INTELLIGENT SYSTEMS LABORATORY

Lesson 4
Probability Distributions

Mathematics and Statistics for Data Science

Tapanan Yeophantong

Vincent Mary School of Science and Technology

Assumption University

Content

- Probability distributions & their applications
- Joint probability distributions

Random Variables

- A random variable assigns a numerical value to each outcome in a sample space.
- There are two types: *discrete* and *continuous*.
  - A discrete random variable is one whose possible values can be ordered, and there are gaps between adjacent values.
  - The possible values of a continuous random variable always contain an interval, that is, all the points between some two numbers.

Probability Distribution (Discrete)

- The list of possible values of a discrete random variable X, along with the probabilities for each, provides a complete description of the population from which X is drawn.
- This is known as probability distribution.
- The probability distribution of a discrete random variable X is the function $p(x) = P(X = x)$.
- A cumulative distribution function specifies the probability that X is *less than or equal to* a given value, i.e. $F(x) = P(X \leq x)$.

$$\mu_X = \sum_x x P(X = x)$$

Mean

(Discrete Random Variable)

$$\sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x)$$

Variance

(Discrete Random Variable)

$$\sigma_X = \sqrt{\sigma_X^2}$$

Standard Deviation

(Discrete Random Variable)

Example 1

- A certain industrial process is brought down for recalibration whenever the quality of the items produced falls below specifications. Let X represent the number of times the process is recalibrated during a week, and assume that X has the following probability distribution.

x	0	1	2	3	4
$p(x)$	0.35	0.25	0.20	0.15	0.05

- Find mean, variance, and standard deviation of X.

Example 1 - Solution

- Find mean, variance, and standard deviation of X.

x	0	1	2	3	4
$p(x)$	0.35	0.25	0.20	0.15	0.05

$$\mu_X = 0(0.35) + 1(0.25) + 2(0.20) + 3(0.15) + 4(0.05) = 1.30$$

$$\begin{aligned}\sigma_X^2 &= (0 - 1.30)^2 P(X = 0) + (1 - 1.30)^2 P(X = 1) + (2 - 1.30)^2 P(X = 2) \\ &\quad + (3 - 1.30)^2 P(X = 3) + (4 - 1.30)^2 P(X = 4) \\ &= (1.69)(0.35) + (0.09)(0.25) + (0.49)(0.20) + (2.89)(0.15) + (7.29)(0.05) \\ &= 1.51\end{aligned}$$

