Recap: Monotone Sequences of Sets

- A sequence, $A_1, A_2, \cdots$, is said to be monotone decreasing if

$$A_{n+1} \subset A_n, \ \forall n \quad \text{(denoted as } A_n \downarrow\text{)}$$

- Limit of a monotone decreasing sequence is

$$A_n \downarrow: \quad \lim_{n\to\infty} A_n = \cap_{k=1}^{\infty} A_k$$

- A sequence, $A_1, A_2, \cdots$, is said to be monotone increasing if

$$A_n \subset A_{n+1}, \ \forall n \quad \text{(denoted as } A_n \uparrow\text{)}$$

- Limit of monotone increasing sequence is

$$A_n \uparrow: \quad \lim_{n\to\infty} A_n = \cup_{k=1}^{\infty} A_k$$

Recap: Monotone Sequential Continuity

- We showed that

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

when $A_n \downarrow$ or $A_n \uparrow$

Recap: Probability Models

- Everything in probability theory is with reference to an underlying probability space: $(\Omega, \mathcal{F}, P)$.
- For a given $\Omega, \mathcal{F}$, there can be different P
- We can say that different P correspond to different models.
- Theory does not tell you how to get the P.
- The theory allows one to derive consequences or properties of the model.
- The random variables provide a convenient language to describe different probability models.

Recap: Random Variable

- A random variable is a real-valued function on Ω: $X : \Omega \to \Re$
- For example, $\Omega = \{H, T\}$, $X(H) = 1, \ X(T) = 0$.
- This random variable results in a new probability space:

$$(\Omega, \mathcal{F}, P) \ \stackrel{X}{\to} \ (\Re, \mathcal{B}, P_X)$$

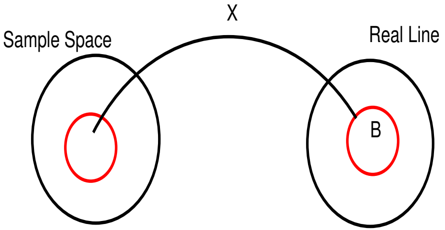
  where $\Re$ is the new sample space and $\mathcal{B} \subset 2^{\Re}$ is the new set of events and P_X (which depends on P and X) is a probability defined on $\mathcal{B}$.
- For now we will assume that any set of $\Re$ that we want would be in $\mathcal{B}$ and hence is an event.

- Given a probability space $(\Omega, \mathcal{F}, P)$, a random variable X

$$(\Omega, \mathcal{F}, P) \stackrel{X}{\rightarrow} (\Re, \mathcal{B}, P_X)$$

- We define P_X:

$$P_X(B) = P\left(\{\omega \in \Omega \ : \ X(\omega) \in B\}\right), \ B \in \mathcal{B}$$



- Given a probability space $(\Omega, \mathcal{F}, P)$, a random variable X

$$(\Omega, \mathcal{F}, P) \stackrel{X}{\rightarrow} (\Re, \mathcal{B}, P_X)$$

- We define P_X:

$$P_X(B) = P\left(\{\omega \in \Omega \ : \ X(\omega) \in B\}\right), \ B \in \mathcal{B}$$

- We use the notation

$$[X \in B] = \{\omega \in \Omega \ : \ X(\omega) \in B\}$$

- So, now we can write

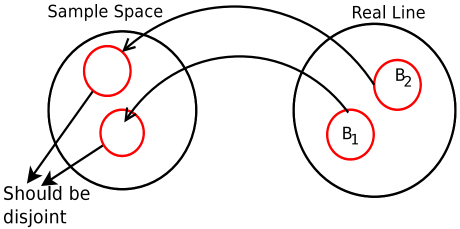
$$P_X(B) = P([X \in B]) = P[X \in B]$$

- For the definition of P_X to be proper, for each $B \in \mathcal{B}$, we must have $[X \in B] \in \mathcal{F}$.
  We will assume that. (This is trivially true if $\mathcal{F} = 2^\Omega$).
- We can easily verify P_X is a probability measure. It satisfies the axioms.

