Random Variables

Lecture Notes of PRP offered by Dr. Prasad Krishnan, IIIT Hyderabad
prepared by Hari Hara Suthan and Sai Praneeth.Ch (corrections, mail to prasad.krishnan@iiit.ac.in)

Monsoon 2018

Class 6. (17/08/18)

Note. *Reader is recommended to prove the statements in Exercise, Lemmas and Theorems.*

1 Random Variables

Consider a probability space corresponding to some random experiment. In many occasions, we may be interested only in events of a particular kind, but not in all events. For instance, consider the sample space of a random experiment in which we switch on a digital source (one which generates 0s and 1s) and obtain 10000 samples. Suppose we are not interested in the precise 10000 samples, but only in the number of 1s we obtain. Thus, considering a probability space over the power set of the sample space containing all the 2^{10000} possible outcomes is overkill, given that we only need to track subsets of outcomes indicating the 10001 different possible number of 1s in the outcome.

For such situations, the notion of a *random variable* is quite useful, as we shall see. Furthermore, we shall also see that we will be able model a variety of random quantities of interest in different disciplines using the idea of random variables, and study them via the well-developed tools of mathematics applicable to the real-number system. We therefore begin our formal study of real-valued random variables with this understanding.

Definition 1.1. *Consider a probability space $(\Omega, \mathcal{F}, P)$. A function $X : \Omega \to \mathbb{R}$ is said to be a (real-valued) random variable if $X^{-1}\big((-\infty, x]\big) \in \mathcal{F}, \forall x \in \mathbb{R}$, where for any $A \subset \mathbb{R}$, $X^{-1}(A)$ is defined as follows*

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\},$$

i.e., $X^{-1}(A)$ is the set of all pre-images of all the individual elements of A.

Equivalently, a random variable X (w.r.t Ω) is also defined as any *measurable* function from $\Omega \to \mathbb{R}$, where the word *measurable* means (roughly) that the preimages of an event-space of $\mathbb{R}$ should be an event space in Ω (this is just a rough definition, unless we want to go deeper into mathematics to formalize this). The formal definition above is more accurate.

Following definition 1.1, we develop the following 'short-hand' for referring to certain events of interest. For any $x \in \mathbb{R}$, let us define

$$(X \leq x) \triangleq \{\omega \in \Omega : X(\omega) \leq x\},$$
$$(X = x) \triangleq \{\omega \in \Omega : X(\omega) = x\},$$

and so on. In other words, the notation $P(X \leq x)$ actually means $P\{\omega \in \Omega : X(\omega) \leq x\}$, which is well-defined according to the probability measure P defined on the event space $\mathcal{F}$.

Remark. *Non-real-valued random variable X can also be defined similarly. For instance, a random variable whose codomain is $\mathbb{C}$ is a complex-value random variable, and so on. However, that the function should necessarily be measurable. We mostly deal with real valued random variables in this course.*

Example 1.1 (Example for random variable and non random variable)**.** *Let $\Omega = \{a, b, c\}$ be the sample space and $\mathcal{F} = \{\phi, \{a\}, \{b, c\}, \Omega\}$ be the event space. Define a function $X : \Omega \to \mathbb{R}$ such that*

$$X(\omega) = \begin{cases} 0 & \omega = a \\ 1 & \omega = b \text{ or } \omega = c \end{cases}$$

It is easy to see that

$$X^{-1}((-\infty, x]) = \begin{cases} \phi & x < 0 \\ \{a\} & 0 \leq x < 1 \\ \Omega & 1 \leq x < \infty \end{cases}$$

Here $X^{-1}\big((-\infty,x]\big) \in \mathcal{F}, \forall x \in \mathbb{R}$. Hence X is a random variable with respect to the given event space $\mathcal{F}$.

Let's define another function $Y : \Omega \to \mathbb{R}$ such that

$$Y(\omega) = \begin{cases} 0 & \omega = b \\ 1 & \omega = a \text{ or } \omega = c \end{cases}$$

It is easy to see that

$$Y^{-1}((-\infty,y]) = \begin{cases} \phi & y < 0 \\ \{b\} & 0 \le y < 1 \\ \Omega & 1 \le y < \infty \end{cases}$$

Here $Y^{-1}((-\infty,0.5]) = \{b\} \notin \mathcal{F}$. Hence Y is not a random variable with respect to the given event space $\mathcal{F}$.

Note. *Here both X, Y are random variables with respect to the event space $\mathcal{F}_1 =$ power set (Ω).*

Remark. *Any real valued function will be a random variable with respect to the largest event space $\big($power set$(\Omega)\big)$. (Why? - Try to answer, dear reader!)*

The following theorem classifies the reason why it is sufficient to consider only intervals of the form $(-\infty,x], \forall x \in \mathbb{R}$ in the definition of random variable.

Theorem 1.1. *Consider a probability space $(\Omega,\mathcal{F},P)$. Let X be the random variable defined on Ω. If $(X \le x) \in \mathcal{F}, \forall x \in \mathbb{R}$, then $\forall a, b \in \mathbb{R}$ such that $a < b$*

1. $(X < a) \in \mathcal{F}$.
2. $(X = a) = (X \le a) \setminus (X < a) \in \mathcal{F}$.
3. $(X \in (a,b]) = (X \le b) \setminus (X \le a) \in \mathcal{F}$.
4. $(X \in [a,b]) = (X \le b) \setminus (X < a) \in \mathcal{F}$.
5. $(X \in [a,b)) = (X < b) \setminus (X < a) \in \mathcal{F}$.
6. $(X \in (a,b)) = (X < b) \setminus (X \le a) \in \mathcal{F}$.

