1

Boolean functions and key concepts

In this first chapter, we set the stage for the book by presenting many of its key concepts of the book and stating a number of important theorems that we prove here.

1.1 Boolean functions

Definition 1.1 A **Boolean function** is a function from the hypercube $\Omega_n := \{-1,1\}^n$ into either $\{-1,1\}$ or $\{0,1\}$.

Ω_n is endowed with the uniform measure $\mathbb{P} = \mathbb{P}^n = (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}$ and $\mathbb{E}$ denotes the corresponding expectation. Occasionally, Ω_n will be endowed with the general product measure $\mathbb{P}_p = \mathbb{P}_p^n = ((1-p)\delta_{-1} + p\delta_1)^{\otimes n}$ but in such cases the p is made explicit. $\mathbb{E}_p$ then denotes the corresponding expectation.

An element of Ω_n is denoted by either ω or ω_n and its n bits by $x_1,\dots,x_n$ so that $\omega = (x_1,\dots,x_n)$.

For the range, we choose to work with $\{-1,1\}$ in some contexts and $\{0,1\}$ in others, and at some specific places we even relax the Boolean constraint (i.e., that the function takes only two possible values). In these cases (which are clearly identified), we consider instead real-valued functions $f\colon \Omega_n \to \mathbb{R}$.

A Boolean function f is canonically identified with a subset A_f of Ω_n via $A_f := \{\omega : f(\omega) = 1\}$.

Remark Often, Boolean functions are defined on $\{0,1\}^n$ rather than $\Omega_n = \{-1,1\}^n$. This does not make any fundamental difference but, as we see later, the choice of $\{-1,1\}^n$ turns out to be more convenient when one wishes to apply Fourier analysis on the hypercube.

1.2 Some examples

We begin with a few examples of Boolean functions. Others appear throughout this chapter.

Example 1.2 (Dictator)

$$\mathbf{DICT}_n(x_1,\dots,x_n) := x_1 .$$

The first bit determines what the outcome is.

Example 1.3 (Parity)

$$\mathbf{PAR}_n(x_1,\dots,x_n) := \prod_{i=1}^n x_i .$$

This Boolean function's output is determined by whether the number of -1's in ω is even or odd.

These two examples are in some sense trivial, but they are good to keep in mind because in many cases they turn out to be the "extreme cases" for properties concerning Boolean functions.

The next rather simple Boolean function is of interest in social choice theory.

Example 1.4 (Majority function) Let n be odd and define

$$\mathbf{MAJ}_n(x_1,\dots,x_n) := \mathrm{sign}\Big(\sum_{i=1}^n x_i\Big) .$$

Following are two further examples that also arise in our discussions.

Example 1.5 (Iterated 3-Majority function) Let $n = 3^k$ for some integer k. The bits are indexed by the leaves of a rooted 3-ary tree (so the root has degree 3, the leaves have degree 1, and all others have degree 4) with depth k. Apply Example 1.4 (with $n = 3$) iteratively to obtain values at the vertices at level $k-1$, then level $k-2$, and so on until the root is assigned a value. The root's value is then the output of f. For example, when $k = 2$, $f(-1,1,1;1,-1,-1;-1,1,-1) = -1$. The recursive structure of this Boolean function enables explicit computations for various properties of interest.

Example 1.6 (Clique containment) If $r = \binom{n}{2}$ for some integer n, then Ω_r can be identified with the set of labeled graphs on n vertices. (Bit x_i is 1 if and only if the ith edge is present.) Recall that a **clique** of size k of a graph $G = (V,E)$ is a complete graph on k vertices embedded in G.

Cambridge University Press
978-1-107-43255-0 — Noise Sensitivity of Boolean Functions and Percolation
Christophe Garban , Jeffrey E. Steif
Excerpt
More Information

Now for any $1 \leq k \leq \binom{n}{2} = r$, let $\mathbf{CLIQ}_n^k$ be the indicator function of the event that the random graph G_ω defined by $\omega \in \Omega_r$ contains a clique of size k. Choosing $k = k_n$ so that this Boolean function is nondegenerate turns out to be a rather delicate issue. The interesting regime is near $k_n \approx 2\log_2(n)$. See Exercise 1.9 for this "tuning" of $k = k_n$. It turns out that for most values of n, the Boolean function $\mathbf{CLIQ}_n^k$ is degenerate (i.e., has small variance) for all values of k. However, there is a sequence of n for which there is some $k = k_n$ for which $\mathbf{CLIQ}_n^k$ is nondegenerate.

