A Mathematical Theory for Mixing of Particulate Materials

by

Ali Ghaderi

Chemical and Process Engineering

School of Engineering

University of Surrey

UK

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

2006

To Loving memory of Khairunnisa

Abstract

A theoretical framework for sampling theory is developed. In this relation, concepts like mixture heterogeneity and representative samples are mathematically defined. Further, the relation between Gy's concepts of accuracy and reproducibility with mixture quality and the entropy of the sample distribution is established. Moreover, it is shown that within the developed framework, Lacey's conjecture is mathematically consistent. It is also shown that a consequence of the theory is the prediction of the number of key components of given size in random binary closed batch systems. It is also shown that this estimate is a function of microstructural properties of the mixture under study. Furthermore, this theory is used to develop a unifying approach to description of RTD of continuous systems. These results are further used to develop a model for RTD of a commercial twin screw extruder.

A new theoretical approach to the dynamics of the mixing processes is developed. In this context, the concept of heterogeneity landscape is introduced. It is argued that the valleys in the heterogeneity landscape correspond to different equilibrium states of the mixture. Further, it is shown that the valleys in the heterogeneity landscape can mathematically be described by heterogeneity equation and this would allow for classification of all the valleys. The characteristic function of the general solution to the heterogeneity equation is also determined. Moreover, it is shown that based on the mathematical model for the valleys, one can deduce that in the case of insufficient information about the mixture structure, the normal distribution, up to the second order; is the best distribution in describing the mixture structure.

Contents

List of Figures

List of Tables

Chapter 1

Foreword

> "Who could ever calculate the path of a molecule? How do we know that the creations of worlds are not determined by falling grains of sand?"
>
> ---
>
> Victor Hugo, Les misérables (1862)

The worldwide annual production of grains and aggregates of various kinds is gigantic, reaching almost ten billion tons (Duran 2000, p.3). It is estimated that the processing of particulate materials consumes approximately 10% of the total energy consumptions in the world. According to US Department of Energy (EIA 2005), the world's total energy consumption in the year 2001 was 119,000 TWh. Hence, roughly speaking, we spent about 12,000 TWh just for processing of particulate materials in year 2001. This was more than the amount of electrical energy produced by all the conventional hydroelectric, Geothermal, Solar and Wind power plants in the world during the same year! This also exceeded the combined energy consumption of Europe's three biggest economies; Germany, France and United Kingdom in the same year. Obviously, this huge figure also contributes significantly to the emission of greenhouse gases, which have great impact on life on the planet. Therefore, advances on understanding of the particulate systems bound to have major global economic and environmental impact.

From scientific point of view, particulate systems offer numerous challenges with far reaching implications for many seemingly different research fields. The reason perhaps lies in some apparently intractable complexities. The particulate materials exhibit some sort of "dual" nature, wherein they simultaneously show both macroscopic discrete organization and microscopic continuous behaviour, in which both chaos and new order may emerge. This has been a source of great consternation for scientists. It is becoming more and more apparent that a satisfactory understanding of particulate systems ought to lead to a paradigm shift in how we conceive the world around us.

An immediate consequence of the complexities involved in particulate systems is that it is rarely possible to have a complete overview of all parameters affecting the system under study. This would usually result in a situation where many systems operate in far from optimal state and consequently may result in severe economic loses. Furthermore, even in the rare cases in which it is possible to identify the parameters affecting the process; it is not always possible to find economically viable measurement methods to directly determine the process parameters of interest. For more than a century, many engineers and scientists have been struggling to tackle such problems with limited success. However, it is still believed by many that the science of particulate materials is in its infancy. Nevertheless, this situation is not unique for this subject. In fact many phenomenon as diverse as biological processes and traffic jams in big cities to earthquake and tsunami forecast share similar resilience in revealing their secrets. It seems that the common denominator in all of these subjects is their complex nature which demands an approach with interdisciplinary flavour.

1.1 The Objective of the Thesis

A bulk of particulate materials usually consists of particles with different sizes and shapes. That is, one should always consider a bulk of solids consisting of mixture of particles with different sizes, shape and so on. Nonetheless, mixing of different powders is also an important operation in a wide range of processes. For these reasons, it does not come as surprise that the characterization of mixing and mixtures of particulate materials is one of the most important issues in bulk solid handling processes.

It has been over sixty years ago since Lacey (1943) gave a probabilistic characterization of mixture of particulate materials. Since than several hundreds of research articles and books, covering various aspects of the subject have been published. But despite this huge wealth of information, we believe that the subject is still lacking a reliable scientific framework. We believe that the function of science, as Braithwaite (1953, p.1) put it, is to establish general laws covering the behaviour of the empirical events or objects with which the science in question is concerned, and thereby to enable us to connect together our knowledge of separately known events, and make reliable predictions of events as yet unknown. However, it is our opinion that most of the work in this subject do not serve this function. Therefore, the main objective of this work is to establish a framework within which a unifying scientific approach to the subject is possible.

The most important step in this direction is to establish a method by which one can unambiguously describe existing information. The common and most successful method of describing our state of knowledge about a phenomenon is by a mathematical model. With a mathematical model one achieves several things. As Zellner (1984, p.14) put it, "Mathematically formulated models do have the advantage that in general their logical consistency can be checked using the available operations of mathematics. While, this task is not always easy, it does appear easier than that associated with checking the logical consistency of many non-mathematical models". Thus, the first and perhaps the most important advantage is that it is an effective way of communicating our state of the knowledge to others. Secondly, our model allows us to take advantage of huge number of methods developed by others outside our field. And the last but not the least, a mathematical model can aid us to get a clearer picture of the consequences of our assumptions and discoveries.

However, despite these advantages, it is difficult to find a single mathematical model capturing every circumstance that is relevant in the real world. This has led us to believe that any mathematical model would be at best a *subjective* description of our state of knowledge of the real world. This is perhaps the main reason that there are so many different models for characterizing mixtures of particulate materials. Indeed, the subjectivity means that each model corresponds to a different state of knowledge about the process under study. Nevertheless, it is reasonable to demand that any one who is given same information about a phenomenon should arrive at the same conclusions. Accordingly, we believe that any unifying approach to the subject of mixing has to be founded on a mathematical framework. However, in order to achieve this, one needs to carefully identify and define the most fundamental concepts of the subject. Indeed, as Jaynes (1967) has pointed out, "you cannot base a general mathematical theory on imprecisely defined concepts. You can make some progress that way; but sooner or later the theory bound to dissolve in ambiguities which prevent you from extending it further ... Unless the conceptual problems of a field have been clearly resolved, you cannot say which mathematical problems are the relevant ones worth working on; and your efforts are more likely to be wasted". Most of the present work is devoted to identification and definition of the fundamental concepts of the mixing of particulate materials. However, some efforts is also made to demonstrate some of the direct implications of these concepts.

The task of formulating a scientific theory based on an acceptable definition of scientific theory is not without its difficulties. The most accepted definition is due to Popper (2002, §22), which requires that all scientific theories should fulfil the *falsifiability* criteria. This basically means that all the "meaningful" statements of a scientific theory should be in such logical form that they can be refuted by experience. Although, it is tempting to adopt this definition, one should bear in mind that Popper dismisses the idea of *induction* (see for example Popper 2002, p.18). That is, he claims that there is no rational justification for reasoning from repeated instances of which one has experience to instances of which one has no experience. However, in the case of complex systems in which one, in general, does not have a complete overview of all the parameters, induction is a valuable tool. Indeed, as Jaynes (2003, p.310) puts it, "the functional use of induction in science is not to tell us what predictions must be true, but rather what predications are most strongly indicated by our present hypotheses and our present information". Accordingly, we found it necessary to replace the falsifiability criteria by *irrationalability* criteria. That is, the statements of a scientific theory should be in such logical form that they can be shown to be irrational by experience. This allows for development of a probabilistic theory in which all the statements fulfil the irrationalability criteria. Indeed, since a probabilistic statement can only be shown to be irrational and almost never possible to falsify.

1.2 The Structure of Thesis

The thesis is divided into three major parts. The first part is devoted to state of the subject and fundamental tools being used in the rest of the thesis. The second and third parts are concerned with development of a mathematical framework for Sampling and Mixing theory, respectively. The main material of the thesis is organized in the following manner:

- **Chapter 2:** A general exposition of the literature on the subject is given. The chapter begins with a short exposition of the parameters which affect the final state of the mixture. The mixers are further divided into two major classes, batch and continuous. In each case, some of the existing models in literature are reviewed.

- **Chapter 3:** Our approach to irrationalability criteria rests on Bayesian interpretation of probability theory. Therefore, in this chapter a short presentation on the most basic concepts of the subject relevant to this work is given. This chapter begins by exploring some of the historical aspects of the subject and the arguments in the favour for revision of the classical interpretation of the probability. This follows by a short introduction to the basic rules of the Bayesian probability theory in which concepts like prior and posterior probability and their relation to Bayes' theorem are explained in some details. It is shown how these basic concepts result into some of the most powerful tools that Bayesian interpretation can offer. More specifically, a short introduction to the principal of maximum entropy, marginalization and model selection is provided. The chapter is concluded by explaining the now celebrated entropy concentration theorem.

- **Chapter 4:** The main objective of this chapter is to develop a consistent model for mixture heterogeneity. In this context, the meaning of representative sample in relation to evaluation of mixture quality is discussed. Further, a mathematical model which quantifies the most important properties of a representative sample, i.e., accuracy and reproducibility, is developed. In this relation, it is shown that reproducibility can be used as a measure for mixedness and mathematically it can be modeled by the entropy of the sample distribution. Further, from the sampling point of view, the mixing systems are divided into two categories of open and closed systems. In each case a mathematical model is developed. It is demonstrated that the same mathematical formalism can be adopted to model both type of systems and the only difference is in how the constraints are assigned. A set-theoretical approach to the concept of mixture heterogeneity is also established and it is demonstrated how mixture heterogeneity can be quantified. This chapter is concluded by determining the relation between the pdfs of different sample spaces of different orders. These results are summarized in two theorems as, fundamental theorem of sample spaces and fundamental theorem of ensembles.

- **Chapter 5:** The objective of this chapter is to demonstrate some of the established facts about sampling of particulate mixtures and hence demonstrate the consistency of the framework developed in Chapter 4. In this context, Lacey's conjecture is stated and is shown that it is mathematically consistent within the developed framework. One major consequence of this analysis is stated in the mixture heterogeneity theorem. Moreover, a new model is developed which can be applied to estimate the number of tracer particles of given size in random binary closed batch systems. It is further shown how the same techniques can be used in determining the RTD of the binary closed continuous systems. This chapter is concluded by demonstrating how the developed model for the RTD of a binary closed continuous system can be applied to model the RTD of a commercial twin screw extruder.

- **Chapter 6:** In this chapter a new approach to the dynamics of the mixing processes is introduced. In order to do so, a short introduction to some of the most commonly used concepts in theory of complex dynamic systems is presented. Based on these concepts, the theory of punctuated equilibrium is introduced and is shown how it can explain the dynamics of mixture of particulate materials. Furthermore, it is demonstrated how all these concepts can be unified under the concept of heterogeneity landscape. A mathematical classification of all the valleys in a heterogeneity landscape is also achieved through the introduction of the heterogeneity equation, which is summarized under the fundamental theorem of mixing. This chapter is concluded by a discussion on a method for determining the coefficients of the heterogeneity operator.

- **Chapter 7:** The objective of this chapter is to demonstrate some of the direct consequences of the analysis of the previous chapter. It is shown that the choice of pdf for heterogeneity function being normal is not an assumption but a direct consequence of insufficient information on the structure of the mixture. It is also shown how inference on the mixture heterogeneity can be conducted and in the case of the pdf for heterogeneity function being normal, Gy's estimate for mixture heterogeneity is reproduced.

In the Appendix A, an alternative approach to determination of the mean residence time is given. This method was discovered during our literature study of the subject. It is surprising that no one in literature has ever considered to study the continuous mixers from the Queueing theory point of view. Therefore, we found it necessary to add this appendix as a supplement to the material in Chapter 2.

We have in many occasions found it necessary to explain the sources of inspiration, historical facts, related results in other fields or philosophical bases for our approach, which in general are not directly related to the subject of this thesis. For this reason we have placed all such material under the sections with the title of "Comments". These sections should not be considered as a part of thesis but as a source of additional information. We hope that this would lead to constructive discussions which would hopefully result into further advances.

1.3 Acknowledgement

I would like to express my gratitude to the late Prof. de Silva for all his support and encouragements during the start of this project. I would also like to thank my supervisor, Prof. Tüzün, for his constructive comments and his patience during the whole period of this work. I am also grateful for all the help I got from people at the University of Surrey during my stay at Guildford. I would like also to thank Norwegian Research Council for their 3 years of financial support for this project. I am also grateful for support I got from Tel-Tek and specially people at the Department of Powder Technology, POSTEC. A special thank to Prof. Datta for his encouragements during the final stage of this project. Also I would like to express my gratitude to Prof. Enstad, although reluctantly, for taking over as my supervisor here in Norway after Prof. de Silva's early death. Also I appreciate all the support I got during my 3 years of employment at the Telemark University College. And last but not least, I would like to thank my two dearest friends Wenche and Simen for putting up with me during my ups and downs for the last 6 years.

Chapter 2

Literature Survey on Particulate Mixing

> The Master said, Yu, shall I tell you what knowledge is? When you know a thing, to know it, and when you do not know a thing, to recognize that you do not know it. That is knowledge.
>
> K'ung Fu-tsu, Analects of Confucius (500 BC)

In general a batch of particulate material can be considered as a set of units which consists of solid fragments. By definition, it is *homogeneous* if all the units are strictly identical. It would be *heterogeneous* if the units are not strictly identical to each other (see Ghaderi 2003). A good example of an apparently homogeneous set is a batch of calibrated ball-bearing balls. Even though the difference in diameter of each ball is in the sub-micron range the balls are not strictly identical. The batch is therefore heterogeneous. Small differences between units, in a batch of particulate material could have a profound effect on the outcome of the process. This is a well established fact which is proven in many applications, see for example Abou-Chakra and Tüzün (1999). Thus, in reality, one always has to consider a batch of particles as a mixture of non-identical solid fragments. Therefore being able to characterize mixtures should be one of the most important objectives of research in bulk solid handling. There is a large number of articles and books which covers different aspects of the subject. In this chapter, however, only a small portion of the literature, which is believed to be most relevant to the present work is reviewed. Accordingly, the chapter begins with a short exposition of the parameters which affect the final state of the mixture. The mixers are further divided into two major classes, batch and continuous. In each case, some of the existing models in literature are reviewed.

2.1 Mixing Parameters

Mixing of particulate materials is a complex process in which characteristics of the particles, mixing equipment and operating conditions can influence the tendency to mix and de-mix. Therefore, the degree of mixedness and the rate of mixing are functions of many variables related to these characteristics. Segregation of particles usually occurs when mixing particles with different sizes. Larger particles stay at the top of the mixture, and smaller particles sink to the bottom. Segregation is made possible by the gaps that open up around particles when they are shaken or made to flow, allowing percolations to occur. Large particles tend to move toward less dense regions of smaller particles and this is why large heavy objects can rise to the top of vertically shaken layers of particulate materials. The above mechanism is known as *percolation.* Campbell and Bridgewater (1973) have investigated the mixing of dry particles by percolation. Percolation is also associated with *convection* and *shear* movements. The convective movement is defined as movement of adjacent particles from one location to another in the mixture, while the shear is defined as slipping movement of particle planes within the whole volume. A recent study by Shinbrot *et al.* (1997) of particles in a vibrating container indicate that the qualitative mechanism of convection may simply be that the penetration of particles into voids in the bulk depends on relative velocities of adjacent particles. The role of this velocity differential is to provide additional scattering opportunities that may permit a particle from one side or the other to enter the void. Shinbrot *et al.* conclude that this observation explains the mechanism that drives particles toward regions of low shear.

Difference in density of particles can also lead to segregation. In a recent experi-

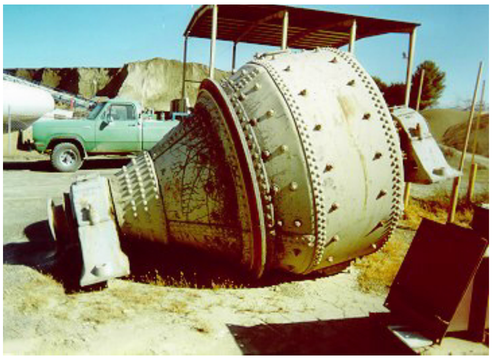

Figure 2.1.1: In a Hardinge Ball Mill, the large pebbles segregate to the part with the larger diameter and small pebbles segregate to the part with smaller diameter.

ment, Burtally *et al.* (2002) demonstrated that when a mixture consisting of particles of equal size but different density is vibrated up and down, it can completely separate out. However, Campbell and Bauer (1966) have reported that for density ratio less than 3:1, size distribution exerts greater influence on mixing and demixing than the differences in density and shape. In this relation, Egermann and Orr (1983) reported suggestions on particle size to reduce or eliminate segregation.

Particles possessing small angles of repose exhibit good flowability and small coefficient of friction. Good flowability is not necessarily of advantage for mixing. Fan *et al.* (1970) have pointed out that segregation may occur due to their rapid movement.

Dimension and geometry of a mixer and those of an agitator have influence on the particle flow pattern and flow velocities. For instance, if one considered a cylinder rotating on its horizontal axis, quite small variations to that shape could convert the machine from a mixer to a separator. For example, if one end of the cylinder were changed to a cone, the resulting system is the very well-known and well-established piece of equipment, the Hardinge Ball Mill (see Figure 2.1.1). It is well-known that during operation, the large pebbles segregate to the part with the larger diameter (mainly the cylindrical part) and small pebbles segregate to the part with smaller diameter (mainly the conical part).

Interactions between particles and the construction materials and surface finishes of a mixer may produce static charges and hence cause agglomeration. Friction between particles and the surface of a mixer has some effect on mixing and demixing. Broadbent *et al.* (1993) have investigated some of these issues. They studied the mixing process in a batch mixer by introducing radioactive tracer particles into the mixture which in turn was monitored by a positron camera. They found, among other things, that there were significant differences in behaviour of the particles in the central region and the two end regions of the mixer, as well as significant differences in the behaviour of particles in the two end wall regions even though these were superficially mechanically similar. They concluded that the mixer performance is extremely sensitive to the mixer design.

Type, location, and number of loading and emptying devices may enhance or hamper the mixing action. The operating conditions of a mixer can also greatly affect the degree of mixedness and the rate of mixing. Carley-Macauly and Donald (1962) studied some of these effects on mixing and demixing in tumbling mixers. However, it appears to be that Rose (1959) was the first investigator to systematically study the relationships between the mixing process and the physical parameters. He attempted a dimensional analysis by assuming that mixing and demixing occur simultaneously in a mixer. He found that the parameters that can be considered to be related to mixing were:

- Diameter of the mixer
- Speed of rotation of the mixer
- Mean diameter of particles
- Mean coefficient of friction of the particles
- Gravitational force

and those parameters considered to be related to demixing were:

- Diameter of the mixer
- Speed of rotation of the mixer

Characteristics of Particles	Characteristics of Mixing Equipment
Particle size distribution	Mixer dimension and geometry
Particle shape and surface characteristics	Agitator dimension
Bulk density and particle density	Construction materials and surface finishes
Moisture content	Type, location and number of loading and emptying devices
Angle of repose	**Operating Conditions**
Coefficient of friction particles	Weight of each constituent added
Friability	Ratio of volume of the mixture to that of the mixer
State of agglomeration	Method, sequence, place and rate of adding constituents
Flowability	Mixer or agitation speed, if any

Table 2.1.1: Variables influencing the state of a mixture, summerized by Fan et al. (1970).

- Mean diameter of the particles
- Difference of the mean sizes of the constituents of the mixture
- Mean density of the mixture
- Difference of the densities of the constituents of the mixture

Fan *et al.* (1970) reported that Rose's analysis has been also confirmed experimentally by other investigators. A list containing some of the above mentioned variables is given in table 2.1.1.

In practice, the above list is incomplete and could be extended indefinitely. In a 1964 review, which was based on a survey for the Institution of Chemical Engineers in the UK, Bourne (1964) pointed out that "far too little is known of how the interactions between the geometry of a mixer, its flow pattern, and physical proportion of materials determine the rate and degree of mixing and the laws of scale-up" and that "experiments with powders will probably be difficult to interpret and generalize quantitatively until the laws of particle flow and force transmission in powders become established, even for simple geometries". After 40 years and hundreds of research papers, most of the Bourne's conclusions and recommendations are still valid. In this sense, any model would necessarily be based on only some of the parameters that influence the final state of a mixture. However, the reliability of models will depend on the degree of independence between the parameters included in the model and the ones that were not included. Nevertheless, the proposed models should be flexible enough in order to be modified to take into account the new knowledge, which is as yet unknown.

2.2 Batch vs. Continuous Mixing

Mixers, in general, can be divided into Continuous and Batch mixers. *Batch* or *discontinuous* is characterized by the fact that the mixer is filled with the ingredients, then started and after a certain mixing time, the mixture is discharged, see for example Weinekötter and Gericke (2000, p.67). The feeding, mixing and discharging operations are performed one after the other. Whilst, in *continuous* mixing process the ingredients are continuously fed into the mixer then mixed and the product prepared for the next processing stage. However, the operations of feeding, mixing and discharging follow each other locally but contemporaneously.

It is difficult to generalize the relative merits of batch and continuous mixing because each particular application has its own special requirements and constrains. However, one can consider the advantages and disadvantages of each process as follows, see de Silva (1997), Schofield (1975) and van Heusden (1993):

- In continuous mixing high throughputs can be achieved with relatively small units.
- Continuous mixers have lower power requirements due to small hold-up.
- Continuous mixing requires less handling than batch mixers.
- In continuous mixing there is less chance of segregation after the mixer.
- Continuous mixers, if properly designed, can have short residence times, resulting in small space requirements, but are unable to deal with fluctuations in feed compositions. Expensive control systems become necessary in order to ensure that such fluctuations do not occur.

- Control loops for continuous mixers must have very short response times.
- Continuous mixers are very sensitive to malfunction of system components.
- The design of a continuous mixer is often restricted to one application and adaptation to other duties is often difficult.
- Successive operations are not possible in continuous mixers.

The relative importance of these various advantages and disadvantages of batch and continuous mixing depend upon the process it is likely to be employed in. Traditionally, mixing has been a batch operation, but requirements for large throughputs with a constant composition in an uninterrupted process are growing. This, in turn, has resulted in an increased interest in continuous mixers/mixing.

2.3 Review of Models on Batch Mixing

Most of the models and characterization of the batch mixers are either empirical or stochastic in nature. This in a sense reflects the difficulties in delineating the inherently complex nature of solid mixing processes by means of the deterministic approaches.

In general, the mixing and segregation mechanisms can be viewed either macroscopically or microscopically, see Fan *et al.* (1990). In the macroscopic approach, the movement of solids are visualized as that of coherent clumps or blocks of powders, whilst in microscopic approach, as interparticle percolation in the regions known as *failure zones* existing between moving blocks, see for example Bridgewater (1976).

Lacey (1954) has suggested three mechanisms contributing to creation of a mixture; (i) *convective mixing*: the transfer of groups of adjacent particles from one location in the mixture to another, (ii) *diffusional mixing*: the interpenetration and random local mingling of individual particles, (iii) *shear mixing*: the slipping of particle planes within the whole mixture, in which it changes the relative position of the constituting particles.

Nonetheless, as Poux *et al.* (1991) have also pointed out, although a specific mechanical action needs to be considered, shear mixing can be considered as a specific combination of the convective and diffusional mixing.

As mixing proceeds, all the aforementioned mechanisms function to some extent. Different mixers give different predominating mechanisms. Bourne (1964) noticed that convection alone always produces a mixture consisting of streaks or patches of individual components, if the size of the regions scrutinized is sufficiently small. On the other hand, diffusion is a mechanism whereby patches are diluted by the other components. Diffusion, which occurs simultaneously with sub-division and convection, acts across the boundaries of patches so that in practice the boundaries are not sharply defined. Moreover, convection increases the rate of diffusion by increasing the surface area of patches and decreasing diffusion paths. Nevertheless, particles in the mixture can only migrate from or to patches while the mixture is agitated. This means that the rate of migration depends on the nature of agitation and so it depends upon more than the physical properties of the material.

2.3.1 Macroscopic Approach

Most macroscopic approaches are based on modelling of above mechanisms. For example, Lacey (1954)argued that the behaviour of materials in mixers that repeatedly spread particles over freshly exposed surface of the mix is very similar to ordinary molecular or thermal diffusion. Based on this argument he applied the *Fick's law* of diffusion (Ficks, 1855) to model diffusional mixing. The *Fickian diffusion* model is mathematically described by

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2}, \tag{2.3.1}$$

where C is the concentration of the key component, D is the diffusivity, x is the distance in the direction of dispersion and t is the time. However, in order for Fickian diffusion model to work, the mixture ingredients should not behave differently in the way they move; i.e. the segregation tendencies should be negligible. In fact, Carstensen and Patel

(1977) have experimentally demonstrated that in the case of binary non-spherical particles having rough surfaces and different mean diameter, the Fickian diffusion model is not valid. Moreover, as Harnby (1967), points out, in general, Fickian diffusion model is not applicable to the mixing of a real system resulting from interaction between the mechanisms of mixing and segregation.

For this reason a new mathematical model, based on Fick's law of diffusion was developed by Strek *et al.* (1978), which was used in describing a two-component system mixed radially in a horizontal rotating drum mixer, accompanied by segregation. In this model, which is also known as *diffusional segregation* model, the concentration of the key component, $C\left(r,t\right)$, is considered to be the sum of two independent component functions called the *homogenization* and the *segregation* functions; i.e.

$$C\left(r,t\right)=C_h\left(r,t\right)+C_s\left(r,t\right). \tag{2.3.2}$$

Applying Eqn. (2.3.1) in cylindrical coordinate to both $C_h\left(r,t\right)$ and $C_s\left(r,t\right)$ leads to

$$\frac{\partial C_h\left(r,t\right)}{\partial t}=D_h\left[\frac{\partial^2 C_h\left(r,t\right)}{\partial r^2}+\frac{1}{r}\frac{\partial C_h\left(r,t\right)}{\partial r}\right] \tag{2.3.3a}$$

$$\frac{\partial C_s\left(r,t\right)}{\partial t}=-\frac{D_s}{t^k}\left[\frac{\partial^2 C_s\left(r,t\right)}{\partial r^2}+\frac{1}{r}\frac{\partial C_s\left(r,t\right)}{\partial r}\right] \tag{2.3.3b}$$

where D_h and D_s are diffusional homogenization and segregation coefficients, respectively, and $h>1$ is a dimensionless constant. These equations can be solved for $C\left(r,t\right)$ under the assumptions of unity loading and complete segregation as the final state of the system.

The diffusional segregation model, besides being specific to certain type of mixers, introduces a new way of considering the mixing and segregation. In this approach, one considers the effect of mixing and segregation to diffuse, independently, through the mixture and depending on the mixture parameters, determine the final state of the mixture. This is indeed a radical approach, which contrary to previous approaches, focuses on modelling the effects using the machinery which was previously only used for studying the causes. However, judging from literature, this model despite having the potential of providing a unified mathematical description of both mixing and segregation processes, did not receive the attention it deserved. It is also difficult to judge if the authors had a full overview of the consequences of their radical approach. They also admit that more experimental investigations are necessary in order to check the model. Unfortunately, not much further work in this direction was ever conducted posterior to their publication.

Sommer (1979) has also noticed that in presence of demixing tendencies, the Fickian diffusion model is inadequate. He suggested that a better description of the mixing process is to assume that the mass transport in the mixer consists of two components; transport by convection and transport by dispersion. Based on this simple assumption, he demonstrated that this model, which is now known as *diffusive-convective* model, is better described by a generalized version of the Eqn. (2.3.1); i.e.

$$\frac{\partial C\left(x,t\right)}{\partial t} = -\frac{\partial}{\partial x}\left[v\left(x\right)C\left(x,t\right)\right] + \frac{\partial^2}{\partial x^2}\left[D\left(x\right)C\left(x,t\right)\right], \tag{2.3.4}$$

where $v\left(x\right)$ and $D\left(x\right)$ are convection and dispersion transport coefficient functions, respectively. Equations of type (2.3.4) are known as Fokker-Planck equations and were first introduced by Fokker (1914) and Planck (1917) in relation to problems in statistical mechanics. In order to validate the model, Sommer observed the mixing of ballotini in a glass drum. Based on the experimental observations he showed that the Eqn. (2.3.4) could be used to describe the mixing mechanisms. He also demonstrated that the diffusive-convective model could sufficiently describe previous experimental findings by Müller (1966). Hwang *et al.* (1980) also used the diffusive-convective model in studying the mixing of dry powders during flow over an inclined surface. The validity of the model was verified through a series of experiments using different chute lengths and different feeding conditions.

Thus, in general, it appears that the diffusive-convective model is better than the Fickian diffusion model. However, although this model has been successful in describing the mixing processes consisting of non-spherical particles, in deriving the Eqn. (2.3.4), Sommer assumed that the mixture consisted of completely identical spheres which were independent of one another. It seems that this assumption, at least in cases studied, is not necessary. Nevertheless, no convincing argument has ever been produced to confirm the redundancy of the assumption.

2.3.2 Microscopic Approach

In microscopic approaches, one constructs models in which interaction between particles and their relation to the bulk properties can be studied. The most common method is known as *discrete element method* (DEM). DEM is a computer based simulation method in which the main goal is to establish a link between experimental observations and behaviour of the bulk at particle level. In DEM, particles are assumed to undergo collisions obeying classical Newtonian mechanics. Hence, the problem reduces to one with N body interaction where no closed form solution exist (except for $N = 2$). Theoretically however, it is possible to integrate the coupled equations of motions in time using numerical techniques. Therefore, the size of the system under study is restricted by the computational power available.

In a recent article, Dury and Ristow (1999), applied DEM in studying the dynamics of the size segregation process of binary particle mixtures in three-dimensional rotating drums, operated in the continuous flow regime. The drum used in the simulation had a diameter of 70 mm and a height of 25 mm. It was filled with a binary mixture of large beads having radius of 1.5 mm and small beads $r \in \{0.75 \text{ mm}, 1.0 \text{ mm}, 1.25 \text{ mm}\}$. The total number of particles used were up to 17000, which is quite small comparing with the mixtures in practice. Nonetheless, they could demonstrate that when a rotating drum is operated in the continuous flow regime, the size segregation will take place for arbitrary small differences in particle size. Moreover, they could show that the highest achievable segregation can be obtained for a slightly more than half-filled cylinder and therefore least mixing. These results which also were previously shown experimentally by Metcalfe *et al.* (1995), could be interpreted as a confirmation that the assumed model in the particle level is correct. However, as Baxter (1998, p.50) points out, it is perfectly possible that quite different models of the mechanics at particle level will result in very similar macroscopic predictions. Nonetheless, DEM is potentially powerful approach in gaining insight into the mechanisms governing mixing and segregation, provided that it can be shown to reproduce physical phenomena of real systems.

2.4 Review of Models on Continuous Mixing

In general the mixing actions involved in continuous mixers are divided in *radial* and *axial* mixing, see for example Schofield (1975). The radial mixing causes the intermingling of the parallel streams, which in turn, reduces the *radial heterogeneity* of the mixture and axial mixing smooths out the time based fluctuations in the effective composition of the input, which reduces the *axial heterogeneity* of the mixture. Since the radial direction is usually much smaller than the length of the mixer, the contribution of heterogeneity along the axial direction is far greater than in the radial direction. Therefore, continuous mixers are characterized by their ability to smoothing out ingoing fluctuations (axial mixing).

In order to reduce the axial heterogeneity, the mixer delays the particles on their way out. This, in practice, means that the mixer forces the particles not to take the geometrically shortest way out but rather more tortuous paths. This can be either achieved by allowing the mixture ingredients to pass over structures which disturb the flow and divert the particles or simply by dilating the mixture in order to initiate the equivalent of Brownian motion in fluids.

2.4.1 Residence Time Distribution

Continuous mixers are, in general, characterized by quality of the mixture that has passed a certain point in the mixer (usually the outlet), within a certain duration of time. In this relation, each slice of material that flows out of the mixer can be thought as a sample. The heterogeneity within each sample is not important and is related to radial mixing. Therefore the mixers are characterized by their ability in smoothing out in going fluctuations; i.e. reduction of axial heterogeneity. This property is often related to *residence time distribution* (RTD); the distribution of times that elements of the flow of material spend in the mixer.

As Shinnar *et al.* (1972) have also pointed out, the RTD-model based approach has become an important analytical tool in the study of various processes. The subject is especially important in design of continuous reactors. Although, the concept was originally developed for fluids, it has been extensively used in the literature on continuous mixing of particulate materials. The reason for this is the natural connection between the time each particle spends in the mixer and the quality of axial mixing.

To this end, notice that each particle in the outflow from the system possesses a previous history. The duration of the particle's stay in the system is definable in statistical term. With each distribution of residence time, one can define a function $F(t)$ with the following property: the fraction of particles possessing a residence time t or less is equal to $F(t)$. The fraction of particles whose residence time exceeds t is given by the complementary function $F^*(t)$, which is defined as

$$F^*(t) = 1 - F(t). \tag{2.4.1}$$

As Shinnar and Naor (1963) have also noticed, the functions $F(t)$ and $F^*(t)$ may also be interpreted as probabilities. Therefore, $F(t)$ can be considered as being the probability of a single particle staying in the system for a time t or less and $F^*(t)$ the probability that the particle's residence time exceeds t. Now, notice that $\Delta F(t) = F(t + \Delta t) - F(t)$, is the probability of a particle having RTD in $< t, t + \Delta t >$. But

$$\Delta F(t) \approx \frac{dF}{dt}\Delta t = f(t)\Delta t. \tag{2.4.2}$$

Hence, if we let $\Delta t \to 0$, then $f(t)dt$ can be interpreted as the probability of a particle having a residence time between t and $t + dt$ or equivalently the probability of a residence time being found between t and $t + dt$. Thus $f(t)$ can be regarded as the probability density for RTD. It should be stressed that usually the functions f, F and F^* refer to the density and distribution of the RTDs of particles as they appear at the location of the exit from the system.

Models for Residence Time Distribution

In principle, all the information about the RTD is embodied in any one of the functions f, F and F^*. The difficulty arises when an attempt is made to translate the somewhat abstract information into more physically meaningful concepts. The most common method is to compare the experimental distributions with functions obtained from some idealized theoretical models. One model which is frequently encountered is based on the assumption that as soon as a particle has entered the system (an intensely agitated vessel, for example) the probability of its leaving in the outflow becomes independent of past history and is thus constant. The *memoryless* property is precisely a property of exponentially distributed random variable which can be used to model RTD, see for example Arnold and Huang (1995):

$$f(t) = \begin{cases} \mu e^{-\mu t} \ , \ t > 0 \\ 0 \ , \ t \leq 0 \end{cases} \tag{2.4.3}$$

where $\mu > 0$, is the reciprocal of the *mean residence time*,

$$\mu = \frac{1}{\bar{t}} \cdot \tag{2.4.4}$$

This type of mixing and the ensuing RTD are sometimes referred to as belonging to the *ideally mixed* or *exponential* vessel.

If there are several identical and independent exponential vessels with each having the same mean residence time, connected in series, then the RTD can be modelled by *n-Erlang distribution* (see Dougherty, 1990, p.181),

$$f(t) = \begin{cases} \frac{(n\mu)^n}{(n-1)!} t^{n-1} e^{-n\mu t} \ , t > 0 \\ 0 \ , t \leq 0 \end{cases} \tag{2.4.5}$$

where $\mu > 0$, is the same as the one in Eqn. (2.4.4) and $n > 0$, is the number of the vessels. Moreover, it can be shown that the mean residence time of a vessel is given by $\bar{t}/n$, in which $\bar{t}$ is the mean residence time of the whole system.

On the other extreme, one has a behaviour characterized as *plug flow.* This situation can be described as first-in-first-out (FIFO). This means that particles spend same amount of time, $\bar{t}$, in the system. Thus, in this case $f(t)$ can be described by a delta function. It is easy to show that a plug flow mixer can be represented as infinite serially connected exponential vessels. Indeed, notice that the Laplace transform of f for n serially connected exponential vessels is

$$\mathcal{L}\{f(t)\} = (\mu n/(\mu n + s))^n. \tag{2.4.6}$$

Then, it can be shown that

$$\lim_{n\to\infty} \mathcal{L}\{f(t)\} = \exp(-s/\mu) \Rightarrow \lim_{n\to\infty} f(t) = \delta\left(t - \mu^{-1}\right). \tag{2.4.7}$$

However, true plug flow does not exist in actual mixers because each particle undergoes a randomwalk brought about by convection and diffusion. Therefore, it is useful to introduce the concept of *near plug flow* to describe such situations. This means that a system in the state of near plug flow can be approximated by a large, but finite, number of serially connected exponential vessels.

The cases in which n vessels are connected in parallel have also being studied. It can be shown that in these cases, the RTD density can be described by *n-Hyperexponential*

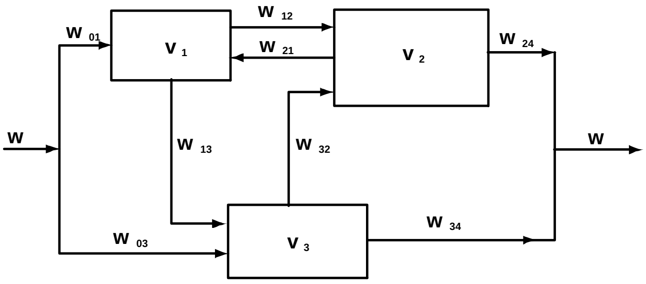

Figure 2.4.1: A typical flow network, Shinnar *et al.* (1967)

distribution (see for example Allen, 1990, p.148),

$$f(t) = \begin{cases} \sum_{i=1}^{n} q_i \mu_i e^{-\mu_i t} \ , \ t > 0 \\ 0 \ , \ t \leq 0 \end{cases} \tag{2.4.8}$$

where

$$\sum_{i=1}^{n} \frac{q_i}{\mu_i} = \frac{1}{\mu}, \tag{2.4.9}$$

and $\mu^{-1} > 0$, is the mean residence time and $q_i > 0$, is the fraction of flow passing through the vessel with μ_i^{-1} as its mean residence time.

The down side to these models is that as soon as the number of elements increase, the calculation becomes more tedious. Mathematical techniques like Laplace transformation makes such calculation easier and can give an alternative picture of the model which is easier to grasp. However, more complicated networks of exponential vessels are possible. In the most general case the model consists of n ideally mixed vessels arbitrarily connected by interstage flows where the volume of the ith vessel is v_i and the volumetric flow rate from the ith to jth vessel is w_{ij} $(i, j = 1, 2, \ldots, n)$, as shown in Figure 2.4.1. The inlet stream is distributed to the vessels arbitrarily. By adjusting these values, one can fit the model with the experimental data. Shinnar *et al.* (1969) have used a method of analysis based on randomwalk model and tracer experiments to determine the model parameters. This method, like most others, has the weakness that it becomes practically impossible to handle when the number of connections and vessels increases beyond a certain limit.

Little else exists in the literature on how to handle such complex models. Nonetheless, as Shinnar and Naor (1963) have pointed out, all actual distribution functions may be approximated by a theoretical model composed of a number of exponential and near plug flow vessels connected in some network. This is of course mathematically equivalent to the statement that all well behaved functions can be approximated by some power series.

The use of other models has also been reported by Fan and Wen (1975). A particular one which is of some potential interest is the *Gamma distribution*, see Dougherty (1990, p.176),

$$f(t) = \begin{cases} \frac{(\alpha\mu)^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\alpha\mu t} \ , \ t > 0 \\ 0 \ , \ t \leq 0 \end{cases} \tag{2.4.10}$$

where $\mu^{-1} > 0$, is the mean residence time of the system and $\Gamma(\alpha)$ is the Gamma function defined as,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \ , \ \alpha > 0. \tag{2.4.11}$$

Gamma distribution is just a generalization of $n-$Erlang distribution. Indeed, by restricting the values of α to positive integers, the Gamma distribution reduces to $n-$Erlang. Quite a variety of shapes can be described by different choices of α and μ. This can be an advantage in fitting the data.

The representations discussed in this section may have a physical justification in the flow characteristics of the system. Alternatively, they may be artifices which help to illustrate the basic features of the system. However, one should be careful not to give undue weight to $f(t)$ obtained by any arbitrary models.

2.4.2 Escape Probability Density

Wang (1986) pointed out that mixing behaviour of a majority of the actual mixers deviates from the ideal mixer (exponential vessel). He argued that this deviation may be caused by non-uniform velocity profile, velocity fluctuation due to diffusion, short-circuiting, by-passing and channelling of particles, by the presence of segregating component or stagnant regions caused by mixer geometrical shape and internals, or by the recycling of solid particles within the mixer as a result of impeller and helix design. These are evidently part of a reason for complicated modelling networks.