- Given a probability space $(\Omega, \mathcal{F}, P)$, a random variable X
- We define P_X:

$$P_X(B) = P[X \in B] = P\left(\{\omega \in \Omega \;:\; X(\omega) \in B\}\right)$$

- Easy to see: $P_X(B) \geq 0, \; \forall B$ and $P_X(\Re) = 1$
- If $B_1 \cap B_2 = \phi$ then $P_X(B_1 \cup B_2) = P[X \in B_1 \cup B_2] = ?$



$$P[X \in B_1 \cup B_2] = P[X \in B_1] + P[X \in B_2] = P_X(B_1) + P_X(B_2)$$

- Let us look at a couple of simple examples.
- Let $\Omega = \{H, T\}$ and $P(H) = p$.
  Let $X(H) = 1; X(T) = 0$.

$$\begin{aligned}
[X \in \{0\}] &= \{\omega \ : \ X(\omega) = 0\} = \{T\} \\
[X \in [-3.14, \ 0.552]\,] &= \{\omega : -3.14 \leq X(\omega) \leq 0.552\} = \{T\} \\
[X \in (0.62, \ 15.5)] &= \{\omega : 0.62 < X(\omega) < 15.5\} = \{H\} \\
[X \in [-2, \ 2)\,] &= \Omega
\end{aligned}$$

- Hence we get

$$P_X(\{0\}) = (1 - p) = P_X([-3.14, \ 0.552])$$

$$P_X((0.6237, \ 15.5)) = p; \ P_X([-2, \ 2)) = 1$$

- Let us look at a couple of simple examples.
- Let $\Omega = \{H, T\}$ and $P(H) = p$.
  Let $X(H) = 1; X(T) = 0$.

$$\begin{aligned} [X \in \{0\}] &= \{\omega \ : \ X(\omega) = 0\} = \{T\} \\ [X \in [-3.14, \ 0.552]\,] &= \{\omega : -3.14 \leq X(\omega) \leq 0.552\} = \{T\} \\ [X \in (0.62, \ 15.5)] &= \{\omega : 0.62 < X(\omega) < 15.5\} = \{H\} \\ [X \in [-2, \ 2)\,] &= \Omega \end{aligned}$$

- Hence we get

$$P_X(\{0\}) = (1 - p) = P_X([-3.14, \ 0.552])$$

$$P_X((0.6237, \ 15.5)) = p; \ P_X([-2, \ 2)) = 1$$

- Let $\Omega = \{H, T\}^3 = \{HHH, HHT, \cdots, TTT\}$.
  Let P be specified through 'equally likely' assignment.
  Let $X(\omega)$ be number of H's in ω. Thus, $X(THT) = 1$.
  (X takes one of the values: 0, 1, 2, or 3)
- We can once again write down $[X \in B]$ for different $B \subset \Re$

$$[X \in (0, 1]\,] = \{HTT, THT, TTH\};$$

$$[X \in (-1.2, 2.78)\,] = \Omega - \{HHH\}$$

- Hence

$$P_X((0, 1]) = \frac{3}{8};\ P_X((-1.2, 2.78)) = \frac{7}{8}$$

- A random variable defined on $(\Omega, \mathcal{F}, P)$ results in a new or induced probability space $(\Re, \mathcal{B}, P_X)$.
- The Ω may be countable or uncountable (even though we looked at only examples of finite Ω).
- Thus, we can study probability models by taking $\Re$ as sample space through the use of random variables.
- However there are some technical issues regarding what $\mathcal{B}$ we should consider.
- We briefly consider this and then move on to studying random variables.

- We want to look at the probability space $(\Re, \mathcal{B}, P_X)$.
- If we could take $\mathcal{B} = 2^{\Re}$ then everything would be simple. But that is not feasible.
- What this means is the following.
- Suppose $\Omega = \Re$. If we want every subset of real line to be an event, we cannot construct a probability measure (to satisfy the axioms).

