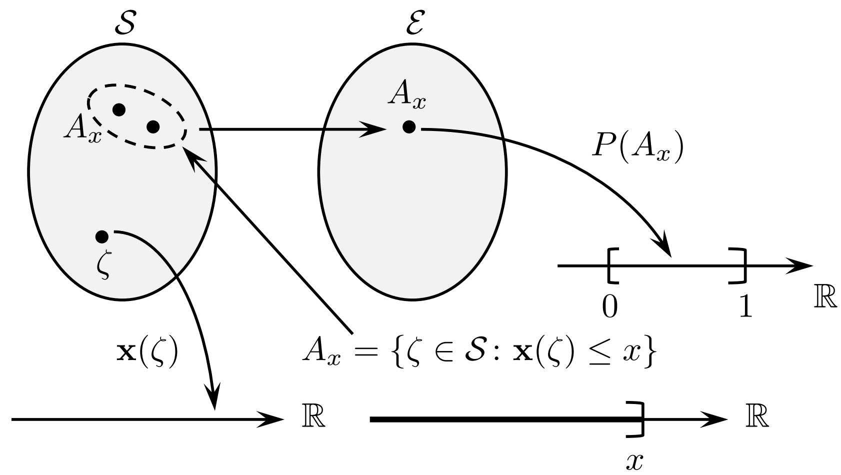
4-1 The Concept of a Random Variable



Definition

- A *random variable* is a *map* from $\mathcal{S}$ to $\mathbb{R}$.
- For every outcome $\zeta \in \mathcal{S}$, $\mathbf{x}(\zeta)$ is a real number.
- When it is understood $\mathbf{x}(\zeta)$ is a random variable, it is customary to drop the notational dependence on ζ and write "$\mathbf{x}$ is a random variable."
- The *event* generated by the random variable $\mathbf{x}(\zeta)$ is

$$\{\zeta \in \mathcal{S} \colon \mathbf{x}(\zeta) \le x\}.$$

- The short-hand notation $\mathbf{x} \le x$ is usually used:

$$\mathbf{x} \le x \quad \text{means} \quad \{\zeta \in \mathcal{S} \colon \mathbf{x}(\zeta) \le x\}$$

"Thus $\{\mathbf{x} \le x\}$ is not a set of numbers but a *set of experimental outcomes*."

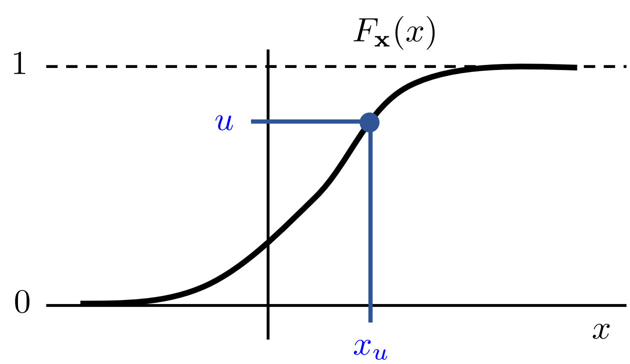
4-2 Distribution and Density Functions

Definitions

- The elements of the set $\mathcal{S}$ contained in the event $\{\zeta\colon \mathbf{x}(\zeta) \leq x\}$ change as the number x takes various values.
- The probability $P(\{\zeta\colon \mathbf{x}(\zeta) \leq x\}) = P(\mathbf{x} \leq x)$ is, therefore, a number that depends on x.
- This number is denoted $F_{\mathbf{x}}(x)$ and is called the (*cumulative*) *distribution function*.
- The distribution function $F_{\mathbf{x}}(x)$ is defined for every $-\infty < x < \infty$.

Percentiles
The u percentile of a rnaom variable $\mathbf{x}$ is the smallest number x_u such that

$$u = P(\mathbf{x} \leq x_u) = F_{\mathbf{x}}(x_u).$$



Properties

1. End points.

$$F_{\mathbf{x}}(+\infty) = 1 \quad F_{\mathbf{x}}(-\infty) = 0.$$

2. $F_{\mathbf{x}}(x)$ is a nondecreasing function of x.

   if $x_1 < x_2$ then $F_{\mathbf{x}}(x_1) \le F_{\mathbf{x}}(x_2)$

3. If $F_{\mathbf{x}}(x_0) = 0$ then $F_{\mathbf{x}}(x) = 0$ for every $x \le x_0$.

4. $P(\mathbf{x} > x) = 1 - F_{\mathbf{x}}(x)$.