Shinnar and Naor (1963) have suggested the *intensity function* or *escape probability density* as a method of visualizing the features of RTD related to stagnancy. In the context of modelling based on n- Erlang and n- Hyperexponential distributions, *stagnancy* is generally associated with systems in which total flow may be decomposed into flows connected in parallel where one of the components has a significantly larger average residence time than the other. Now, on viewing a particle which has just entered the system, as mentioned previously, the probability of its leaving the system within the time interval $< t, t + dt >$ is equal to $f(t)dt$. However, a slightly different problem may be posed; a particle has already stayed in the system for a time t, one wishes to know the probability of the particle leaving the system within the next time element dt. To this end, let this probability be denoted by $\lambda(t)dt$. This function $\lambda(t)$ may be evaluated by the following reasoning. Notice that on one hand, the probability of a particle leaving the system within t and $t + dt$ equals $f(t)dt$; whilst on the other hand, this probability is the product of two other terms:

- $F^*(t)$ is the fraction of particles whose RTD exceed t.
- The probability of a particle leaving between t and $t + dt$, assuming it is of age t, that is $\lambda(t)dt$.

Therefore

$$f(t)dt = F^*(t)\lambda(t)dt, \tag{2.4.12}$$

which implies

$$\lambda(t) = \frac{f(t)}{F^*(t)} = -\frac{d}{dt}\ln F^*(t). \tag{2.4.13}$$

Analogous transforms of probability density functions are commonly called intensity functions and are well known in different statistical contexts, see for example Johnson *et al.* (1995). In general, the intensity function $\lambda(t)$, reveals more features of RTD and , as Shinnar and Naor have also demonstrated, it is a more natural function to study than $f\left(t\right)$. Moreover, there is a one to one correspondence between $\lambda(t)$ and $f(t)$. Indeed, it is straightforward to show that

$$f\left(t\right)=\lambda\left(t\right)\exp\left[-\int_{0}^{t}\lambda\left(\tau\right)d\tau\right]. \tag{2.4.14}$$

Physically $\lambda(t)$ is a measure of the probability of escape for a particle which has stayed in the system for a period t. Thus, by definition, $\lambda(t)$ should be constant in the case of ideally mixed vessel as the chance of escape is independent of previous history. This is indeed the case, which can be deduced from Eqn. (2.4.3). Thus any departure of $\lambda(t)$ from constancy is an indication of ill-mixedness. Indeed, as Shinnar and Naor (1963) have demonstrated, a system with stagnancy is one in which the escape probability (or the intensity function) decreases in time over some interval. For example, imagine a system in which a considerable fraction of the particles moves in near plug flow, whereas the remaining fraction is absorbed into a stagnant phase from which it is exuded later into the main stream. Any particle which has remained in the vessel for a time exceeding the time of the plug flow has a high chance of being in the stagnant phase and, therefore, a low escape probability and intensity function. In comparison, a particle which has stayed in the system for duration slightly shorter than the plug flow time possesses a high escape probability. The corresponding λ- curve would show a definite maximum even though the plug flow may be somewhat distorted by the mixer. The same result is obtained, if a fraction of the particles passes through a region with an extended residence time.

2.4.3 Mean Residence Time

Another useful quantity is the mean residence time, $\bar{t}$. This value corresponds to the first moment of t, i.e.:

$$\bar{t} = \int_0^\infty t f(t) dt = \int_0^\infty F^*(t) dt, \tag{2.4.15}$$

where $F^*(t)$ is as defined in § 2.4.1 and the last equality achieved by noticing that

$$\int F^*(t) dt = \int t f(t) dt + t F^*(t) \text{ and } \lim_{t \to \infty} t F^*(t) = 0. \tag{2.4.16}$$

The mean residence time $\bar{t}$, is basically the average time that each element spends in the vessel. This number is often used to characterize the continuous mixers.

Now, let $V(t)$ denote the total volume of the mixture, in the mixer at time t, with inlet and outlet flow rates $u_i(t)$ and $u_o(t)$, respectively. From the conservation of mass follows that

$$\frac{dV(t)}{dt} = u_i(t) - u_0(t). \tag{2.4.17}$$

Based on the above equation, Shinnar *et al.* (1969) deduced that

$$\bar{V} = V_0 + \bar{u_o} \bar{t}, \tag{2.4.18}$$

where V_0 denotes the total stagnation volume in the system (assumed to be constant). The most noticeable aspect of their approach is that they assume no statistical considerations, nor any general considerations of stationarity but instead carry out simple averaging in time, over what may be regarded as a very long transient from the moment the mixing vessel is brought (empty) on stream till it is retired (empty) from service. However, as the authors also admit, their proof is more intuitive and based on some heuristic arguments. In the Appendix A a more rigorous proof of this result based on some well established results from the Queueing theory is given.

2.4.4 Variance Reduction Ratio

Another method of the characterization which is related to RTD of a continuous mixer is, *variance reduction ratio* or in short VRR. It is defined as

$$VRR = \frac{\sigma_{in}^2}{\sigma_{out}^2}, \tag{2.4.19}$$

where σ_{in}^2 and σ_{out}^2 are the variance in concentration of the key component at the input and output of the mixer, respectively. Danckwerts (1953) demonstrated that VRR, in the case of steady-state flow, could be used to characterize how effectively a mixer reduces the time-based input fluctuations. Indeed, let $R_{in}(\tau)$ denote the *autocorrelation* function of the input and $f(t)$ the density of residence time of the mixer. Danckwerts showed that

$$\frac{1}{VRR} = \frac{\sigma_{out}^2}{\sigma_{in}^2} = 2\int_0^\infty \int_0^\infty \rho_{in}(\tau) f(t) f(t+\tau)\, dt d\tau, \tag{2.4.20}$$

where the *autocorrelation coefficient,* $\rho_{in}(\tau)$ is defined as

$$\rho_{in}(\tau) = \frac{R_{in}(\tau)}{\sigma_{in}^2}, \tag{2.4.21}$$

which can be shown to have the following property (Dougherty 1990, p.245),

$$|\rho_{in}(\tau)| \leq 1 \text{ for all } \tau. \tag{2.4.22}$$

Assuming that $R_{in}(\tau)$ and $f(t)$ are known, the variance of outgoing composition can be determined by Eqn. (2.4.20). Wang (1986) has also suggested that the variance obtained by this method can be used to construct the so called *quality control chart* in order to monitor the quality of the mixture, which could be interesting information from the process control point of view.

As it was already noticed by Lacey (1943), mixture heterogeneity reaches its minimum for a random mixture and is always nonzero. Based on this observation, Weinekötter and Reh (1995) suggested that Danckwerts model is only valid in the case of fluids and has to be modified in order to be valid for solid mixing. They suggested the following modification:

$$\frac{1}{VRR_{solids}} = \frac{1}{VRR_{fluid}} + \frac{\sigma^2_{out,\text{ideal feed}}}{\sigma^2_{in}}. \tag{2.4.23}$$

The first term on the right side is calculated from Eqn. (2.4.20) and the second term stands for feed in absence of fluctuations. However, Weinekötter and Reh only produced a plausible argument in defence of their suggested modification. They neither show or suggest, in a consistent way, how one can determine the value of each term in Eqn. (2.4.23).

To this end, Ghaderi (2003) demonstrated that Eqn. (2.4.23) naturally follows by assuming that the mixer being a *linear time invariant system* (LTI). Moreover, he showed how the second term in Eqn. (2.4.23) can be estimated from the input *variogram* and that mixer efficiency is properly characterized by (2.4.20); i.e. Danckwerts original formula.

2.4.5 Dispersion Model

The current research is mainly concentrated around what is known as the *dispersion model*. This model is an attempt to describe the mixing process based on the mixing mechanisms in the mixer. As mentioned earlier, in a mixer particles are forced to not take the geometrically shortest way out but rather more tortuous paths. These paths, which can be modeled by *randomwalks*, are the essential mechanisms in reducing the axial heterogeneity of the mixture. The dispersion model relates some of the statistics of these randomwalks to the performance of the mixer. This model is described by the transport equation:

$$\frac{\partial c_i}{\partial t} = -u_i \frac{\partial c_i}{\partial x} + D_i \frac{\partial^2 c_i}{\partial x^2}, \tag{2.4.24}$$

where $c_i\,(t)$ is the weight concentration of the key component, u_i is the axial transport velocity (also known as convective transport coefficient) and D_i is the dispersion coefficient. This is essentially the Fokker-Planck equation on page 18, with u_i and D_i being independent of x.

Several investigators have demonstrated that the mixing mechanisms in continuous mixers can be described by the dispersion model (Sommer 1979, Weinekötter and Reh 1995). This model is based on two assumptions:

1. Axial dispersion takes place only inside the mixer (closed vessel as boundary condition).
2. The heterogeneity in radial direction can be disregarded.

Sommer (1996) suggested that each key component in the mixture can be described by a separate Fokker-Planck equation. Furthermore, he points out that each component has its own unique dispersive coefficient and axial transport velocity. Obviously, in this context, it is reasonable to assume that if the components consist of nearly identical or very similar substances and differ, for instance, only in their colour, (u_i, D_i) pair for each component may be assumed to be equal.

The concentration c_i at point x changes within a certain time interval Δt. During this period of time, each particle receives several blows, either from other particles or from the mixer, which may change their direction of the motion. Now, let Δx denote the amount of displacement of a given key particle in axial direction within time Δt. Each particle at point x is displaced by a different amount Δx. Therefore, the displacements Δx are statistically distributed with a probability density function of $p\left(\Delta x\right)$. If the first and second moments of $p\left(\Delta x\right)$ are finite then

$$u_i = \lim_{\Delta t \to 0} \frac{E\left(\Delta x\right)}{\Delta t} \text{ and } D_i = \lim_{\Delta t \to 0} \frac{E\left(\Delta x^2\right)}{2\Delta t}, \tag{2.4.25}$$

where

$$\sigma^2\left(\Delta x\right) = E\left(\Delta x^2\right) - \left[E\left(\Delta x\right)\right]^2. \tag{2.4.26}$$

Thus u_i describes the mean rate of the displacement and D_i is deviation from this mean value caused by random fluctuations. Notice that in case, $D_i = 0$ all the particles are displaced equally and thus the plug flow behaviour can be expected.

If the Eqn. (2.4.24) is normalized with the length of the mixer l and the mean residence time of the key component, μ_i^{-1}, it follows that

$$\frac{\partial c_i}{\partial \tau} = \left[\frac{D_i}{u_i l}\right] \frac{\partial^2 c_i}{\partial \eta^2} - \frac{\partial c_i}{\partial \eta}, \tag{2.4.27}$$

where $\eta = x/l$; $\tau = t\mu_i$; $\mu_i = u_i/l$. The inverse of the expression in brackets is also known as *Peclet* or *Bodenstein* number (See for example, Weineкötter and Gericke 2000, p.91). Weinekötter (1993) has numerically solved this equation for a set of Peclet numbers. Based on these solutions he was able to calculate RTD for each corresponding Peclet number. Thus by comparing the experimental and theoretical RTDs, he achieved the characterization of the continuous mixers by their Peclet numbers. For very large Peclet numbers the mixer shows near plug flow characteristics whereas for small ones, the RTD resembles that of an ideal mixer. The RTDs determined from the dispersion model are similar to RTDs of serial cascade of exponential vessels presented in § 2.4.1. If this is true then there should be a relation between the Peclet number and the long-term behaviour of the escape probability density of particles, presented in § 2.4.2. However, no such relation has ever being established.

One of the biggest challenges a designer can face is the scale-up. As Fan *et al.* (1990) have also pointed out, it is not possible to totally rely on experiences gained from experiments with a pilot-scale mixer for scale-up. As one proceeds with scale-up, the hypotheses developed from and verified in pilot-scale mixer may become invalid. In order to overcome this problem, Fan and Wang (1974) suggested a procedure known as *principle of similarity*, in which the knowledge of Peclet number and mean residence time may be very useful. Indeed, the principle of similarity requires that the Peclet number and the mean residence time of the both pilot-scale and plant-scale mixers to be equal. However, in practice, this is only possible if the relations between the Peclet number, mean residence time, mixer parameters and the mixture characteristics are known. Unfortunately, judging from literature, at present this has not been fully achieved.

2.5 Epilogue: Comments

In a review of the subject, Fan *et al.* (1990) concluded that although considerable progress has been made in our understanding of solid mixing processes, the design of mixers for particulate solids has mainly been carried out heuristically. They attribute this to the complexity of particulate mixing behaviour which is only describable by a large number of parameters. In a recent book, Kaye (1997, p.8) claims that the failure of academic study to significantly improve the performance of mixing equipment is due to fundamental philosophical problems and not the inadequacies of the research investigations. He also points out that the source of the problem is the assumption that the problems of efficient design of mixing equipment are solvable in a deterministic manner, provided that we gained more understanding of the causes which contribute to the performance of the mixer. He concludes that systematic study of the dynamics of the mixing equipment is properly a branch of *deterministic chaos* or in short *chaos.* In general, processes which are very sensitive to small fluctuations are called *chaotic.* This is because their trajectories are in general very irregular, so that they give the impression of being random, even though they are driven by deterministic forces.

A quick review of literature supports some of Kaye's claims. First of all, there is no philosophical discussion on the issues that are faced by most of scientist in this field. Secondly, there exists a great deal of valuable results, mostly experimental, which would eventually fade into oblivion by time. The most important reason for this is that, as yet, there is no so-called *grand theory* of mixing. This means that we do not yet have the possibility of synthesizing our knowledge into a single theory. Bourne (1964) has pointed out that there are in general two difficulties in any study of mixing:

(a) the absence of theoretical background to aid the deduction of further generalizations;

(b) the absence of widely applicable criteria of mixing.

He warned the investigators that as long as above two difficulties are not resolved, one should not produce extensive experimental data which remain specific to the conditions studied. This is a great philosophical dilemma. Mixing of particulate materials has important applications in the modern process industry. However, in order for it to flourish as a science, it has to find a way to synthesize the existing knowledge in a unifying manner. This discussion is mainly absent from the main stream literature. It appears that a discussion on this issue is not fashionable and certainly not publishable. This is perhaps the main reason why many outsiders do not consider the mixing of particulate materials as a science. A similar situation has been experienced by biologists. For a long time many people did not recognize biology as a science. However, recent developments, which are as much philosophical as theoretical, have not only drastically reduced the number of skeptics but also turned biology to a discipline which has been making valuable contributions to our understanding of the dynamics of complex systems. Chaos theory has found its place in the biological systems. This is perhaps what Kaye also suggests that should happen in mixing.

However, some would perhaps disagree with Kaye that Laplacian determinism should be completely abandoned. In fact, most people find the idea that all events in the universe are connected by a cause-and-effect chain very compelling. It is still believed that theories should be based on principals which are not in conflict with common sense. Nevertheless, in practice a mixture in a given state is a result of interactions of many causes, which for all useful purposes, are difficult to determine. Therefore, it is perhaps meaningless to talk about a grand theory of mixing. However, despite this, it is still reasonable to seek a grand theory of existing knowledge which also should be flexible enough to be modified in order to explain new knowledge which is not known to us yet.

Roughly speaking, as Jeffreys (1973, p.190) put it, a theory that explains more than one fact is *useful*, one that explains no facts is *redundant*, and one that explains just as much as it assumes is *ad hoc* and tells us no more than the fact itself does. In order to have a useful theory of mixing, many researchers only consider hypothesis that are actual logical constructs from observations on the structure of the mixture. It appears that this idea is inspired from a form of idealism known as *phenomenalism,* which may be defined by the statement that nothing is to be supposed to exist that cannot be reduced to descriptions of sensations. According to Thorburn (1918), phenomenalism may be tracked back to the medieval writer William of Occam[1]. In the current scientific literature, this idea is usually known as *Occam's razor* or the law of *parsimony.* It is interpreted to mean that the simplest of two or more competing theories is preferable and that an explanation for unknown phenomena should first be attempted in terms of what is already known. However, in the literature on mixing of particulate materials, phenomenalism has been adopted to mean the following:

1. The analysis of suggested hypotheses should show what they actually say about experience.
2. If a hypothesis contains reference to quantities whose values do not affect the prediction of experience, then the hypothesis should be restated in such a way that these quantities do not appear in it.

Although the above points are quite compelling, unfortunately, in the literature the second point is overstated to the effect that everything mentioned in a hypothesis must be separately observable. This means that a hypothesis can never be stated until we have knowledge, by experience, of all its aspects, perceived and unperceived. Now, the question is why this obvious contradiction left unnoticed by most researchers in this field. One might conjecture that the main reason for this is that we, as yet, are not in possession of complete knowledge of the parameters which affect the state of a mixture. Therefore,

---

[1] William of Occam (also Ockham or any of several other spellings) (ca. 1285-1349) was a Franciscan friar and philosopher, from Occam a small village near Ripley, in Surrey, south-west of London, England.

most models, described in scientific papers on mixing, are at best semi-empirical. These models, in general, are constructed to describe the mixing operation under particular circumstances. For this reason, any inconsistencies are usually attributed to our incomplete knowledge. Indeed, in general, we have to restrict ourselves to theories which describe our incomplete state of knowledge. However, despite this, it is still reasonable to demand that our theory should have the flexibility of being modified to explain new facts, which are not known to us yet.

It appears that theories which are result of pure phenomenalism, in general, do not fulfil our flexibility criteria. In fact, Ramsey (1931, p.212) has argued that if the theoretical terms of a theory are logically constructed only out of observable entities, the theory will be incapable of being modified to explain new sorts of facts. It seems to us that Braithwaite (1953, p.62) have produced such type of admissible scientific hypotheses in support of Ramsey's argument. This implies that phenomenalism, in general, is not adequate for our needs and it requires further development or modification, before it can deal with problems of mixing.

Thus based on the above discussion, in order to have a useful theory of mixing, it must necessarily be abstract. Because only in this way, the theory will be capable of taking into account all the important aspects of the mixing process. Moreover, since this theory is based on incomplete knowledge of the parameters that affect the final state of a mixture, it cannot have any specific reference to particular circumstances in which the mixing operation takes place. In the following chapters, it is demonstrated how this theory can be constructed.

Chapter 3

Bayesian Probability Theory

Probability theory is nothing but common sense reduced to calculation.

Pierre S. Laplace, A Philosophical Essay on Probabilities (1814)

The major part of the present work, both from the philosophical and technical point of view, rests on Bayesian interpretation of probability theory. Therefore, despite a great number of excellent expositions of the subject, it was necessary to give a short presentation on the most basic concepts of the subject relevant to this work. This Chapter begins by exploring some of the historical aspects of the subject and the arguments in the favour for revision of the classical interpretation of the probability. This follows by a short introduction to the basic rules of the Bayesian probability theory in which concepts like prior and posterior probability and their relation to Bayes' theorem are explained in some details. It is shown how these basic concepts result into some of the most powerful tools that Bayesian interpretation can offer. More specifically, a short introduction to the principal of maximum entropy, marginalization and model selection is provided. The chapter is concluded by explaining the now celebrated entropy concentration theorem.

3.1 The Necessity of New Interpretation

There is a popular jargon that says “there are three kinds of lies: lies, damned lies and statistics”, see Twain (2000). Obviously, one hopes that scientists would never use data in such a selective manner to suit their own ends. However, the analysis of data is often the source of some frustration even in an academic context. Sivia (1996, p.1) explains this in the following manner:

> “The sense of unease, which many of us have towards the subject of statistics is largely a reflection of the inadequacies of the "cook book" approach to data analysis that we are taught as undergraduates. Rather than being offered a few clear principles, we are usually presented with a maze of tests and procedures; while most seem to be intuitively reasonable individually, their interrelations are not obvious. The apparent lack of a coherent rationale leads to considerable apprehension because we have little feeling for which test to use or, importantly, why.”

It is believed that the above quote summarizes the concerns of many scientists. As Loredo (1990) points out, this has led to a situation such that many scientists are dubious about results obtained using any but the simplest statistical methods, and some even openly assert, “If it takes statistics to show it, I don’t believe it.” However, it seems to be that a more unified and logical approach to the whole subject already exists and, in effect, is provided by the probability formulations of Bayes and Laplace.

The probability theory, as traditionally interpreted, treats all probabilities as frequencies (here the word frequency is used in the sense of the number of times an event occurs). This interpretation of probability theory is often known as *frequency interpretation.* For example to address the problem of estimating the value of a mixture parameter within this framework, one must imagine one is estimating the distribution of a random parameter within an ensemble of data sets. One then tries to determine the mean and standard deviation of this parameter within the ensemble. However, this is not the problem, which is usually encountered in practice. Indeed, typically the problem is that there is only a single data set, and one is trying to determine the value that the parameter had at the time the data was taken. But frequentists have no way to reason from an observed frequency in a finite number of trials to the value of the probability (identified as long-term frequency). This is an awkward situation, because to frequentists, probabilities by definition deal with long-term frequencies, and therefore there is no way of inferring their values from actual data.

Jacob Bernoulli clearly recognized the distinction between probability and frequency, deriving the relationship between probability of occurrence in a single trial and frequency of occurrence in a large number of independent trials now known as *Bernoulli's theorem* or the *law of the large numbers* (see for example Uspensky 1937, p.96). Bernoulli's theorem tells us that, if the probability of obtaining a particular outcome in a single trial is known to be p, the relative frequency of occurrence of that outcome in a large number of trials converges to p. This theorem is an example of reasoning from probability to frequency. However, the inverted version of this problem was also of interest to Bernoulli. That is, supposing the probability of occurrence in a single trial is unknown, what does the observation of the outcome n times in N repeated, independent trials tells us about the value of the probability? Bernoulli never solved this problem, but his interest in it further emphasizes the distinction made by him and his contemporaries between probability and frequency.

The aforementioned problem was first addressed by Rev. Thomas Bayes (1763) and later in a much more systematic way by Laplace. Laplace (1814) interpreted a probability as a reasonable degree of belief; not a frequency of occurrence. However, it appears that his work was ignored for over a century until Jeffreys (1998), in the mid 1930, rediscovered it and derived probability theory as an axiomatic theory of inference. Cox (1946) while studying the question of plausible reasoning from the perspective of logical consistency, discovered that the only rules which met his requirements of consistency were those of probability theory. That is, the rules of probability theory rather than being restricted to just frequencies, constitutes the basic calculus for logical and consistent plausible reasoning (see for example Cox 1961). This finally led Jaynes (1957a) to prove that if one represents a reasonable degree of belief as a real number, then the only consistent rules for manipulating probabilities are those given by Laplace. This gave birth to a wider interpretation of probability theory, called *Bayesian probability theory* (BPT). For Bayesians, probability theory is a kind of "quantitative epistemology", a numerical encoding of one's state of knowledge. The Bayesian interpretation is founded on three simple *desiderata,* see for example Bretthorst (1996) and Loredo (1990):

- The degrees of belief should be represented by real numbers.
- One should reason consistently, i.e., if a conclusion can be reasoned out in more than one way, every possible way must lead to the same result.
- The theory should reduce to Aristotelian logic when the truth values of the hypotheses are known. Effectively, this desideratum will ensure that the resulting theory is consistent with deductive logic in the limit that propositions are certainly true or false.

In this context, problems of the form "what is the best estimate of a parameter (say, mixture heterogeneity) one can make from the data and one's prior information?" make perfect sense. Indeed, BPT allows one to reason from observed frequency to probability. The observed frequency constitutes data, which one can use to estimate the value of the simple trial probability. Such a calculation can be done for any number of trials and it is not restricted to the infinite case.

However, as Loredo (1990) puts it, it is futile to argue over which of the interpretations are correct. The different interpretations merely reflect different choices for the types of problems the probability theory can address, and it seems possible that either interpretation could lead to a consistent mathematical theory. But though this is true, it leaves open the question of which approach is more useful or appropriate, or which approach addresses the types of problems actually encountered by scientists in the most straightforward manner. It appears to us that BPT is much more closely related to how we intuitively reason in the presence of uncertainty or incomplete knowledge and hence, at least in case of mixture of particulate materials, often ought to lead to a simpler description.

3.2 The Rules of Probability Theory

The mathematical content of the probability theory of Bernoulli, Bayes and Laplace is founded on two basic rules, the product rule and the sum rule; all other rules may be derived from these (see Jaynes 2003). If A, B, and C stand for arbitrary hypotheses, then the product rule states

$$p\,(A, B\,|C\,) = p\,(A\,|C\,)\,p\,(B\,|A, C\,) \tag{3.2.1}$$

where $p\,(A, B\,|C\,)$ is the joint probability that "A and B are true given that C is true", $p\,(A\,|C\,)$ is the probability that "A is true given C is true", and $p\,(B\,|A, C\,)$ is the probability that "B is true given that A and C are true". The vertical bar is conditional symbol, indicating what information is assumed for the assignment of a probability. In fact in BPT all probabilities are conditional.

The second rule of probability theory, the sum rule, relates the probability for "A or B". The operation "or" is indicated by a "+" inside a probability symbol. The probability for "A or B given C" is

$$p\,(A + B\,|C\,) = p\,(A\,|C\,) + p\,(B\,|C\,) - p\,(A, B\,|C\,)\,. \tag{3.2.2}$$

If the hypotheses A and B are mutually exclusive, i.e., they are independent, the probability $p\,(A, B\,|C\,)$ is zero. It is important to keep in mind that the arguments for a probability symbol are propositions, not numbers, and that operations inside the parentheses are logical operations.

Theorem 3.1 (Bayes) *Let A, B, and C stand for arbitrary hypotheses. Then*

$$p(A\,|B,C) = \frac{p(A\,|C)\,p(B\,|A,C)}{p(B\,|C)} \tag{3.2.3}$$

Proof. *In Aristotelian logic, the hypothesis "A and B" is same as "B and A", so the numerical value assigned to the probabilities for these hypotheses must be the same. This commutativity property of Aristotelian logic implies that the order in the product rule (3.2.1) may be rearranged to obtain*

$$p(B,A\,|C) = p(B\,|C)\,p(A\,|B,C) \tag{3.2.4}$$

which implies that

$$p(B\,|C)\,p(A\,|B,C) = p(A\,|C)\,p(B\,|A,C) \tag{3.2.5}$$

and hence the statement of the theorem. ■

Notice that above theorem refers to probabilities, not probability densities. Thus when considering continuous parameters, we should write

$$p(A\,|B,C)\,dA = p(A\,|C)\,dA \times \frac{p(B\,|A,C)\,dB}{p(B\,|C)\,dB} \tag{3.2.6}$$

where p's are understood to be densities. But the differentials cancel, so equation (3.2.3) is correct for densities as well as probabilities.

The above general result is always called *Bayes' Theorem*, after Rev. Thomas Bayes (1763), who derived a special case of the theorem. However, as Gillispie (1997, Ch.10) points out, it was Laplace and not Bayes who first saw the result in its generality and showed how to use it in real problems of inference.

In a sense, Bayes' theorem (BT) represents the learning process. It tells us how to adjust our plausibility assessments when our state of knowledge regarding an hypothesis changes through the acquisition of data. Indeed, let $A = H$, an hypothesis we want to assess, $B = D$, some data we have that is relevant to the hypothesis and $C = I$, some background information that describes the way in which H and D are related, and also any alternatives we may have to H. Then (3.2.3) implies that

$$p\,(H\,|D, I\,) = p\,(H\,|I\,)\,\frac{p\,(D\,|H, I\,)}{p\,(D\,|I\,)}. \tag{3.2.7}$$

The above equation tells us that our "after data" or *posterior probability* of H is obtained by multiplying our "before data" or *prior probability* $p\,(H\,|I\,)$ by the probability of the data assuming the truth of the hypothesis, $p\,(D\,|H, I\,)$, and dividing it by the probability that we would have seen the data anyway, $p\,(D\,|I\,)$, also called *global likelihood*[1]. The factor $p\,(D\,|H, I\,)$ is called the *sampling distribution* when considered as a function of the data, or the *likelihood function*, $\mathcal{L}\,(H)$, when considered as a function of the hypothesis.

Sampling distribution basically represents *the process of reasoning from some specified hypothesis to potentially observable data*, whether the link between hypothesis and data is logical or causal (see for example Jaynes 2003, p.84). However, although the determination of sampling distributions plays an important role in probability theory, in the real world such problems are an almost negligible minority. In virtually all real problems of scientific inference, one is in just the opposite situation; the data D are known but the correct hypothesis H is not. Then the problem facing the scientist is of the inverse type: "Given the data D, what is the probability that some specified hypothesis H is true?"

This is exactly sort of problems that Bernoulli could not find a satisfactory solution for and, as mentioned above, was first addressed by Bayes and Laplace. Indeed, it turns out that Bayes' theorem provides the necessary theoretical tool to handle this sort of problems.

---

[1] For reasons that will become clear later, $p\,(D\,|I\,)$ usually plays the role of an ignorable normalization constant.

3.3 Interpretation of Prior Probabilities

A closer look at the Bayes' theorem seems to indicate the presence of a new feature absent from the frequency theory. Indeed, Bayes' theorem indicates that the question: "what do you know about the hypothesis H after seeing the data D?", i.e., $p\left(H\,|D,I\,\right)$, can not have any defensible answer unless we take into account: "what did you know about H before seeing D?", which is represented by prior probability, $p\left(H\,|I\,\right)$. The concept of prior probability is completely missing from frequency Theory. Prior probability as Jeffreys (1973, p.31) put it; "is intended to express simply the probability at the start of an investigation and may have been influenced by many previous investigations.". If we interpret probability as a representation of our state of knowledge then Bayes' theorem is indeed in accordance with the aforementioned learning process. However, Bayes's theorem, despite this intuitively appealing interpretation has been a source of controversy and dispute for over two hundred years. The dispute is not on the truth of the theorem but on the concept of prior probability. Loredo (1990, p.88) blames this on the lack of a compelling rationale for some of the practices of Bernoulli, Bayes, Laplace and their contemporaries in assigning the prior probabilities.

Indeed, there were problems associated with how prior probabilities should be assigned. The probability axioms described how to manipulate probabilities, but did not specify how to assign the probabilities that were being manipulated. In most problems, it seemed clear how to assign the sampling probability, given some model for the phenomenon being studied. But finding compelling assignments of prior probabilities proved more difficult. In a certain class of problems, Bernoulli and his successors found an intuitively reasonable principle for such an assignment, which following Keynes (1921, p.41), we shall call the *principle of indifference* (PI), also known as *principle of insufficient reason.* It is a rule for assignment of probabilities to a finite, discrete set of mutually exclusive and exhaustive propositions (i.e., one proposition, and only one, must be true). The PI asserts that if the available evidence does not provide any reason for considering proposition A_1 to be more or less likely than proposition A_2, then this state of knowledge should be described by assigning the propositions equal probabilities. It follows that in a problem with N mutually exclusive and exhaustive propositions and no evidence distinguishing them, each proposition should be assigned probability $1/N$.

The domain of useful applications of PI is certainly not zero, for Laplace was led to some of the most important discoveries in celestial mechanics by using it in analysis of astronomical data (for more detail see Gillispie 1997, Ch.16). However, while PI seemed compelling for dealing with probability assignments on discrete finite sets of propositions, it was not clear how to extend it to cases where there were infinitely many propositions of interest. Such cases arise frequently in science, whenever one wants to estimate the value of a continuous parameter, θ. In this case, θ is a label for a continuous infinity of propositions about the true value of the parameter and we need to assign a prior probability (density) to all values of θ in order to use Bayes' theorem. However, Bayes' theorem has the obvious difficulty that it is not, in general, invariant under a change of parameters and there seems to be no criterion for telling us which parameterization to use. For this reason, the assignment of prior probability seem to have a disturbing subjectivity, since different investigators choosing to label hypotheses differently by using different parameters could come to different conclusions.

The statisticians of the late nineteenth and early twentieth centuries dealt with this legitimate problem by surgical removal. They drastically restricted the domain of the theory by asserting that probability had to be interpreted as relative frequency of occurrence in an ensemble or in repeated random experiments. As a by-product, the problem with arbitrariness of assignment of prior probability disappeared. Because the frequency interpretation of probability made the concept of the probability of an hypothesis illegitimate. This is because the frequency interpretation can only describe the probability of a *random variable*: a quantity that can meaningfully be considered to take on various values throughout an ensemble or a series of repeated experiments. An hypothesis, being either true or false for every element of an ensemble or every repetition of an experiment is not a random variable; its "relative frequency of occurrence" throughout the ensemble or sequence of experiments is either 0 or 1.

Assessing hypotheses was one of the principle aims of probability theory. Denied the use of Bayes's theorem for this task, frequency theory had to develop ways to accomplish it without actually calculating probabilities of hypotheses. As Loredo (1990, p.89) points out, the frequentist solution to this problem was the creation of the discipline of *statistics*. In statistical analysis one basically constructs some function of observable random variables that is somehow related to what one wishes to measure; such a function is called a *statistic*. Familiar statistics include the sample mean and variance, the χ^2 statistic and F statistic. Since a statistic is a function of random variables, its probability distribution, assuming the truth of the hypothesis of interest, can be calculated. A hypothesis is assessed by comparing the observed value of the statistic with the long-run frequency distribution of the values of the statistic in hypothetical repetitions of the experiment. However, for complicated problems, there is seldom a compelling "natural" choice for a statistic. To provide a rational for statistic selection, many principles and criteria have been added to classical theory, including unbiasedness, efficiency, consistency, coherence, the conditionality principle, sufficiency and likelihood principle.

Once a statistic is selected, it must be decided how its frequency distribution will be used to assess a hypothesis. To replace the Bayesian notion of the probability of a hypothesis, other real number measures of the plausibility of an hypothesis are introduced, including confidence regions, significance levels, type I and II error probabilities, test size and power and so on. These all require the consideration of hypothetical data for their definitions. These hypothetical data sets are predicted by the hypothesis but not seen. This is as Loredo (1990, p.90) puts it, as if a juror tried to decide guilt or innocence by taking into consideration a mass of evidence that might possibly have been presented at the trial but which was not.

The frequency interpretation was introduced to eliminate apparent arbitrariness and subjectivity of Bayesian interpretation. Yet a large degree of arbitrariness must enter the frequency theory to allow it to address the problems Laplace could address directly. For a more comprehensive discussion and illustrative examples on this topic we refer to Jaynes (2003, Ch.5) and also Good (1983, Ch.6).

3.4 Assignment of Prior Probabilities

In the mean time, developments in the seemingly unrelated fields outside the traditional domain of probability theory changed the odds in favour of the Bayesian interpretation. The introduction of the theory of *transformation groups* during the second half of the 19th century and *information theory* during the first half of the 20th century made the necessary tools available in order to overcome the problems of assignment of prior probabilities.

3.4.1 Least Informative Probabilities

In probability theory a problem is considered to be *well-posed* if it contains enough information to allow unique, unambiguous probability assignments. The simplest kind of information we can have about some proposition A_1 is a specification of alternatives to it. That is, we can only be uncertain of A_1 if there are alternatives $A_2, A_3, \ldots$ that may be true instead of A_1; and the nature of the alternatives will have a bearing on the plausibility of A_1. Probability assignments that make use of only this minimal amount of information are referred to as *least informative probabilities* (LIP). The following example by Loredo (1990, p.99) sheds some light on the concept.

Consider a problem where probabilities must be assigned to two propositions, A_1 and A_2. Suppose we know from the very nature of the alternatives that they form an exclusive and exhaustive set (one of them, and only one, must be true), but that is all we know. We might indicate this symbolically by writing our conditioning information as $B = A_1 + A_2$. Since the propositions are exclusive, $p\left(A_1, A_2 \middle| B\right) = 0$, so the sum rule (3.2.2) implies that $p\left(A_2 \middle| B\right) = 1 - p\left(A_1 \middle| B\right)$. Now imagine someone else addressing this problem, but labeling the propositions differently, writing $A'_1 = A_2$ and $A'_2 = A_1$. Obviously, $p\left(A'_1 \middle| B\right) = p\left(A_2 \middle| B\right)$, and $p\left(A'_2 \middle| B\right) = p\left(A_1 \middle| B\right)$. But now note that since B is indifferent to A_1 and A_2, the state of knowledge of this second person regarding A'_1 and A'_2, including their labeling is the same as that in the original problem. Since equivalent states of knowledge must be represented by equivalent probability assignments, then $p\left(A'_1 \middle| B\right) = p\left(A_1 \middle| B\right)$. But this means that $p\left(A_1 \middle| B\right) = p\left(A_2 \middle| B\right)$, which through the sum rule implies $p\left(A_1 \middle| B\right) = p\left(A_2 \middle| B\right) = 1/2$. In the similar manner, this line of thought can be generalized to a set of N mutually exclusive and exhaustive propositions A_i ($i = 1$ to N), leading to the LIP assignments $p\left(A_i \middle| B\right) = 1/N$. This is just the principle of indifference mentioned earlier, now seen to be a consequence of consistency when all the information we have is an enumeration of a mutually exclusive and exhaustive set of possibilities, with no information leading us to prefer some possibilities over the others. The philosophy of above approach can be traced back to David Hume. Indeed, this idea was formulated by Hume in the following passage (see Hume 1740, p.86):

> "As chance is nothing real in itself, and, properly speaking, is merely the negation of a cause, its influence on the mind is contrary to that of causation; and it is essential to it, to leave the imagination perfectly indifferent, either to consider the existence or non-existence of that object, which is regarded as contingent. A cause traces the way to our thought, and in a manner forces us to survey such certain objects, in such certain relations. Chance can only destroy this determination of the thought, and leave the mind in its native situation of indifference; in which, upon the absence of a cause, it is instantly re-instated. Since therefore an entire indifference is essential to chance, no one chance can possibly be superior to another, otherwise than as it is composed of a superior number of equal chances. For if we affirm that one chance can, after any other manner, be superior to another, we must at the same time affirm, that there is something, which gives it the superiority, and determines

> the event rather to that side than the other: That is, in other words, we must allow of a cause, and destroy the supposition of chance; which we had before established. A perfect and total indifference is essential to chance, and one total indifference can never in itself be either superior or inferior to another."

When the set of possibilities is infinite, as when it is desirable to assign probabilities to the possible values of continuous parameters, the analysis becomes more complicated. This is because it may not be obvious how to transform the original problem to an equivalent one that will help us determine the probability assignment. Indeed, in the finite discrete case, the only transformation that preserves the identity of the possibilities is permutation, leading to PI. But in the continuous case, there is an infinite number of possible reparameterizations.

The key to resolving this dilemma is to realize that specifying the possibilities not only provides labels for them, but it tells us also about their nature. As Jaynes (1968, p.239) put it, "if we approach a problem with the charitable presumption that it has a definite solution, then every circumstance left unspecified in the statement of the problem defines an invariance property (i.e., a transformation to an equivalent problem) which that solution must have.". In this sense, in problems with continuous parameters, transformations that lead to equivalent problems that can help one assign a LIP can often be identified by the nature of the parameters themselves. Information unspecified in the problem statement can be as important for this identification as the specified information itself, for problems that differ with respect to unspecified details are equivalent.

It is easy to show that in any specific case, mathematically, the collection of all the transformations that transform the problem to an equivalent one poses a *group* structure[2].

---

[2] A nonempty set G with a binary operation $(\cdot)$ on G is called a group if the following three axioms hold (see e.g. Bhattacharya *et al.* 1994, p.62):

1. Associativity: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in G$.
2. Identity: there exists $e \in G$ such that $e \cdot a = a$ for all $a \in G$.
3. Inverse: for every $a \in G$ there exists $a' \in G$ such that $a' \cdot a = e$.

In the discrete case, this group is the *permutation group* S_n where n is the number of propositions (Bhattacharya *et al.* 1994, p.84). In the continuous case, in general, the group of transformations is a *Lie group* (for more detail see Onishchiks 1993, Warner 1983 or Weyl 1961). Usually, the group of transformations can be related to the symmetries of the system under study. Symmetries are often related to invariant properties of a given system. It turns out that there is a profound relation between the symmetries of a system and its behaviour. This is currently a very active area of research and there are reasons to believe that it might be fruitful to study particulate systems from this point of view (for some illustrative examples on the concepts of symmetry and transformation groups, see Ghaderi and Naqvi 1995).

3.4.2 Informative Probabilities

More than often, beside the specification of possibilities, I_1, we may have some additional information I_2 that should lead us to probability assignments different from LIP. Rather than $p\,(A_i|\,I_1)$, we seek $p\,(A_i|\,I_1, I_2)$, an *informative probability* (IP) assignment.

One way to find $p\,(A_i|\,I_1, I_2)$ is to use Bayes' theorem, i.e., Eqn. (3.2.7), to update our assignments for each of the A_i one at a time. To do this, as Loredo (1990, p.101) points out, the additional information $D \equiv I_2$ must be able to play the role of data, that is, it must be meaningful to consider for each A_i the "sampling probability" $p\,(D|\,I_1, A_i)$ in

$$p\,(A_i|\,I_1, D) = p\,(A_i\,|I_1\,)\,\frac{p\,(D\,|I_1, A_i\,)}{p\,(D\,|I_1\,)}. \tag{3.4.1}$$

Specifically, D has to be a possible consequence of one or more of the A_i considered individually, since each application of Bayes' theorem will require us to assume that one of the A_i is true to calculate the likelihood of the additional information. However, this is not the only kind of information we may have about the various possibilities. Our information may refer directly to the possibilities themselves, rather than to their consequences. In these cases the Bayes' theorem cannot be used. Yet such information is clearly relevant for assessing the plausibility of the propositions. Therefore, there was a need for finding rules that could allow one to use information of this kind to make probability assignments.

