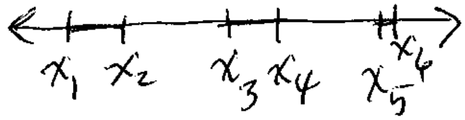
9/20/2022

We have defined the cdf $F_X(x)$ of a rv X as $F_X(x) = P(X \leq x)$, $\forall x \in \mathbb{R}$

But it can be unwieldy to use the cdf for some $A \in \mathcal{B}(\mathbb{R})$, e.g.,



if $A = (x_1, x_2) \cup (x_3, x_4) \cup (x_5, x_6)$,

$P(X \in A)$ can be written in terms F_X, but the form for $P(X \in A)$ would have to be found for this specific.

Instead, often use the following:

Defn. A <u>probability density function</u>, or simply density function or pdf, of a rv X is a function $f_X(x)$ satisfying $P(X \in A) = \int_A f_X(x)\,dx$.

Extra discussion:

If the Riemann integral (RI) is used in the above defn, there will be a problem.

There are valid pdfs (defined later) and $A \in \mathcal{B}(\mathbb{R})$ where

$\int_A f_X(x)\,dx$ does not exist under RI

For example, consider the pdf

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

and $A = \mathbb{Q} = \bigcup_{n=1}^{\infty} \{q_n\} \in \mathcal{B}(\mathbb{R})$

q_n ← nth rational

Then ~~under~~

$$\int_A f_X(x)\,dx = \int_0^1 1\, I_{\mathbb{Q}}(x)\,dx = \int_0^1 I_{\mathbb{Q}}(x)\,dx$$

where $I_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$

Indicator function on $B \subset \mathbb{R}$

It can be shown that
the RI does not exist in this case. (See supplementary pages)

~~the~~ of

Instead the Lebesgue integral is used in prob. This can be learned in ~~real~~ real analysis/ measure theory.

When the ~~LI~~ RI exists, it is equal to the LI, so you can do integration as you always have.

An alternative definition of the pdf f_X is

$$f_X(x) = \frac{dF_X(x)}{dx},$$

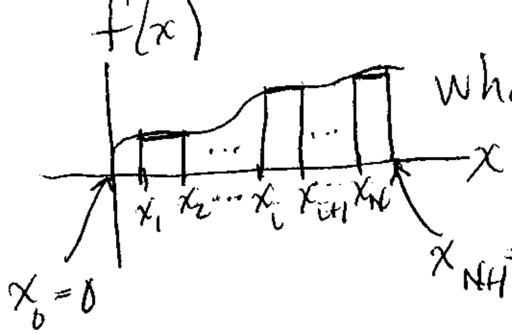
$\forall x \in \mathbb{R}$ where x is differentiable

At points where F_X has a jump discontinuity, we let $f_X(x) = P(X=x)\delta(x)$,

where $\delta(x)$ is the Dirac delta fn.

Prove that if
$$f_X(x) = \begin{cases} 1 & \text{if } x \in Q \cap [0,1] \\ 0 & \text{otherwise,} \end{cases}$$
then the RI
$$RI \int_Q f_X(x)\,dx \quad \text{does not exist.}$$

Proof: Recall that the RI of a function f (over $[0,1]$ for this example) is
$$\int_0^1 f(x)\,dx = \lim_{N \to \infty} \sum_{i=0}^{N} f(\tilde{x}_i)(x_{i+1} - x_i),$$



where $\tilde{x}_i$ is an arbitrary element of $[x_i, x_{i+1}]$.

The RI exists if for any choice of $\tilde{x}_0, \ldots, \tilde{x}_N$ the limit exists and is equal to the limit for any other choice of $\tilde{x}_0, \ldots, \tilde{x}_N$.

We can write
$$\int_Q f_X(x)\,dx = \int_0^1 1 \cdot I_Q(x)\,dx = \int_0^1 I_Q(x)\,dx,$$
where I_Q is the indicator fn. on the set Q.

(We write it this way because we know how to integrate over $[0,1]$, but not over Q. Actually, if we treat elements of $\mathbb{Q}$ as zero-width intervals, we could argue that the integral is $0 \cdot \infty$, or undefined, but the integral is defined for non-zero-width intervals, so that would be incorrect.) (52)

So $\int_{\mathbb{Q}} f_X(x)\,dx = \lim_{N\to\infty} \sum_{i=0}^{N} I_{\mathbb{Q}}(\tilde{x}_i)(x_{i+1}-x_i)$

First, let $\tilde{x}_i \in \mathbb{Q}$ $\forall i = 0, \dots, N$.

(Note that every interval in $\mathbb{R}$ has at least one rational and one irrational number.)

Then

$$\lim_{N\to\infty} \sum_{i=0}^{N} I_{\mathbb{Q}}(\tilde{x}_i)(x_{i+1}-x_i) =$$

$$\lim_{N\to\infty} \sum_{i=0}^{N} 1 \cdot (x_{i+1}-x_i) =$$

$$\lim_{N\to\infty} \sum_{i=1}^{N} (x_{i+1}-x_i) = \lim_{N\to\infty} |[0,1]| = 1$$

It can similarly be shown that if we let $\tilde{x}_i \notin \mathbb{Q}$ $\forall i = 0, \dots, N$, then $\lim_{N\to\infty} \sum_{i=0}^{N} I_{\mathbb{Q}}(\tilde{x}_i)(x_{i+1}-x_i) = 0$

Since the limits are not equal, the RI does not exist defined as the function satisfying: (i) $\cancel{\delta(0) = 0}$ $\delta(x) = 0 \quad \forall x \neq 0$

(ii) $\int_{-\infty}^{\infty} \delta(x)\,dx = \int_{-\varepsilon}^{\varepsilon} \delta(x)\,dx = 1$

$\forall \varepsilon > 0$.

