IST1990 Probability And Statistics

- **Lecture 5**

- Cumulative Distribution Functions of Discrete R.V's.
- Continuous Random Variables
- Probability Density Function of Continuous R.V's.
- Cumulative Distribution Functions of Continuous R.V's.

Cumulative Distribution Function

For any random variable the possible outcomes of a random variable are real numbers. It is then often useful to describe the probability distribution using a different function...

Definition: Cumulative Distribution Function (c.d.f.)

The **cumulative distribution function** of a random variable X is denoted by $F(x) : \mathbb{R} \to [0,1]$, and is given by

$$F(x) = P(X \leq x), \qquad \text{where } x \in \mathbb{R}$$

Cumulative Distribution Function

Properties:

Let X be a discrete random variable X with p.m.f. given by $f(\cdot)$. For any $x \in \mathbb{R}$ we have

$$F(x) = P(X \leq x) = \sum_{i: x_i \leq x} f(x_i)$$

Furthermore $F(x)$ satisfies the following properties

a. F is increasing: if $x \leq y$ then $F(x) \leq F(y)$.
b. $F(x^+) = F(x)$ for $x \in \mathbb{R}$. Thus, F is *continuous from the right.*
c. $F(x^-) = \mathbb{P}(X < x)$ for $x \in \mathbb{R}$. Thus, F has *limits from the left.*
d. $F(-\infty) = 0$.
e. $F(\infty) = 1$.

$$F(x^+) = \lim_{t \downarrow x} F(t),\ F(x^-) = \lim_{t \uparrow x} F(t),\ F(\infty) = \lim_{t \to \infty} F(t),\ F(-\infty) = \lim_{t \to -\infty} F(t)$$

The following result shows how the distribution function can be used to compute the probability that X is in an interval →

Suppose that F(x) is the c.d.f. of a real-valued discrete random variable X. If a,b are Real numbers such that a<b, then

a. $\mathbb{P}(X = a) = F(a) - F(a^-)$

b. $\mathbb{P}(a < X \le b) = F(b) - F(a)$

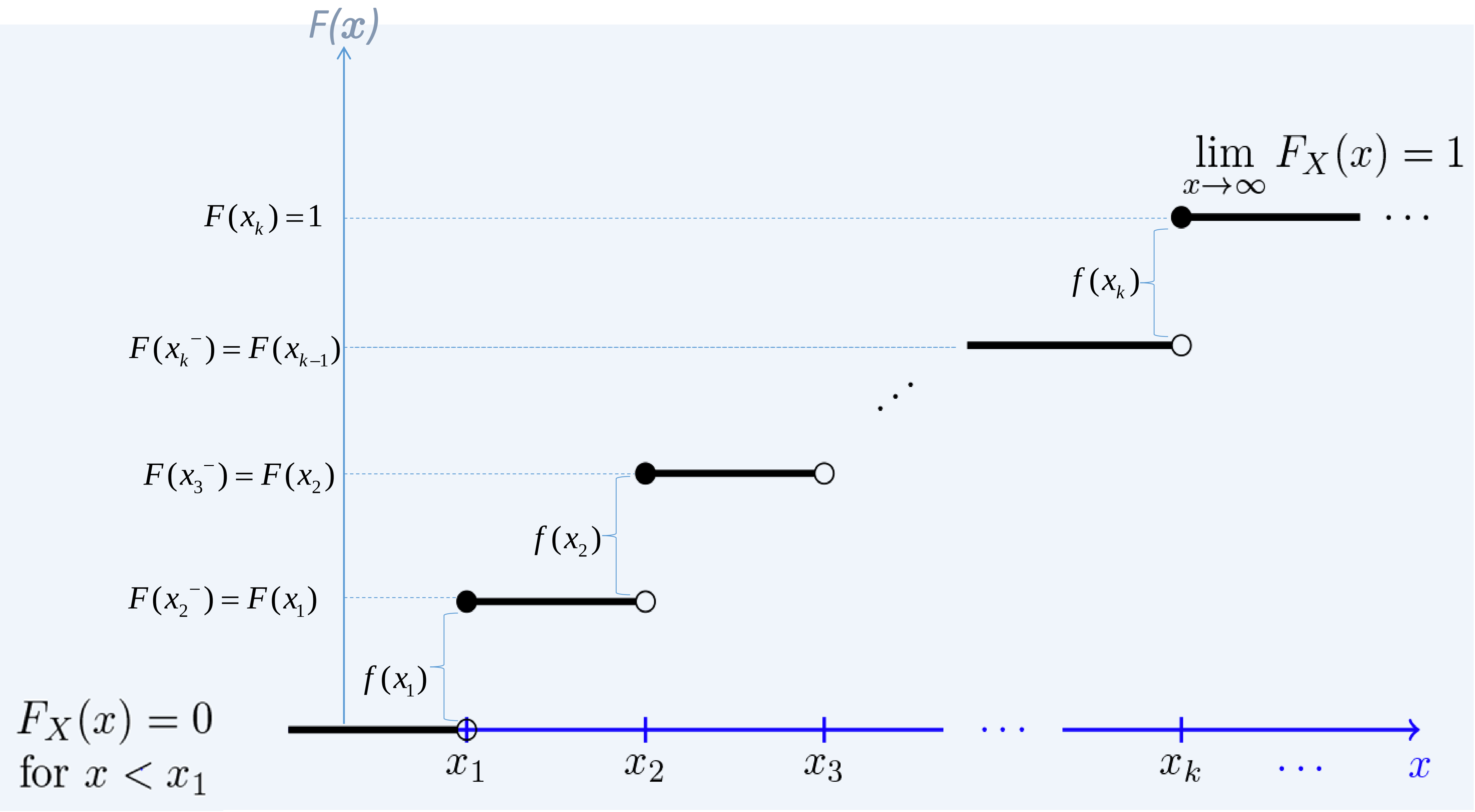
c. $\mathbb{P}(a < X < b) = F(b^-) - F(a)$