The clues to a possible solution of this problem revealed itself after development of information theory by Shannon (1948). He introduced a measure for uncertainty which he named *entropy*[3]. Later, Jaynes (1957b; 1957c) showed that the concept of entropy is useful in converting certain types of information called *testable information* to a probability assignment. Following Jaynes (1968, p.230) a testable information is defined in the following manner:

Definition 3.1 *A piece of information I concerning a parameter θ is called testable if, given any proposed prior probability assignment $f(\theta)\,d\theta$, there is a procedure which will determine unambiguously whether $f(\theta)$ does or does not agree with the information I.*

Jaynes demonstrated that if I is testable, then in accordance with Bayesian desiderata, one should select from among all the possible normalized distributions satisfying the constraints imposed by I, the one with maximum entropy. The entropy of a finite discrete distribution over mutually exclusive and exhaustive alternatives is defined by

$$H = -\sum_{i=1}^{N} p_i \ln p_i, \tag{3.4.2}$$

and that of a continuous distribution is defined analogously by

$$H = -\int p(\theta) \ln\left[\frac{p(\theta)}{m(\theta)}\right] d\theta, \tag{3.4.3}$$

with $m(\theta)$ the LIP assignment for the parameter θ. This rule is now called the principle of *maximum entropy* (MaxEnt). In Section 4.3, based on combinatorial arguments, both

---

[3] In science, the term *entropy* is generally interpreted in three distinct, but semi-related, ways, i.e. from macroscopic viewpoint (classical thermodynamics), a microscopic viewpoint (statistical thermodynamics), and an information viewpoint (information theory). The thermodynamic interpretations, generally, differ substantially from the information theory interpretation and are only related in namesake, although there is not complete agreement on this issue. It was Rudolf Clausius (1865, p. 355) who first introduced the phrase entropy (or equivalence-value as he called it in 1854) and gave a consistent mathematical formulation of it within the classical thermodynamics theory.

above entropy expressions for the mixture of particulate materials are developed.

Shore and Johnson (1980) have shown that MaxEnt is an uniquely correct method for inductive inference when the constraints on pdf are given in the form of expected values. In fact, they showed that maximizing any functional but entropy will lead to inconsistency unless that functional and entropy have identical maxima. In other words, given information in the form of constraints on the expected values, there is only one pdf satisfying the constraints that can be chosen by a procedure that satisfies the consistency axioms; this unique pdf can be obtained by MaxEnt. These axioms maybe phrased as follows.

I. *Uniqueness*: The result should be unique.

II. *Invariance*: The choice of coordinate system should not matter.

III. *System Independence*: It should not matter whether one accounts for independent information about independent systems separately in terms of different densities or together in terms of a joint density.

IV. *Subset Independence*: It should not matter whether one treats an independent subset of system states in terms of a separate conditional density or in terms of the full system density.

For further readings on MaxEnt the reader is referred to Jaynes (2003, Ch.11) and Gregory (2005, Ch. 8).

3.5 Marginalization

From the above discussions it is clear that any problem one wishes to address with BPT must be well-posed, in the sense that enough information must be provided to allow unambiguous assignment of all probabilities required in a calculation. As a bare minimum, this means that an exhaustive set of possibilities must be specified at the start of every problem. This set is called *sample space* if it refers to possible outcomes of an experiment, or *hypothesis space* if it specifies possible hypotheses one wishes to assess. For example in an estimation problem the hypothesis space is simply the set of possible values of the parameter, say $\mathcal{H} = \{\theta_i\}$ and the sample space, say $\mathcal{S} = \{s_i\}$, consists of set of possible data. The hypothesis space, the sample space, or both can be either discrete or continuous.

Let the unknown true value of a parameter be Θ. Bayes' theorem can be used to address an estimation problem by calculating the probability that each of the possible parameter values $\mathcal{H}$ is the true value. To this end, let, in Eqn. (3.2.7), D represent a proposition asserting the values of the data actually observed and H be the proposition $\Theta = \theta$, asserting that one of the possible parameter values, θ is the true value. Then the Bayes' theorem reads,

$$p\left(\theta\,|D, I\right) = p\left(\theta\,|I\right)\frac{p\left(D\,|\theta, I\right)}{p\left(D\,|I\right)}. \tag{3.5.1}$$

The prior $p\left(\theta\,|I\right)$ and the likelihood $p\left(D\,|\theta, I\right)$ are both *direct probabilities*; i.e., their values are assigned directly, rather than derived from other probabilities using the product and sum rules. The direct probabilities can be assigned using the methods described previously. The term in the denominator is independent of θ and given the prior and the likelihood, its value can be calculated using the rules of probability theory as follows.

Recall that the assumption is that the model is true for some value of its parameters. Thus the logical proposition $(\theta_1 + \theta_2 + \cdots)$ is true and so has a probability of one for the given I. Then from the product rule (3.2.1) follows,

$$\begin{aligned} p\left(D, \theta_1 + \theta_2 + \cdots |I\right) &= p\left(D\,|I\right) p\left(\theta_1 + \theta_2 + \cdots |I\right) \\ &= p\left(D\,|I\right). \end{aligned} \tag{3.5.2}$$

But by expanding the logical product on the left and again using (3.2.1), we also have

$$\begin{aligned} p\,(D, \theta_1 + \theta_2 + \cdots \,|I\,) &= \sum_i p\,(D, \theta_i\,|I\,) \\ &= \sum_i p\,(\theta_i\,|I\,)\,p\,(D\,|\theta_i, I\,)\,. \end{aligned} \tag{3.5.3}$$

Eqns. (3.5.2) and (3.5.3) together imply that

$$p\,(D\,|I\,) = \sum_i p\,(\theta_i\,|I\,)\,p\,(D\,|\theta_i, I\,)\,, \tag{3.5.4}$$

which can easily extended to the case when we are dealing with continuous parameters to yield

$$p\,(D\,|I\,) = \int p\,(\theta\,|I\,)\,p\,(D\,|\theta, I\,)\,d\theta. \tag{3.5.5}$$

This expresses $p\,(D\,|I\,)$ in terms of the prior and the likelihood. Thus in an estimation problem, the denominator of Bayes' theorem is just the normalization constant for the posterior. In these cases the Bayes's theorem takes the simple form

$$p\,(\theta\,|D, I\,) = N p\,(\theta\,|I\,)\,p\,(D\,|\theta, I\,)\,, \tag{3.5.6}$$

where N denotes the normalization constant.

The trick just used to calculate $p\,(D\,|I\,)$ arises frequently in BPT. It is usually referred to as *marginalization.* Marginalization is of great practical and theoretical importance, because it can often be used to significantly reduce the dimensionality of a problem by eliminating the so-called *nuisance parameters,* i.e., quantities which necessarily enter the analysis but are of no intrinsic interest. For example if a problem has two parameters, θ and ϕ, but we are interested only in θ, then we can calculate $p\,(\theta\,|D, I\,)$ from the full posterior $p\,(\theta, \phi\,|D, I\,)$ by using the trick we used to calculate $p\,(D\,|I\,)$. Hence

$$\begin{aligned} p\,(\theta\,|D, I\,) &= \int p\,(\theta, \phi\,|D, I\,)\, d\phi \\ &= \tfrac{1}{p(D|I\,)} \int p\,(\phi\,|I\,)\, p\,(\theta\,|\phi, I\,)\, p\,(D\,|\theta, \phi, I\,)\, d\phi. \end{aligned} \tag{3.5.7}$$

The above integral can sometimes be evaluated analytically which can greatly reduce the computational aspects of the problem especially when many parameters are involved. A good example of this can be found in Gregory and Loredo (1992).

3.6 Model Comparison

If a model is inadequate, then some alternative model must be better and so BPT assesses a model by comparing it to one or more alternatives. This is done by assuming that some member of a set of competing models is true. Then one uses the Bayes' theorem to calculate the probability of each model given the observed data.

To this end, let I denote that one of a set of models is true and the information about the model number k be I_k, where $k = 1$ to m. Then

$$I = I_1 + I_2 + \cdots + I_m, \tag{3.6.1}$$

where as usual "+" denotes the logical operation "or". Moreover let D stand for the data and H_k stand for the hypothesis, "model number k is true". Then Bayes' theorem reads,

$$p\,(H_k\,|D, I\,) = p\,(H_k\,|I\,)\,\frac{p\,(D\,|H_k, I\,)}{p\,(D\,|I\,)}. \tag{3.6.2}$$

Furthermore, note that since H_k asserts the truth of model number k, only information I_k in I is relevant and hence

$$H_k I = H_k\,(I_1 + I_2 + \cdots + I_m) = I_k, \tag{3.6.3}$$

where H_kI denotes the logical product of H_k and I. This implies that

$$p(D|H_k, I) = p(D|I_k). \tag{3.6.4}$$

Now, the ratios of the probabilities of the models are called *odds.* The odds in favour of model k over model j is defined as,

$$O_{kj} = \frac{p(H_k|D, I)}{p(H_j|D, I)}, \tag{3.6.5}$$

which together with (3.6.2)-(3.6.4) imply that

$$O_{kj} = \left[\frac{p(H_k|I)}{p(H_j|I)}\right]\left[\frac{p(D|I_k)}{p(D|I_j)}\right]. \tag{3.6.6}$$

In this work, it is more convenient to take the logarithm of the odds because of the fact that one can then add up terms. In the spirit of Jaynes (2003, p.90), this new function is called as *evidence,*

$$\psi_{kj} = 10\log_{10} O_{kj}, \tag{3.6.7}$$

where the base 10 logarithms is used. By using the base 10 and putting the factor 10 in front, one measures the evidence in *decibels* (dB). If $\psi_{kj} > 0$ then there are good reasons to prefer model number k over model number j. The opposite would be true if $\psi_{kj} < 0$. Theoretically, the model k is as good as model j if $\psi_{kj} = 0$.

3.7 Entropy Concentration Theorem

In many statistical problems, one usually has information which places some kind of restriction on a probability distribution without completely determining it. According to MaxEnt, if two distributions satisfy the information at hand, one should choose the one with greater entropy. In other words, the conversion of prior information into definite prior probability assignment becomes a variational problem in which the prior information plays the role of constraint. This means that the notion of entropy defines a kind of measure on the space of probability distributions such that those of high entropy are in some sense favoured over others.

However, at this point a fair question is that how far distributions of lower entropy are from the one determined by MaxEnt? And more importantly, how can one by comparing observed entropy and the MaxEnt entropy accept or reject a hypothesis (model)? In the later case, using entropy analysis for hypothesis testing, frequency of different events are known experimentally. A successful hypothesis about the systematic influences is than one for which the experimentally observed entropy is sufficiently close to the MaxEnt entropy. In this context, the notion of distance or "being sufficiently close" was first properly explored by Jaynes which led to the result which is known as *entropy concentration theorem*, see Jaynes (1982; 1983).

Theorem 3.2 (Entropy concentration) *Let* $\Delta H = H_{\max} - H$ *denote the difference between the MaxEnt entropy and the observed entropy. Moreover, let* N *be the total number of observations,* n *total number of outcomes and* m *total number of constraints. Then for large* N, $2N\Delta H$ *is distributed as chi-square with* $\nu = n - m - 1$ *degrees of freedom, independently of the nature of constraints.*

For example, the intervals

$$\left[H_{\max} - \frac{\chi^2_\nu\,(0.95)}{2N}, H_{\max}\right] \tag{3.7.1}$$

and

$$\left[H_{\max} - \frac{\chi^2_\nu\,(0.99)}{2N}, H_{\max}\right] \tag{3.7.2}$$

are known as 95% and 99% *entropy intervals.* If one takes any probability distribution satisfying the same constraints as the maximum entropy distribution, there is a 95% chance that its entropy will be greater than $H_{\max} - \frac{\chi^2_\nu(0.95)}{2N}$ and a 99% chance that its entropy will be greater than $H_{\max} - \frac{\chi^2_\nu(0.99)}{2N}$. Consequently, for large values of N, entropies of most of the probability distributions satisfying a given set of constrains will be concentrated near the maximum entropy value. In fact, it is possible to show that the length of the entropy interval (see e.g. Kapur and Kesavan 1992, p.51):

- decreases fast as N increases;
- increases with confidence level;
- increases with n and decreases with m.

Since for large N, most distributions have entropies very near the maximum entropy, most distributions satisfying the given constraints are very close to MaxEnt distribution. As a result, MaxEnt distribution is the best choice for a unique distribution. Moreover, one can also conclude that if the entropy of a distribution satisfying the given constraints lies outside, say 99% entropy interval, it indicates the possibility of an additional constraint, which has to be taken into account.

Chapter 4

Sampling Theory

> Quality estimation is a chain and sampling is its weakest link.
>
> P.M. Gy, Sampling for Analytical Purposes (1998)

The characterization of mixtures is closely related to some of the topics in sampling theory. Therefore, sampling is an integral part of the any theory on the mixing of particulate materials. In this chapter some of the important issues relevant to the present work are addressed. The aim is to develop a consistent model for mixture heterogeneity. In this context, the meaning of a representative sample in relation to evaluation of mixture quality is discussed. Further, a mathematical model which quantifies the most important properties of a representative sample, i.e., Accuracy and Reproducibility, is developed. In this relation, it is shown that Reproducibility can be used as a measure for mixedness and mathematically it can be modeled by the entropy of the sample distribution. Further, from the sampling point of view, the mixing systems are divided into two categories of open and closed systems. In each case a mathematical model is developed. It is demonstrated that the same mathematical formalism can be adopted to model both type of systems and the only difference is in how the constraints are assigned. A set-theoretical approach to the concept of mixture heterogeneity is also established. In this relation, it is demonstrated how mixture heterogeneity can be quantified. This chapter is concluded by determining the relation between the pdf of different sample spaces of different orders. These results are summarized in two theorems as, fundamental theorem of sample spaces and fundamental theorem of ensembles.

4.1 Introduction

Usually the analysis of the whole mixture, due to economic reasons or destructiveness of the process is not possible. In these cases, the mass of the mixture under consideration is first reduced to tiny amount before taken for analysis. The process of mass reduction is known as *sampling*. Obviously, this mass reduction should be of such nature that retains the important features of the original lot. Thus, an important aim of any *sampling theory* should be to describe methods of reducing the mass of a lot without significantly changing its other important properties. In order for any sampling theory to achieve this objective, it needs to identify the sources which contribute to divergence of sample properties from the properties of the original lot. In literature, a sample which has approximately similar properties as the original lot is known as *representative sample.* However, more often most authors only describe the method and the locations of samples, the sample size and the number of samples (see for example Fan *et al.* (1970) and references therein). But in the absence of compelling criteria for representativeness, it is difficult to see the usefulness of such information. Therefore, producing a set of compelling criteria should also be one of the chief objectives of any successful sampling theory.

It appears that the first documented attempt to develop a theory for sampling of particulate materials was by D. Brunton (1895). However, the most noteworthy of all the sampling theories in the past hundred years, is the one developed by P. Gy. It appears that he has been very successful in handling the challenges one faces in sampling. Gy (1998, p.30) defines a sample to be representative when it is taken by a selection method that is both *accurate* and *reproducible.* The Accuracy is defined as the absence of bias or systematic error and reproducibility is defined as a low dispersion of the sample values about their mean. According to Gy, accuracy is achieved whenever the selection method is *correct.* That is, all the constituent elements of the lot have an equal probability of being selected. He also demonstrates how this can be determined by studying the sampling procedure. Similarly, reproducibility can be checked as the condition in which the sampling variance is minimal, i.e., below certain prescribed value (see Gy 1998, p.32). The reproducibility is improved as the mixture quality improves. The best Reproducibility is achieved whenever the mixture components are randomly distributed through the mixture. Thus the value of the variance would vary depending on the quality of the mixture and need not necessarily be small. In this sense, in the case of particulate mixtures, the reproducibility is, in general, less important than accuracy. This also means that in general a representative sample, from particulate mixing point of view, is a sample which is collected by a correct selection method. However, if the samples show good reproducibility, this would imply that mixture ingredients are evenly distributed. Thus, reproducibility could in a sense be used as a measure for mixedness.

Although, in general, Gy's theory gives a compelling qualitative explanation of sampling process, to some extent, it fails on the quantitative part. Ironically, the source of the problem is not due to lack of understanding of fundamentals of the subject but is caused by the statistical concepts used in the theory. Gy, as most other researchers in this field, applies the frequency interpretation of the probability theory (see also Sommer 1986). But, as explained previously, from a mathematical point of view, this has led to that, the easy and intuitive concepts become more complicated and similarly the more complicated problems become hopelessly beyond reach. Nevertheless, once one understands Gy's sampling theory, description of a suitable mathematical framework would be within reach. In this chapter, affords would be concentrated on demonstrating how this framework can be constructed.

4.2 Closed vs. Open System

In order to check the composition and distribution of the components in a mixture, one needs to define the smallest scale at which the desired component is to be observed. This scale is known as *scale of scrutiny.* The scale of scrutiny defines the limit of resolution of information. This limit is imposed either by the application or by measurement instruments and is specified by the volume of the sample. For example, producers of a drug in tablet form need to make sure that composition of their tablets are within the desired limits otherwise it can have a catastrophic result. Therefore, in these cases, the scale of scrutiny is chosen to be equal to the volume of a typical tablet (see e.g. Fan *et al.* 1970). Any variations within smaller volumes are ignored, while the variations between samples give an indication on how well different components in the mixture are distributed. Thus, the *structure of a mixture* can be defined as the relation between the compositions of different samples in which their volume is specified by the scale of scrutiny.

From an analysts point of view, mixing systems can be divided into two classes, *closed* and *open* systems. In a closed system, depending on the scale of scrutiny, the whole batch of the mixture is divided into a number of samples, which are further analyzed. Whereas, in an open system, only a fraction of the samples are analyzed[1]. Thus, in reality, sampling, as defined previously, is conducted only in the case of open systems and in the case of closed systems, *splitting* is a more suitable description. In this work, sampling will be used also to mean splitting. However, at each case it would be clear from the context if one is working with a closed or open system.

In a closed system, properties like amount of different components in the mixture can be determined exactly, whereas in an open system, same properties can only be estimated (unless there is some cogent information, which allows determining these properties in some other way). In other words, in general, models for open systems are more uncertain than for corresponding closed systems.

4.2.1 Closed System

A mixture of particulate materials could in general consists of many components. However, binary mixtures are also frequently encountered. For example, in the pharmaceutical industry usually only one active ingredient (key component) and several diluents (lactose,

---

[1] It should be emphasised that our definition of closeness and openness of a system depends solely on the information about the system. This does not necessarily coincide with the common use of these words in physics and chemistry.

starch, glucose, etc.) are in a dosage; therefore, the homogeneity of the active ingredient is the main concern (see e.g. Kristensen 1973). Nevertheless, from the sampling point of view, there is no conceptual differences between binary and multicomponent mixtures. However, mathematically, the difference is in the dimension of the *sample space*; i.e., the collection of all possible samples of given size and shape. Indeed, following Wang *et al.* (1978), assume that the mixture consists of $(k+1)$-components. Then for a given sample size and shape, one can define a k- vector in which each of its component equals the number of particles of a given mixture component[2]. Now, suppose that one divides the whole mixture into n samples. Then for each sample a k- vector $\mathbf{r}_i'$ can be obtained, which can be used to construct the so-called *sample matrix*,

$$\mathbf{X} = \begin{pmatrix} \mathbf{r}_1' \\ \vdots \\ \mathbf{r}_n' \end{pmatrix} = \begin{pmatrix} r_{11} & \cdots & r_{1k} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nk} \end{pmatrix}, \tag{4.2.1}$$

where r_{ij} stands for number of particles of type j in sample i and $\mathbf{r}_i'$ is the transpose of the column vector $\mathbf{r}_i$. In other words, each row of the sample matrix contains the information on the number of particles of each type in a given sample. Thus, in the case of binary mixtures, the above matrix consists of only one column vector.

The above procedure is also valid for mixtures produced by continuous mixers. Indeed, according to Gy (1998, Ch.6), the best way to sample a mixture flowing out of a continuous mixer is by directing the mixture onto a conveyer belt with a constant velocity. Then, each sample can be taken as all the material occupying a fixed length in the moving direction of the conveyer belt. In this case, one can construct a similar sample matrix as

---

[2] It is assumed that the total number of particles in each sample is known and is equal to r_i. Now, let r_{ij} denote the number of particles of type j in sample i. Since

$$\sum_{j=1}^{k+1} r_{ij} = r_i$$

then only k of $(k+1)$ components need to be determined.

in (4.2.1), with the difference that in this case the distance between samples is measured by time contra spatial distance in the case of a batch mixer. The closedness of the system is guaranteed if all the material flowing out of the mixer during its operation is equal the combined content of the analyzed samples.

4.2.2 Open System

As Too *et al.* (1980) point out, in many mixing problems, the true proportions of components in a mixture are known. However, it is more than often not possible to analyze the whole mixture. In these cases, one takes so called *spot samples*, which their combined content only constitutes a fraction of the whole mixture. In other words, the sampling procedure is exactly same as in the case of closed systems with the only difference that just a fraction of the mixture is analyzed. This is the true sampling in the sense which was defined previously. However, in the case of open systems, the acquired information about the structure of the mixture is less reliable compared with the closed systems. Moreover, in the case of an open system the sample matrix has fewer rows than the corresponding closed system.

4.3 Models for Binary Closed Systems

In order to keep the matters simple, for the rest of this section, it is assumed that the system under study is a binary closed system. After establishing a model for this system, it will be demonstrated how the results can be extended to other cases.

4.3.1 Batch Systems

Suppose, for the sake of the argument, that the mixture and the mixing equipment, to start with are in a definite state and it is always possible to rearrange the particles and the machine parts in order to achieve this initial state on demand. Now, begin the mixing operation and let it last for a definite amount of time, after which the mixer is brought to rest. The mixture is then divided into n samples, in which the *sample volume fraction* is denoted by v_i such that

$$\sum_{i=1}^{n} v_i = 1, \tag{4.3.1}$$

where the sample volume fraction is defined as the sample volume divided by the total volume of the mixture. Furthermore, assume that there are totally N tracer particles (key component particles) in the mixture. After analyzing the samples, suppose that we find r_1 tracer particles in the first sample, r_2 in the second sample and so on. Thus

$$N = \sum_{i=1}^{n} r_i, \tag{4.3.2}$$

which is usually much larger than the number of samples. The distribution $\{r_i\}$ gives rise to a probability distribution $\{p_i\}$, in which p_i denotes the probability of finding a tracer particle in sample i. Indeed, the natural candidates are,

$$p_i = r_i/N. \tag{4.3.3}$$

If the experiment stops here, the above candidate is the best choice. However, it might happen that one simply would like to explore the possibility of other candidates. For this reason, the samples are combined and the mixture and the mixing equipment are brought to the same initial state before the first experiment. The mixture is then mixed and sampled according to the same procedure as the previous experiment. Each time one conducts this experiment, a potential candidate for the distribution of the tracer particles can be found. After many trails, some distributions will be found to come up more often than others. Accordingly, the one that occurs most frequently would be the best candidate representing the observers state of knowledge. Now, the question is what the expected frequency, say F, of each candidate distribution is. If one could determine this from the information about the system, then it would be possible to pick the one with the highest frequency.

To this end, note that the probability of a tracer particle being found in the ith sample is v_i. Now, since in reality the volume of the sample is far greater than the size of each particle, the probability of finding more than one tracer particle in sample i is independent of the number of tracer particles already in the sample and hence, the probability of finding the distribution $\{r_i\}$ is

$$v_1^{r_1} \cdot v_2^{r_2} \cdots v_n^{r_n} = \prod_{i=1}^{n} v_i^{r_i}. \tag{4.3.4}$$

Obviously, $\{p_i\}$ is not affected by exchanging the tracer particles among samples as long as the distribution $\{r_i\}$ remains the same. Thus, the expected frequency F with which $\{p_i\}$ will arise is given by

$$F\left(\{p_i\}\right) = (\text{number of ways of obtaining } \{r_i\}) \times \prod_{i=1}^{n} v_i^{r_i}. \tag{4.3.5}$$

The number of ways to distribute N particles among n samples such that r_i particles are in the sample i, for $i = 1, \ldots, n$, is given by the *multinomial coefficients* (see Dougherty 1990, p.45):

$$\begin{pmatrix} N \\ r_1, \ldots, r_n \end{pmatrix} = \frac{N!}{r_1! \cdots r_n!}, \tag{4.3.6}$$

and hence

$$F\left(\{p_i\}\right) = \begin{pmatrix} N \\ r_1, \ldots, r_n \end{pmatrix} \prod_{i=1}^{n} v_i^{r_i} = \frac{N!}{(Np_1)! \cdots (Np_n)!} \prod_{i=1}^{n} v_i^{Np_i}. \tag{4.3.7}$$

Now, notice that the functional F achieves its maximum at the same $\{p_i\}$ as $\ln(F)$. Moreover, recall that for large x, the Stirling's formula yields the following approximation (see Arfken 2001, p.650):

$$\ln(x!) \approx x \ln x + \ln \sqrt{2\pi x} - x. \tag{4.3.8}$$

In fact, it is easy to show that Stirling's approximation is quite good for even numbers as small as $x = 10$. Hence,

$$\ln\left[F\left(\{p_i\}\right)\right] \approx -N\sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{v_i}\right) - (n-1)\ln\sqrt{2\pi N} - \sum_{i=1}^{n}\ln\sqrt{p_i}. \tag{4.3.9}$$

Since any other monotonically increasing function of $F\left(\{p_i\}\right)$ also achieves its maximum at the same point, in particular $N^{-1}\ln\left[F\left(\{p_i\}\right)\right]$, then from Eqn. (4.3.9) follows that

$$D\left(\mathbf{p} : \mathbf{v}\right) = \lim_{N\to\infty}\frac{1}{N}\ln\left[F\left(\{p_i\}\right)\right] = -\sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{v_i}\right). \tag{4.3.10}$$

In literature, $D\left(\mathbf{p} : \mathbf{v}\right)$ is known as *cross-entropy* or Kullback-Leibler entropy, due to the work by Kullback and Leibler (1951)[3]. Thus, Eqn. (4.3.10) implies that the maximum value of $F\left(\{p_i\}\right)$ is attained for the set $\{p_i\}$ which maximizes $D\left(\mathbf{p} : \mathbf{v}\right)$.

As mentioned previously, one of the conditions that is necessary in order for a sample to be representative, is that the selection method has to be Correct. As Gy (1998, p.31) points out, the Correctness of a sampler is the result of its design, construction, installation, usage and maintenance. Nonetheless, when sampling is done Correctly, all the constituent elements of the lot have an equal probability of being selected to each sample. In the present case under study, this means that the volume of samples have to be equal, i.e.

$$v_1 = v_2 = \cdots = v_n = \frac{1}{n}, \tag{4.3.11}$$

which implies that

$$D\left(\mathbf{p} : \left\{\frac{1}{n}\right\}\right) = -\sum_{i=1}^{n} p_i \ln p_i - \ln n. \tag{4.3.12}$$

Since n is constant then this means that, in the case of Correct sampling, the maximum value of $F\left(\{p_i\}\right)$ is attained for the set $\{p_i\}$ which maximizes its entropy $H\left(\{p_i\}\right)$ (see Chapter 3).

---

[3] It seems that this measure was first appeared in Gibbs (1902, Ch.11)

The *Lagrangian multipliers* method (Arfken 2001, p.1039), can be used to determine the vector $\mathbf{p}$ which maximizes $H(\mathbf{p})$, subject to the obvious constraint,

$$\sum_{i=1}^{n} p_i = 1. \tag{4.3.13}$$

In this case, the Lagrangian is

$$L = -\sum_{i=1}^{n} p_i \ln p_i - \lambda \left(\sum_{i=1}^{n} p_i - 1 \right), \tag{4.3.14}$$

in which λ denotes the Lagrangian multiplier. Differentiating the above equation with respect to $p_1, \ldots, p_n$, yield

$$-\left(1 + \ln p_i\right) - \lambda = 0, \text{ for } i = 1, \ldots, n, \tag{4.3.15}$$

which, subject to the constraint (4.3.13) implies that

$$p_1 = \cdots = p_n = \frac{1}{n}\cdot \tag{4.3.16}$$

Moreover, the *second-derivative* matrix or *Hessian* matrix of the functional $H(\mathbf{p})$ is (see Strang 1988, p.327),

$$\begin{pmatrix} -1/p_1 & 0 & \cdots & 0 \\ 0 & -1/p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1/p_n \end{pmatrix}, \tag{4.3.17}$$

which is always negative definite. Thus, for the uniform probability distribution given by (4.3.16), $H(\mathbf{p})$ has a local maximum. The strict negative definiteness of the Hessian matrix also indicates that $H(\mathbf{p})$ is a *concave* function, i.e., $H(\mathbf{p})$ cannot have more than one maximum and if it has a local maximum, then that is also the global maximum (Kapur 1989, p.199). In other words, for the uniform probability distribution, $H(\mathbf{p})$ is maximum and all other distributions have lower entropy. Moreover, notice that from (4.3.10) follows that for any other set of probabilities, say $\{q_i\}$, $F\left(\{p_i = n^{-1}\}\right)/F\left(\{q_i\}\right)$ increases asymptotically as

$$\frac{F\left(\{p_i = n^{-1}\}\right)}{F\left(\{q_i\}\right)} \to \exp\left\{N\left[\ln\left(n\right) - H\left(\mathbf{q}\right)\right]\right\}, \tag{4.3.18}$$

and passes all bounds as N grows large. This means that number of configurations which results into assignment of uniform probability by far exceeds all other alternatives.

The above discussion shows that the maximum entropy is achieved whenever the ingredients of the mixture are randomly distributed among the samples. Thus, this implies that Gy's Reproducibility concept is equivalent to maximization of the sample entropy. Indeed, this result is not surprising, since the distribution of the particles at maximum entropy can be realized in greatest number of ways. This is what Reproducibility is all about. However, sampling real mixtures rarely results in uniform distribution. As Fan *et al.* (1979) have pointed out, this is due to either the mixing process is incomplete or because the segregation occurs by differences in the physical properties of the mixture components. As it will be shown later, these phenomenons enter the model as constraints which are expressed in the form of expectations of certain functions. This would, in general, result in probabilities other than the uniform probability. Nonetheless, based on the above discussions, it is evident that the entropy of sample distribution can be used as a measure of Reproducibility. Indeed, this can be done by calculating the difference between the observed entropy and entropy corresponding to the uniform distribution. As was mentioned previously, this difference times $2N$ is χ^2 distributed and therefore can easily be evaluated (see Section 3.7 on entropy concentration theorem). Thus, the greater the difference is, the less is the Reproducibility of the samples and vice versa.

4.3.2 Continuous Systems

In a continuous system, the samples can be thought as a discrete time series. The simplest example of such a system is a binary closed continuous system. Each element in the time series, say $r(t)$, represents the number of tracer particles in a sample. Each sample consists of the portion of the mixture which left the mixer between t and $t+\Delta t$, where Δt is a predetermined time length. Therefore, the entire time series $\{r(0), \ldots, r(T)\}$ must be considered as a single trial and combinatorial arguments refer to a collection of many different realization of it.

Now, recall that the continuous mixers are characterized by their ability to reduce the axial fluctuations feed into the mixer, which in turn is related to RTD of the mixer. The RTD of a mixer is determined by imposing an interference factor on the mixer whilst it is in stationary operation and observing how this interference is broken down inside it. The tracer must be quantitatively measurable in small concentrations in order not to change the flow pattern. Its motion should correspond to that of the mixture. One common method of measuring RTD of a mixer is by determining its impulse response. Indeed, a slug of concentrated tracer is injected into the feed at a specified time t_0 (reference time), and the concentration of the tracer in the outlet stream is measured at various times, t. The time of injection should be virtually instantaneous, that is, within a period much shorter than the mean residence time. Since, all the tracer material enters the mixer at the same time then each sample from the outlet only contains the tracer particles with residence time between t_i and $t_i+\Delta t$[4]. Following the same arguments as § 4.3.1, it follows that the maximum value of $F(\{p(t_i)\})$ is attained for the set $\{p(t_i)\}$ which maximizes

$$D(\mathbf{p}:\mathbf{v}) = -\sum_{i=1}^{n} p(t_i) \ln\left(\frac{p(t_i)}{v(t_i)}\right), \tag{4.3.19}$$

where $t_i = t_{i-1} + \Delta t$. Now, recall that in the above context, $p(t_i)$ is the probability of finding a tracer particle in sample i. However, this can also be interpreted as the

---

[4] For the sake of simplicity in notation, we shall assume that $t_0 = 0$.

probability of finding a tracer particle having a residence time in $< t_i, t_{i+1} >$. This implies that

$$p\left(t_i\right) \approx f\left(t_i\right)\Delta t, \tag{4.3.20}$$

where $f\left(t_i\right)$ is the probability density for the RTD of the system. Moreover, it is obvious that

$$v\left(t_i\right) \approx \frac{u\left(t_i\right)}{V}\Delta t, \tag{4.3.21}$$

where $u\left(t_i\right)$ is the volumetric flow rate of the mixture out of the mixer at time t_i and V is the total volume of the mixture at the end of the trial. Substituting the above equations back into Eqn. (4.3.19) and letting $\Delta t \to 0$, one gets

$$D\left(f{:}u\right) = -\int_0^\infty f\left(t\right)\ln\left[\frac{f\left(t\right)}{u\left(t\right)/V}\right]dt. \tag{4.3.22}$$

The above equation is the continuous version of the Eqn. (4.3.10) for binary closed continuous systems. However, note that, in general, $u\left(t\right)$ is not constant and hence it affects the choice of $f\left(t\right)$. Indeed, Eqn. (4.3.10) can be rewritten as

$$D\left(f{:}u\right) = -\left(\int_0^\infty f\left(t\right)\ln f\left(t\right)dt - \int_0^\infty f\left(t\right)\ln\left[\frac{u\left(t\right)}{V}\right]dt\right). \tag{4.3.23}$$

That is, maximizing $D\left(f{:}u\right)$, maximizes our ignorance about the details which do not contain any regularity at the outlet of the mixer.

Now, there is another way of looking at this. Note that since non-tracer components are assumed to be stationary and per definition each sample contains only tracer particles of similar residence time then

$$v\left(t_i\right) \propto \frac{t_i}{n\bar{t}}, \tag{4.3.24}$$

where n is the number of samples and $\bar{t}$ is the mean residence time of the system. Let T denote the total mixing time during the whole cycle of our tracer experiment. Then since per definition Δt is fixed,

$$n = \frac{T}{\Delta t} \Rightarrow v\left(t_i\right) \propto \frac{t_i}{\bar{t}T}\Delta t, \tag{4.3.25}$$

from which follows that

$$D\left(f{:}u\right) = -\int_0^\infty f\left(t\right)\ln f\left(t\right)dt + \int_0^\infty f\left(t\right)\ln t dt - C \tag{4.3.26}$$

in which C is a constant. Hence, it is reasonable to assume that

$$\int_0^\infty f\left(t\right)\ln\left[\frac{u\left(t\right)}{V}\right]dt \propto \ln g = \int_0^\infty f\left(t\right)\ln t dt, \tag{4.3.27}$$

where g is the *geometric mean residence time*. In other words, any choice of $f\left(t\right)$ should also fulfil the constraint on the geometric mean residence time of the system. Thus, in the case of closed binary continuous systems, the probability density for the system RTD is determined by maximizing

$$-\int_0^\infty f\left(t\right)\ln\left[f\left(t\right)\right]dt, \tag{4.3.28}$$

subject to the constraints

$$\int_0^\infty f\left(t\right)dt = 1 \text{ and } \int_0^\infty f\left(t\right)\ln t dt = \ln g. \tag{4.3.29}$$

Moreover, one should also in addition take into account any other possible constraints.

4.4 Interlude: Comments

The combinatorial argument that resulted into Eqn. (4.3.10) is based on a more general argument known as *monkey argument* (see Sivia 1996, p.116). The argument, in its more general form, goes something like this that various propositions (samples, in our case) could be represented by different boxes, which might have different sizes, into which pennies (tracer particles, in our case) are thrown at random. This job is often delegated to a hypothetical team of monkeys, to denote the fact that there should be no underlying bias in the procedure. Following almost the same argument as above, one reaches at principle of maximum entropy. In fact as Jaynes (1982) pointed out, this rational was well understood by Jacob Bernoulli and Laplace, although they did not use the principle in its logarithmic form. It might probably be that the Stirling's approximation was unknown at the time (Stirling's formula was first published in his most important work *Methodus Differentialis* in 1730. This formula appears as Example 2 to Proposition 28).

We need to be also clear about the rationale of the limit $N \to \infty$, in relation (4.3.10). We pass to the limit, not because we believe that N is infinite; we know that it is not. We pass to the limit rather because we know that this will simplify the calculation without affecting the final result. Indeed, as was explained in Chapter 3, p_i simply represents a reasonable degree of belief. It is easy to convince oneself that for sufficiently large N, the relation in (4.3.10) should be a good approximation and there is no need to let $N \to \infty$. In fact, the exact value of N is not important. This is of great advantage since in reality N is usually unknown.

As mentioned previously, it is important to distinguish between frequency and probability. A frequency is something that can be measured, whereas a probability represents a state of knowledge. As Tribus (1961, p.43) also pointed out, a statement about frequencies can, by direct measurement, be shown to be wrong. However, a statement about probability can only be demonstrated to be irrational. The combinatorial argument leading to Eqn. (4.3.10) was based on imagining many identical experiments were conducted. Such a thought experiment is not proof but rather an attempt to clarify the plausibility of certain outcomes, while be able to reason consistently without violating common sense.

4.5 Modelling of Multicomponent Closed and Open Systems

It is instructive to explain in detail how the model for binary closed batch and continuous systems can be extended to multicomponent closed/open systems. However, the underlying arguments remain similar to the combinatorial argument of Section 4.3. Therefore, in the following most of the calculations are omitted.

4.5.1 Multicomponent Batch Systems

Suppose that the system under study is a $(k+1)$-component closed batch system which is divided into n samples. Then the sample matrix consists of n rows and k columns. Let N_j denote the number of particles of type j in the mixture. Then

$$\sum_{i=1}^{n} r_{ij} = N_j \text{ for } j = 1, \ldots, k \tag{4.5.1}$$

where r_{ij} denotes the number of particles of type j in sample i. That is, the sum of elements in each column of the sample matrix is predetermined. Thus the sample matrix can be considered as a *contingency table*. The applications of contingency tables have had a long history and are well established in statistics (see e.g. Dougherty 1990, p.509). However, there is a vital difference between the sample matrix and the contingency table. In general, the sum of elements of each row of the sample matrix is unknown. The reason is that the volume of the sample is fixed but the particles are not mono-size. Therefore, the total number of particles in each sample may vary. Nonetheless, other constraints like the mean number of particles in each sample is possible.

To this end, let

$$p\left(r, j \middle| i, I\right) = p_{rij}, \tag{4.5.2}$$

where p_{rij} is the probability of finding r particles of type j in a given sample i and

$$N = \sum_{j=1}^{k} N_j. \tag{4.5.3}$$

Now, following similar combinatorial arguments as in § 4.3.1, it can be shown that, in the case of correct sampling of $(k+1)$-component closed batch system, the maximum value of $F(\{p_{rij}\})$ is attained for the set $\{p_{rij}\}$ which maximizes

$$H(\{p_{rij}\}) = -\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{r=0}^{s_j} p_{rij} \ln p_{rij}, \tag{4.5.4}$$

subject to

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{r=0}^{s_j} p_{rij} = 1, \quad \sum_{i=1}^{n}\sum_{r=0}^{s_j} r p_{rij} = N_j \text{ and } \sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{r=0}^{s_j} r p_{rij} = N, \tag{4.5.5}$$

where s_j is the maximum possible number of particles of type j in a sample. Of course if there are any additional constraints, they should also be taken into account.

Now, suppose that the system under study is a $(k+1)$-component open batch system. Since, in the case of open systems, one have only information on a small fraction of the system, the only available informations are in the form of expectancy constraints. This means that, mathematically, the analysis is similar to the closed system with the only difference that the constraints are

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{r=0}^{s_j} p_{rij} = 1, \quad \sum_{i=1}^{n}\sum_{r=0}^{s_j} r p_{rij} = \bar{N}_j \text{ and } \sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{r=0}^{s_j} r p_{rij} = \bar{N}, \tag{4.5.6}$$

where $\bar{N}_j$ and $\bar{N}$ are expected or average number of particles of type j and the total expected number of all k- type particles in the mixture, respectively. It is assumed that one has prior knowledge of the number of components in the mixture. If this information is not available, the value of k has to be estimated from the samples. The parameter s_j

can also be estimated from the volume of the sample and the particles of type j.

