CWI Quarterly Volume 11 (4) 1998, pp. 371 – 384

Averaging of Random Sets and Binary Images

Ilya S. Molchanov
Department of Statistics, University of Glasgow,
Glasgow G12 8QW, Scotland, U.K.
e-mail: ilya@stats.gla.ac.uk

Such a usual concept like expectation or average is not straightforward to define for random sets. This survey provides an overview of existing definitions of expectations for random sets and figures with emphasis on their use in image analysis.

1. Introduction

A sample of real numbers is the classical object of statistics. Such a sample can be summarised in many different ways, for example, through its average, sample median, sample variance, etc. These quantities estimate respectively the mean, the median and the variance of the underlying distribution.

These classical numerical summaries have been successfully extended to the area of *signal processing*, see, e.g., [6]. In signal processing values of a signal around a point (in time) or pixel (in space) are used to produce summary statistics. These statistics are often used to filter signals, for example, the famous median filter returns the median of a signal's values acquired through a neighbourhood of a given time moment. Similar methods are used in image analysis, since an image can be represented as a signal evolving in space rather than time.

The very fact that a signal is a function explains close links with the *functional data analysis* [26]. The latter tends to consider functions 'in the whole' rather than 'locally' or pixelwise. One of typical problems of the functional data analysis is related to realignment (or synchronisation) of functions (curves) using the dynamic time warping [20].

A *binary image* can be also represented as a signal with two possible values 0 and 1. Unfortunately, this rigid structure of the set of possible values makes it very difficult to apply effectively both filtering (averaging methods) from

signal processing and the functional data analysis approaches. For example, the average of several 0- or 1-valued functions may take values between 0 and 1.

When dealing with binary images it is often quite natural to view them as indicators of *sets* which are also referred to as binary images. A single binary image (or set) is usually filtered and summarised using methods coming from mathematical morphology [14]. These methods are essentially non-statistical and do not impose assumptions on the distribution of the image, although the obtained summaries (e.g., granulometry) can be used as visual summaries for classification purposes [2, 28]. If, additionally, we assume that the binary image is obtained by sampling a 'large' stationary image through a bounded sampling window, then it is possible to use statistical methods developed for some models of stationary random sets and point processes [23, 30]. The relevant objective is to estimate parameters of the underlying probability model for the image. Some of the methods are designed for discretised images [12], while other assume that the image is a closed set with non-empty interior [23].

A *random closed set*, X, is a random element whose values are sets; the details are described in [30]. In this survey we always consider random closed sets in the Euclidean space $\mathbb{R}^d$. Note that a random *compact* set is a random closed set which is almost surely bounded (and thereupon compact).

While images representing a part of a 'stationary' or spatially homogeneous patterns are common in texture analysis, material science and microscopy, there are many cases when images cannot be represented as observations of stationary random sets. Therefore, statistical analysis of such images should rely on several independent (or weakly dependent) observations rather than a single realisation within a big window. In other words, the aim is to analyse a *sample of sets* corresponding to a sample of binary images and interpreted as independent identically distributed realisations of a certain random closed set. Purposes of such statistical analysis could be: to average the sets, to explore their variability, to fit a probability model, etc.

The situation is easy if the images are *aligned*, i.e. all features are in the same relative positions with respect to the boundaries of observation windows or other reference points. Small displacements represent little problem if they do not have a systematic component and can cancel through the law of large numbers.

However, in many interesting situations images are not aligned. For instance, if a sample of sand grains is considered, then location or orientation of a particular grain is of no relevance for the subsequent analysis. What matter are the grains' *shapes* and sizes. Roughly speaking, the shapes represent information left after neglecting positions and orientation (and sometimes size) of sets, so that two shapes are identical if the corresponding sets can be superimposed using rigid motions (and sometimes scale transformations). Instead of observing shapes we deal with sets which represent relevant shapes. In this case we speak about a sample of *figures* rather than a sample of sets, so that each set 'represents' the corresponding figure.

A sample of non-aligned images (or shape representatives) must be first

realigned with subsequent application of statistical methods for a sample of sets. It should be noted that there are various alignment methods and the choice of a particular method usually depends on the statistical (or averaging) techniques applied afterwards [31]. In statistical shape analysis [29] the concept of shape roughly corresponds to our concept of figure, apart from the fact that we usually exclude scale transformations from consideration.