1.3 Pivotality and influence

This section contains our first fundamental concepts. We abbreviate $\{1, \dots, n\}$ as $[n]$.

Definition 1.7 Given a Boolean function f from Ω_n into either $\{-1, 1\}$ or $\{0, 1\}$ and a variable $i \in [n]$, we say that i **is pivotal for** f **for** ω if $f(\omega) \neq f(\omega^i)$ where ω^i is ω but flipped in the ith coordinate. Note that this event $\{f(\omega) \neq f(\omega^i)\}$ is measurable with respect to $\{x_j\}_{j \neq i}$.

Definition 1.8 The **pivotal set**, $\mathcal{P}$, for f is the random set of $[n]$ given by

$$\mathcal{P}(\omega) = \mathcal{P}_f(\omega) := \{i \in [n] : \ i \text{ is pivotal for } f \text{ for } \omega\}.$$

Expressed in words, the pivotal set is the (random) set of bits with the property that if you flip the bit, then the function output changes.

Definition 1.9 The **influence** of the ith bit, $\mathbf{I}_i(f)$, is defined by

$$\mathbf{I}_i(f) := \mathbb{P}(\ i \text{ is pivotal for } f\ ) = \mathbb{P}(f(\omega) \neq f(\omega^i)) = \mathbb{P}(i \in \mathcal{P})$$

Let also the **influence vector**, $\mathbf{Inf}(f)$, be the collection of all the influences: i.e. $\{\mathbf{I}_i(f)\}_{i \in [n]}$.

Expressed in words, the influence of the ith bit is the probability that, on flipping this bit, the function output changes. This concept originally arose in political science to measure the power of different voters and is often called the Banzhaf power index (see (B65)) but in fact the concept arose earlier (see (P46)) in the work of L. Penrose.

Definition 1.10 The **total influence**, $\mathbf{I}(f)$, is defined by

$$\mathbf{I}(f) := \sum_i \mathbf{I}_i(f) = \|\mathbf{Inf}(f)\|_1 = \mathbb{E}(|\mathcal{P}|).$$

It would now be instructive to compute these quantities for Examples 1.2–1.4. See Exercise 1.1.

Later, we will need the last two concepts in the context where our probability measure is $\mathbb{P}_p$ instead. We now give the corresponding definitions.

Definition 1.11 The **influence vector at level** p, $\{\mathbf{I}_i^p(f)\}_{i\in[n]}$, is defined by

$$\mathbf{I}_i^p(f) := \mathbb{P}_p(\ i \text{ is pivotal for } f\ ) = \mathbb{P}_p(f(\omega) \neq f(\omega^i)) = \mathbb{P}_p(i \in \mathcal{P}).$$

Definition 1.12 The **total influence at level** p, $\mathbf{I}^p(f)$, is defined by

$$\mathbf{I}^p(f) := \sum_i \mathbf{I}_i^p(f) \ = \mathbb{E}_p(|\mathcal{P}|).$$

It turns out that the total influence has a geometric-combinatorial interpretation as the size of the so-called edge boundary of the corresponding subset of the hypercube. See Exercise 1.4.

Remark Aside from its natural definition as well as its geometric interpretation as measuring the edge boundary of the corresponding subset of the hypercube, the notion of *total influence* arises very naturally when one studies **sharp thresholds** for *monotone functions* (to be defined in Chapter 3). Roughly speaking, as we see in detail in Chapter 3, for a monotone event A, $d\mathbb{P}_p[A]/dp$ is the total influence at level p (this is the Margulis–Russo formula). This tells us that the speed at which things change from the event A “almost surely” *not* occurring to the case where it “almost surely” *does* occur is very sudden if the Boolean function happens to have a large total influence.

1.4 The Kahn–Kalai–Linial theorems

This section addresses the following question. Does there always exist some variable i with (reasonably) large influence? In other words, for large n, what is the smallest value (as we vary over Boolean functions) that the largest influence (as we vary over the different variables) can take on?