- Let us consider $\Omega = [0, 1]$.
- This is the simplest example of uncountable Ω we considered.
- We also saw that this sample space comes up when we consider infinite tosses of a coin.
- The simplest extension of the idea of 'equally likely' is to say probability of an event (subset of Ω) is the length of the event (subset).
- But not all subsets of $[0, 1]$ are intervals and length is defined only for intervals.
- We can define length of countable union of disjoint intervals to be sum of the lengths of individual intervals.
- But what about subsets that may not be countable unions of disjoint intervals ?
- Well, we say those can be assigned probability by using the axioms.

- Thus the question is the following:
- Can we construct a function $m : 2^{[0,1]} \to [0,1]$ such that
  1. $m(A) =$ length(A) if $A \subset [0,1]$ is an interval
  2. $m(\cup_i A_i) = \sum_i m(A_i)$ where $A_i \cap A_j = \phi$ whenever $i \neq j,\ (A_1, A_2, \cdots \subset [0,1])$
- The surprising answer is 'NO'
- This is a fundamental result in real analysis.
- Hence for the probability space $(\Re, \mathcal{B}, P_X)$ we cannot take $\mathcal{B} = 2^{\Re}$.
  (Recall that for countable Ω we can take $\mathcal{F} = 2^{\Omega}$).
- Now the question is what is the best $\mathcal{B}$ we can have?

σ-algebra

- Consider a set of subsets of Ω: $\mathcal{F} \subset 2^\Omega$
- An $\mathcal{F} \subset 2^\Omega$ is called a σ-algebra (also called σ-field) on Ω if it satisfies the following:
  1. $\Omega \in \mathcal{F}$
  2. $A \in \mathcal{F} \;\; \Rightarrow \;\; A^c \in \mathcal{F}$
  3. $A_1, A_2, \cdots \in \mathcal{F} \;\; \Rightarrow \;\; \cup_i A_i \in \mathcal{F}$
- Thus a σ-algebra is a collection of subsets of Ω that is closed under complements and countable unions (and hence countable intersections because $\cap_i A_i = (\cup_i A_i^c)^c$).
- Note that 2^Ω is obviously a σ-algebra
- In a Probability space $(\Omega, \mathcal{F}, P)$, if $\mathcal{F} \neq 2^\Omega$ then we want it to be a σ-algebra. (Why?)

- Easy to construct examples of σ-algebras
  Let $A \subset \Omega$.

$$\mathcal{F} = \{\Omega, \phi, A, A^c\} \text{ is a } \sigma\text{-algebra}$$

- For example, with $\Omega = \{1, 2, 3, 4, 5, 6\}$,

$$\mathcal{F} = \{\Omega, \phi, \{1, 3, 5\}, \{2, 4, 6\}\} \text{ is a } \sigma\text{-algebra}$$

- There can be many other σ-algebras on the Ω.
- Suppose on this Ω we want to make a σ-algebra containing $\{1, 2\}$ and $\{3, 4\}$.

$$\{\Omega, \phi, \{1, 2\}, \{3, 4\}, \{3, 4, 5, 6\}, \{1, 2, 5, 6\}, \{1, 2, 3, 4\}, \{5, 6\}\}$$

- This is the 'smallest' σ-algebra containing $\{1, 2\}$, $\{3, 4\}$

- Let $\mathcal{F}_1, \mathcal{F}_2$ be σ-algebras on Ω.
- Then, so is $\mathcal{F}_1 \cap \mathcal{F}_2$.
- It is simple to show.
  (E.g., $A \in \mathcal{F}_1 \cap \mathcal{F}_2 \;\Rightarrow\; A \in \mathcal{F}_1, A \in \mathcal{F}_2 \;\Rightarrow\; A^c \in \mathcal{F}_1, A^c \in \mathcal{F}_2 \;\Rightarrow\; A^c \in \mathcal{F}_1 \cap \mathcal{F}_2$)
- Let $G \subset 2^{\Omega}$. We denote by $\sigma(G)$ the smallest σ-algebra containing G.
- It is defined as the intersection of all σ-algebras containing G (and hence is well defined).