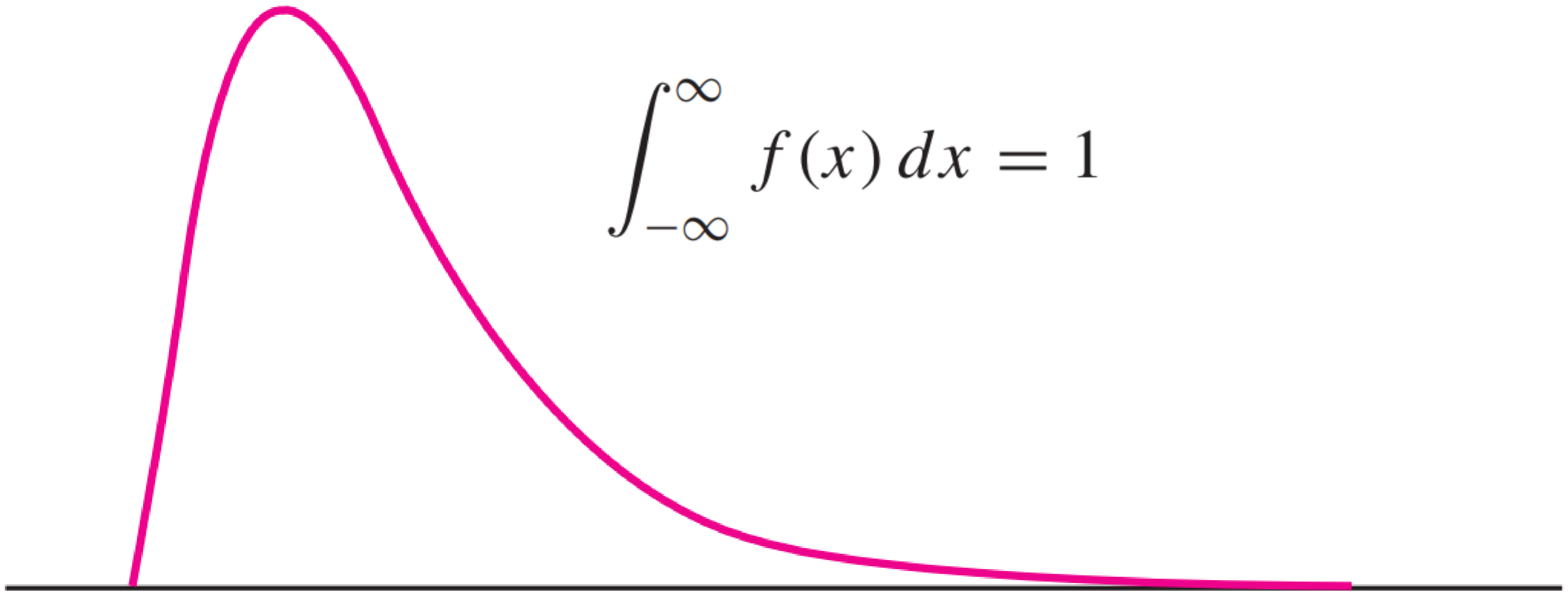
$$\sigma_X = \sqrt{1.51} = 1.23$$

Probability Distribution (Continuous)

- A random variable is continuous if its probabilities are given by areas under a curve.
- The curve is called a probability density function (or probability distribution) for the random variable.
- Let X be a continuous random variable with probability density function *f(x)*. Then:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Continuous Distribution - Example



$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Cumulative Distribution Function

- Let X be a continuous random variable with probability density function f(x), the cumulative distribution function of X is:

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

$$\mu_X = \int_{-\infty}^{\infty} x f(x)\, dx$$

Mean

(Continuous Random Variable)

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\, dx$$

Variance

(Continuous Random Variable)

$$\sigma_X = \sqrt{\sigma_X^2}$$

Standard Deviation

(Continuous Random Variable)

Joint Probabilities – An Example

x	y	$P(X = x \text{ and } Y = y)$
129	15	0.12
129	16	0.08
130	15	0.42
130	16	0.28
131	15	0.06
131	16	0.04

Jointly Discrete

- If X and Y are jointly discrete random variables, the joint probability distribution of X and Y is:

$$p(x, y) = P(X = x \text{ and } Y = y)$$

- Marginal probability distribution of X and of Y:

$$p_X(x) = P(X = x) = \sum_y p(x, y)$$

$$p_Y(y) = P(Y = y) = \sum_x p(x, y)$$

Jointly Continuous

- If X and Y are jointly continuous random variables, the joint probability distribution of X and Y is:

$$P(a \leq X \leq b \text{ and } c \leq Y \leq d) = \int_a^b \int_c^d f(x,y)\, dy\, dx$$

- Marginal probability distribution of X and of Y:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

Example 2

- Given the discrete joint probability distribution below:

x	y	$P(X = x \text{ and } Y = y)$
129	15	0.12
129	16	0.08
130	15	0.42
130	16	0.28
131	15	0.06
131	16	0.04

  - Find the probability that X = 129.
  - Find the probability that Y = 16.

Example 2 - Solution

- Given the discrete joint probability distribution below:

x	y	$P(X = x \text{ and } Y = y)$
129	15	0.12
129	16	0.08
130	15	0.42
130	16	0.28
131	15	0.06
131	16	0.04

- $P(X = 129) = 0.20$
- $P(Y = 16) = 0.40$