Because for the constant function all influences are 0, and the function that is 1 only if all the bits are 1 has all influences $1/2^{n-1}$, clearly we want to deal with functions that are reasonably balanced (meaning having variances not so close to 0) or, alternatively, obtain lower bounds on the maximal influence in terms of the variance of the Boolean function.

The first result in this direction is the following. A sketch of the proof is given in Exercise 1.5.

Theorem 1.13 (Discrete Poincaré) *If f is a Boolean function mapping Ω_n into $\{-1,1\}$, then*

$$\mathrm{Var}(f) \le \sum_i \mathbf{I}_i(f).$$

It follows that there exists some i such that

$$\mathbf{I}_i(f) \ge \mathrm{Var}(f)/n.$$

This gives a first answer to our question. For reasonably balanced functions, there is some variable whose influence is at least of order $1/n$. *Can we find a better "universal" lower bound on the maximal influence?* Note that for Example 1.4 all the influences are of order $1/\sqrt{n}$ (and the variance is 1). Therefore, in terms of our question, the universal lower bound we are looking for should lie somewhere between $1/n$ and $1/\sqrt{n}$. The following celebrated result improves by a logarithmic factor on the $\Omega(1/n)$ lower bound.

Theorem 1.14 (KKL88) *There exists a universal $c > 0$ such that if f is a Boolean function mapping Ω_n into $\{0,1\}$, then there exists some i such that*

$$\mathbf{I}_i(f) \ge c\mathrm{Var}(f)(\log n)/n.$$

What is remarkable about this theorem is that this "logarithmic" lower bound on the maximal influence turns out to be *sharp*! This is shown by the following example by Ben-Or and Linial.

Example 1.15 (Tribes) Partition $[n]$ into disjoint blocks of length $\log_2(n) - \log_2(\log_2(n))$ with perhaps some leftover debris. Define f_n to be 1 if there exists at least one block that contains all 1's, and 0 otherwise.

One can check that the sequence of variances stays bounded away from 0 and that all the influences (including of course those belonging to the debris which are equal to 0) are smaller than $c(\log n)/n$ for some $c < \infty$. See Exercise 1.3. Hence Theorem 1.14 is indeed sharp. We mention that in (BOL87), the example of Tribes was given and the question of whether $\log n/n$ was sharp was asked.

Our next result tells us that if all the influences are "small," then the total influence is large.

Theorem 1.16 (KKL88) *There exists $c > 0$ such that if f is a Boolean function mapping Ω_n into $\{0,1\}$ and $\delta := \max_i \mathbf{I}_i(f)$, then*

$$\mathbf{I}(f) \ge c\,\mathrm{Var}(f)\log(1/\delta).$$

Or equivalently,

$$\|\mathbf{Inf}(f)\|_1 \geq c\,\mathrm{Var}(f)\log\frac{1}{\|\mathbf{Inf}(f)\|_\infty}.$$

One can in fact talk about the influence of a set of variables rather than the influence of a single variable.

Definition 1.17 Given $S \subseteq [n]$, the **influence of** S, $\mathbf{I}_S(f)$, is defined by

$$\mathbf{I}_S(f) := \mathbb{P}(\,f \text{ is not determined by the bits in } S^c).$$

It is easy to see that when S is a single bit, this corresponds to our previous definition. The following is also proved in (KKL88). We do not give the proof in this book.

Theorem 1.18 (KKL88) *Given a sequence f_n of Boolean functions mapping Ω_n into $\{0,1\}$ such that $0 < \inf_n \mathbb{E}_n(f) \leq \sup_n \mathbb{E}_n(f) < 1$ and any sequence a_n going to ∞ arbitrarily slowly, then there exists a sequence of sets $S_n \subseteq [n]$ such that $|S_n| \leq a_n n/\log n$ and $\mathbf{I}_{S_n}(f_n) \to 1$ as $n \to \infty$.*

Theorems 1.14 and 1.16 are proved in Chapter 5.

1.5 Noise sensitivity and noise stability

This section introduces our second set of fundamental concepts.