- Let us get back to the question we started with.
- In the probability space $(\Re, \mathcal{B}, P)$ what is the $\mathcal{B}$ we should choose.
- We can choose it to be the smallest σ-algebra containing all intervals
- That is called Borel σ-algebra, $\mathcal{B}$.
- It contains all intervals, all complements, countable unions and intersections of intervals and all sets that can be obtained through complements, countable unions and/or intersections of such sets and so on.

Borel σ-algebra

- Let $\quad G = \{(-\infty,\ x]\ :\ x \in \Re\}$
- We can define the Borel σ-algebra, $\mathcal{B}$, as the smallest σ-algebra containing G.
- We can see that $\mathcal{B}$ would contain all intervals.
  1. $(-\infty,\ x) \in \mathcal{B}$ because $(-\infty,\ x) = \cup_n (-\infty,\ x - \frac{1}{n}]$
  2. $(x,\ \infty) \in \mathcal{B}$ because $(x,\ \infty) = (-\infty,\ x]^c$
  3. $[x,\ \infty) \in \mathcal{B}$ because $[x,\ \infty) = \cap_n (x - \frac{1}{n},\ \infty)$
  4. $(x,\ y] \in \mathcal{B}$ because $(x,\ y] = (-\infty,\ y] \cap (x,\ \infty)$
  5. $[x,\ y] \in \mathcal{B}$ because $[x,\ y] = \cap_n (x - \frac{1}{n},\ y]$
  6. $[x,\ y), (x,\ y) \in \mathcal{B}$, similarly
- Thus, $\sigma(G)$ is also the smallest σ-algebra containing all intervals.

Borel σ-algebra

- We have defined $\mathcal{B}$ as

$$\mathcal{B} = \sigma\left(\{(-\infty,\ x]\ :\ x \in \Re\}\right)$$

- It is also the smallest σ-algebra containing all intervals.
- Elements of $\mathcal{B}$ are called Borel sets
- Intervals (including singleton sets), complements of intervals, countable unions and intersections of intervals, countable unions and intersections of such sets and so on are all Borel sets.
- Borel σ-algebra contains enough sets for our purposes.
- Are there any subsets of real line that are not Borel?
- YES!! Infinitely many non-Borel sets exist!

Random Variables

- Given a probability space $(\Omega, \mathcal{F}, P)$, a random variable is a real-valued function on Ω.
- It essentially results in an induced probability space

$$(\Omega, \mathcal{F}, P) \overset{X}{\rightarrow} (\Re, \mathcal{B}, P_X)$$

where $\mathcal{B}$ is the Borel σ-algebra.
- $\mathcal{B}$ contains all subsets of $\Re$ that are intervals, and complements of intervals, and countable unions and intersections of such sets, and $\cdots$
- We define P_X: for all Borel sets, $B \subset \Re$,

$$P_X(B) = P[X \in B] = P\left(\{\omega \in \Omega \;:\; X(\omega) \in B\}\right)$$

- For X to be a random variable, the folowing should also hold

$$[X \in B] = \{\omega \in \Omega \;:\; X(\omega) \in B\} \in \mathcal{F},\ \forall B \in \mathcal{B}$$

- We always assume this.

- Let X be a random variable.
- It represents a probability model with $\Re$ as the sample space.
- The probability assigned to different events (Borel subsets of $\Re$) is

$$P_X(B) = P[X \in B] = P\left(\{\omega \in \Omega \ : \ X(\omega) \in B\}\right)$$

- How does one represent this probability measure

Distribution function of a random variable

- Let X be a random variable on $(\Omega, \mathcal{F}, P)$. The distribution function of X is:
  $F_X : \Re \rightarrow \Re$ defined by
  $$F_X(x) = P[X \in (-\infty, x]] = P(\{\omega \in \Omega \ : \ X(\omega) \leq x\})$$
- We write the event $\{\omega \ : \ X(\omega) \leq x\}$ as $[X \leq x]$.
  We follow this notation with any such relation statement involving X
  e.g., $[X \neq 3]$ represents the event $\{\omega \in \Omega \ : \ X(\omega) \neq 3\}$.
- Thus we have

$$F_X(x) = P[X \leq x] = P(\{\omega \in \Omega \ : \ X(\omega) \leq x\}) = P_X(\ (-\infty, \ x]\ )$$

- The distribution function, F_X completely specifies the probability measure, P_X.
  (The proof of this important theorem is beyond the scope of this course)

- The distribution function of X is given by

$$F_X(x) = P[X \leq x] = P(\{\omega \in \Omega \ : \ X(\omega) \leq x\})$$

- This is also sometimes called the cumulative distribution function.
- F_X is a real-valued function of a real variable.
- Let us look at a simple example.