d. $\mathbb{P}(a \le X \le b) = F(b) - F(a^-)$

e. $\mathbb{P}(a \le X < b) = F(b^-) - F(a^-)$

Remember that: $F(x) = \mathbb{P}(X \le x), \quad x \in \mathbb{R}$ and $F(x^-) = \mathbb{P}(X < x)$ for $x \in \mathbb{R}$.

Note that: c.d.f is defined for any real number, and not only for the values X can take!.



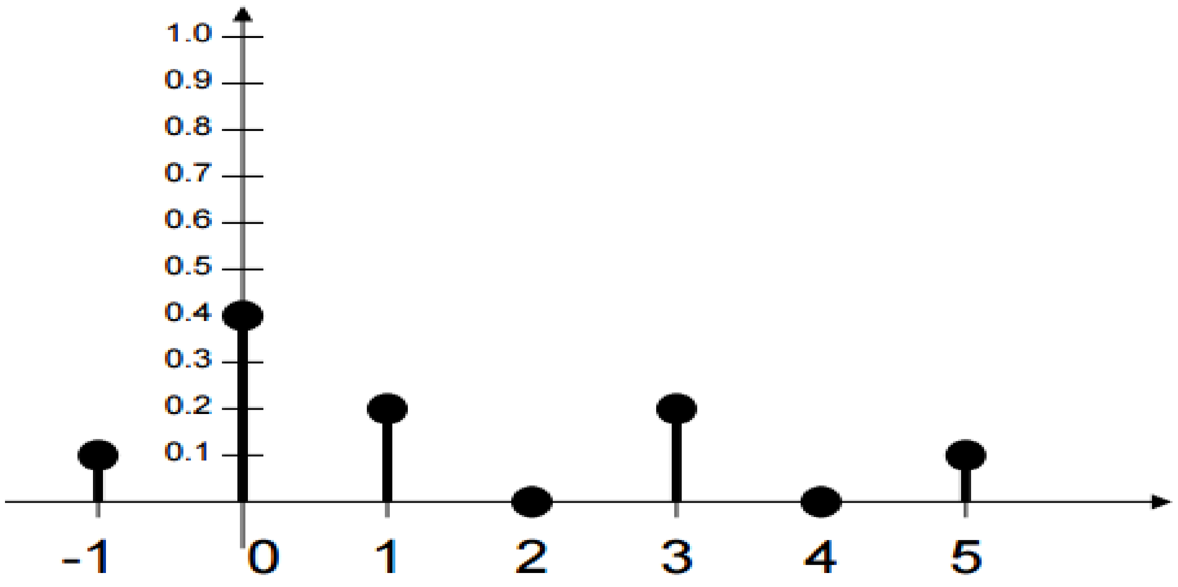
F(x)
$\lim_{x\to\infty} F_X(x) = 1$
$F(x_k) = 1$
$f(x_k)$
$F(x_k^-) = F(x_{k-1})$
$F(x_3^-) = F(x_2)$
$f(x_2)$
$F(x_2^-) = F(x_1)$
$f(x_1)$
$F_X(x) = 0$
for $x < x_1$
x_1
x_2
x_3
x_k
x

