1

The Buffon needle problem

We begin with what is probably the best-known problem of geometric probability, the Buffon needle problem. This solution of the needle problem via the characterization of an additive set functional serves to motivate the study of valuations on lattices, the topic of Chapter 2. Variations and generalizations of the Buffon needle problem are presented in Chapters 8 and 9.

1.1 The classical problem

Parallel straight lines are drawn on the plane $\mathbf{R}^2$, at a distance d from each other. A needle of length L is dropped at random on the plane. What is the probability that the needle shall meet at least one of the lines?

This problem can be solved by computations with conditional probability (Feller, for example, solved it in this way in his well known treatise [23, p. 61]). It is, however, more instructive to solve it by another method, one that minimizes the amount of computation and maximizes the role of probabilistic reasoning.

Let X_1 be the number of intersections of a randomly dropped needle of length L_1 with any of the parallel straight lines. If the needle is long enough, the random variable X_1 can take several integer values, whereas if the needle is short, it can take only the values 0 or 1.

If p_n is the probability that the needle meets exactly n of the straight lines, and if $E(X_1)$ denotes the expectation of the random variable X_1, then we have

$$E(X_1) = \sum_{n \geq 0} np_n.$$

Thus, if $L_1 < d$, then

$$E(X_1) = 0p_0 + 1p_1 = p_1,$$

and p_1 is the probability we seek. Therefore, it is sufficient to compute the expectation $E(X_1)$. Suppose that another needle of length L_2 is dropped at random. The number of intersections of this second needle with any of the parallel straight lines drawn on $\mathbf{R}^2$ is another random variable, say X_2. The random variables X_1 and X_2 are independent, unless the needles are welded together. Suppose that the needles are rigidly bound at one of their endpoints. They may form a straight line, or they may be at an angle. In either case, if the two rigidly bound needles are simultaneously dropped on $\mathbf{R}^2$, their total number of intersections will still be $X_1 + X_2$. The random variables X_1 and X_2 will no longer be independent, but their expectation will remain additive:

$$E(X_1 + X_2) = E(X_1) + E(X_2). \tag{1.1}$$

The same reasoning applies to the random variable $X_1 + X_2 + \cdots + X_k$, for the case in which k needles are welded together to form a polygonal line of arbitrary shape.

Since $E(X_1)$ clearly depends on the length L_1, we can write $E(X_1) = f(L_1)$, where f is a function to be determined. By welding together two needles so that they form one straight line we find that $E(X_1 + X_2) = f(L_1 + L_2)$, and we infer from (1.1)

$$f(L_1 + L_2) = f(L_1) + f(L_2).$$

It then follows that f is linear when restricted to rational values of L. Since f is clearly a monotonically increasing function with respect to L, we infer that $f(L) = rL$ for all $L \in \mathbf{R}$, where the constant r is to be determined.

If C is a rigid wire of length L, dropped randomly on $\mathbf{R}^2$, and if Y is the number of intersections of C with any of the straight lines, then C can be approximated by polygonal wires, so that Y is approximately equal to $X_1 + X_2 + \cdots + X_k$. Passing to the limit, we find that

$$E(Y) = rL. \tag{1.2}$$

This allows us to determine the value of the constant r, by choosing a wire of suitable shape. Let C be a circular wire of diameter d. Obviously $E(Y) = 2$, and $L = \pi d$. It then follows from (1.2) that

$$2 = r\pi d,$$

whence $r = 2/(\pi d)$. Thus, for a short needle, we have

$$E(X_1) = p_1 = \frac{2L}{\pi d}.$$

This result has been used (rather inefficiently) to compute the value of π. Instead, we shall use it as the theorem leading into the heart of geometric probability, following the ideas of Crofton and Sylvester.

1.2 The space of lines

Let Graff(2,1) denote the set of all straight lines in $\mathbf{R}^2$ (the reason for this notation shall be made clearer in Chapter 7). It is well known that this set enjoys some notable properties.

To this end, denote by Z_1 the number of intersections of a straight line taken at random with a straight line segment of length L_1, and let λ_1^2 denote the invariant measure on Graff(2,1). The integral

$$\int_{\mathrm{Graff}(2,1)} Z_1 \, \mathrm{d}\lambda_1^2$$

depends only on L_1. Since Z_1 takes only the values 0 or 1, this integral is equal to the measure of the set of all straight lines that meet the given straight line segment. Since the value of the integral depends only on the length L_1 of the straight line segment, denote this value by $f(L_1)$. We can now repeat the argument we used for the Buffon needle problem: given a polygonal line consisting of segments of length $L_1, L_2, \ldots$, the number of intersections of a randomly chosen straight line with the polygonal line is

$$\int_{\mathrm{Graff}(2,1)} (Z_1 + Z_2 + \cdots) \, \mathrm{d}\lambda_1^2 = f(L_1 + L_2 + \cdots).$$

Since integrals are linear, this becomes

$$\int_{\mathrm{Graff}(2,1)} Z_1 \, \mathrm{d}\lambda_1^2 + \int_{\mathrm{Graff}(2,1)} Z_2 \, \mathrm{d}\lambda_1^2 + \cdots = f(L_1) + f(L_2) + \cdots,$$

and we again conclude that $f(L) = rL$. We shall *not* normalize the measure λ_1^2 by setting $r = 1$; rather, we shall decide later what the 'right' normalization should be.