A simple random set is obtained as an unordered collection of random points $X = \{\xi_1, \dots, \xi_k\}$. These k points can be interpreted as landmarks; if a specific order of them is prescribed, then we end up in the framework usual in the statistical theory of shape. However, random sets offer substantially more flexibility in comparison with the classical statistical studies of shapes given as collections of points [7, 19, 29]. For instance, a random set framework can be used when it is difficult to identify natural landmarks (so that, perhaps, the whole set should be a 'landmark'), or when the numbers of landmarks are different for different shapes (like polygons with different numbers of vertices), or when some landmarks are missing.

2. Expectations of random sets

A general construction. Such a natural concept like expectation or average is not straightforward to define for random sets. The space $\mathcal{F}$ of closed sets (as well as the space $\mathcal{K}$ of compact sets) is non-linear, so that conventional concepts of expectations in linear spaces are not applicable directly to deal with random closed (or compact) sets. Sets have different features (that often are difficult to express numerically) and particular definitions of expectations highlight particular features important in different contexts.

Further, X is an almost surely non-empty random compact set in $\mathbb{R}^d$. Different realisations of X may be quite different and have no 'common features' which could be used as a 'platform' for averaging. For example, if X takes values $\{0\}$ and $\{0, 1\}$ with probabilities $1/2$, or X takes values $\{0, 1\}$ and $[0, 1]$, with equal probabilities, then it is quite difficult to come up with an 'intuitive' definition of expectation for X.

Clearly, different definitions of expectations utilise different features of the realisations of the random set X. The situation can be explained by the following lucid example. Imagine that X is a 'cat' with probability $1/2$ and a 'dog' otherwise. Clearly, it is pointless to average them, there is no known animal that might serve as their average. However, on the second thought, the question becomes sensible if we aim to average several features of a 'cat' and a 'dog' (weight, tail length, etc.) and then find an existing animal with the features matching the obtained averages as exactly as possible.

The above described approach is the most common to define expectations of random sets. With a random compact set X we associate a random element ξ_X taking values in a Banach space $\mathcal{E}$. This is done by mapping $\mathcal{K}$ into $\mathcal{E}$, so that ξ_X becomes the image of X under this embedding. For example, $\mathcal{K}$ can be mapped into the line $\mathbb{R}$ (so that $\mathcal{E} = \mathbb{R}$) by taking Lebesgue measures of the

elements of $\mathcal{K}$.

Then the expectation of ξ_X is defined in $\mathcal{E}$ with the aim to 'map it back' into $\mathcal{K}$, as the following diagram shows:

$$\begin{array}{ccc} X - \text{random set},\ X \in \mathcal{K} & \longrightarrow & \xi_X - \text{random element in } \mathcal{E} \\ & & \downarrow \\ \text{inverse image of } \mathbf{E}\xi_X & \longleftarrow & \mathbf{E}\xi_X \end{array} \tag{1}$$

However, in many interesting cases $\mathbf{E}\xi_X$ has no inverse image, so that instead of the 'exact' (or 'ideal') inverse image of $\mathbf{E}\xi_X$ one takes a deterministic set K such that ξ_K is as close as possible to $\mathbf{E}\xi_X$. Several examples of this construction are given below.

Aumann expectation. This definition relies on representation of X through its support function

$$h(X,u) = \sup\{\langle x,u\rangle :\ x \in X\}\,, \quad u \in \mathbb{R}^d\,,$$

where $\langle x,u\rangle$ is the conventional scalar product. Assume that $\mathbf{E}\|X\|$ is finite, where $\|X\| = \sup\{\|x\| :\ x \in X\}$ is the norm of X. Then

$$\mathbf{E}h(X,u) = h(\mathbf{E}_{\mathrm{A}}(X),u)$$

is again the support function of a compact set $\mathbf{E}_{\mathrm{A}}(X)$ called the *Aumann expectation* of X, see [4, 32, 33]. The corresponding space $\mathcal{E}$ is the space of square integrable functions on the unit sphere (it suffices to consider u with the unit norm only); and the expectation $\mathbf{E}\xi_X = \mathbf{E}h(X,\cdot)$ has the unique inverse image in the space $\mathcal{K}$.

Alternatively, the Aumann expectation of X is defined as the set of all expectations $\mathbf{E}\zeta$, where ζ is a random vector such that $\zeta \in X$ almost surely. Such a random element ζ is called the *selection* of X. It follows from Lyapounov's theorem on vector-valued measures that both definitions yield the same set if the underlying probability space is non-atomic, see also [27]. This definition works well for convex random sets, while if X is non-convex (and even non-random), then $\mathbf{E}_{\mathrm{A}}(X)$ coincides with the expectation of the convex hull of X. For instance, if $X = \{0,1\}$, then $\mathbf{E}_{\mathrm{A}}X = [0,1]$.