Let ω be uniformly chosen from Ω_n and let ω_ϵ be ω but with each bit independently "re-randomized" with probability ϵ. To rerandomize a bit means that, independently of everything else, the value of the bit is rechosen to be 1 or -1, each with probability $1/2$. Note that ω_ϵ then has the same distribution as ω.

The following definition is *central*. Let m_n be an increasing sequence of integers and let $f_n : \Omega_{m_n} \to \{\pm 1\}$ or $\{0,1\}$.

Definition 1.19 The sequence $\{f_n\}$ is **noise sensitive** if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{E}[f_n(\omega) f_n(\omega_\epsilon)] - \mathbb{E}[f_n(\omega)]^2 = 0. \tag{1.1}$$

Because f_n takes just two values, this definition says that the random variables $f_n(\omega)$ and $f_n(\omega_\epsilon)$ are asymptotically independent for $\epsilon > 0$ fixed and n large. We see later that (1.1) holds for one value of $\epsilon \in (0,1)$ if and only if it holds for all such ϵ. The following notion captures the opposite situation, where the two events are close to being the same event if ϵ is small, uniformly in n.

Definition 1.20 The sequence $\{f_n\}$ is **noise stable** if

$$\lim_{\epsilon \to 0} \sup_n \mathbb{P}(f_n(\omega) \neq f_n(\omega_\epsilon)) = 0.$$

It is an easy exercise to check that a sequence $\{f_n\}$ is both noise sensitive and noise stable if and only if it is degenerate in the sense that the sequence of variances $\{\mathrm{Var}(f_n)\}$ goes to 0. Note also that a sequence of Boolean functions could be neither noise sensitive nor noise stable (see Exercise 1.11).

It is also an easy exercise to check that Example 1.2 (Dictator) is noise stable and Example 1.3 (Parity) is noise sensitive. We see later, when Fourier analysis is brought into the picture, that these examples are the two opposite extreme cases. For the other examples, it turns out that Example 1.4 (Majority) is noise stable, while Examples 1.5,1.6, and 1.15 are all noise sensitive. See Exercises 1.6–1.9. In fact, there is a deep theorem (see (MOO10)) that says in some sense that, among all low-influence Boolean functions, Example 1.4 (Majority) is the most stable.

In Figure 1.1, we give a slightly *impressionistic* view of what "noise sensitivity" is.

1.6 The Benjamini–Kalai–Schramm noise sensitivity theorem

We now come to the main theorem concerning noise sensitivity.

Theorem 1.21 (BKS99) *If*

$$\lim_n \sum_k \mathbf{I}_k(f_n)^2 = 0, \tag{1.2}$$

then $\{f_n\}$ is noise sensitive.

Remark The converse of Theorem 1.21 is clearly false, as shown by Example 1.3. However, the converse is true for **monotone functions** (defined in the next chapter), as we see in Chapter 4.

Theorem 1.21 allows us to conclude noise sensitivity of many of the examples introduced in this first chapter. See Exercise 1.10. This theorem is proved in Chapter 5.

1.7 Percolation crossings: Our final and most important example

We have saved our most important example to the end. This book would not have been written were it not for this example and for the results that have been proved about it.

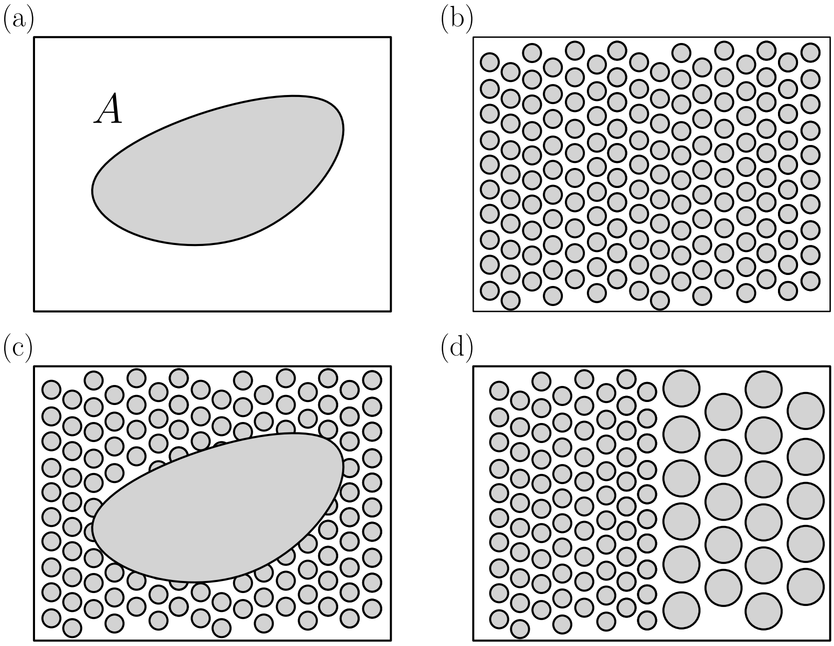
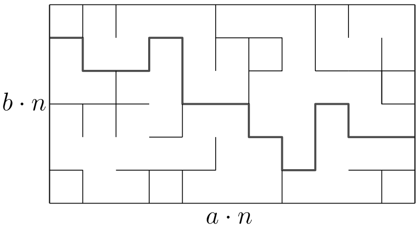


