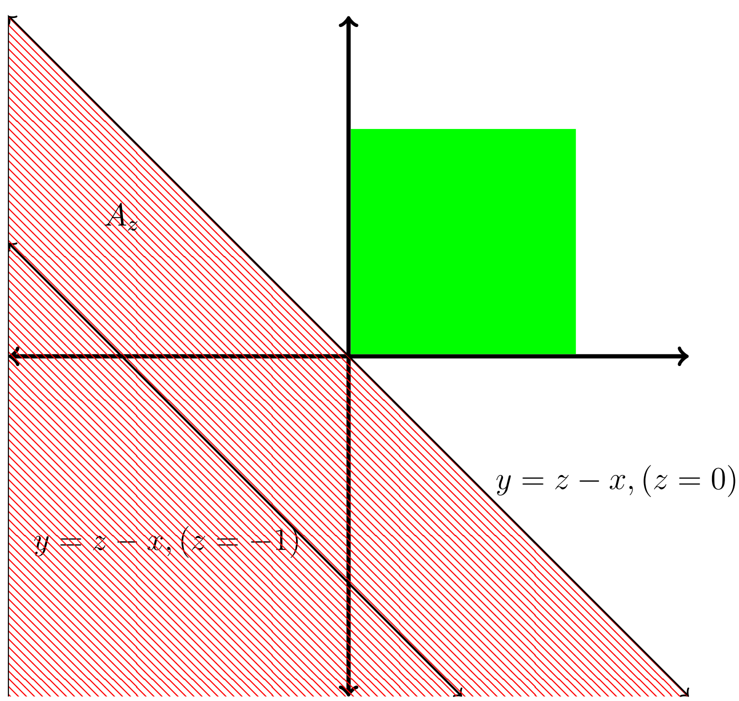
Example 34.2 *Let X and Y be iid uniform random variables distributed over interval $(0, 1)$. Find the density of $X + Y$ (if it exists).*

Solution: Define $Z := X + Y$. For fixed $z \in \mathbb{R}$ the event $\{Z \leq z\}$ is equivalent to the event $\{(X, Y) \in A_z\}$, where A_z is the subset of $\mathbb{R}^2$ defined by $A_z = \{(x, y) \in \mathbb{R}^2 | x + y \leq z\}$. Thus

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P((X, Y) \in A_z) \\ &= \iint\limits_{A_z} f(x, y)dxdy \end{aligned}$$

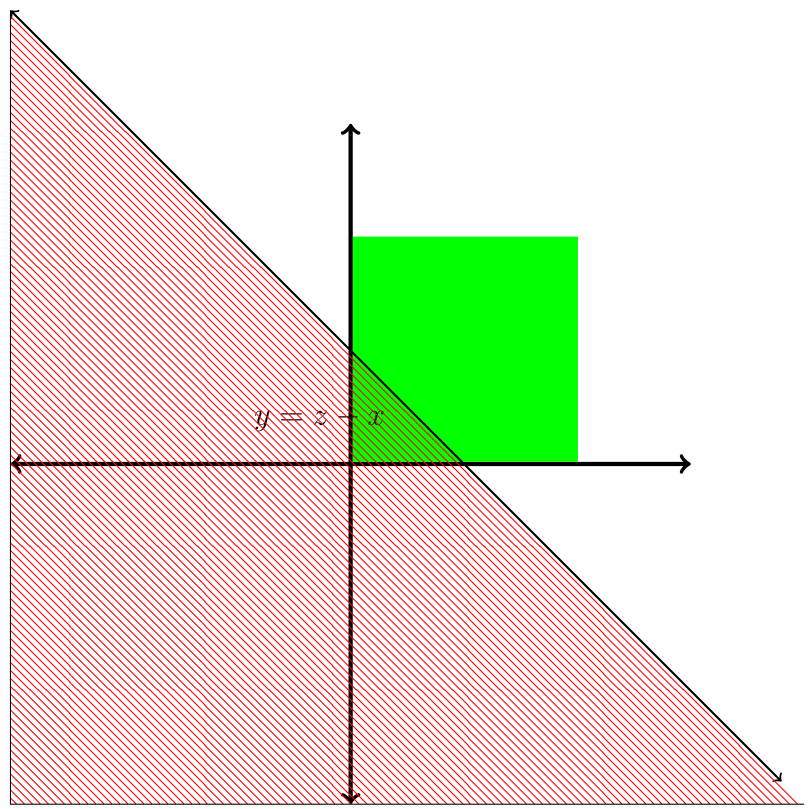
Since our joint density is non-zero only on unit square therefore we analyse the set A_z for various values of z.

1. **If $-\infty < z \leq 0$:**



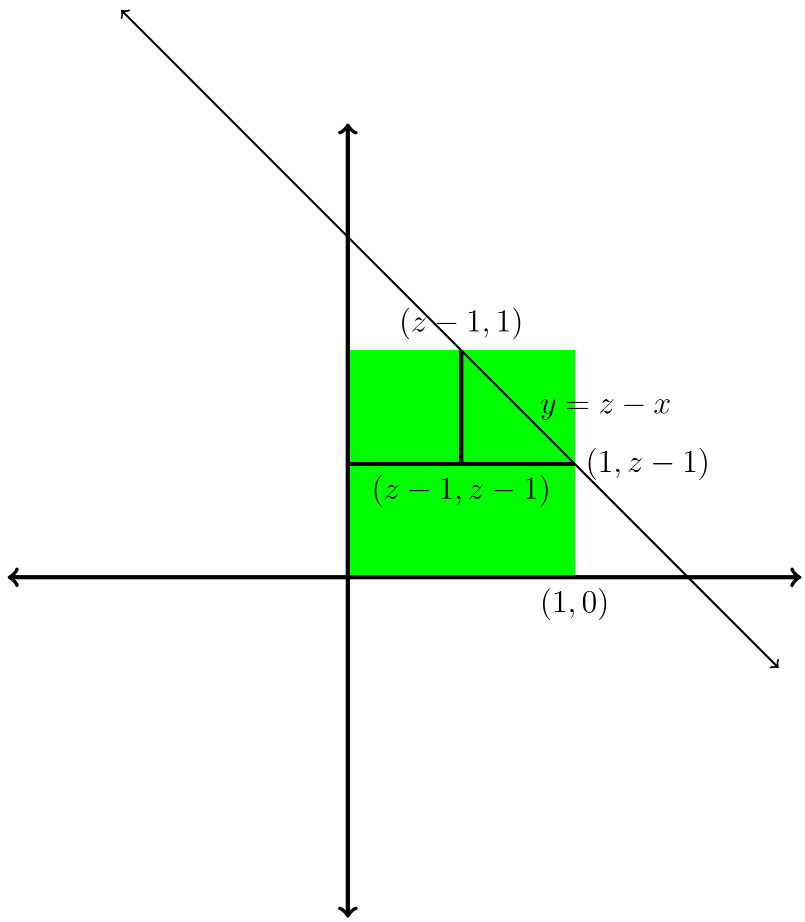
$$\iint_{A_z} f(x,y)dxdy = 0$$

2. **If** $0 < z \leq 1$**:**



$$\iint_{A_z} f(x,y)dxdy = \text{ Area of the triangle with vertices}(0,0),(z,0),(0,z) = \frac{z^2}{2}$$

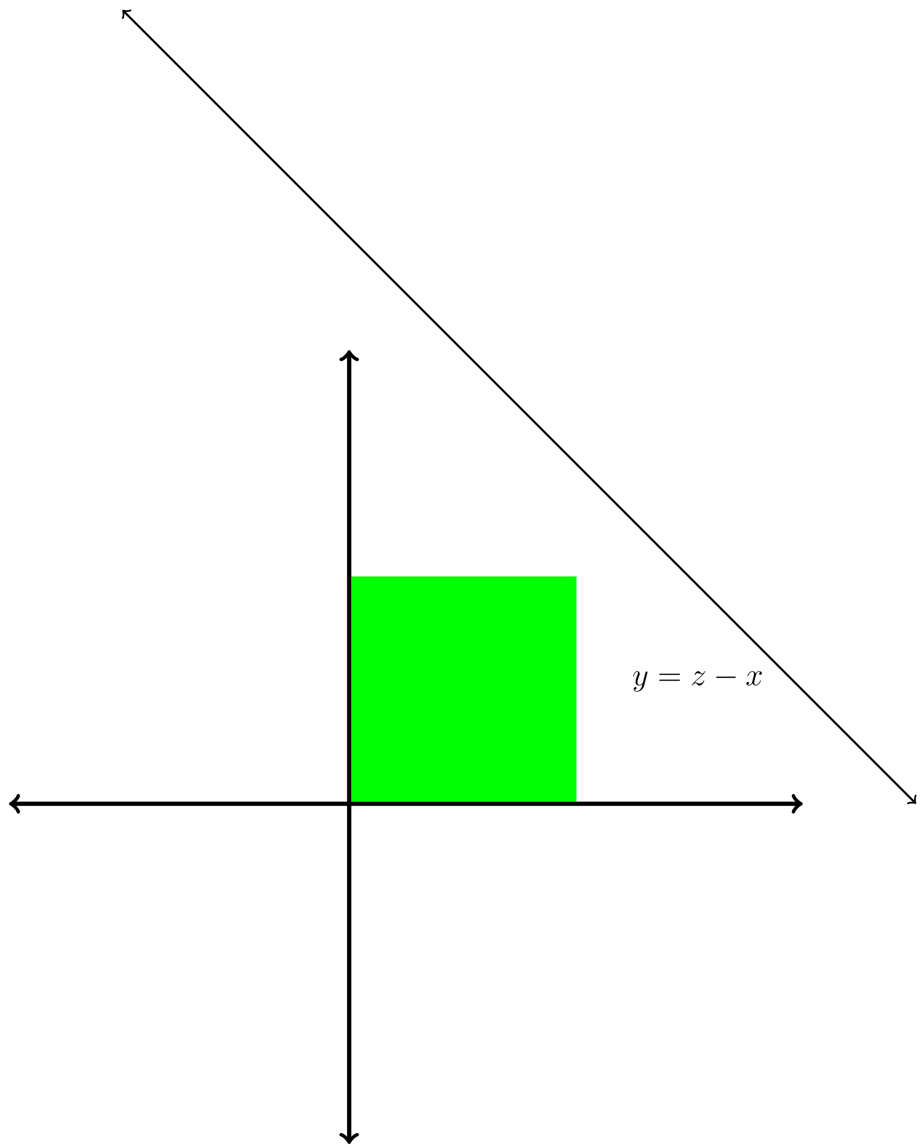
3. **If** $1 < z \leq 2$**:**



Thanks to Anshu Musaddi for the following short cut:

$$\begin{aligned}\iint_{A_z} f(x,y)dxdy = &\text{ Area of the unit square}\\ &- \text{ Area of the triangle with vertices } (1,z-1),(1,1),(z-1,1)\\ = 1-\frac{1}{2}\times(2-z)\times(2-z) = 1-\frac{(2-z)^2}{2}\end{aligned}$$

4. **If $z \geq 2$:**



$$\iint_{A_z} f(x,y)dxdy = 1$$

Hence

$$F_Z(z) = \begin{cases} 0 & \text{if} \quad z \leq 0 \\ \dfrac{z^2}{2} & \text{if} \quad 0 < z \leq 1 \\ 1 - \dfrac{(2-z)^2}{2} & \text{if} \quad 1 < z \leq 2 \\ 1 & \text{if} \quad z > 2 \end{cases}.$$

It is continuous everywhere. Hence we may differentiate it to get the density

$$F'_Z(z) = \begin{cases} 0 & \text{if} \quad z < 0 \\ z & \text{if} \quad 0 < z < 1 \\ 2 - z & \text{if} \quad 1 < z < 2 \\ 0 & \text{if} \quad z < 2 \end{cases}.$$

This tells is that F_Z is actually differentiable at $z = 0, 1, 2$ also. Hence we get the following pdf of the random variable Z.

$$f_Z(z) = \begin{cases} 0 & \text{if} \quad z \leq 0 \\ z & \text{if} \quad 0 \leq z < 1 \\ 2 - z & \text{if} \quad 1 \leq z < 2 \\ 0 & \text{if} \quad z \geq 2 \end{cases}.$$

■

34.1 Expectation of function of two random variables

Theorem 34.3 *1. Let X, Y be two discrete random variable with joint pmf $f(x, y)$. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a function, then*

$$E[g(X,Y)] = \sum_{x \in R_X} \sum_{y \in R_Y} g(x,y) f(x,y), \tag{34.1}$$

provided $\displaystyle\sum_{x \in R_X} \sum_{y \in R_Y} |g(x,y)| f(x,y) < \infty.$

2. Let X, Y be two absolutely continuous random variable with joint pdf $f(x, y)$. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a Borel measurable function (e.g., continuous, indicators of reasonable sets (Borel subsets of $\mathbb{R}^2$), and functions that are continuous except across some smooth boundaries) , then