The Aumann expectation is very popular in the studies of multivalued functions. It also can be used to define set-valued martingales, see [1, 8, 18, 24, 25]. A generalisation which handles more naturally both atomic and non-atomic probability spaces was suggested in [34].

Vorob'ev expectation. This definition is based on representation of X through its *indicator function* $\xi_X(x) = \mathbb{1}_X(x)$, $x \in \mathbb{R}^d$, so that $\mathcal{E}$ is the space of integrable functions on $\mathbb{R}^d$. The expectation of the indicator function is the *coverage function*

$$p_X(x) = \mathbf{P}\{x \in X\}\,.$$

Assume that $\mu(X)$, the Lebesgue measure of X, has a finite expectation, so that

$$\mathbf{E}\mu(X) = \int_{\mathbb{R}^d} p_X(x)dx < \infty\,.$$

It should be noted that $p_X(x)$ is not an indicator function anymore, so that $\mathbf{E}\xi_X$ has no 'exact' inverse image in the space $\mathcal{K}$. This situation is quite typical in many definitions of expectations and is usually circumvented by finding a deterministic compact set K such that ξ_K 'mimics' $\mathbf{E}\xi_X$ as exactly as possible. Since it is very difficult to solve minimisation problems parametrised by K running through the whole family $\mathcal{K}$, one usually searches through a sub-family of $\mathcal{K}$, which is often determined by $\mathbf{E}\xi_X$ itself. In the current setup, possible 'candidates' are provided by thresholded sets of the coverage function,

$$L_p = \{x \in \mathbb{R}^d : \; p_X(x) \geq p\}\,.$$

The *Vorob'ev expectation* is defined by $\mathbf{E}_{\mathrm{V}}(X) = L_p$ for p determined from the inequality

$$\mu(L_q) \leq \mathbf{E}(\mu(X)) \leq \mu(L_p)\,, \quad \text{for all} \quad q > p\,,$$

see [32]. In other words, $\mathbf{E}_{\mathrm{V}}X$ is the thresholded set of the coverage function such that its Lebesgue measure is as closest as possible to $\mathbf{E}\mu(X)$. Note that this definition treats singletons as well as sets of almost surely vanishing Lebesgue measure as uninteresting.

Radius-vector expectation. The *radius-vector expectation* [32] is applicable for star-shaped sets. This expectation is defined by the expected values of the radius-vector function, $r_X(u)$, of X (that represent the boundary of X in polar or spherical coordinates) for u from the unit sphere. The major shortcoming is the necessity to work with star-shaped sets and the non-linearity with respect to translations of the sets. This means that the expectation is essentially different for different choices of the reference point (or the polar origin) within X, see [21].

Distance average. A random closed set X corresponds to its distance function $\xi_X(x) = \rho(x, X)$ which equals the Euclidean distance from x to the nearest point of X, so that $\mathcal{E}$ becomes the space of continuous functions on $\mathbb{R}^d$. It is easy to show that the expected distance function

$$\bar{d}(x) = \mathbf{E}(\rho(x, X))$$

is not a distance function of a deterministic set (unless X is deterministic itself).

A suitable thresholded set of the expected distance function $\bar{d}(x)$ can serve as the mean of X. The *distance average* $\bar{X}$ is the set

$$X(\varepsilon) = \{x : \bar{d}(x) \leq \varepsilon\}\,,$$

where $\varepsilon \geq 0$ is chosen to minimise

$$\mathfrak{m}(\bar{d}, \rho(\cdot, X(\varepsilon))) = \sup\{|\bar{d}(x) - \rho(x, X(\varepsilon))| : \ x \in \mathbb{R}^d\}, \tag{2}$$

see [5] for details and further generalisations. In particular, it is possible to use the signed distance function $\rho_{\mathrm{s}}(x, X)$ instead of $\rho(x, y)$, so that $\rho_{\mathrm{s}}(x, X)$ is equal to the difference between the distance function of X and its complement. Furthermore, the distance $\mathfrak{m}$ in (2) could be the L^2 distance between the corresponding functions instead of L^∞ metric used in (2). If X is a subset of $W \subset \mathbb{R}^d$, then it is possible to rewrite the definition for the functions defined on W.