Again we may pass to the limit. Recall that a subset K of the plane is *convex* if any two points x and y in K are the endpoints of a line segment lying inside K. A curve C in the plane is called *convex* if C encloses a convex subset. Let C be a convex curve in the plane of length L, and let

Z_C be the number of intersections of C with a randomly chosen straight line. Then

$$\int_{\mathrm{Graff}(2,1)} Z_C \, \mathrm{d}\lambda_1^2 = rL.$$

In particular, let K_1 and K_2 be compact convex sets in the plane with non-empty interiors, and with boundaries $C_1 = \partial K_1$ and $C_2 = \partial K_2$ of length L_1 and L_2. For each i, we have

$$\int_{\mathrm{Graff}(2,1)} Z_{C_i} \, \mathrm{d}\lambda_1^2 = rL_i.$$

On the other hand, since K_i is convex, a straight line meets K_i either twice or not at all (excluding the limiting cases of tangents, which can be shown to have measure zero). Thus, the function Z_{C_i} takes either the value 2 or the value 0. If we denote by D_i the set of all straight lines in $\mathbf{R}^2$ that meet K_i, then we have

$$\int_{\mathrm{Graff}(2,1)} Z_{C_i} \, \mathrm{d}\lambda_1^2 = 2\lambda_1^2(D_i).$$

To re-state these results in terms of probability, assume that $K_1 \subseteq K_2$. The conditional probability that a straight line shall meet the compact convex set K_1, given that it meets K_2, is the ratio

$$\frac{\lambda_1^2(D_1)}{\lambda_1^2(D_2)}.$$

The computation above shows that this ratio is equal to

$$\frac{L_1}{L_2} = \frac{\mathrm{length}(\partial K_1)}{\mathrm{length}(\partial K_2)}.$$

Note that the value of the normalization constant r is irrelevant to the computation of this conditional probability.

The results above (sometimes designated *Sylvester's theorem*) can be compared to the analogous result for points: if $K_1 \subseteq K_2$, the conditional probability that a point taken at random shall belong to K_1, given that it belongs to K_2, is

$$\frac{\mathrm{area}(K_1)}{\mathrm{area}(K_2)}.$$

Thus, we see a striking analogy: replacing every occurrence of the word 'point' by the word 'line' corresponds to replacing the word 'area' by the word 'perimeter.' This analogy suggests that a generalization of Sylvester's theorem to arbitrary dimension may prove worthwhile.

1.3 Notes

The solution to Buffon's needle problem presented here is due to Barbier [5], and was later generalized still further by Crofton in [14, 15, 16]. Crofton's main paper, which set geometric probability on its modern footing, is the Encyclopaedia Britannica article [17]. It is still an excellent reference.

In [95] Sylvester considered a variation of the Buffon needle problem in which the needle is replaced by a finite rigid collection of compact convex (and possibly disjoint) sets $K_1, \ldots, K_m$ tossed randomly into a plane tiled by evenly spaced lines. Sylvester then considered the cases in which a line meets one, some, or all of the sets K_i. In the previous section we measured the set of all lines meeting a compact convex set K in the plane. When dealing with multiple convex sets Sylvester was led to consider also the measure of the set of lines that *separate* two disjoint compact convex regions of the plane. This theme has also been pursued extensively in the work of Ambartzumian [1, 2].

Buffon's result gives a very inefficient means of approximating the number π; for a history of this technique, see [30]. For additional modern treatments of geometric probability in the plane, see also [1, 2, 49, 82, 90].

Cambridge University Press
978-0-521-59362-5 - Introduction to Geometric Probability
Daniel A. Klain and Gian-Carlo Rota
Excerpt
More information

2

Valuation and integral

In Chapter 1 we expressed the Buffon needle problem in terms of a set functional (1.1) on a certain collection of sets in the plane satisfying a certain kind of additivity. We then solved the problem by characterizing this additive functional in (1.2), using in this case the fact that the functional was monotonically increasing and invariant with respect to certain motions of sets in the plane.

In this chapter we make more precise the notion of 'additive set functional', or *valuation*, on a lattice of sets. The abstract notions developed in this chapter will then be specialized to several different specific lattices in the chapters following, leading in turn to similarly elegant solutions to generalizations and analogues of Buffon's original problem. Section 2.2 is devoted to Groemer's integral theorem, which is needed to prove Groemer's extension theorems in Sections 4.1, 5.1, and 11.1.