- Consider tossing of a fair coin: $\Omega = \{T, H\}$, $P(\{T\}) = P(\{H\}) = 0.5$.
- Let $X(T) = 0$ and $X(H) = 1$. We want to calculate F_X
- For this we want the event $[X \leq x]$, for different x
- Let us first look at some examples:

$$\begin{aligned} [X \leq -0.5] &= \{\omega \ : \ X(\omega) \leq -0.5\} = \phi \\ [X \leq 0.25] &= \{\omega \ : \ X(\omega) \leq 0.25\} = \{T\} \\ [X \leq 1.3] &= \{\omega \ : \ X(\omega) \leq 1.3\} = \Omega \end{aligned}$$

- Thus we get

$$\begin{aligned} [X \leq x] &= \{\omega \ : \ X(\omega) \leq x\} \\ &= \begin{cases} \phi & \text{if } x < 0 \\ \Omega & \text{if } x \geq 1 \\ \{T\} & \text{if } 0 \leq x < 1 \end{cases} \end{aligned}$$

- We are considering: $\Omega = \{T, H\}$, $P(\{T\}) = P(\{H\}) = 0.5$.
- $X(T) = 0$ and $X(H) = 1$. We want to calculate F_X
- We showed

$$\begin{aligned}
[X \leq x] &= \{\omega \;:\; X(\omega) \leq x\} \\
&= \begin{cases} \phi & \text{if } x < 0 \\ \{T\} & \text{if } 0 \leq x < 1 \\ \Omega & \text{if } x \geq 1 \end{cases}
\end{aligned}$$

- Hence $F_X(x) = P[X \leq x]$ is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.5 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Please note that x is a 'dummy variable'

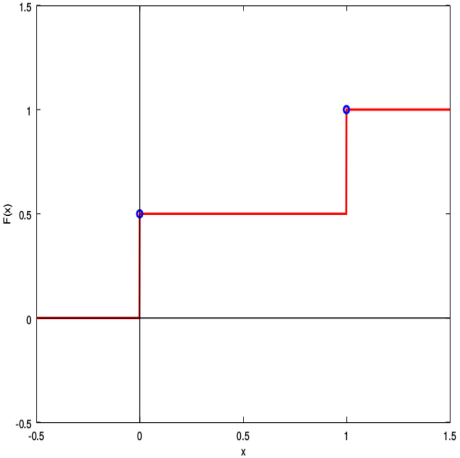
- We are considering: $\Omega = \{T, H\}$, $P(\{T\}) = P(\{H\}) = 0.5$.
- $X(T) = 0$ and $X(H) = 1$. We want to calculate F_X
- We showed

$$\begin{aligned} [X \leq x] &= \{\omega \ : \ X(\omega) \leq x\} \\ &= \begin{cases} \phi & \text{if } \ x < 0 \\ \{T\} & \text{if } \ 0 \leq x < 1 \\ \Omega & \text{if } \ x \geq 1 \end{cases} \end{aligned}$$

- Hence $F_X(y) = P[X \leq y]$ is given by

$$F_X(y) = \begin{cases} 0 & \text{if } \ y < 0 \\ 0.5 & \text{if } \ 0 \leq y < 1 \\ 1 & \text{if } \ y \geq 1 \end{cases}$$

- A plot of this distribution function:



- Let us look at another example.
- Let $\Omega = [0,\ 1]$ and take events to be Borel subsets of $[0,\ 1]$. (That is, $\mathcal{F} = \{B \cap [0,\ 1]\ :\ B \in \mathcal{B}\}$).
- We take P to be such that probability of an interval is its length.
- This is the 'usual' probability space whenever we take $\Omega = [0,\ 1]$.
- Let $X(\omega) = \omega$.
- We want to find the distribution function of X.