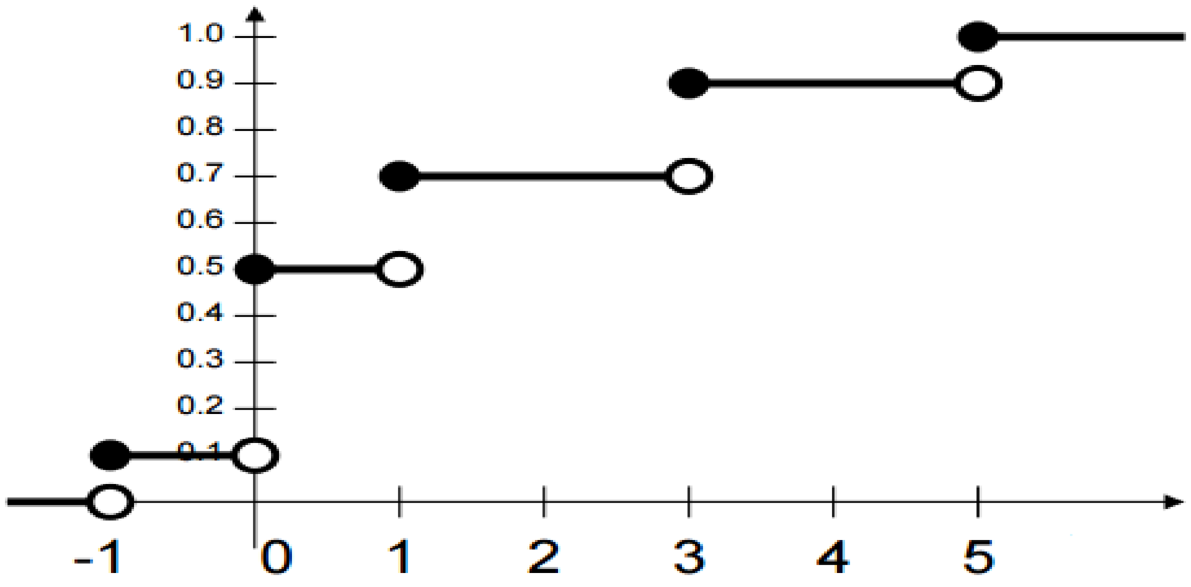
C.D.F. Example

Let X be a discrete random variable X taking values $\{-1, 0, 1, 3, 5\}$, and having p.m.f. given by

$$f(x) = \begin{cases} 0.1 & \text{if } x = -1 \\ 0.4 & \text{if } x = 0 \\ 0.2 & \text{if } x = 1 \\ 0.2 & \text{if } x = 3 \\ 0.1 & \text{if } x = 5 \end{cases} \qquad F(x) = \begin{cases} 0 & \text{if } x < -1 \\ 0.1 & \text{if } -1 \le x < 0 \\ 0.5 & \text{if } 0 \le x < 1 \\ 0.7 & \text{if } 1 \le x < 3 \\ 0.9 & \text{if } 3 \le x < 5 \\ 1 & \text{if } x \ge 5 \end{cases}$$

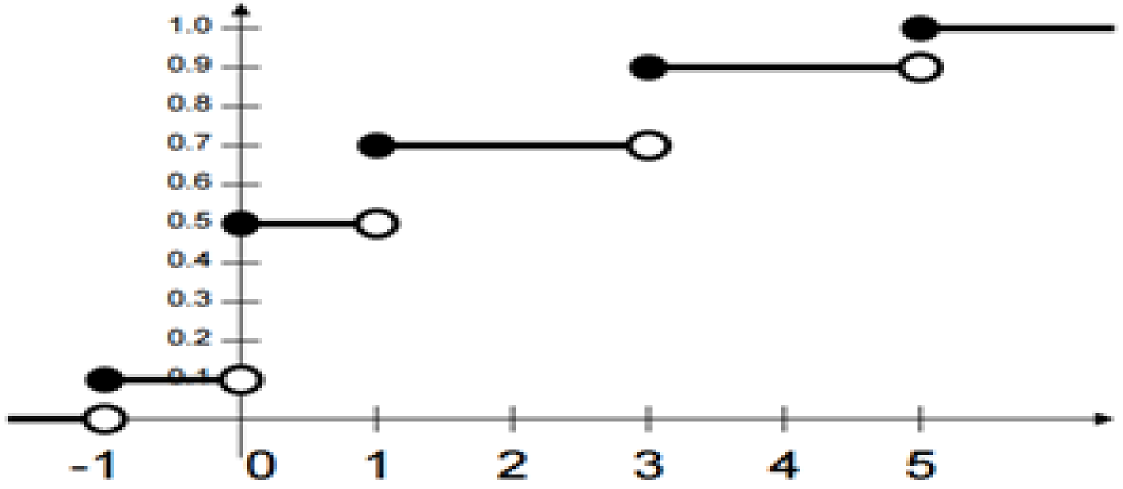
Let's plot the p.m.f. and c.d.f.:





C.D.F. Example

$$F(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ 0.1 & \text{if} \quad -1 \leq x < 0 \\ 0.5 & \text{if} \quad 0 \leq x < 1 \\ 0.7 & \text{if} \quad 1 \leq x < 3 \\ 0.9 & \text{if} \quad 3 \leq x < 5 \\ 1 & \text{if} \quad x \geq 5 \end{cases}$$



$$P(X \leq 4) = 0.9$$

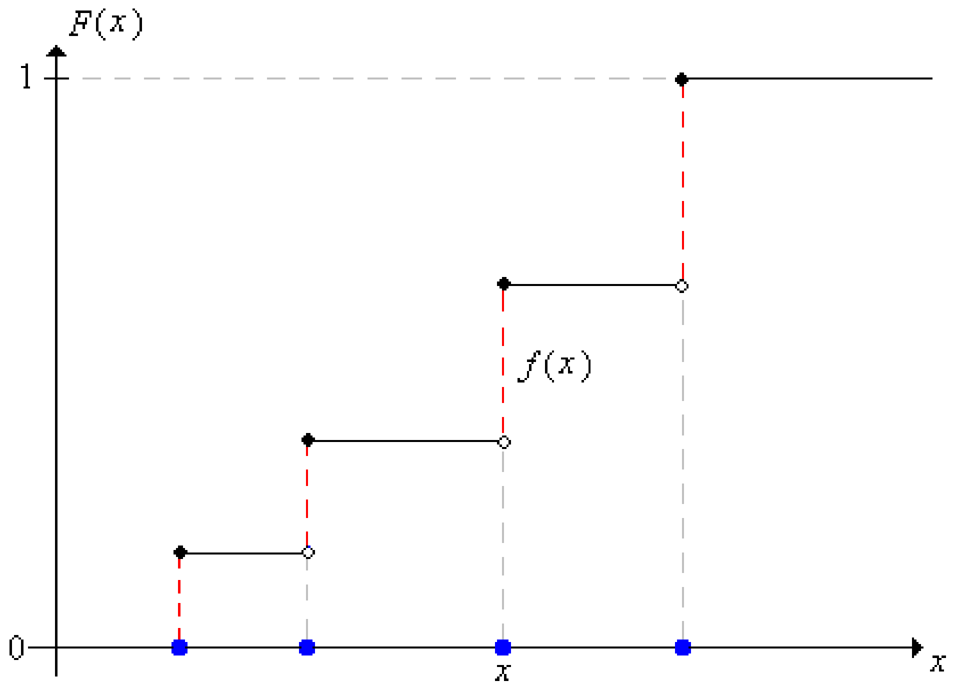
$$P(1 < X \leq 4) = P(X \leq 4) - P(X \leq 1) = 0.9 - 0.7 = 0.2$$

$$\begin{aligned} P(1 \leq X \leq 4) = P(X \leq 4) - P(X < 1) &= P(X \leq 4) - P(X \leq 0.9999999) \\ &= 0.9 - 0.5 = 0.4 \end{aligned}$$

From c.d.f to p.m.f

There is a simple relationships between the cumulative distribution function and the probability mass function. If c.d.f is given, then we can easily obtain the p.m.f of discrete random variable X:

$$f(x) = F(x) - F(x^-) \text{ for } x \in S$$



$$F(x) = \begin{cases} 0 & x < 0 \\ 0,10 & 0 \le x < 1 \\ 0,25 & 1 \le x < 2 \\ 0,45 & 2 \le x < 3 \\ 0,75 & 3 \le x < 4 \\ 0,9 & 4 \le x < 5 \\ 1 & 5 \le x \end{cases}$$

$$f(x) = \begin{cases} 0,10 & x = 0 \\ 0,25 - 0,10 = 0,15 & x = 1 \\ 0,45 - 0,25 = 0,2 & x = 2 \\ 0,75 - 0,45 = 0,3 & x = 3 \\ 0,90 - 0,75 = 0,15 & x = 4 \\ 1.00 - 0,90 = 0,1 & x = 5 \end{cases}$$

Continuous Random Variables - Motivation

Discrete random variables are not the end of the story...

