Probability Theory for EOR

Random variables and their distributions II

Distributions: probability values of events associated with random variables

Real-valued r.v. X maps elements from S to real numbers!
Now we can work with functions/numbers:

$$X : S \mapsto \mathbb{R}$$

- We can now use $\{X \in C\}$ with some subsets $C \subseteq \mathbb{R}$ to represent different random events (not all subsets C generate well-defined events, and usually we focus on different intervals: e.g., $(-\infty, a], [a, b] \cap (a, e) \cdots$).

  The distribution of a given r.v. needs to specify all probability values of all those events generated by the r.v.:

  $$P : \{\{X \in C\} : \text{some subsets } C \subseteq \mathbb{R}\} \mapsto [0, 1].$$

  For example, a distribution needs to be able to answer $P(X \in [a, b] \cup [c, d]), P(X = e), \cdots$

- Though there are so many associated events, **we only need to know probability values of some selected events** and the rest probability values can be inferred from these values.

Definition (CDF of real-valued r.v.)

The **cumulative distribution function (CDF)** of an **real-valued** r.v., $X : S \to \mathbb{R}$, is the function, usually denoted by F_X, which maps numbers to numbers within [0,1].

$$F_X(x) = P(X \leq x).$$

- We only need to know probability values of the following types of events: $\{X \leq x\}, x \in \mathbb{R}$.

Properties:

- **Increasing function from zero to one:**
  $0 \leq F_X(x_1) \leq F_X(x_2) \leq 1, \forall -\infty < x_1 \leq x_2 < \infty.$
  (a): $\lim_{x \to -\infty} F_X(x) = 0$; think about $\emptyset$!
  (b): $\lim_{x \to \infty} F_X(x) = 1$; think about S!
- **Right-continuous:** $F_X(a) = \lim_{x \downarrow a} F(x)$.

Simple example revisits: waiting time

Customers come in according to a Poisson process such that:

Poisson arrivals.
The number of arrivals that occur in an interval of length t, N_t, is a Pois(λt) r.v.;

Independence condition.
The number of arrivals that occur in disjoint intervals are independent from each other.

How long do you need to wait until the first customer?

- Denote T_1 the time until the first arrival.
- The CDF of T_1 then

$$F_{T_1}(t) = P(T_1 \leq t) = 1 - P(T_1 > t) = 1 - P(N_t = 0) = 1 - e^{-\lambda t}$$

$$= \int_0^t \lambda e^{-\lambda s} ds.$$

From CDF to PDF

Definition (CDF of real-valued r.v.)

The **cumulative distribution function (CDF)** of an **real-valued** r.v., $X : S \to \mathbb{R}$, is the function, usually denoted by F_X:

$$F_X(x) = P(X \leq x).$$

Definition (PDF of continuous real-valued r.v.)

The **probability density function (PDF)** of a **continuous** real-valued r.v. is a non-negative function f_X on the real line such that via the Riemann integral:

$$\int_{-\infty}^{x} f_X(s)ds = P(X \leq x).$$

For a **continuous** random variable with differentiable CDF F_X, conventionally, $f_X(x) = F'_X(x)$.

Definition (PDF of continuous real-valued r.v.)

The **probability density function (PDF)** of a **continuous** real-valued r.v. is a non-negative function f_X on the real line such that via the Riemann integral:

$$\int_{-\infty}^{x} f_X(s)ds = P(X \leq x).$$

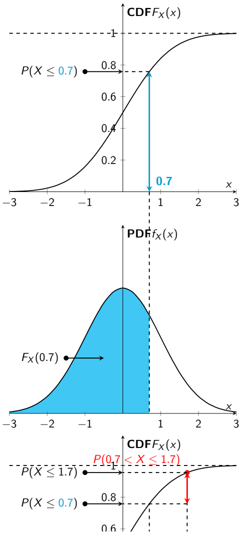
For a **continuous** random variable with differentiable CDF F_X, conventionally, $f_X(x) = F'_X(x)$.

- f_X can be used to calculate the CDF function values, and thus f_X is also a way describing the distribution of **continuous real-valued** random variables.
  Properties:
- **Non-negativity**: $f_X(x) > 0$ if $x \in \mathrm{Supp}(X)$; $f_X(x) = 0$ otherwise.
- **Integrate to 1.** $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

E.g., $\lambda = 3, 3e^{-3t} > 0, t > 0, 3 * e^{-0.03} > 2.911$.

The distribution tell us almost all we need to know about a random variable.

Example I: standard normal distribution, $N(0,1)$, CDF PDF



The standard normal distribution has PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

and CDF

$$F(x) = \int_{-\infty}^{x} f(s)ds.$$

Suppose X follows the standard normal distribution.

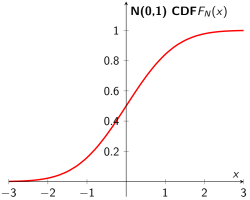
- $P(X \leq 0.7)$?

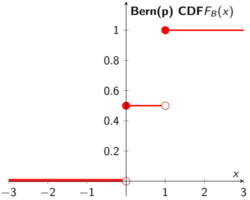
$$P(X \leq 0.7) = F_X(0.7);$$
$$P(X \leq 0.7) = \int_{-\infty}^{0.7} f_X(s)ds.$$

- $P(0.7 < X \leq 1.7)$?

Example II: compare standard normal distribution $N(0,1)$ with Bernoulli distribution

Bern(p) distribution





The $N(0,1)$ distribution has CDF

$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds.$

- **No jumps**, continuous function.
- $F'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} > 0$ **at some open intervals.**
- **There exists** f (e.g., $F'(x)$) such that $F(x) = \int_{-\infty}^{x} f(s)ds$.

The Bern(p) distribution has CDF
$F(x) = 0I_{(-\infty,0)}(x) + 1/2I_{[0,1)} + 1I_{[1,\infty)}$.

- **Jumps** at 0, and 1.
- $F'(x) = 0$ **almost everywhere**, and not well-defined at 0 and 1.
- **No such** f such that F can be rewritten as a Riemann integral of f.