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dx dy \tag{34.2}$$

provided $\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x,y)| f(x,y) dx dy < \infty.$

Example 34.4 *Recall Example 32.7, X and Y were random variables with the joint pmf given by the following table.*

X \ Y	-1	0	2	6
-2	$\frac{1}{9}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{9}$
1	$\frac{2}{9}$	0	$\frac{1}{9}$	$\frac{1}{9}$
3	0	0	$\frac{1}{9}$	$\frac{4}{27}$

Then PMF of $Z := |Y - X|$*, is*

$$P(Z=1)=\frac{1}{3},\quad P(Z=2)=\frac{7}{27},\quad P(Z=3)=\frac{4}{27},\quad P(Z=4)=\frac{1}{27},\quad P(Z=5)=\frac{1}{9},\quad P(Z=8)=\frac{1}{9}$$

So

$$E[Z]=\sum_{z\in R_Z} zP(Z=z)=1\times\frac{1}{3}+2\times\frac{7}{27}+3\times\frac{4}{27}+4\times\frac{1}{27}+5\times\frac{1}{9}+8\times\frac{1}{9}=\frac{26}{9}$$

Suppose we are interested only in $E|Y-X|$ *then by formula (34.1), we have*

$$\begin{aligned}
E|Y-X| &= \sum_y \sum_x |y-x| f(x,y) \\
&= \sum_y |y-(-2)| f(-2,y) + \sum_y |y-1| f(1,y) + \sum_y |y-3| f(3,y)
\end{aligned}$$

$$\begin{aligned}
\sum_y |y-(-2)| f(-2,y) &= 1\times\frac{1}{9}+2\times\frac{1}{27}+4\times\frac{1}{27}+8\times\frac{1}{9}=\frac{11}{9} \\
\sum_y |y-1| f(1,y) &= 2\times\frac{2}{9}+1\times 0+1\times\frac{1}{9}+5\times\frac{1}{9}=\frac{10}{9} \\
\sum_y |y-3| f(3,y) &= 4\times 0+3\times 0+1\times\frac{1}{9}+3\times\frac{4}{27}=\frac{5}{9}
\end{aligned}$$

Lecture 35: Expectation of Function of Random Vector & Conditional Distribution

29 March, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Example 35.1 *Let X and Y be independent and exponential random variables with parameters λ and μ, respectively. Find the mean of* $\max\{X, Y\}$ *(if it exists).*

Solution: Since joint pdf is non-zero in first quadrant only and function $\max\{x, y\}$ is also non-negative in first quadrant hence absolute convergence of improper integral is equivalent to the conditional convergence.

$$
\begin{aligned}
E[\max\{X,Y\}] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \max\{x,y\} f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \max\{x,y\} f_X(x) f_Y(y)dxdy \quad (\because X, Y \text{ are independent}) \\
&= \iint\limits_{\{x\geq 0, y\geq 0\}} \max\{x,y\}\lambda\ \mu\ e^{-\lambda x}e^{-\lambda y}dxdy \\
&= \iint\limits_{\{x\geq 0, y\geq 0\}\cap\{y\geq x\}} \max\{x,y\}\lambda\ \mu\ e^{-\lambda x}e^{-\lambda y}dxdy \\
&\quad + \iint\limits_{\{x\geq 0, y\geq 0\}\cap\{y< x\}} \max\{x,y\}\lambda\ \mu\ e^{-\lambda x}e^{-\mu y}dxdy \\
&= \lambda\ \mu\left[\int_0^{\infty} e^{-\lambda x}\left(\int_x^{\infty} ye^{-\mu y}dy\right)dx + \int_0^{\infty} xe^{-\lambda x}\left(\int_0^{x} e^{-\mu y}dy\right)dx\right] \\
&= \lambda\ \mu\left[\int_0^{\infty} e^{-\lambda x}\left(\frac{xe^{-\mu x}}{\mu} + \frac{e^{-\mu x}}{\mu^2}\right)dx + \int_0^{\infty} xe^{-\lambda x}\left(\frac{1-e^{-\mu x}}{\mu}\right)dx\right] \\
&= \lambda\ \mu\left[\int_0^{\infty} e^{-\lambda x}\frac{e^{-\mu x}}{\mu^2}dx + \int_0^{\infty} xe^{-\lambda x}\frac{1}{\mu}dx\right] \\
&= \lambda\ \mu\left[\frac{1}{\mu^2}\left[\frac{-e^{-(\lambda+\mu)x}}{\lambda+\mu}\right]_0^{\infty} + \frac{1}{\mu}\frac{1}{\lambda^2}\right] = \frac{\lambda}{\mu(\lambda+\mu)} + \frac{1}{\lambda} = \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda+\mu}
\end{aligned}
$$

Now

$$\int_x^\infty ye^{-\mu y}dy = \left[\frac{ye^{-\mu y}}{-\mu}\right]_{y=x}^{y=\infty} + \int_x^\infty \frac{e^{-\mu y}}{\mu}dy = \frac{xe^{-\mu x}}{\mu} - \frac{1}{\mu^2}\left[e^{-\mu y}\right]_x^\infty = \frac{xe^{-\mu x}}{\mu} + \frac{e^{-\mu x}}{\mu^2}$$

$$\int_0^x e^{-\mu y}dy = \left[\frac{e^{-\mu y}}{-\mu}\right]_0^x = \frac{1-e^{-\mu x}}{\mu}$$

$$\int_0^\infty xe^{-\lambda x} = -\frac{1}{\lambda}[xe^{-\lambda x}]_0^\infty + \frac{1}{\lambda}\int_0^\infty e^{-\lambda x}dx = -\frac{1}{\lambda^2}[e^{-\lambda x}]_0^\infty = \frac{1}{\lambda^2}$$

■

Theorem 35.2 *Let X and Y be two random variable on a probability space $(\Omega, \mathcal{F}, P)$ such that both have finite mean. Then*

(a) $E[X+Y] = EX + EY$.

(b) *If X and Y are independent, then*

$$E[XY] = EX\ EY.$$

Proof: We shall prove it only in the case when (X, Y) have joint density, though both the results are true even if X is discrete and Y has pdf.

(a)

$$\begin{aligned}
E[X+Y] &= \int_{-\infty}^\infty\int_{-\infty}^\infty (x+y)f(x,y)dxdy \\
&= \int_{-\infty}^\infty\int_{-\infty}^\infty xf(x,y)dxdy + \int_{-\infty}^\infty\int_{-\infty}^\infty yf(x,y)dxdy \\
&= \int_{-\infty}^\infty x\left(\int_{-\infty}^\infty f(x,y)dy\right)dx + \int_{-\infty}^\infty y\left(\int_{-\infty}^\infty f(x,y)dx\right)dy \\
&= \int_{-\infty}^\infty xf_X(x)dx + \int_{-\infty}^\infty yf_Y(y)dy \\
&= EX + EY
\end{aligned}$$

(b)

$$\begin{aligned}
E[XY] &= \int_{-\infty}^\infty\int_{-\infty}^\infty x\,yf(x,y)dxdy \\
&= \int_{-\infty}^\infty\int_{-\infty}^\infty x\,yf_X(x)f_Y(y)dxdy \\
&= \left(\int_{-\infty}^\infty xf_X(x)dx\right)\left(\int_{-\infty}^\infty yf_Y(y)dy\right) \\
&= EXEY
\end{aligned}$$

Quiz – II syllabus over.

■

35.1 Conditional Distributions

Conditional probability $P(A|B)$, is probability of event A in the new universe (or sample space) B. Now we extend this idea to conditioning one random variable on another in order to give a quantification of dependence of one random variable over the other if the random variables are not independent. We first look at discrete random variables.

Conditional PMF

Definition 35.3 *Let X and Y be two discrete random variable associated with the same random experiment. Then the conditional pmf $f_{X|Y}$ of X given $Y = y$, is defined as*

$$f_{X|Y}(x|y) = \begin{cases} P\{X = x|Y = y\} & \text{if} \quad P\{Y = y\} > 0 \\ 0 & \text{if} \quad P\{Y = y\} = 0 \end{cases}$$

In the original sample space Ω, random variable X has some probability distribution. Now we are told that event $\{Y = y\}$ has occurred. Since X depend on Y, this new information provides partial knowledge about value of X. Hence the probability distribution of X in the new universe determined by the event $\{Y = y\}$ should change. This change is captured by conditional pmf.

A conditional pmf can be thought of as an ordinary pmf over a new universe determined by the conditioning event. For this, note that for fixed y, $f_{X|Y}(x|y) \geq 0$ for all $x \in R_X$. Also if $P(Y = y) > 0$ then

$$\sum_{x \in R_X} f_{X|Y}(x|y) = \sum_{x \in R_X} P(X = x|Y = y) = P\left(\bigcup_{x \in R_X} \{X = x\}\Big|Y = y\right) = P(\Omega|Y = y) = 1$$

If X, Y have joint pmf f, then using the definition of conditional probability we obtain

$f_{X|Y} = \dfrac{f(x,y)}{f_Y(y)}$

$$f_{X|Y}(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)} & \text{if} \quad f_Y(y) > 0 \\ 0 & \text{if} \quad f_Y(y) = 0 \end{cases}$$

pmf

Conditional Distribution Function

Recall that the distribution function F_X of any random variable X (discrete, continuous or mixed) is defined as

$$\boxed{F_X(x) = P\{X \leq x\}, \quad \forall x \in \mathbb{R}.}$$

We define conditional distribution function of X given $Y = y$ as

$$F_{X|Y}(x|y) := P(X \le x | Y = y).$$

So conditional distribution function is an ordinary (or unconditional) distribution function in new universe determined by the conditioning event.

Recall that if X is a discrete random variable with pmf f_X then

$$F_X(x) = \sum_{t \in R_X : t \le x} f_X(t).$$

Similarly, if X is a discrete random variable with conditional pmf $f_{X|Y}$ then

$$F_{X|Y}(x|y) = \sum_{t \in R_X : t \le x} f_{X|Y}(t|y).$$

Recall that if X is a discrete random variable with pmf f_X then and $A \subset \mathbb{R}$ then

$$P(X \in A) = \sum_{x \in A \cap R_X} f_X(x).$$

Similarly, if X is a discrete random variable with conditional pmf $f_{X|Y}$ and $A \subset \mathbb{R}$, then we have

$$P(X \in A | Y = y) = \sum_{x \in R_X \cap A} f_{X|Y}(x|y)$$

Example 35.4 *Let the joint pmf of X and Y is given as follows:*

X \ Y	-1	0	1
-1	0	$\frac{1}{4}$	0
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{4}$	0

Then compute the conditional pmf of X given $Y = 0$. Also compute the conditional distribution function of the same.

Solution: Note that $P(Y = 0) = \dfrac{1}{2}$. Hence

$$\boxed{f_{X|Y}(x|0) = \begin{cases} \frac{1}{2} & \text{if} \quad x = -1, 1 \\ 0 & \text{if} \quad x = 0 \end{cases}}$$

Now the conditional distribution function

$$F_{X|Y}(x|0) = \begin{cases} 0 & \text{if} \quad x < -1 \\ \frac{1}{2} & \text{if} \quad -1 \le x < 1 \\ 1 & \text{if} \quad x \ge 1 \end{cases}$$

■

Remark 35.5 *We have said that conditional pmf is a pmf in the new universe determined by the conditioning event. In Example 35.4, the probability distribution of X is*

$$P(X = -1) = P(X = 1) = \frac{1}{4}, P(X = 0) = \frac{1}{2}.$$

Where as in new universe determined by the event $\{Y = 0\}$, the probability distribution of X is revised as

$$P(X = -1|Y = 0) = P(X = 1|Y = 0) = \frac{1}{2}.$$

Similarly, the distribution function of X is

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ \dfrac{1}{4} & \text{if} \quad -1 \le x < 0 \\ \\ \dfrac{3}{4} & \text{if} \quad 0 \le x < 1 \\ \\ 1 & \text{if} \quad x \ge 1 \end{cases}$$

F_X got revised as $F_{X|Y}(x|0)$ in the new universe determined by the event $\{Y = 0\}$. Also note that $F_{X|Y}(x|0)$ satisfies all the properties of a distribution function:

1. $\lim_{x \to -\infty} F_{X|Y}(x|0) = 0, \lim_{x \to +\infty} F_{X|Y}(x|0) = 1.$

2. *$F_{X|Y}(\cdot|0)$ is non-decreasing on $\mathbb{R}$.*

3. *$F_{X|Y}(\cdot|0)$ is right-continuous on $\mathbb{R}$.*

The conditional PMF can also be used to calculate the marginal PMFs. In particular, we have by using the definitions,

$$\boxed{f_X(x) = \sum_y f(x, y) = \sum_y f_{X|Y}(x|y) f_Y(y)}$$

Example 35.6 *Suppose*

$$f_Y(y) = \begin{cases} \frac{5}{6} & \text{if} \quad y = 10^2 \\ \frac{1}{6} & \text{if} \quad y = 10^4 \end{cases}, \; f_{X|Y}(x|10^2) = \begin{cases} \frac{1}{2} & \text{if} \quad x = 10^{-2} \\ \frac{1}{3} & \text{if} \quad x = 10^{-1} \\ \frac{1}{6} & \text{if} \quad x = 1 \end{cases}, \; f_{X|Y}(x|10^4) = \begin{cases} \frac{1}{2} & \text{if} \quad x = 1 \\ \frac{1}{3} & \text{if} \quad x = 10 \\ \frac{1}{6} & \text{if} \quad x = 100 \end{cases},$$

Then find the pmf of X.