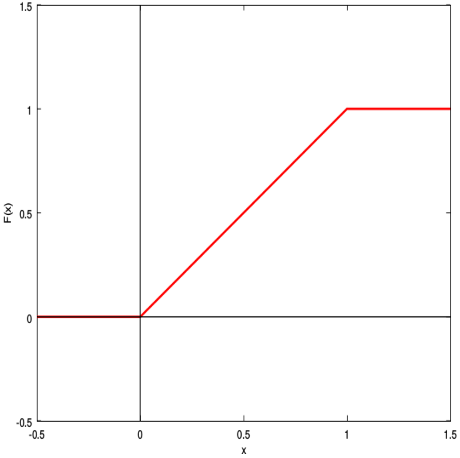
- Once again we need to find the event $[X \leq x]$ for different values of x.
- Note that the function X takes values in $[0,\ 1]$ and $X(\omega) = \omega$.

$$
\begin{aligned}
[X \leq x] &= \{\omega \in \Omega \ : \ X(\omega) \leq x\} = \{\omega \in [0,\ 1] \ : \ \omega \leq x\} \\
&= \begin{cases} \phi & \text{if } \ x < 0 \\ \Omega & \text{if } \ x \geq 1 \\ [0, x] & \text{if } \ 0 \leq x < 1 \end{cases}
\end{aligned}
$$

- Hence $F_X(x) = P[X \leq x]$ is given by

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

- The plot of this distribution function:



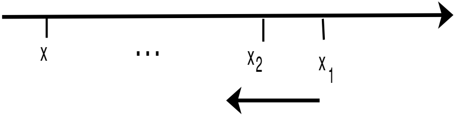
Properties of Distribution Functions

- The distribution function of random variable X is given by

$$F_X(x) = P[X \leq x] = P(\{\omega\ : X(\omega) \leq x\}) = P_X(\ (-\infty,\ x]\ )$$

- Any distribution function should satisfy the following:
  1. $0 \leq F_X(x) \leq 1,\ \forall x$
  2. $F_X(-\infty) = 0;\ F_X(\infty) = 1$
  3. F_X is non-decreasing: $x_1 \leq x_2\ \Rightarrow\ F_X(x_1) \leq F_X(x_2)$
     (Why?)
     $x_1 \leq x_2\ \Rightarrow\ (-\infty,\ x_1] \subset (-\infty,\ x_2]$
     $\Rightarrow\ \ P_X(\ (-\infty,\ x_1]\ ) \leq P_X(\ (-\infty,\ x_2]\ )$
     $\Rightarrow\ \ F_X(x_1) \leq F_X(x_2)$
  4. F_X is right continuous and has left-hand limits.

- Right continuity of F_X: $x_n \downarrow x \;\Rightarrow F_X(x_n) \rightarrow F_X(x)$
  - $x_n \downarrow x$ implies the sequence of events $(-\infty,\ x_n]$ is monotone decreasing.



  - Also, $\lim_n (-\infty,\ x_n] = \cap_n (-\infty,\ x_n] = (-\infty,\ x]$
  - This implies

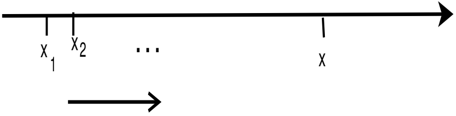
$$\lim_n P_X((-\infty,\ x_n]) = P_X(\lim_n (-\infty,\ x_n]) = P_X((-\infty,\ x])$$

  - This in turn implies

$$\lim_{x_n \downarrow x} F_X(x_n) = F_X(x)$$

  - Using the usual notation for right limit of a function, we can write $F_X(x^+) = F_X(x), \forall x$.