Let X the be a random variable representing the temperature in a museum room (in centigrade degrees)

$$P(X \leq 3000) = 1$$
$$P(10 \leq X \leq 29) = 0.99$$
$$P(X = 19) = 0$$
$$P(X = 19.01) = 0$$
$$P(X = 29.035) = 0$$

It seems the probability of X taking any specific value is always be equal to zero, but the probability of X being in an interval is often strictly positive..

Continuous Random Variables - Motivation

It seems the probability that X is in some small interval should be small (but not zero)…

$$P(22.0 \leq X < 22.1) = a, \text{ where } a \text{ is small}$$

The smaller the interval, the smaller the probability, so perhaps we can write it as

$$P(22.0 \leq X < 22.0 + \Delta) = c\Delta, \quad c \geq 0 \text{ and } \Delta \text{ small}$$

So if we want to compute the probability that X is in an arbitrary interval we can perhaps break it into small pieces…

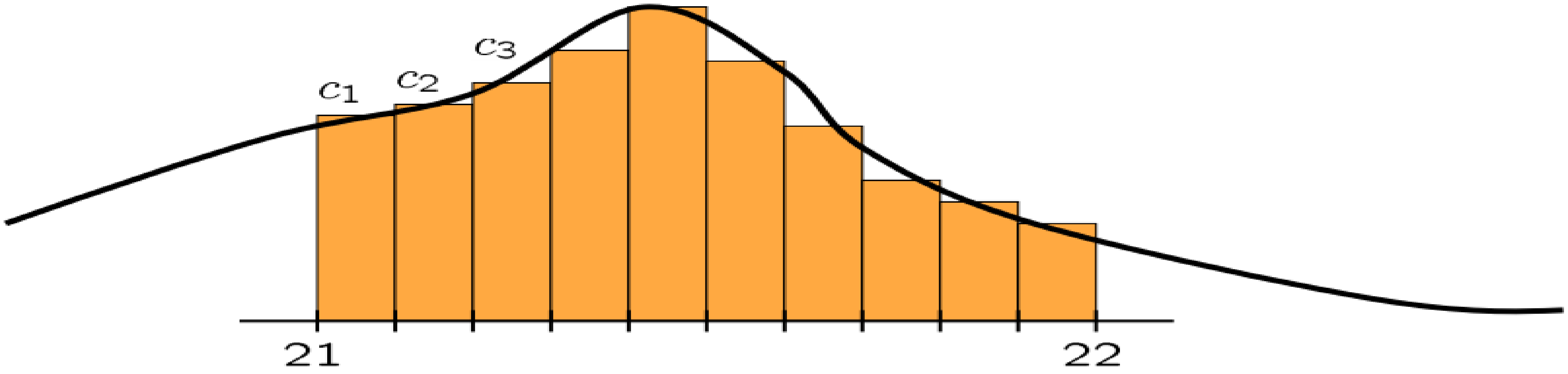
$$\begin{aligned} & P(21 \leq X \leq 22) \\ = \quad & P\left(X \in [21, 21.1) \cup [21.1, 21.2) \cup \cdots \cup [22.9, 22)\right) \end{aligned}$$

Probabilities as Areas

$$P(21 \le X \le 22) = \sum_{i=1}^{10} P(X \in [21 + 0.1(i-1), 21 + 0.1i))$$

$$= \sum_{i=1}^{10} c_i 0.1$$

area of a rectangle with sides 0.1 and c_i



So computing these probabilities seems to be essentially the same as computing areas under functions !!!

Probability Density Functions

For continuous random variables it doesn't make sense to talk about probability mass functions, as the probability of such variables to take a specific value is always zero. However, we can define a **density**, which plays a similar role...

Definition: Probability Density Function (p.d.f.)