Solution: First of all by looking at conditional pmf $f_{X|Y}$ we see that X takes 5 values $10^{-2}, 10^{-1}, 1, 10, 100$. Now

$$\begin{aligned} f_X(10^{-2}) &= \frac{1}{2} \times \frac{5}{6} = \frac{5}{12} \\ f_X(10^{-1}) &= \frac{1}{3} \times \frac{5}{6} = \frac{5}{18} \\ f_X(1) &= \frac{1}{6} \times \frac{5}{6} + \frac{1}{2} \times \frac{1}{6} = \frac{8}{36} \\ f_X(10) &= \frac{1}{3} \times \frac{1}{6} = \frac{1}{18} \\ f_X(100) &= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \end{aligned}$$

■

Lecture 36: Conditional PDF & Law of Total Probability

1 April, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Definition 36.1 (Conditional Densities) *Let X and Y be two random variables with joint pdf f. The conditional density of X given $Y = y$ is defined as*

$$f_{X|Y}(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)} & \text{if} \quad f_Y(y) > 0 \\ 0 & \text{if} \quad f_Y(y) = 0 \end{cases}$$

As the case of conditional pmf, conditional pdf can be thought of as an ordinary pdf over a new universe determined by the conditioning event. For this, note that for fixed y, $f_{X|Y}(x|y) \geq 0$ for all $x \in \mathbb{R}$. Also if $f_Y(y) > 0$ then

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x,y)dx = \frac{f_Y(y)}{f_Y(y)} = 1$$

Suppose that in sampling from a human population X denotes person's weight and Y denotes persons hight. Surely we would think it more likely that $X > 200$ pounds if we were told that $Y = 6$ feet than if we were told that $Y = 4$ feet. Knowledge about the value of Y gives us some information about the value of X even if it does not tell us the value of X exactly. If we model (X, Y) as continuous random variables then $P(Y = 6) = 0$, yet in actuality a value of Y is observed. If, to the limit of our measurement we see $Y = 6$, this knowledge might give us information about X. The conditional pdf allows to define appropriately the conditional distribution of X given $Y = y$.

Recall that if X is a continuous random variable with pdf f_X and B is any Borel subset of $\mathbb{R}$, then

$$P(X \in B) = \int_B f_X(x)dx.$$

The above motivated the following definition.

Definition 36.2 *Let X,Y be jointly continuous random variables and $f_{X|Y}(\cdot)$ denotes the conditional density of X given Y. Then for any Borel subset B of $\mathbb{R}$, we have*

$$P(X \in B|Y = y) = \int_B f_{X|Y}(x|y)dx \tag{36.1}$$

Remark 36.3 *Conditional probability $P(X \in B|Y = y)$ were left undefined if the $P\{Y = y\} = 0$. But the above formula provides a natural way of defining such conditional probabilities in the present context . In addition, it allows us to view the conditional PDF $f_{X|Y}$ (as a function of x) as a description of the probability law of X, given that the event $\{Y = y\}$ has occurred.*

In view of equation (36.1) the conditional CDF of X given $Y = y$ is

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(t|y)dt.$$

We have

$$\frac{d}{dx}F_{X|Y}(x|y) = f_{X|Y}(x|y),$$

where equality holds at points (x, y) at which joint pdf f is continuous and $f_Y(y) > 0$ and $f_Y(\cdot)$ is continuous at y.

Example 36.4 *Let X and Y be two random variables having the joint probability density function*

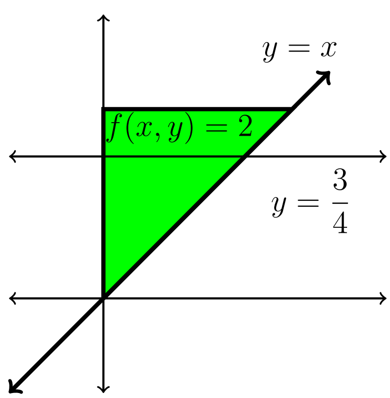
$$f(x,y) = \begin{cases} 2 & \text{if} \quad 0 < x < y < 1 \\ 0 & \quad \textit{elsewhere} \end{cases}$$

Then find the conditional probability $P\left(X \le \frac{2}{3}\middle|Y = \frac{3}{4}\right)$.

Solution: We are suppose to use the following definition

$$P(X \in B|Y = y) = \int_B f_{X|Y}(x|y)dx,$$

i.e., we need to compute the conditional density $f_{X|Y}\left(x\middle|\frac{3}{4}\right)$ and for this we need to compute $f_Y\left(\frac{3}{4}\right)$.



$$f_Y\left(\frac{3}{4}\right) = \int_{-\infty}^{\infty} f\left(x, \frac{3}{4}\right) dx = \int_0^{\frac{3}{4}} f\left(x, \frac{3}{4}\right) dx = \int_0^{\frac{3}{4}} 2dx = 2 \times \frac{3}{4} = \frac{3}{2}.$$

Since $f_Y\left(\frac{3}{4}\right) > 0$, therefore

$$f_{X|Y}\left(x \Big| \frac{3}{4}\right) = \begin{cases} \dfrac{f\left(x, \frac{3}{4}\right)}{f_Y\left(\frac{3}{4}\right)} = \dfrac{2}{\frac{3}{2}} = \dfrac{4}{3} & \text{if} \quad 0 < x < \dfrac{3}{4} \\ 0 & \text{elsewhere} \end{cases}$$

Hence

$$\begin{aligned} P\left(X \le \frac{2}{3} \Big| Y = \frac{3}{4}\right) &= \int_{-\infty}^{\frac{2}{3}} f_{X|Y}\left(x \Big| \frac{3}{4}\right) dx \\ &= \int_0^{\frac{2}{3}} \frac{4}{3}\, dx \\ &= \frac{8}{9} \end{aligned}$$

■

Example 36.5 *Let X and Y be independent continuous random variables with pdf f_X and f_Y respectively. Let $Z = X + Y$. Determine conditional density of Z given X.*

Solution: Basically we first determine the conditional distribution function of Z given X, i.e., $P(Z \le z|X = x)$. Then we have the relation

$$P(Z \le z|X = x) = \int_{-\infty}^{z} f_{Z|X}(t|x)dt$$

Now

$$\begin{aligned} P(Z \le z|X = x) &= P(X + Y \le z|X = x) \\ &= P(x + Y \le z|X = x) \\ &= P(x + Y \le z) \quad (\because X, Y \text{ are indepedent}) \\ &= P(Y \le z - x) \\ &= \int_{-\infty}^{z-x} f_Y(y)dy \\ &= \int_{-\infty}^{z} f_Y(t - x)dt \qquad (\text{put } y = t - x) \end{aligned}$$

Hence $f_{Z|X}(z|x) = f_Y(z - x)$. ■

Remark 36.6 *In Example 36.5, if we try to compute conditional density of $X+Y$ given X by definition then we require to compute the joint density of $X+Y$ and X. This type of problem we have not studied.*

Rather than going by definition, we adopt the technique of finding pdf of a real-valued function of two random variables. We first compute the conditional distribution function and differentiate it to obtain the conditional density.

36.1 Law of Total Probability

Proposition 36.7 (Law of total probability) *Let Y be a discrete random variable on the probability space $(\Omega, \mathcal{F}, P)$. Then for any event $B \in \mathcal{F}$,*

$$\boxed{P(B) = \sum_{y \in R_Y} P(B|Y=y) f_Y(y),} \tag{36.2}$$

where f_Y is the pmf of Y.

Proof: If Y is a discrete random variable with range $R_Y \subset \mathbb{R}$, then the collection of events $\{\{Y=y\}\}_{y \in R_Y}$ form a partition of the sample space Ω. Thus, we can use the total probability theorem.

$$P(B) = \sum_{y \in R_Y} P(B|Y=y)P(Y=y) = \sum_{y \in R_Y} P(B|Y=y) f_Y(y).$$

■

We state the law of total probability for continuous random variable, which is completely analogous to the discrete case.

Theorem 36.8 *Let X be a random variable (on the probability space $(\Omega, \mathcal{F}, P)$) with the pdf f_X. Then for any event $B \in \mathcal{F}$,*

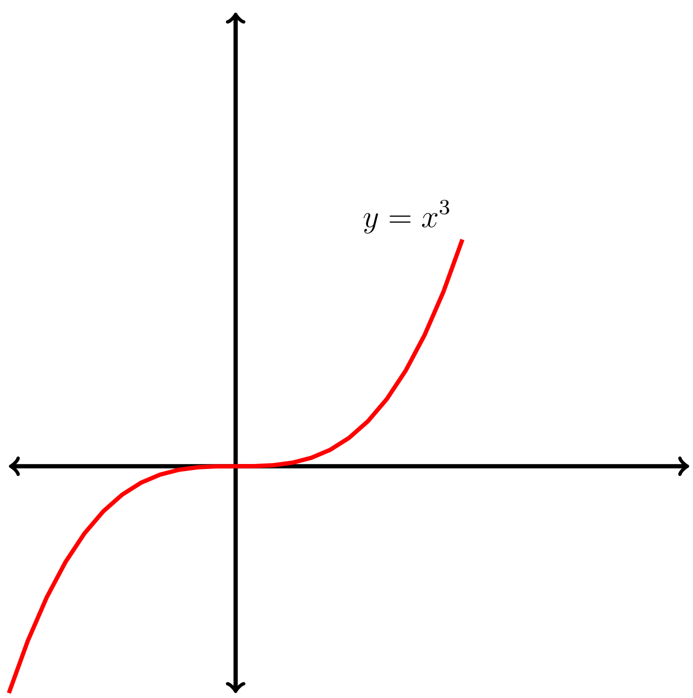
$$\boxed{P(B) = \int_{-\infty}^{\infty} P(B|X=x) f_X(x) dx.}$$

Example 36.9 *Let X and Y be two independent uniform(0,1) random variables. Find $P(X^3+Y>1)$.*

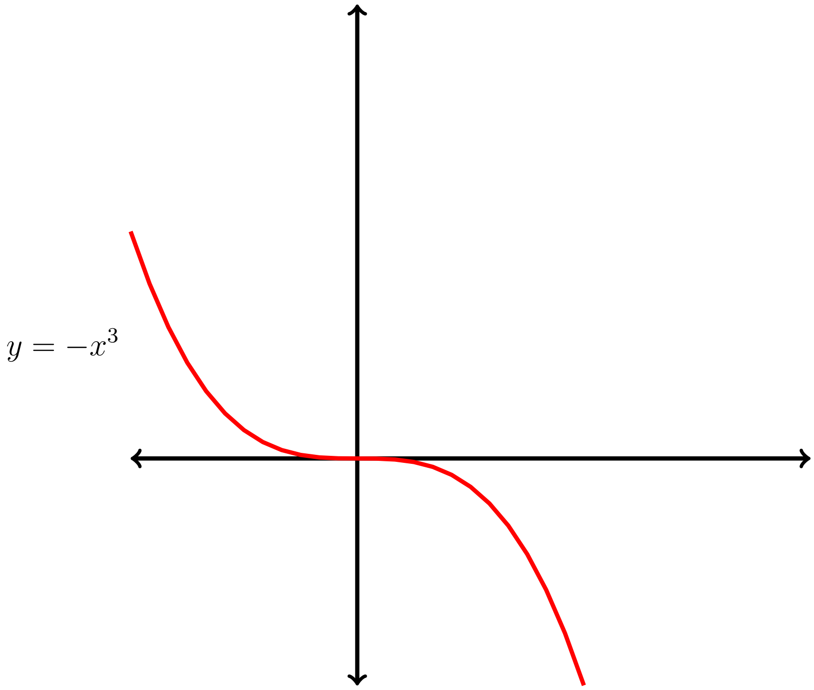
Solution: We discuss two solution methods, which one makes our life simple, we shall see it.

1. One approach would be form the joint pdf $f(x, y) = f_X(x)f_Y(y)$ and then compute the following integral:
$$P(X^3 + Y > 1) = \iint_A f(x, y)dxdy,$$
where $A := \{(x, y) \in \mathbb{R}^2 | x^3 + y > 1\}$. Since X and Y are independent, the joint pdf of X and Y is $f(x, y) = f_X(x)f_Y(y)$, where
$$f_X(x) = \begin{cases} 1 & \text{if} \quad 0 < x < 1 \\ 0 & \quad\text{otherwsie} \end{cases}, \; f_Y(y) = \begin{cases} 1 & \text{if} \quad 0 < y < 1 \\ 0 & \quad\text{otherwsie} \end{cases},$$
Hence
$$f(x, y) = \begin{cases} 1 & \text{if} \quad 0 < x < 1, 0 < y < 1 \\ 0 & \quad\text{otherwsie} \end{cases}.$$
We need to plot the graph of the function $f(x) = 1 - x^3$.