- F_X is right-continuous at all x.
- Next, let us look at the lefthand limits: $\lim_{x_n \uparrow x} F_X(x_n)$
- When $x_n \uparrow x$, the sequence of events $(-\infty,\ x_n]$ is monotone increasing



$$\lim_n (-\infty,\ x_n] = \cup_n (-\infty,\ x_n] = (-\infty,\ x)$$

- By sequential continuity of probability, we have

$$\lim_n P_X((-\infty,\ x_n]) = P_X(\lim_n (-\infty,\ x_n]\ ) = P_X(\ (-\infty,\ x)\ )$$

- Hence we get

$$F_X(x^-) \;\; = \;\; \lim_{x_n \uparrow x} F_X(x_n) = \lim_n P_X((-\infty,\ x_n]) = P_X(\ (-\infty,\ x)\ )$$

- Thus, at every x the left limit of F_X exists.

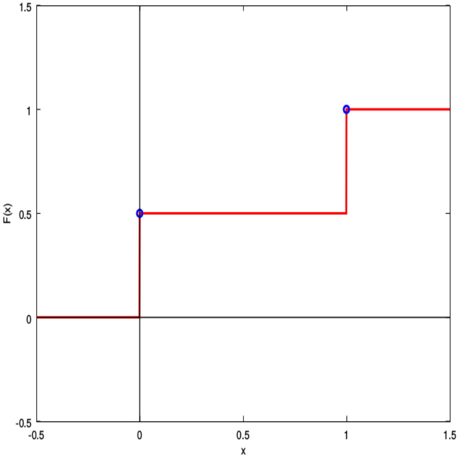
- F_X is right-continuous:
  $F_X(x^+) = F_X(x) = P_X(\ (-\infty,\ x]\ )$
- It has left limits: $F_X(x^-) = P_X(\ (-\infty,\ x)\ )$
- If $A \subset B$ then $P(B - A) = P(B) - P(A)$
- We have $(-\infty,\ x] - (-\infty,\ x) = \{x\}$. Hence

$$P_X(\ (-\infty,\ x]\ ) - P_X(\ (-\infty,\ x)\ ) = P_X(\{x\}) = P(\{\omega\ :\ X(\omega) = x\})$$

- Thus we get

$$F_X(x^+) - F_X(x^-) = P(\{\omega\ :\ X(\omega) = x\}) = P[X = x]$$

- When F_X is discontinuous at x the height of discontinuity is the probability that X takes that value.
- And, if F_X is continuous at x then $P[X = x] = 0$


1.5
1
0.5
0
-0.5
F(x)
-0.5
0
0.5
1
1.5
x

Distribution Functions

- Let X be a random variable.
- Its distribution function, $F_X : \Re \rightarrow \Re$ is given by $F_X(x) = P[X \leq x]$
- The distribution function satisfies
  1. $0 \leq F_X(x) \leq 1, \ \forall x$
  2. $F_X(-\infty) = 0; \ F_X(\infty) = 1$
  3. F_X is non-decreasing: $x_1 \leq x_2 \ \Rightarrow \ F_X(x_1) \leq F_X(x_2)$
  4. F_X is right continuous and has left-hand limits.
- We also have $F_X(x^+) - F_X(x^-) = P[X = x]$
- Any real-valued function of a real variable satisfying the above four properties would be a distribution function of some random variable.

- $F_X(x) = P[X \leq x] = P[X \in (-\infty,\ x]\ ]$
- Given F_X, we can, in principle, find $P[X \in B]$ for all Borel sets.
- In particular, for $a < b$,

$$
\begin{aligned}
P[a < X \leq b] &= P[X \in (a,\ b]\ ] \\
&= P[X \in (\ (-\infty,\ b] - (-\infty,\ a]\ )\ ] \\
&= P[X \in (-\infty,\ b]\ ] - P[X \in (-\infty,\ a]\ ] \\
&= F_X(b) - F_X(a)
\end{aligned}
$$

- There are two classes of random variables that we would study here.
- These are called discrete and continuous random variables.
- There can be random variables that are neither discrete nor continuous.
- But these two are important classes of random variables that we deal with in this course.
- Note that the distribution function is defined for **all** random variables.