Figure 1.1 Consider the following "experiment": take a bounded domain in the plane, say a rectangle, and consider a measurable subset A of this domain. What would be an analog of the definitions of *noise sensitive* or *noise stable* in this case? Start by sampling a point x uniformly in the domain according to Lebesgue measure. Then apply some noise to this position x to end up with a new position x_ϵ. One can think of many natural "noising" procedures here. For example, let x_ϵ be a uniform point in the ball of radius ϵ around x, conditioned to remain in the domain. (This is not quite perfect as this procedure does not exactly preserve Lebesgue measure, but don't worry about this.) The natural analog of Definitions 1.19 and 1.20 is to ask whether $1_A(x)$ and $1_A(x_\epsilon)$ are decorrelated or not.
Question: According to this analogy, what are the sensitivity and stability properties of the sets A sketched in pictures (a) to (d)? Note that to match with Definitions 1.19 and 1.20, one should consider sequences of subsets $\{A_n\}$ instead, as noise sensitivity is an asymptotic notion.

Consider percolation on $\mathbb{Z}^2$ at the critical point $p_c(\mathbb{Z}^2) = 1/2$. (See Chapter 2 for a brief review of the model.) At this critical point, there is no infinite cluster, but somehow clusters are "large" and there are clusters at all scales. This can be seen using duality or with the RSW Theorem 2.1. To understand the geometry of the critical picture, the following large-scale *observables* turn out to be very useful: Let Ω be a piecewise smooth domain with two disjoint open arcs ∂_1 and ∂_2 on its boundary $\partial\Omega$. For each $n \geq 1$, we consider the scaled domain $n\Omega$. Let A_n be the event that there is an open path in ω from $n\partial_1$ to $n\partial_2$ which stays inside $n\Omega$. Such events are called **crossing events**. They are naturally associated with Boolean functions whose entries are indexed by the set of edges inside $n\Omega$ (there are $O(n^2)$ such variables).

For simplicity, consider the particular case of rectangle crossings.

Example 1.22 (Percolation crossings)



Let $a, b > 0$ and let us consider the rectangle $[0, a \cdot n] \times [0, b \cdot n]$. The left-to-right crossing event corresponds to the Boolean function $f_n : \{-1,1\}^{O(1)n^2} \to \{0,1\}$ defined as follows:

$$f_n(\omega) := \begin{cases} 1 & \text{if there is a left–right crossing} \\ 0 & \text{otherwise} \end{cases}$$

We later prove that this sequence of Boolean functions $\{f_n\}$ is noise sensitive. This means (see Exercise 4.7) that if a percolation configuration $\omega \sim \mathbb{P}_{p_c=1/2}$ is given, one typically cannot predict anything about the large-scale clusters of the slightly perturbed percolation configuration ω_ϵ where only an ϵ-fraction of the edges has been resampled.

Remark The same statement holds for the more general crossing events described above (i.e., in $(n\Omega, n\partial_1, n\partial_2)$).