We know the graph of $y = x^3$.

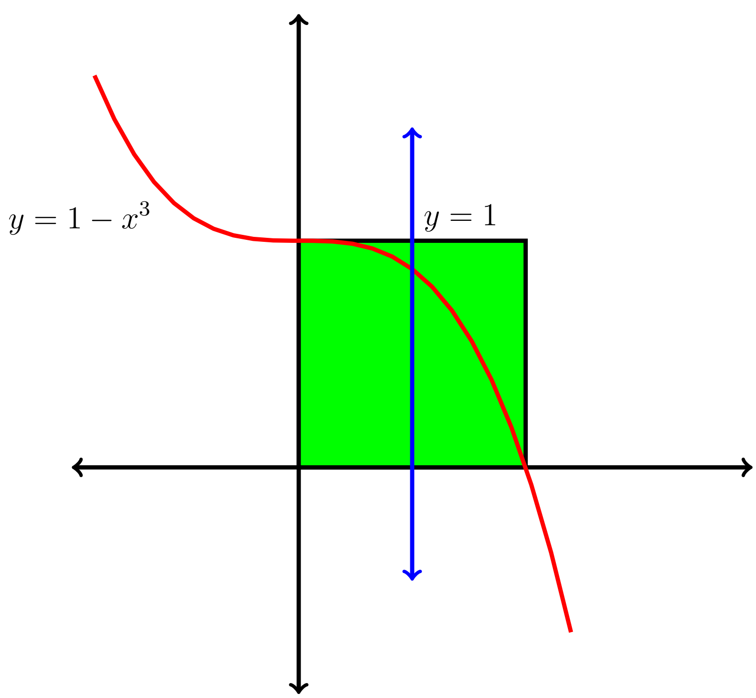


Now $y = -x^3$ is reflecting it w.r.t. y axis.



Now $y = 1 - x^3$ is lifting the above graph by unit 1.

Gathering all the above information we plot the graph of $f(x) = 1 - x^3$.



Setting up the limits as in figure:

$$\iint_A f(x,y)dxdy = \int_0^1 \left(\int_{1-x^3}^1 dy\right) dx = \int_0^1 \left(1-(1-x^3)\right) dx = \int_0^1 x^3 dx = \frac{1}{4}.$$

2. Now we illustrate how the conditioning is useful: One can condition either w.r.t. $Y = y$ or $X = x$.

   (a) We condition w.r.t. Y. Hence by total probability law:

$$\begin{aligned}
P(X^3+Y>1) &= \int_{-\infty}^{\infty} P(X^3+Y>1|Y=y)f_Y(y)dy = \int_0^1 P(X^3+y>1|Y=y)dy \\
&= \int_0^1 P(X^3>1-y|Y=y)dy = \int_0^1 P(X^3>1-y)dy \\
&= \int_0^1 P(X>(1-y)^{\frac{1}{3}})dy \; (\because 0<y<1, X^3>1-y \iff X>(1-y)^{\frac{1}{3}}\,) \\
&= \int_0^1 \left(\int_{\sqrt[3]{1-y}}^1 dx\right) dy \\
&= \int_0^1 \left[1-\sqrt[3]{1-y}\right] dy = \int_1^0 (1-\sqrt[3]{u})(-du) = \left[u - \frac{3}{4}u^{\frac{4}{3}}\right]_0^1 = \frac{1}{4}
\end{aligned}$$

   (b) We condition w.r.t. X. Hence by total probability law:

$$\begin{aligned}
P(X^3+Y>1) &= \int_{-\infty}^{\infty} P(X^3+Y>1|X=x)f_X(x)dx = \int_0^1 P(X^3+Y>1|X=x)dx \\
&= \int_0^1 P(x^3+Y>1|X=x)dx = \int_0^1 P(Y>1-x^3|X=x)dx \\
&= \int_0^1 P(Y>1-x^3)dx = \int_0^1 \left(\int_{1-x^3}^1 dy\right) dx \\
&= \int_0^1 x^3 dx = \frac{x^4}{4}\Big|_0^1 = \frac{1}{4}
\end{aligned}$$

■

Lecture 37: Conditional Expectation

2 April, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Now we define conditional expectation denoted by $E[X|Y = y]$ of the random variable X given the information $Y = y$. A conditional expectation is the same as an ordinary expectation, except that it refers to the new universe and pmf/pdf is replaced by their conditional counterparts.

Definition 37.1 *Let X and Y be discrete random variables with conditional pmf $f_{X|Y}$ of X given Y. Then conditional expectation of X given $Y = y$ is defined as*

$$E[X|Y = y] = \sum_{x \in R_X} x f_{X|Y}(x|y),$$

provided $\sum_{x \in R_X} |x| f_{X|Y}(x|y) < \infty$.

Example 37.2 *Let X, Y be independent random variables with geometric distribution of parameter $0 < p < 1$. Calculate $E[Y|X + Y = n]$ where $n \geq 2$.*

Solution: First we find the conditional pmf of Y given $X + Y = n$ where $n \geq 2$.

Since range of X and Y is $\mathbb{N}$, hence the range of the random variable $Z := X+Y$ is $\{2, 3, \cdots\}$. Let $n \geq 2$ be given. So if $X + Y = n$ then Y can only assume values in $\{1, 2, \cdots, n - 1\}$. Therefore

$$P(Y = y|Z = n) = 0, \quad \text{for } y = n, n + 1, n + 2, \cdots$$

For $y \in \{1, 2, \cdots, n-1\}$

$$P(Y = y | Z = n) = \frac{P(Y = y, X + Y = n)}{P(X + Y = n)} = \frac{P(Y = y, X + y = n)}{P\left(\bigcup_{k=1}^{n-1} \{X = k, Y = n - k\}\right)}$$

$$= \frac{P(Y = y, X = n - y)}{\sum_{k=1}^{n-1} P(X = k, Y = n - k)} \quad \text{(By finite additivity of Probability measure)}$$

$$= \frac{P(Y = y)P(X = n - y)}{\sum_{k=1}^{n-1} P(X = k)P(Y = n - k)} \quad (\because X, Y \text{ are independent })$$

$$= \frac{p(1-p)^{y-1}p(1-p)^{n-y-1}}{\sum_{k=1}^{n-1} p(1-p)^{k-1}p(1-p)^{n-k-1}} = \frac{p^2(1-p)^{n-2}}{\sum_{k=1}^{n-1} p^2(1-p)^{n-2}} = \frac{1}{n-1}$$

This shows that

$$f_{Y|X+Y}(y|n) = \begin{cases} \dfrac{1}{n-1} & \text{if} \quad y = 1, \cdots, n-1, \\ 0 & \text{if} \quad y \geq n \end{cases}$$

Hence Y is geometrically distributed in the original universe, but in the new universe determined by the event $X + Y = n$, Y is uniformly (discrete) distributed over the set $\{1, 2, \cdots, n-1\}$.

Hence

$$\begin{aligned} E[Y|Z = n] &= \sum_y y f_{Y|Z}(y|n) \\ &= \sum_{y=1}^{n-1} y \frac{1}{n-1} \\ &= \frac{1}{n-1} \times \frac{(n-1)(n-1+1)}{2} \\ &= \frac{n}{2} \end{aligned}$$

■

Definition 37.3 *Let X and Y be random variables with conditional pdf $f_{X|Y}$ of X given Y. The conditional expectation of X given $\{Y = y\}$ is defined as*

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx,$$

provided $\int_{-\infty}^{\infty} |x| f_{X|Y}(x|y)dx < \infty$.

Theorem 37.4 *1. Let X, Y be discrete random variables with joint pmf f. If Y has finite mean then*

$$E[Y] = \sum_x E[Y|X = x] f_X(x)$$

2. Let X, Y be random variables with joint pdf f. If Y has finite mean then

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

Proof:

$$\begin{aligned}
\sum_x E[Y|X = x] f_X(x) &= \sum_x \left(\sum_y y f_{Y|X}(y|x) \right) f_X(x) \\
&= \sum_x \sum_y y f(x, y) \\
&= \sum_y y \sum_x f(x, y) \\
&= \sum_y y f_Y(y) \\
&= EY
\end{aligned}$$

■

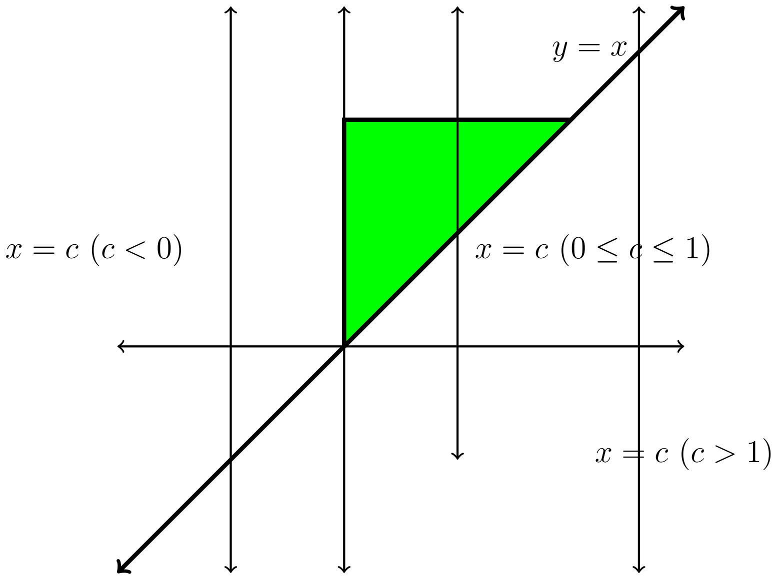
Remark 37.5 *The above theorem is called total expectation theorem. It express the fact that "the unconditional average can be obtained by averaging the conditional averages". They can be used to calculate the unconditional expectation $E[X]$ from the conditional expectation.*

Example 37.6 *Let X, Y be continuous random variables with joint pdf given by*

$$f(x, y) = \begin{cases} 6(y - x) & ; \quad 0 \le x \le y \le 1 \\ 0 & ; \quad \textit{otherwise} \end{cases}$$

Find $E[Y|X = x]$ and hence calculate EY.

Solution: In order to calculate $E[Y|X = x]$ we need to find $f_{Y|X}$, which is by definition equal to $\frac{f(x, y)}{f_X(x)}$.



Note that

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \begin{cases} \displaystyle\int_x^1 f(x,y)dy & ; \quad 0 \le x \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases} \\
&= \int_x^1 6(y-x)dy \\
&= 6\left[\frac{y^2}{2} - xy\right]_x^1 \\
&= 6\left[\frac{x^2}{2} - x + \frac{1}{2}\right] \\
&= \begin{cases} 3(x-1)^2 & ; \quad 0 \le x \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases}
\end{aligned}
$$

This implies that

$$
f_{Y|X}(y|x) = \begin{cases} \dfrac{2(y-x)}{(x-1)^2} & ; \quad 0 \le x \le y < 1 \\ 0 & ; \quad \text{otherwise} \end{cases}
$$

Hence $E[Y|X = x]$ would be non-zero only if $0 \leq x < 1$.

$$\begin{aligned}
E[Y|X = x] &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy \\
&= \int_x^1 y \frac{2(y-x)}{(x-1)^2} dy \\
&= \frac{2}{(x-1)^2} \int_x^1 (y^2 - xy)dy \\
&= \frac{2}{(x-1)^2} \left[\frac{y^3}{3} - x\frac{y^2}{2} \right]_x^1 \\
&= \frac{2}{(x-1)^2} \left[\frac{1}{3} - \frac{x}{2} + \frac{x^3}{6} \right] \\
&= \frac{2(x^3 - 3x + 2)}{6(x-1)^2} \\
&= \frac{x^2 + x - 2}{3(x-1)}
\end{aligned}$$

Therefore

$$\begin{aligned}
EY &= \int_{-\infty}^{\infty} E[Y|X = x] f_X(x)dx \\
&= \int_0^1 \frac{x^2 + x - 2}{3(x-1)} \times 3(x-1)^2 dx \\
&= \int_0^1 (x^2 + x - 2)(x-1)dx \\
&= \int_0^1 (x^3 - 3x + 2)dx \\
&= \frac{3}{4}
\end{aligned}$$

■

Theorem 37.7 *1. Let X and Y be discrete random variables with joint pmf f. If g is a function then*

$$E[g(X)|Y = y] = \sum_x g(x) f_{X|Y}(x|y),$$

provided $\sum_x |g(x)| f_{X|Y}(x|y) < \infty$.