Note. *From axioms of event space we have that, if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$. If $A, B \in \mathcal{F}$ then $A \cup B, A \cap B \in \mathcal{F}$. We can write $A \setminus B$ as $A \cap B^c$.*

Proof. Consider two events $(X \le a)$ and $(X \le b)$ in the event space $\mathcal{F}$.

1. Let $a_i = a - \dfrac{1}{i}, \forall i \in \{1,2,\cdots\}$, then $(X \le a_i) \in \mathcal{F}, \forall i \in \{1,2,\cdots\}$. From axioms of event space we have $\bigcup_{i=1}^{\infty}(X \le a_i) \in \mathcal{F}$. It is clear that $\bigcup_{i=1}^{\infty}(X \le a_i) = (X < a)$. Therefore $(X < a) \in \mathcal{F}$. Simillarly $(X < b) \in \mathcal{F}$.
2. Since $(X \le a), (X < a) \in \mathcal{F}$, we can write $(X = a) \in \mathcal{F}$.
3. Since $(X \le a), (X \le b) \in \mathcal{F}$, we can write $(X \in (a,b]) \in \mathcal{F}$.
4. Since $(X < a), (X \le b) \in \mathcal{F}$, we can write $(X \in [a,b]) \in \mathcal{F}$.
5. Since $(X < a), (X < b) \in \mathcal{F}$, we can write $(X \in [a,b)) \in \mathcal{F}$.
6. Since $(X \le a), (X < b) \in \mathcal{F}$, we can write $(X \in (a,b)) \in \mathcal{F}$.

□

Lemma 1.2. *Consider a probability space $(\Omega,\mathcal{F},P)$. Let X be the random variable defined on Ω. Let $(X \le x) \in \mathcal{F}, \forall x \in \mathbb{R}$. Prove that it is sufficient to know $P(X \le x), \forall x \in \mathbb{R}$ to calculate probability of any interval in $\mathbb{R}$.*

Proof. From the axioms of probability measure we have $P(A\dot\cup B) = P(A) + P(B)$, where $A, B \in \mathcal{F}$ and $\dot\cup$ denotes the disjoint union. Here we have $P(X \le x), \forall x \in \mathbb{R}$.

1. $P(X < a) = P\left(\dot\bigcup_{i=1}^{\infty}\left(X \le a - \dfrac{1}{i}\right)\right) = \lim_{i\to\infty} P\left(X \le a - \dfrac{1}{i}\right)$. (from Continuity of probability theorem).
2. $P(X \le a) = P((X < a)\dot\cup(X = a)) = P(X < a) + P(X = a)$. Hence $P(X = a) = P(X \le a) - P(X < a)$.

3. $P(X \leq b) = P\left((X \leq a)\dot{\cup}(X \in (a,b])\right) = P(X \leq a) + P(X \in (a,b])$. Hence $P(X \in (a,b]) = P(X \leq b) - P(X \leq a)$.

4. $P(X \leq b) = P\left((X < a)\dot{\cup}(X \in [a,b])\right) = P(X < a) + P(X \in [a,b])$. Hence $P(X \in [a,b]) = P(X \leq b) - P(X < a)$.

5. $P(X < b) = P\left((X < a)\dot{\cup}(X \in [a,b))\right) = P(X < a) + P(X \in [a,b))$. Hence $P(X \in [a,b)) = P(X < b) - P(X < a)$.

6. $P(X < b) = P\left((X \leq a)\dot{\cup}(X \in (a,b))\right) = P(X \leq a) + P(X \in (a,b))$. Hence $P(X \in (a,b)) = P(X < b) - P(X \leq a)$.

Therefore it is sufficient to know $P(X \leq x), \forall x \in \mathbb{R}$ to calculate probability of any interval in $\mathbb{R}$. $\square$

Corollary 1.2.1. *With respect to the Borel $\sigma-$ algebra of $\mathbb{R}$ to ensure pre-images of all intervals of $\mathbb{R}$ as events in the probability space $(\Omega, \mathcal{F}, P)$ and to calculate there respective probabilities it is sufficient*

- *to ensure $(X \leq x) \in \mathcal{F}, \forall x \in \mathbb{R}$*
- *to know $P(X \leq x), \forall x \in \mathbb{R}$.*

Class 7. (24/08/18)

2 Cumulative Distribution Function(CDF)(F_X)

From the above corollary it is sufficient to know $P(X \leq x)$ to calculate all relevant properties. Hence we give a special name to this function as Cumulative Distribution Function as it finds cumulative(collective) probability of of all the events upto a particular point x. Even though the random experiment, $\Omega, \mathcal{F}, P$ are unknown we can completely characterize X by just knowing its CDF

Definition 2.1. *Consider a probability space $(\Omega, \mathcal{F}, P)$. Let X be a real valued random variable. A function $F_X : \mathbb{R} \to \mathbb{R}$ is said to be a Cumulative Distribution Function(CDF) of the random variable X if $F_X(x) = P(X \leq x)$.*

2.1 Properties of the CDF of a RV

Theorem 2.1 (Properties of CDF)**.** *The CDF $F_X(x) = P(X \leq x)$ of the random variable X satisfies the following properties*

1. *F_X is a non-decresing funciton of $x \in \mathbb{R}$.*
2. *$\lim_{x \to -\infty} F_X(x) = 0$.*
3. *$\lim_{x \to \infty} F_X(x) = 1$.*
4. *F_X is a right continuous function.*

Proof. 1. Let $x_1, x_2 \in \mathbb{R}$ such that $x_1 < x_2$. Consider $F_X(x_2) = P(X \leq x_2) = P((X \leq x_1)\dot{\cup}(X \in (x_1, x_2])) = P((X \leq x_1) + P(X \in (x_1, x_2]) = F_X(x_1) + P(X \in (x_1, x_2])$, Since $P(X \in (x_1, x_2]) \geq 0$ we can write $F_X(x_2) \geq F_X(x_1)$. Therefore If $x_1 < x_2$ then $F_X(x_1) \leq F_X(x_2)$. Hence F_X is a non-decreasing function of $x \in \mathbb{R}$.