1.8 A dynamical consequence of noise sensitivity

One can consider a continuous time random walk $\{\omega_t\}_{t\geq 0}$ (implicitly depending on n which we suppress) on $\Omega_n := \{-1,1\}^n$ obtained by having each variable independently re-randomize at the times of a rate 1 Poisson process (so that the times between rerandomizations are independent exponential times with parameter 1). The stationary distribution is of course our usual probability measure, which is a product measure with 1 and -1 equally likely. Starting from stationarity, observe that the joint distribution of ω_s and ω_{s+t} is the same as the joint distribution of ω and ω_ϵ introduced earlier where $\epsilon = 1 - e^{-t}$.

Considering next a sequence of Boolean functions $\{f_n\}_{n\geq 1}$ mapping Ω_n into, say, $\{0,1\}$, we obtain a sequence of processes $\{g_n(t)\}$ defined by $g_n(t) := f_n(\omega_t)$. The following general result was proved in (BKS99) for the specific case of percolation crossings; however, their proof applies verbatim in this general context.

Theorem 1.23 (BKS99) *Let $\{f_n\}_{n\geq 1}$ be a sequence of Boolean functions that is noise sensitive and satisfies $\delta_0 \leq \mathbb{P}(f_n(\omega) = 1) \leq 1 - \delta_0$ for all n for some $\delta_0 > 0$. Let S_n be the set of times in $[0,1]$ at which $g_n(t)$ changes its value. Then $|S_n| \to \infty$ in probability as $n \to \infty$.*

Proof We first claim that for all $0 \leq a < b \leq 1$,

$$\lim_{n\to\infty} \mathbb{P}(S_n \cap [a,b] = \emptyset) = 0. \tag{1.3}$$

Let $W_{n,\epsilon} := \{\omega : \mathbb{P}[f_n(\omega_\epsilon) = 1|\omega] \in [0, \delta_0/2] \cup [1-\delta_0/2, 1]\}$. The noise sensitivity assumption, the (δ_0-)nondegenericity assumption, and Exercise 4.7 (which gives an alternative description of noise sensitivity) imply that for each $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(W_{n,\epsilon}) = 0.$$

Fix $\gamma > 0$ arbitrarily. Choose k so that $(1-\delta_0/2)^k < \gamma/2$ and then choose $\epsilon := (b-a)/k$. Finally choose N so that for all $n \geq N$, $\mathbb{P}(W_{n,\epsilon}) \leq \gamma\delta_0/4$. Let $a = t_0 < t_1 < t_2 < \cdots < t_k = b$, where each $t_i - t_{i-1}$ equals $(b-a)/k$. We then have for $n \geq N$,

$$\begin{aligned}
&\mathbb{P}(S_n \cap [a,b] = \emptyset) \\
&\leq \mathbb{P}(\omega_{t_{k-1}} \in W_{n,\epsilon}) + \mathbb{E}[I_{\{\omega_{t_{k-1}} \notin W_{n,\epsilon}\}} \mathbb{P}[S_n \cap [a,b] = \emptyset \mid \omega_{t_{k-1}}]] \\
&\leq \frac{\gamma\delta_0}{4} \mathbb{E}[I_{\{\omega_{t_{k-1}} \notin W_{n,\epsilon}\}} \mathbb{P}[S_n \cap [t_{k-1}, b] = \emptyset \mid \omega_{t_{k-1}}] \mathbb{P}[S_n \cap [a, t_{k-1}] = \emptyset \mid \omega_{t_{k-1}}]]
\end{aligned}$$

using the Markov property. If $\omega_{t_{k-1}} \notin W_{n,\epsilon}$, then

$$\mathbb{P}[S_n \cap [t_{k-1}, b] = \emptyset \mid \omega_{t_{k-1}}] \leq \mathbb{P}[g_n(b) = g_n(t_{k-1}) \mid \omega_{t_{k-1}}] \leq 1 - \delta_0/2.$$

This yields

$$\mathbb{P}(S_n \cap [a,b] = \emptyset) \leq \frac{\gamma\delta_0}{4} + \left(1 - \frac{\delta_0}{2}\right) \mathbb{P}(S_n \cap [a, t_{k-1}] = \emptyset).$$

Continuing by induction $k-1$ more times yields

$$\mathbb{P}(S_n \cap [a,b] = \emptyset) \leq \frac{\gamma}{2} + \left(1 - \frac{\delta_0}{2}\right)^k < \gamma.$$