Statistics for Data Science - 2

Week 4 Notes

Continuous Random Variables

1. **Cumulative distribution function:**
   A function $F : \mathbb{R} \to [0, 1]$ is said to be a Cumulative Distribution Function (CDF) if
   (i) F is a non-decreasing function taking values between 0 and 1.
   (ii) As $x \to -\infty$, $F \to 0$
   (iii) As $x \to \infty$, $F \to 1$
   (iv) Technical: F is continuous from the right.

2. **CDF of a random variable:**
   Cumulative distribution function of a random variable X is a function $F_X : R \to [0, 1]$ defined as
   $$F_X(x) = P(X \leq x)$$
   Properties of CDF
   - $F_X(b) - F_X(a) = P(a < X \leq b)$
   - F_X is a non-decreasing function of x.
   - F_X takes non-negative values.
   - As $x \to -\infty$, $F_X(x) \to 0$
   - As $x \to \infty$, $F_X(x) \to 1$

3. **Theorem: Random variable with CDF F(x)**
   Given a valid CDF $F(x)$, there exists a random variable X taking values in $\mathbb{R}$ such that
   $$P(X \leq x) = F(x)$$
   - If F is not continuous at x and $F(X)$ rises from F_1 to F_2 at x (jump at x), then
     $$P(X = x) = F_2 - F_1$$
   - If F is continuous at x, then
     $$P(X = x) = 0$$

4. **Continuous random variable:**
   A random variable X with CDF $F_X(x)$ is said to be a continuous random variable if $F_X(x)$ is continuous at every x.
   Properties of continuous random variables
   - CDF has no jumps or steps.
   - $P(X = x) = 0$ for all x.

- Probability of X falling in an interval will be nonzero

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

- Since $P(X = a) = 0$ and $P(X = b) = 0$, we have

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

5. **Probability density function (PDF):**
   A continuous random variable X with CDF $F_X(x)$ is said to have a PDF $f_X(x)$ if, for all x_0,
   $$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)dx$$
   - CDF is the integral of the PDF.
   - Derivative of the CDF (wherever it exists) is usually taken as the PDF.
   - Value of PDF around $f_X(x_0)$ is related to X taking a value around x_0.
   - Higher the PDF, higher the chance that X lies there.

6. For a random variable X with PDF f_X, an event A is a subset of the real line and its probability is computed as
   $$P(A) = \int_A f_X(x)dx$$
   - $P(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$

7. **Density function:**
   A function $f : \mathbb{R} \to \mathbb{R}$ is said to be a density function if
   (i) $f(x) \geq 0$
   (ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$
   (iii) $f(x)$ is piece-wise continuous

8. Given a density function f, there is a continuous random variable X with PDF as f.

9. **Support of random variable** X
   Support of the random variable X with PDF f_X is
   $$\text{supp}(X) = \{x : f_X(x) > 0\}$$
   - $\text{supp}(X)$ contains intervals in which X can fall with positive probability.

10. **Continuous Uniform distribution:**
    - $X \sim \text{Uniform}[a, b]$
    - PDF:
    $$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$
    - CDF:
    $$F_X(x) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$
11. **Exponential distribution:**
    - $X \sim \text{Exp}(\lambda)$
    - PDF:
    $$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$
    - CDF:
    $$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$
12. **Normal distribution:**
    - $X \sim \text{Normal}[\mu, \sigma^2]$
    - PDF:
    $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \qquad -\infty < x < \infty$$
    - CDF:
    $$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$
    - CDF has no closed form expression.
    - Standard normal: $Z = \text{Normal}(0,\, 1)$
      - PDF: $f_Z(z) = \dfrac{1}{\sqrt{2\pi}} \exp\left(\dfrac{-z^2}{2}\right) \qquad -\infty < z < \infty$
13. **Standardization:**
    If $X \sim \text{Normal}(\mu, \sigma^2)$, then
    $$\frac{X-\mu}{\sigma} = Z \sim \text{Normal}(0, 1)$$
14. To compute the probabilities of the normal distribution, convert probability computation to that of a standard normal.