For a continuous random variable X, a **probability density function** is a function $f(x)$ such that

(i) $f(x) \geq 0$ for all $x \in \mathbb{R}$

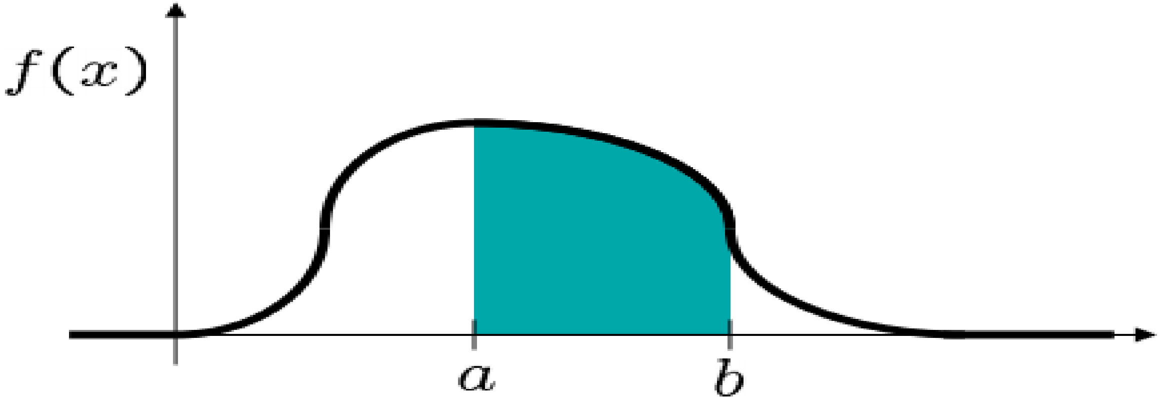
(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

(iii) $P(a \leq X \leq b) = \int_a^b f(x)dx$ for any $a, b \in \mathbb{R}$ such that $a \leq b$

A valid p.d.f. must always satisfy (i) and (ii) !

Probability Density Functions

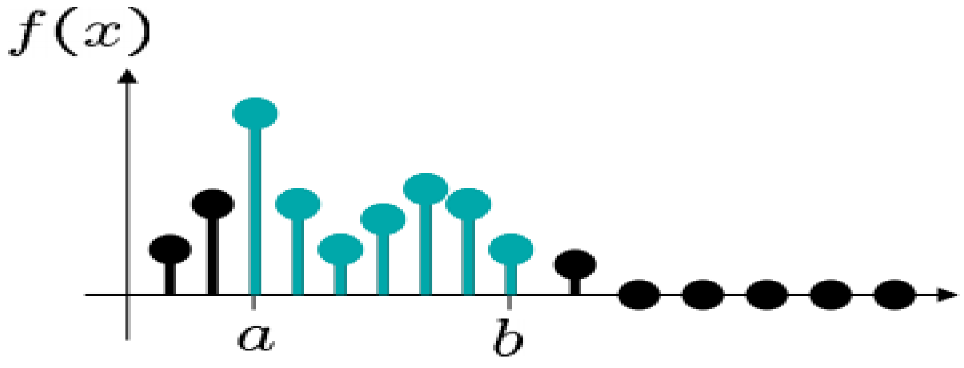
Probability Density Function
(continuous random variables)



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

= area under the curve between a and b

Probability Mass Function
(discrete random variables)



$$P(a \leq X \leq b) = \sum_{x=a}^{b} f(x)$$

= sum of point masses from a to b

There is a strong parallel between discrete and continuous random variables

Probability Density Functions

Important Note:

For a continuous random variable X we have, for any a and b

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

This is because $P(X = x) = 0$ for any $x \in \mathbb{R}$.

It is often difficult to get your head around the fact that despite the fact the probability that a continuous random variable takes any specific value is always zero, but that the probability that the random variable takes a value within an interval is generally non-zero...

Example

Let the random variable X denote the time it takes (in days) for a hard drive to fail. We have reason to believe that this random variable has the following density

$$f(x) = \begin{cases} \frac{1}{5000} e^{-x/5000} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

What is the probability that the hard drive fails in the first three years?

$$\begin{aligned} P(X \leq 365 \times 3) &= P(X \leq 1095) = \int_{-\infty}^{1095} f(x)dx \\ &= \int_{0}^{1095} f(x)dx = \int_{0}^{1095} \frac{1}{5000} e^{-x/5000} dx \\ &= -e^{-x/5000} \Big|_{x=0}^{1095} \\ &= (-e^{-1095/5000}) - (-e^{-0/5000}) \\ &= 1 - e^{-1095/5000} \approx 0.1967 \end{aligned}$$

Cumulative Distribution Function

We already saw that we can consider also different descriptions of a random variable, namely we defined in the first lecture the

Definition: Cumulative Distribution Function (c.d.f.)