Properties of the pdf:

1. $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$,

since $\int_{-\infty}^{\infty} f_X(x)\,dx = P(X \in \mathbb{R}) = 1$

It can be shown that a fn. f_X satisfying Properties 1 and 2 is a valid pdf, meaning that the measure P obtained using

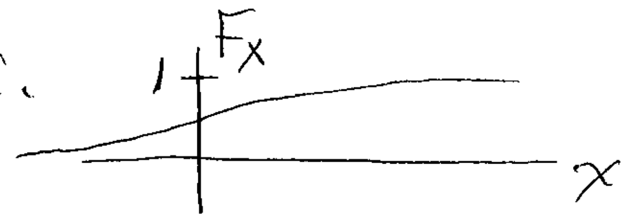
$$P(A) = \int_A f_X(x)\,dx \qquad \forall A \in \mathcal{B}(\mathbb{R})$$

~~is~~ satisfies the axioms

Types of Random Variables

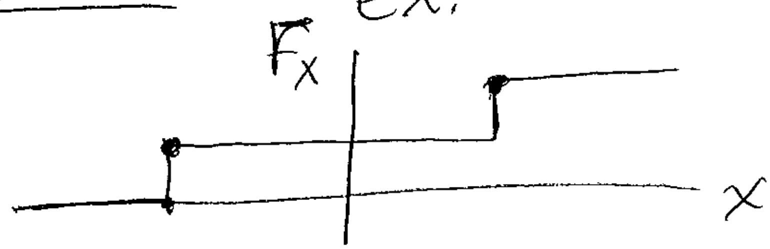
Defn. A rv X having a cdf that is continuous everywhere is called a <u>continuous rv</u>.

Ex.



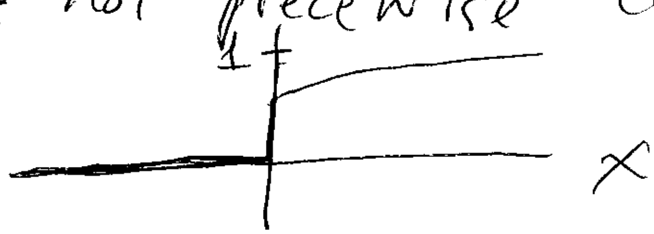
Defn. A rv X having a piece-wise constant cdf is called a <u>discrete rv</u>

Ex.



Note that a <u>mixed rv</u> has a cdf that is continuous everywhere except possibly on a ~~set~~ finite or countably infinite set of points and is ~~not~~ not piecewise constant

Ex.

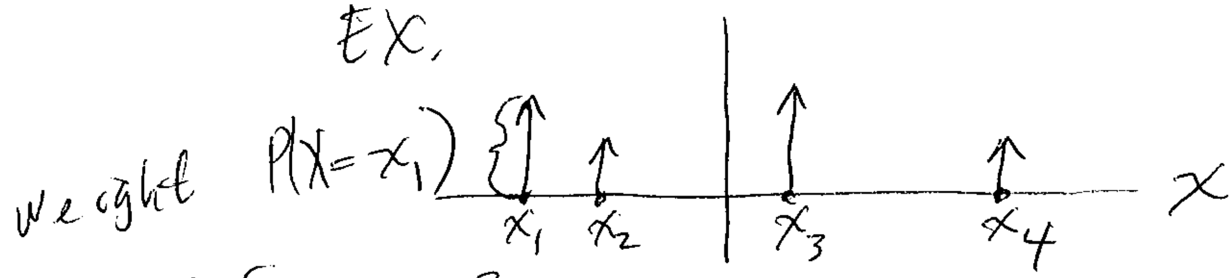


Note that the density fn. of a discrete rv X is written as

$$f_X(x) = \sum_{x_0 \in R_X} P(X=x_0)\,\delta(x-x_0)$$

where R_X is the set of points where F_X is discontinuous

Ex.



$R_X = \{x_1, x_2, x_3, x_4\}$

Defn. The <u>probability mass function</u>, or simply mass function, or pmf of a discrete rv X is

$$p_X(x) = P(X=x) \quad \forall x \in R_X$$

Note that for a continuous rv

$P(X \in A) = \int_A f_X(x)\,dx$ is an integral of the type you are used to seeing

In the discrete case:

$$P(X \in A) = \int_A f_X(x)\,dx =$$

$$\int_A \sum_{x_0 \in \mathcal{R}_X} P(X = x_0)\,\delta(x - x_0)\,dx$$

$$= \sum_{x_0 \in \mathcal{R}_X} P(X = x_0) \int_A \delta(x - x_0)\,dx$$

$$= \sum_{x_0 \in \mathcal{R}_X} P(X = x_0)\, \mathbb{1}_{\cancel{A}}(x_0)$$

The pmf of discrete X can be read directly from the pdf and vice versa. For previous example,