2.1 Valuations

We now introduce a class of set functions that comprise the most basic and important tools of geometric probability, namely *valuations*. We begin with partially ordered sets and lattices. A partial ordering $\leq$ on a set L is a relation satisfying the following conditions for all $x, y, z \in L$.

(i) $x \leq x$.
(ii) If $x \leq y$ and $y \leq x$ then $x = y$.
(iii) If $x \leq y$ and $y \leq z$ then $x \leq z$.

The partially ordered set L is called a *lattice* if, for all $x, y \in L$, there exist a greatest lower bound (or *meet*) $x \wedge y \in L$ and a least upper bound (or *join*) $x \vee y \in L$. A lattice L is said to be *distributive* if, for all $x, y, z \in L$, we have the following.

(i) $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$.
(ii) $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$.

Let S be a set, and let L be a family of subsets of S closed under finite unions and finite intersections. Such a family is clearly a distributive lattice, in which the partial ordering is given by subset inclusion, while the meet and join are given by intersection and union of sets, respectively.

A *valuation* on a lattice L of sets is a function μ defined on L that takes real values, and that satisfies the following conditions:

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B), \tag{2.1}$$

$$\mu(\emptyset) = 0, \text{ where } \emptyset \text{ is the empty set.} \tag{2.2}$$

By iterating the identity (2.1) we obtain the *inclusion–exclusion principle* for a valuation μ on a lattice L, namely

$$\mu(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_i \mu(A_i) - \sum_{i<j} \mu(A_i \cap A_j) + \sum_{i<j<k} \mu(A_i \cap A_j \cap A_k) + \cdots \tag{2.3}$$

for each positive integer n.

If A is any subset of S, the *indicator function* (or simply the *indicator*) of A, denoted by I_A, is the function on S given by

- $I_A(s) = 1$; $s \in A$,
- $I_A(s) = 0$; $s \notin A$.

A finite linear combination

$$f = \sum_{i=1}^{k} \alpha_i I_{A_i}, \tag{2.4}$$

where $\alpha_i \in \mathbf{R}$, and $A_i \in L$, is said to be an L-*simple* function, or a *simple* function for short. The set of all L-simple functions forms a ring under the usual operations on functions.

Indicator functions satisfy the following properties:

$$I_{A \cap B} = I_A I_B, \tag{2.5}$$

$$I_{A \cup B} = I_A + I_B - I_A I_B = 1 - (1 - I_A)(1 - I_B). \tag{2.6}$$

By iteration of the identities (2.5) and (2.6) we obtain the inclusion–exclusion formula for indicators,

$$I_{A_1 \cup A_2 \cup \cdots \cup A_n} = 1 - (1 - I_{A_1})(1 - I_{A_2}) \cdots (1 - I_{A_n})$$
$$= \sum_i I_{A_i} - \sum_{i<j} I_{A_i \cap A_j} + \sum_{i<j<k} I_{A_i \cap A_j \cap A_k} + \cdots . \qquad (2.7)$$

A subset G of L that is closed under finite intersections is said to be a *generating set* of L when every element of L is a finite union of elements of G. Using the inclusion–exclusion formula for indicators, it can be shown that every L-simple function can be written as a finite linear combination

$$f = \sum_{i=1}^{r} \beta_i I_{B_i}, \qquad (2.8)$$

where $B_i \in G$. A real-valued function ν on G is called a *valuation* on G provided that ν satisfies identities (2.1) and (2.2) for all sets $A, B \in G$ such that $A \cup B \in G$ as well. Note that, since G need not be closed under unions, identity (2.1) does not make sense for all pairs of sets $A, B \in G$. Hence, there is no reason to assume that the identities (2.3) should hold for ν if $n > 2$.

Since every element $B \in L$ can be expressed as a union $B = B_1 \cup \cdots \cup B_n$ with $B_1, \ldots, B_n \in G$, we can attempt to extend ν to a valuation μ on all of L by setting

$$\mu(B) = \sum_i \nu(B_i) - \sum_{i<j} \nu(B_i \cap B_j) + \cdots , \qquad (2.9)$$

as is suggested by (2.3). There remains to check that $\mu(B)$ is well defined, in the case that B could be expressed as a union of elements of G in more than one way.

Given a valuation μ on G, define the *integral* with respect to μ as follows. For an L-simple function $f = \alpha_1 I_{A_1} + \cdots + \alpha_k I_{A_1}$, with $A_i \in G$ for $1 \leq i \leq k$, define

$$\int f \, d\mu = \sum_{i=1}^{k} \alpha_i \mu(A_i). \qquad (2.10)$$

In general, a simple function f has infinitely many expressions of the form (2.4), for A_i in G. Consequently we must check that the integral in (2.10) is well defined.