5. $F_{\mathbf{x}}(x)$ is continuous from the right:

$$\lim_{\epsilon \to 0} F_{\mathbf{x}}(x + \epsilon) = F_{\mathbf{x}}(x).$$

$F_{\mathbf{x}}(x)$ is always continuous from the right.

6. $P(x_1 < \mathbf{x} \le x_2) = F_{\mathbf{x}}(x_2) - F_{\mathbf{x}}(x_1)$

7. $P(\mathbf{x} = x) = F_{\mathbf{x}}(x) - \lim_{\epsilon \to 0} F_{\mathbf{x}}(x - \epsilon)$.

8. $P(x_1 \le \mathbf{x} \le x_2) = F_{\mathbf{x}}(x_2) - \lim_{\epsilon \to 0} F_{\mathbf{x}}(x_1 - \epsilon)$.

$F_{\mathbf{x}}(x)$ is does not have to be continuous from the left.

If $F_{\mathbf{x}}(x)$ is not continuous, then the left and right limits are different:

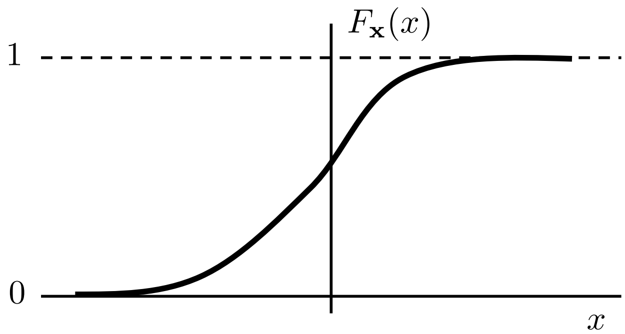
$$\lim_{\epsilon \to 0} F_{\mathbf{x}}(x + \epsilon) - \lim_{\epsilon \to 0} F_{\mathbf{x}}(x - \epsilon) > 0$$

The difference is > 0 because of Property 2.

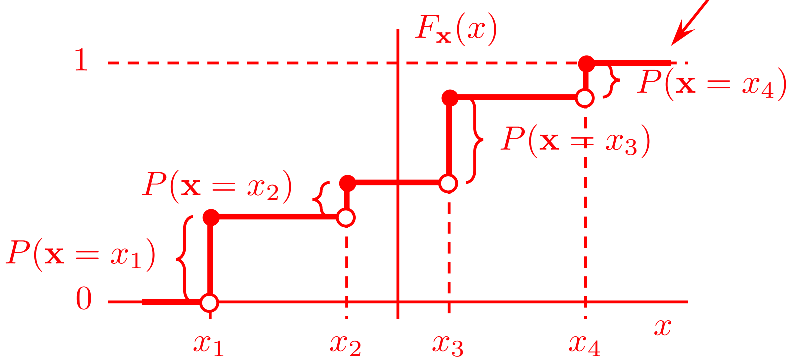
From Property 7, the difference is $P(\mathbf{x} = x)$

The only discontinuities of $F_{\mathbf{x}}(x)$ are of the jump type.

continuous distribution function



discontinuous distribution function



$F_{\mathbf{x}}(x)$ is continuous: $P(\mathbf{x} = x) = 0$

$F_{\mathbf{x}}(x)$ is discontinuous at x_1, x_2, x_3, x_4:

$$P(\mathbf{x} = x) = \begin{cases} 0 & x \notin \{x_1, x_2, x_3, x_4\} \\ > 0 & x \in \{x_1, x_2, x_3, x_4\} \end{cases}$$