It should be noted that this approach allows us to deal with sets of zero Lebesgue measure, since the distance function is non-trivial in this case. Distance average has a serious potential in imaging applications because it can deal with non-convex sets and also is not computationally expensive, see [5]. It is also possible to use an analogue of distance average to threshold grey-scale images, see [10]. Then the threshold appears at the level such that the corresponding distance function approximates the expected distance function of a random set obtained by thresholding at a random height U. Note that the Vorob'ev expectation in this context leads to naive thresholds, where the threshold level is chosen in such a way that the measure of the thresholded set equals the integral of the image (if U is uniformly distributed over the range of grey levels) or half the measure of the window (if U is distributed according to the histogram of the image).

Evaluations and expectations. Another approach to define expectations of general lattice-valued random elements (in particular, random sets) was elaborated in [15], see also [22, Chapter 2]. Let $\mathbb{U}$ be a family of *evaluations* which are functions mapping $\mathcal{F}$ into $\mathbb{R}$. Then $\delta_u(F) = \{u(x) : \ x \in F\}$, $F \in \mathcal{F}$, and $\varepsilon_u(y) = \{x \in \mathbb{R}^d : \ u(x) \leq y\}$, $y \in \mathbb{R}$, form a pair $(\varepsilon_u, \delta_u)$, which is an *adjunction* between $\mathcal{F}$ and $\mathbb{R}$, see [14]. The expectation is defined as

$$\mathbf{E}_{\mathbb{U}} X = \bigcap_{u \in \mathbb{U}} \varepsilon_u(\mathbf{E}\delta_u(X)). \tag{3}$$

If X is a deterministic set, then $\mathbf{E}_{\mathbb{U}} X$ is called the $\mathbb{U}$-closure of X. For instance, the Aumann expectation is obtained if $\mathbb{U}$ is the family of support functions, and the $\mathbb{U}$-closure becomes the convex hull. In this approach $\mathcal{E}$ is the space of all real-valued functions on $\mathbb{U}$. However, instead of finding the inverse of $\mathbf{E}\xi_X$ by minimisation, the approximate inverse image is computed directly by (3).

Doss expectation. The *Doss expectation*, $\mathbf{E}_{\mathrm{D}}(X)$, is defined by

$$\mathbf{E}_{\mathrm{D}}(X) = \{y : \ \rho(x, y) \leq \mathbf{E}\rho_{\mathrm{H}}(\{x\}, X) \text{ for all } x \in \mathbb{R}^d\},$$

see [9, 16]. The Doss expectation is a particular case of $\mathbf{E}_{\mathbb{U}} X$, if $\mathbb{U}$ is the family of functions $\rho(x, u)$, $u \in \mathbb{R}^d$. Since $\mathbf{E}_{\mathrm{D}}(X)$ equals the intersection of all balls

with radius $\mathbf{E}\rho_{\mathrm{H}}(\{x\}, X)$ centred at x with x running through $\mathbb{R}^d$, it is easy to see that $\mathbf{E}_{\mathrm{D}}(X)$ is convex and the Doss expectation is 'compatible' with the standard definition of the expectation if X is a random singleton. The Doss expectation can be also used to define set-valued martingales, see [17].

Fréchet expectation. It is possible as well to explore the metric structure on the family $\mathcal{K}$ of compact subsets of $\mathbb{R}^d$ that is determined by the *Hausdorff metric*

$$\rho_{\mathrm{H}}(K, K_1) = \inf\{r > 0 : \ K \subset K_1^r, K_1 \subset K^r\},$$

where K^r is the r-neighbourhood of K. Now a set $K = K_0 \in \mathcal{K}$ that minimises $\mathbf{E}(\rho_{\mathrm{H}}(X, K)^2)$ for $K \in \mathcal{K}$ is said to be the *Fréchet expectation* of X. This approach is very general and can be used if ρ_{H} is replaced by another metric on $\mathcal{K}$. Unfortunately, in most practical cases it is not possible to solve the basic minimisation problem, since the parameter space $\mathcal{K}$ is too rich. Also the Fréchet expectation can be non-unique.

A toy example. Let X be a random subset of the real line so that X takes values $[0,1]$ and $\{0,1\}$ with probabilities $1/2$. Then $\mathbf{E}_{\mathrm{A}}X = \mathbf{E}_{\mathrm{V}}X = \mathbf{E}_{\mathrm{D}}X = [0,1]$, the radius-vector expectation is not applicable, the distance average is $[-1/8, 1/8] \cup [7/8, 9/8]$, and the Fréchet expectation is $[0, 0.4] \cup [0.6, 1]$. In this case, the latter seems to reflect the best the true character of the mean for X, since the Aumann, Vorob'ev and Doss expectation do not respect the possibility of X being $\{0,1\}$, while the distance average is not a subset of $[0,1]$. Unfortunately, for more complicated examples the Fréchet expectation becomes very difficult to compute.