4.5.2 Multicomponent Continuous Systems

In a continuous system, the sample matrix for a $(k+1)$-component closed continuous system is

$$\begin{pmatrix} \mathbf{r}'(0) \\ \vdots \\ \mathbf{r}'(T) \end{pmatrix} = \begin{pmatrix} r_1(0) & \cdots & r_k(0) \\ \vdots & \ddots & \vdots \\ r_1(T) & \cdots & r_k(T) \end{pmatrix},$$

where $T = n\Delta t$. Following similar arguments as in the case of closed batch system, it is easy to show that in the case of $(k+1)$-component closed continuous system, the Eqn. (4.3.19) can be written as

$$D(\mathbf{p}:\mathbf{v}) = -\sum_{j=1}^{k}\sum_{i=1}^{n} p_j(t_i)\ln\left(\frac{p_j(t_i)}{v_j(t_i)}\right), \tag{4.5.7}$$

where $p_j(t_i)$ is the probability of finding a tracer particle of size class j in sample i. As it was mentioned in § 4.3.2, $p_j(t_i)$ can be related to the the density of the RTD and keeping in mind that the RTD of the system, in general, depends on the size of the particles, one gets

$$p_j(t_i) \approx f(x_j, t_j)\,\Delta x \Delta t, \tag{4.5.8}$$

where x_j is the equivalent diameter of the particle of size class j. Moreover, it is easy to show that

$$v_j(t_i) \approx \frac{u(x_j, t_i)}{V}\Delta x \Delta t, \tag{4.5.9}$$

where $u(x_j, t_i)$ is the volumetric flow rate of particles of size x_j and V is the total volume of the mixture at the end of the trial. Substituting the above equations in (4.5.7) and letting $\Delta x \to 0$ and $\Delta t \to 0$,

$$D(f{:}u) = -\int_0^\infty \int_0^\infty f(x,t)\ln\left[\frac{f(x,t)}{u(x,t)/V}\right] dx dt. \tag{4.5.10}$$

Now, as in § 4.3.2, it follows that

$$\int_0^\infty \int_0^\infty f(x,t) \ln \left[\frac{u(x,t)}{V} \right] dxdt \propto \int_0^\infty \int_0^\infty f(x,t) \ln [t(x)] \, dxdt, \tag{4.5.11}$$

where $t(x)$ is the residence time of the particles of size x in the mixer. The right hand side of the above equation is just the geometric mean residence time which is a measure of fluctuations at the outlet of the mixer. Therefore, in the case of closed multicomponent continuous systems, the probability density for the system RTD is determined by maximizing

$$-\int_0^\infty \int_0^\infty f(x,t) \ln [f(x,t)] \, dxdt, \tag{4.5.12}$$

subject to the constraints

$$\int_0^\infty \int_0^\infty f(x,t) \, dxdt = 1 \text{ and } \int_0^\infty \int_0^\infty f(x,t) \ln [t(x)] \, dxdt = \ln g. \tag{4.5.13}$$

One should also in addition take into account any other possible constraints. In the case where the system is open, average values of the constraints, which can be determined experimentally, should be used.

4.6 A Model for Mixture Heterogeneity

The mathematical model presented in previous section models the accuracy and reproducibility of samples and gives a simple but accurate mathematical description of each of these terms. However, it is rarely possible to count the number of particles in each sample. The content of each sample is usually described by the amount of relative weight of each mixture component in the sample. Although, usually the density of each component is known, it is almost impossible to get an exact figure on how many particles it corresponds to. The reason for this is that, in general, the size of particles of any given mixture component varies over a wide range. This means that for a given sample weight, there can be no unique number of particles of any type. Therefore, it is desired to characterize the mixture based on weight of components rather than the number of particles. In the following it will be demonstrated how this can be achieved.

4.6.1 Sample Space

As it was argued earlier, one of the important conditions for correctness of samples is that the volume of samples are equal which in turn is determined by the scale of scrutiny. However, this in general does not put any significant restriction on the shape of the sample. It is conceivable that the shape of the samples can in some way affect the observed structure of the mixture. However, except for regular shapes, it is difficult to find a practical way of describing an irregular shape. Nevertheless, it is always possible to check the similarity between two shapes. Since, the number of particles in a mixture is finite then there would be only finite number of possible shapes to be considered. Hence, one can talk about the set of all possible sample shapes, without being specific about a particular way of describing them (For further discussion see Section 5.1).

Based on the above discussion, in order to avoid favouring any particular shape, in the following a set-theoretical approach is adapted.

Definition 4.1 *Let Ω_L be the collection of all partitions of the lot L into disjoint non-empty subsets. A sub-collection is called* ***sample space*** *of order $n \in \mathbb{N}$, denoted as $\Omega^n_\alpha \subseteq \Omega_L$, if it is the collection of all possible partitions of L, each of which consists of n equivolume samples (subsets) with predetermine sets of shape, denoted by index α, and is a result of repetition of the same mixing experiment.*

The mixing experiment referred to in the above definition is similar to the one described in § 4.3.1. In this relation, each partition corresponds to all the samples taken in a single sampling trial, which fulfils the system constraints. Moreover, notice that the position of samples in each partition does matter. That is, the neighbourhood that each sample resides in is important and therefore the sampling process should not disturb it. In fact, this is what determines the structure of the mixture. However, the positions of particles within each sample are irrelevant and do not affect the structure of the mixture. Accordingly, let the *state of a mixture* be the spatial configuration of constituents of the mixture at a particular instant in time and the set of all such states as *state space.* Note that the position of each particle, with respect to a common reference, can be uniquely determined by three numbers. This implies that a state of the mixture is determined by a set of $3N$ numbers, where N denotes the total number of particles in the mixture. In other words, the state space can be considered as a sub-space of $\mathbb{R}^{3N}$.

Two states of a mixture are *equivalent modulo* n, if the collection of their respective n samples of given shape are equal. This is in accordance with the notion of scale of scrutiny. For a given scale of scrutiny, in general, many states of the mixture are indistinguishable, i.e., equivalent modulo n. Nevertheless, for the sake of mathematical consistency, it is assumed that the state space only consists of those states which are attainable by the aforementioned mixing experiments, modulo n.

Definition 4.2 *Let* Ω^r_α *and* Ω^s_β *be two sample spaces with predetermined shape indexes. Furthermore, let* $\mathcal{A} = \{A_i\} \in \Omega^r_\alpha$ *and* $\mathcal{B} = \{B_j\} \in \Omega^s_\beta$ *denote the partitions* $L = \cup_{i=1}^r A_i$ *and* $L = \cup_{j=1}^s B_j$*, respectively. Then* $\mathcal{B}$ ***refines*** $\mathcal{A}$*, written* $\mathcal{A} \leq \mathcal{B}$*, if each* $A_i \in \mathcal{A}$ *is union of some samples* $B_j \in \mathcal{B}$*. Similarly,* Ω^s_β *refines* Ω^r_α*, denoted* $\Omega^r_\alpha \leq \Omega^s_\beta$*, if every partition in* Ω^r_α *has a refinement in* Ω^s_β*, i.e., for every* $\mathcal{A} \in \Omega^r_\alpha$*, there exists* $\mathcal{B} \in \Omega^s_\beta$ *such that* $\mathcal{A} \leq \mathcal{B}$*. Furthermore, if for every* α *there exists a* β *such that* $\Omega^r_\alpha \leq \Omega^s_\beta$*, then we shall say that the* ***ensembles*** $\Omega^r \leq \Omega^s$*, where* $\Omega^r = \cup_\alpha \Omega^r_\alpha$ *and* $\Omega^s = \cup_\beta \Omega^s_\beta$*.*

Strictly speaking, by the above definition, an ensemble is not a sample space, because samples in a sample space are result of repetition of the same mixing experiment. This also means that the sample shapes remain the same in every trial, whereas, in the case of ensemble, there is no restriction on the shape of the samples. An immediate consequence of the above definition is stated in the following proposition.

Proposition 4.1 *Let Ω^r_α and Ω^s_β be two non-empty sample spaces of Ω_L where α and β denote the corresponding shape indexes. Then, $\Omega^r_\alpha \leq \Omega^s_\beta$ if and only if $s = qr$, where q is a positive integer.*

Proof. *Assume that $s = qr$ for some positive integer q. Now, let $\mathcal{A}$ be any partition in Ω^r_α. Then each sample $A_i \in \mathcal{A}$ can be divided into q equivolume samples of given shapes. This new partition corresponds to the same mixture state and hence must belong to Ω^s_β for some β. In the similar manner, it is easy to see that by splitting each sample into q equivolume samples of shape index β, one can find a refinement for any partition $\mathcal{A} \in \Omega^r_\alpha$ in Ω^s_β and hence, by definition, $\Omega^r_\alpha \leq \Omega^s_\beta$. Conversely, assume that $\Omega^r_\alpha \leq \Omega^s_\beta$ for some α and β. Since by hypothesis the sampling is Correct, then the samples are equivolume. In other words there exists a positive integer q such that $s = qr$.* ■

Corollary 4.1 *Let Ω^r and Ω^s be two non-empty ensembles of Ω_L. Then $\Omega^r \leq \Omega^s$ if and only if $s = qr$, where q is a positive integer.*

Proof. *This is a direct consequence of Proposition (4.1) and Definition (4.2).* ■

Until now, the sample space was associated to physical collection of particles. But in order to take full advantage of this abstract definition, one needs to work with numbers. That is, if one likes, it is necessary to assign coordinates to each partition, which in a later stage can be used to quantify their properties. In general, there is no unique way to assign coordinates to a partition. Therefore, there is a great degree of freedom in choosing any consistent procedure which fits the purpose of the application of interest. Of course, one's conclusions should be independent of any particular choice.

4.6.2 Mixture Heterogeneity

Let c_i and c_L denote the relative weight of the key component in sample i and lot L, respectively. Notice that depending on the system being batch or continuous, the index i determines the position or the time the sample is taken, respectively. Moreover, the exact value of c_L is only known in the case of closed systems and in the case of open systems it has to be estimated from the corresponding sample space. Accordingly, if it was possible to have a homogeneous mixture, then all the samples would have been identical to each other. In particular, if M_i denotes the mass of the sample i, then $c_i M_i = c_L M_i$, irrespective of i[5]. Therefore, contribution of heterogeneity by sample i should be proportional to $(c_i - c_L)M_i$. That is

$$h_i = K(c_i - c_L)M_i, \tag{4.6.1}$$

where K is an arbitrary non-zero constant. In order to be able to compare the contribution of heterogeneity from different samples, it is assumed that h_i is dimensionless. Since, K is arbitrary and its value should not affect the quality of the mixture, it can be chosen such that

$$h_i = \frac{(c_i - c_L)M_i}{c_L M^*}, \tag{4.6.2}$$

where $c_L M^*$ denotes the mean mass of the key component in the samples. Similarly h_i can be defined for each component in the mixture. Thus one can construct a so-called *heterogeneity* matrix similar to the sample matrix. That is

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}'_1 \\ \vdots \\ \mathbf{h}'_n \end{pmatrix} = \begin{pmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nk} \end{pmatrix}, \tag{4.6.3}$$

where n stands for the number of samples and $k + 1$ is the number components of the mixture. Accordingly, the matrix element h_{ij} corresponds to contribution of heterogeneity of component j of the mixture in sample i, to the mixture.

Mathematically, construction of each heterogeneity matrix can be considered as assignment of coordinates to each partition in the sample space. The collection of all these

---

[5] Note that the following equality always holds, irrespective of mixture structure,

$$\sum_{i=1}^{n} c_i M_i = \sum_{i=1}^{n} c_L M_i = c_L M_L$$

coordinate points constitute $\Gamma^{n\times k} \subseteq \mathbb{R}^{n\times k}$.

Definition 4.3 *Let* $\mathbf{H} : \Omega_\alpha^{n\times k} \to \Gamma_\alpha^{n\times k}$*, where*

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}_1' \\ \vdots \\ \mathbf{h}_n' \end{pmatrix} = \begin{pmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nk} \end{pmatrix},$$

and $\Gamma_\alpha^{n\times k}$ *is some discrete or continuous subset of* $\mathbb{R}^{n\times k}$*. Then* h_{ij} *is called a random variable if it is a measurable function. The kernel of* $\mathbf{H}$ *is the partition of* $\Omega_\alpha^{n\times k}$ *into pre-images under* $\mathbf{H}$*.*

Notice, to say that h_{ij} is a measurable function means that the pre-images are *measurable sets*. In this work, the class of relevant measures are the so-called *probability measures*.

It appears that Gy (1998, p.64) was the first investigator to introduce the concept of heterogeneity function, which also plays a major role in his approach to sampling theory. Nonetheless, the degree of success would greatly depend on how good this model coincides with the real notion of heterogeneity. In order to be able to demonstrate this clearly, for the rest of this chapter, it is assumed that the mixture under study is binary. This means that the heterogeneity matrix consists of a single column. Hence, the heterogeneity functions or coordinates over Ω_α^n can be defined as

$$\mathbf{h} : \Omega_\alpha^n \to \mathbb{R}^n,$$

where

$$\mathcal{A} \mapsto \mathbf{h}'(\mathcal{A}) = (h(A_1), \ldots, h(A_n))' = (h_1, \ldots, h_n)' \in \mathbb{R}^n,$$

for a given key component.

Heterogeneity functions contain much useful information about the structure of the mixture. In particular, their average properties are of great importance. The reason for this is that most of the observable physical phenomenon that affect the structure of the mixture, also affect the average properties of the heterogeneity functions. But, in general, the knowledge about the range of each heterogeneity function is incomplete. Therefore, one needs to somehow estimate the average properties of the heterogeneity functions. This is where the probability theory enters. In order to be able to assign a probability distribution to $\mathbf{h}$, the following procedure can be followed:

1. Use the prior information to assign a probability distribution to $\mathbb{R}^n$.
2. Use the experimental value of $\mathbf{h}$ and the probability distribution, determined in the previous point, to estimate the average properties of $\Gamma_\alpha^n \subseteq \mathbb{R}^n$.

A general discussion of the underlying concepts behind the above procedure was given in Chapter 3 and further elaborated in Section 7.2. For the rest of this Section it is assumed that a *probability density function* (pdf) which adequately models the variations in $\mathbf{h}$ is known.

Now, let $p\left(\mathbf{h}|\, I\right)$ denote the conditional probability of $\mathbf{h}$ given the prior information I about the structure and shape of the samples. As was argued previously, $p\left(\mathbf{h}|\, I\right)$ assigns a reasonable degree of belief to subsets of $\mathbb{R}^n$. If there exists sufficient amount of information about Γ_α^n, $p\left(\mathbf{h}|\, I\right)$ assigns higher probabilities to elements in Γ_α^n than to its complement $\mathbb{R}^n \backslash \Gamma_\alpha^n$. The lack of knowledge about the range of $\mathbf{h}$ leads usually to assume that it can attain any value in $\mathbb{R}^n$. This is the reason that the assigned distributions are usually continuous. Based on experience, there is never any need to consider infinite sets or measure theory in real, exact problems. Indeed, any data set that can actually be recorded and analyzed is digitized into multiples of some smallest element. But the continuity assumption makes the mathematics more manageable and unless there is some strong evidence to believe otherwise, it is a good approximation.

Since the information contained in a pdf can be extracted by using usual rule for estimating its moments, then, in general, it is reasonable to believe that the moments for $\mathbf{h}$ with respect to $p\left(\mathbf{h}|\, I\right)$, if they exist, contain some useful information about the average properties of Γ_{α}^{n}[6]. In the case of multivariable distributions, the *power law moments* of the form

$$\mathbb{M}_s = \frac{1}{n}\int_{\mathbb{R}^n}\left(\sum_{i=1}^{n} h_i^s\right) p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h}, \tag{4.6.4}$$

where $d\mathbf{h} = dh_1 \cdots dh_n$, are the most interesting objects. In particular, in the context of present chapter, the first and second moments for $\mathbf{h}$ with respect to the distribution $p\left(\mathbf{h}\middle|\, I\right)$ are of special interest.

Definition 4.4 *let $p\left(\mathbf{h}\middle|\, I\right)$ denote the conditional probability of $\mathbf{h}$, given the prior information I. Then the first and second moments for $\mathbf{h}$ are*

1. $\mathbb{M}_1 = \mu_{\mathbf{h}} = \frac{1}{n}\int_{\mathbb{R}^n}\left(\sum_{i=1}^{n} h_i\right) p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h} = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^n} h_i p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h} = \frac{1}{n}\sum_{i=1}^{n}\langle h_i\rangle$.
2. $\mathbb{M}_2 = \sigma_{\mathbf{h}}^2 + \mu_{\mathbf{h}}^2 = \frac{1}{n}\int_{\mathbb{R}^n}\left(\sum_{i=1}^{n} h_i^2\right) p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h} = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^n} h_i^2 p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h} = \frac{1}{n}\sum_{i=1}^{n}\langle h_i^2\rangle$.

There are some general properties of $\mathbb{M}_1$ and $\mathbb{M}_2$, independent of the pdf assigned to $\mathbf{h}$, which are useful in the proceedings.

Proposition 4.2 *let $\mathbb{M}_1$ and $\mathbb{M}_2$ denote the first and the second moments for $\mathbf{h}$, respectively. Then*

1. $\mathbb{M}_1 = \mu = 0$
2. $\mathbb{M}_2 = 0$ *if and only if the cumulative probability* $P\left(\mathbf{h} \neq 0\middle|\, I\right) = 0$

Proof. *Let $p\left(\mathbf{h}\middle|\, I\right)$ denote the probability density for $\mathbf{h}$. Then from the Definition (4.4) follows that*

$$\mathbb{M}_1 = \mu_{\mathbf{h}} = \frac{1}{n}\int_{\mathbb{R}^n}\left(\sum_{i=1}^{n} h_i\right) p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h} \tag{4.6.5a}$$

$$= \frac{1}{n}\int_{\mathbb{R}^n}\left(\sum_{i=1}^{n} \frac{(c_i - c_L)M_i}{c_L M^*}\right) p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h} \tag{4.6.5b}$$

$$= \frac{1}{nc_L M^*}\int_{\mathbb{R}^n}\left(\sum_{i=1}^{n} c_i M_i\right) p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h} - \frac{1}{nc_L M^*}\int_{\mathbb{R}^n}\left(\sum_{i=1}^{n} c_L M_i\right) p\left(\mathbf{h}\middle|\, I\right) d\mathbf{h} \tag{4.6.5c}$$

---

[6] For a discussion on the existence of moments with respect to p see Jeffreys (1998, p.86). See also Prohorov and Rozanov (1969, § 4.3) for more details.

but by definition

$$\sum_{i=1}^{n} c_i M_i = c_L M_L \text{ and } \sum_{i=1}^{n} c_L M_i = c_L M_L \tag{4.6.6}$$

and hence statement (1) follows. Note that this also implies that $\mathbb{M}_2 = \sigma^2$*, where* σ^2 *is the variance for* $\mathbf{h}$*. Now, to prove the second statement, assume that* $\mathbb{M}_2 = 0$ *and suppose on the contrary that* $P\left(\mathbf{h} \neq 0 | \, I\right) \neq 0$*. Then it is possible to find an* $\varepsilon > 0$ *such that*

$$P\left(\|\mathbf{h}\| > \varepsilon | \, I\right) = \int_{\|\mathbf{h}\|>\varepsilon} p\left(\mathbf{h} | \, I\right) d\mathbf{h} \neq 0, \tag{4.6.7}$$

where $\|\mathbf{h}\|$ *denotes the length of the vector* $\mathbf{h}$*. However,*

$$\mathbb{M}_2 = \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}^n} h_i^2 p\left(\mathbf{h} | \, I\right) d\mathbf{h} \tag{4.6.8a}$$

$$\geq \frac{1}{n} \sum_{i=1}^{n} \int_{\|\mathbf{h}\|>\varepsilon} h_i^2 p\left(\mathbf{h} | \, I\right) d\mathbf{h} \tag{4.6.8b}$$

$$\geq \frac{\varepsilon^2}{n} \int_{\|\mathbf{h}\|>\varepsilon} p\left(\mathbf{h} | \, I\right) d\mathbf{h} > 0, \tag{4.6.8c}$$

which clearly contradicts the assumption that $\sigma^2 = 0$*. Hence, the statement* $P\left(\mathbf{h} \neq 0 | \, I\right) = 0$*, is true. Conversely, if* $P\left(\mathbf{h} \neq 0 | \, I\right) = 0$ *then* $p\left(\mathbf{h} | \, I\right) = \prod_{i=1}^{n} \delta\left(h_i\right)$*, where* δ *denotes the impulse function. This is because the total probability over the whole space should always be equal to unity. Thus*

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}^n} h_i^2 \prod_{i=1}^{n} \delta\left(h_i\right) d\mathbf{h} = 0. \tag{4.6.9}$$

■

From the above proposition, it is evident that σ^2 can be used as a global measure for heterogeneity. In the future, this measure would be referred to as *mixture heterogeneity.* Mixture heterogeneity is a measure of *heterogeneous properties* of the mixture and the contribution due to error in estimation based on insufficient knowledge over the range of heterogeneity functions. The heterogeneous properties of the mixture can be divided into two categories, static and dynamic. The static properties are properties like size and shape of the particles in the mixture, whilst, the dynamic properties depend on if the mixture is at equilibrium or not. In general, the mixture heterogeneity is higher when the mixture is far from equilibrium as compared to being at equilibrium. Further discussion on this topic can be found in § 6.1.2.

As was mentioned above, the range of $\mathbf{h}$ can be considered as coordinates for each sampling trail in $\Gamma_\alpha^n \subseteq \mathbb{R}^n$. The distance of each point to the origo in $\mathbb{R}^n$, corresponds to the Euclidian length $\|\mathbf{h}\|$. Thus, the mixture heterogeneity is equivalent to the average square of the distances of all the points from the origo in $\mathbb{R}^n$. Geometrically, the set of all possible points in Γ_α^n constitute a hyper-plane $\mathcal{P}^n$, in which

$$h_1 + \cdots + h_n = 0 \text{ for all } (h_1, \ldots, h_n)' \in \mathbb{R}^n. \tag{4.6.10}$$

The mixture heterogeneity gives an indication on the region where, in general, one can expect to find majority of points belonging to Γ_α^n. Indeed, the majority of these regions are clustered within the closed set which is created by the intersection of the hyper-plane $\mathcal{P}^n$ and the hyper-sphere S^{n-1}, with the radius σ. Similarly, all the above results can be extended to the case of ensembles.

Now, notice that

$$\begin{aligned} h_i + h_j &= \frac{(c_i - c_L)\, M_i}{c_L M^*} + \frac{(c_j - c_L)\, M_j}{c_L M^*} \\ &= \frac{(c_{i+j} - c_L)\, M_{i+j}}{c_L M^*}, \end{aligned} \tag{4.6.11}$$

in which

$$M_{i+j} c_{i+j} = c_i M_i + c_j M_j \text{ and } M_{i+j} = M_i + M_j. \tag{4.6.12}$$

Assume that

$$\mathbf{h} = (h_1, \ldots, h_{2r})' \in \Gamma_\beta^{2r} \text{ and } \mathbf{g} = (g_1, \ldots, g_r)' \in \Gamma_\alpha^r, \tag{4.6.13}$$

such that $\Omega^r_\alpha \leq \Omega^{2r}_\beta$. Then a sample, say k, in a partition in Ω^r_α is a combination of samples, say i and j, in a partition in Ω^{2r}_β. From the above equations, it is easy to show that

$$g_k = \frac{1}{2}\left(h_i + h_j\right). \tag{4.6.14}$$

This result can easily be generalized as in the following proposition.

Proposition 4.3 *Let* $\Omega^r_\alpha \leq \Omega^s_\beta$*, such that* $\mathbf{g}$ *and* $\mathbf{h}$ *denote the heterogeneity functions on* Ω^r_α *and* Ω^s_β*, respectively. Then for any given partition* $\mathcal{A} \in \Omega^r_\alpha$*, there exists a partition* $\mathcal{B} \in \Omega^s_\beta$*, such that the heterogeneity functions on* $\mathcal{A}$ *and* $\mathcal{B}$ *are related in the following manner*

$$g_i = \frac{1}{q} \sum_{j=k_{i-1}+1}^{k_i} h_j, \tag{4.6.15}$$

where $s = qr$ *and* $k_i = iq$*.*

Proof. *The prove follows from Proposition (4.1) and induction on the order of the sample spaces.* ■

Corollary 4.2 *Let* Ω^r *and* Ω^s *be two non-empty ensembles of* Ω_L *such that* $\Omega^r \leq \Omega^s$*. Furthermore, let* $\mathbf{g}$ *and* $\mathbf{h}$ *denote the heterogeneity functions defined on* Ω^r *and* Ω^s*, respectively. Then for any given partition* $\mathcal{A} \in \Omega^r$*, there exists a partition* $\mathcal{B} \in \Omega^s$*, such that the heterogeneity functions on* $\mathcal{A}$ *and* $\mathcal{B}$ *are related in the following manner*

$$g_i = \frac{1}{q} \sum_{j=k_{i-1}+1}^{k_i} h_j, \tag{4.6.16}$$

where $s = qr$ *and* $k_i = iq$*.*

Proof. *The prove follows from Corollary (4.1) and Proposition (4.3).* ■

Proposition (4.3) indicates that as scale of scrutiny becomes coarser, i.e., each sample contains more particles; the observer becomes more ignorant of finer variations. In fact, it demonstrates that a coarsening of scale of scrutiny is equivalent to the average of the contribution of heterogeneity of the q neighbouring samples of the finer partition. In other words, the heterogeneity function of the coarser sample space can be found by letting the heterogeneity function of the finer sample space pass through a suitable low-pass filter. As a result of this filtering operation, one would expect that the mixture becomes less heterogeneous. Note that this process is asymmetric. Indeed, the low-pass filtering is an irreversible process. Therefore, unless there is sufficient extra information, it is impossible to recover the information about the finer sample space from the coarser one.

4.6.3 The Relation between Pdfs of Non-relatively Prime Sample Spaces and Ensembles

Two integers are *relatively prime* if they share no common positive factors (divisors) except 1. For example, 4 and 9 are relatively prime, whereas 4 and 8 are non-relatively prime.

Definition 4.5 *Two sample spaces are* ***relatively prime*** *if their orders are relatively prime. Similarly, two ensembles are relatively prime if their orders are relatively prime.*

If two sample spaces are non-relatively prime then by Proposition (4.1), there exists two shape indexes α and β such that the sample space with the greater order is a refinement of the sample space with the smaller order. Thus, in the case of non-relatively prime sample spaces, it would be interesting to determine the relation between the pdf of a sample space and the pdf of any of its refinements.

A meaningful relation between a sample space and its refinements can only exist if they share the same mixture state. This means that in moving from a sample space to any of its refinements, one is not allowed to physically move the particles or exchange the position of the samples, which in reality is only possible by convective type operations. Moreover, the transformation is assumed to be correct. That is, in any intermediate stage of the transformation, the samples are equivolume. This restriction does not affect the final structure of the sample space but mathematically, it implies that the information

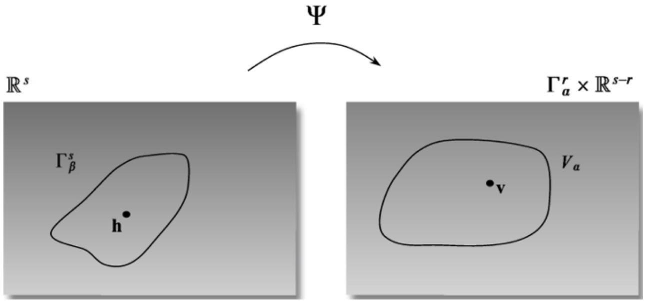

Figure 4.6.1: Schematic representation of the map $\Psi : \Gamma^s_\beta \rightarrow \Gamma^r_\alpha \times R^{s-r}$.

on the structure of a sample space can only be related to its refinements in a continuous manner and vice versa. Accordingly, the information on heterogeneity functions can only be mapped by continuous maps from a sample space to its coarser contra part.

To this end, let $\mathbf{g}\left(\Omega^r_\alpha\right) = \Gamma^r_\alpha$ and $\mathbf{h}\left(\Omega^s_\beta\right) = \Gamma^s_\beta$ such that $s = qr$, in which q is a positive integer and α and β are shape indexes such that $\Omega^r_\alpha \leq \Omega^s_\beta$. Then following Proposition (4.3), one can define a continuous map

$$\Psi : \Gamma^s_\beta \rightarrow \Gamma^r_\alpha \times \mathbb{R}^{s-r} \tag{4.6.17}$$

such that

$$\mathbf{h} = (h_1, \cdots, h_s)' \overset{\Psi}{\mapsto} \left(\frac{1}{q} \sum_{j=1}^{k_1} h_j, \cdots, \frac{1}{q} \sum_{j=k_{r-1}+1}^{k_r} h_j, \frac{1}{q} h_2, \cdots, \frac{1}{q} h_{k_1}, \frac{1}{q} h_{k_1+2}, \cdots, \frac{1}{q} h_s \right)', \tag{4.6.18}$$

where $k_i = iq$ (see Fig. 4.6.1). In a sense, $\Psi\left(\Gamma^s_\beta\right) = V_\alpha$ can be considered as a coordinate patch induced from Γ^s_β into $\Gamma^r_\alpha \times \mathbb{R}^{s-r}$. It is easy to show that Ψ is an *injective* map[7]. Thus, any element, say $\mathbf{v} \in V_\alpha$, has a coordinate defined by (4.6.18) in terms of a unique point on Γ^s_β, i.e., $\mathbf{h} = \Psi^{-1}\left(\mathbf{v}\right)$.

---

[7] A map $\varphi : X \rightarrow Y$ is injective if and only if $\varphi\left(x_1\right) = \varphi\left(x_2\right) \Rightarrow x_1 = x_2$.

Let $f\left(\mathbf{v}|\,\alpha, I\right)$ denote the induced pdf on the patch V_α. The corresponding pull-back, $\Psi^* f$ is defined as

$$\Psi^* f = \Psi^*\left[f\left(\mathbf{v}|\,\alpha, I\right)\right] = f \circ \Psi\left(\mathbf{h}|\,\beta, I\right) = p\left(\mathbf{h}|\,\beta, I\right) \text{ such that } \mathbf{v} \in V_\alpha \text{ and } \mathbf{h} = \Psi^{-1}\left(\mathbf{v}\right). \tag{4.6.19}$$

Then for any $D \subseteq V_\alpha$ such that the *support*[8] of f on D is contained in D, i.e., $\operatorname{supp}_D\left(f\right) \subseteq D$,

$$\int_D f\left(\mathbf{v}|\,\alpha, I\right) d\mathbf{v} = \int_{\Psi^{-1}(D)} \Psi^*\left[f\left(\mathbf{v}|\,\alpha, I\right)\right] d\Psi = \int_{\Psi^{-1}(D)} p\left(\mathbf{h}|\,\beta, I\right) d\Psi, \tag{4.6.20}$$

where $\mathbf{v} = \Psi\left(\mathbf{h}\right)$. The last integral in the above equation is known as the *Stieltjes integral* (Kestelman 1960, p.247), which can be shown to be

$$\int_{\Psi^{-1}(D)} p\left(\mathbf{h}|\,\beta, I\right) d\Psi = \int_{\Psi^{-1}(D)} p\left(\mathbf{h}|\,\beta, I\right) \left|\det J_\Psi\left(\mathbf{h}\right)\right| d\mathbf{h}, \tag{4.6.21}$$

where $\det J_\Psi$ denotes the *Jacobian determinant* of the map Ψ, i.e., $\det\left(\partial\Psi^i/\partial h_j\right)$. Since the region D was an arbitrary subset in which $\operatorname{supp}_D\left(f\right)$ was non-empty, then for every $\mathbf{v} = \Psi\left(\mathbf{h}\right)$,

$$f\left(\mathbf{v}|\,\alpha, I\right) = p\left(\mathbf{h}|\,\beta, I\right) \left|\det J_\Psi\left(\mathbf{h}\right)\right|. \tag{4.6.22}$$

Furthermore, from (4.6.18), it is easy to see that

$$\frac{\partial\Psi^i}{\partial h_j} \in \left\{0, \frac{1}{q}\right\} \text{ for all } i = 1, \ldots, r. \tag{4.6.23}$$

Notice that with appropriate row exchange operations, the matrix J_Ψ can be made triangular with the diagonal entries equal to q^{-1}. It is a well known property of the determinants that the row exchange only changes the sign of the determinant and not its magnitude (Strang 1988, p.214). Thus, since the resultant matrix is triangular then its determinant

---

[8] Support of f on D is defined as the closure of a subset of D in which f is nonzero (see e.g. Loomis and Sternberg 1990, p.336).

is equal to the product of its diagonal elements (Strang 1988, p.216). In other words,

$$\left|\det J_{\Psi}\left(\mathbf{h}\right)\right| = q^{-s}. \tag{4.6.24}$$

Substituting this result back into (4.6.22), yields

$$f\left(\mathbf{v}|\,\alpha, I\right) = q^{-s} p\left(\mathbf{h}|\,\beta, I\right) \text{ such that } \mathbf{v} \in V_{\alpha} \text{ and } \mathbf{h} = \Psi^{-1}\left(\mathbf{v}\right), \tag{4.6.25}$$

which defines the value of the induced pdf for any given element in the patch V_{α}, in term of the pdf on Γ_{β}^{s}.

Now, by marginalizing the nuisance part of V_{α} (see Section 3.5 on marginalization), one arrives at

$$p\left(\mathbf{g}|\,\alpha, I\right) = \int_{V_{\alpha}\cap\mathbb{R}^{s-r}} f\left(\mathbf{v}|\,\alpha, I\right) d\mathbf{v} = q^{-s}\int_{U_{\beta}} p\left(\mathbf{h}|\,\beta, I\right) d\mathbf{h}, \text{ where } U_{\beta} = \Psi^{-1}\left(V_{\alpha}\cap\mathbb{R}^{s-r}\right). \tag{4.6.26}$$

The above equation establishes the relation between the pdf of a sample space and the pdfs of its refinements. Since the integration is an irreversible operation, as was mentioned previously, it is impossible to recover $p\left(\mathbf{h}|\,\beta, I\right)$ from the mere knowledge of $p\left(\mathbf{g}|\,\alpha, I\right)$.

The above result can readily be extended to ensembles. Indeed, by Corollary (4.2), one can extend the map defined in (4.6.18) to yield,

$$\Psi : \Gamma^{s} \to \Gamma^{r} \times \mathbb{R}^{s-r}. \tag{4.6.27}$$

This map is also injective and with similar arguments as above it can be shown that

$$p\left(\mathbf{g}|\,I\right) = \int_{V\cap\mathbb{R}^{s-r}} f\left(\mathbf{v}|\,I\right) d\mathbf{v} = q^{-s}\int_{U} p\left(\mathbf{h}|\,I\right) d\mathbf{h}, \tag{4.6.28}$$

where

$$\Psi\left(\Gamma^{s}\right) = V \text{ and } U = \Psi^{-1}\left(V\cap\mathbb{R}^{s-r}\right). \tag{4.6.29}$$

The above results demonstrate that the relations between the pdfs of sample spaces and pdfs of ensembles are essentially the same and in the case of sample spaces, also independent of sample shapes. However, the domain of integration in the later case is determined by the shape of the samples.

4.6.4 The Relation between Pdfs of Relatively Prime Sample Spaces and Ensembles

In the cases where the sample spaces are relatively prime, Proposition (4.1) can not be used directly as in the above. Nevertheless, as it will be shown shortly, Proposition (4.1) still plays an important role in establishing the relation between pdfs of relatively prime sample spaces and ensembles.

Suppose r and s are relatively prime integers such that $r < s$ and $t = sr$. Then by Proposition (4.1), there exists shape indexes α, β and γ such that $\Omega^s_\beta \leq \Omega^t_\alpha$ and $\Omega^r_\gamma \leq \Omega^t_\alpha$. Furthermore, let $\mathbf{k}$, $\mathbf{h}$ and $\mathbf{g}$ be the heterogeneity functions defined on Ω^t_α, Ω^s_β and Ω^r_γ, respectively. Then, by Proposition (4.3), one can define the maps φ and ψ,

$$\varphi : \Gamma^t_\alpha \to \Gamma^s_\beta \times \mathbb{R}^{t-s} \text{ and } \psi : \Gamma^t_\alpha \to \Gamma^r_\gamma \times \mathbb{R}^{t-r} \tag{4.6.30}$$

such that

$$\mathbf{k} = (k_1, \cdots, k_t)' \overset{\varphi}{\mapsto} \left(\frac{1}{r}\sum_{j=1}^{n_1} k_j, \cdots, \frac{1}{r}\sum_{j=n_{s-1}+1}^{n_s} k_j, \frac{1}{r}k_2, \cdots, \frac{1}{r}k_{n_1}, \frac{1}{r}k_{n_1+2}, \cdots, \frac{1}{r}k_t \right)' \tag{4.6.31a}$$

$$\mathbf{k} = (k_1, \cdots, k_t)' \overset{\psi}{\mapsto} \left(\frac{1}{s}\sum_{j=1}^{m_1} k_j, \cdots, \frac{1}{s}\sum_{j=m_{r-1}+1}^{m_r} k_j, \frac{1}{s}k_2, \cdots, \frac{1}{s}k_{m_1}, \frac{1}{s}k_{m_1+2}, \cdots, \frac{1}{s}k_t \right)' \tag{4.6.31b}$$

where $n_i = ir$ and $m_i = is$ (see Fig. 4.6.2). It is easy to show that φ and ψ are injective.

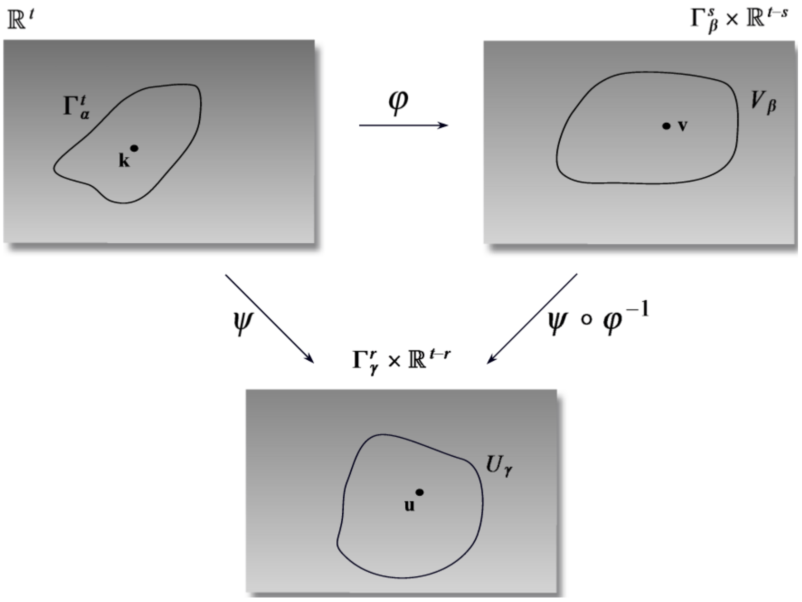

Figure 4.6.2: Schematic representation of the map $\varphi : \Gamma^t_\alpha \longrightarrow \Gamma^s_\beta \times R^{t-s}$ and $\psi : \Gamma^t_\alpha \longrightarrow \Gamma^r_\gamma \times R^{t-r}$.

That is for each element $\mathbf{v} \in V_\beta = \varphi\left(\Gamma^t_\alpha\right)$, there exists an unique element $\mathbf{k} \in \Gamma^t_\alpha$ such that $\mathbf{k} = \varphi^{-1}\left(\mathbf{v}\right)$. Similarly, for each element $\mathbf{u} \in U_\gamma = \psi\left(\Gamma^t_\alpha\right)$, there exits an unique element $\mathbf{k} \in \Gamma^t_\alpha$ such that $\mathbf{k} = \psi^{-1}\left(\mathbf{u}\right)$. Thus, the following map is well-defined,

$$\psi \circ \varphi^{-1} : V_\beta \longrightarrow U_\gamma \tag{4.6.32}$$

and easy to show that it is *bijective*[9]. In other words, there is a one-to-one correspondence between the elements in V_β and U_γ. Following similar arguments leading to (4.6.22), it can be shown that

$$p\left(\mathbf{u}|\,\gamma, I\right) = p\left(\mathbf{v}|\,\beta, I\right) \left|\det J_{\psi\circ\varphi^{-1}}\left(\mathbf{v}\right)\right|, \tag{4.6.33}$$

where $\mathbf{u} = \psi \circ \varphi^{-1}\left(\mathbf{v}\right)$ and $\det J_{\psi\circ\varphi^{-1}}\left(\mathbf{v}\right)$ denotes the Jacobian determinant of the map $\psi \circ \varphi^{-1}$.