The **cumulative distribution function** of a random variable X is denoted by $F(x) : \mathbb{R} \to [0, 1]$, and is given by

$$F(x) = P(X \leq x), \qquad \text{where } x \in \mathbb{R}$$

This seemingly simple object is extremely powerful, as all the information contained in the probability density function is also contained in the c.d.f.

This is EXACTLY the same definition we had for discrete random variables!!!

Cumulative Distribution Function

Let X be a **continuous random variable** with density $f(x)$. For any $x \in \mathbb{R}$ we have

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \ .$$

Furthermore any c.d.f. $F(x)$ satisfies the following properties

a. F is increasing: if $x \leq y$ then $F(x) \leq F(y)$.
b. $F(x^+) = F(x)$ for $x \in \mathbb{R}$. Thus, F is *continuous from the right.*
c. $F(x^-) = \mathbb{P}(X < x)$ for $x \in \mathbb{R}$. Thus, F has *limits from the left.*
d. $F(-\infty) = 0$.
e. $F(\infty) = 1$.

All the information in the density is also contained in the cumulative distribution function (also referred to simply as the distribution function)

From c.d.f to p.d.f

If $F(x)$ is the cummulative distribution function of a continuous random variable then the corresponding derivative

$$\frac{d}{dx}F(x) = F'(x)$$

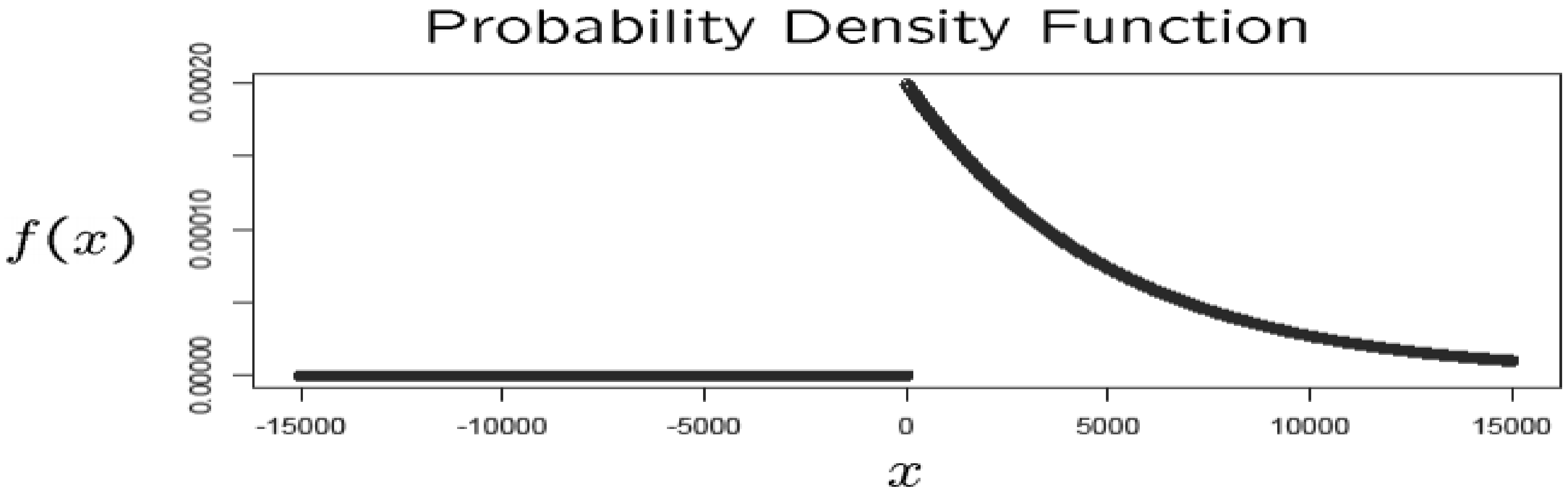
is precisely the probability density function.

Note that the cumulative distribution function might not be differentiable at all points. However, this is not a bit problem, provided the derivative exists for all but a finite number of points (all those densities are equivalent).

