Lecture 9 - Continuous random variables I

Motivation

- We often deal with variables which take uncountably many values
  - Value of voltage in an electric network
  - Waiting time for the arrival of a bus
  - Lifespan of a hard drive
  - Height of a randomly selected person
- Probabilities of taking any particular value by such a variable are zero.
- How can we handle a probability distribution of such variables?

Continuous random variable

- A random variable $X : \Omega \to \mathbb{R}$ is called continuous, if there exists a function $f : \mathbb{R} \to \mathbb{R}_+$, such that for any $A \subseteq \mathbb{R}$ it holds:

$$P(X \in A) = \int_A f(x)dx$$

- Function f is called probability density function of a random variable X.
- The probability density function has the following properties:
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
  - $f(x) \geq 0$ for all x
  - It doesn't need to be smaller or equal to 1 everywhere, it can actually exceed 1 (or even go to infinity)

Uniform distribution

- A continuous random variable X has uniform distribution on interval $[a, b]$ (denoted $X \sim \text{Unif}[a, b]$) if:

$$f(x) = \left\{ \begin{array}{ll} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{array} \right\}$$

- For $a \leq c \leq d \leq b$:

$$P(c \leq X \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$$

Exponential distribution

- A continuous random variable X has exponential distribution with parameter $\lambda > 0$ (denoted $X \sim \text{Exp}(\lambda)$) if:

$$f(x) = \left\{ \begin{array}{ll} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{array} \right\}$$

- It is a continuous version of a geometric distribution.
- $\int \lambda e^{-\lambda x} dx = -e^{-\lambda x}$
- Exponential distribution exhibits the memorylessness property.

Cumulative distribution function

- A cumulative distribution function if a random variable X is defined as:

$$F(x) = P(X \leq x)$$

- For a continuous random variable:

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

- It is true that $F'(x) = f(x)$

Distribution of a function of continuous random variable

- X — continuous random variable with density $f_X(x)$ taking values in $[a, b]$
- $Y = g(X)$ — function of the random variable, taking values in $[c, d]$ where $g : [a, b] \to [c, d]$ is differentiable and invertible
- Then the density $f_Y(y)$ of Y can be calculated:

$$f_Y(y) = f_X(h(y))|h'(y)|$$

- For $y \in [c, d]$ and where $h = g^{-1}$ is the inverse function of g.
- Example, let $X \sim \mathrm{Unif}[0, 1]$, calculate the density f_Y of $Y = X^2$
    - $f_X(x) = 1$ for $0 \le x \le 1$
    - $g(x) = x^2$, so $h(y) = \sqrt{y}$.
    - $h'(y) = \frac{1}{2\sqrt{y}}$
    - Therefore $f_Y(y) = 1 \cdot \frac{1}{2\sqrt{y}}$ for $0 \le y \le 1$
    - Explanation: Lets say that we want to calculate the probability that X is between 0 and 0.1. It is equal to $1 \cdot (0.1 - 0) = 0.1$, but when X is in that range Y must be in the range 0 to 0.01. Since Y is just a function of X, the event that X is between 0 and 0.1 and the event that Y is between 0 and 0.01 must have the same probability. Therefore Y is distributed according to some function that gives probability $\sqrt{b} - \sqrt{a}$ to an event that Y is between b and a. We can see that $\frac{1}{2\sqrt{y}}$ exhibits that property, since the integral of that function from a to b is just $\sqrt{b} - \sqrt{a}$.

Drawing numbers from a distribution

- How to draw a number from a discrete distribution P_X only having a generator of random real numbers in the range $[0,1]$ (uniform)?
- Theorem: let $U \sim \text{Unif}[0,1]$. Then a variable

$$X = F^{-1}(U)$$

- has a distribution for which the cumulative distribution function has the form $F_X = F$. Where F is a function of X with a value set $(0,1)$.

Expected value

- For a discrete random variable X:

$$EX = \sum_x xP(X = x)$$

- For a continuous random variable X:

$$EX = \int_{-\infty}^{\infty} xf(x)dx$$

- Remark: it suffices to integrate only over the set for which $f(x) > 0$.
- Properties of an expected value
  - If $a \leq X \leq b$ then $a \leq EX \leq b$
  - If $Y = g(X)$ then:

$$EY = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

  - The expected value is a linear operation:
  - $E(aX + B) = aEX + b$
  - $E(X_1 + ... + X_n) = EX_1 + ... + EX_n$

Variance

$$D^2(X) = E((X - EX)^2) = \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx$$

- It also holds that:

$$D^2(X) = E(X^2) - (EX)^2$$
$$D^2(aX + b) = a^2 D^2(X)$$

Normal distribution

- A random variable $X \in \mathbb{R}$ has normal (Gaussian) distribution, if:

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

- with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$.
- We denote: $X \sim N(\mu, \sigma^2)$.
- It is a very popular distribution in nature as is it models random phenomena that are result of averaging many independent factors.
- It is normalized since:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

- If $X \sim N(\mu, \sigma^2)$ then $EX = \mu$ and $D^2(X) = \sigma^2$
- If $X \sim N(\mu, \sigma^2)$ then $Y = aX + b$ has distribution $N(\mu a + b, \sigma^2 a^2)$
- Therefore a linear function of a variable with normal distribution has normal distribution.
- Let $X \sim N(0, 1)$. The function:

$$\phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

- is called the standard normal c.d.f.
- $\phi(x)$ does not have an analytical form (cannot be expressed with the usual functions and operations on numbers), so we can only compute it using numerical methods.
- $\phi(-x) = 1 - \phi(x)$
- If $X \sim N(\mu, \sigma^2)$ then $P(a \leq X \leq b) = \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$