3. Properties of expectations

A classification of basic properties. It is possible to formulate several basic properties of a 'reasonable' expectation $\mathbf{E}(X)$ of X. The first group of the properties is related to inclusion relationships.

A1 If X is deterministic, then $\mathbf{E}X = X$.

A2 If $K \subseteq X$ a.s., where K is deterministic, then $K \subseteq \mathbf{E}X$.

A3 If $X \subseteq W$ a.s. for a deterministic set W (perhaps, from some special family), then $\mathbf{E}X \subseteq W$.

A4 If $X \subseteq Y$ a.s., then $\mathbf{E}X \subseteq \mathbf{E}Y$.

Clearly, **A2** and **A3** imply **A1**, while **A1** and **A4** yield both **A2** and **A3**.

The second group consists of the properties related to invariance with respect to some transformations.

B1 If X is distribution-invariant with respect to a certain group G (which means that gX and X have the same distribution for each $g \in G$), then the expectation of X must be invariant with respect to G.

B2 Translation-equivariance: $\mathbf{E}(X + x) = \mathbf{E}X + x$.

B3 Homogeneity: $\mathbf{E}(cX) = c\mathbf{E}(X)$.

The third group of properties relates expectations of sets and 'usual' expectations of random variables and vectors.

C1 If $X = \{\xi\}$ is a random singleton, then $\mathbf{E}X = \{\mathbf{E}\xi\}$.

C2 If $X = B_\eta(\xi)$ is a ball of random radius η and centre ξ, then $\mathbf{E}X = B_{\mathbf{E}\eta}(\mathbf{E}\xi)$.

C3 If $X = \mathrm{conv}(\xi_1, \ldots, \xi_n)$ is the convex hull of a finite number of random points, then $\mathbf{E}X = \mathrm{conv}(\mathbf{E}\xi_1, \ldots, \mathbf{E}\xi_n)$.

C4 If $X = \{\xi_1, \ldots, \xi_n\}$ is of a finite set of random points, then $\mathbf{E}X = \{\mathbf{E}\xi_1, \ldots, \mathbf{E}\xi_n\}$.

Note that some of these natural properties are non-compatible and have far-reaching consequences. For example, **A4** and **C1** imply that $\mathbf{E}X \ni \mathbf{E}\xi$ for each selection $\xi \in X$, so that $\mathbf{E}X$ should contain the Aumann expectation of X. For instance, it is possible to see that the Doss expectation satisfies **C1** and **A4**, whence it contains the Aumann expectation. However convexity of both Aumann and Doss expectations severely restricts possible applications, e.g. in image analysis, where most images are non-convex. The distance average seems to be the most versatile expectation, which can have a wide range of possible applications in image analysis.

Averaging and laws of large numbers While the above mentioned expectations are derived from the probability distributions of random sets, in practice they are estimated as averages if a sample of independent identically distributed realisations is given. Such a sample of sets $X_1, \ldots, X_n$ can be interpreted as a random closed set X which takes the enlisted values with equal probabilities $1/n$. This allows us to reformulate all expectations for samples of sets. In statistical language this approach means substituting of the empirical distribution instead of the theoretical distribution of X. The corresponding 'naive' estimators are unbiased for the case of Aumann expectation and radius-vector average and asymptotically unbiased in other cases.

For instance, if the Aumann expectation is chosen, then it is estimated by

$$\hat{X}_n = \frac{X_1 \oplus \cdots \oplus X_n}{n},$$

since the Minkowski addition of convex sets corresponds to addition of their support functions. If $\mathbf{E}\|X_1\| < \infty$, then the strong law of large numbers for random sets [3] states that $\hat{X}_n$ converges to $\mathbf{E}_{\mathrm{A}}X$ almost surely in the Hausdorff metric as $n \to \infty$.

4. Averaging of figures

Typically, the starting point for statistical analysis is a sample of i.i.d. realisations of a random compact set. If positions of the sets are known, then we speak about statistics of *sets*, in contrast to statistics of *figures* when locations/orientations of sets are not specified. This means that the positions of the sets are irrelevant for the problem and the aim is to find the average shape of the sets in the sample. Such a situation appears in studies of particles (dust powder, sand grains, abrasives etc.).

