# CHEM. ENG. II : PROBABILITY AND STATISTICS

## Chapter 2. Random Variables and Probability Distributions

### 2.1 Motivation: Generalization of Notation

The probability definitions, rules, and theorems given previously are all framed in terms of events in a sample space. For example, for an experiment with possible sample outcomes denoted by the *sample space* $S$, an *event* $E$ was defined as any collection of sample outcomes, that is, any subset of the set $S$.

$$\begin{array}{ccccccc} \text{EXPERIMENT} & \longrightarrow & \text{SAMPLE OUTCOMES} & \longrightarrow & \text{EVENTS} & \longrightarrow & \text{PROBABILITY FUNCTION} \\ & & \{s_1, s_2, ...\} = S & \longrightarrow & E \subseteq S & \longrightarrow & \mathrm{P}(E). \end{array}$$

In this framework, it is necessary to consider each experiment with its associated sample space separately - the nature of sample space $S$ is typically different for different experiments.

**EXAMPLE 1**: Count the number of days in February which have zero precipitation.

**SAMPLE SPACE**: $S = \{0, 1, 2, ..., 28\}$.

Let $E_i$ = "$i$ days have zero precipitation". Then

$$E_i \cap E_j = \emptyset \quad \text{and} \quad \bigcup_{i=0}^{28} E_i = S$$

so $E_0, ..., E_{28}$ form a partition of $\Omega$. We must assign $\mathrm{P}(E_i) = p_i$ for $i = 0, 1, ..., 28$ such that, by the axioms,

$$0 \leq p_i \leq 1 \quad \text{and} \quad \sum_{i=0}^{28} p_i = 1$$

**EXAMPLE 2**: Count the number of goals in a football match.

**SAMPLE SPACE**: $S = \{0, 1, 2, 3, ...\}$.

Let $E_i$ = "$i$ goals in the match". Then, as above

$$E_i \cap E_j = \emptyset \quad \text{and} \quad \bigcup_{i=0}^{\infty} E_i = S$$

so $E_0, E_1, E_2, ...$ form a partition of $\Omega$. We assign $\mathrm{P}(E_i) = p_i$ for $i = 0, 1, 2, 3, ...$ such that, by the axioms,

$$0 \leq p_i \leq 1 \quad \text{and} \quad \sum_{i=0}^{\infty} p_i = 1$$

In both of these examples, we need a formula to specify each $p_i$.

**EXAMPLE 3**: Measure the operating temperature of an experimental process.

**SAMPLE SPACE**: $S = \{ x \ : \ x > T_{min} \}$.

Here it is difficult to express

$$\mathrm{P}[\text{ “Measurement is } x \text{ ” }]$$

but possible to think about

$$\mathrm{P}[\text{ “Measurement is } \leq x \text{ ” }] = F(x), \text{ say,}$$

and again we seek a formula for $F(x)$.

A general notation useful for all such examples can be obtained by considering a sample space that is **equivalent** to $S$ for a general experiment, but whose form is more familiar. For example, for a general sample space $S$, if it were possible to associate a subset of the **integer** or **real** number systems, $\mathbb{X}$ say, with $S$, then attention could be restricted to considering events in $\mathbb{X}$, whose structure is more convenient, as then

*events in $S$ are collections of sample outcomes of the experiment*

*events in $\mathbb{X}$ are intervals of the real numbers*

**EXAMPLE** : Consider an experiment involving counting the number of breakdowns of a production line in a given month. The experimental sample space $S$ is therefore the collection of sample outcomes $s_0, s_1, s_2, ...$ where $s_i$ is the outcome "there were $i$ breakdowns"; events in $S$ are collections of the $s_i$s. Then a useful equivalent sample space is the set $\mathbb{X} = \{0, 1, 2, ...\}$, and events in $\mathbb{X}$ are collections of non-negative integers.

Formally, therefore, we seek a function or **map** from $S$ to $\mathbb{X}$. This map is known as a **random variable**.

## 2.2 Random Variables

**DEFINITION**

A **random variable** $X$ is a function from experimental sample space $S$ to some set of real numbers $\mathbb{X}$ that maps $s \in S$ to a unique $x \in \mathbb{X}$

$$X : \quad \begin{aligned} S &\longrightarrow \mathbb{X} \subseteq \mathbb{R} \\ s &\longmapsto x \end{aligned}$$

**Interpretation** A random variable is a shorthand way of describing the outcome of an experiment in terms of real numbers.