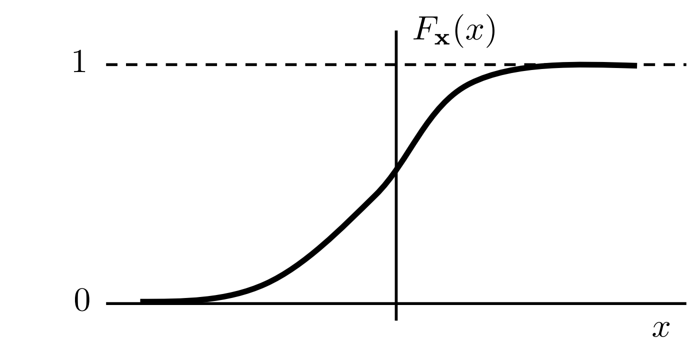
Probability Density Function (pdf)

Definition $f_{\mathbf{x}}(x) = \dfrac{d}{dx} F_{\mathbf{x}}(x)$

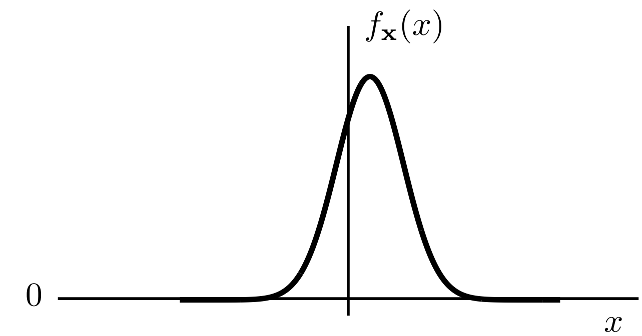
Properties

1. $f_{\mathbf{x}}(x) > 0$ for all x.
2. $F_{\mathbf{x}}(x) = \displaystyle\int_{-\infty}^{x} f_{\mathbf{x}}(u) du.$
3. $\displaystyle\int_{-\infty}^{\infty} f_{\mathbf{x}}(u) du = 1.$
4. $P(x_1 < \mathbf{x} \le x_2) = \displaystyle\int_{x_1}^{x_2} f_{\mathbf{x}}(u) du.$

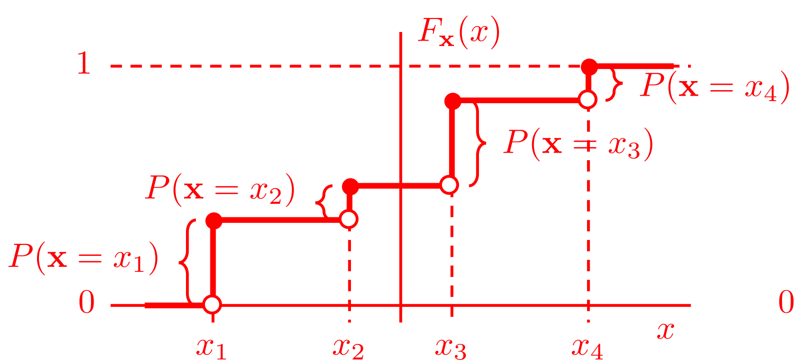
continuous distribution function



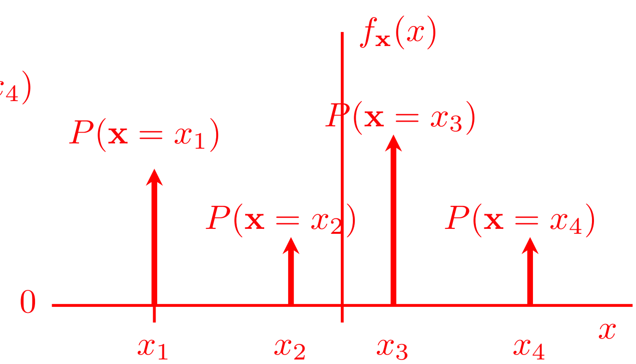
“continuous random variable”



discontinuous distribution function



“discrete random variable”



Probability Density Function: $f_{\mathbf{x}}(x) = \displaystyle\sum_{i=1}^{4} P(\mathbf{x} = x_i)\delta(x - x_i)$

Probability Mass Function: $P(\mathbf{x} = x_1),\ P(\mathbf{x} = x_2),\ P(\mathbf{x} = x_3),\ P(\mathbf{x} = x_4)$