2. Let X and Y be random variables with joint pdf f. If $g : \mathbb{R} \to \mathbb{R}$ is a Borel function then

$$E[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx,$$

provided $\displaystyle\int_{-\infty}^{\infty} |g(x)| f_{X|Y}(x|y) dx < \infty.$

It is immediate that conditional expectation satisfies the usual properties of an expectation. The following results are easy to prove. We assume existence of indicated expectations.

Theorem 37.8 (Properties of Conditional Expectation) *Let X and Y be two random variable on a probability space $(\Omega, \mathcal{F}, P)$. and let a, b and c be real numbers. Suppose g_1 and g_2 are real-valued function of one real variable such that $E\left[g_1(X)\right] < \infty$ and $E\left[g_2(X)\right] < \infty$. Then*

(a) $E\left[a g_1(X) + b g_2(X) + c | Y = y\right] = aE\left[g_1(X) | Y = y\right] + bE\left[g_2(X) | Y = y\right] + c.$

(b) *If $g_1(x) \geq g_2(x)$ for all x, then $E\left[g_1(X) | Y = y\right] \geq E\left[g_2(X) | Y = y\right]$.*

(c) *Let Z be another random variable then, $E[X + Y | Z = z] = E[X | Z = z] + E[Y | Z = z]$.*

(d) *If X and Y are independent, then*

$$E[X | Y = y] = EX, \quad E[Y | X = x] = E[Y].$$

Lecture 38: Covariance & Correlation

3 April, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

When two random variables are dependent, we say there exist a relationship between them. But if there is a relationship, the relationship may be strong or weak. Now we discuss two numerical measure of the strength of a relationship between two random variables, the covariance and correlation.

38.1 Covariance

Definition 38.1 *Let X and Y be jointly distributed on the probability space $(\Omega, \mathcal{F}, P)$. The covariance of two random variables X and Y, denoted by* $\operatorname{cov}(X, Y)$ *is defined by*

$$\operatorname{cov}(X, Y) = E\left[(X - EX)(Y - EY)\right],$$

provided $E\left|(X - EX)(Y - EY)\right| < \infty$. When $\operatorname{cov}(X, Y) = 0$, we say that X and Y are uncorrelated.

The covariance gives information about how random variable X and Y are linearly related. Intuitively, the covariance between X and Y indicates how the values of X and Y move relative to each other. If large values of X tend to happen with large values of Y, then the covariance is positive (or small values of X tend to happen with small values of Y) and we say X and Y are positively correlated. On the other hand, if X tends to be small when Y is large (or vice-versa), then the covariance is negative and we say X and Y are negatively correlated.

Alternate expression for Covariance: By appealing to the linearity of the expectation,

$$\begin{aligned}
\operatorname{cov}(X, Y) &= E\left[XY - XEY + YEX - EXEY\right] \\
&= E[XY] - EXEY + EYEX - EXEY \\
&= E[XY] - E[X]E[Y].
\end{aligned}$$

Independence & Covariance If X and Y are independent then $E[XY] = EXEY$, therefore $\operatorname{cov}(X, Y) = 0$. But converse is not true in general.

Example 38.2 *Let the joint probabilities of random variables X and Y are given by the following table.*

X \ Y	*-1*	*0*	*1*
-1	0	$\frac{1}{4}$	0
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{4}$	0

Then X and Y are identically distributed and X has the following pmf

$$P(X=-1) = P(X=1) = \frac{1}{4} \quad \textit{and } P(X=0) = \frac{1}{2}$$

Also, it is easy to see that $E[X] = E[Y] = 0$. Furthermore, random variable XY takes values $\{-1, 0, 1\}$ with the pmf

$$P(XY=1) = 0 = P(XY=-1) \textit{ and } P(XY=0) = 1.$$

Therefore, $E[XY] = 0$, which in turn implies $\operatorname{cov}(X,Y) = 0$. However, X and Y are not independent since

$$P(X=-1, Y=-1) = 0 \neq \frac{1}{16} = P(X=-1)P(Y=-1)$$

Proposition 38.3 *For any random variable X, Y and Z, and any $a, b \in \mathbb{R}$,*

1. $\operatorname{cov}(X,X) = \operatorname{var}(X)$. ✓
2. $\operatorname{cov}(X,Y) = \overset{\text{cov}}{\cancel{\text{var}}}(Y,X)$. ✓
3. $\operatorname{cov}(X, aY+b) = a \operatorname{cov}(X,Y)$. ✓
4. $\operatorname{cov}(X, Y+Z) = \operatorname{cov}(X,Y) + \operatorname{cov}(X,Z)$. ✓

Proof: All the proof follows from definition of covariance and linearity of expectation.

1. $$\operatorname{cov}(X,X) = E[X^2] - EX\ EX = var(X)$$

2. immediate from definition.

3. $$\begin{aligned}\operatorname{cov}(X, aY+b) &= E[X(aY+b)] - EXE[aY+b] = E[aXY+bX] - EX[aEY+b] \\ &= aE[XY] + bEX - aEXEY - bEX = a\left[E[XY] - EXEY\right] \underset{a}{=} \operatorname{cov}(X,Y)\end{aligned}$$

4.

$$\begin{aligned}\mathrm{cov}(X, Y+Z) &= E[X(Y+Z)] - EXE(Y+Z) = E[XY+XZ] - EX[EY+EZ] \\ &= E[XY] + E[XZ] - EXEY - EXEZ = \mathrm{cov}(X,Y) + \mathrm{cov}(X,Z)\end{aligned}$$

■

Example 38.4 *Let X and Y be two independent $N(0,1)$ random variables and $Z = 1 + X + XY^2, W = 1 + X$. Find $\mathrm{cov}(Z, W)$.*

Solution:

$$\begin{aligned}\mathrm{cov}(Z,W) &= \mathrm{cov}(1 + X + XY^2, 1 + X) \\ &= \mathrm{cov}(X + XY^2, X) \quad (\because \text{adding a constant to any rv does not affect the covariance}) \\ &= \mathrm{cov}(X,X) + \mathrm{cov}(XY^2, X) \\ &= var(X) + E[XY^2X] - E[XY^2]EX = 1 + E[X^2]E[Y^2] = 2\end{aligned}$$

■

Example 38.5 *For any random variables X, Y, show that*

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y).$$

Solution:

$$\begin{aligned}\mathrm{var}(X+Y) &= \mathrm{cov}(X+Y, X+Y) = \mathrm{cov}(X+Y, X) + \mathrm{cov}(X+Y, Y) \\ &= \mathrm{cov}(X,X) + \mathrm{cov}(Y,X) + \mathrm{cov}(X,Y) + \mathrm{cov}(Y,Y) \\ &= \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y).\end{aligned}$$

■

If X and Y are independent then from Example 38.5, it follows that

$$var(X+Y) = var(X) + var(Y).$$

So variance of sum of independent random variables is the sum of variance of random variables.

38.2 Correlation

The sign of $\text{cov}(X, Y)$ gives information about the linear relationship of X and Y; however, its actual magnitude does not have much meaning since it depends on the variability of X and Y. Therefore $\text{cov}(X, Y)$ the number itself does not give information about the strength of the relationship between X and Y.

The correlation coefficient removes, in a sense, the individual variability of each X and Y by dividing the covariance by the product of the standard deviations, and thus the correlation coefficient is a better measure of the linear relationship of X and Y than is the covariance. Also, the correlation coefficient is unitless.

Definition 38.6 *The correlation coefficient of two random variables X and Y, denoted by $\rho(X, Y)$ is defined as*

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}},$$

provided $var(X) > 0$ and $var(Y) > 0$.

$$\boxed{\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)\, var\, Y}}}$$

Lecture 39: Correlation & Characteristics Function

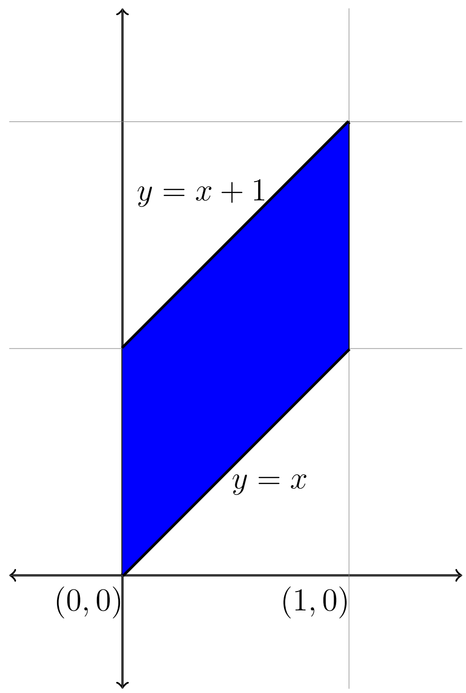
5 April, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Example 39.1 *Let the joint pdf of* (X, Y) *be*

$$f(x,y) = \begin{cases} 1 & ; \quad 0 < x < 1, x < y < x+1 \\ 0 & ; \quad \textit{otherwise} \end{cases}$$



$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \begin{cases} 0 & ; \quad x \leq 0 \\ \displaystyle\int_x^{x+1} 1 \; dy = 1 & ; \quad 0 < x < 1 \\ 0 & ; \quad x \geq 1 \end{cases}$$

Then $X \sim U(0,1)$. *Hence* $EX = \frac{1}{2}$ *and* $var(X) = \frac{1}{12}$. *The marginal pdf of* Y *is*

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \begin{cases} \displaystyle\int_0^y 1 \; dx = y & ; \quad 0 < y < 1 \\ \displaystyle\int_{y-1}^1 1 \; dx = 2 - y & ; \quad 1 \leq y < 2 \\ 0 & ; \quad \textit{otherwise} \end{cases}$$

Hence $EY = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^1 y^2 dy + \int_1^2 (2y - y^2)dy = \frac{y^3}{3}\Big|_0^1 + \left[y^2 - \frac{y^3}{3}\right]_1^2 = \frac{1}{3} + \frac{2}{3} = 1$
and $var(Y) = E[Y^2] - 1 = \frac{1}{6}$, *because*

$$E[Y^2] = \int_{-\infty}^{\infty} y^2 f_Y(y)dy = \int_0^1 y^3 dy + \int_1^2 (2y^2 - y^3)dy = \frac{y^2}{3}\Big|_0^1 + \left[\frac{2y^3}{3} - \frac{y^4}{4}\right]_1^2 = \frac{1}{3} + \frac{4}{3} - \frac{5}{12} = \frac{1}{3} + + \frac{11}{12} = \frac{7}{6}$$

We also have

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_0^1 \left[\int_x^{x+1} xydy\right]dx = \int_0^1 x\left[\frac{y^2}{2}\right]_x^{x+1} dx = \int_0^1 x\frac{(x+1)^2 - x^2}{2}dx$$
$$= \int_0^1 x\frac{(2x+1)}{2}dx = \int_0^1 \left(x^2 + \frac{x}{2}\right)dx = \left[\frac{x^3}{3} + \frac{x^2}{4}\right]_0^1 = \frac{7}{12}$$

Hence

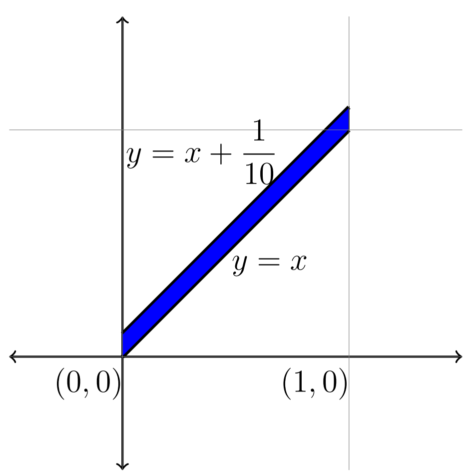
$$Cov(X,Y) = E[XY] - EXEY = \frac{7}{12} - \frac{1}{2} = \frac{1}{12}.$$

The correlation is

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{\frac{1}{12}}{\sqrt{\frac{1}{6} \times \frac{1}{12}}} = \frac{1}{12} \times 6\sqrt{2} = \frac{1}{\sqrt{2}}.$$

Example 39.2 *Let the joint pdf of* (X,Y) *be*

$$f(x,y) = \begin{cases} 10 & ; \quad 0 < x < 1, x < y < x + \frac{1}{10} \\ 0 & ; \quad otherwise \end{cases}$$



$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \begin{cases} 0 & ; \quad x \leq 0 \\ \displaystyle\int_x^{x+\frac{1}{10}} 10\ dy = 1 & ; \quad 0 < x < 1 \\ 0 & ; \quad x \geq 1 \end{cases}$$

Then $X \sim U(0,1)$. *Hence* $EX = \frac{1}{2}$ *and* $var(X) = \frac{1}{12}$. *The marginal pdf of* Y *is*