**EXAMPLE 1** $X =$"the number of days in Feb. with zero precipitation"

**EXAMPLE 2** $X =$"the number of goals in a football match"

**EXAMPLE 3** $X =$"the measured operating temperature"

Our objective is to find (or assume) a formula for

**EXAMPLE 1** P[ $X = x$ ] for $x = 0, 1, 2, ..., 28$

**EXAMPLE 2** P[ $X = x$ ] for $x = 0, 1, 2, 3, ...$

**EXAMPLE 3** P[ $X \leq x$ ] for $x > T_{min}$.

Therefore $X$ is merely the count/number/measured value corresponding to the outcome of the experiment.

Depending on the type of experiment being carried out, there are two possible forms for the set of values that $X$ can take:

A random variable is **DISCRETE** if the set $\mathbb{X}$ is of the form

$$\mathbb{X} = \{x_1, x_2, ..., x_n\} \text{ or } \mathbb{X} = \{x_1, x_2, ...\},$$

that is, a finite or infinite set of **distinct** values $x_1, x_2, ..., x_n, ....$ Discrete random variables are used to describe the outcomes of experiments that involve **counting** or **classification**.

A random variable is **CONTINUOUS** if the set $\mathbb{X}$ is of the form

$$\mathbb{X} = \bigcup_i \{x : a_i \leq x \leq b_i\}$$

for real numbers $a_i, b_i$, that is, the union of **intervals** in $\mathbb{R}$. Continuous random variables are used to describe the outcomes of experiments that involve **measurement**.

## 2.3 Probability distributions

A **probability distribution** is a function that assigns probabilities to the possible values of a random variable. When specifying a probability distribution for a random variable, two aspects need to be considered. First, the range of the random variable (that is, the values of the random variable which have positive probability) must be specified. Secondly, the method via which the probabilities are assigned to different values in the range must be specified; typically this is achieved by means of a function or formula.

In summary, we need to find a function or formula via which

$$\mathrm{P}[\ X = x\ ] \qquad \text{or} \qquad \mathrm{P}[\ X \leq x\ ]$$

can be calculated for each $x$ in a suitable range $\mathbb{X}$.

## 2.4 Discrete probability distributions

For discrete random variables there are two routes via which the probability distribution can be specified.

### 2.4.1. The probability mass function

The probability distribution of a *discrete* random variable $X$ is described by the **probability mass function** $f_X$, specified by

$$f_X(x) = \mathrm{P}[X = x] \quad \text{for } x \in \mathbb{X} = \{x_1, x_2, ..., x_n, ...\}$$

### 2.4.2. Properties of the mass function

The mass function $f_X$ must exhibit the following properties:

(i) $f_X(x_i) \geq 0$ for all $i$ (ii) $\sum_i f_X(x_i) = 1.$

### 2.4.3. The cumulative distribution function

The **cumulative distribution function** or **cdf**, $F_X$, is defined by

$$F_X(x) = \mathrm{P}[X \leq x] \quad \text{for } x \in \mathbb{R}$$

### 2.4.4. Properties of the (discrete) distribution function

The cdf $F_X$ must exhibit the following properties:

(i) $\lim_{x \longrightarrow -\infty} F_X(x) = 0$

(ii) $\lim_{x \longrightarrow \infty} F_X(x) = 1$

(iii) $\lim_{h \to 0^+} F_X(x+h) = F_X(x)$ [i.e. $F_X$ is continuous from the right]

(iv) $a < b \Rightarrow F_X(a) \leq F_X(b)$ [i.e. $F_X$ is non-decreasing]

(v) $\mathrm{P}[a < X \leq b] = F_X(b) - F_X(a)$

The cumulative distribution function defined in this way is a "**step function**".

The functions $f_X$ and/or $F_X$ can be used to describe the **probability distribution** of random variable $X$.