2. Define $B_i = (X \leq i), \forall i \in \mathbb{Z}$. Then we can write $B_{-1} \supset B_{-2} \supset \cdots$. It is clear that $\left(\bigcap_{i=1}^{\infty} B_{-i}\right) = \phi$. By continuity of probability theorem we have $\lim_{n\to\infty} P(B_{-n}) = P\left(\bigcap_{i=1}^{\infty} B_{-i}\right) = P(\phi) = 0$. Therefore, we can write $\lim_{n\to\infty} P(B_{-n}) = \lim_{n\to\infty} P(X \leq -n) = \lim_{n\to-\infty} P(X \leq n) = \lim_{x\to-\infty} P(X \leq x)$. Now, since $P(X \leq x) = F_X(x)$. therefore we have proved $\lim_{x\to-\infty} F_X(x) = 0$.

3. Define $B_i = (X \leq i), \forall i \in \mathbb{Z}$. Then we can write $B_1 \subset B_2 \subset \cdots$. It is clear that $\left(\bigcup_{i=1}^{\infty} B_{-i}\right) = \Omega$. By continuity of probability theorem we have $\lim_{n\to\infty} P(B_n) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = P(\Omega) = 1$. We can write $\lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} P(X \leq n) = \lim_{x\to\infty} P(X \leq x)$. Using the fact that $P(X \leq x) = F_X(x)$, therefore we have proved $\lim_{x\to\infty} F_X(x) = 1$.(Here $n \to \infty$ is a countable sequence and $x \to \infty$ is a uncountable sequence).

4. Let $B_i = \left(X \leq x + \frac{1}{i}\right), i \in \mathbb{N}$. Then $B_1 \supset B_2 \supset \cdots$. It is clear that $\bigcap_{i=1}^{\infty} B_i = (X \leq x)$. By using continuity of probability we can write $F_X(x) = P(X \leq x) = P\left(\bigcap_{i=1}^{\infty} B_i\right) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} P\left(X \leq x + \frac{1}{n}\right) = \lim_{\epsilon\to 0} P(X \leq x + \epsilon) = \lim_{\epsilon\to 0} F_X(x+\epsilon)$. Therefore $F_X(x) = \lim_{\epsilon\to 0} F_X(x+\epsilon)$. Hence F_X is a right continuous function.

□

Why should we specifically claim only right-continuity? We now show that left continuity does not automatically hold, using the following lemma and corollary.

Lemma 2.2. *For any $x \in \mathbb{R}$,*

$$P(X = x) = F_X(x) - P(X < x) = F_X(x) - \lim_{\epsilon\to 0} F_X(x-\epsilon)$$

Proof. The first inequality in the statement follows because $(X = x)\dot{\cup}(X < x) = (X \leq x)$. Thus we prove the second equality only.

Let $B_i = \left(X \leq x - \frac{1}{i}\right), i \in \mathbb{N}$. Then $B_1 \subset B_2 \subset \cdots$. It is clear that $\bigcup_{i=1}^{\infty} B_i \neq (X \leq x)$, since the event $(X = x)$ is not included in any of the B_is. On the other hand, $\bigcup_{i=1}^{\infty} B_i = (X < x)$, as for any small $\epsilon > 0$, the event $(X \leq x - \epsilon)$ is included in the event $(X \leq x - \frac{1}{i})$ for any $i \geq \frac{1}{\epsilon}$.

Thus, by using the theorem of continuity of probability, we can write

$$P(X < x) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} P\left(X \leq x - \frac{1}{n}\right) = \lim_{\epsilon\to 0} P(X \leq x - \epsilon) = \lim_{\epsilon\to 0} F_X(x-\epsilon).$$

Therefore $P(X < x) = \lim_{\epsilon\to 0} F_X(x-\epsilon)$. Hence F_X need not be a left continuous function unless $P(X$.

□

By the above lemma, we thus have the following corollary.

Corollary 2.2.1. *The CDF $F_X(x)$ is continuous (i.e., left-continuous) if and only if $P(X = x) = 0, \forall x \in \mathbb{R}$.*

Proof. $F_X(x)$ is already right-continuous. For left continuity, we need that

$$F_X(x) = \lim_{\epsilon\to 0} F_X(x-\epsilon),$$

where $P(X < x) = \lim_{\epsilon\to 0} F_X(x-\epsilon)$. The proof follows from Lemma 2.2. □

Indeed, in the next section, we will study random variables which are *continuous* and others which are discrete (having non-continuous CDFs).

2.2 Why we can study RVs through their CDFs alone?

The following theorem shows that *any* function which satisfies the CDF properties, is infact a CDF of *some* random variable.

Theorem 2.3. *Let $F : \mathbb{R} \to \mathbb{R}$ be any function satisfying*

1. *For $a \leq b$, $a, b \in \mathbb{R}, F_X(a) \leq F_X(b)$.*
2. *$\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.*
3. *$\lim_{\epsilon\to 0} F_X(x+\epsilon) = 0$. (right continuous).*

then there exists some random variable X such that $F_X = F$.