At the first approximation, one can characterise the shape of a compact set X by numerical parameters, called *shape ratios*, see [32]. For example, the *area-perimeter ratio* (or compacity) is given by $4\pi\mathrm{area}(X)/\mathrm{perimeter}(X)^2$, *circularity shape ratio* is the quotient of the diameter of the circle with the same area as X and the diameter of the smallest circumscribed circle of X. All these shape ratios are motion- and scale-invariant, so that their values do not depend on translations/rotations and scale transformations of X. In the engineering literature it is usual to perform statistical analysis of a sample of sets $X_1, \dots, X_n$ by computing several shape ratios for each set from the sample. This yields a multivariate sample, which can be analysed using multivariate statistical methods. It should be noted that the distributions of the relevant measurements are not known in most cases, so that non-parametric statistical methods are the best suitable for statistical analysis of such samples.

If an observer deals with a sample of figures rather than sets, then the definitions of expectation of a random compact set are not directly applicable or are not informative. For instance, the images of particles are isotropic sets, whence the corresponding set-expectations are balls or discs.

The approach below can be found in [31]. It is inspired by the studies of shapes and landmark configurations, see [7, 29]. Two compact sets are equivalent if they can be superimposed by a rigid motion (scale transformations are excluded). Then a sample of sets (that represent the corresponding figures) must be transformed in order to place them 'close together', and then a set-theoretic mean could be determined for the transformed sets. If $X_1, \dots, X_n$ is a sample of sets which 'represent' the corresponding figures, then the aim is to 'move' these sets in such way that the results $X_1^*, \dots, X_n^*$ are 'closed together' or 'aligned'.

Remember that (1) defines expectation of X through a map into a linear space $\mathcal{E}$. Assume that $\mathcal{E}$ is a Hilbert space. Motions of sets (translations, rotations, etc.) give rise to transformations of functions in $\mathcal{E}$. For example, if ω is rotation, then the support function of X is transformed as

$$h(\omega X + z, u) = h(X, \omega^{-1}u) + \langle z, u\rangle\,. \tag{4}$$

Then motions of sets $X_1, \dots, X_n$ correspond to transformations of the associated functions $\xi_{X_1}, \dots \xi_{X_n}$. Such associated functions are typically used to defined the average

$$\hat{\xi} = n^{-1}(\xi_{X_1} + \dots + \xi_{X_n})$$

and then find a set $\hat{X}$ such that $\xi_{\hat{X}}$ is the closest (in some sense) to $\hat{\xi}$. This implies that the functions rather that the sets themselves should be 'closed together' in the corresponding norm. This allows us to use simple geometrical facts of Hilbert space theory (applied to elements of $\mathcal{E}$) in order to characterise the families of functions in 'optimal relative position'.

To be general enough, let $\mathcal{H}$ be an abstract Hilbert space, and let G be a group of isometric transformations acting on $\mathcal{H}$. Two elements of $\mathcal{H}$ are equivalent if they can be superimposed by actions of G, or, in other words, if they belong to the same orbit generated by G. The procrustean distance between two elements is given by

$$\rho_{\mathrm{P}}(x,y) = \inf_{g\in G} \|x - gy\| .$$

This definition goes back to the theory of shape where the group G often includes non-isometric scale transformations, see [11, 13].

A finite set of points $\mathbf{x} = \{x_1, \ldots, x_n\} \subset \mathcal{H}$ is called a *configuration*. Now the aim is to determine transformations $\mathbf{g} = \{g_1, \ldots, g_n\} \subset G$ such that the elements of the transformed configuration $\mathbf{gx} = \{g_1x_1, \ldots, g_nx_n\}$ are close together. A possible goal functional is the *inertia* of configuration

$$I(\mathbf{x}) = \sum_{1\le i<j\le n} \|x_i - x_j\|^2 .$$

If

$$I(\mathbf{x}) = \inf_{\mathbf{g}\subset G} I(\mathbf{gx}) , \tag{5}$$

then the configuration $\mathbf{x}$ is said to be in *optimal relative position* (with respect to G).