**EXAMPLE** An electrical circuit comprises six fuses.

let $X =$"number of fuses that fail within one month". Then

$$\mathbb{X} = \{0, 1, 2, 3, 4, 5, 6\}$$

To specify the probability distribution of $X$, can use the mass function $f_X$ or the cdf $F_X$. For example,

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $f_X(x)$ | $\frac{1}{16}$ | $\frac{2}{16}$ | $\frac{4}{16}$ | $\frac{4}{16}$ | $\frac{2}{16}$ | $\frac{2}{16}$ | $\frac{1}{16}$ |
| $F_X(x)$ | $\frac{1}{16}$ | $\frac{3}{16}$ | $\frac{7}{16}$ | $\frac{11}{16}$ | $\frac{13}{16}$ | $\frac{15}{16}$ | $\frac{16}{16}$ |

as $F_X(0) = \mathrm{P}[\ X \le 0\ ] = \mathrm{P}[\ X = 0\ ] = f_X(0)$, $F_X(1) = \mathrm{P}[\ X \le 1\ ] = \mathrm{P}[\ X = 0\ ] + \mathrm{P}[\ X = 1\ ] = f_X(0) + f_X(1)$, and so on. Note also that, for example,

$$\mathrm{P}[\ X \le 2.5\ ] \equiv \mathrm{P}[\ X \le 2\ ]$$

as the random variable $X$ only takes values $0, 1, 2, 3, 4, 5, 6$.

**EXAMPLE** A computer is prone to crashes.

Suppose that P[ "Computer crashes on any given day" ] $= \theta$, for some $0 \le \theta \le 1$, independently of crashes on any other day.

Let $X =$"number of days until the first crash". Then

$$\mathbb{X} = \{1, 2, 3, ...\}$$

To specify the probability distribution of $X$, can use the mass function $f_X$ or the cdf $F_X$. Now,

$$f_X(x) = \mathrm{P}[\ X = x\ ] = (1-\theta)^{x-1}\theta$$

for $x = 1, 2, 3, ...$ (if the first crash occurs on day $x$, then we must have a sequence of $x - 1$ crash-free days, followed by a crash on day $x$). Also

$$F_X(x) = \mathrm{P}[\ X \le x\ ] = \mathrm{P}[\ X = 1\ ] + \mathrm{P}[\ X = 2\ ] + ... + \mathrm{P}[\ X = x\ ] = 1 - (1-\theta)^x$$

as the terms in the summation are merely a geometric progression with first term $\theta$ and common term $1 - \theta$.

### 2.4.5 Fundamental relationship between $f_X$ and $F_X$

The fundamental relationship between $f_X$ and $F_X$ is obtained by noting that if $x_1 \le x_2 \le ... \le x_n \le ...$, then

$$\mathrm{P}[X \le x_i] = \mathrm{P}[X = x_1] + ... + \mathrm{P}[X = x_i],$$

so that

$$F_X(x) = \sum_{x_i \le x} f_X(x_i)\ ,$$

and

$$f_X(x_1) = F_X(x_1)$$

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \qquad \text{for } i \ge 2$$

so $\mathrm{P}[c_1 < X \le c_2] = F_X(c_2) - F_X(c_1)$ for any real numbers $c_1 < c_2$.

Hence, in the discrete case, we can calculate $F_X$ from $f_X$ by **summation**, and calculate $f_X$ from $F_X$ by **differencing**.

## 2.5 Continuous probability distributions

For discrete random variables there are two routes via which the probability distribution can be specified.

### 2.5.1. The cumulative distribution function

The probability distribution of a *continuous* random variable $X$ is defined by the continuous **cumulative distribution function** or **c.d.f.**, $F_X$, specified by

$$F_X(x) = \mathrm{P}[X \leq x] \quad \text{for } x \in \mathbb{X}$$

that is, an identical definition to the discrete case.

### 2.5.2. Properties of the (continuous) distribution function

The continuous cdf $F_X$ must exhibit the same properties: as for the discrete cdf, except

(iii) $\lim_{h \to 0} F_X(x+h) = F_X(x)$ [i.e. $F_X$ is continuous]

### 2.5.3. The probability density function

The **probability density function**, or **pdf**, $f_X$, is defined by

$$f_X(x) = \frac{d}{dx}\{F_X(x)\}$$

so that, by a fundamental calculus result,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

### 2.5.4. Properties of the density function

The pdf $f_X$ must exhibit the following properties:

(i) $f_X(x) \geq 0$ for $x \in \mathbb{X}$, (ii) $\int_{\mathbb{X}} f_X(x)\, dx = 1$.

In the continuous case, we calculate $F_X$ from $f_X$ by **integration**, and $f_X$ from $F_X$ by **differentiation**.