2.2 Groemer's integral theorem

The existence of the extension (2.9) and the integral (2.10) turn out to be equivalent properties of μ, a nontrivial fact stated formally as follows.

Theorem 2.2.1 (Groemer's integral theorem) *Let G be a generating set for a lattice L, and let μ be a valuation on G. The following statements are equivalent.*

(i) *μ extends uniquely to a valuation on L.*

(ii) *μ satisfies the inclusion–exclusion identities*

$$\mu(B_1 \cup B_2 \cup \cdots \cup B_n) = \sum_i \mu(B_i) - \sum_{i<j} \mu(B_i \cap B_j) + \cdots, \quad (2.11)$$

whenever $B_i \in G$ and $B_1 \cup B_2 \cup \cdots \cup B_n \in G$, and for all $n \geq 2$.

(iii) *μ defines an integral on the vector space of linear combinations of indicator functions of sets in L.*

Proof We prove the implications **(i)** ⇒ **(ii)** ⇒ **(iii)** ⇒ **(i)**.

If μ extends uniquely to a valuation on all of L, then **(ii)** follows from an iteration of identity (2.1). Therefore, **(i)** implies **(ii)**.

To show that **(ii)** implies **(iii)**, suppose there exist non-empty distinct $K_1, \ldots, K_m \in G$ and nonzero real numbers $\alpha_1, \ldots, \alpha_m$ such that

$$\sum_{i=1}^{m} \alpha_i I_{K_i} = 0, \quad (2.12)$$

while

$$\sum_{i=1}^{m} \alpha_i \mu(K_i) \neq 0. \quad (2.13)$$

Let $L_1 = K_1$, ..., $L_m = K_m$, $L_{m+1} = K_1 \cap K_2$, $L_{m+2} = K_1 \cap K_3$, and so on, to define a list $L_1, L_2, \ldots, L_p$, comprising all possible intersections of the sets K_i. Since G is closed under intersections, $L_i \in G$ for all i. Note also that the collection $\{L_i\}$ is closed under intersections.

Suppose that

$$\sum_{i=q}^{p} \alpha_i I_{L_i} = 0, \quad (2.14)$$

while

$$\sum_{i=q}^{p} \alpha_i \mu(L_i) \neq 0, \quad (2.15)$$

where $\alpha_q \neq 0$. Choose an instance of these equations such that q is maximal. It follows from (2.12) and (2.13) that $q \geq 1$, while the conditions (2.14) and (2.15) imply that $q < p$.

Suppose that $x \in L_q - \bigcup_{j=q+1}^{p} L_j$. Then (2.14) implies that

$$\alpha_q = \sum_{i=q}^{p} \alpha_i I_{L_i}(x) = 0,$$

contradicting our assumption. It follows that

$$L_q \subseteq L_{q+1} \cup \cdots \cup L_p,$$

so that

$$L_q = L_q \cap (L_{q+1} \cup \cdots \cup L_p) = (L_q \cap L_{q+1}) \cup \cdots \cup (L_q \cap L_p).$$

For $i > q$, note that $L_q \cap L_i = L_j$, where $j > q$. Using the principle of inclusion–exclusion **(ii)** we obtain

$$\sum_{i=q}^{p} \alpha_i \mu(L_i) = \alpha_q \mu \left(\bigcup_{i=q+1}^{p} (L_q \cap L_i) \right) + \sum_{i=q+1}^{p} \alpha_i \mu(L_i) = \sum_{i=q+1}^{p} \beta_i \mu(L_i),$$

so that

$$\sum_{i=q+1}^{p} \beta_i \mu(L_i) \neq 0 \tag{2.16}$$

by (2.15), where each β_i is obtained by collecting the terms containing $\mu(L_i)$. Meanwhile, application of the same inclusion–exclusion procedure to the indicator functions yields

$$\sum_{i=q}^{p} \alpha_i I_{L_i} = \alpha_q I_{\cup_{i=q+1}^{p}(L_q \cap L_i)} + \sum_{i=q+1}^{p} \alpha_i I_{L_i} = \sum_{i=q+1}^{p} \beta_i I_{L_i}$$

so that

$$\sum_{i=q+1}^{p} \beta_i I_{L_i} = 0 \tag{2.17}$$

by (2.14). Together (2.16) and (2.17) contradict the maximality of q. This completes the proof that **(ii)** implies **(iii)**.

To show that **(iii)** implies **(i)**, suppose that the function μ defines an integral on the space of L-simple functions. For $A \in L$ define

$$\mu(A) = \int I_A \, d\mu.$$

The linearity of the integral together with the identity (2.6) implies that this extension of μ is a valuation on L. □