Proof. proof is beyond the scope of this course. □

Remark. *Theorem 2.3 is a special case of the so-called Kolmogorov's extension theorem. It is important to mention this here, because of the following reason. Since we can capture the essential probabilistic structure in the random variable by its cdf, and hence discussing about functions which satisfy the properties as in Theorem 2.3, we are indirectly discussing random variables themselves.*

Class 8. (28/08/18)

3 Types of random variables

As with all things under the sun, we like to classify random variables as well. This presentation of the classification is based on the structure of the CDF. However, many books do a simpler classification based on the values taken by the random variable (we mention this in parantheses).

3.1 Continuous Random Variables

Definition 3.1. *A random variable X is said to be a* ***continuous type random variable*** *if $F_X(x)$ is a continuous function of x (roughly speaking, X can take values in an uncountable set).*

We mostly consider continuous random variables whose F_X is also differentiable, in which case $\frac{dF_X(x)}{dx}$ is well defined. Therefore we give it a special name called the **probability density function(pdf)**, and a special notation $f_X(x)$. Thus we have the pdf of a continuous random variable X as

$$f_X(x) \triangleq \frac{dF_X(x)}{dx}.$$

We can show without much difficulty that $f_X(x)$ satisfies the following properties.

- $f_X(x) \geq 0, \forall x.$
- $\int\limits_{-\infty}^{\infty} f_X(x)dx = 1$ (this property means that the integral can be computed and the value is equal to 1).

And indeed, as we would expect, similar to Theorem 2.3, if we have a function $f(x)$ satisfying the above two properties, then there exists some random variable X such that $f(x)$ is the pdf of the random variable X. Therefore, more often than not, we rely upon the pdf $f_X(x)$ to describe the continuous random variable X.

By definition, we know how to obtain $f_X(x)$ if we know the CDF $F_X(x)$. Now, suppose we the pdf $f_X(x)$. Then the following expressions naturally follow by definition of $f_X(x)$.

We can further obtain the following properties for a continuous random variable X.

1. $P(X = x) = 0, \forall x \in \mathbb{R}$. (From Lemma 2.2.1 we have, $F_X(x)$ is continuous iff $P(X = x) = 0, \forall x \in \mathbb{R}$).
2. For any $a \in \mathbb{R}, F_X(a) = \int\limits_{-\infty}^{a} f_X(x)dx.$
3. $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a) = \int\limits_{-\infty}^{b} f_X(x)dx - \int\limits_{-\infty}^{a} f_X(x)dx = \int\limits_{a}^{b} f_X(x)dx.$
4. $P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = \int\limits_{a}^{b} f_X(x)dx.$ (by the first property).
5. $P(x < X \leq x+\Delta x) \approx f_X(x)\Delta x$, where Δx is infinitesimally small (This is simply by definition of the integral (in the Reimann sense)).

Remark. *With respect to point (1) above, note that $P(event) = 0$, doesn't necessarily mean that the event never happens. For instance, consider a random experiment of selecting a point in the interval $[0.1]$ of real numbers and thus $\Omega = [0, 1]$. The performance of the experiment ensures some outcome in $[0, 1]$. Consider the uniform probability measure on $[0, 1]$. Hence $P(any\ outcome) = 0$, however we do get some outcomes.*

Note (Very Important). *The pdf at $x \in \mathbb{R}$, $f_X(x)$, is* ***not*** *equal to $P(X = x)$. The pdf $f_X(x)$ is rather like the rate of change of probability of the random variable X with x. That is why if we need the probability of X being in some interval, we have to integrate the pdf $f_X(x)$.*

3.2 Discrete Random Variables

We now come to the notion of discrete random variables.

Definition 3.2. *A random variable X is said to be a* ***discrete type random variable*** *if $F_X(x)$ is a step function (or equivalently, a staircase function) (roughly speaking, X can take a finite or a countably infinite number of values).*

Since $F_X(x)$ is in the shape of a staircase function, it takes countably many steps (or discontinues), and is flat elsewhere. This is the reason we say that X takes only a countable set of values. Let us call these values as $\{x_i : i \in S\}$, where S is some countable set.

Since X takes values only in $\{x_i : i \in S\}$, we have $P(X = x) = 0, \forall x \notin \{x_i : i \in S\}$. What about $P(X = x_i)$? Clearly, we know from Lemma 2.2 that $P(X = x_i) = F_X(x_i) - P(X < x_i)$. Thus, it is only for some $x_i \in \{x_i : i \in S\}$ that we can have discontinuities in $F_X(x)$, i.e., $P(X = x_i) \geq 0$.

These values of $P(X = x_i)$ are thus special for us, and at least some of them are non-zero, and hence we have to define a *probability mass function* to capture these.

Definition 3.3 (Probability mass function (PMF))**.** *Consider a discrete type random variable X taking values in $\{x_i : i \in S\}$ for some countable set S. The* ***probability mass function*** *of X is denoted by $P_X : \mathbb{R} \to \mathbb{R}$ and is defined as $P_X(x_i) \triangleq P(X = x_i)$ and $P_X(x) = 0, \forall x \notin \{x_i : i \in S\}$.*

Clearly, the PMF satisfies the following properties, and indeed any function which satisfies the below properties can serve as the PMF of some discrete random variable X which takes values from a countable set of values $\{x_i : i \in S\}$ (where S is a countable set)

- $P_X(x_i) \geq 0, \forall x_i$.
- $\sum\limits_{i \in S} P_X(x_i) = 1$.

The following properties of the PMF of a discrete random variable X and its CDF can be easily proved from the definition.

