Chapter 3: Random Variables and Probability Distributions

Juhyung Lee

Department of Statistics
University of Florida

Random Variable

Definition (Random Variable)

A **random variable** is a function from a sample space S into the real numbers.

Example (Example 3.1 of WMMY)

Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. Define the random variable Y to be the number of red balls. State the sample space and the possible values of Y.

Example (Example 3.7 of WMMY)

Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. The random variable X takes on all values x for which $x \geq 0$.

Random Variable

Definition (Discrete and Continuous Random Variables)

A **discrete random variable** takes a set of separate values, such as $0, 1, 2, \ldots$ (e.g., Example 3.1 of WMMY).
A **continuous random variable** has possible values that form an interval (e.g., Example 3.7 of WMMY).

Discrete Probability Distributions

Definition (Probability Mass Function)

The function $f(x)$ is a **probability mass function (pmf)** of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum_x f(x) = 1$,
3. $P(X = x) = f(x)$.

Definition (Cumulative Distribution Function: Discrete Case)

The **cumulative distribution function (cdf)** $F(x)$ of a discrete random variable X with pmf $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \text{ for } -\infty < x < \infty.$$

Discrete Probability Distributions

Example (Examples 3.9 and 3.10 of WMMY)

Suppose a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags. Let X be the number of cars with side airbags among the next 4 cars sold by the agency. Find the pmf and cdf of X.

See also Figures 3.1 and 3.3 of WMMY.

Continuous Probability Distributions

Definition (Probability Density Function)

The function $f(x)$ is a **probability density function (pdf)** of the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in \mathbb{R}$,
2. $\int_{-\infty}^{\infty} f(x)dx = 1$,
3. $P(a < X < b) = \int_a^b f(x)dx$.

Definition (Cumulative Distribution Function: Continuous Case)

The **cumulative distribution function (cdf)** $F(x)$ of a continuous random variable X with pdf $f(x)$ is

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

Continuous Probability Distributions

Example (Examples 3.11 and 3.12 of WMMY)

Suppose that the error in the reaction temperature, in $^\circ\mathrm{C}$, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \dfrac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

a. Verify that $f(x)$ is a density function.
b. Find $F(x)$.
c. Find $P(0 < X \leq 1)$.

See also Figure 3.6 of WMMY.

Joint Probability Distributions

Definition (Joint PMF)

The function $f(x, y)$ is a **joint pmf** of the discrete random variables X and Y if

1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

For any region A in the xy plane, $P[(X, Y) \in A] = \sum \sum_A f(x, y)$.

Definition (Joint PDF)

The function $f(x, y)$ is a **joint pdf** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$,
3. $P[(X, Y) \in A] = \int \int_A f(x, y) dxdy$, for any region A in the xy plane.

Joint Probability Distributions

Definition (Marginal PMF and PDF)

For the discrete random variables X and Y, the **marginal pmfs** of X alone and of Y alone are

$$g(x) = \sum_y f(x,y) \text{ and } h(y) = \sum_x f(x,y).$$

For the continuous random variables X and Y, the **marginal pdfs** of X alone and of Y alone are

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy \text{ and } h(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

Joint Probability Distributions

Definition (Conditional PMF and PDF)

Let X and Y be two random variables, discrete or continuous. The **conditional pmf** or **pdf** of Y given that $X = x$ is

$$f(y|x) = \frac{f(x, y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the **conditional pmf** or **pdf** of X given that $Y = y$ is

$$f(x|y) = \frac{f(x, y)}{h(y)}, \text{ provided } h(y) > 0.$$

Joint Probability Distributions

Definition (Independence of Two Random Variables)

Let X and Y be two random variables, discrete or continuous, with joint pmf or pdf $f(x, y)$ and marginal pmfs or pdfs $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

Joint Probability Distributions

Example (Example 3.20 of WMMY)

Given the joint pdf

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere}, \end{cases}$$

find $g(x)$, $h(y)$, $f(x|y)$, and evaluate

$$P\left(\frac{1}{4} < X < \frac{1}{2} \middle| Y = \frac{1}{3}\right).$$

Are X and Y independent?

See also Examples 3.14, 3.16, 3.18, and 3.21 of WMMY for the discrete case.