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \begin{cases} \displaystyle\int_0^y 10\ dx = 10y & ; \quad 0 < y < \frac{1}{10} \\ \displaystyle\int_{y-\frac{1}{10}}^{y} 10\ dx = 1 & ; \quad \frac{1}{10} \leq y < 1 \\ \displaystyle\int_{y-\frac{1}{10}}^{1} 10\ dx = 10\left(1 - y + \frac{1}{10}\right) = 11 - 10y & ; \quad 1 \leq y < 1 + \frac{1}{10} \\ 0 & ; \quad otherwise \end{cases}$$

Hence

$$\begin{aligned} EY &= \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^{1/10} 10y^2 dy + \int_{1/10}^{1} y dy + \int_1^{1+1/10} (11y - 10y^2)dy \\ &= \frac{10y^3}{3}\Big|_0^{1/10} + \frac{y^2}{2}\Big|_{1/10}^{1} + \left[\frac{11y^2}{2} - \frac{10y^3}{3}\right]_1^{1+1/10} \\ &= \frac{11}{20} \end{aligned}$$

and $var(Y) = \frac{29}{75} - \frac{121}{400} = \frac{101}{1200}$, *because*

$$E[Y^2] = \int_{-\infty}^{\infty} y^2 f_Y(y)dy = \int_0^{1/10} 10y^3 dy + \int_{1/10}^{1} y^2 dy + \int_1^{1+1/10} (11y^2 - 10y^3)dy$$

$$= \frac{1}{4000} + \frac{333}{1000} + \frac{641}{12000} = \frac{29}{75}$$

We also have

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = 10\int_0^1 x\left[\int_x^{x+\frac{1}{10}} ydy\right]dx = 10\int_0^1 x\left[\frac{y^2}{2}\right]_x^{x+\frac{1}{10}} dx \\ &= 10\int_0^1 x\frac{(x+\frac{1}{10})^2 - x^2}{2}dx \\ &= 10\int_0^1 x\frac{(2x+\frac{1}{10})}{2} \times \frac{1}{10}dx = \int_0^1 \left(x^2 + \frac{x}{20}\right)dx = \left[\frac{x^3}{3} + \frac{x^2}{40}\right]_0^1 = \frac{43}{120} \end{aligned}$$

Hence

$$Cov(X,Y) = E[XY] - EXEY = \frac{43}{120} - \frac{11}{40} = \frac{1}{12}.$$

The correlation is

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{\frac{1}{12}}{\sqrt{\frac{1}{12} \times \frac{101}{1200}}} = \frac{1}{12} \times \frac{120}{\sqrt{101}} = \frac{10}{\sqrt{101}} = \sqrt{\frac{100}{101}}.$$

Superiority of Correlation over Covariance

Observe that in both Examples 39.1 and 39.2, the Cov(X,Y) is same but $\rho(X,Y)$ in Example 39.2 is much larger than the $\rho(X,Y)$ in Example 39.1. These two examples together tells us that the magnitude of covariance does not give any information about the strength of the linear relationship between random variable X and Y but correlation does give that information.

In Example 39.1, look at the region in which in the joint density is non-zero. If we draw a line $x = c$ passing in this reason then random variable Y could take many other values.

But in Example 39.2, if draw a line line $x = c$ passing in this reason then random variable Y takes values in very restricted manner. This suggest that there is a very strong relationship in this case. Though in both cases there is a linearly increasing relationship between X and Y.

The nature of the linear relationship measured by the covariance and correlation is somewhat explained in the following theorem.

Proposition 39.3 *The correlation coefficient between two random variables X and Y satisfies the following properties.*

1. $|\rho(X,Y)| \leq 1$.
2. $|\rho(X,Y)| = 1$ *if and only if there exits real numbers a, b with $a \neq 0$ such that $Y = aX + b$. If $\rho(X,Y) = 1$ then $a > 0$ and if $\rho(X,Y) = -1$, then $a < 0$.*

Remark 39.4 *Intuitively, if there is a line $y = ax + b$, with $a \neq 0$, such that values of (X,Y) have high probability being near to this line, then the correlation between X and Y will be near 1 or -1. But if no such line exists, the correlation will be near zero.*

Example 39.5 *A standard normal random variable X satisfies: $EX = 0, EX^2 = 1, EX^3 = 0, EX^4 = 3$. Let $Y = a + bX + cX^2$. Find the correlation coefficient $\rho(X,Y)$.*

Solution:

$$\begin{aligned}
\mathrm{cov}(X,Y) &= E[XY] - E[X]E[Y] = E\left[aX + bX^2 + cX^3\right] - 0 \times E[Y] \\
&= aEX + bEX^2 + cEX^3 = b \\
\mathrm{var}(X) &= EX^2 - (EX)^2 = 1 \\
\mathrm{var}(Y) &= EY^2 - (EY)^2 = E(a^2 + b^2X^2 + 2abX + c^2X^4 + 2c(a+bX)X^2) - (a+c)^2 \\
&= a^2 + b^2 + 3c^2 + 2ac - a^2 - c^2 - 2ac = b^2 + 2c^2
\end{aligned}$$

Therefore

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} = \frac{b}{\sqrt{b^2+2c^2}}.$$

■

When Correlation fails

Covariance and correlation measure only a particular kind of linear relationship. But it may happen that X and Y have a strong relationship but their covariance and correlation are small or even zero, because the relationship is not linear. In fact in Example 39.5, we see that $|\rho(X,Y)| \le \frac{|b|}{\sqrt{2}|c|}$. Therefore if b is small and c is large then correlation is small. If $b = 0$, then $\mathrm{cov}(X,Y) = 0$ and $\rho(X,Y) = 0$ but $Y = a + cX^2$.

39.1 Complex-valued Random Variables

A complex-valued random variable $Z : \Omega \to \mathbb{C}$ can be written in the form $Z = X + iY$, where X and Y are real-valued random variables. Its expectation EZ is defined as $EZ = E(X+iY) = EX + iEY$ whenever EX and EY are well defined and finite. The formula $E(a_1Z_1 + a_2Z_2) = aEZ_1 + a_2EZ_2$ is valid whenever a_1 and a_2 are complex constants and Z_1 and Z_2 are complex-valued random variables having finite expectation.

39.2 Characteristic Function

We introduce the notion of characteristic function of a random variable and study its properties. Characteristic function serves as an important tool for analyzing random phenomenon.

Definition 39.6 *The characteristic function of a random variable X is defined*

$$\boxed{\phi_X(t) = E\left[e^{itX}\right],} \quad \underline{t \in \mathbb{R}}$$

So basically $\phi_X : \mathbb{R} \to \mathbb{C}$.

The advantage of the characteristic function is that it is defined for all real-valued random variables. Because for any real-valued random variable X and for any real number t, the random variables $\cos tX, \sin tX$ are bounded by 1. Therefore, both have finite expectation bounded by 1, hence $\phi_X(t)$ is defined for all t and for all X.

Characteristic Function of a discrete random variable: If X is a discrete random variable then

$$\begin{aligned}
\phi_X(t) &= E\left[e^{itX}\right] = E[\cos(tX)] + iE[\sin(tX)] \\
&= \sum_{x \in R(X)} \cos(tx)P(X = x) + i \sum_{x \in R(X)} \sin(tx)P(X = x) \\
&= \sum_{x \in R(X)} \left[\cos(tx) + i\sin(tx)\right] P(X = x) = \sum_{x \in R(X)} e^{itx} P(X = x)
\end{aligned}$$

Characteristic Function of a random variable with density: If random variable X has density f_X then

$$\begin{aligned}
\phi_X(t) &= E\left[e^{itX}\right] = E[\cos(tX)] + iE[\sin(tX)] \\
&= \int_{-\infty}^{\infty} \cos(tx) f_X(x)dx + i \int_{-\infty}^{\infty} \sin(tx) f_X(x)dx \\
&= \int_{-\infty}^{\infty} \left[\cos(tx) + i\sin(tx)\right] f_X(x)dx = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx
\end{aligned}$$

Example 39.7 *Let $X \sim$ Bernoulli (p). Find its characteristic function.*

Solution:

$$\begin{aligned}
\phi_X(t) &:= E\left[e^{itX}\right] \\
&= e^{it}P(X = 1) + e^0 P(X = 0) \\
&= e^{it}p + (1 - p)
\end{aligned}$$

■

Lecture 40: Characteristics Function

8 April, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Example 40.1 *Find the characteristic function of the Poisson(λ) distribution.*

Solution:

$$\begin{aligned}\phi_X(t) = E\left[e^{itX}\right] &= \sum_{k=0}^{\infty} e^{itk} P(X=k) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = e^{-\lambda} \exp(e^{it}\lambda) \\ &= \exp\left(\lambda(e^{it}-1)\right)\end{aligned}$$

$\phi_{a+bX} = e^{ita} \phi(bt)$

∎

Example 40.2 *Let $X \sim N(0,1)$. Find it's characteristic function.*

$\phi_X(t) = e^{-t^2/2}$

Solution:

$$\begin{aligned}\phi_X(t) &= E[\cos tX + i \sin tX] = E[\cos tX] + iE[\sin tX] \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cos tx \; e^{-\frac{x^2}{2}} dx + i \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \sin tx \; e^{-\frac{x^2}{2}} dx\end{aligned}$$

Since characteristic function exists for every random variable, therefore both the improper integral exists. So value both improper integrals agrees with their Cauchy principle value. We have

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \sin tx \; e^{-\frac{x^2}{2}} dx = \lim_{a\to\infty} \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} \sin tx \; e^{-\frac{x^2}{2}} dx = 0,$$

because sin is an odd function. Also

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cos tx \; e^{-\frac{x^2}{2}} dx = \lim_{a\to\infty} \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} \cos tx \; e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} \cos tx \; e^{-\frac{x^2}{2}} dx = e^{-\frac{t^2}{2}}$$

where the last integral can computed using differentiation under integration. Let $t \in \mathbb{R}$ be given. Define

$$\begin{aligned}I(t) = \int_0^{\infty} \cos tx \; e^{-\frac{x^2}{2}} dx \implies I'(t) &= -\int_0^{\infty} x \sin tx \; e^{-\frac{x^2}{2}} dx \\ &= -\left[-\sin tx e^{-\frac{x^2}{2}} \Big|_0^{\infty} + \int_0^{\infty} t \cos tx \; e^{-\frac{x^2}{2}} dx \right] \\ &= 0 - tI(t)\end{aligned}$$

Therefore $\ln I(t) = -\frac{t^2}{2} + C \implies I(t) = Ke^{-\frac{t^2}{2}}$. Also $I(0) = \int_0^\infty e^{-\frac{x^2}{2}}dx = \frac{\sqrt{2\pi}}{2}$. So $K = \sqrt{\frac{\pi}{2}}$. ■

Example 40.3 *Let X be a random variable and a and b are real constants, then*

$$\phi_{a+bX}(t) = E\left[e^{it(a+bX)}\right] = E\left[e^{ita}e^{itbX}\right] = e^{ita}E\left[e^{itbX}\right] = e^{ita}\phi_X(bt)$$

Example 40.4 *Let $X \sim N(\mu, \sigma^2)$. Then it is implicit that $\sigma > 0$. Then $Y = \frac{X-\mu}{\sigma}$ has mean zero and variance 1. Also $Y \sim N(0,1)$. Hence by Example 40.3, $X = \sigma Y + \mu$ has the characteristic function*

$$\phi_X(t) = \phi_{\sigma Y+\mu}(t) = e^{it\mu}\phi_Y(\sigma t) = e^{it\mu}e^{-\frac{\sigma^2 t^2}{2}}$$

Example 40.5 *Let X and Y be independent random variables. Show that*

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$$

Solution:

$$\begin{aligned}\phi_{X+Y}(t) &= E[e^{it(X+Y)}] \\ &= E[e^{itX}e^{itY}] = E[e^{itX}]E[e^{itY}] = \phi_X(t)\phi_Y(t)\end{aligned}$$

■

More generally, if $X_1, X_2, \cdots, X_n$ are n independent random variables, then

$$\phi_{X_1+X_2+\cdots+X_n}(t) = \phi_{X_1}(t)\phi_{X_2}(t)\cdots\phi_{X_n}(t).$$

Example 40.6 *Compute the characteristic function of a Binomial(n,p) random variables.*

Solution: A Binomial(n,p) random variable is a sum of n independent Bernoulli(p) random variables. Therefore it's characteristic function is

$$\left[e^{it}p + (1-p)\right]^n.$$

■

Theorem 40.7 (Uniqueness Theorem) *Let X_1 and X_2 be two random variables such that $\phi_{X_1} = \phi_{X_2}$. Then X_1 and X_2 have same distribution.*

Example 40.8 *Let $X \sim B(n_1,p)$ and $Y \sim B(n_2,p)$ be two independent Binomial random variables. Show that $X+Y$ is a Binomial(n_1+n_2,p) random variable.*

Solution: Let $X \sim B(n_1, p)$ and $Y \sim B(n_2, p)$ be two independent random variables. Therefore the characteristic function of $X + Y$ is