C.D.F. Example

Recall the density of our previous example

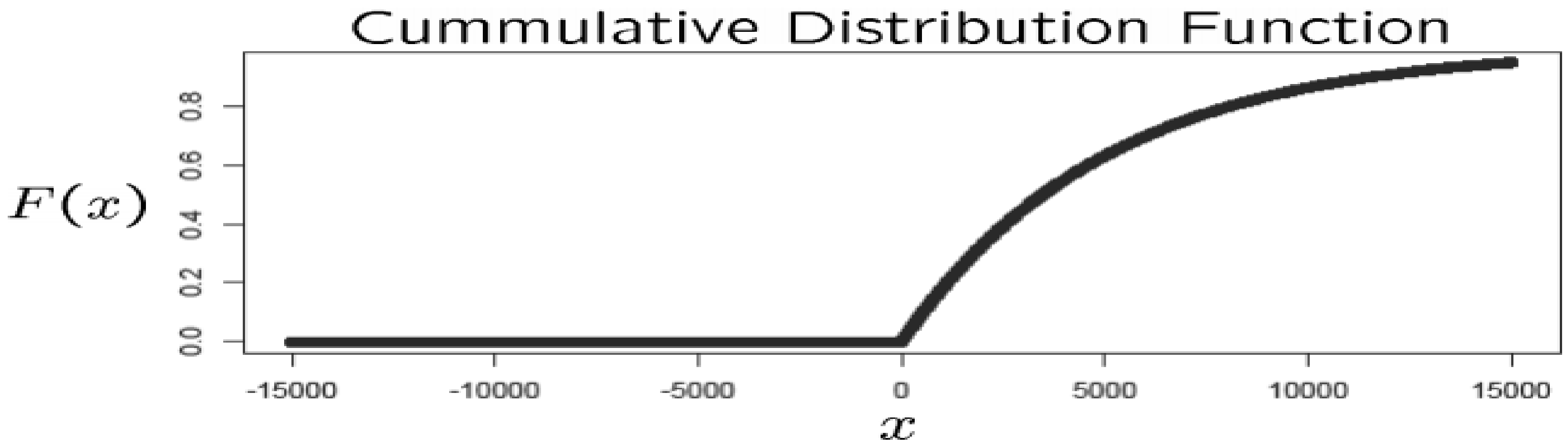
$$f(x) = \begin{cases} \frac{1}{5000} e^{-x/5000} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x/5000} & \text{if } x \geq 0 \end{cases}$$

C.D.F. Example

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x/5000} & \text{if } x \ge 0 \end{cases}$$



$$\begin{aligned} P(365 \le X \le 1095) &= F(1095) - F(365) \\ &= e^{-365/5000} - e^{-1095/5000} \approx 0.1263 \end{aligned}$$

Note: the area under the distribution function is NOT a probability!!!

EXAMPLE: Suppose that the error in the reaction temperature for a controlled laboratory experiment is a continuous random variable *X* having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that *f*(*x*) is a density function.

$$\int_{-\infty}^{\infty} f(x)\ dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9} |_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1.$$

(b) Find $P(0 < X \leq 1)$.

$$P(0 < X \leq 1) = \int_0^1 \frac{x^2}{3} dx = \left. \frac{x^3}{9} \right|_0^1 = \frac{1}{9}.$$

Example: Consider the density function

$$f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Evaluate k.

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

$$\int_{-\infty}^{+\infty} kx^{1/2}dx = \int_0^1 kx^{1/2}dx = k\frac{2}{3}\,x^{3/2}|_0^1 = k\frac{2}{3}\,(1)^{3/2} - k\frac{2}{3}\,(0)^{3/2} = k\frac{2}{3}$$

$$k\frac{2}{3} = 1 \rightarrow k = \frac{3}{2}$$

(b) Find $F(x)$ and use it to evaluate $P(0.3 < X < 0.6)$

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

$$F(x) = \int_0^x \frac{3}{2}t^{1/2}dt = \frac{3}{2} * \frac{2}{3}\,t^{3/2}|_0^x = x^{3/2}$$

$$P(0.3 < X \le 0.6) = F(0.6) - F(0.3) = (0.6)^{3/2} - (0.3)^{3/2} = 0.30$$

HOMEWORK: Consider the following probability density function

$$f(y)=\begin{cases} k * e^{-y/4} & y \geq 0, \\ 0 & elsewhere. \end{cases}$$

Find;

a) value of k.

b) $P(Y < 2)$

c) $P(Y \leq 2)$

d) $P(Y = 2)$

e) $P(0 < Y < 2)$