From the multiplicative property of determinants (Strange 1988, p.217) and the chain

---

[9] A map $\phi : X \rightarrow Y$ is bijective if and only if $\phi\left(x_1\right) = \phi\left(x_2\right) \Rightarrow x_1 = x_2$ and for any element $y \in Y$, there exists an unique element $x \in X$ such that $\phi\left(x\right) = y$.

rule, one can show that

$$\left|\det J_{\psi\circ\varphi^{-1}}\left(\mathbf{v}\right)\right| = \left|\det J_{\psi}\left(\mathbf{k}\right)\right| \left|\det J_{\varphi^{-1}}\left(\mathbf{v}\right)\right| \tag{4.6.34}$$

where

$$\left|\det J_{\varphi}\left(\mathbf{k}\right)\right| \left|\det J_{\varphi^{-1}}\left(\mathbf{v}\right)\right| = 1. \tag{4.6.35}$$

This implies that

$$\left|\det J_{\psi\circ\varphi^{-1}}\left(\mathbf{v}\right)\right| = \frac{\left|\det J_{\psi}\left(\mathbf{k}\right)\right|}{\left|\det J_{\varphi}\left(\mathbf{k}\right)\right|}. \tag{4.6.36}$$

Using Eqn. (4.6.24), follows that

$$\left|\det J_{\psi\circ\varphi^{-1}}\left(\mathbf{v}\right)\right| = \left(\frac{r}{s}\right)^{t}, \tag{4.6.37}$$

and hence

$$p\left(\mathbf{u}\middle|\, \gamma, I\right) = \left(\frac{r}{s}\right)^{t} p\left(\mathbf{v}\middle|\, \beta, I\right), \tag{4.6.38}$$

where $\mathbf{u} = \psi\circ\varphi^{-1}\left(\mathbf{v}\right)$ and $t = rs$. Then by marginalization,

$$p\left(\mathbf{g}\middle|\, \gamma, I\right) = \int_{U_{\gamma}\cap\mathbb{R}^{t-r}} p\left(\mathbf{u}\middle|\, \gamma, I\right) d\mathbf{u} = \left(\frac{r}{s}\right)^{t} \int_{\varphi\circ\psi^{-1}\left(U_{\gamma}\cap\mathbb{R}^{t-r}\right)} p\left(\mathbf{v}\middle|\, \beta, I\right) d\mathbf{v}. \tag{4.6.39}$$

Note that if Ω_{γ}^{r} and Ω_{β}^{s} are not relatively prime, one can easily deduce Eqn. (4.6.26) from Eqn. (4.6.39). Similarly, in the case of relatively prime ensembles, it can be shown that

$$p\left(\mathbf{g}\middle|\, I\right) = \int_{U\cap\mathbb{R}^{t-r}} p\left(\mathbf{u}\middle|\, I\right) d\mathbf{u} = \left(\frac{r}{s}\right)^{t} \int_{\varphi\circ\psi^{-1}\left(U\cap\mathbb{R}^{t-r}\right)} p\left(\mathbf{v}\middle|\, I\right) d\mathbf{v}, \tag{4.6.40}$$

where $\mathbf{u} = \psi\circ\varphi^{-1}\left(\mathbf{v}\right)$, $t = rs$, $\varphi\left(\Gamma^{t}\right) = V$ and $\psi\left(\Gamma^{t}\right) = U$. It can also be shown that Eqn. (4.6.28) follows from Eqn. (4.6.40) subject to Ω^{r} and Ω^{s} not being relatively prime.

4.6.5 Fundamental Theorems of Sampling

As it was shown, Eqns. (4.6.39) and (4.6.40) summarize the relation between the pdfs of sample spaces and ensembles, respectively.

Theorem 4.1 (Fundamental theorem of sample spaces) *Suppose there exists shape indexes* α, β *and* γ *such that* $\Omega_\beta^s \leq \Omega_\alpha^t$, $\Omega_\gamma^r \leq \Omega_\alpha^t$ *and* $t = sr$ *with* $r < s$*. Furthermore, let* $\mathbf{k}$, $\mathbf{h}$ *and* $\mathbf{g}$ *be the heterogeneity functions defined on* Ω_α^t, Ω_β^s *and* Ω_γ^r*, respectively. Then*

$$p\left(\mathbf{g}|\,\gamma, I\right) = \int_{U\cap\mathbb{R}^{t-r}} p\left(\mathbf{u}|\,\gamma, I\right) d\mathbf{u} = \left(\frac{r}{s}\right)^t \int_{\varphi\circ\psi^{-1}(U\cap\mathbb{R}^{t-r})} p\left(\mathbf{v}|\,\beta, I\right) d\mathbf{v}, \qquad (4.6.41)$$

where the maps φ *and* ψ *are defined by (4.6.31). If* Ω_β^s *and* Ω_γ^r *are non-relatively prime sample spaces, then in Eqn. (4.6.41)* $\varphi = \mathbf{1}$*, i.e., the identity map,* $\Gamma_\alpha^t = \Gamma_\beta^s$ *and* $r/s = q^{-1}$ *in which* q *is an positive integer.*

Theorem 4.2 (Fundamental theorem of ensembles) *Suppose that* $\Omega^s \leq \Omega^t$, $\Omega^r \leq \Omega^t$ *and* $t = sr$ *with* $r < s$*. Furthermore, let* $\mathbf{k}$, $\mathbf{h}$ *and* $\mathbf{g}$ *be the heterogeneity functions defined on* Ω^t, Ω^s *and* Ω^r*, respectively. Then*

$$p\left(\mathbf{g}|\,I\right) = \int_{U\cap\mathbb{R}^{t-r}} p\left(\mathbf{u}|\,I\right) d\mathbf{u} = \left(\frac{r}{s}\right)^t \int_{\varphi\circ\psi^{-1}(U\cap\mathbb{R}^{t-r})} p\left(\mathbf{v}|\,I\right) d\mathbf{v}, \qquad (4.6.42)$$

where the maps φ *and* ψ *are defined by extension of the domain of (4.6.31) to the whole space* Γ^t*. If* Ω^s *and* Ω^r *are non-relatively prime sample spaces, then in Eqn. (4.6.42)* $\varphi = \mathbf{1}$*, i.e., the identity map,* $\Gamma^t = \Gamma^s$ *and* $r/s = q^{-1}$ *in which* q *is an positive integer.*

One consequence of the above Theorems is that the pdf of two sample spaces (ensembles), independent of their shape indexes and orders, can be related to each other by (4.6.41) and (4.6.42), under the condition that they both have a common refinement. This is a useful result which is very valuable in theoretical study of sample spaces (ensembles).

4.7 Epilogue: Comments

In the previous Section, we assumed that the mixing system under study is binary. This assumption was necessary in order to avoid mathematical complication. Indeed, this in one respect would have dragged us into the subject of *random matrices* (Mehta 2004). This is currently a very active field of research with range of applications covering a great number of subjects. However, there are still many issues remain to be resolved before it can be applied to the problems we are dealing with in the present work. Nevertheless, there is another less elegant approach to modelling of multicomponent mixtures. Indeed, this can be done by stacking each column of the heterogeneity matrix on another. This operation transforms a $n \times k$ heterogeneity matrix into a column vector consisting of nk entries. Then similar mathematical approach as above can be applied to model multicomponent mixtures. It remains to be seen which approach is simpler and more effective.

As it was mentioned previously, the knowledge about the range of the heterogeneity functions, due to insufficient information about all the system parameters, is limited. This results in a sort of "fuzziness" of assumed range of heterogeneity functions. That is, if an element $\mathbf{h} \in \Gamma^n_\alpha \subseteq \mathbb{R}^n$, then a subset of $\mathbb{R}^n$ containing $\mathbf{h}$ also belongs to Γ^n_α. Mathematically this fuzziness property can be thought of as instead of working with points in the space, one is forced to work with so-called open sets. In this context, it seems that an appropriate framework can be founded on the subject of *general topology*. This subject can be divided into two broad areas. The first, which could be called *continuous topology*, centres on the concepts like *compactness* and *metrization* which are the indispensable tools of modern mathematics. The second area, which might be called *geometric topology*, is primarily concerned with the *connectivity* properties of topological spaces (see e.g. Willard 1970). It appears that the first attempt to develop such a theory resulted into creation of *random set topology* by Matheron (1975). However, despite its success in Geostatistics and other related subjects, it seems that adaptation of frequency interpretation of the probability theory have had a limiting effect on its applicability. Nevertheless, the subject of random set topology offers an alternative approach which is certainly worth some attention.

Chapter 5

Application of the Sampling Theory

> The whole is more than the sum of the parts.
>
> Aristotle, Metaphysica (350 BC)

The main objective of this chapter is to demonstrate some of the established facts about sampling and mixing of particulate mixtures. Lacey's conjecture is stated and is shown that it is mathematically consistent within the mathematical framework of sampling of particulate materials, developed in Chapter 4. One major consequences of this analysis is stated in the mixture heterogeneity theorem. Moreover, the modelling techniques of Chapter 4 is used in developing a new model which can be applied to estimate the number of tracer particles of given size in random binary closed batch systems. It is further shown how the same techniques can be used in determining the RTD of the binary closed continuous systems. This is further deepen by applying these methods to construct a model for RTD of commercial twin screw extruder. Also a short discussion on the concept of Constitutional heterogeneity and its relation to the topics discussed in the present chapter is given.

5.1 Lacey's Conjecture

Lacey (1943) developed a model for mixtures based on the assumption that the constituent particles are identical and only differ in colour. He demonstrated that, in these cases, for a completely random mixture the variance of the sample composition decays inversely with the sample size. He, although in different wordings, also made the following conjecture.

Conjecture 5.1 (Lacey) *The mixture heterogeneity increases as the sample size decreases, independent of the distribution of the constituents and sample shape.*

Poux *et al.* (1991) pointed out that this conjecture is reasonable since every mixture can be made homogeneous provided that samples selected for analysis are large enough. The experimental studies by Poole *et al.* (1964) indicated that in the case of non-random mixtures that they studied, there was a correlation between the dependence of the heterogeneity on sample size and the degree of randomness attained in the mixture. Nonetheless, they could also confirm that the mixture heterogeneity decreases as the sample size increases, although, not necessarily as in the case of Lacey's ideal mixture.

Several workers have suggested that the correlation between the sample size and degree of mixedness is due to the shape of the samples. Indeed, Bourne (1967; 1968) showed that the sample variance-size relationship in two dimensional samples, is influenced by the type of correlation inherent in the mixture and the type of sample shape. Cooke and Bridgwater (1977) extended Bourne's results and determined general expressions for one, two and three dimensional sample shape-variance relation. However, beside empirical evidence, they could not produce any convincing rational for their choice of correlation coefficient.

It is possible to calculate the correlation coefficient from a experimentally determined curve of sample variance vs. sample size. However, as Scott and Bridgewater (1974) have pointed out, the correlation coefficient is actually related to the rate of change of the variance-sample size curve. It has been for long known that large experimental errors may mask the mixture variance (Orr 1979). For example, Lai and Hersey (1981) confirmed this point experimentally for at least several types of micro-dose-mixing. The effect of these errors on the derivative of a variance-sample size curve is quite large. Thus as Tucker (1981) has also pointed out, it has been convenient to assume a certain shape for the curve describing the correlation coefficient as a function of distance, which is also known as *correlogram*, and use sample variance measurements to determine its validity. Although this method has its advantages, there appears to be no other reason for preferring any particular correlogram than empirical evidence. Moreover, because of considerable experimental errors in sample variance which come from using a finite number of samples, any choice of correlogram which is solely based on empirical evidences is doubtful.

Tucker (1981) realizing the weakness of the above approach, suggested a method of choosing the correlogram function based on physical consideration. He showed that the variance in concentration of the key component in the mixture is determined by the integral of the product of the mixture correlation function and a function depending on the sample shape. This sample shape function can only be evaluated analytically for simple shapes and for more complicated shapes is determined numerically. Further, he argued that for small distances, the sample shape function can be approximated by the surface area to volume ratio of the sample. Using this and the correlogram function determined by physical considerations, it is possible to show the truth of Lacey's conjecture. However, this is only possible if the correlation function is non-zero for small distances in which the aforementioned approximation for the shape function is valid. Therefore, for the more general mixtures, the status of the conjecture remains undetermined.

Gy (1979, Ch.19), based on some logical arguments, did get closest in theoretically demonstrating the truth of Lacey's conjecture. But in proving the conjecture, he assumed that $p\left(\mathbf{h}|\, I\right)$ is a normal distribution. Although, based on experience, he argued that this assumption is not necessary, however, he never proved the redundancy of the normality assumption. In fact as Ashton and Valentin (1966) have pointed out, under conditions usually encountered in practice, the distribution of mixture ingredients and hence $p\left(\mathbf{h}|\, I\right)$, often deviates from the normal distribution. That is, the third or higher moments of the distribution are non-zero. Nevertheless, Pearson (1922) has shown that in the absence of a logical relation, because of sampling and analytical errors, the determination of moments higher than the second is precluded, since several hundred samples would be required for their accurate determination.

The above discussion demonstrates that despite strong logical and empirical evidences, it appears that there is no model for mixture heterogeneity that could confirm the truth of the Lacey's conjecture. Indeed, this is the first test which any model should pass in order to be taken seriously. In the following, it will be shown that the model which was developed for mixture heterogeneity in Section 4.6, does indeed confirm the truth of Lacey's conjecture. Since this heterogeneity model is identical to Gy's model then in a sense it proves also that Gy's conclusion was right.

5.1.1 Proof of Lacey's Conjecture for Non-relatively Prime Sample Spaces and Ensembles

As it was explained earlier, the volume of the samples decreases as the order of the sample space increases. Thus sample space of higher order correspond to samples of smaller volume. Now, for the sake of clarity, Lacey's conjecture is first proved for Ω_α^r and Ω_β^s such that $\Omega_\alpha^r \leq \Omega_\beta^s$. In the next Sub-Section, it will be shown how it can be proven for relatively prime spaces.

From Definition (4.4) and Theorem (4.1), it follows that

$$\begin{aligned}\sigma^2\left(\mathbf{g}\right) &= \frac{1}{r}\sum\nolimits_{i=1}^{r}\int_{\Gamma_\alpha^r} g_i^2 p\left(\mathbf{g}|\,\alpha, I\right)d\mathbf{g} \\ &= \frac{1}{r}\sum\nolimits_{i=1}^{r}\int_{V_\alpha}\left(v^i\right)^2 f\left(\mathbf{v}|\,\alpha, I\right)d\mathbf{v} \\ &= \frac{1}{rq^s}\sum\nolimits_{i=1}^{r}\int_{\Psi^{-1}(V_\alpha)}\left(q^{-1}\sum\nolimits_{j=k_{i-1}+1}^{k_i} h_j\right)^2 p\left(\mathbf{h}|\,\beta, I\right)d\mathbf{h},\end{aligned} \tag{5.1.1}$$

where all the parameters are same as in (4.6.26). Moreover, notice that $\Psi^{-1}\left(V_\alpha\right) = \Gamma_\beta^s$, and hence

$$\begin{aligned}\sigma^2(\mathbf{g}) &= \frac{1}{rq^{s+2}}\sum\nolimits_{i=1}^{r}\int_{\Gamma_\beta^s}\left(\sum\nolimits_{j=k_{i-1}+1}^{k_i} h_j\right)^2 p(\mathbf{h}|\,\beta, I)\,d\mathbf{h} \qquad (5.1.2)\\ &= \frac{q^{-(s+1)}}{s}\sum\nolimits_{i=1}^{r}\left\langle\left(\sum\nolimits_{j=k_{i-1}+1}^{k_i} h_j\right)^2\right\rangle.\end{aligned}$$

Now, by Schwarz inequality (Arfken 2001, p.607), it follows that

$$\left(\sum\nolimits_{j=k_{i-1}+1}^{k_i} h_j\right)^2 \le (k_i - k_{i-1})\sum\nolimits_{j=k_{i-1}+1}^{k_i} h_j^2 = q\sum\nolimits_{j=k_{i-1}+1}^{k_i} h_j^2, \qquad (5.1.3)$$

which implies that

$$\sigma^2(\mathbf{g}) \le \frac{q^{-(s+1)}}{s}\sum\nolimits_{i=1}^{r}\left\langle q\sum\nolimits_{j=k_{i-1}+1}^{k_i} h_j^2\right\rangle. \qquad (5.1.4)$$

Moreover, it is evident that

$$\frac{1}{s}\sum\nolimits_{j=k_{i-1}+1}^{k_i}\left\langle h_j^2\right\rangle = \sigma^2(\mathbf{h}) - \frac{1}{s}\left(\sum\nolimits_{j=1}^{k_{i-1}}\left\langle h_j^2\right\rangle + \sum\nolimits_{j=k_i+1}^{s}\left\langle h_j^2\right\rangle\right), \qquad (5.1.5)$$

and hence

$$\begin{aligned}\sigma^2(\mathbf{g}) &\le q^{-(s+1)}\sum\nolimits_{i=1}^{r} q\left[\sigma^2(\mathbf{h}) - \frac{1}{s}\left(\sum\nolimits_{j=1}^{k_{i-1}}\left\langle h_j^2\right\rangle + \sum\nolimits_{j=k_i+1}^{s}\left\langle h_j^2\right\rangle\right)\right] \qquad (5.1.6a)\\ &\le q^{-(s+1)}\left[s\sigma^2(\mathbf{h}) - \frac{1}{s}\sum\nolimits_{i=1}^{r}\left(\sum\nolimits_{j=1}^{k_{i-1}}\left\langle h_j^2\right\rangle + \sum\nolimits_{j=k_i+1}^{s}\left\langle h_j^2\right\rangle\right)\right]. \qquad (5.1.6b)\end{aligned}$$

It is easy to show that

$$\sum\nolimits_{i=1}^{r}\left(\sum\nolimits_{j=1}^{k_{i-1}}\left\langle h_j^2\right\rangle + \sum\nolimits_{j=k_i+1}^{s}\left\langle h_j^2\right\rangle\right) = s(r-1)\,\sigma^2(\mathbf{h}), \qquad (5.1.7)$$

which if substituted back into the inequality (5.1.6b), yields

$$\sigma^2(\mathbf{g}) \le q^{-(s+1)}(s-r+1)\,\sigma^2(\mathbf{h}). \qquad (5.1.8)$$

Now, consider the function

$$m\left(s\right)=\left(\frac{r}{s}\right)^{s+1}\left(s-r+1\right). \tag{5.1.9}$$

The finite Taylor's expansion of $m\left(s\right)$ with respect to s around r is (Arfken 2001, p.334)

$$m\left(s\right)=m\left(r\right)+\Delta\left.\frac{dm}{ds}\right|_{s=r}+\cdots+\frac{\Delta^{k-1}}{\left(k-1\right)!}\left.\frac{d^{(k-1)}m}{ds^{(k-1)}}\right|_{s=r}+R_k, \tag{5.1.10}$$

where

$$R_k=\frac{\Delta^k}{k!}\left.\frac{d^k m}{ds^k}\right|_{s=r+\epsilon\Delta}, \text{ for some } \epsilon\in[0,1] \tag{5.1.11}$$

is the Lagrange remainder of order k and $\Delta=s-r$. It is easy to show that for $k=1$,

$$m\left(s\right)=1-\Delta\left(\frac{r}{r+\epsilon\Delta}\right)\left[\frac{r}{r+\epsilon\Delta}\ln\left(\frac{r+\epsilon\Delta}{r}\right)+\frac{r\left(1+r+\epsilon\Delta\right)}{\left(r+\epsilon\Delta\right)}\right], \tag{5.1.12}$$

which evidently implies that $m\left(s\right)\leq 1$ and hence by (5.1.8), it follows that,

$$\sigma^2\left(\mathbf{g}\right)\leq\sigma^2\left(\mathbf{h}\right). \tag{5.1.13}$$

In other words, Lacey's conjecture is true in the cases where $\Omega_\alpha^r\leq\Omega_\beta^s$ for some α and β. By Theorem (4.2), the above result can be extended to ensembles. This means that Lacey's conjecture, in the case of the non-relatively prime ensembles, is also true independent of sample shape.

5.1.2 Proof of Lacey's Conjecture for Relatively Prime Sample Spaces and Ensembles

In order to prove Lacey's Conjecture, one has to show the truth of the conjecture independent of order of sample spaces or the shape of the samples. Accordingly, let $\Omega_\gamma^r\leq\Omega_\alpha^t$ and $\Omega_\beta^s\leq\Omega_\alpha^t$ for some shape indexes α,β and γ such that $t=rs$, for $1<r<s$ and r and

s being relatively prime. Furthermore, let $\mathbf{k}$, $\mathbf{h}$ and $\mathbf{g}$ denote the heterogeneity functions on Ω_α^t, Ω_β^s and Ω_γ^r, respectively. Then by Definition (4.4) and Theorem (4.1), it follows

$$\begin{aligned}
\sigma^2(\mathbf{g}) &= \frac{1}{r}\sum\nolimits_{i=1}^{r}\int_{\Gamma_\gamma^r} g_i^2 p(\mathbf{g}|\,\gamma, I)\,d\mathbf{g} && \text{(5.1.14a)}\\
&= \frac{1}{r}\sum\nolimits_{i=1}^{r}\int_{U_\gamma} \left([\mathbf{u}_r]_i\right)^2 p(\mathbf{u}|\,\gamma, I)\,d\mathbf{u} && \text{(5.1.14b)}\\
&= \frac{1}{r}\left(\frac{r}{s}\right)^t\sum\nolimits_{i=1}^{r}\int_{V_\beta} \left(\left[\psi\circ\varphi^{-1}(\mathbf{v}_r)\right]_i\right)^2 p(\mathbf{v}|\,\beta, I)\,d\mathbf{v}, && \text{(5.1.14c)}
\end{aligned}$$

where, since $\psi\circ\varphi^{-1}$ is a bijective map, $\varphi\circ\psi^{-1}(U_\gamma) = \varphi(\Gamma_\alpha^t) = V_\beta$ and $\mathbf{u}_r$ denotes

$$\psi\circ\varphi^{-1}(\mathbf{v}_r) = \mathbf{u}_r = (g_1,\ldots,g_r,0,\ldots,0)\in U_\gamma \tag{5.1.15}$$

Now, from (4.6.31) follows that

$$\psi^{-1}(\mathbf{u}_r) = \left(\underbrace{sg_1,0,\ldots,0}_{s},\underbrace{sg_2,0,\ldots,0}_{s},\ldots,\underbrace{sg_r,0,\ldots,0}_{s}\right). \tag{5.1.16}$$

Since r and s are relatively prime, it is easy to show that

$$\left[\varphi\circ\psi^{-1}(\mathbf{u}_r)\right]_i = v_i = \begin{cases} \frac{s}{r}g_j & \text{for an unique } j\in\{1,\ldots,r\}\\ \text{or} & \\ 0 & \end{cases} \tag{5.1.17}$$

Thus

$$\sigma^2(\mathbf{g}) \le \frac{1}{r}\left(\frac{r}{s}\right)^t\sum\nolimits_{i=1}^{r}\int_{V_\beta}\left(\frac{r}{s}v_i\right)^2 p(\mathbf{v}|\,\beta, I)\,d\mathbf{v}, \tag{5.1.18}$$

which implies

$$\begin{aligned}
\sigma^2(\mathbf{g}) &\leq \left(\frac{r}{s}\right)^{t+1} \frac{1}{s} \sum_{i=1}^{r} \int_{V_\beta} (v_i)^2 \, p(\mathbf{v}|\beta, I)\, d\mathbf{v} && \text{(5.1.19a)} \\
&\leq \left(\frac{r}{s}\right)^{t+1} \frac{1}{s} \sum_{i=1}^{s} \int_{V_\beta} (v_i)^2 \, p(\mathbf{v}|\beta, I)\, d\mathbf{v} && \text{(5.1.19b)} \\
&\leq \left(\frac{r}{s}\right)^{t+1} \sigma^2(\mathbf{h}) && \text{(5.1.19c)}
\end{aligned}$$

and since $1 < r < s < t$, then

$$\sigma^2(\mathbf{g}) \leq \sigma^2(\mathbf{h})\,. \tag{5.1.20}$$

The above result can similarly be extended to the case of relatively prime ensembles. In other words, Lacey's conjecture is proved for relatively prime ensembles.

5.1.3 Fundamental Theorem of Mixture Heterogeneity

The above results can now be summarized as follows.

Theorem 5.1 (Fundamental theorem of mixture heterogeneity) *Let* Ω_α^r *and* Ω_β^s *be two sample spaces for some* α *and* β *such that* $r < s$. *Furthermore, let* $\mathbf{g}$ *and* $\mathbf{h}$ *denote the heterogeneity functions on* Ω_α^r *and* Ω_β^s, *respectively. Then* $\sigma^2(\mathbf{g}) \leq \sigma^2(\mathbf{h})$. *This is also true in the case of ensembles.*

Corollary 5.1 *The mixture heterogeneity increases as the sample size decreases, independent of the distribution of the constituents and sample shape.*

Proof. *This is a direct consequence of Theorem (5.1).* ■

The above results show that Lacey's conjecture is a direct consequence of how one conceives the concept of heterogeneity and no additional assumptions were necessary in order to show its truth. In fact the proof is more general. Indeed, it is easy to see that Theorem (5.1) is valid for any model of heterogeneity that possesses the same property as was described in Proposition (4.3). In other words, this property is more fundamental to the notion of heterogeneity than the form of its representation in algebraic form. This is also consistent with the condition that one's conclusions should be independent of choice of the algebraic form of the heterogeneity function, as long as it satisfies certain conditions.

Another consequence of Theorem (5.1), in the case of closed systems, is that the following series

$$0 = \sigma_1^2 (\mathbf{h}) \leq \cdots \leq \sigma_i^2 (\mathbf{h}) \leq \cdots \leq \sigma_N^2 (\mathbf{h}) , \tag{5.1.21}$$

where σ_i^2 and N denote the mixture heterogeneity over the ensemble Ω^i and number of particles in the mixture, respectively, make perfect sense. That is, if the order of the ensemble is one, i.e., the whole mixture consists of one sample which contains all the particles in the mixture, the mixture heterogeneity is nil. Physically, this means that by choosing the scale of scrutiny to be the whole mixture, one becomes blind to all the finer details in the mixture. On the other extreme, if the order is equal to the number of particles in the mixture, it is evident that mixing or segregation has no effect on the value of the mixture heterogeneity. That is, at this scale of scrutiny, the mixture looks as heterogeneous as it can be.

5.2 Interlude: Comments

In the above, it was demonstrated that if we assume the validity of the model for mixture heterogeneity developed in Chapter 4, then, mathematically, Lacey's conjecture has to be true. However, as Jeffreys (1973, p.11) put it, mathematics can only display connections between scientific statements; it does not prove the statement by itself. Indeed, the only thing which is achieved by proving the Lacey's conjecture is to demonstrate that our model is consistent with respect to empirical evidences.

As mentioned earlier, it was Gy who first proposed the model for heterogeneity which is also used in this work. However, Gy could not produce convincing theoretical evidence in favour of Lacey's conjecture. In fact, he could only partially demonstrate its plausibility by assuming that the sample distribution is Normal. As long as the mixture components are randomly distributed, i.e., the mixture heterogeneity is at its minimum or equivalently the sample Reproducibility is at its maximum, this assumption seems to be reasonable. But if the mixture is partially mixed, as mentioned previously, the correlation among the samples and their respective shape affect the level of heterogeneity. Bretthorst (1996) points out that if a pdf does not contain any information on correlations, it would be more correct to say that this pdf makes allowances for every possible correlation. Evidently, in the case of partially mixed mixtures, this does not apply and proper constraints on the correlations or higher moments should be taken into consideration. In Gy's modelling approach one assumes that the samples are reproducible. As was explained previously, this assumption, in general, is not valid in particulate mixing.

5.3 An Estimate for the Number of Tracer Particles of Given Size in Random Binary Closed Batch Systems

In the following, the modelling technique of Chapter 4 is applied to estimate the number of tracer particles of different sizes in a random closed binary batch system. In this context, by the phrase "random" one means that the mixture is at the state of minimum heterogeneity.

As was argued previously, the assignment of the probability depends on the amount of testable information about the mixture. This information is usually expressed in the form of physical constraints which restricts the degrees of freedom of the mixture parameters. Since, in the present case, the system is a closed batch system, two obvious constraints yield. The first one is the conservation of the number of tracer particles in the mixture and the second one is the conservation of the mass of the tracer particles in the mixture.

Now, let $p\left(r, j|\, i, I\right)$ be the probability of finding r tracer particles of size class j in a given sample i. Then by the product rule (3.2.1),

$$p\left(r, j \middle| i, I\right) = p\left(j \middle| i, I\right) p\left(r \middle| i, j, I\right). \tag{5.3.1}$$

The terms on the right hand side of the above equation can be interpreted as follows. The first term $p\left(j \middle| i, I\right)$ is probability of finding size class j in a given sample i. Since the samples are equivolume and the state of minimum heterogeneity is assumed, $p\left(j \middle| i, I\right)$ is independent of choice of particular i and therefore contains no additional information which is not already in I. Thus

$$p\left(j \middle| i, I\right) = p\left(j \middle| I\right) = q_j, \tag{5.3.2}$$

where q_j is the probability of finding a tracer particle of size class j in a randomly chosen sample and is assumed to be known[1]. For the sake of simplicity, it is assumed that q_j's represent the distribution of the particle volumes and therefore by size of a particle one means its volume. The second term $p\left(r \middle| i, j, I\right)$, is the probability of finding r particles of given size class j in a given sample i. Again, since the samples are equivolume and the mixture is at the state of minimum heterogeneity, $p\left(r \middle| i, j, I\right)$ is independent of choice of i and hence

$$p\left(r \middle| i, j, I\right) = p\left(r \middle| j, I\right) = p_{rj}, \tag{5.3.3}$$

where p_{rj} is the probability of finding r tracer particles of given size class j in a randomly chosen sample. Thus

$$p\left(r, j \middle| i, I\right) = q_j p_{rj}. \tag{5.3.4}$$

Now, if one assumes that the number of tracer particles of given size class j in any sample

---

[1] If there is no information on the size distribution of the tracer particles, uniform distribution should be assumed (see Section 3.4).

can be somewhere between 0 and s_j, then

$$\sum_{r=0}^{s_j} p_{rj} = 1, \text{ for each given } j = 1, \ldots, k \tag{5.3.5}$$

which implies

$$\sum_{j=1}^{k} \sum_{r=0}^{s_j} p\left(r, j \middle| i, I\right) = \sum_{j=1}^{k} q_j \sum_{r=0}^{s_j} p_{rj} = \sum_{j=1}^{k} q_j = 1. \tag{5.3.6}$$

The total entropy of the distribution is

$$-\sum_{j=1}^{k} \sum_{r=0}^{s_j} p\left(r, j \middle| i, I\right) \ln \left[p\left(r, j \middle| i, I\right)\right] = -\sum_{j=1}^{k} q_j \ln q_j - \sum_{j=1}^{k} q_j \sum_{r=0}^{s_j} p_{rj} \ln p_{rj}. \tag{5.3.7}$$

As it was shown in Chapter 4, the most probable distribution is the one which maximizes the above entropy functional subject to appropriate constraints. Since q_j's are fixed, this implies that the best candidate for p_{rj} is the one which maximizes

$$-\sum_{j=1}^{k} q_j \sum_{r=0}^{s_j} p_{rj} \ln p_{rj}. \tag{5.3.8}$$

Now, in the probabilistic framework, the physical constraints affect the expected values of parameters. Therefore, the conservation of the number of tracer particles in the mixture is expressed as

$$n \sum_{j=1}^{k} q_j \sum_{r=0}^{s_j} r p_{rj} = N \tag{5.3.9}$$

and similarly the conservation of the mass of the tracer particles in the mixture as

$$n\rho \sum_{j=1}^{k} v_j q_j \sum_{r=0}^{s_j} r p_{rj} = c_L M_L, \tag{5.3.10}$$

where v_j denotes the volume of a particle of size class j, $c_L M_L$ the total mass of the tracer particles in the mixture of total mass of M_L and ρ, the density of the material that the tracer particles are composed of. Thus, the Lagrangian in this case is

$$\begin{aligned} L = &-\sum_{j=1}^{k} q_j \sum_{r=0}^{s_j} p_{rj} \ln p_{rj} - \left(\sum_{j=1}^{k} \kappa_j - q_j \right) \left(\sum_{r=0}^{s_j} p_{rj} - 1 \right) \\ &- \lambda \left(\sum_{j=1}^{k} q_j \sum_{r=0}^{s_j} r p_{rj} - \frac{N}{n} \right) - \mu \left(\rho \sum_{j=1}^{k} v_j q_j \sum_{r=0}^{s_j} r p_{rj} - \frac{c_L M_L}{n} \right), \end{aligned} \tag{5.3.11}$$

where κ_j, λ and μ are the Lagrange multipliers. Now, maximizing L for variations in p_{rj}'s, one gets

$$\frac{\partial L}{\partial p_{ij}} = 0 \Rightarrow -q_j \left(1 + \ln p_{rj}\right) - \left(\kappa_j - q_j\right) - \lambda r q_j - \mu \rho r v_j q_j = 0, \tag{5.3.12}$$

which implies

$$p_{rj} = a_j \exp\left[-r\left(\lambda + \mu \rho v_j\right)\right]. \tag{5.3.13}$$

Using (5.3.5) and observing that it is the sum of a Geometric progression of s_j terms,

$$1 = \sum_{r=0}^{s_j} a_j \exp\left[-r\left(\lambda + \mu\rho v_j\right)\right] = a_j \frac{1 - \exp\left[-\left(s_j + 1\right)\left(\lambda + \mu\rho v_j\right)\right]}{1 - \exp\left[-r\left(\lambda + \mu\rho v_j\right)\right]}, \tag{5.3.14}$$

which implies that

$$a_j = \frac{1 - \exp\left[-r\left(\lambda + \mu\rho v_j\right)\right]}{1 - \exp\left[-\left(s_j + 1\right)\left(\lambda + \mu\rho v_j\right)\right]}. \tag{5.3.15}$$

The parameters λ and μ can be determined by using the remaining constraints. Indeed, Eqn. (5.3.9) implies that

$$\frac{N}{n} = \sum_{j=1}^{k} q_j \sum_{r=0}^{s_j} r p_{rj} = \sum_{j=1}^{k} q_j \bar{N}_j, \tag{5.3.16}$$

where $\bar{N}_j$ is the expected number of particles of size class j in a randomly chosen sample. But

$$\bar{N}_j = \sum_{r=0}^{s_j} r p_{rj} = \frac{1 - z_j}{1 - z_j^{s_j+1}} \sum_{r=0}^{s_j} r z_j^r, \tag{5.3.17}$$

in which $z_j = \exp\left[-\left(\lambda + \mu\rho v_j\right)\right]$. Since the last term in the above equation is an Arithmetic-Geometric progression (see e.g. Gradshteyn and Ryzhik 2000, p.1), one gets

$$\sum_{r=0}^{s_j} r z_j^r = \frac{z_j}{\left(1 - z_j\right)^2}\left(1 - \left(s_j + 1\right) z_j^{s_j} + s_j z_j^{s_j+1}\right), \tag{5.3.18}$$

and hence

$$\bar{N}_j = \frac{z_j}{1 - z_j}\,\frac{1 - \left(s_j + 1\right) z_j^{s_j} + s_j z_j^{s_j+1}}{1 - z_j^{s_j+1}}, \tag{5.3.19}$$

where $z_j \neq 1$ and $s_j > 0$ are assumed. Hence, the Lagrange multipliers λ and μ can uniquely be determined by solving the following system of equations

$$\frac{N}{n} = \sum_{j=1}^{k} q_j \frac{z_j}{1 - z_j}\,\frac{1 - \left(s_j + 1\right) z_j^{s_j} + s_j z_j^{s_j+1}}{1 - z_j^{s_j+1}} \tag{5.3.20a}$$

$$\frac{c_L M_L}{n\rho} = \sum_{j=1}^{k} v_j q_j \frac{z_j}{1 - z_j}\,\frac{1 - \left(s_j + 1\right) z_j^{s_j} + s_j z_j^{s_j+1}}{1 - z_j^{s_j+1}} \tag{5.3.20b}$$

Thus, the expression in (5.3.19) describes the expected number of tracer particles of given size in a randomly chosen sample, subject to the conservation of number of tracer particles and their total mass given by (5.3.20).

The parameter s_j is the maximum possible number of tracer particles of size class j in a randomly chosen sample, which is obviously application dependent. For example, s_j is depending on the ratio of sample volume to particle volume. This in turn is also related to the number of particles in the closest packing configuration of particles of size class j in the sample. For particles with irregular shapes there is, as yet, no consistent theory which can be used to determine this upper bound. Moreover, even if such theory existed, the shape of samples has also to be taken into account, which further complicate the calculations. Nevertheless, it is always possible to estimate s_j experimentally. In the following two distinct possibilities of great importance are considered which can function as limiting cases of the expression in (5.3.19).

5.3.1 Case I: $s_j = 1$

As it was argued in Chapter 4, with respect to size of particles in a mixture, no particles are exactly alike. That is, each possible size class in the interval between the minimum and maximum size of the tracer particles is represented by at most 1 particle and hence $s_j = 1$. Then by Eqn. (5.3.19), one gets

$$\bar{N}_j = \frac{z_j}{1+z_j} = \frac{1}{\exp\left(\lambda + \mu m_j\right) + 1}, \tag{5.3.21}$$

where $m_j = \rho v_j$ is the mass of a single particle of size class j. Since in most practical cases the number of tracer particles and hence particles in the mixture is quite large, in the cases in which the size distribution is narrow one can assume that it is continuous, i.e.,

$$\frac{1}{\rho}\int\limits_{m_{\min}}^{m_{\max}} q\left(m\right) dm = 1, \tag{5.3.22}$$

where $m_{\min}$ and $m_{\max}$ stand for the minimum and maximum mass of the tracer particles in the mixture, respectively. Hence, the constraints assume the following form,

$$\begin{aligned} \frac{N}{n} &= \frac{1}{\rho}\int\limits_{m_{\min}}^{m_{\max}} \frac{q\left(m\right)}{\exp\left(\lambda+\mu m\right)+1} dm \\ \frac{c_L M_L}{n} &= \int\limits_{m_{\min}}^{m_{\max}} \frac{m q\left(m\right)}{\exp\left(\lambda+\mu m\right)+1} dm. \end{aligned} \tag{5.3.23}$$

From Eqn. (5.3.21) it is clear that in order to have dimensional consistency, μ must be expressed in reciprocal mass units; that is, Kg^{-1}, g^{-1}, etc. Moreover, from (5.3.21) follows that

- $\mu = 0$: $\bar{N}_j$ is independent of the mass of the tracer particles
- $\mu > 0$: $\bar{N}_j$ decreases as m_j increases
- $\mu < 0$: $\bar{N}_j$ increases as m_j increases

Mixtures of particles of different sizes will, in general, have voids formed by the larger particles that are partially filled by the smaller ones (e.g. Furnas, 1931). In order to increase the packing density of the mixture, large number of fine particles are mixed with much smaller number of coarse particles. If the proportion of fines in the mixture is sufficiently high the mixture is known as being *fines continuous*. This, as Arteaga and Tüzün (1990) have also pointed out, depends on the size ratio of the smallest to largest particles and the weight fraction of fines. If the non-tracer component of the mixture consists of coarser particles then one needs to have $\mu > 0$ in order to achieve the closest packing. On the contrary, if the non-tracer component of the mixture is finer, then one needs to have $\mu < 0$. However, as Heywood (1961) have also noticed, although smaller particles can fit in the voids between larger ones, decreasing the voidage, fine particles tend to cohere, forming loose chains with high voidage, and thus nullifying the expected voidage decrease. In fact in some of these cases, as Heywood also discovered, if the mixture is *coarse continuous*, a greater packing density might be attained. In these cases the choice of μ should be opposite of the aforementioned cases.

Now, from (5.3.20) follows

$$\frac{c_L M_L}{N} = \frac{\sum_{j=1}^{k} q_j m_j \bar{N}_j}{\sum_{j=1}^{k} q_j \bar{N}_j} = \bar{m}, \tag{5.3.24}$$

where $\bar{m}$ is the weighted mean of $m_1, \ldots, m_k$ in which the weights being proportional to $q_j \bar{N}_j$. If $\mu = 0$, these weights are equal and $c_L M_L / N$ coincides with the mean particle mass of the tracer particles. This means that contrary to the above two cases, the size classes in the neighbourhood of the mean particle size are more dominantly represented in the mixture than the ones in the lower and upper tail of the tracer size distribution. Moreover, since in this case $\bar{N}_j$ is also independent of the mass of the tracer particles, then the size distribution of the tracer particles have to be uniformly distributed.

Based on the above discussion, it is reasonable to assume that μ depends on the microstructural properties of the mixture. Moreover, it is easy to show that

$$\frac{\partial}{\partial\mu}\left(\frac{c_L M_L}{n}\right)=\frac{\partial}{\partial\mu}\left(\sum_{j=1}^{k} q_j m_j \bar{N}_j\right)=-\sum_{j=1}^{k} q_j m_j^2 \frac{z_j}{(1+z_j)^2}<0 \tag{5.3.25}$$

and

$$\frac{\partial}{\partial\mu}\left(\frac{N}{n}\right)=\frac{\partial}{\partial\mu}\left(\sum_{j=1}^{k} q_j \bar{N}_j\right)=-\sum_{j=1}^{k} q_j m_j \frac{z_j}{(1+z_j)^2}<0, \tag{5.3.26}$$

that is, both $c_L M_L/n$ and N/n are monotonic decreasing functions of μ.