**NOTES**

(i) In both discrete and continuous cases, $F_X(x)$ id defined for all $x \in \mathbb{R}$, and $f_X(x)$ also defined for all $x$ but may be zero for some values of $x$.

(ii) If $X$ is continuous, we have

$$\mathrm{P}[\, a \leq X \leq b \,] = F_X(b) - F_X(a) \longrightarrow 0$$

as $b \longrightarrow a$. Hence, for each $x$, we must have

$$\mathrm{P}[\, X = x \,] = 0$$

if $X$ is continuous. Therefore must use $F_X$ to specify the probability distribution initially, although it is often easier to think of the "shape" of the distribution via the pdf $f_X$.

Any function that satisfies the properties for a pdf can be used to construct a probability distribution. Note that, for a continuous random variable

$$f_X(x) \neq \mathrm{P}[X = x].$$

**EXAMPLE** Failure times.

A component is installed at time $x = 0$ and continues to operate until failure. Let $X =$"failure time of the component. then $\mathbb{X} = \{ x \ : \ x > 0\} = \mathbb{R}^+$. Suppose that

$$\mathrm{P}[\ X > x\ ] = \frac{1}{(1+x)^2} \qquad x > 0$$

Then

$$F_X(x) = \mathrm{P}[\ X \leq x\ ] = 1 - \mathrm{P}[\ X > x\ ] = 1 - \frac{1}{(1+x)^2}$$

for $x > 0$. By differentiation, we have the pdf

$$f_X(x) = \frac{d}{dx}\{F_X(x)\} = \frac{2}{(1+x)^3}$$

**EXERCISE** Sketch the two functions $f_X$ and $F_X$, and verify that the required properties detailed in sections 2.4.2., 2.4.4., 2.5.2 and 2.5.4..

## 2.6. Joint Probability Distributions

Consider a vector of $k$ random variables, $\mathbf{X} = (X_1, ..., X_k)$, (representing the outcomes of $k$ different experiments carried out once each, or of one experiment carried out $k$ times). The probability distribution of $\mathbf{X}$ is described by a **joint** probability mass or density function.

e.g. Consider the particular case $k = 2$, $\mathbf{X} = (X_1, X_2)$. Then the following functions are used to specify the probability distribution of $\mathbf{X}$;

### 2.6.1. Joint probability mass/density function

The joint mass/density function is denoted $f_{X_1,X_2}(x_1, x_2)$

- assigns probability to the joint space of outcomes

- in the discrete case, $f_{X_1,X_2}(x_1, x_2) = \mathrm{P}[(X_1 = x_1) \cap (X_2 \cap x_2)]$

- need

(i) $f_{X_1,X_2}(x_1, x_2) \geq 0$ for all possible outcomes $x_1, x_2$.

(ii) $\sum\sum f_{X_1,X_2}(x_1, x_2) = 1$ or $\int\int f_{X_1,X_2}(x_1, x_2)\ dx_1 dx_2 = 1$

where the double summation/integration is over all possible values of $(x_1, x_2)$.

Typically, such a specification is represented by a probability table; for example for discrete random variables $X_1$ and $X_2$, we could have

| | | $X_1$ | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| $X_2$ | 1 | 0.100 | 0.200 | 0.000 | 0.000 |
| | 2 | 0.200 | 0.250 | 0.050 | 0.000 |
| | 3 | 0.000 | 0.050 | 0.050 | 0.025 |
| | 4 | 0.000 | 0.000 | 0.025 | 0.050 |

where the entry in column $i$, row $j$ is $f_{X_1,X_2}(i, j) = \mathrm{P}[\ (X_1 = i) \cap (X_2 = j)\ ]$, which we may write $\mathrm{P}[\ X_1 = i, X_2 = j\ ]$.

### 2.6.2. Marginal probability mass/density functions

The joint mass function automatically defines the probability distribution of the individual random variables. For example, if $k = 2$, then we have the two marginal mass/density functions are $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$

- in the discrete case,

$$f_{X_1}(x_1) = \sum_{x_2} f_{X_1,X_2}(x_1, x_2)$$

and

$$f_{X_2}(x_2) = \sum_{x_1} f_{X_1,X_2}(x_1, x_2)$$

- in the continuous case,

$$f_{X_1}(x_1) = \int f_{X_1,X_2}(x_1, x_2)\ dx_2$$

$$f_{X_2}(x_2) = \int f_{X_1,X_2}(x_1, x_2)\ dx_1.$$

so the marginal mass/density function for random variable $X_1$ is obtained by summing/integrating out the joint mass/density function for $X_1$ and $X_2$ over all possible values of random variable $X_2$. In the discrete case

$$\text{P}[X_1 = x_1] = \sum_{x_2} \text{P}[(X_1 = x_1) \cap (X_2 = x_2)]$$

which is a result that is justified by the Theorem of Total Probability.