It is easy to show (see [31]) that $I(\mathbf{x}\cup\{y\})$ is minimal over $y \in \mathcal{H}$ if and only if

$$y = \bar{\mathbf{x}} = n^{-1}(x_1 + \cdots + x_n)$$

is the (arithmetic) mean of the points from $\mathbf{x}$. However, it is not the arithmetic mean which is of interest when computing the average of several figures. In the latter case one seeks the *relative mean* $\mathbf{x}_G$ which is a set of all points $b \in \mathcal{H}$ such that

$$\inf_{\mathbf{g}\subset G}(I(\mathbf{gx}\cup\{b\}) - I(\mathbf{gx})) = \inf_{y\in\mathcal{H}} \inf_{\mathbf{g}\subset G}(I(\mathbf{gx}\cup\{y\}) - I(\mathbf{gx})) .$$

It is shown in [31] that $\mathbf{x}$ is in optimal relative position if and only if $\bar{\mathbf{x}} \in \mathbf{x}_G$ and for each $i = 1, \ldots, n$

$$\|x_i - \bar{\mathbf{x}}\| = \rho_{\mathrm{P}}(x_i, \bar{\mathbf{x}}) .$$

The latter means that each of x_i is in (pairwise) optimal relative position with the arithmetic mean $\bar{\mathbf{x}}$. This result explains that finding the relative mean is equivalent to finding the arithmetic mean if the configuration is in optimal relative position.

In order to bring $\mathbf{x}$ to its optimal relative position, one has to find $\mathbf{g} \subset G$ which provides the global minimum for the right-hand side of (5). This complicated optimisation problem can be simplified in several cases. In the most versatile case, let $g \in G$ allow representation as $g = g'g''$ with g' and g'' belonging to other groups G' and G'' respectively. Furthermore, assume that each $g' \in G'$ corresponds to a point $l = l(g') \in L \subset \mathcal{H}$ such that $g'x = x + l$ for all $x \in \mathcal{H}$ with L being a linear subspace in $\mathcal{H}$; and G'' consists of linear operators such that $g''L \subset L$. Then $\mathbf{x}$ is in optimal relative position with respect to G if and only if

$$\mathrm{pr}_L\, x_1 = \mathrm{pr}_L\, x_2 = \cdots = \mathrm{pr}_L\, x_n\,,$$

where $\mathrm{pr}_L\, x$ denotes the orthogonal projection of x onto L; and the configuration of 'residuals':

$$\mathbf{x}^o = \{x_1 - \mathrm{pr}_L\, x_1, \ldots, x_n - \mathrm{pr}_L\, x_n\}$$

is in optimal relative position with respect to G''. In applications, G' becomes the group of translations and G'' the group of rotations. This result allows us to find the optimal translations explicitly while the rotations are to be determined via a numerical optimisation problem.

For example, if convex compact sets are described by support functions, then the group of proper motions of sets corresponds to a group acting on the space of support functions as shown in (4). It was proven in [31] that the 'optimal' translations of convex sets $X_1, \ldots, X_n$ superimpose their Steiner points

$$s(X_i) = \frac{1}{b_d} \int_{\|u\|=1} h(X_i, u)\, u\, du\,, \quad i = 1, \ldots, n\,,$$

where b_d is the volume of the unit ball in $\mathbb{R}^d$.

To find the optimal transformations in the general case, it is possible to use an iterative algorithm, described in [31]. On each step it computes the arithmetic mean of $\mathbf{x}$ and then brings each point x_i as close as possible to the computed mean using admissible transformations from G in order to obtain another configuration which is used to initiate the next step of the algorithm.

5. Concluding remarks

Since there is no universally applicable concept of expectation for random sets, the choice of a particular definition depends on the objective. In many cases, geometric properties of realisations of random sets entail the first choice of expectation and at the same time rule out some other definitions. For example, if

realisations are convex, then the Aumann expectation is the only natural choice, since it works very well for convex random sets and at the same time has desirable theoretical properties. If realisations are star-shaped, the radius-vector expectation is quite natural, but a lack of natural candidates for the reference points (polar origins) within the sets can cause considerable problems in practice. If all realisations are accessible through some numerical measurements only, then the approach based on evaluations becomes the first choice. In image analysis, the distance average yields sufficiently good results, as it does not rely on specific geometric properties of the sets (or images), which often are impossible to impose in real imaging applications. The Fréchet expectation is a very general and natural definition, but it is very difficult to use because of computational problems.

