Probability Theory for EOR

Random variables (continuous)

Random variables are functions

Definition (Random variables (r.v.'s) (preliminary))

A random variable (r.v.) is a function maps elements in sample space S to numbers in $\mathcal{C}$, e.g., $\mathcal{C} \subseteq \mathbb{R}$.

There are functions from a sample space to a set of numbers that can be listed with natural numbers: $x_1, x_2, x_3, \cdots$ (which contains no non-empty open subsets/intervals).

Definition (Discrete random variable)

A random variable X is a **discrete** random variable if there exists a set C with at most countably many numbers (so you can index all the elements within C using a set of natural numbers):

$$P(\{X \in C\}) = 1, \text{ or equivalently, } P(X \in C) = 1.$$

The **support** of a **discrete** random variable is the set of all numbers, x's, such that $P(X = x) > 0$ **(PMF)**:

$$supp(X) = \{x \in C : P(X = x) > 0\}.$$

The support of a discrete random variable has empty interior (it contains no open intervals at all).

Continuous random variable

There are functions from a sample space to a set of numbers that contains non-empty open sets: e.g., random variables could take any positive values from the interval $(0, \infty)$.

Example: 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 is a Pois(λt) r.v.;

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

The distribution of the waiting time until the first customer?

Example: waiting time

The distribution of the waiting time until the first customer?

- Denote N_t the number of arrivals within $[0, t]$, then N_t follows Pois(λt).
- Denote T_1 the time until the first arrival.
- Note that $\{T_1 > t\} = \{N_t = 0\}$.
- $P(T_1 > t) = P(N_t = 0) = e^{-\lambda t}$.
- The CDF of T_1 then

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

- **Note that T_1 can take any non-negative values, and its CDF is differentiable which can be expressed as a Riemann integral.**

Example: waiting time

We have two ways to specify the distribution of T_1.

$$F_{T_1}(t) = 1 - e^{-\lambda t} = \int_0^t f_{T_1}(t)dt.$$

$$f_{T_1}(t) = F'_{T_1}(t) = \lambda e^{-\lambda t}, t \in (0, \infty).$$

- f_{T_1} is called the **probability density function (PDF)** of T_1.
- T_1 may take any value t where $f_{T_1}(t) > 0$:$(0, \infty)$.
- Integrate the (probability) **density** via the Riemann integral we get the (probability) **value**.
- If we work with Riemann integral, we can not find such f for discrete random variables.

  E.g., For a Bern(p) distributed X: $F_X(t) = \begin{cases} 0; & t < 0 \\ 1/2; & 0 \leq t < 1 \\ 1; & t \geq 1 \end{cases}$.

Definition (Continuous random variable)

A **real-valued** random variable X is a **continuous** random variable if there exists a non-negative function f_X such that:

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

If F_X is differentiable, we can choose $f_X(x) = F'_X(x)$.

The **support** of a **continuous** random variable with differentialble CDF F_X, conventionally, is the set of all numbers, x's, such that $f_X(x) = F'_X(x) > 0$ **(PDF)**:

$$supp(X) = \{x \in C : F'_X(x) > 0\}.$$

The support of a continuous random variable has **non-empty interior** (it contains open intervals): e.g., $(0, \infty)$.

- A **real-valued** random variable X is a **discrete** random variable if there exists a set C with at most countably many numbers (so you can index all the elements within C using a set of natural numbers):

  $$P(\{X \in C\}) = 1, \text{ or equivalently, } P(X \in C) = 1.$$

  The **support** of a **discrete real-valued** random variable is the set of all numbers, x's, such that $P(X = x) > 0$ **(PMF)**:

  $$supp(X) = \{x \in C : P(X = x) > 0\}.$$

  The support of a discrete random variable has **empty interior** (it contains no open intervals at all): e.g., $\{0, 1, 2, 3, \cdots\}$.

- A **real-valued** random variable X is a **continuous** random variable if there exists a non-negative function f_X such that:

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

  If F_X is differentiable, we can choose $f_X(x) = F_X'(x)$.

  The **support** of a **continuous** random variable with differentialble CDF F_X, conventionally, is the set of all numbers, x's, such that $f_X(x) = F_X'(x) > 0$ **(PDF)**:

  $$supp(X) = \{x \in C : F_X'(x) > 0\}.$$

  The support of a continuous random variable has **non-empty interior** (it contains open intervals): e.g., $(0, 1)$.