1. For any $a \in \mathbb{R}$,
$$F_X(a) = P(X \leq a) = \sum_{x_i \leq a} P(X = x_i) = \sum_{x_i \leq a} P_X(x_i).$$
2. $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$.
3. $P(a \leq X \leq b) = F_X(b) - F_X(a) + P(X = a)$.
4. $P(a < X < b) = F_X(b) - F_X(a) - P(X = b)$.
5. $P(a \leq X < b) = F_X(b) - F_X(a) + P(X = a) - P(X = b)$.

3.3 Mixed Random Variable

Definition 3.4. *A random variable X is said to be* ***mixed type random variable*** *if $F_X(x)$ has both countable number of discontinuities and is also continuous in some intervals.*

We will not pursue Mixed Random Variables much in this course, but essentially the techniques applied to both discrete and continuous random variables apply in this case.

Class 9. (31/08/18) In many situations, we may have some function of a random variable, and then we are interested in studying the probabilities of values taken by such a function. Hence we give the following definition.

4 Measurable function of random variable

Lemma 4.1 (A (measurable) function of random variable is a random variable)**.** *Let $X : \Omega \to \mathbb{R}$ be a random variable. Let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function (i.e., a function such that preimages of event spaces in $\mathbb{R}$ gives an event space of $\mathbb{R}$). Consider the composition $g(X)$ denoted by Y, which is a function from $\Omega \to \mathbb{R}$ defined as $Y(\omega) = g(X(\omega))$. Then Y is a random variable.*

We will not prove this lemma, since we are going to study only measurable functions g here. However the proof is reasonably straightforward.

Note. *Note that we can break-down Y as $Y : \Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$. Hence $\Omega \xrightarrow{Y = g \circ X} \mathbb{R}$.*

4.1 The CDF of $Y = g(X)$

To capture the behaviour of Y, we look at the CDF of Y. Clearly, since $Y = g(X)$, it is clear that the CDF of Y depends on that of X. Hence, for any $y \in \mathbb{R}$, we have the following expressions.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P\left(\{\omega \in \Omega : Y(\omega) \leq y\}\right) \\ &= P\left(\{\omega \in \Omega : g(X(\omega)) \leq y\}\right) \\ &= P(g(X) \leq y) \end{aligned}$$

Example 4.1. *Let* $Y = g(X) = X^2$. *Then*

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P\left(\{\omega \in \Omega : Y(\omega) \leq y\}\right) \\ &= P\left(\{\omega \in \Omega : (X(\omega))^2 \leq y\}\right) \\ &= P\left(\{\omega \in \Omega : -\sqrt{y} \leq X(\omega) \leq \sqrt{y}\}\right) \\ &= P(X \in [-\sqrt{y}, \sqrt{y}]) \\ &= P(X \leq \sqrt{y}) - P(X < -\sqrt{y}) \\ &= F_X(\sqrt{y}) - \lim_{\epsilon \to 0} F_X(-\sqrt{y} - \epsilon). \end{aligned}$$

Note that the above method is a general method for understanding the distribution of the random variable Y. However, in the special cases of discrete and continuous random variables, we may **also** be interested in focusing on the PMF or pdf directly, rather than computing the CDF first. This is what we do now. We consider three cases.

Discrete-to-Discrete: If X is discrete type random variable then $Y = g(X)$ is also a discrete type random variable irrespective of g.

Since X is a discrete type random variable, it takes values in some countable set, say $\{x_i : i \in S\}$ (for some countable set S). Thus $Y = g(X)$ must also be a discrete random variable. Hence, we can focus on the PMF of Y.

$$P_Y(y) = P(Y = y) = P(\{x_i : g(x_i) = y\}) = \sum_{x_i : g(x_i) = y} P(X = x_i) = \sum_{x_i : g(x_i) = y} P_X(x_i).$$

Continuous-to-Discrete: If X is a continuous type random variable and $g : \mathbb{R} \to \mathbb{R}$ is a discrete measurable function then $Y = g(X)$ is a discrete type random variable.

Since X is a continuous type random variable we consider its pdf $f_X(x)$. Since g is a discrete measurable function so $Y = g(X)$ is a discrete type random variable. The pmf of Y is given by

$$\begin{aligned} P_Y(y) &= P(Y = y) = P(g(x) = y) = P(\{X = x : g(x) = y\}) \\ &= P(\{\omega \in \Omega : X(\omega) = x \text{ and } g(x) = y\}) = \int_{\text{all intervals } : g(x) = y} f_X(x)dx. \end{aligned}$$

Class 10. (04/09/18)

Continuous-to-Continuous:If X is a continuous type random variable and $g : \mathbb{R} \to \mathbb{R}$ is a continuous measurable function then $Y = g(X)$ is a continuous type random variable.

Since X is a continuous type random variable we consider its pdf $f_X(x)$. Since g is a continuous measurable function so $Y = g(X)$ is a continuous type random variable. The pdf of Y is given by

$$f_Y(y) = P(y < Y \leq \Delta y) = \int_y^{y+\Delta y} f_Y(y)dy \approx f_Y(y)\Delta y. \tag{1}$$

$$P(y \leq Y \leq y + \Delta y) = P\left(\dot{\bigcup_i}\big(X \in [x_i, x_i + \Delta x_i]\big)\right) = \sum_i P\big(X \in [x_i, x_i + \Delta x_i]\big) \approx \sum_i f_X(x_i)|\Delta x_i| \tag{2}$$

Note. *In the interval* $y \leq Y \leq y + \Delta y$ *the pre-images of* y *under* g *are* x_i *and the pre-images of* $y + \Delta y$ *under* g *are* $x_i + \Delta x_i$. *Here* Δy *and each* Δx_i *is sufficiently small such that all the events* $\big(X \in [x_i, x_i + \Delta x_i]\big)$ *are disjoint.*

From the Equations (1) and (2) we have

$$f_Y(y)\Delta y = \sum_i f_X(x_i)|\Delta x_i|$$

$$f_Y(y) = \sum_i f_X(x_i)\left|\frac{\Delta x_i}{\Delta y}\right| = \sum_i f_X(x_i)\left|\frac{\Delta y}{\Delta x_i}\right|^{-1}$$

$$\textit{We have } \lim_{\Delta x_i \to 0}\left|\frac{\Delta y}{\Delta x_i}\right| = \left|\frac{dg(x)}{dx}\right|_{x=x_i} = |g'(x_i)|$$

Therefore the pdf of the random variable $Y = g(X)$ is given as $f_Y(y) = \sum\limits_{x_i:g(x_i)=y} f_X(x_i)|g'(x_i)|^{-1}$.