References

1. R.A. Aló, A. de Korvin, and C. Roberts (1979). The optional sampling theorem for convex set valued martingales. *J. Reine Angew. Math.* **310**, 1–6.
2. S. Archambault and M. Moore (1993). Statistiques morphologiques pour l'ajustement d'images. *Internat. Statist. Rev.* **61**, 283–297.
3. Z. Artstein and R.A. Vitale (1975). A strong law of large numbers for random compact sets. *Ann. Probab.* **3**, 879–882.
4. R.J. Aumann (1965). Integrals of set-valued functions. *J. Math. Anal. Appl.* **12**, 1–12.
5. A.J. Baddeley and I.S. Molchanov (1998). Averaging of random sets based on their distance functions. *Journal of Mathematical Imaging and Vision* **8**, 79–92.
6. S.P. Banks (1990). *Signal Processing, Image Processing and Pattern Recognition.* Prentice Hall, New York.
7. F.L. Bookstein (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statist. Sci.* **1**, 181–242.
8. B. van Cutsem (1972). Martingales de convexes fermés aléatoires en dimension finie. *Ann. Inst. H.Poincaré, Sect. B, Prob. et Stat.* **8**, 365–385.
9. S. Doss (1949). Sur la moyenne d'un élément aléatoire dans un espace distancié. *Bull. Sci. Math.* **73**, 48–72.
10. N. Friel and I.S. Molchanov (1999). A new thresholding technique based on random sets. *Pattern Recognition.* To appear.
11. C.R. Goodall (1991). Procrustes methods in the statistical analysis of shape (with discussion). *J. Roy. Statist. Soc. Ser. B* **53**, 285–339.
12. J. Goutsias (1993). Binary random fields, random set theory, and the morphological analysis of shape. *Image Algebra and Morphological Image Processing, IV (San Diego, CA, 1993)*, volume **2030** of *Proc. SPIE*, pages 54–64. SPIE, Bellingham, WA.
13. J.C. Gower (1975). Generalized procrustes analysis. *Psychometrika* **40**, 33–51.

14. H.J.A.M. Heijmans (1994). *Morphological Image Operators.* Academic Press, Boston.
15. H.J.A.M. Heijmans and I.S. Molchanov (1998). Morphology on convolution lattices with applications to the slope transform and random set theory. *Journal of Mathematical Imaging and Vision* **8**, 199–214.
16. W. Herer (1986). Esperance mathematique au sens de Doss d'une variable aleatoire a valeurs dans un espace metrique. *C. R. Acad. Sci., Paris, Ser. I* **302**, 131–134.
17. W. Herer (1987). Martingales with values in closed bounded subsets of a metric space. *C. R. Acad. Sci., Paris, Ser. I* **305** (6), 275–278. In French.
18. F. Hiai and H. Umegaki (1977). Integrals, conditional expectations, and martingales of multivalued functions. *J. Multivariate Anal.* **7**, 149–182.
19. D.G. Kendall (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16**, 81–121.
20. A. Kneip, T. Gasser (1992). Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20**, 1266–1305.
21. A. Mancham and I.S. Molchanov (1996). Stochastic models of randomly perturbed images and related estimation problems. In K.V. Mardia and C.A. Gill, editors, *Image Fusion and Shape Variability Techniques*, pages 44–49, Leeds, 3–5 July. Leeds University Press.
22. I.S. Molchanov (1993). *Limit Theorems for Unions of Random Closed Sets*, volume **1561** of *Lect. Notes Math.* Springer, Berlin.
23. I.S. Molchanov (1997). *Statistics of the Boolean Model for Practitioners and Mathematicians.* Wiley, Chichester.
24. N.S. Papageorgiou (1985). On the theory of Banach space valued multifunctions I, II. *J. Multivariate Anal.* **17**, 185–206, 207–227.
25. N.S. Papageorgiou (1987). A convergence theorem for set-valued supermartingales with values in a separable Banach space. *Stochastic Anal. Appl.* **5**, 405–422.
26. J.O. Ramsay and B.W. Silverman (1997). *Functional Data Analysis.* Springer, Ney York.
27. H. Richter (1963). Verallgemeinerung eines in der Statistik benötigten Satzes der Maßtheorie. *Math. Ann.* **150**, 85–90 and 440–441.
28. B.D. Ripley (1986). Statistics, images and pattern recognition. *Canad. J. Statist.* **14**, 83–102.
29. C.G. Small (1996). *The Statistical Theory of Shape.* Springer, New York.
30. D. Stoyan, W.S. Kendall and J. Mecke (1995). *Stochastic Geometry and Its Applications.* Wiley, Chichester, second edition.
31. D. Stoyan and I.S. Molchanov (1997). Set-valued means of random particles. *Journal of Mathematical Imaging and Vision* **7**, 111–121.
32. D. Stoyan and H. Stoyan (1994). *Fractals, Random Shapes and Point Fields.* Wiley, Chichester.

33. R.A. VITALE (1988). An alternate formulation of mean value for random geometric figures. *J. Microscopy* **151**, 197–204.
34. R.A. VITALE (1991). The translative expectation of a random set. *J. Math. Anal. Appl.* **160**, 556–562.