Data Mining for Business Intelligence (IBM 312)

Sumit Kumar Yadav

Department of Management Studies

January 31, 2023

Uniform Distribution

Example: Uniform Distribution in $[0, 1]$.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}, \quad \rho(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x > 1 \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\,dx = \int_0^1 x\,dx = \frac{1}{2},$$

$$\mathrm{var}(X) = \int_0^1 x^2\,dx - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Uniform Distribution

Example: Uniform Distribution in $[0, 1]$.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}, \quad \rho(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{if } x > 1 \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\, dx = \int_0^1 x\, dx = \frac{1}{2},$$

$$\text{var}(X) = \int_0^1 x^2\, dx - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Practice Problem

$$P(Z \le \cancel{0.7}\, z \mid Y \in (0.3, 0.3 + dy)) = P(X + Y \le \cancel{0.7}\, z \mid Y \in \ldots)$$

$$= P(\cancel{0.4}\, X \le z - 0.3)$$

Random Variable X follows uniform distribution from [0,1], Random Variable Y follows same distribution and is independent of X. What is the distribution of X+Y?

$$(z - 0.3)$$

Case - 1: $z \le 1$

$$P(Z \le z) = \sum_{y \in} P(Z \le z \mid Y \in (y, y+dy)) \cdot P(Y \in_{y, \ldots})$$

$$\int_0^z (z - y)\, dy = (z^2/2)$$

Normal Distribution

Example: Normal (Gaussian) Distribution, Mean μ, Variance σ^2.

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

Why is this weird density called normal?

Normal Distribution

Example: Normal (Gaussian) Distribution, Mean μ, Variance σ^2.

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

Why is this weird density called normal?

Central Limit Theorem

Theorem

Let $\{X_k\}$ be a sequence of n mutually independent random variables having a common distribution, and mean (μ) and variance (σ^2) exists. Assuming, n to be large, the average of these random variables $\overline{X}$ follows approximately normal distribution with

1. *mean* = μ
2. *variance* = $\frac{\sigma^2}{n}$

What is meant by large n? Typically, $n \geq 30$

Central Limit Theorem - Special Case

Theorem

If the sample size is large, for WITH REPLACEMENT and independent sampling, the sample mean $\overline{X}$ is approximately normal with

1. *mean* = μ
2. *variance* = $\frac{\sigma^2}{n}$

What is meant by large n? Typically, $n \geq 30$

Simulation of Random numbers in Python

$f(x)$ — PDF

$F(x)$ — CDF : $P(X \leq x)$

$Z \sim unif(0,1)$

Inverse Transform Method??

$F^{-1}(x)$

$F^{-1}(Z)$

$\boxed{P(F^{-1}(Z) \leq z)} = P(Z \leq F(z)) = F(z)$