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \left[e^{it}p + (1-p)\right]^{n_1} \left[e^{it}p + (1-p)\right]^{n_2} = \left[e^{it}p + (1-p)\right]^{n_1+n_2}.$$

RHS is a characteristic function of a Binomial$(n_1 + n_2, p)$ random variable, therefore by uniqueness theorem $X + Y \sim Binomial(n_1 + n_2, p)$. ■

Example 40.9 *Let $X \sim Poisson(\lambda)$ and $Y \sim Poisson(\mu)$ be two independent Poisson random variables. Show that $X + Y$ is a $Poisson(\lambda + \mu)$ random variable.*

Solution: The characteristic function of $X + Y$ is

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \exp\left(\lambda(e^{it} - 1)\right) \exp\left(\mu(e^{it} - 1)\right) = \exp\left[(\lambda + \mu)(e^{it} - 1)\right]$$

RHS is a characteristic function of a Poisson$(\lambda+\mu)$ random variable, therefore by uniqueness theorem $X + Y \sim$Poisson$(\lambda + \mu)$. ■

Example 40.10 *Let $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent normal random variable. Then show that $X + Y$ is a $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Solution: Hence we have

$$\phi_X(t) = e^{it\mu_1} e^{-\frac{\sigma_1^2 t^2}{2}}, \quad \phi_Y(t) = e^{it\mu_2} e^{-\frac{\sigma_2^2 t^2}{2}}.$$

Now

$$\begin{aligned}\phi_{X+Y}(t) &:= E\left[e^{it(X+Y)}\right] = E\left[e^{itX}e^{itY}\right] = E\left[e^{itX}\right] E\left[e^{itY}\right] = \phi_X(t)\phi_Y(t) \\ &= e^{it(\mu_1+\mu_2)} e^{-\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}\end{aligned}$$

Now right hand side is the characteristic function of a normal random variable with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Therefore by uniqueness theorem, we conclude $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ ■

Lecture 41: Inequalities

9 April, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Definition 41.1 *Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be a function. We say that*

1. *f is convex on I or concave upward on I if for any $x_1, x_2 \in I$ and any $t \in (0,1)$ we have*
$$f((1-t)x_1 + tx_2) \le (1-t)f(x_1) + tf(x_2).$$

2. *f is concave on I or concave downward on I if for any $x_1, x_2 \in I$ and any $t \in (0,1)$ we have*
$$f((1-t)x_1 + tx_2) \ge (1-t)f(x_1) + tf(x_2).$$

It follows from the definition that f is convex iff $-f$ is concave, (reflection about the x-axis).

Theorem 41.2 (Jensen's Inequality) *Let $f : I \to \mathbb{R}$ be a convex function where $I \subset \mathbb{R}$ is an interval and X be a random variable such that X and $f(X)$ has finite mean. Then*

$$f(EX) \le E[f(X)].$$

If f is a concave function then $-f$ is convex so by Jensen's inequality

$$\begin{aligned} -f(EX) &\le E(-f(X)) \\ &= -E[f(X)] \quad \text{(By linearity of expectation)} \\ \implies f(EX) &\ge E[f(X)] \end{aligned}$$

Example 41.3 *Note that $f(x) = |x|$ is a convex function hence by Jensen's inequality*

$$EX \le |EX| \le E|X|.$$

Definition 41.4 *Let r be a positive real number and X be a random variable. Then $E[X^r]$ is called the r-th moment of X about origin or central moment of X of order r.*

$E|X|^r$ is called the r-th absolute moment of X about origin or central moment of X of order r.

We know from definition of expectation that, $E\left[X^r\right]$ exists and is a finite number if $E\left[|X^r|\right] < \infty$. Therefore from above observation

$$E[X^r] \leq |E[X^r]| \leq E\left[|X|^r\right].$$

Example 41.5 *If the moment of order $q > 0$ exists for a random variable X, then show that moments of order p, where $0 < p < q$ exist.*

Solution: Let $f : (0, \infty) \to \mathbb{R}$ be defined as $f(x) = x^r$, where $r > 1$ is a real number. Then $f'(x) = rx^{r-1}$, $f''(x) = r(r-1)x^{r-2}$. Since $r > 1$, $f''(x) > 0$ on $(0, \infty)$, i.e., f is a convex function on $(0, \infty)$. Hence by Jensen's inequality,

$$[E|X|]^r \leq E\left[|X|^r\right] \implies [E|X|] \leq \left(E\left[|X|^r\right]\right)^{\frac{1}{r}}. \tag{41.1}$$

Let $0 < p < q$. Then we take $r = \frac{q}{p} > 1$ in (41.1) and we get

$$[E|X|] \leq \left(E\left[|X|^{\frac{q}{p}}\right]\right)^{\frac{p}{q}}. \tag{41.2}$$

Now replacing $|X|$ by $|X|^p$ in (41.2), we get

$$[E|X|^p] \leq \left(E\left[|X|^q\right]\right)^{\frac{p}{q}}$$

If $E|X|^q < \infty$ then $(E|X|^q)^{\frac{p}{q}} < \infty$ and therefore $[E|X|^p] < \infty$. ∎

Example 41.6 *Let X be a random variable with $EX = 10$. Show that $E\left[\ln\sqrt{X}\right] \leq \frac{1}{2}\ln 10$.*

Solution: Consider $f(x) = \ln\sqrt{x} = \frac{1}{2}\ln x$, for $x \in (0, \infty)$. Then $f'(x) = \dfrac{1}{2x}$ and $f''(x) = -\dfrac{1}{2x^2} < 0$ on $(0, \infty)$. Hence f is a concave function. Therefore by Jensen's inequality

$$\frac{1}{2}\ln 10 = f(EX) \geq E[f(X)] = E\left[\ln\sqrt{X}\right].$$

∎

Now we derive some important inequalities. These inequalities use the mean and possibly the variance of a random variable to draw conclusions on the probabilities of certain events. They are primarily useful in situations where exact values or bounds for the mean and variance of a random variable X are easily computable, but the distribution of X is either unavailable or hard to calculate.

Theorem 41.7 (Markov Inequality) *Let X be a non-negative random variable with finite nth moment. Then we have for each $\epsilon > 0$,*

$$P\{X \geq \epsilon\} \leq \frac{E[X^n]}{\epsilon^n}$$

Loosely speaking, Markov inequality asserts that if a nonnegative random variable has a small nth central moment, then the probability that it takes a large value must also be small.

As a corollary we have the Chebyshev's inequality.

Corollary 41.8 (Chebyshev's inequality) *Let X be a random variable with finite mean μ and finite variance σ^2. Then for every $\epsilon > 0$,*

$$P\{|X - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}$$

Proof: The proof of Chebyshev's inequality follows by replacing X by $|X - \mu|$ in the Markov inequality and realizing that $|X - \mu|^2 = [X - \mu]^2$. ■

Remark 41.9 *Loosely speaking, Chebyshev's inequality asserts that if a random variable has small variance, then the probability that it takes a value far from its mean is also small. Note that the Chebyshev inequality does not require the random variable to be nonnegative.*

Example 41.10 (Illustrating Chebychev) *If we take $\epsilon = 2\sigma$, then*

$$P\{|X - \mu| \geq 2\sigma\} \leq \frac{\sigma^2}{4\sigma^2} = 0.25,$$

so there is at least a 75% chance that a random variable will be within 2σ of its mean, no matter what the distribution of X.

Similarly if we take $\epsilon = 3\sigma$, then

$$P\{|X - \mu| \geq 3\sigma\} \leq \frac{\sigma^2}{9\sigma^2} = 0.1111111\cdots,$$

so there is at least a 89% chance that a random variable will be within 3σ of its mean, no matter what the distribution of X. Recall that for the normal distribution it is 99% chance that the random variable will be within 3σ of its mean. So the estimates provided by Chebychev's inequality could be crude. One reason for is that it puts no restrictions on the underlying distributions. Though Chebyshev's inequality is necessarily conservative but its applicability is wide. In particular, we can often get tighter bounds for some specific distributions.

Example 41.11 *Let $X \sim B(n,p)$. Estimate $P(X \geq \alpha n)$, where $p < \alpha < 1$ using Markov (for first moment) and Chebyshev's inequality. Compare both the estimates for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$.*

Solution: Note that X takes values $\{0, 1, \cdots, n\}$, hence is a nonnegative random variable and $EX = np$. Applying Markov's inequality, we obtain

$$P(X \geq \alpha n) \leq \frac{EX}{\alpha n} = \frac{pn}{\alpha n} = \frac{p}{\alpha}$$

Chebyshev's inequality gives estimate for $P(|X - EX| \geq \alpha n)$ so we have rewrite the event $\{X \geq \alpha n\}$ so that we can use the Chebushev's inequality.

$$\begin{aligned} P\{X \geq \alpha n\} &= P\{X - np \geq \alpha n - np\} \\ &\leq P\left(|X - np| \geq \alpha n - np\right) \quad (\because \{|Y| \geq a\} = \{Y \leq -a\} \cup \{Y \geq a\}) \\ &\leq \frac{\text{var(X)}}{(\alpha n - np)^2} = \frac{np(1-p)}{n^2(\alpha - p)^2} = \frac{p(1-p)}{n(\alpha - p)^2} \end{aligned}$$

By Markov inequality for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$, we

$$P\left(X \geq \frac{3n}{4}\right) \leq \frac{2}{3}$$

By Chebyshev's inequality for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$, we

$$P\left(X \geq \frac{3n}{4}\right) \leq \frac{4}{n}$$

If $n \geq 6$ then estimate given by Chebyshev's are sharper than the estimates provided by Markov inequality. Also as n increases, estimate given by Chebyshev's inequality decreases, i.e., gives much information whereas the estimates provided by Markov inequality remains constant as n varies.

■

Lecture 42: Law of Large Numbers

10 April, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Theorem 42.1 (Weak Law of Large numbers) *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having finite mean μ. Then for every $\delta > 0$,*

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \mu\right| \geq \delta\right\} = 0, \quad \text{or equivalently} \quad \lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \mu\right| < \delta\right\} = 1 \tag{42.1}$$

where $S_n = X_1 + X_2 + \cdots + X_n$.

Remark 42.2

1. *The weak law of large numbers states that for large n, the bulk of the distribution of $\dfrac{S_n}{n}$ is concentrated near μ . That is, if we consider a positive length interval $[\mu-\delta, \mu+\delta]$ around μ, then there is high probability that S_n/n will fall in that interval; as $n \to \infty$, this probability converges to 1 . Of course, if δ is very small, we may have to wait longer (i.e. , need a larger value of n) before we can assert that S_n/n is highly likely to fall in that interval.*

2. *To understand the convergence in weak law, think in terms of PMF (if X_i are discrete random variables) or PDF (if X_i's have the pdf then we know that S_n will posses a pdf) of random variable S_n/n. Weak law states that "almost all" of the PMF or PDF of S_n/n is concentrated within δ neighborhood of μ for large values of n.*

3. *The limit in (42.2) means: $\forall\ \delta, \epsilon > 0$, there exists $n_0(\epsilon, \delta)$ such that for all $n \geq n_0(\epsilon, \delta)$ we have*

$$P\left\{\omega : \left|\frac{S_n}{n} - \mu\right| < \delta\right\} > 1 - \epsilon.$$

If we refer to δ as the accuracy level and ϵ as the confidence level, the weak law takes the following intuitive form: for any given level of accuracy and confidence, S_n/n will be equal to μ, within these levels of accuracy and confidence, provided n is large enough.

Example 42.3 *Let $X_1, X_2, \cdots$ be independent and identically distributed random variables with $E[X_i] = 0$ and $Var(X_i) = 1$ for all i. Let $S_n = X_1 + X_2 + \cdots + X_n$. Then, for any $x > 0$, compute $\lim_{n\to\infty} P\left(-nx < S_n < nx\right)$.*

Solution: For any $x > 0$, we have

$$P(-nx < S_n < nx) = P\left(-x < \frac{S_n}{n} < x\right) = P\left(\left|\frac{S_n}{n} - 0\right| < x\right)$$
$$= 1 - P\left(\left|\frac{S_n}{n} - 0\right| \geq x\right)$$

By weak law of large numbers, we have

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - 0\right| \geq x\right) = 0$$

■

Random Sampling

Let $X_1, \cdots, X_n$ be n independent random variables having the same distribution. These random variables may be thought of as n independent measurements of some quantity that is distributed according to their common distribution (e.g., height of students in LNMIIT campus). In this sense we sometimes speak of the random variables $X_1, \cdots, X_n$ as constituting a random sample of size n from this distribution.

Suppose that the common distribution of these random variables has finite mean μ. Then for n sufficiently large we would expect that the sample mean $\frac{S_n}{n} = (X_1 + \cdots + X_n)/n$ should be close to ture mean μ.

The weak law of large numbers asserts that the sample mean of a large number of independent identically distributed random variables is very close to the true mean, with high probability.