5.3.2 Case II: $s_j = \infty$

On the other extreme, If the size of particles is much smaller than the sample volume and if in addition it can be assumed that the mixture consists of a large number of particles then it is reasonable to assume that the number of particles belonging to each size class is somewhere between 0 and infinity, that is $s_j \to \infty$. Thus

$$\bar{N}_j=\frac{z_j}{1-z_j}=\frac{1}{\exp\left(\lambda+\mu m_j\right)-1}, \tag{5.3.27}$$

where $m_j=\rho v_j$ is the mass of a single particle of size class j. From the above expression it is evident that $\lambda+\mu m_j>0$. Moreover, a similar analysis as the above reveals that for all the pairs (λ,μ) that satisfy this constraint, $\bar{N}_j$ decreases as m_j increases, unless $\mu=0$ in which λ must be greater than zero and hence $\bar{N}_j$ is independent of the mass of the tracer particles. Both $c_L M_L/n$ and N/n are also monotonic decreasing functions of μ.

5.4 Interlude: Comments

In statistical mechanics the distributions in (5.3.21) and (5.3.27) are known as *Fermi-Dirac* and *Bose-Einstein*, respectively (e.g. Trevena 1993, Ch.5). The distribution in (5.3.19) is known as the *Intermediate Statistics*. It appears that this distribution was first proposed by Gentile (1940) in relation to physics of elementary particles in which he discussed the possibility of the existence of particle statistics intermediate between Fermi-Dirac and Bose-Einstein statistics. However, in the context of quantum mechanics, it appears to be no physical situation which follows the Intermediate Statistics (see e.g. Lavenda and Dunning-Davies 1989). Nonetheless, in the light of the above results, the mixture of particulate materials may serve as model for studying the properties of the Intermediate Statistics.

It seems to be that Jaynes (1957b-1957c) was the first investigator to suggest that these statistics can be derived from the maximum entropy principle (see also Chapter 3). Similar mathematical approach as the one in previous Section is also discussed by Forte and Sempi (1976), Kapur (1983) and Kapur (1994, Ch. 41) in relation to other problems in statistical mechanics and traffic theory.

5.5 Constitutional Heterogeneity

In the case of closed systems, the series (5.1.21) and Theorem (5.1) have another interesting application. Indeed, consider what will happen if the particles agglomerate in the mixture. This, in practice means that the number of distinguishable constituents are reduced to say N'. That is, in a sense, the effect of agglomeration on $\sigma_N^2(\mathbf{h})$, can be considered as an increase in the sample size and thus by Theorem (5.1),

$$\sigma_{N'}^2\left(\bar{\mathbf{h}}\right) \leq \sigma_N^2(\mathbf{h}). \tag{5.5.1}$$

In a similar manner, comminution increases $\sigma_N^2(\mathbf{h})$. For this reason, in the case of closed systems, $\sigma_N^2(\mathbf{h})$ is known as *Constitutional heterogeneity* of a mixture. Clearly, the Constitutional heterogeneity of a mixture is not affected by mixing or segregation. It is only changed by physically changing the size of the constituents. This fact was already known to Aristotle for more than two thousand years ago. Indeed, while he was studying the sand particles, he concluded his findings in a now famous sentence "the whole is more than the sum of the parts". It is impossible to determine the quality of the mixture by examining each of its constituents separately.

In the context of the sampling theory, Gy (1979, p.225) noticed these facts which also play an important part in his approach to sampling theory. However, in practice, the Constitutional heterogeneity of a mixture is difficult to determine. Since, as was mentioned earlier, any knowledge on the coarser sample space is insufficient to uniquely determine the structure of its refinements. Nonetheless, since the Constitutional heterogeneity is related to some physical concepts, it should be possible to estimate. Indeed, it seems to be that the distribution (5.3.21) can be used to estimate the Constitutional heterogeneity of a mixture. Further investigation into this matter is needed and is left to a later time.

5.6 Residence Time Distribution for Closed Binary Continuous Systems

As it was shown in § 4.3.2, in the case of closed binary continuous systems, the probability density for the system RTD is determined by maximizing

$$-\int_0^\infty f(t)\ln[f(t)]\,dt, \tag{5.6.1}$$

subject to the constraints

$$\int_0^\infty f(t)\,dt = 1 \text{ and } \int_0^\infty f(t)\ln t dt = \ln g. \tag{5.6.2}$$

Hence, in this case, the Lagrangian is

$$\begin{aligned} L = &-\int_0^\infty f(t)\ln[f(t)]\,dt - (\lambda_0 - 1)\left(\int_0^\infty f(t)\,dt - 1\right) \\ &- \lambda_1\left(\int_0^\infty f(t)\ln t dt - \ln g\right), \end{aligned} \tag{5.6.3}$$

where $\lambda_0 - 1$ is used as the first Lagrange multiplier instead of λ_0 as a matter of convenience. Now, maximizing L with respect to $f(t)$ is equivalent of solving the corresponding *Euler-Lagrange* equation[2] of the *Calculus of Variations*, which in this case implies that

$$-\ln\left[f(t)\right] - 1 - (\lambda_0 - 1) - \lambda_1 \ln t = 0, \tag{5.6.4}$$

and hence

$$f(t) = e^{-\lambda_0} t^{-\lambda_1}. \tag{5.6.5}$$

However, a closer look at the constraints revels that, in general, one can not uniquely determine the parameters λ_0 and λ_1. The reason for this is that $f(t)$ has a singularity at $t = 0$. This singularity can be avoided by changing the definition domain of the $f(t)$ to $[t_0, \infty]$, where $t_0 > 0$. Physically, this means that the residence time of any particle in the mixture is greater than zero, which is obviously true. If this condition is fulfilled, the parameters λ_0 and λ_1 can be uniquely determined and accordingly

$$f(t) = \mu t_0^{\mu} t^{-(\mu+1)}, \ t \geq t_0 > 0 \tag{5.6.6}$$

where

$$\mu^{-1} = \ln\left(\frac{g}{t_0}\right), \ g > t_0. \tag{5.6.7}$$

In literature, the distributions of type given by Eqn. (5.6.6) are known as the *Pareto distribution.*

From the above analysis it is clear that the choice of the parameters λ_0 and λ_1 depend on t_0. Since, in general, there is a great degree of freedom in choosing t_0; there would be an infinite family of Pareto distributions, all satisfying the above constraints. In other

---

[2] Let $L = \int_a^b F\left[t, f(t), f'(t)\right] dt$ where F is a known function. Then according to Euler-Lagrange equation, the function $f(t)$, which maximizes or minimizes L is given by

$$\frac{\partial F}{\partial f} - \frac{d}{dt}\left(\frac{\partial F}{\partial f'}\right) = 0$$

words, in order to achieve uniqueness, further constraints have to be imposed. Now, as it was discussed in § 2.4.3, the conservation of mass of the mixture imposes a constraint on the mean residence time of the system which has to be taken into account. That is, in this case the Lagrangian is

$$L = -\int_0^\infty f\left(t\right)\ln\left[f\left(t\right)\right]dt - \left(\lambda_0 - 1\right)\left(\int_0^\infty f\left(t\right)dt - 1\right) - \lambda_1\left(\int_0^\infty f\left(t\right)\ln t dt - \ln g\right) \\ - \lambda_2\left(\int_0^\infty t f\left(t\right)dt - \bar{t}\right). \tag{5.6.8}$$

Now, maximizing L with respect to $f\left(t\right)$ yields

$$f(t) = \begin{cases} \frac{(\alpha\mu)^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\alpha\mu t} \ , \ t > 0 \\ 0 \ , \ t \leq 0 \end{cases} \tag{5.6.9}$$

where $\mu^{-1} = \bar{t}$ and α is determined from

$$\ln g = \int_0^\infty \frac{(\alpha\mu)^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\alpha\mu t} \ln t dt. \tag{5.6.10}$$

The distribution in (5.6.9) is the well-known *gamma distribution*. Furthermore, it can be shown that (see e.g. Gradshteyn and Ryzhik 2000, p.572)

$$\ln g = \int_0^\infty \frac{(\alpha\mu)^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\alpha\mu t} \ln t dt = \frac{(\alpha\mu)^\alpha}{\Gamma(\alpha)} \frac{d}{d\alpha}\left(\frac{\Gamma(\alpha)}{(\alpha\mu)^\alpha}\right), \tag{5.6.11}$$

and hence

$$\ln g = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \ln\left(\alpha\mu\right) = \psi\left(\alpha\right) - \ln\left(\alpha\mu\right), \tag{5.6.12}$$

where ψ is the *Euler psi* function also known as *digamma* function (Gradshteyn and Ryzhik 2000, p.892). This implies that

$$g\mu = \frac{1}{\alpha}\exp\left[\psi\left(\alpha\right)\right]. \tag{5.6.13}$$

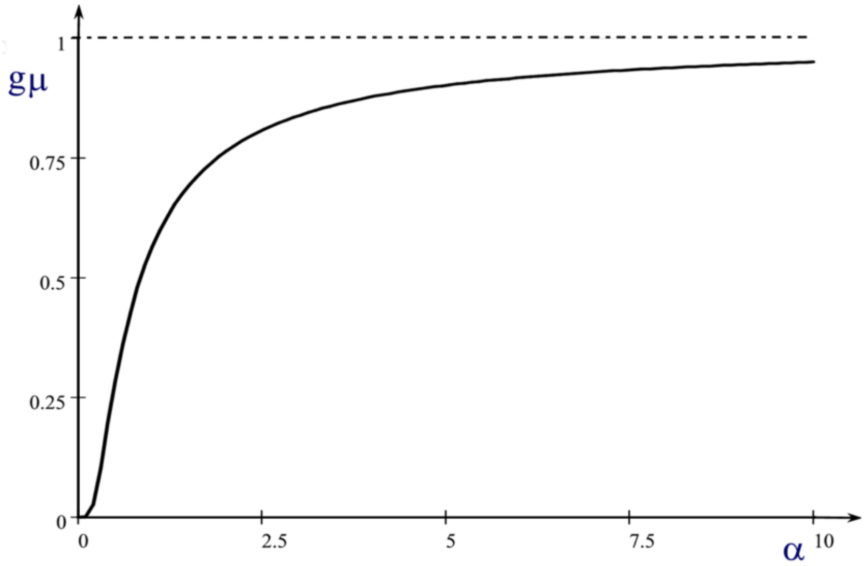

Figure 5.6.1: The relation between the ratio of the geometric mean to the mean residence time of a mixing vessel and its order.

Now, according to the modelling techniques of § 2.4.1, α is the number of exponential vessels in series. However, until now there has not been any constructive method for determining α. Nonetheless, Eqn. (5.6.13) can now be applied to determine the order of the mixer. Indeed, since the mean and geometric mean residence time of a mixer can experimentally be determined; Eqn. (5.6.13) can be solved for α. A plot of the order of the mixer versus the ratio of the geometric mean to the mean residence time is given in Fig. (5.6.1). It is easy to show that

$$\lim_{\alpha\to\infty} g\mu = 1. \tag{5.6.14}$$

Indeed, if $g = \bar{t}$, the distribution given by (5.6.9) is the *degenerate Dirac's delta* distribution centred at the common value of g and $\bar{t}$. This corresponds, as also mentioned in § 2.4.1, to the plug flow regime. On the other extreme, small values of μ imply that particles, on average, spend longer time in the mixer and small values of g would mean that the average axial fluctuations at the outlet of the mixer are small. Therefore, in general, small values of $g\mu$ correspond to better axial mixing as compared to larger ones. Thus, the ratio of the geometric mean to the mean residence time can be used as a measure for how capable a mixer is to disperse the incoming stream and hence a measure for axial mixing.

Now, suppose that the volumetric flow rate at the outlet of the mixer is constant. That is, the mixer smooths out all the incoming axial fluctuations. Then the geometric mean residence time does not impose any constraint on the RTD of the system and therefore the Lagrangian in this case would be

$$L = -\int_0^\infty f(t)\ln[f(t)]\,dt - (\lambda_0 - 1)\left(\int_0^\infty f(t)\,dt - 1\right) - \lambda\left(\int_0^\infty tf(t)\,dt - \bar{t}\right). \tag{5.6.15}$$

It can easily be shown that in this case maximizing L with respect to $f(t)$ would result into exponential distribution given by

$$f(t) = \begin{cases} \mu e^{-\mu t}\ ,\ t > 0 \\ 0\ ,\ t \leq 0 \end{cases} \tag{5.6.16}$$

which as mentioned previously, has been used as a model for an ideal mixer. Moreover, notice that for $\alpha = 1$, the distribution in (5.6.9) reduces also to the exponential distribution. Hence form this and Eqn. (5.6.13), it can be said that the exponential distribution is a special case of gamma distribution in which

$$g\mu = e^{-\gamma} = 0.561\,46, \tag{5.6.17}$$

where γ is the so called *Euler's constant* (Gradshteyn and Ryzhik 2000, p.xxxii). Thus, in the case of an ideal mixer the geometric mean of the system is uniquely determined by the processes which impose other constraints on the system. That is, in the above case, the geometric mean is solely determined by the same processes that restrict the mean residence time of the system. In other words, one can design a mixer which is very similar to an ideal mixer by choosing the mean residence time in such a way that Eqn. (5.6.17) is fulfilled. Moreover, since an ideal mixer is the best achievable mixer, for any given mixer

$$e^{-\gamma} = 0.561\,46 \leq g\mu \leq 1 \tag{5.6.18}$$

or equivalently

$$\alpha \geq 1. \tag{5.6.19}$$

The above analysis offers an insight into the dynamics of the system. As it was explained in § 2.4.1, the mixing behaviour of a majority of the actual mixers deviates from the ideal mixer (exponential vessel). This deviation is usually believed to be caused by the non-uniform velocity profile, velocity fluctuation due to diffusion, short-circuiting, by-passing and channelling of particles, by the presence of segregating component or stagnant regions caused by mixer geometrical shape and internal, or by the recycling of solid particles within the mixer as a result of impeller and helix design. It is as yet to be determined how much each of the aforementioned mechanisms affects the mean and the geometric mean residence time of the mixer.

Finally, as was also explained in § 2.4.1, in literature the sole reason for using the gamma distribution is due to the variety of shapes which can be described by different choices of α and μ. This can certainly be an advantage in fitting the data, but it is doubtful if it can contribute to gaining new insight into the processes involved. The above analysis not only explains why the gamma distribution is the natural choice but also relates the parameters α and μ to some physically measurable characteristics of the continuous systems.

5.6.1 Application to Modelling of Plasticating Twin-Screw Extruder

The above modelling technique can be applied to wide variety of continuous systems in industry. As it was mentioned earlier, it has been long known that gamma distribution can be used as a density for RTD of many continuous systems (e.g. see Wen and Fan 1975, Ch. 8). In order to demonstrate this, in the following, the parameters of the gamma distribution for a twin screw extruder is determined and compared with the experimental RTD published in literature.

Twin screw extrusion is one of the core operations in polymer processing and is also a key component in many other processing operations. An extrusion process consists of a solid conveying section, a melting section with a mixture of solids and liquid material, and finally a metering zone where only liquid is pumped. The foremost goal of a twin screw extrusion process is to build pressure in a polymer melt so that it can be extruded through a die or injected into a mould. The screw design should ensure good mixing conditions and a uniform temperature distribution of the melt, in addition to serving as a good positive conveying and pumping device (see e.g. Tadmor and Klein 1970). When uniform retention times are desired, the system should provide a uniform distribution of the material, which means that plug flow conditions should prevail with no dead zone regions (Todd 1978). In the following, based on the published data by Wolf *et al.* (1986), the RTD of a commercial counterrotating twin-screw extruder is studied.

In order to determine the RTD of the extruder, Wolf and co-workers used a special injector to inject an impulse stimulus of the radioactive tracer in the feed stream. This was done after the extruder was at steady state with regard to flow rate, pressure and temperature conditions. The time of injection was recorded and the tracer gamma-ray radiations, $C(t)$ (number of counts/min) at the die was continuously monitored by a sodium iodide probe (see the first two columns of the Table 5.6.1).

Now, in the case of the impulse response, the concentration of the injected tracer at the outlet of the system is

$$C(t) = \int_0^\infty q\delta(\tau) f(t-\tau)\, d\tau = qf(t), \tag{5.6.20}$$

where q is the quantity of the tracer injected and $f(t)$ is the density for RTD. This implies that

$$f\left(t\right)=\frac{C\left(t\right)}{\int_0^\infty C\left(t\right)dt}. \tag{5.6.21}$$

Similarly, the cumulative RTD is

$$F\left(t\right)=\int_0^t f\left(\tau\right)d\tau=\frac{\int_0^t C\left(\tau\right)d\tau}{\int_0^\infty C\left(t\right)dt}. \tag{5.6.22}$$

The experimental distribution functions can be calculated by using the data of output concentration vs. time as given below:

$$\begin{aligned} f\left(t\right) &= \tfrac{C(t)}{\sum_{t=0}^{\infty}C(t)\Delta t} \\ F\left(t\right) &= \textstyle\sum_0^t f\left(t\right)\Delta t \end{aligned}$$

Partial calculations of these functions can be found in Table 5.6.1.

Accordingly, the mean and geometric mean residence time of the system are

$$\begin{aligned} \bar{t} &= \tfrac{\sum_{t=0}^{\infty}t\times C(t)\Delta t}{\sum_{t=0}^{\infty}C(t)\Delta t}=3.3377 \\ g &= \exp\left(\tfrac{\sum_{t=0}^{\infty}\ln(t)\times C(t)\Delta t}{\sum_{t=0}^{\infty}C(t)\Delta t}\right)=3.3272 \end{aligned} \tag{5.6.23}$$

Now, putting the above values in Eq. (5.6.13) and solving for α, one gets

$$\alpha=199. \tag{5.6.24}$$

Hence, according to Eq. (5.6.9), the density for RTD of the system is

$$f(t)=\begin{cases} \left(1.0203\times10^{-17}\right)t^{198}e^{-59.62t}\ , t>0 \\ 0\ , t\leq 0 \end{cases} \tag{5.6.25}$$

Both theoretical and experimental density and cumulative RTD of the system are shown in Figs. 5.6.2 and 5.6.3. As it was mentioned previously, large values of α, which in this case is 199, correspond to more plug flow behaviour as compared to ideal mixer model for

t [min]	$C(t)$ [Cnt/min]	$C(t)\,\Delta t$	$\ln(t)\,C(t)\,\Delta t$	$tC(t)\,\Delta t$	$f(t)$	$f(t)\,\Delta t$	$F(t)$
0-2.8	0	0	0	0	0	0	0
2.8-2.9	120	12	12.568	34.2	0.0897	0.0090	0.0090
2.9-3.0	600	60	64.908	177	0.0448	0.0448	0.0538
3.0-3.1	1550	155	172.847	472.75	1.1584	0.1158	0.1697
3.1-3-2	2350	235	269.640	740.25	1.7564	0.1756	0.3453
3.2-3.3	2460	246	289.949	799.5	1.8386	0.1839	0.5291
3.3-3.4	1900	190	229.703	636.5	1.42	0.1420	0.6712
3.4-3.5	1360	136	168.419	469.2	1.0164	0.1016	0.7728
3.5-3.6	980	98	124.161	347.9	0.7324	0.0732	0.846
3.6-3.7	700	70	90.631	255.5	0.5232	0.0523	0.8984
3.7-3.8	500	50	66.088	187.5	0.3737	0.0374	0.9357
3.8-3.9	320	32	43.138	123.2	0.2392	0.0239	0.9596
3.9-4.0	200	20	27.474	79	0.1495	0.0149	0.9746
4.0-4.1	120	12	16.785	48.6	0.0897	0.0090	0.9836
4.1-4.2	70	7	9.962	29.05	0.0523	0.0052	0.9888
4.2-4.3	50	5	7.235	21.25	0.0374	0.0037	0.9925
4.3-4.4	40	4	5.881	17.4	0.0299	0.0030	0.9955
4.4-4.5	30	3	4.479	13.35	0.0224	0.0022	0.9978
4.5-4.6	20	2	3.030	9.1	0.0149	0.0015	0.9993
4.6-4.7	10	1	1.537	4.65	0.0075	0.0007	1
4.7-4.8	0	0	0	0	0	0	1

Table 5.6.1: Experimental data by Wolf et al. (1986) and calculated parameters of the tracer experiment on a commercial counterrotating twin-screw extruder.

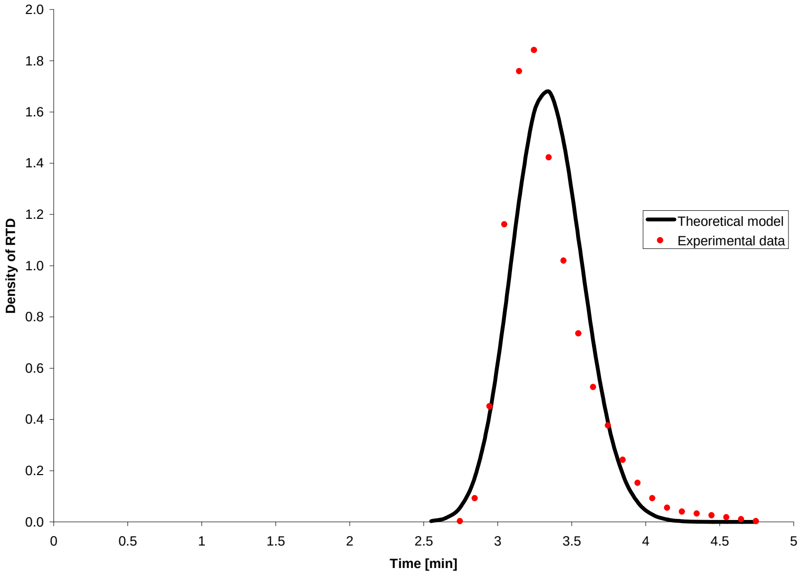

Figure 5.6.2: The theoretical model for the density of RTD versus corresponding experimental values from Wolf et al. (1986)

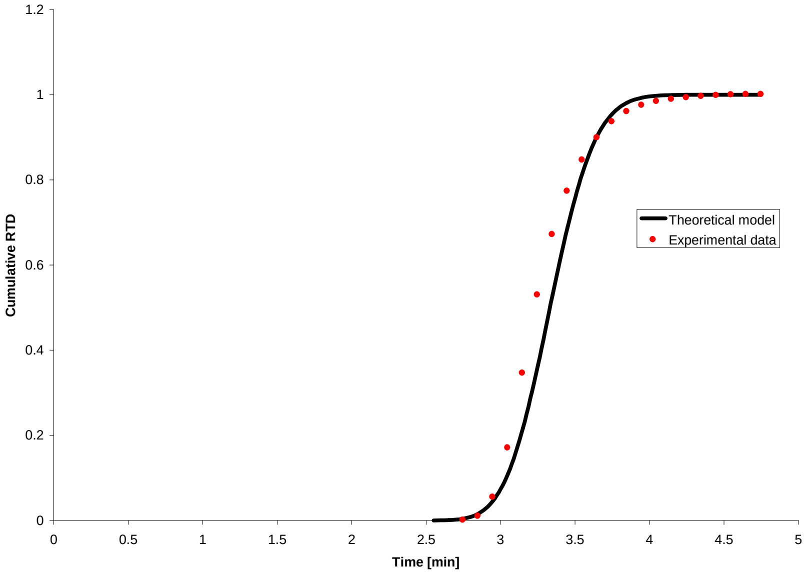

Figure 5.6.3: The theoretical model for the cumulative RTD versus corresponding experimental values from Wolf et al. (1986)

which α is equal one. The plug flow features can also be seen from the Figs. 5.6.2 and 5.6.3.

It is obvious that the gamma distribution captures the most essential features of the experimental data. However, it is easy to show that large values of α are very sensitive to errors in measured mean and geometric mean residence time. Indeed, in the above case, when comparing experimental data with the theoretical model, two things have to be kept in mind. First of all, since Wolf *et al.* used radioactive tracers of relatively short half-life[3], the effect of decay on the true measurement had to be taken into account. Although Wolf *et al.* had noticed this and pointed out that they have taken this into consideration, no figures on the expected error after this correction is given. Secondly, and perhaps more importantly, the experimental errors due to concentration measurements and flow rate fluctuation are magnified in the f-curve and causes more scatter of data points.

---

[3] The tracers had a half-life of 2.576 h. This would correspond to a reduction of counts by 6% during a period of 16 min (see Wolf and White 1976).

Chapter 6

A Model for Dynamics of the Mixing Processes

> One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity.
>
> Willard Gibbs, Rumford Medal Ceremoney (1881)

In this chapter a new approach to the dynamics of the mixing processes is introduced. This chapter begins by a short introduction to some of the most commonly used concepts in theory of complex dynamic systems. Based on these concepts, the theory of punctuated equilibrium is introduced and is shown how it can explain the dynamics of mixture of particulate materials. Furthermore, it is demonstrated how all these concepts can be unified under the concept of heterogeneity landscape. A mathematical classification of all the valleys in a heterogeneity landscape is also achieved through the introduction of the heterogeneity equation, which is summarized under the fundamental theorem of mixing. This chapter is concluded by a discussion on a method for determining the quotients of the heterogeneity equation.

6.1 A Qualitative Approach to Mixing Processes

A mixer acts on a mixture in a deterministic way. However, in general, it is difficult to predict the final state of the mixture, in which both chaos and new order may emerge.

Therefore, a mixing system can be considered as a *complex dynamic system.* According to Heylighen (1996), a system would be more complex if more parts could be distinguished, and if more connections between them existed. More parts to be represented means more extensive models, which require more time to be searched or computed. Since the components of a complex entity cannot be separated without destroying it, the method of analysis or decomposition into independent modules cannot be used to develop or simplify such models. This implies that complex entities will be difficult to model, that eventual models will be difficult to use for prediction or control, and that problems will be difficult to solve. In this relation, Grassberger (1989) defines the complexity as midpoint between *order* and *disorder,* which is also known as being on the *edge of chaos.* However, Edmonds (1996) points out that Grassberger's definition of complexity depends on the level of representation: what seems complex in one representation, may seem ordered or disordered in a representation at a different scale. Indeed, in the present work, this role is played by the scale of scrutiny. As was explained previously, the perception of the mixture quality is greatly influenced by the size of samples. As the size of samples decrease, the mixture becomes more heterogeneous. At the extreme limit, i.e., when each sample only consists of one particle, it would be impossible to say anything about the quality of the mixture. Thus, for a given scale of scrutiny, a mixture could posses a simple structure, while changing the scale might result in a more complex structure.

Evidently, the quality of the mixture depends on the state of the mixture. Recall that the state of a mixture is defined as spatial configuration of the constituents of the mixture at a particular instant in time and the set of all such states as state space (see also § 4.6.1). Accordingly, the evolution of the mixture can then be described by a time-parametrized trajectory in the state space, representing the states of the mixture at subsequent instants. Obviously, the trajectory to a great extent is affected by the mixing parameters, some of which were mentioned previously. According to Heylighen (1992), in order to determine possible trajectories in the state space, one needs two further structures: operators and dynamical constraints. An *operator* is a transformation or transition rule mapping initial states onto subsequent states[1]. A *dynamical constraint* is a selection criterion which determines which of the possible state transitions corresponding to different operators will actually take place. Dynamical constraints, in general, can be found in the following forms:

1. Differential (or difference) equations, relating the predicted state transition to the present state.
2. Conservation principles, stating that a certain global property of the system must be conserved during the transition.
3. Variation or optimization principles, stating that transition will occur which minimizes (or maximizes) a certain function of the transition parameters.

However, as mentioned previously, in general, one does not have all the information necessary to give a precise description of evolution of the state of the mixture. This means that the best one can achieve is to make the best estimate based on whatever information available. In other words, it is only possible to describe the most probable states based on the available information. Thus, in this work, the dynamical constraints can be found in the following forms:

1. Differential (or difference) equations, relating the most probable state transition to the present state of the mixture.
2. The information that has led to assignment of the pdf to the states of the mixture is testable, i.e., constraints on the moments of the pdf are known. The conservation principles affect the moments of the pdf.
3. Maximum entropy principle, stating that transition will occur which maximizes the entropy of the assigned pdf.

---

[1] It should be emphasized that operator is here used in the restricted meaning of a *time evolution operator*.

In Chapters 4 and 5, it was shown how points 2 and 3 in the above list can be utilized. The objective of this chapter is to demonstrate a possible approach to point 1. However, to start with, one needs a more intuitive model which can help to visualize the problem one is facing.

6.1.1 Attractors and Basins

The most fundamental concept in complex dynamics is that of an *attractor*. An attractor is a region of state space such that the trajectory of a dynamical system can enter but cannot leave, and which contains no smaller such region (Heylighen 1998). In a sense, an attractor can be considered as the collection of neighbouring equilibrium states which the mixture state converges to. The simplest attractor consists of only one state, and is known as *point attractor*. However, this is only possible in the hypothetical case of *ideally perfect mixture*, which as defined by Egermann (1980), is a mixture which consists of identical particles arranged in a perfectly ordered manner. In the case of realistic mixtures, an attractor can have a much more complex structure. For example, as Abouzeid (1989) has also pointed out, in any mixing operation, the mixing and demixing mechanisms will be operational. The combination of mixing-demixing, in which one is dampening the heterogeneity and the other one is amplifying it, can produce the most complicated behaviours. The simplest non-point attractor is a one-dimensional *limit cycle attractor* (see Fig. 6.1.1). But it is also possible to have attractors with multiple dimensions or even *fractal* dimensions, which are known as *strange attractors*.

Another characteristic of highly non-linear systems is that they have in general several attractors. This situation is usually encountered in the case of cohesive mixtures. Indeed, as Harnby (1985) also noticed, a cohesive mixture usually possesses a natural structure which has to be repeatedly broken down in order to give individual particles within that structure an opportunity of relocating themselves. The nature and strength of the interparticulate forces acting within the cohesive mixture will determine the ease,

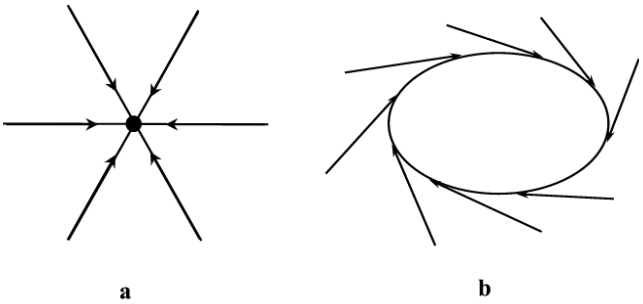

Figure 6.1.1: a) A point attractor: the arrows represent trajectories starting from different points but all converging in the same equilibrium state. b) A one-dimensional limit cycle attractor: the arrows correspond to trajectories starting outside the attractor, but ending up in a continuing cycle along the attractor.

or difficulty, likely to be experienced in re-locating individual particles within a mixture and also determines the number and the shape of the system attractors.

When there are more than one attractor, the main question is in which of those attractors the system will end up. Imagine, as Heylighen (1998) argues, that each attractor corresponds to a lake, and that the trajectories leading into an attractor correspond to the rivers and streams flowing into those lakes. Depending on where it falls, rainwater will follow either one river or another, ending up in either one lake or another. The complete area drained by a river is called its basin. Similarly, each attractor has a *basin*, which is the surrounding region in state space such that all trajectories starting in that region end up in the attractor. The basins belonging to different attractors are separated by a narrow boundary. However, the boundary separating the two basins will in general be very difficult to discriminate exactly (it may for example have a fractal shape). This means that for initial positions close to the boundary, it is very difficult to determine to which attractor they lead. Small fluctuations can push the system either into one or into the other basin, and therefore either into the one or into the other attractor (see Figure 6.1.2). Thus, close to the border, the system behaves chaotically, whilst inside the basin it moves predictably towards the attractor.

In cohesive mixtures, the interparticulate forces cause the agglomeration of constituent particles into small but stable groupings of like constituents (Harnby 1985). These, in turn

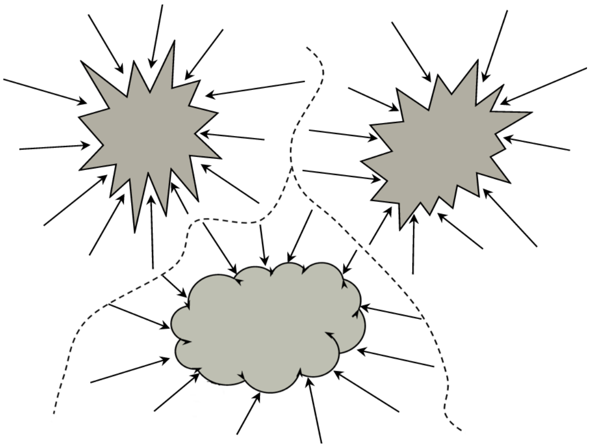

Figure 6.1.2: Three attractors with some of the trajectories leading into them. Their respective basins are separated by a dotted line.

may lead to macroscopically *self-organized* structures which are stable. Self-organization is basically a process where the effect of the environment is minimal, i.e., where the development of new, complex structures take place primarily in and through the system itself (Heylighen 1997). For example, when mixing a very cohesive mixture, there is a strong possibility that mixture will adhere to the mixer walls or form dead spots within the mixer. This usually results into a self-organized structure where there are a large number of stationary islands separated by small but dynamic regions. In these cases, the mixing process instead of being a slow and continuous process, tends to consist of long periods of virtual standstill, “equilibrium”, “punctuated” by avalanche like episodes of very fast development to new heterogeneity levels. For this reason, in this work whenever a mixture achieves this type of self-organized structure, it is said that the mixture self-organized itself into state of *punctuated equilibrium*.

In general, at the punctuated equilibrium, typical variations in mixture heterogeneity tend to be small. This is because the changes either cover small regions of the mixture and therefore their global impacts are small, or system has entered an attractor. However, occasionally big changes do take place without necessarily being caused by any external triggering mechanism. This is mainly caused by interaction among the constituents of the mixture through the mixing equipment in which cascade of local changes result in overall global change in the heterogeneity level of the mixture. In the aforementioned terminology, the big changes usually occur whenever a bifurcation takes place in which it has a drastic effect on the mixture heterogeneity.

For example, agglomeration of the particles in a mixture may result in a stable structure, i.e., the mixture ends up in an attractor, in which the changes in the level of heterogeneity of the mixture are quite small. However, mixers with high shearing or impaction characteristics are capable of breaking down agglomerates (Harnby 1985). In these cases, if the mixing process persists, eventually the structure of the mixture will lose its stability and breaks down. The mixture may then evolve toward another attractor. This results in changes which affects the whole mixture and thus, in general, a change in the mixture heterogeneity.

In the aforementioned example, it is evident that shearing or impaction action results in an increase in the number of attractors of the system. One way to understand this is by observing that destruction of agglomerates results in an increase in the number of free constituents of the mixture. The greater number of free constituents means greater number of possible mixture states and thus, in general, greater number of attractors. Hence, in the above example, as mixing process continuous, the system attractor splits in two (or more). Prigogine (1984, p.160) calls the splitting of attractors as *bifurcation.* After the first bifurcation, the system achieves two stable patterns of behaviour and the mixture will evolve very rapidly toward one of them, depending on which basin its state falls into right after bifurcation. If the mixing continuous, more bifurcation takes place, and the attractors split up further. At certain point, the number of attractors becomes very large and the system is erratically jumping from the one to the other all the time. This is true chaos. At this point, the behaviour of the system has become totally unpredictable and no structure can be sustained. That is, it is like finding an attractor arbitrarily close to any point of state space. At this stage, if the mixing process is halted, the final resting state of the mixture could be anywhere in the state space. An overview of entire range of possible states is illustrated in Fig. 6.1.3, in what is termed a *bifurcation diagram.*

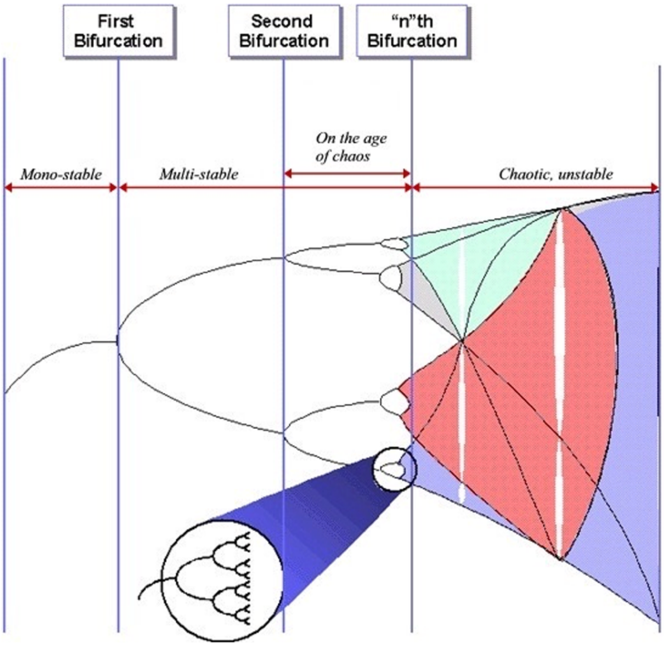

Figure 6.1.3: Bifurcation diagram: it represents the entire range of possible mixture states.

In general, the true chaos is not desirable. Since, pushing the state of the mixture toward the far right hand side of the bifurcation diagram implies that the mixture heterogeneity can practically attain every possible value, something which makes the mixing operation redundant. However, in some cases, it is an effective way to destroy the undesired stable structures. Nonetheless, under normal conditions, most mixtures remain in a region in the bifurcation diagram known as *edge of chaos*. In this region, the mixture will eventually self-organizes itself into the punctuated equilibrium.

Of course for a given instance of time, a mixture in the punctuated equilibrium does not necessarily posses a good quality. It may appear that the concept of punctuated equilibrium suggests that if the mixing operation continuous, eventually over time, the quality of the mixture may improve by occasional bifurcation events. However, there are at least two reasons to believe that this process, in general, is not efficient. First of all, since the time between each bifurcation could be arbitrarily long, there is no guarantee that one can achieve the desired mixture quality within the reasonable time interval. Secondly, sensitivity of the system to fluctuations at each bifurcation point may in effect result in rising of the level of mixture heterogeneity. Indeed, the effects of heterogeneous nature of the constituents of the mixture are negligible within an attractor. But since the bifurcation is mainly induced due to interaction between the constituents of the mixture, the heterogeneous nature of the mixture plays a significant role at the close vicinity of the bifurcation point. This means that in practice, right after a bifurcation, one has little control over which attractor the system will evolve into. Thus in practice, one needs to change some of the mixing parameters in order to improve the quality of the mixture. The change of mixing parameters may result in getting the system out of equilibrium. Of course, if the mixing persists, the mixture enters another equilibrium state which on average might posses a better quality. Otherwise, same process has to be repeated again until desired quality is attained.

What the concept of punctuated equilibrium seems to suggest is that there are virtually same mechanisms which are responsible for small as well as big changes in the mixture heterogeneity. The self-organization of the mixture into the punctuated equilibrium is established solely because of the dynamical interactions among individual elements of the mixture. Therefore, the concept of punctuated equilibrium could only be able to describe the general features of a mixing process, not detailed account of utterly accidental details of that process.

6.1.2 Heterogeneity Landscape

Recall that a sample space consists of collection of all partitions of the mixture into a predetermined number of samples with predefined shapes, at a given instance of time (see also § 4.6.1). Thus, the dynamics of the mixture can be described by changes in the structure of the sample space over time. As it was shown in Section 4.6, one way to characterize the structure of a sample space is by the concept of mixture heterogeneity. Hence, the dynamics of the mixture can be described by the changes in the mixture heterogeneity over time.

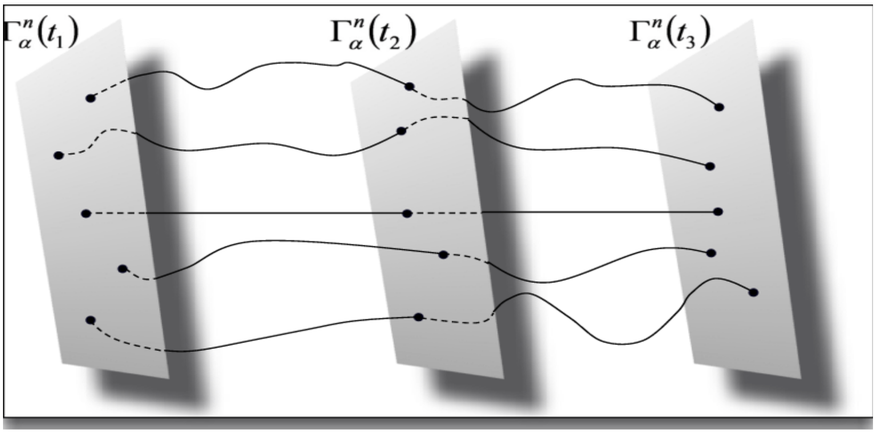

Figure 6.1.4: Schematic representation of the state space in which t_i denote the time and $t_1 < t_2 < t_3$. The points in the principal sub-space $\Gamma^n_\alpha(t_i)$ represent the different states of the mixture corresponding to the sample space $\Omega^n_\alpha(t_i)$.