Note. *Discrete-to-Continuous case is not possible. (Check !)* ***Class 11. (11/09/18)***

5 Mean(Expectation) and Variance of a random variable

Definition 5.1 (Mean of discrete random variable)**.** *Consider a discrete type random variable X taking values in $\{x_i : i \in S\}$ for some countable set S with PMF $P_X(x_i)$. The Mean or Expectation of X is denoted by $\mathbb{E}(X)$ and is defined as $\mathbb{E}(X) = \sum\limits_{x_i \in S} x_i P_X(x_i)$ provided $\sum\limits_{x_i \in S} |x_i| P_X(x_i) < \infty$.*

Definition 5.2 (Mean of continuous random variable)**.** *Consider a continuous type random variable X with pdf $f_X(x)$. The Mean or Expectation of X is denoted by $\mathbb{E}(X)$ and is defined as $\mathbb{E}(X) = \int\limits_{-\infty}^{\infty} x f_X(x)dx$ provided $\int\limits_{-\infty}^{\infty} |x| f_X(x)dx < \infty$.*

$\mathbb{E}(X)$ is usually denoted by the symbol μ.

Definition 5.3 (Variance of a random variable)**.** *The variance of a random variable is denoted by $Var(X)$ and is defined as $Var(X) = \mathbb{E}((X-\mu)^2)$.*

$Var(X)$ is usually denoted as σ^2.

$$Var(X) = \mathbb{E}((X-\mu)^2) = \mathbb{E}(X^2 + \mu^2 - 2\mu X) = \mathbb{E}(X^2) - \mu^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Note. *$\mathbb{E}((X-\mu)^n)$ is called as " n^{th} order moment of X about the mean" or " n^{th} order central moment of X" (for $n \in \{1, 2, \cdots\}$). So $Var(X)$ is the second order central moment. $\mathbb{E}(X^n)$ is called as n^{th} order moment of X about the origin (for $n \in \{1, 2, \cdots\}$).*

Standard deviation of X is defined as $+\sqrt{Var(X)}$. Therefore $\sigma = +\sqrt{Var(X)}$.

5.1 Mean of function of random variable

Lemma 5.1. *Let X be a discrete type random variable with PMF $P_X(x)$. Let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function and $Y = g(X)$ then $\mathbb{E}(Y) = \sum\limits_{x_i} g(x_i)P_X(x_i)$.*

Proof. Since X is a discrete type random variable, consider X takes values in the countable set $\{x_i : i \in S\}$. We know that Y is a discrete type random variable with PMF $P_Y(y) = \sum\limits_{x_i:g(x_i)=y} P_X(x_i)$.

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{y_i} y_i P_Y(y_i) = \sum_{y_i} y_i \sum_{g(x_{i,j})=y_i} P_X(x_{i,j}) = \sum_{y_i}\sum_{g(x_{i,j})=y_i} y_i P_X(x_{i,j}) \\
&= \sum_{y_i}\sum_{g(x_{i,j})=y_i} g(x_{i,j})P_X(x_{i,j}) \\
&= \sum_{g(x_{1,j})=y_1} g(x_{1,j})P_X(x_{1,j}) + \sum_{g(x_{2,j})=y_2} g(x_{2,j})P_X(x_{2,j}) + \cdots + \sum_{g(x_{s,j})=y_s} g(x_{s,j})P_X(x_{s,j}) \\
&= g(x_{1,j})\sum_{g(x_{1,j})=y_1} P_X(x_{1,j}) + g(x_{2,j})\sum_{g(x_{2,j})=y_2} P_X(x_{2,j}) + \cdots + g(x_{s,j})\sum_{g(x_{s,j})=y_s} P_X(x_{s,j}) \\
&= \sum_{x_j} g(x_j)P_X(x_j)
\end{aligned}$$

$$\textit{Therefore } \mathbb{E}(Y) = \sum_{x_i} g(x_i)P_X(x_i)$$

□

Lemma 5.2. *Let X be a continuous type random variable with pdf $f_X(x)$. Let $g : \mathbb{R} \to \mathbb{R}$ be a discrete measurable function and $Y = g(X)$ then $\mathbb{E}(Y) = \int\limits_{-\infty}^{\infty} g(x) f_X(x) dx$.*

Proof. We know that Y is a discrete type random variable with PMF $P_Y(y) = \int\limits_{all\ intervals\ :g(x)=y} f_X(x)dx$.

