

Continuous Random Variables

General Information

- A function $f: \mathbb{R} \rightarrow \mathbb{R}$ is a *probability mass function* (pdf) of a continuous random variable X iff f is nonnegative and $\int_{-\infty}^{\infty} f(x) dx = 1$.
- For any probability mass function f , we have $P(a \leq X \leq b) = \int_a^b f(x) dx$. Whether the inequality is strict or nonstrict does not affect the above identity.
- A *mode* of X is any value m such that $f(m)$ is maximum.
- A *cumulative distribution function* (cdf) $F: \mathbb{R} \rightarrow [0, 1]$ of a random variable X is defined by

$$F(x) := P(X \leq x) = \int_{-\infty}^x f(x) dx.$$

- When writing out the cdf as a piecewise function, we explicitly write out the range of values for each case. We reserve the use of “otherwise” for pdf’s.
- Any cdf is continuous and nondecreasing.
- Let X be a continuous random variable with cdf F . To find the pdf g of any $y(X)$, we first find its cdf, then differentiate. We achieve this by reverse engineering $y(X) \leq y$ to find an inequality that relates X with y . E.g. $e^X \leq y$ iff $X \leq \ln(y)$.
- A *median* of X is any value m such that $P(X \leq m) = F(m) = 1/2$.
- Mean/Expectation:

$$\mu = E(X) := \int_{-\infty}^{\infty} x f(x) dx \quad \text{and} \quad E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

- Important property:

$$E(ag(X) \pm bh(x)) = a E(g(X)) \pm E(h(X)).$$

- Variance:

$$\text{Var}(X) := E(X^2) - [E(X)]^2.$$

- Important property:

$$\text{Var}(aX \pm b) = a^2 \text{Var}(X).$$

- A continuous random variable X has a *uniform distribution* over the interval $[a, b]$ iff its pdf f is such that

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Special Continuous Random Variables

Definition 2.1

A continuous random variable X has a *normal distribution* with mean μ and standard deviation σ , denoted by $X \sim N(\mu, \sigma^2)$, iff its pdf f is such that

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

General Information

1. A normal distribution is symmetrical about the line $x = \mu$. That is

$$P(X \leq \mu - \delta) = P(X \geq \mu + \delta)$$

for each $\delta > 0$. Note that the mean, median, and mode coincide with μ .

2. Properties of the normal distribution. Let X and Y be independent, such that $X \sim N(\mu, \sigma^2)$ and $Y \sim N(m, s^2)$. Then, for any $n \in \mathbb{N}$ and $x, y \in \mathbb{R}$,

(a) $nX \sim N(n\mu, n^2\sigma^2)$,

(b) $X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$,

(c) $aX \pm bY \sim N(a\mu \pm bm, a^2\sigma^2 + b^2s^2)$.

3. A variable $Z \sim N(0, 1)$ is said to follow the *standard* normal distribution.

Note: Z is reserved for this purpose.

4. Let $X \in N(\mu, \sigma^2)$. Then, $\frac{X-\mu}{\sigma}$ follows the standard normal distribution.

Correlation and Linear Regression

Note

A good scatter diagram should follow the guidelines below.

- The relative position of each point on the scatter diagram should be clearly shown.
- The range of values for the set of data should be clearly shown by marking out the extreme x and y values on the corresponding axis.
- The axes should be labeled clearly with the variables.

General Information