Now, as mentioned above, the dynamics of the mixture can also be described by a time-parametrized trajectory in the state space, representing the states of the mixture at subsequent instants, modulo n (see § 4.6.1). Then each point on such a trajectory can be identified by a partition in a sample space. This means that the state space can be partitioned into subsets, each of which can be identified by certain sample space. Indeed, in Fig. 6.1.4 a schematic representation of the state space is given in which five possible trajectories in the state space are depicted. Consequently, each path can be parametrized by time and hence for a given time, the set of all the points belonging to different paths constitute a sub-space, which for convenience is denoted as *principal sub-space* of the state space. In Fig. 6.1.4, the principal sub-spaces are represented by the planes $\Gamma^n_\alpha(t_i)$. Since every point in each $\Gamma^n_\alpha(t_i)$ can be identified by a partition in the sample space $\Omega^n_\alpha(t_i)$ and vice versa, then $\Gamma^n_\alpha(t_i)$ can also be identified by the same mixture heterogeneity as $\Omega^n_\alpha(t_i)$. However, the reader is warned that in general the topology of the principal sub-spaces $\Gamma^n_\alpha(t_i)$ is more complicated than the ones depicted in Fig. 6.1.4. For example $\Gamma^n_\alpha(t_i)$ might not even be connected, i.e. it consists of several isolated sub-spaces or some paths may even be self-intersecting, resulting in more complicated topology.

Thus, from the state space point of view, the mixture heterogeneity can be interpreted as a measure for how spread the paths are in the state space. The more spread the paths in the state space are in a given time, the larger the mixture heterogeneity would be at that given time. Accordingly, the paths in the state space converge in the attractors and hence the mixture heterogeneity is low in the vicinity of attractors. On the other extreme, the paths in the state space are more spread in the vicinity of bifurcation points and hence correspond to much higher mixture heterogeneity. In § 4.6.2, the heterogeneous properties of the mixture where divided into two categories, static and dynamic. The contribution of dynamic properties to mixture heterogeneity is only significant in the vicinity of bifurcation points, whilst in attractors the static properties are dominant.

Now, the rather abstract and mathematically complex structure of a system of attractors and basins can be replaced by the more intuitive model of *heterogeneity landscape.* A heterogeneity landscape is determined by the collection of principal sub-spaces $\mathcal{M}$ of the mixture states, a definition of the neighbourhood of principal sub-spaces, and the mixture heterogeneity function $\sigma^2 : \mathcal{M} \to \mathbb{R}$. On the heterogeneity landscape every principal sub-space has a certain height corresponding to a particular value of σ^2. This landscape has peaks and valleys. In this relation, one can imagine that there are two forces[2] which are operating in the heterogeneity landscape; mixing and demixing. Mixing forces the mixture state toward a lower level of heterogeneity and therefore one can assume that it acts downward. On the other hand, demixing increases the mixture heterogeneity and thus can be regarded as a force which acts upward. Evidently, the strength of these forces depends on the mixing parameters.

Now, it is desirable that the trajectory of the system through the state space always follow the path of steepest descent, i.e., move from a given principal sub-space state to that neighbouring one for which σ^2 is minimal. This is exactly what one is trying to achieve by designing a good mixer. Moreover, from the above discussion, it is evident that the attractors of the state space correspond to the local minima of σ^2. This means

---

[2] In this context, the word force is a metaphor to illustrate a concept and is quite different from its usual definition in Newtonian mechanics.

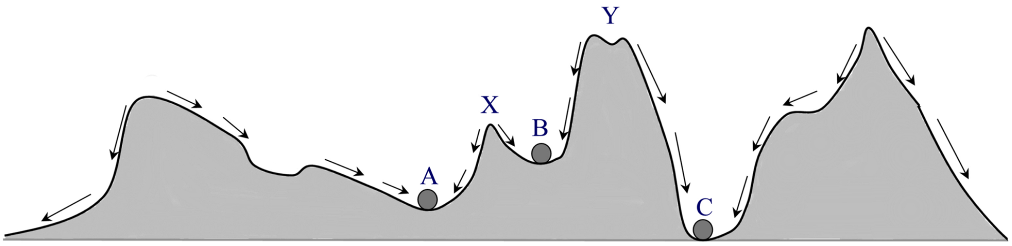

Figure 6.1.5: A Heterogeneity landscape: the arrows denote the directions in which the mixture state will evolve. The height of a position corresponds to the value of the mixture heterogeneity in that state.

that, in general, the system would move downward in the heterogeneity landscape. When it has reached the locally lowest point, all remaining directions would point upward, and therefore the system would not be able to leave the bottom of the valley. The local maxima of the heterogeneity (peaks) are points that separate the basins of the attractors (valleys) that lie in between the peaks. In general, the steeper the slope, the faster would be the descent of the system along that slope. An imaginary heterogeneity landscape is depicted in Fig. 6.1.5, in which the arrows denote the directions to which the mixture state could evolve. In this context, the mixture state A corresponds to a lower heterogeneity level than B. The bottoms of the valleys A, B and C are the local minima of the mixture heterogeneity function, i.e., the attractors. The peaks X and Y delimit the basin of the attractor B. X separates the basins of A and B, whilst Y separates the basins of B and C.

In general, mixing operation does not only affect the state of the mixture but also changes the heterogeneity landscape. For example, as mentioned previously, when mixing very cohesive mixtures, a mixer with high shearing characteristic is capable of breaking down the agglomerates. The destruction of agglomerates results in an increase in the number of free constituents of the mixture, which in turn increases the number of available states. Moreover, destruction of the agglomerates means also that the mixture state is changed. This means that mixture has moved to a new neighbouring state that did not exist before comminution, i.e., as a result of mixing operation, the heterogeneity landscape is changed.

6.2 Interlude: Comments

Just a quick study of literature reveals that although different terminologies are used in different subjects, many people working with complex phenomenon are familiar with the sensitivity of the systems in the vicinity of a bifurcation. For example, in a recently published best-seller, Gladwell (2002) uses the term *tipping point* instead of bifurcation to describe how a small but precisely targeted push cause a fashion trend, the popularity of a new product, or a drop in the crime rate. In a now classical work in Sociology, Schelling (1971) demonstrated, by cellular automata modelling, how at the tipping point a small preference for one's neighbours to be of the same colour could lead to total segregation of a society. Similar approaches could be found to almost all complex problems in different scientific disciplines.

The theory of punctuated equilibrium was first proposed by Gould and Eldredge (1977) as a criticism of the traditional Darwinian theory of evolution. The traditional Darwinian theory of evolution saw evolution as a slow, continuous process, without sudden jumps. However, if one studies the fossils of organisms found in subsequent geological layers, it appears that long intervals in which nothing changed (equilibrium), punctuated by short, revolutionary transitions, in which species became extinct and replaced by wholly new forms. It accounts for what the fossil record appears to suggest that evolution of species consists of a series of irregularly spaced periods of chaotic and rapid evolutionary changes in what are otherwise dominated by long periods of evolutionary stasis.

It appears that in general, large complex systems with many components have a tendency to evolve toward the state of punctuated equilibrium. In the context of this work, this happens when a mixture self-organizes itself into a structure in which small changes in mixture heterogeneity are much more frequent over time than the large ones. In fact, Bak *et al.* (1988) claim that many complex systems have tendency to self-organize to a critical state, where the size of disturbances obeys a *Pareto like distribution*, in which the large disturbances being less frequent than small ones. They called this phenomenon *self-organized criticality* (SOC). In the context of present chapter, according to Bak (1997, Ch.9), the state of punctuated equilibrium is achieved whenever the mixture self-organizes itself into the critical state. However, although this idea is quite compelling, we believe that at punctuated equilibrium, how an observer perceives the distribution of the mixture heterogeneity depends on the available information on the evolution of the mixture structure over time. This means that the distribution of the mixture heterogeneity does not necessarily need to be Pareto like. If we are supplied with sufficient amount of information we might discover details that cannot be modeled by Pareto like distributions. Moreover, as it was demonstrated in Section 5.6, the Pareto like distributions might not even be able to uniquely describe the phenomenon under study. Nonetheless, if one chooses to look beyond this minor disagreement, the following quote by Bak captures the core essence of systems in punctuated equilibrium (Bak 1997, p.143):

> "Systems with punctuated equilibria combine features of frozen, ordered systems, with those of chaotic, disordered systems. Systems can remember the past because of the long periods of stasis allowing them to preserve what they have learned through history, mimicking the behaviour of frozen systems; they can evolve because of the intermittent bursts of activity."

The concept of heterogeneity landscape of previous section was inspired by a similar idea in biology which was first introduced by Wright (1982) and is known as *fitness landscape*. The fitness landscape depicts the process of adaptive evolution as that of climbing peaks and passing over saddles to reach higher peaks in the fitness landscape. This picture proved to tremendously simplify the topology of genetic space. However, Wright's fitness landscape is static contrary to our heterogeneity landscape which is sort of "deformable rubber landscape". This results in that the heterogeneity landscape is much richer and more general a concept than the fitness landscape. In fact, there are reasons to believe that deformable rubber fitness landscape is also more appropriate than the Wright's original one, in describing the evolution of species. For a more detail discussion of this topic we refer to Kauffman (1993).

6.3 Valleys in Heterogeneity Landscape

As it was mentioned previously, the heterogeneity landscape also changes over time and for all practical reasons, it is not possible to have complete information over its geography. However, as soon as the mixture state reaches the bottom of a valley and the mixture achieves equilibrium, the mixture heterogeneity exhibits only small fluctuations. As it was noticed by Lacey (1943), although stated differently, for any given mixer, the best possible mixture is obtained whenever the mixture reaches a state in which the observer's information about its structure is insufficient to confidently claim that further mixing has any significant structural effect on the mixture. This is what is known as a *random* mixture. In the context of the present work, this corresponds to bottom of a valley in the heterogeneity landscape with lowest height. However, in general, heterogeneity landscape might also contain several other valleys which do not correspond to the state of minimum heterogeneity. Theoretically, at equilibrium, the state of the mixture could be in any one of these valleys and therefore it is important to have a model which is general enough to be able to describe any valley in the heterogeneity landscape. In the following, a mathematical model is proposed which fulfills this condition.

To this end, let $p\left(\mathbf{h}\middle|\,\sigma_t^2, t, I\right)$ denote the pdf describing the conditional probability of $\mathbf{h}$ at a given time t where σ_t^2 and I stand for mixture heterogeneity at time t and information which indicates that mixture is at equilibrium, respectively. Since, it is assumed that the mixer is still processing the mixture, it is reasonable to assume that the level of mixture heterogeneity is fluctuating, but in a small neighbourhood of bottom of the given valley. Now, as explained previously, p represents the state of the knowledge of the observer. At equilibrium, although σ_t^2 fluctuates over time, since p describes the same equilibrium state and hence same state of knowledge, the functional form of p should remain same. Consequently, one has a hierarchy of distributions $p\left(\mathbf{h}\middle|\,\sigma_t^2, t, I\right)$

characterized by a single scale parameter σ_t^2. Whenever the mixture heterogeneity is changed by adding or subtracting small increments of heterogeneity to components of $\mathbf{h}$, the probability distribution, which is a model for the state of observer's knowledge, will keep its functional form, but is only moved up the hierarchy to a new value of σ_t^2.

Now, suppose that the pdfs $p\left(\mathbf{h}\left(t\right)\middle|\,\sigma_t^2, t, I\right)$ and $f\left(\mathbf{h}\left(t+\tau\right) = \bar{\mathbf{h}}\middle|\,\sigma_{t+\tau}^2, t+\tau, I\right)$ are assigned to the heterogeneity functions $\mathbf{h}\left(t\right)$ and $\bar{\mathbf{h}}$, respectively. Let $\mathbf{h}\left(t\right) = \bar{\mathbf{h}} - \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ has a probability distribution $q\left(\boldsymbol{\varepsilon}\middle|\,\sigma_{\boldsymbol{\varepsilon}}^2, t+\tau, I\right)$, which for convenient is denoted as *transition probability* and where $\sigma_{\boldsymbol{\varepsilon}}^2$ is the contribution to the mixture heterogeneity by $\boldsymbol{\varepsilon}$. By definition, at equilibrium, q must be independent of p. Furthermore, $\boldsymbol{\varepsilon}$ is evidently function of τ and q must fulfil the condition

$$\lim_{\tau\to 0} q\left(\boldsymbol{\varepsilon}\middle|\,\sigma_{\boldsymbol{\varepsilon}}^2, t+\tau, I\right) = \prod_{i=1}^{n} \delta\left(\left\langle \varepsilon_i \right\rangle\right), \tag{6.3.1}$$

where δ denotes the impulse function and $\langle \varepsilon_i \rangle$ denotes the expected value of ε_i.

By the product rule (3.2.1), one gets

$$f\left(\bar{\mathbf{h}}, \boldsymbol{\varepsilon}\middle|\,\sigma_{t+\tau}^2, t+\tau, I\right) = f\left(\bar{\mathbf{h}}\middle|\,\sigma_{t+\tau}^2, t+\tau, I\right) f\left(\boldsymbol{\varepsilon}\middle|\,\bar{\mathbf{h}}, \sigma_{t+\tau}^2, t+\tau, I\right). \tag{6.3.2}$$

But since by assumption $\boldsymbol{\varepsilon}$ is independent of $\mathbf{h}$ then

$$\sigma_{t+\tau}^2 = \sigma_{\mathbf{h}}^2\left(t\right) + \sigma_{\boldsymbol{\varepsilon}}^2\left(t+\tau\right) \tag{6.3.3}$$

and hence

$$f\left(\boldsymbol{\varepsilon}\middle|\,\bar{\mathbf{h}}, \sigma_{t+\tau}^2, t+\tau, I\right) = f\left(\boldsymbol{\varepsilon}\middle|\,\sigma_{\boldsymbol{\varepsilon}}^2, t+\tau, I\right) = q\left(\boldsymbol{\varepsilon}\middle|\,\sigma_{\boldsymbol{\varepsilon}}^2, t+\tau, I\right). \tag{6.3.4}$$

Moreover, note that for a specific $\boldsymbol{\varepsilon}$, the probability of the heterogeneity function having the value $\bar{\mathbf{h}}$ is just the probability that $\mathbf{h}$ should have the value $\left(\bar{\mathbf{h}}-\boldsymbol{\varepsilon}\right)$. Thus

$$f\left(\bar{\mathbf{h}}\middle|\,\sigma_{t+\tau}^2,t+\tau,I\right)=p\left(\bar{\mathbf{h}}-\boldsymbol{\varepsilon}\middle|\,\sigma_t^2,t+\tau,I\right), \tag{6.3.5}$$

which combined by (6.3.2) and (6.3.4) yields

$$f\left(\bar{\mathbf{h}},\boldsymbol{\varepsilon}\middle|\,\sigma_{t+\tau}^2,t+\tau,I\right)=p\left(\bar{\mathbf{h}}-\boldsymbol{\varepsilon}\middle|\,\sigma_t^2,t+\tau,I\right)q\left(\boldsymbol{\varepsilon}\middle|\,\sigma_{\boldsymbol{\varepsilon}}^2,t+\tau,I\right). \tag{6.3.6}$$

Finally, by marginalizing the nuisance parameter $\boldsymbol{\varepsilon}$ (see Section 3.5 on Marginalization), one gets,

$$f\left(\bar{\mathbf{h}}\middle|\,\sigma_{t+\tau}^2,t+\tau,I\right)=\int\limits_{-\infty}^{\infty}p\left(\bar{\mathbf{h}}-\boldsymbol{\varepsilon}\middle|\,\sigma_t^2,t+\tau,I\right)q\left(\boldsymbol{\varepsilon}\middle|\,\sigma_{\boldsymbol{\varepsilon}}^2,t+\tau,I\right)d\boldsymbol{\varepsilon}. \tag{6.3.7}$$

Now, for small τ it is reasonable to assume that $\sigma_{\boldsymbol{\varepsilon}}^2\left(t+\tau\right)\ll\sigma_{\mathbf{h}}^2\left(t\right)$. The Taylor series expansion of $p\left(\bar{\mathbf{h}}-\boldsymbol{\varepsilon}\middle|\,\sigma_t^2,t+\tau,I\right)$ with respect to $\mathbf{h}$ is, (Arfken 2001, p.340)

$$p\left(\bar{\mathbf{h}}-\boldsymbol{\varepsilon}\middle|\,\sigma_t^2,t+\tau,I\right)=\sum_{k=0}^{\infty}\frac{\left(-1\right)^k}{k!}\left(\boldsymbol{\varepsilon}\cdot\boldsymbol{\nabla}_{\mathbf{h}}\right)^k p\left(\mathbf{h}\middle|\,\sigma_t^2,t,I\right) \tag{6.3.8}$$

where

$$\boldsymbol{\nabla}_{\mathbf{h}}=\left(\frac{\partial}{\partial h_1},\ldots,\frac{\partial}{\partial h_n}\right). \tag{6.3.9}$$

Substituting (6.3.8) back into (6.3.7) and rearranging the terms

$$\begin{aligned}f\left(\bar{\mathbf{h}}\middle|\,\sigma_{t+\tau}^2,t+\tau,I\right)=\;&p\left(\mathbf{h}\middle|\,\sigma_t^2,t,I\right)-\sum_{j=1}^{n}\left\langle\varepsilon_j\right\rangle\tfrac{\partial}{\partial h_j}p\left(\mathbf{h}\middle|\,\sigma_t^2,t,I\right)\\&+\tfrac{1}{2}\sum_{j=1}^{n}\left\langle\varepsilon_j^2\right\rangle\tfrac{\partial^2}{\partial h_j^2}p\left(\mathbf{h}\middle|\,\sigma_t^2,t,I\right)+\sum_{i<j}^{n}\left\langle\varepsilon_i\varepsilon_j\right\rangle\tfrac{\partial^2}{\partial h_i\partial h_j}p\left(\mathbf{h}\middle|\,\sigma_t^2,t,I\right)\\&-\tfrac{1}{6}\sum_{j=1}^{n}\left\langle\varepsilon_j^3\right\rangle\tfrac{\partial^3}{\partial h_j^3}p\left(\mathbf{h}\middle|\,\sigma_t^2,t,I\right)-\cdots\end{aligned} \tag{6.3.10}$$

where $\langle\cdot\rangle$ is the expected value operator.

As discussed earlier, at equilibrium the functional form of the probability distribution remains the same except possibly moving up the σ_t^2 hierarchy. Accordingly, since by assumption $\sigma_{\boldsymbol{\varepsilon}}^2(t+\tau) \ll \sigma_{\mathbf{h}}^2(t)$, the tangent-plane approximation (see Loomis and Sternberg 1990, p.140) implies

$$f\left(\bar{\mathbf{h}}\middle|\,\sigma_{t+\tau}^2, t+\tau, I\right) = p\left(\mathbf{h}\middle|\,\sigma_t^2, t, I\right) + \sigma_{\boldsymbol{\varepsilon}}^2 \frac{\partial}{\partial \sigma_t^2} p\left(\mathbf{h}\middle|\,\sigma_t^2, t, I\right) + \mathcal{O}\left(\sigma_{\boldsymbol{\varepsilon}}^2\right) \tag{6.3.11}$$

where $\mathcal{O}$ is a function of higher orders of $\sigma_{\boldsymbol{\varepsilon}}^2$ such that

$$\lim_{\sigma_{\boldsymbol{\varepsilon}}^2 \to 0} \frac{\mathcal{O}\left(\sigma_{\boldsymbol{\varepsilon}}^2\right)}{\sigma_{\boldsymbol{\varepsilon}}^2} = 0. \tag{6.3.12}$$

Comparing (6.3.11) and (6.3.10), yields

$$\begin{aligned} \tfrac{\partial}{\partial \sigma^2} p\left(\mathbf{h}\middle|\,\sigma^2, t, I\right) + \tfrac{\mathcal{O}\left(\sigma_{\boldsymbol{\varepsilon}}^2\right)}{\sigma_{\boldsymbol{\varepsilon}}^2} = & -\sum_{j=1}^{n} \tfrac{\langle \varepsilon_j \rangle}{\sigma_{\boldsymbol{\varepsilon}}^2} \tfrac{\partial}{\partial h_j} p\left(\mathbf{h}\middle|\,\sigma^2, t, I\right) \\ & + \tfrac{1}{2} \sum_{j=1}^{n} \tfrac{\left\langle \varepsilon_j^2 \right\rangle}{\sigma_{\boldsymbol{\varepsilon}}^2} \tfrac{\partial^2}{\partial h_j^2} p\left(\mathbf{h}\middle|\,\sigma^2, t, I\right) + \cdots, \end{aligned} \tag{6.3.13}$$

where $\sigma^2 = \sigma_t^2$. At this point, in order to make the notation more manageable, it is convenient to adapt the Einstein's summation convention, i.e., the summation is performed over the Latin indices appearing twice in the equations without writing down the summation sign. In this new notation, the Eqn. (6.3.13) can be written as

$$\frac{\partial}{\partial \sigma^2} p\left(\mathbf{h}\middle|\,\sigma^2, t, I\right) + \frac{\mathcal{O}\left(\sigma_{\boldsymbol{\varepsilon}}^2\right)}{\sigma_{\boldsymbol{\varepsilon}}^2} = \sum_{\nu=1}^{\infty} \frac{M_{j_1, j_2, \cdots, j_\nu}^{(\nu)}(\boldsymbol{\varepsilon}, t+\tau)}{\nu! \sigma_{\boldsymbol{\varepsilon}}^2} \frac{(-\partial)^\nu}{\partial h_{j_1} \cdots \partial h_{j_\nu}} p\left(\mathbf{h}\middle|\,\sigma^2, t, I\right), \tag{6.3.14}$$

where

$$M_{j_1, \cdots, j_\nu}^{(\nu)}(\boldsymbol{\varepsilon}, t+\tau) = \left\langle \varepsilon_{j_1} \cdots \varepsilon_{j_\nu} \right\rangle = \int_{-\infty}^{\infty} \left(\varepsilon_{j_1} \cdots \varepsilon_{j_\nu}\right) q\left(\boldsymbol{\varepsilon}\middle|\,\sigma_{\boldsymbol{\varepsilon}}^2, t+\tau, I\right) d\boldsymbol{\varepsilon}, \tag{6.3.15}$$

denotes the νth moment of $\boldsymbol{\varepsilon}$. Moreover, from (6.3.1) follows that

$$\lim_{\tau \to 0} M_{j_1, \cdots, j_\nu}^{(\nu)}(\boldsymbol{\varepsilon}, t+\tau) = 0 \text{ and } \lim_{\tau \to 0} \sigma_{\boldsymbol{\varepsilon}}^2(t+\tau) = 0. \tag{6.3.16}$$

Now, If τ is small enough it is reasonable to assume that $\sigma_{\boldsymbol{\varepsilon}}^2(t+\tau)$ is continuously differentiable function of τ with a non-vanishing derivative in a neighbourhood of $\tau = 0$. Then by the so-called *inverse-function theorem* of calculus (Loomis and Sternberg 1990, p.167), there exists a unique function, say α, on a neighbourhood of $\tau = 0$ such that

$$\tau = \alpha\left(\sigma_{\boldsymbol{\varepsilon}}^2\right)\cdot \tag{6.3.17}$$

Thus, the moments $M_{j_1,\cdots,j_\nu}^{(\nu)}(\boldsymbol{\varepsilon}, t+\tau)$ can be expanded into a Taylor series with respect to $\sigma_{\boldsymbol{\varepsilon}}^2$ to yield

$$\frac{M_{j_1,\cdots,j_\nu}^{(\nu)}(\boldsymbol{\varepsilon}, t+\tau)}{\nu!} = \frac{M_{j_1,\cdots,j_\nu}^{(\nu)}[\boldsymbol{\varepsilon}, t+\alpha]}{\nu!} = D_{j_1,\cdots,j_\nu}^{(\nu)}(t)\,\sigma_{\boldsymbol{\varepsilon}}^2 + \Delta\left(\sigma_{\boldsymbol{\varepsilon}}^2\right) \tag{6.3.18}$$

where Δ is a function of higher orders of $\sigma_{\boldsymbol{\varepsilon}}^2$. Now, let $\sigma_{\boldsymbol{\varepsilon}}^2 \to 0$ in (6.3.14) and taking into account only the linear terms in (6.3.18),

$$\frac{\partial}{\partial \sigma^2} p\left(\mathbf{h}\middle|\, \sigma^2, t, I\right) = L(\mathbf{h}, t)\, p\left(\mathbf{h}\middle|\, \sigma^2, t, I\right) \tag{6.3.19}$$

where

$$L(\mathbf{h}, t) = \sum_{\nu=1}^{\infty} D_{j_1,\cdots,j_\nu}^{(\nu)}(t)\,\frac{(-\partial)^\nu}{\partial h_{j_1}\cdots\partial h_{j_\nu}}\cdot \tag{6.3.20}$$

The solution of Eqn. (6.3.19), i.e., $p(\mathbf{h}|\, \sigma^2, t, I)$ with respect to the obvious initial condition (see Proposition 4.2)

$$p\left(\mathbf{h}\middle|\, \sigma^2 = 0, t, I\right) = \prod_{j=1}^{n} \delta(h_j), \tag{6.3.21}$$

represents the state of observer's knowledge about the mixture at equilibrium. Thus, by assigning appropriate values to the coefficients of the operator L, small neighbourhoods of the bottom of any valley in the heterogeneity landscape can be characterized by the solutions of the Eqn. (6.3.19). The results of this section are summarized in the following theorem.

Theorem 6.1 (Fundamental Theorem of Mixing) *In the heterogeneity landscape, the pdf* $p\left(\mathbf{h}\middle|\,\sigma^2,t,I\right)$ *in the small neighbourhoods of the bottom of any valley satisfies the heterogeneity equation*

$$\frac{\partial}{\partial\sigma^2}p\left(\mathbf{h}\middle|\,\sigma^2,t,I\right)=L\left(\mathbf{h},t\right)p\left(\mathbf{h}\middle|\,\sigma^2,t,I\right) \tag{6.3.22}$$

where

$$L\left(\mathbf{h},t\right)=\sum_{\nu=1}^{\infty}D^{(\nu)}_{j_1,\cdots,j_\nu}\left(t\right)\frac{\left(-\partial\right)^{\nu}}{\partial h_{j_1}\cdots\partial h_{j_\nu}} \tag{6.3.23}$$

and

$$D^{(\nu)}_{j_1,\cdots,j_\nu}\left(t\right)=\frac{1}{\nu!}\lim_{\tau\to0}\frac{M^{(\nu)}_{j_1,\cdots,j_\nu}\left(\boldsymbol{\varepsilon},t+\tau\right)}{\sigma^2_{\boldsymbol{\varepsilon}}\left(t+\tau\right)}, \tag{6.3.24}$$

in which the pdf of $\boldsymbol{\varepsilon}$ *is given by the transition probability* $q\left(\boldsymbol{\varepsilon}\middle|\,\sigma^2_{\boldsymbol{\varepsilon}},t+\tau,I\right)$*, independent of* p *and* $\sigma^2_{\boldsymbol{\varepsilon}}$ *is the contribution to the mixture heterogeneity by* $\boldsymbol{\varepsilon}$*. Furthermore,* p *has to fulfill the following boundary condition*

$$p\left(\mathbf{h}\middle|\,\sigma^2=0,t,I\right)=\prod_{j=1}^{n}\delta\left(h_j\right), \tag{6.3.25}$$

where δ *is the impulse function.*

In proving the theorem (6.1), only some general assumptions are made and therefore, in a sense, it can be considered as the fundamental theorem for mixture of particulate materials at equilibrium. Thus, all the information about the structure of the mixture in the equilibrium are encoded in the operator L.

In order to solve the heterogeneity Eqn. (6.3.22), one natural question would be if it is possible to simplify the equation by ignoring some of the higher terms of the right hand side. However, it is conceivable that by truncating the expansion after certain order, the solution might not even be positive, a condition which is demanded by the axioms of the probability theory. Indeed, in the one dimensional case, for mathematically similar equations, there exists a theorem due to Pawula (1967) which in the present context states that in order for the probability $p\left(\mathbf{h}|\,\sigma^2, t, I\right)$ to be positive, the terms on the right hand side of (6.3.22) may stop either after the first term or after the second term, if it does not stop after the second term it must contain infinite number of terms[3].

6.4 Interlude: Comments

In the most frequently encountered form of heterogeneity equation, σ^2 and h_j are analogous to time and distance, respectively. In these cases, the heterogeneity equation is known as n- dimensional Kramers-Moyal expansion. Similarly, the operator of type $L\left(\mathbf{h}, t\right)$ is known as the n- dimensional Kramers-Moyal operator with $D^{(\nu)}_{j_1,\cdots,j_\nu}\left(t\right)$ being the so-called Kramers-Moyal coefficients. It is just a matter of proper transformation of variables and therefore, mathematically the solutions are similar. These type of equations play a crucial role in statistical mechanics and some of their applications first appeared in the work of Kramers (1941) and Moyal (1949). The Kramers-Moyal coefficients with the upper indices $\nu = 1$ and $\nu = 2$ are known as *drift coefficient* or *drift vector* and *diffusion coefficient* or *diffusion matrix*, respectively (see Risken 1996, p.84). Moreover, it is evident from (6.3.24) that the diffusion matrix is a symmetric matrix which in general is semi-definite. In general, if the Eqn. (6.3.22) stops after the second order terms, the resulting heterogeneity equation is known as the n- variable Fokker-Planck equation.

Our approach and the context that it was used is very different from the most known text-book approaches (see e.g. Risken 1996, Ch.4). The derivation of the heterogeneity equation was to certain degree inspired by the work of Jaynes (2003, Ch.7) in driving the one variable Fokker-Planck equation and his proof of *central limit theorem*. Nevertheless, the first appearance of similar treatment, as Jaynes also acknowledges, is given in a paper by Landon (1941). Vernon D. Landon was an electrical engineer studying properties of noise in communication circuits. He discovered that the samples of electrical noise produced by widely different sources could not be distinguished one from the other

---

[3] This has not stopped people from exploring the non-positive solutions, see e.g. Risken and Vollmer (1987).

by any known test. This led Landon to argue that if the distribution of noise voltage is so universal, then it must be better determined theoretically than empirically. This led him to similar equation as (6.3.22). Jaynes, by applying the Bayesian's interpretation of probability was able to generalize Landon's discovery and shed some light on its significance.

6.5 Determination of The Heterogeneity Operator

The *characteristic function* of $\mathbf{h}$ with respect to $p\left(\mathbf{h}|\,\sigma^2, I\right)$ is defined as

$$\Phi\left(\mathbf{k},\sigma^2\right)=\int_{\mathbb{R}^n} e^{i\mathbf{k}'\mathbf{h}} p\left(\mathbf{h}|\,\sigma^2,I\right)d\mathbf{h} \tag{6.5.1}$$

where $\mathbf{k}'$ stands for transpose of the vector $\mathbf{k}$ and $i=\sqrt{-1}$. Similarly

$$p\left(\mathbf{h}|\,\sigma^2,I\right)=(2\pi)^{-n}\int_{\mathbb{R}^n} e^{-i\mathbf{k}'\mathbf{h}}\Phi\left(\mathbf{k},\sigma^2\right)d\mathbf{k}. \tag{6.5.2}$$

It can be shown that a pdf of a random variable is uniquely determined by its characteristic function (see for example Feller 1971, Sect. XV.7).

It turns out that $\Phi\left(\mathbf{k},\sigma^2\right)$ can be used to determine the coefficients of the heterogeneity operator. Indeed, recall that the heterogeneity equation is

$$\frac{\partial}{\partial\sigma^2}p\left(\mathbf{h}|\,\sigma^2,t,I\right)=\sum_{\nu=1}^{\infty}D^{(\nu)}_{j_1,\cdots,j_\nu}(t)\,\frac{(-\partial)^\nu}{\partial h_{j_1}\cdots\partial h_{j_\nu}}p\left(\mathbf{h}|\,\sigma^2,t,I\right) \tag{6.5.3}$$

where

$$D^{(\nu)}_{j_1,\cdots,j_\nu}(t)=\frac{1}{\nu!}\lim_{\tau\to 0}\frac{M^{(\nu)}_{j_1,\cdots,j_\nu}(\boldsymbol{\varepsilon},t+\tau)}{\sigma^2_{\boldsymbol{\varepsilon}}(t+\tau)} \tag{6.5.4}$$

in which

$$M^{(\nu)}_{j_1,\cdots,j_\nu}(\boldsymbol{\varepsilon},t+\tau)=\langle\varepsilon_{j_1}\cdots\varepsilon_{j_\nu}\rangle=\int_{-\infty}^{\infty}(\varepsilon_{j_1}\cdots\varepsilon_{j_\nu})\,q\left(\boldsymbol{\varepsilon}|\,\sigma^2_{\boldsymbol{\varepsilon}},t+\tau,I\right)d\boldsymbol{\varepsilon}. \tag{6.5.5}$$

From Eqn. (6.5.1), it is easy to show that by replacing the operator $\partial/\partial h_j$ by ik_j in equation (6.5.3), one arrives at the corresponding equation in the characteristic function of $\mathbf{h}$, which is

$$\frac{\partial}{\partial\sigma^2}\Phi\left(\mathbf{k},\sigma^2\right)=\sum_{\nu=1}^{\infty}\left(\prod_{s=1}^{\nu}k_{j_s}\right)D_{j_1,\cdots,j_\nu}^{(\nu)}\left(t\right)\left(-i\right)^{\nu}\Phi\left(\mathbf{k},\sigma^2\right). \tag{6.5.6}$$

This equation can easily be solved to yield

$$\Phi\left(\mathbf{k},\sigma^2\right)=C\exp\left\{\sum_{\nu=1}^{\infty}\left(-i\right)^{\nu}\left(\prod_{s=1}^{\nu}k_{j_s}\right)D_{j_1,\cdots,j_\nu}^{(\nu)}\left(t\right)\sigma^2\right\} \tag{6.5.7}$$

where C is a constant. Furthermore, from (6.5.1) follows that

$$\Phi\left(\mathbf{0},\sigma^2\right)=1. \tag{6.5.8}$$

Therefore, $C=1$ and hence

$$\Phi\left(\mathbf{k},\sigma^2\right)=\exp\left\{\sum_{\nu=1}^{\infty}\left(-i\right)^{\nu}\left(\prod_{s=1}^{\nu}k_{j_s}\right)D_{j_1,\cdots,j_\nu}^{(\nu)}\left(t\right)\sigma^2\right\}. \tag{6.5.9}$$

It can be shown that the mixed moments of $\mathbf{h}$ can be obtained by differentiation of the characteristic function (see for example Prohorov and Rozanov 1969, Sect.4.3), i.e.,

$$M_{r_1,\ldots,r_n}\left(\mathbf{h},t\right)=\langle h_1^{r_1}\cdots h_n^{r_n}\rangle=\left(\frac{\partial}{\partial ik_1}\right)^{r_1}\cdots\left(\frac{\partial}{\partial ik_n}\right)^{r_n}\Phi\left(\mathbf{k},\sigma^2\right)\big|_{k_1=\cdots=k_n=0} \tag{6.5.10}$$

where $M_{0,\ldots,0}=1$. From the above equation, it is evident that the mixed moments of $\mathbf{h}$ are a function of the coefficients of the heterogeneity operator, i.e. $D_{j_1,\cdots,j_\nu}^{(\nu)}\left(t\right)$. This implies that any information on the mixed moments of $\mathbf{h}$ can be converted into information on $D_{j_1,\cdots,j_\nu}^{(\nu)}\left(t\right)$ through the Eqn. (6.5.10).

Chapter 7

Application of Theory of Mixing

> All the laws of physics are not properties of the world, but arise from conventions introduced in the analysis of the data.
>
> ---
>
> Arthur Eddington, Fundamental Theory (1946)

The objective of this chapter is to demonstrate some of the direct consequences of the analysis of the previous chapter. It is shown that on the contrary to generally accepted dogma that choice of pdf for **h** being Normal is an assumption; it is a direct consequence of insufficient information on the structure of the mixture. It is also shown how inference on the mixture heterogeneity can be conducted and in the case of the pdf for **h** being Normal, Gy's estimate for mixture heterogeneity is reproduced.

7.1 Insufficient Information

Mixture heterogeneity σ^2 is not a real property of $\mathbf{h}$, but only a property of the probability distribution p that one assigns to represent one's state of knowledge about $\mathbf{h}$. As it was shown in the previous chapter, at equilibrium, the mixture heterogeneity is related to $\mathbf{h}$ through the pdf $p\left(\mathbf{h}\right|\sigma^2,t,I)$, which satisfies the heterogeneity equation

$$\frac{\partial}{\partial\sigma^2}p\left(\mathbf{h}\right|\sigma^2,t,I\left.\right)=\sum_{\nu=1}^{\infty}D^{(\nu)}_{j_1,\cdots,j_\nu}\left(t\right)\frac{\left(-\partial\right)^\nu}{\partial h_{j_1}\cdots\partial h_{j_\nu}}p\left(\mathbf{h}\right|\sigma^2,t,I\left.\right) \tag{7.1.1}$$

where

$$D^{(\nu)}_{j_1,\cdots,j_\nu}\left(t\right)=\frac{1}{\nu!}\lim_{\tau\to 0}\frac{M^{(\nu)}_{j_1,\cdots,j_\nu}\left(\boldsymbol{\varepsilon},t+\tau\right)}{\sigma^2_{\boldsymbol{\varepsilon}}\left(t+\tau\right)} \tag{7.1.2}$$

in which

$$M^{(\nu)}_{j_1,\cdots,j_\nu}\left(\boldsymbol{\varepsilon},t+\tau\right)=\langle\varepsilon_{j_1}\cdots\varepsilon_{j_\nu}\rangle=\int\limits_{-\infty}^{\infty}\left(\varepsilon_{j_1}\cdots\varepsilon_{j_\nu}\right)q\left(\boldsymbol{\varepsilon}\middle|\,\sigma^2_{\boldsymbol{\varepsilon}},t+\tau,I\right)d\boldsymbol{\varepsilon}. \tag{7.1.3}$$

However, a necessary condition for solving the general heterogeneity equation is the complete knowledge of the transition probability in the given valley, which of course depends on the amount of information available to the observer. From this point of view, any model or description of the structure of a mixture would necessarily be subjective. Nevertheless, despite the subjectivity of the description, it is still reasonable to demand that any one who is given same information about the structure of the mixture should arrive at the same conclusions. In other words, the demand is on the objectivity in drawing conclusions based on the given information and not on the information itself. In the context of the heterogeneity equation, as was demonstrated previously, this information is encoded in the coefficients of the equation through Eqn. (6.5.10), which constitute an unique pdf.

In general, when a mixture is at equilibrium, any observed fine details of the past structure of the mixture are irrelevant for predicting fine details of the future structure, but that coarser features may be expected reasonably to persists, and thus be relevant for predicting future values of mixture heterogeneity. It is like studying a landscape from a far distance in which only the general features of the landscape is visible. In some applications, this description is sufficient enough.

Cogent information on the higher mixed moments of $\boldsymbol{\varepsilon}$ might generally result in a more accurate description of the mixture structure. Thus, in order to describe the major features of the mixture structure, it is sufficient to only look at the first and second order mixed moments of $\boldsymbol{\varepsilon}$. Consequently, in Eqn. (7.1.1) for $\nu\geq 3$, it is assumed that all the mixed moments of $\boldsymbol{\varepsilon}$ are identically zero. This implies that

$$\frac{\partial}{\partial \sigma^2} p\left(\mathbf{h}\middle|\, \sigma^2, t, I\right) = -D_i^{(1)} \frac{\partial}{\partial h_i} p\left(\mathbf{h}\middle|\, \sigma^2, t, I\right) + D_{ij}^{(2)} \frac{\partial^2}{\partial h_i \partial h_j} p\left(\mathbf{h}\middle|\, \sigma^2, t, I\right). \tag{7.1.4}$$

Then by Eqn. (6.5.9), the characteristic function of $\mathbf{h}$ is

$$\Phi\left(\mathbf{k}, \sigma^2\right) = \exp\left\{-ik_r D_r^{(1)} \sigma^2 - k_r k_s D_{rs}^{(2)} \sigma^2\right\}, \tag{7.1.5}$$

which in the vector form can be written as

$$\Phi\left(\mathbf{k}, \sigma^2\right) = \exp\left\{i\mathbf{k}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{k}'\boldsymbol{\Sigma}\mathbf{k}\right\}, \tag{7.1.6}$$

where

$$\boldsymbol{\mu}' = \left(-D_1^{(1)}\sigma^2, \ldots, -D_n^{(1)}\sigma^2\right) \tag{7.1.7}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2D_{11}^{(2)}\sigma^2 & \cdots & 2D_{1n}^{(2)}\sigma^2 \\ \vdots & \ddots & \vdots \\ 2D_{n1}^{(2)}\sigma^2 & \cdots & 2D_{nn}^{(2)}\sigma^2 \end{pmatrix}. \tag{7.1.8}$$

Accordingly, it can be shown that $p\left(\mathbf{h}\middle|\, \sigma^2, t, I\right)$ is a *multivariate Normal distribution*, which is uniquely determined by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (Fang and Zhang 1990, p.43). Furthermore, if $\boldsymbol{\Sigma}$ is non-singluar, then

$$p\left(\mathbf{h}\middle|\, \sigma^2, t, I\right) = (2\pi)^{-\frac{n}{2}} \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{h} - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{h} - \boldsymbol{\mu}\right)\right\}, \tag{7.1.9}$$

where $|\boldsymbol{\Sigma}|$ denotes the determinant of the matrix $\boldsymbol{\Sigma}$.