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{y_i} y_i P_Y(y_i) = \sum_{y_i} y_i \int\limits_{all\ intervals\ :g(x)=y_i} f_X(x)dx \\
&= \sum_{y_i} \int\limits_{all\ intervals\ :g(x)=y_i} y_i f_X(x)dx \\
&= \sum_{y_i} \int\limits_{all\ intervals\ :g(x)=y_i} g(x) f_X(x)dx \\
&= \int\limits_{all\ intervals\ :g(x)=y_1} g(x) f_X(x)dx + \int\limits_{all\ intervals\ :g(x)=y_2} g(x) f_X(x)dx + \cdots \int\limits_{all\ intervals\ :g(x)=y_s} g(x) f_X(x)dx \\
&= \int\limits_{-\infty}^{\infty} g(x) f_X(x)dx.
\end{aligned}$$

□

Lemma 5.3. *Let X be a continuous type random variable with pdf $P_X(x)$. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous measurable function and $Y = g(X)$ then $\mathbb{E}(Y) = \int\limits_{-\infty}^{\infty} g(x) f_X(x)$.*

Proof. The proof is beyond the scope of this course. □

Class 12. (14/09/18)

6 Standard/Common random variables and their distributions

In many real world problems some random variables are often relevant to model several situations. So we give some formal names to them and study their properties like CDF,pdf, PMF, mean, variance, moment generating function. Note that in general, the CDF $F_X(x)$ is sufficient to give the description of the random variable. In the case of continuous random variables, the pdf suffices, and in the case of discrete random variables, the PMF is sufficient.

Note. *for sketches of CDF, pdf, PMF of standard random variables refer to any standard text book (or draw them yourself or using a software like Octave to see how they look, how they change with the parameters, etc.).*

6.1 Standard continuous distributions

6.1.1 Uniform random variable (or Uniform Distribution on one dimension)

A continuous type random variable is said to be a *uniform random variable* if its pdf is

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

for some interval $[a, b]$.

If X is a uniformly distributed random variable in the interval $[a, b]$, then we represent it as $X \sim U(a, b)$.

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{1}{b-a}\int_a^b x dx = \frac{1}{2*(b-a)} * \left[x^2\right]_a^b = \frac{b+a}{2} \\
\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_a^b \frac{x^2}{b-a}dx = \frac{1}{b-a}\int_a^b x^2 dx = \frac{1}{3*(b-a)} * \left[x^3\right]_a^b = \frac{b^2+ba+a^2}{3} \\
Var(X) &= \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{b^2+ba+a^2}{3} - \Big(\frac{b+a}{2}\Big)^2 = \frac{(b-a)^2}{12}
\end{aligned}$$

6.1.2 Exponential Random variable (Exponential Distribution)

A continuous type random variable is said to be a *exponential random variable* if its pdf is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x < \infty \\ 0 & otherwise \end{cases}$$

for some $\lambda > 0$.

If X is a exponentially distributed random variable with parameter λ, then we represent it as $X \sim exp(\lambda)$.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \frac{1}{\lambda}$$
$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \frac{2}{\lambda^2}$$
$$Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{1}{\lambda^2}$$

6.1.3 Normal Distribution or Gaussian Distribution

A continuous type random variable is said to be a *Gaussian random variable* if its pdf is

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

for some $\mu \geq 0, \quad \sigma^2 > 0$.

If X is a Gaussian distributed random variable with parameters μ, σ^2, then we represent it as $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$\mathbb{E}(X) = \mu$$
$$Var(X) = \sigma^2$$

Note. *The special case of the Gaussian distribution with $\mu = 0, \sigma^2 = 1$ is called the Standard Normal Distribution. A random variable $X \sim \mathcal{N}(0,1)$ is called a Standard Normal Variable.*

Class 13. (18/09/18)

6.2 Standard discrete distributions

6.2.1 Constant random variable

The constant random variable takes only one value, say $a \in \mathbb{R}$. The PMF is given by

$$P_X(x) = \begin{cases} 1 & x = a \\ 0 & otherwise \end{cases}$$

Thus the CDF, $F_X(x) = \begin{cases} 0 & x < a. \\ 1 & x \geq a. \end{cases}$

$$\mathbb{E}(X) = \sum_{x_i} x_i P_X(x_i) = aP(a) + \sum_{x_i \neq a} x_i P_X(x_i) = a(1) + 0 = a$$
$$\mathbb{E}(X^2) = \sum_{x_i} x_i^2 P_X(x_i) = a^2 P(a) + \sum_{x_i \neq a} x_i^2 P_X(x_i) = a^2$$
$$Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = a^2 - (a)^2 = 0$$

This can be easily understood by intuition. The random variable takes a single value i.e it does not *deviate from its mean*, and so the variance will be zero.

Exercise : Let X be a non-negative random variable. If $Var(X) = 0$, show that $P(X = c) = 1$ for some $c \geq 0$.

6.2.2 Bernoulli random variable (or Bernoulli distribution)

A Bernoulli random variable is one which can take two values (typically these two values are 0 and 1). For $p \in [0,1]$ the pmf of a Bernoulli random variable are given below

$$P_X(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & otherwise \end{cases}$$

Thus the CDF is $F_X(x) = \begin{cases} 0 & x < 0. \\ 1-p & 0 \le x < 1. \\ 1 & x \ge 1. \end{cases}$

Note that this distribution is parameterized by the value p. If X is a Bernoulli random variable, we say $X \sim Bernoulli(p)$.

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x_i} x_i P_X(x_i) = 0 * P(0) + 1 * P(1) = 0 * (1-p) + 1 * p = p \\ \mathbb{E}(X^2) &= \sum_{x_i} x_i^2 P_X(x_i) = 0 * P(0) + 1 * P(1) = 0 * (1-p) + 1 * p = p \\ Var(X) &= \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = p - p^2 = p(1-p) \end{aligned}$$

6.2.3 Binomial random variable

Definition 6.1. *A discrete type random variable is said to be Binomial random variable if its pmf is*

$$P_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^x & x \in \{0,1,2,\cdots,n\} \\ 0 & otherwise \end{cases}$$

for some $p \in [0,1]$.