Weak law of large numbers says that for any $\delta > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \mu\right| \geq \delta\right\} = 0. \tag{42.2}$$

We may interpret (42.2) in the following way. The number δ can be thought of as the desired accuracy in the approximation of μ by S_n/n. Equation (42.2) assures us that no matter how small δ may be chosen the probability that S_n/n approximates μ to within this accuracy, that is, $P\left\{\left|\frac{S_n}{n} - \mu\right| < \delta\right\}$, converges to 1 as the number of observations gets large.

Example 5.4. Probabilities and Frequencies. Consider an event A defined in the context of some probabilistic experiment. Let $p = \mathbf{P}(A)$ be the probability of this event. We consider n independent repetitions of the experiment, and let M_n be the fraction of time that event A occurs; in this context, M_n is often called the **empirical frequency** of A. Note that

$$M_n = \frac{X_1 + \cdots + X_n}{n},$$

where X_i is 1 whenever A occurs, and 0 otherwise; in particular, $\mathbf{E}[X_i] = p$. The weak law applies and shows that when n is large, the empirical frequency is most likely to be within ϵ of p. Loosely speaking, this allows us to conclude that empirical frequencies are faithful estimates of p. Alternatively, this is a step towards interpreting the probability p as the frequency of occurrence of A.

Theorem 42.4 (**Strong law of large numbers**) *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having finite mean μ. Then*

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1, \tag{42.3}$$

where $S_n = X_1 + X_2 + \cdots + X_n$.

Lecture 43: Central Limit Theorem

12 April, 2019

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Theorem 43.1 (Central Limit Theorem) *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having finite mean μ and non-zero variance σ^2. Define*

$$S_n := X_1 + X_2 + \cdots + X_n, \quad Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Then

$$\lim_{n\to\infty} P(Z_n \leq x) = N(x), \ \forall x \in \mathbb{R},$$

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

Another frequently used notation for $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ is $\Phi(x)$.

The central limit theorem is surprisingly general. Besides independence, and the implicit assumption that the mean and variance are finite, it places no other requirement on the distribution of the X_i, which could be discrete, continuous, or mixed.

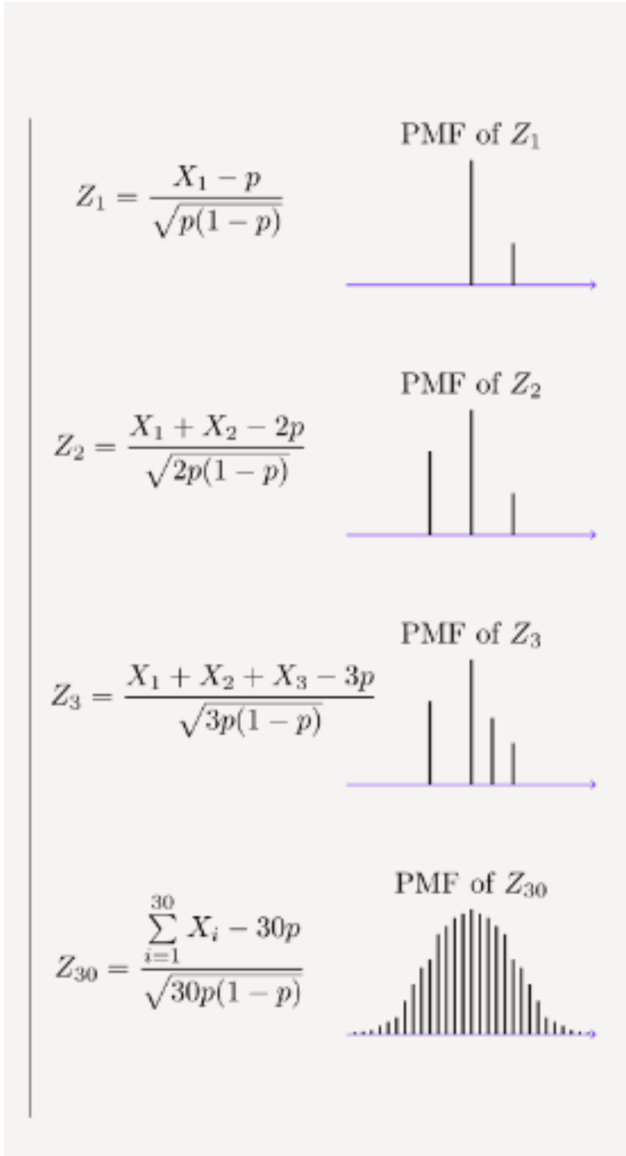
Although CLT gives us a useful general approximation, we have no automatic way of knowing how good the approximation is in general. In fact the goodness of the approximation is a function of the original distribution, and so must be checked case by case.

To get a feeling for the CLT, let us look at some examples.

Example 43.2 *Let X_i's be independent Bernoulli(p). Then $EX_i = p, Var(X_i) = p(1-p)$. Also, $S_n = X_1 + X_2 + \cdots + X_n$ has Binomial(n,p) distribution. Thus,*

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

We plot the PMF of Z_n for different values of n by choosing $p = \frac{1}{3}$.



As you see, the shape of the PMF gets closer to a normal PDF curve as n increases. Hence, the CDF of Z_n will converges to the standard normal CDF.

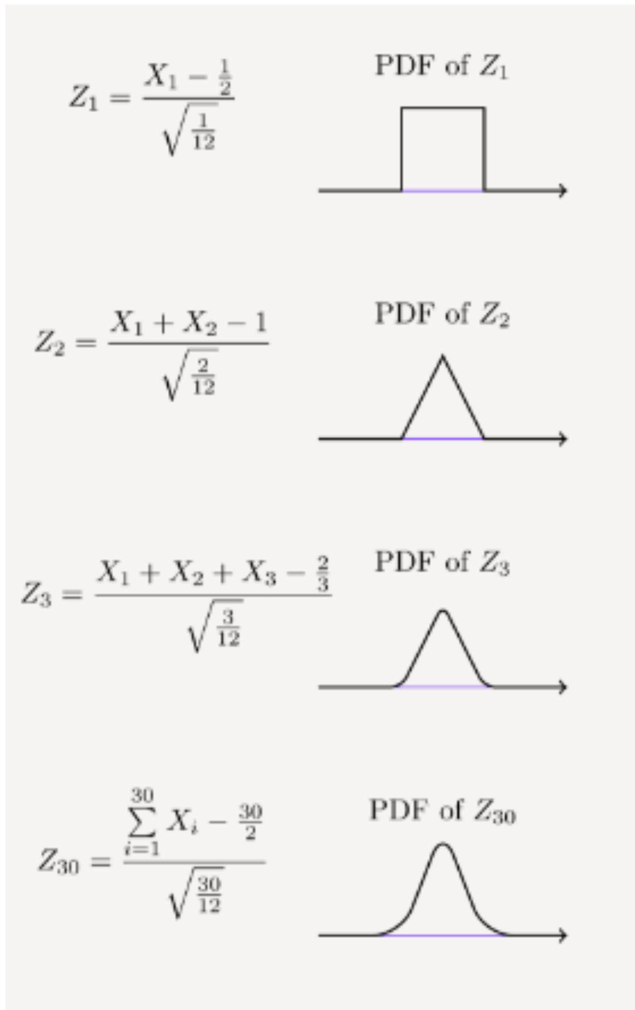
$$F_{Z_n}(x) = \sum_{z \in R_{Z_n}: z \leq x} f_{Z_n}(z) \to N(x) \text{ or } \Phi(x)$$

That what the CLT states.

Example 43.3 *Let X_i's be independent Uniform$(0,1)$. Then $EX_i = \frac{1}{2}, Var(X_i) = \frac{1}{12}$. Let $S_n = X_1 + X_2 + \cdots + X_n$. In this case,*

$$Z_n = \frac{S_n - \frac{n}{2}}{\sqrt{n/12}}.$$

We have derived a general formula for the pdf of sum of two independent random variables with pdf's. Hence random variable Z_n has pdf. We plot the PDF of Z_n for different values of n



As you see, the shape of the PDF gets closer to a normal PDF curve as n increases. Hence, the CDF of Z_n will converges to the standard normal CDF.

$$F_{Z_n}(x) = \int_{-\infty}^{x} f_{Z_n}(z)dz \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz.$$

That what the CLT states.

Question: Suppose (X_n) is sequence of iid discrete (continuous) random variables with finite mean and non-zero variance. Then CLT says that pmf (pdf) of Z_n converges to pdf of standard normal random variable ?

Answer: No. CLT does not says that pmf (pdf) of Z_n converges to pdf of standard normal random variable. CLT is about convergence of distribution functions. Suppose (X_i) are continuous with pdf f. Then S_n has pdf $(*f)^n$. CLT says that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} \int_{-\infty}^{\frac{x-np}{\sigma\sqrt{n}}} (*f)^n(t)dt = \int_{-\infty}^{x} \frac{1}{2\pi} e^{-\frac{t^2}{2}} dt$$

Normal Approximation Based on the Central Limit Theorem: The central limit theorem allows us to calculate probabilities related to Z_n as if Z_n were normal, CLT says $P(Z_n \leq x) \approx N(x)$ for large values of n. Note that $S_n = \sigma\sqrt{n}Z_n + n\mu$. Since normality is preserved under linear transformations, Sicne $Z_n \sim N(0, 1)$, this is equivalent to treating S_n as a normal random variable with mean $n\mu$ and variance $n\sigma^2$.

Let $S_n = X_1 + \cdots + X_n$, where the X_i are independent identically distributed random variables with mean μ and variance σ^2. If n is large, the probability $P(S_n \leq c)$ can be approximated by treating S_n as if it were normal, according to the following procedure.

Step 1 Calculate the mean $n\mu$ and the variance $n\sigma^2$ of S_n.

Step 2 Use the approximation

$$P(S_n \leq c) \approx N\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right)$$

Example 43.4 *We load on a plane* 100 *packages whose weights are independent random variables that are uniformly distributed between* 5 *and* 50 *kg. What is the probability that the total weight will exceed 3000 kg?*

Solution: Let us translate the problem in probabilistic model. Let X_i denotes the weight of ith packages. $X_1, X_2, \cdots, X_{100}$ are iid uniform random variables with density

$$f(x) = \begin{cases} \frac{1}{45}, & \text{if } 5 \leq x \leq 50 \\ 0, & \text{otherwsie} \end{cases}$$

Let $S = X_1 + X_2 + \cdots + X_{100}$ denote the total weight. Then question is to calculate the $P(S > 3000)$. Let f, g, and h be functions on the reals, and suppose the convolutions $(f * g) * h$ and $f * (g * h)$ exist. Then we have $(f * g) * h = f * (g * h)$. Using this result and S is sum of independent random variable we can see that pdf of S is 100-fold convolution of f. It is not easy to calculate this one. Hence it is very difficult to find the desired probability, but an approximate answer can be quickly obtained using the central limit theorem. Treat

S as normal random variable. So now we find it's mean and variance, which is 100μ and $100\sigma^2$ where $\mu = E[X_i], \sigma^2 = var(X_i)$.

$$\begin{aligned}
E(X_i) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \frac{1}{45}\int_{5}^{50} x dx \\
&= \frac{5+50}{2} = 27.5 \\
E(X_i^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\
&= \frac{1}{45}\int_{5}^{50} x^2 dx \\
&= \frac{(50)^3 - 5^3}{3 \times 45} \\
&= \frac{(50)^2 + 50 \times 5 + 5^2}{3} \\
&= 925 \\
var(X_i) &= 925 - (27.5)^2 = 925 - 756.25 = 168.75
\end{aligned}$$

Now

$$P(S > 3000) = 1 - P(S \leq 3000) = 1 - N\left(\frac{3000 - 2750}{10\sqrt{168.75}}\right) = 1 - N(1.92)$$

■

Example 43.5 *Let $X_1, X_2, \cdots$ be iid Poisson(λ) RVs. Then we know $EX_1 = var(X_1) = \lambda$. Hence by CLT, $S_n = X_1 + X_2 + \cdots + X_n$ has approximately an $N(n\lambda, n\lambda)$ distribution for large n. Let $n = 64$, $\lambda = 0.125 \implies n\lambda = 8$. Also sum of independent Poisson is again Poisson hence exact distribution of $S_{64} \sim$ Poisson$(64 \times 0.125 = 8)$ and from Poisson distribution tables $P(S_{64} = 10) = \frac{(8)^{10} e^{-8}}{10!} = 0.099261534$. Using normal approximation*

$$\begin{aligned}
P(S_n = 10) = P(9.5 < S_n < 10.5) &= P\left(\frac{9.5 - 8}{\sqrt{0.125 \times 8}} < \frac{S_n - n\lambda}{\lambda\sqrt{n}} < \frac{10.5 - 8}{\sqrt{0.125 \times 8}}\right) \\
&= P(0.530330086 < Z < 0.883883476) = 0.1087.
\end{aligned}$$

Here we have used the "continuity correction" to compute the $P(S_n = 10)$ which would be zero if S_n is taken to be normal random variable.