Now, assume that the components of $\boldsymbol{\varepsilon}$ are independent. It should be emphasized that here by independence one means *logical* independence and not necessarily physical *causal* independence[1]. Indeed, any two components of $\boldsymbol{\varepsilon}$ may be in fact causally depen-

---

[1] It seems that John M. Keynes (1921, p.164) was the first author who clearly expressed this distinction

dent, i.e., one influences the other, but for an observer who has not yet discovered this or is incapable of observing this finer detail, the probabilities representing his state of knowledge are independent. This would evidently result in a more conservative estimate of the mixture structure. The independence of the components of $\boldsymbol{\varepsilon}$ implies that $\boldsymbol{\Sigma}$ is diagonal. Furthermore, in the absence of credible information on the behaviour of the first moments of the components of $\boldsymbol{\varepsilon}$, it is reasonable to assume that they are equal. Then by independence and Proposition (4.2),

$$0 = \mathbb{M}_1 \left[\mathbf{h}\left(t+\tau\right)\right] = \mathbb{M}_1 \left[\mathbf{h}\left(t\right) + \boldsymbol{\varepsilon}\left(\tau\right)\right] = \mathbb{M}_1 \left[\mathbf{h}\left(t\right)\right] + \mathbb{M}_1 \left[\boldsymbol{\varepsilon}\left(\tau\right)\right] = \mathbb{M}_1 \left[\boldsymbol{\varepsilon}\left(\tau\right)\right], \tag{7.1.10}$$

which implies that

$$\langle \varepsilon_j \rangle = 0 \text{ for all } j \tag{7.1.11}$$

and hence

$$\boldsymbol{\mu} = \mathbf{0}. \tag{7.1.12}$$

Thus the equality and independence of the components of $\boldsymbol{\varepsilon}$ imply that each one of them being equally likely to be positive as negative. Otherwise, one would have, for each component, a systematic build up resulting in a non-zero drift vector, which is certainly not true in the present case.

Next, it is important to determine the rank of the matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma}$ is singular then the Eqn. (7.1.4) would be an *underdetermined* equation which in this case would have infinitely many solutions. In order to avoid disagreement among observers, more detail structural information on the mixture structure is needed in order to settle all disputes. On the other hand, if $\boldsymbol{\Sigma}$ is *positive definite* which in this case basically means that the diagonal entries are greater than zero, Eqn. (7.1.4) has an unique solution[2]. Therefore, no additional information is necessary. This is exactly in accordance with the aforementioned

---

. See also Jaynes (2003, p.92) for further discussions.

[2] Note that $\boldsymbol{\Sigma}$ can not be negative definite. Since, the second moment of a quantity is always non-negative.

argument on the relation between the major features of the mixture structure and amount of information available to the observer. One needs less amount of information in order to describe the coarser as compared to finer structural details. Further, if one assumes that the second moments of the components of $\boldsymbol{\varepsilon}$ are all equal, i.e.

$$\left\langle \varepsilon_i^2 \right\rangle = \left\langle \varepsilon_j^2 \right\rangle \text{ for all } i \text{ and } j, \tag{7.1.13}$$

then it is easy to show that

$$\sigma_{\boldsymbol{\varepsilon}}^2 = \left\langle \varepsilon_j^2 \right\rangle \text{ for all } j \tag{7.1.14}$$

and hence

$$\Sigma_{ij} = \sigma^2 \delta_{ij}, \tag{7.1.15}$$

in which

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \tag{7.1.16}$$

is the *Kronecker delta* function (Arfken 2001, p.11). In other words, Eqn. (7.1.9) attains a simpler form,

$$p\left(\mathbf{h} \middle| \sigma^2, t, I\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} h_i^2\right), \tag{7.1.17}$$

which is the multivariate Normal distribution in n independent and identically distributed parameters of zero mean and variance of σ^2.

The analysis in this section demonstrates that, up to the second order, the distribution of the heterogeneity function attains a Normal form at equilibrium. From this point on, the mixing process can be stopped at any time and as long as the mixture is at equilibrium, the resulting distribution is still Normal. There is also another aspect of this analysis which sheds some light on the long standing debate on Normal distribution being the right choice for description of the mixture structure. Indeed, in general, the Normal distribution, up to the second moment, adequately describes the mixture structure at equilibrium. Moreover, in the event of insufficient information about the mixture structure, the above analysis seems to suggest (but far from proof) that the choice of Normal distribution results in a more conservative estimate of the structural parameters as compared to the case that a more detailed knowledge about the mixture structure is available. This observation could be very valuable, especially for pharmaceutical companies for which there are stringent demands on the quality of the mixtures. Further investigation on this issue is necessary and is left to a later time.

7.2 Inference on Mixture Heterogeneity

In general the value of the mixture heterogeneity is unknown. The objective of this section is to demonstrate how one can estimate its true value based on the available data on the heterogeneity function **h**. The basic mathematical tools for this purpose are described in Chapter 3.

In a sense, given any information on the distribution of the heterogeneity function **h**, one has to reason backward, i.e., from the pdf assigned to **h**, to a pdf for mixture heterogeneity σ^2. This so-called "inverse" reasoning can be achieved by simple application of Bayes' theorem, which implies that

$$p\left(\sigma^2\middle|\,\mathbf{d},t,I\right) = Np\left(\mathbf{d}\middle|\,\sigma^2,t,I\right) \times p\left(\sigma^2\middle|\,t,I\right) \tag{7.2.1}$$

where N is the normalization constant and **d** is the available data. The quantity on the far right, $p\left(\sigma^2\middle|\,t,I\right)$, is the prior probability; it represents the state of knowledge about the mixture heterogeneity before the observer has analyzed the data. This is modified by the experimental measurements through the likelihood function, $p\left(\mathbf{d}\middle|\,\sigma^2,t,I\right)$, and yields the posterior probability, $p\left(\sigma^2\middle|\,\mathbf{d},t,I\right)$, representing the observers state of knowledge about the mixture heterogeneity in the light of the data. The most probable estimate for σ^2, given the measurement vector **d** is the one which maximizes the posterior probability.

That is

$$\frac{\partial}{\partial \sigma^2} p\left(\sigma^2 \middle| \mathbf{h}, t, I\right)\Big|_{\mathbf{h}=\mathbf{d}} = 0 \tag{7.2.2}$$

or equivalently

$$\frac{\partial}{\partial \sigma^2}\left[p\left(\mathbf{h} \middle| \sigma^2, t, I\right) \times p\left(\sigma^2 \middle| t, I\right)\right]\Big|_{\mathbf{h}=\mathbf{d}} = 0. \tag{7.2.3}$$

Now, for the sake of the argument, assume that the heterogeneity function has a multivariate Normal distribution in n independent and identically distributed parameters of zero mean and variance of σ^2, i.e.,

$$p\left(\mathbf{h} \middle| \sigma^2, t, I\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} h_i^2\right). \tag{7.2.4}$$

Moreover, in the absence of credible information on σ^2, it is reasonable to assume that $p\left(\sigma^2 \middle| t, I\right)$ is uniformly distributed, i.e.

$$p\left(\sigma^2 \middle| t, I\right) = \begin{cases} \frac{1}{\sigma^2_{\max} - \sigma^2_{\min}} & \sigma^2_{\min} \le \sigma^2 \le \sigma^2_{\max} \\ 0 & \text{Otherwise} \end{cases}. \tag{7.2.5}$$

Then from Eqn. (7.2.3) follows that

$$\frac{1}{2\sigma^4}\left(n\sigma^2 - \sum_{i=1}^{n} d_i^2\right) p\left(\mathbf{d} \middle| \sigma^2, t, I\right) = 0, \tag{7.2.6}$$

which implies that

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} d_i^2. \tag{7.2.7}$$

The above expression for mixture heterogeneity was first introduced by Gy (1979, p.218). In his work, whenever samples only consist of one particle each, σ^2 is called *Constitutional heterogeneity* and *Distributional heterogeneity* otherwise. From the aforementioned analysis one can deduce the following:

1. Any estimate for the mixture heterogeneity σ^2 has to fulfil the Eqn. (7.2.3).

2. At equilibrium, the likelihood function is a solution of heterogeneity equation.
3. In the absence of sufficient information on the mixture structure at equilibrium, Eqn. (7.2.7) represents, up to the second moment, the best estimate for mixture heterogeneity.

In general, one should be able by combining the solution of the heterogeneity equation, the prior probability on σ^2 and the Eqn. (7.2.3), to produce the best estimate for mixture heterogeneity at equilibrium. However, in many cases it might be difficult to find an expression in closed form for this estimate. In these cases a numerical approach would be more suitable.

The choice of the prior pdf in (7.2.5) is just a way of encoding a lot of initial ignorance about the mixture heterogeneity. Since it was constant and uniformly distributed, it could be cancelled out of the Eqn. (7.2.3) and therefore did not play part in calculation of the estimate for mixture heterogeneity. However, a fair question is how the estimate (7.2.7) would have changed if one had chosen a different prior. In general, it is intuitively evident that as the empirical evidence grows, one should eventually arrive at the same conclusions irrespective of the initial beliefs. This means that the posterior pdf is then dominated by the likelihood function, and the choice of prior becomes largely irrelevant. Accordingly, the choice of prior is quite important whenever the number of trials are small, which is usually the case in practice. For a more general and illustrative discussion of this issue the reader is referred to Sivia (1996, p.15).

Chapter 8

Concluding Words

> Scientific progress never achieves finality; it is a method of succesive approximation.
>
> ---
>
> Harold Jeffreys, Scientific Inference (1931)

8.1 Concluding Remarks

The main objective of this work was to establish a theoretical framework within which a unifying scientific approach to the subject was possible. It was argued that this theoretical framework has to fulfil the irrationalability criteria, which basically means that the statements of the theory should be in such logical form that they can be shown to be irrational by experience. This paved the way for development of a probabilistic theory in which all the statements fulfil the irrationalability criteria.

Our approach to study of the mixtures can basically be divided into two main categories. The first approach (see Sections 4.3-4.5) was based on simply counting the number of the particles of interest in each sample. Using some simple combinatorial arguments, led us to a mathematical theory for sampling of particulate materials. This also resulted in a new measure for mixedness which could be modelled by entropy of the sample distribution. Based on these new findings, a new unifying approach to modelling of RTD of continuous systems is developed (see Section 5.6). Furthermore, this new mathematical approach to sampling theory not only confirms what was previously known, it also resulted in new predictions, which was unknown until now. Indeed, it allowed us, for example, to predict the number of key components of given size in binary closed batch systems (see Section 5.3). In the case of the continuous systems, the theory also gives a compelling explanation for deviation of real systems from the ideal exponential systems. As a result of this analysis a new measure for axial mixing was discovered.

The Second approach was based on characterizing the mass ratio of the particles of interest in each sample. This led us to Gy's definition of the heterogeneity. Based on this, we developed a mathematical theory which relates the sample heterogeneity to mixture heterogeneity (see Section 4.6). In developing this theory, we demanded that the theory should not violate common sense. In our case, this common sense was formulated by Lacey's conjecture. Indeed, within the framework of our mathematical theory it was shown that the Lacey's conjecture is consistent, independent of the shape of the samples (see theorem 5.1). This is a much stronger statement than any ever made by existing theories. The reason for this lies in the nature of this statement. Indeed, no amount of empirical evidence can ever demonstrate its validity and therefore should be considered as the first of its kind. Following the same line of thought, we were able to relate the mixture heterogeneity of a mixture to an abstract object which we have called the heterogeneity landscape (see § 6.1.2). The advantage of this abstract approach was that it made it possible by applying the basic rules of the probability theory to give a mathematical description of all the possible equilibrium states of a mixture (see Section 6.3). This new mathematical theory led us to rediscovery of known facts which previously were only taken to be true by assumption (see Section 7.2).

8.2 Future Work

This work is far from being complete. We merely managed to get beyond the basic definitions and concepts. However, we hope that we managed to demonstrate some of the potentials of this approach. Here are some suggestions for future projects:

1. It is important to determine a possible relation between the entropy of the sample

distribution and the heterogeneity equation describing the dynamics of the mixture at equilibrium. We believe that this should lead to discovery of similar laws as the laws of thermodynamics.

2. Due to lack of right instruments, it was not possible for us to experimentally verify the prediction of the theory on the number of the key components of given size in binary closed batch systems. As it was also shown in Section 5.3, the parameters of this estimate depend on the microstructural properties of the mixture under study. Further investigation is needed to determine the nature of these parameters. This we believe should serve as the foundation of a theory of statistical mechanics for particulate materials.

3. In § 2.4.5, it was argued that the RTDs determined from the dispersion model are similar to RTDs of serial cascade of exponential vessels presented in § 2.4.1. Now, in Section 5.6, it was shown that the ratio of the mean and geometric mean residence time of a continuous mixer is a function of order of the mixer which is also equal to the number of exponential vessels in serial cascade. This observation implies that there should be a relation between the Peclet number and hence the dispersion coefficient and the geometric mean residence time of a continuous system. Further investigation is needed to determine the nature of this relation.

4. The general solution of the heterogeneity equation could only be given in the form of its characteristic function. There is a class of pdfs known as *Stable distributions*. It seems to us that the solution of the heterogeneity equation could serve as a generalization of these types of distributions. Further investigation is necessary in order to verify this. For more readings about Stable distribution, we refer to Schneider (1986) and references therein.

We believe that answer to above questions should result into further development of the theory and a better understanding of complex systems like particulate solids.

Appendix A

On a Problem related to the Mean Residence Time

A major topic of Applied Mathematics that deals with the phenomenon of waiting is called Queueing theory. Queueing theory arises from the use of powerful mathematical analysis to theoretically describe production processes along with statistical/probabilistic techniques to account for varying dynamic patterns within the stages of a productive process. The origins of the formal study of Queueing theory is credited to A. K. Erlang, a Danish telephone engineer who in the 1920's was attempting to predict telephone call service[1].

One of the foundations of Queueing theory is the *Little's law* which is expressed as

$$L = \lambda W \tag{A.0.1}$$

where λ is the average arrival rate, W is the average time a customer spends in the system, and L is the average number of customers in the system. This result was first proved by Little (1961). Little's proof, despite the fact that Eqn. (A.0.1) is easy to state and intuitively reasonable, was difficult. Later Stidham (1974) published a simpler proof that is quite general and more intuitive than Little's proof. Stidham's proof also confirmed the intuition of most researchers that Little's law is essentially a deterministic relation and that the probabilistic assumptions imposed in previous proofs are necessary

---

[1] As a matter of fact most of the model work in § 2.4.1 can be deduced directly from the results in Queueing theory.

only for the existence of the relevant limits, but not for the relation itself. For the sake of completeness, Stidham's version of Little's theorem is stated here without proof.

Theorem A.1 *Let $L(t)$ be the number of customers present at time t. Define L by*

$$L = \lim_{t\to\infty} \frac{1}{t} \int_0^t L(\tau)\, d\tau \tag{A.0.2}$$

and λ by

$$\lambda = \lim_{t\to\infty} \frac{N(t)}{t} \tag{A.0.3}$$

where $N(t)$ is the number of customers who arrive in the interval $[0, t]$. Let W_i be the time in the system for the ith customer and define the mean time in the system W by

$$W = \lim_{n\to\infty} \frac{1}{n} \sum\nolimits_{i=1}^{n} W_i \tag{A.0.4}$$

If λ and W exist and are finite, then so does L, and $L = \lambda W$.

Now, a continuous mixer can be considered as a queueing system. Each particle plays the role of a customer which enters the system through the inlet, processed by the mixer and departs the mixer via the outlet. Therefore, the Eqn. (A.0.1) is also valid in case of continuous mixers. Evidently, the number of particles is proportional to their volume. Substituting this in (A.0.1) and replacing λ by corresponding average volumetric flow rate, one arrives at

$$\bar{V} = \bar{u}_i W = \bar{u}_o W, \tag{A.0.5}$$

where $\bar{u}_i$ and $\bar{u}_o$ are average inlet and outlet volumetric flow rate, respectively. The conservation of mass implies that average inlet flow rate should be equal to average outlet flow rate (from the moment the mixing vessel is brought (empty) on stream till it is retired (empty) from service). Furthermore, if it is assumed that there is a constant dead volume in the mixer, say V_0, then for the case of continuous mixers one gets

$$\bar{V} = V_0 + \bar{u_o}\,\bar{t}, \tag{A.0.6}$$

which confirms all the observations made by Shinnar *et al.* (1969).

Bibliography

Abou-Chakra, H. and Tüzün, U. 1999. Coefficient of friction of binary granular mixtures in contact with a smooth wall, *Chemical Engineering Science* **54**(24): 5901–5912.

Abouzeid, A. Z. M. 1989. Mixing and demixing of particulate solids, *Powder Handling Processing* **1**(2): 173–177.

Allen, A. O. 1990. *Probability, Statistics and Queueing Theory with Computer Science Applications*, second edn, Academic Press Inc.

Arfken, G. B. and Weber, H. J. 2001. *Mathematical Methods for Physicists*, fifth edn, Harcourt/Academic Press.

Arnold, B. C. and Huang, J. S. 1995. Characterizations, *in* N. Balakrishnan and A. P. Basu (eds), *The Exponential Distribution, Theory, Methods and Applications*, Gordon and Beach Publishers, pp. 185–203.

Arteaga, P. and Tüzün, U. 1990. Flow of binary mixtures of equal-density granules in hoppers: Size segregation, flowing density and discharge rates, *Chemical Engineering Science* **45**(1): 205–223.

Ashton, M. D. and Valentin, F. H. H. 1966. The mixing of powders and particles in industrial mixers, *Transactions of Institution of Chemical Engineers* **44**: T166–T188.

Bak, P. 1997. *How Nature Works: The Science of Self-Organized Criticality*, Oxford University Press.

Bak, P., Tang, C. and Weisenfeld, K. 1988. Self-organized criticality, *Physical Review* **A38**: 364–374.

Baxter, J. 1998. *Mechanics of Granular Heaps*, PhD thesis, University of Surrey.

Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London* **53**: 370–418.

Bhattacharya, P. B., Jain, S. K. and Nagpaul, S. R. 1994. *Basic Abstract Algebra*, second edn, Cambridge University Press.

Bourne, J. R. 1964. The mixing of powders, pastes, and non-newtonian fluids, *The Chemical Engineer: The Bulletin of the Institution of Chemical Engineers* **181**: CE202–CE216.

Bourne, J. R. 1967. Variance-sample size relationships for incomplete mixtures, *Chemical Engineering Science* **22**(4): 693–700.

Bourne, J. R. 1968. Statistical relations for square samples, *Chemical Engineering Science* **23**(4): 339–343.

Braithwaite, R. B. 1953. *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*, Cambridge, The University Press.

Bretthorst, G. L. 1996. An introduction to model selection using probability theory as logic, *in* G. R. Heidbreder (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, Dordrecht, pp. 1–42.

Bridgewater, J. 1976. Fundamental powder mixing mechanisms, *Powder Technology* **15**(2): 215–236.

Broadbent, C. J., Bridgwater, J., Parker, D. J., Keningley, S. T. and Knight, P. 1993. A phenomenological study of a batch mixer using a positron camera, *Powder Technology* **76**(3): 317–329.

Brunton, D. W. 1895. The theory and practice of ore sampling, *Transactions of the American Institute of Mining Engineers* **25**: 836–844.

Burtally, N., King, P. J. and Swift, M. R. 2002. Spontaneous air-driven separation in vertically vibrated fine granular mixtures, *Science* **295**: 1877–1879.

Campbell, A. P. and Bridgwater, J. 1973. The mixing of dry solids by percolation, *Transactions of the Institution of Chemical Engineers* **51**(1): 72–74.

Campbell, H. and Bauer, W. C. 1966. Cause and cure of demixing in solid-solid mixers, *Chemical Engineering* **73**: 179–185.

Carley-Machauly, K. W. and Donald, M. B. 1962. The mixing of solids in tumbling mixers-i, *Chemical Engineering Science* **17**: 493–506.

Carstensen, J. T. and Patel, M. R. 1977. Blending of irregularly shaped particles, *Powder Technology* **17**(3): 273–282.

Clausius, R. 1865. On several convenient forms of fundamental equations of machanical theory of heat, *in* T. A. Hirst (ed.), *The Mechanical Theory of Heat: With Application to the Steam Engine and to Physical Properties of Bodies*, John Van Voorst, 1 Patemoster Row, chapter 9, pp. 327–365.

Cooke, M. H. and Bridgwater, J. 1977. The relationship between variance and sample size for mixtures, *Chemical Engineering Science* **32**: 1353–1357.

Courtney, L. H. 1895. To my fellow-disciples at saratoga springs, *The National Review* **26**: 21–26. London.

Cox, R. T. 1946. Probability, frequency and reasonable expectation, *American Journal of Physics* **14**: 1–13.

Cox, R. T. 1961. *The Algebra of Probable Inference*, The Johns Hopkins Press, Baltimore.

Danckwerts, P. V. 1953. Continuous flow systems, distribution of residence times, *Chemical Engineering Science* **2**(1): 1–13.

de Silva, S. 1997. Mixing and segregation in industrial processes, *Technical Report 410037-1*, Tel-Tek, Kjølnes Ring, N-3918, Porsgrunn, Norway.

Dougherty, E. R. 1990. *Probability and Statistics for the Engineering, Computing, and Physical Sciences*, Prentice-Hall Inc.

Duran, J. 2000. *Sands, Powders, and Grains: An Introduction to the Physics of Granular Materials*, Partially Ordered Systems, Springer-Verlag, New York.

Dury, C. M. and Ristow, G. H. 1999. Competition of mixing and segregation in rotating cylinders, *Physics of Fluids* **11**(6): 1387–1394.

Edmonds, B. 1996. What is complexity?- the philosophy of complexity per se with application to some examples in evolution, *in* F. Heylighen and D. Aerts (eds), *The Evolution of Complexity*, Kluwer Academic Publishers, Dordrecht.

Egermann, H. 1980. Suggestions on the nomenclature of powder mixtures, *Powder Technology* **26**(2): 235–237.

Egermann, H. and Orr, N. A. 1983. Ordered mixtures- interactive mixtures, *Powder Technology* **36**(1): 117–118.

EIA 2005. International energy annual 2003, *Technical report*, Energy Information Administration, http://www.eia.doe.gov.

Fan, L. T., Chen, S. J. and Watson, C. A. 1970. Solids mixing, *Industrial and Engineering Chemistry* **62**(7): 53–69.

Fan, L. T., Chen, Y. and Lai, F. S. 1990. Recent developments in solid mixing, *Powder Technology* **61**(3): 255–287.

Fan, L. T., Too, J. R., Rubison, R. M. and Lai, F. S. 1979. Studies on multicomponent solids mixing and mixtures, part III. mixing indices, *Powder Technology* **24**(1): 73–89.

Fan, L. T. and Wang, R. H. 1974. Methods for scaling-up tumbling mixers, *Chemical Engineering* pp. 88–94.

Fan, L. T. and Wen, C. Y. 1975. *Models for Flow Systems and Chemical Reactors*, Vol. 3 of *Chemical Processing and Engineering*, Marcel Dekker Inc.

Fang, K.-T. and Zhang, Y.-T. 1990. *Generalized Multivariate Analysis*, Springer-Verlag.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn, John Wiley Sons, New York.

Fick, A. E. 1855. über diffusion, *Poggendorff's Annalen der Physike und Chemie* **94**: 59–86.

Fokker, A. D. 1914. Die mittlere energie rotierender elektrischer dipole im strahlungsfeld, *Annalen der Physik* **43**: 810–820.

Forte, B. and Sempi, C. 1976. Maximising conditional entropy, *Rend Condi de Mathematics* **9**: 551–566.

Furnas, C. C. 1931. Grading aggregates I: Mathematical relations for beds of broken solids of maximum density, *Industrial and Engineering Chemistry* **23**(9): 1052–1058.

Gentile, G. 1940. Osservazioni sopra le statistiche intermedie, *Nuovo Cimento* **17**: 493.

Ghaderi, A. 2003. On characterization of continuous mixing of particulate materials, *Particulate Science and Technology* **21**(3): 271–282.

Ghaderi, A. and Naqvi, R. 1995. Just SO(2) stories: How Lie theory was applied to a problem with great past, *Symmetri* **2**(3): 45–49.

Gibbs, J. W. 1902. *Elementary Principles in Statistical Mechanics*, Yale University Press, New Haven, Connecticut.

Gillispie, C. C. 1997. *Pierre-Simon Laplace, 1749-1827: A Life in Exact Science*, Princeton University Press.

Gladwell, M. 2002. *The Tipping Point: How Little Things Can Make a Big Difference*, 2nd edn, Back Bay Books.

Good, I. J. 1983. *Good Thinking; The Foundations of Probability and Its Applications*, The University of Minnesota Press.

Gould, S. J. and Eldredge, N. 1977. Punctuated equilibria: The tempo and mode of evolution reconsidered, *Paleobiology* **3**: 115–151.

Gradshteyn, I. S. and Ryzhik, I. M. 2000. *Table of Integrals, Series, and Products*, sixth edn, Academic Press, San Diego.

Grassberger, P. 1989. Problems in quantifying self-organized complexity, *Helvetica Physica Acta* **62**: 498–511.

Gregory, P. C. 2005. *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach With Mathematica® Support*, first edn, Cambridge University Press, Cambridge.

Gregory, P. C. and Loredo, T. J. 1992. A new method for the detection of a periodic signal of unknown shape and period, *Astrophysical Journal* **398**: 146–168.

Gy, P. M. 1979. *Sampling of Particulate Materials: Theory and Practice*, Developments in Geomathematics, Elsevier Scientific Publication Compony, Amsterdam.

Gy, P. M. 1998. *Sampling for Analytical Purposes*, John Wiley Sons.

Harnby, N. 1967. A comparison of the performance of industrial solids mixers using segregating materials, *Powder Technology* **1**(2): 94–102.

Harnby, N. 1985. The mixing of cohesive powders, *in* N. Harnby, M. Edwards and A. Nienow (eds), *Mixing in the Process Industries*, Butterworth Co (Publishers) Ltd, pp. 78–94.

Heylighen, F. 1992. Distinction dynamics: From mechanical to self-organizing evolution, *in* E. Gindev (ed.), *Proceedings of the International Workshop on Analysis and Control of Dynamical Systems*, Bulgarian Academy of Sciences, CLCS, Sofia, http://pespmc1.vub.ac.be/papers/DistinctionDynamics.html.

Heylighen, F. 1996. What is complexity?, *Principia Cybernetica Web*, Brussels, http://pespmc1.vub.ac.be/complex1.html.

Heylighen, F. 1997. Self-organization, *Principia Cybernetica Web*, Brussels, http://pespmc1.vub.ac.be/selforg.html.

Heylighen, F. 1998. Attractors, *Principia Cybernetica Web*, Brussels, http://pespmc1.vub.ac.be/attracto.html.

Heywood, H. 1961. Techniques for the evaluation of powders I.-Fundamental properties of particles and methods of sizing analysis, *Powder Metallurgy* **7**: 1–28.

Hume, D. 1740. *A Treatise of Human Nature*, Oxford Philosophical Texts, Oxford University Press, 2000.

Hwang, C. L. and Hogg, R. 1980. Diffusive mixing in flowing powders, *Powder Technology* **26**: 93–101.

Jaynes, E. T. 1957a. How does the brain do plausible reasoning?, *in* G. J. Erickson and C. R. Smith (eds), *Maximum-Entropy and Bayesian Methods in Science and Engineering, Vol.1*, Kluwer Academic Publishers, Dordrect, 1988, pp. 1–24.

Jaynes, E. T. 1957b. Information theory and statistical mechanics I, *Physical Review* **106**: 620–630.

Jaynes, E. T. 1957c. Information theory and statistical mechanics II, *Physical Review* **108**: 171–190.

Jaynes, E. T. 1967. Foundations of probability theory and statistical mechanics, *in* I. M. Munge (ed.), *Delaware Seminar on the Foundations of Science, Vol. I*, Springer Verlag, pp. 77–101.

Jaynes, E. T. 1968. Prior probabilities, *IEEE Transactions on Systems Science and Cybernetics* **4**(3): 227–241.

Jaynes, E. T. 1982. On the rational of maximum-entropy methods, *Proceedings of The IEEE* **70**(9): 939–952.

Jaynes, E. T. 1983. Concentration of distributions at entropy maxima, *in* R. D. Rosenkrantz (ed.), *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Kluwer Academic Publishers, Dordrecht, chapter 11, pp. 315–336.

Jaynes, E. T. 2003. *The Probability Theory: The Logic of Science*, Cambridge University Press.

Jeffreys, H. 1973. *Scientific Inference*, third edn, Cambridge University Press.

Jeffreys, H. 1998. *Theory of Probability*, third edn, Oxford University Press.

Johnson, N. L., Kotz, S. and Balakrishnan, N. 1995. Genesis, *in* N. Balakrishnan and A. P. Basu (eds), *The Exponential Distribution, Theory, Methods and Applications*, Gordon and Beach Publishers, pp. 1–6.

Kapur, J. 1989. *Maximum-Entropy Models in Science and Engineering*, Wiley Eastern Limited, New Delhi.

Kapur, J. N. 1983. Non-additive measures of entropy and distributions of statistical mechanics, *Indian journal of pure and applied mathematics* **14**(11): 1372–1384.

Kapur, J. N. 1994. *Measures of Information and Their Applications*, Wiley Eastern Limited, New Delhi, India.

Kapur, J. N. and Kesavan, H. K. 1992. *Entropy Optimization Principles with Applications*, Academic Press, INC.

Kauffman, S. A. 1993. *The Origins of Order*, Oxford University Press, Oxford.

Kaye, B. H. 1997. *Powder Mixing*, first edn, Chapman Hall.

Kestelman, H. 1960. *Modern Theories of Integration*, second rev. edn, Dover.

Keynes, J. M. 1921. *A Treatise on Probability*, Macmillan and Co.

Kramers, H. A. 1941. Brownian motion in a field of force and the diffusion model of chemical reactions, *Physica* **7**: 284–304.

Kristensen, G. H. 1973. Statistical properties of random and non-random mixtures of dry solids. part I. A general expression for the variance of the composition of samples, *Powder Technology* **7**(5): 249–257.

Kullback, S. and Leibler, R. A. 1951. On information and sufficiency, *Annals of Mathematical Statistics* **22**: 79–86.

Lacey, P. M. C. 1943. The mixing of solid particles, *Transactions of the Institution of Chemical Engineers* **21**: 53–59.

Lacey, P. M. C. 1954. Developments in the theory of particle mixing, *Journal of Applied Chemistry* **4**: 257–268.

Lai, F. and Hersey, J. A. 1981. Variance-sample size relationship and the effects of magnesium stearate on ordered powder mixtures, *Chemical Engineering Science* **36**(7): 1133–1137.

Landon, V. D. 1941. The distribution of amplitude with time in fluctuation noise, *Proceedings of the Institute of Radio Engineers* **29**: 50–55.

Laplace, P. S. 1814. *A Philosophical Essay on Probabilities*, Dover Publications, 1995. The french version "Essai Philosophique sur les Probabilités" was first published in 1814.

Lavenda, B. H. and Dunning-Davies, J. 1989. The case against intermediate statistics, *Journal of Mathematical Physics* **30**(5): 1117–1121.

Little, J. D. C. 1961. A proof for the queuing formula: L= λw, *Operations Research* **9**(3): 383–387.

Loomis, L. H. and Sternberg, S. 1990. *Advanced Calculus*, Jones and Bartlett Publishers.

Loredo, T. J. 1990. From laplace to supernova SN 1987a: Bayesian inference in astrophysics, *in* P. F. Fougére (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publisher, Dordrecht, pp. 81–142.

Matheron, G. 1975. *Random Sets and Integral Geometry*, Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Mehta, M. L. 2004. *Random Matrices*, Vol. 142 of *Pure and Applied Mathematics Series*, 3 edn, Elsevier, Amsterdam.

Metcalfe, G., Shinbrot, T., McCarthy, J. J. and Ottino, J. M. 1995. Avalanche mixing of granular solids, *Nature* **374**: 39–41.

Moyal, J. E. 1949. Stochastic processes and statistical physics, *Journal of Royal Statistical Society B(Methodological)* **11**: 150–210.

Müller, W. 1966. *Untersuchungen über Mischzeit, Mischgüte und Arbeitsbedarf in Mischtrommeln mit Rotierenden Mischelementen*, PhD thesis, Karlsruhe Technical College.

Onishchik, A. L. 1993. *Lie Groups and Lie Algebras I*, Springer-Verlag, Berlin.

Orr, N. 1979. Assessment of an ordered mix, *Powder Technology* **24**(1): 105–106.

Pawula, R. F. 1967. Approximation of the linear boltzmann equation by the fokker-planck equation, *Physical Review* **162**: 186–188.

Pearson, E. S. 1922. A further development of tests for normality, *Biometrika* **22**: 239–249.

Planck, M. 1917. Über Einen Satz der Statistischen Dynamik und Seine Erweiterung in der Quantentheorie, *Sitzungsberichte der Preußischen Akademie der Wissenschaften* **24**: 324–341.

Poole, K. R., Taylor, R. F. and Wall, G. P. 1964. Mixing powders to fine-scale homogeneity: Studies of batch mixing, *Transactions of Institution of Chemical Engineers* **42**: T305–T315.

Popper, K. 2002. *The Logic of Scientific Discovery*, Routledge Classics, 6 edn, Routledge. This book was first published in Germany in 1935 under the title "Logik der Forschung".

Poux, M., Fayolle, P., Bertrand, J., Bridoux, D. and Bousque, J. 1991. Powder mixing: Some practical rules applied to agitated systems, *Powder Technology* **68**(3): 213–234.

Prigogine, I. and Stengers, I. 1984. *Order Out of Chaos; Man's New Dialogue with Nature*, Bantam Books, New York.

Prohorov, Y. V. and Rozanov, Y. A. 1969. *Probability Theory: Basic Concepts, Limit Theorems, Random Processes*, Vol. 157 of *Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen*, Springer-Verlag, Berlin.

Ramsey, F. P. 1931. Theories, *in* R. B. Braithwaite (ed.), *The Foundations of Mathematics and Other Logical Essays*, Routledge Keegan Paul, London.

Risken, H. 1996. *The Fokker-Planck Equation, Methods of Solution and Applications*, third edn, Springer-Verlag, Berlin.

Risken, H. and Vollmer, H. D. 1987. On solution of truncated kramers moyal-expansions; continuum approximations to the poission process, *Z. Physik* **B66**: 257–262.

Rose, H. E. 1959. A suggested equation relating to the mixing of powders and its application to the study of the performance of certain types of machines, *Transactions of the Institution of Chemical Engineers* **37**(4): 47–64.

Schelling, T. C. 1971. Dynamic models of segregation, *Journal of Mathematical Sociology* **1**: 143–186.

Schneider, W. R. 1986. Stable distributions: Fox function representation and generalization, *in* S. Albeverio, G. Casati and D. Merlini (eds), *Stochastic Processes in Classical and Quantum Systems*, Ascona, Switzerland 1985, Springer-Verlag, Berlin, pp. 497–511.

Schofield, C. 1975. *The Continuous Mixing of Particulate Solids*, PhD thesis, Loughborough University of Technology.

Scott, A. M. and Bridgwater, J. 1974. The characterization of patchy mixtures, *Chemical Engineering Science* **29**(8): 1789–1800.

Shannon, C. E. 1948. A mathematical theory of communication, *The Bell System Technical Journal* **27**: 379–423 and 623–656.

Shinbrot, T., Khakhar, D. V., McCarthy, J. J. and Ottino, J. M. 1997. Role of voids in granular convection, *Physical Review E* **55**(5): 6121–6133.

Shinnar, R., Kats, S. and Krambeck, F. 1967. Stochastic mixing models for chemical reactors, *Industrial Engineering Chemistry, Fundamentals* **6**(2): 276–288.

Shinnar, R., Kats, S. and Krambeck, F. 1969. Interpretation of tracer experiments in systems with fluctuating throughput, *Industrial Engineering Chemistry, Fundamentals* **8**(3): 431–441.

Shinnar, R. and Naor, P. 1963. Representation and evaluation of residence time distributions, *Industrial Engineering Chemistry, Fundamentals* **2**(4): 278–286.

Shinnar, R., Naor, P. and Kats, S. 1972. Interpretation and evaluation of multiple tracer experiments, *Chemical Engineering Science* **27**: 1627–1642.

Shore, J. E. and Johnson, R. W. 1980. Axiomatic derivation of the principle of maximum entropy and the principle of maximum cross-entropy, *IEEE Transactions on Information Theory* **IT-26**(1): 26–37.

Sivia, D. S. 1996. *Data Analysis: A Bayesian Tutorial*, Clarendon Press, Oxford.

Sommer, K. 1979. Powder mixing mechanisms, *Journal of Powder Bulk Solids Technology* **3**(4): 2–9.

Sommer, K. 1986. *Sampling of Powders and Bulk Materials*, Springer-Verlag, Berlin.

Sommer, K. 1996. Mixing of particulate solids, *KONA* **14**: 73–78.

Stidham, S. 1974. A last word on L= λw, *Operations Research* **22**(2): 417–421.

Strang, G. 1988. *Linear Algebra and Its Applications*, third edn, Harcourt Brace Jovanovich, Inc.

Strek, F., Rochowiecki, A. and Karcz, J. 1978. A mathematical model of mixing of particulate solids, *Powder Technology* **20**(2): 243–248.

Tadmor, Z. and Klein, I. 1970. *Engineering Principles of Plasticating Extruders*, Van Nostrand Co., New York.

Thorburn, W. M. 1918. The myth of occam's razor, *Mind* **27**: 345–353.

Todd, D. and Baumann, D. 1978. Twin screw reinforced plastics compounding, *Polymer Engineering and Science* **18**(4): 321–325.

Too, J. R., Fan, L. T., Rubison, R. M. and Lai, F. S. 1980. Applications of nonparametric statistics of multicomponent solids mixing, *Powder Technology* **26**(2): 131–146.

Trevena, D. H. 1993. *Statistical Mechanics: An Introduction*, Ellis Horwood Series in Physics and Its Application, Ellis Horwood Limited.

Tribus, M. 1961. *Thermostatics and Thermodynamics; An Introduction to Energy, Information and States of Matter, with Engineering Applications*, University Series in Basic Engineering, D. Van Nostrand Company, INC., New York.

Tucker, C. L. 1981. Sample variance measurement of mixing, *Chemical Engineering Science* **36**(11): 1829–1839.

Twain, M. 2000. *The Autobiography of Mark Twain*, Perennial Classics. In American literature, this popular remark is usually credited to Mark Twain, due he attributed this to a 19th century politician, Benjamin Disraeli, the 1st Earl of Beaconsfield (see p. 246). However, to the best of our knowledge, this remark first appeared in an article by Leonard Courtney (1895).

Uspensky, J. V. 1937. *Introduction to Mathematical Probability*, McGraw-Hill Book Company Inc.

van Heusden, K. 1993. New continuous mixer from hosokawa, *Bulk Solids Handling* **13**(1): 87–90.

Wang, R. H. 1986. Residence time distribution models for continuous solids mixer, *Proceedings of the 11th Annual Powder and Bulk Solids Handeling Conference*, Rosemont, USA, pp. 8–17.

Wang, R. H., Fan, L. T. and Too, J. R. 1978. Multivariate statistical analysis of solids mixing, *Powder Technology* **21**(2): 171–182.

Warner, F. W. 1983. *Foundations of Differentiable Manifolds and Lie Groups*, Springer-Verlag, New York.

Weinekötter, R. 1993. *Kontinuierliches Mischen Feiner Feststoffe*, PhD thesis, Karlsruhe Technical College.

Weinekötter, R. and Gericke, H. 2000. *Mixing of Solids*, Particle Technology, Kluwer Academic Publishers.

Weinekötter, R. and Reh, R. 1995. Continuous mixing of fine particles, *Particle and Particulate Systems Characterization* **12**: 46–53.

Weyl, H. 1961. *The Classical Groups; Their Invariants and Representations*, Princeton University Press.

Willard, S. 1970. *General Topology*, Addison-Wesley Series in Mathematics, Addison-Wesley Publishing Company, Inc.

Wolf, D., Holin, N. and White, D. H. 1986. Residence time distribution in a commercial twin-screw extruder, *Polymer Engineering and Science* **26**(9): 640–646.

Wolf, D. and White, D. H. 1976. Experimental study of the residence time distribution in plasticating screw extruders, *American Institute of Chemical Engineers (AIChE)* **22**(1): 122–131.

Wright, S. 1982. Character change, speciation, and higher taxa, *Evolution* **36**: 427–443.

Zellner, A. 1984. *Basic Issues in Econometrics*, The University of Chicago Press.

Index