Here n, p are the parameters of the Binomial random variable. If X is binomial random variable with parameters n, p then we represent it as $X \sim B(n,p)$.

$$\begin{aligned} \mathbb{E}(X) &= np \\ Var(X) &= np(1-p) \end{aligned}$$

Note. *Bernoulli random variable is a special case of Binomial random variable with* $n = 1$*. Therefore if* X *is a Bernoulli random variable with parameter* p*, then it can be repreented as* $X \sim B(1,p)$*. More interestingly, at some point later, we will see that the Binomial random variable is equal to the sum of independent Bernoulli random variables.*

6.2.4 Poisson random variable (countably infinite jumps)

Definition 6.2. *A discrete type random variable is said to be Poisson random variable if its pmf is*

$$P_X(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & x \in \{0,1,2,\cdots\} \\ 0 & otherwise \end{cases}$$

for some $\lambda > 0$.

Here λ is the parameter of the Poisson random variable. If X is Poisson random variable with parameter λ then we represent it as $X \sim Poisson(\lambda)$.

$$\begin{aligned} \mathbb{E}(X) &= \sum_k k P_X(k) = \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = \lambda \sum_{(k-1)=0}^{\infty} \frac{e^{-\lambda}\lambda^{(k-1)}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda \\ \mathbb{E}(X^2) &= \sum_k k^2 P_X(k) = \sum_k (k^2 - k + k) P_X(k) = \sum_{k=0}^{\infty} k(k-1) \frac{e^{-\lambda}\lambda^k}{k!} + \mathbb{E}(X) \\ &= e^{-\lambda}\left(0 + 0 + \left(2 * 1 * \frac{\lambda^2}{2!}\right) + \left(3 * 2 * \frac{\lambda^3}{3!}\right) + \cdots\right) + \lambda \\ &= e^{-\lambda}\lambda^2(1 + \lambda + \lambda^2 + \cdots) + \lambda = e^{-\lambda}\lambda^2 e^{\lambda} + \lambda = \lambda^2 + \lambda \\ Var(X) &= \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

Class 14. (25/09/18)

7 Jointly distributed RVs

Definition 7.1. *Consider the probability space* $(\Omega, \mathcal{F}, P)$. *Let* $X_i : \Omega \to \mathbb{R}, \forall i \in \{1, 2, \cdots, n\}$ *be RVs, then* $X_1, X_2, X_3, ...X_n$ *are said to be jointly distributed with* ***joint CDF*** *denoted as* $F_{X_1,X_2,...X_n}(x_1, x_2, ...x_n)$ *and is defined as*

$$F_{X_1,X_2,...X_n}(x_1, x_2, ...x_n) = P\big((X_1 \le x_1), (X_2 \le x_2), ...(X_n \le x_n)\big)$$

Note. *In the right hand side of the above equation comma* $(,)$ *represents intersection* $(\cap)$.

Therefore

$$\begin{aligned} F_{X_1,X_2,...X_n}(x_1, x_2, ...x_n) &= P\big((X_1 \le x_1) \cap (X_2 \le x_2) \cap \cdots (X_n \le x_n)\big) \\ &= P\left(\{\omega \in \Omega : X_1(w) \le x_1\} \cap \{\omega \in \Omega : X_2(w) \le x_2\} .... \cap \{\omega \in \Omega : X_n(w) \le x_n\}\right) \\ &= P\left(\{\omega \in \Omega : X_1(\omega) \le x_1, X_2(\omega) \le x_2, \cdots, X_n(\omega) \le x_n\}\right) \end{aligned}$$

7.1 Properties of Joint CDF

Let X and Y be jointly distributed RVs. Then the joint CDF will be $F_{X,Y}(x, y) = P(X \le x, Y \le y)$.

1. $F_{X,Y}(x, y)$ is non-decreasing in both X and Y.
   - If $x_1 < x_2$, then $F_{X,Y}(x_1, y) \le F_{X,Y}(x_2, y)\ \forall\ y \in \mathbb{R}$.
   - If $y_1 < y_2$, then $F_{X,Y}(x, y_1) \le F_{X,Y}(x, y_2)\ \forall\ x \in \mathbb{R}$.
   - If $x_1 < x_2$, and $y_1 < y_2$, then $F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_1) \le F_{X,Y}(x_2, y_2)$. where $x_1, x_2, y_1, y_2 \in \mathbb{R}$
2. - $\lim\limits_{\substack{x\to\infty \\ y\to\infty}} F_{X,Y}(x, y) = 1$
   - $\lim\limits_{\substack{x\to-\infty \\ y\to-\infty}} F_{X,Y}(x, y) = 0$
   - $\lim\limits_{y\to-\infty} F_{X,Y}(x, y) = 0$
   - $\lim\limits_{x\to-\infty} F_{X,Y}(x, y) = 0$
3. $F_{X,Y}(x, y)$ is Right continuous in x,y.
   - $\lim\limits_{\epsilon_1\to 0+} F_{X,Y}(x + \epsilon_1, y) = F_{X,Y}(x, y)$
   - $\lim\limits_{\epsilon_2\to 0+} F_{X,Y}(x, y + \epsilon_2) = F_{X,Y}(x, y)$
   - $\lim\limits_{\substack{\epsilon_1\to 0+ \\ \epsilon_2\to 0+}} F_{X,Y}(x + \epsilon_1, y + \epsilon_2) = F_{X,Y}(x, y)$
4. Marginal CDF
   - $F_X(x) = \lim\limits_{y\to\infty} F_{X,Y}(x, y)$
   - $F_Y(y) = \lim\limits_{x\to\infty} F_{X,Y}(x, y)$