### 2.6.3. Conditional mass/density functions

In the discrete two variable case, consider the probability

$$\text{P}[\ X_1 = x_1 \mid X_2 = x_2\ ]$$

that is, the conditional probability distribution of $X_1$, given that $X_2 = x_2$. This conditional distribution is easily computed from the conditional probability definition, that is

$$\text{P}[\ X_1 = x_1 \mid X_2 = x_2\ ] = \frac{\text{P}[\ X_1 = x_1, X_2 = x_2\ ]}{\text{P}[\ X_2 = x_2\ ]} = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

that is, proportional to the $x_2$ row of the table.

By extending these concepts, we may define the conditional probability distributions for both variables in the discrete and continuous cases; The two conditional mass/density functions are $f_{X_1|X_2}(x_1|x_2)$ and $f_{X_2|X_1}(x_2|x_1)$

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}\ , \quad \text{if } f_{X_2}(x_2) > 0.$$

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)}\ , \quad \text{if } f_{X_1}(x_1) > 0.$$

In the discrete case, this result becomes

$$f_{X_1|X_2}(x_1|x_2) = \text{P}[X_1 = x_1|X_2 = x_2] = \frac{\text{P}[(X_1 = x_1) \cap (X_2 = x_2)]}{\text{P}[X_2 = x_2]}$$

if $\text{P}[X_2 = x_2] > 0$, which is justified by the definition of conditional probability.

**EXAMPLE** : Suppose that $X_1$ and $X_2$ are discrete random variables that take values $\{1, 2, ..., n\}$ and $\{1, 2, ..., m\}$ respectively. Then the joint mass function can be displayed as a table with $n$ rows and $m$ columns, where

- the $(i, j)$th cell contains $\mathrm{P}[(X_1 = i) \cap (X_2 = j)]$
- the marginal mass function for $X_1$ is given by the **row** totals
- the marginal mass function for $X_2$ is given by the **column** totals
- the conditional mass function for $X_1$ given $X_2 = j$ is given by the $j$th column divided by the sum of the $j$th column
- the conditional mass function for $X_2$ given $X_1 = i$ is given by the $i$th row divided by the sum of the $i$th row

for $i = 1, ..., n$ and $j = 1, ..., m$.

**EXAMPLE** Suppose that the joint density of continuous variables $X_1$ and $X_2$ is given by

$$f_{X_1,X_2}(x_1, x_2) = x_2^2 e^{-x_2(1+x_1)}$$

for $x_1, x_2 \geq 0$ and zero otherwise. It can be shown that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\ dx_1 dx_2 = 1$$

and that the marginal pdf for $X_1$ is given by

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} F_{X_1,X_2}(x_1, x_2)\ dx_2 = \int_0^{\infty} x_2^2 e^{-x_2(1+x_1)}\ dx_2 = \frac{2}{(1 + x_1)^3}$$

for $x_1 \geq 0$, and zero otherwise.

**EXERCISE** Check these calculations, and compute the marginal pdf for $X_2$.

**DEFINITION**

Random variables $X_1$ and $X_2$ are **independent** if

(i) the joint mass/density function of $X_1$ and $X_2$ factorizes into the product of the two marginal pdfs, that is,

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

(ii) the range of $X_1$ does not conflict/influence/depend on the range of $X_2$ (and *vice verse*).

The concept of independence for random variables is closely related to the concept of independence for events.

**EXERCISE** A point is to be selected from the interior of the unit circle. Let $X$ and $Y$ be the continuous random variables corresponding to the $x$- and $y$-coordinates respectively. Suppose that all points within the circle are equally likely to be selected.

(i) write down the joint pdf of $X$ and $Y$.

(ii) find the marginal pdfs of $X$ and $Y$.

(iii) state whether $X$ and $Y$ are independent.