Functions arising by coin flipping

Johan Wästlund

May 1999 [*]

Abstract

We repeatedly toss a biased coin where the probability is p of observing "heads". The probability of a certain event can be regarded as a function of p. L. Lovász asked for a characterization of the functions occurring as the probability of some "reasonable" event. We state this problem in precise terms by introducing the concept of an "observable" event. Our main theorem states that a function $(0,1) \to (0,1)$ represents the probability of an observable event if and only if it is continuous, and does not tend to 0 or 1 faster than polynomially, as p tends to 0 or 1. We also give several simple examples of events having probabilities given by elementary functions.

1 Introduction

Suppose that we repeatedly toss a coin which is unfair in the sense that the probability of observing "heads" is p, and the probability of observing "tails"

[*] Note: A first version of this paper was written in 1997 and included in my PhD thesis in 1999. After my PhD dissertation the last section on "finite" events was added. Later I was informed that my main result, Theorem 1.2, was established already in 1994 by M. S. Keane and G. L. O'Brien, *A Bernoulli factory*, ACM Transactions on Modeling and Computer Simulation **4**, Issue 2, 213–219 (thanks to Yuval Peres for pointing this out). Of course I do not blame Lovász for giving me a problem that was already solved, he mentioned it very informally at a dinner in his house. Actually it is a very nice problem and I do not at all regret having spent time on it. Anyway, this is the reason the paper was never published in a journal. However, in the years that have passed, several people have asked me about this paper. Therefore I here provide it just as it was when I wrote it in 1999, without any further apologies for omitting necessary references. Anyone interested in pursuing the subject further should also consult the more recent publications by Elchanan Mossel, Serban Nacu, Yuval Peres and others.

is $1 - p$. The probability of a given event can then be regarded as a function of p.

Some polynomials and rational functions occur as probabilities of events which are easy to describe. For example, the probability of observing two heads in the first two tosses is p^2, and the probability that the first tails occurs after an odd number of tosses is $1/(1 + p)$.

L. Lovász [5] posed the problem of determining the class of functions representing the probability of some "reasonable" event. In particular, he asked if the function $p \mapsto 2p$, for p in some suitable interval, say $0 < p < 0.4$, will arise in this way. In this paper we state the problem in precise terms, and give a complete solution. We also give some simple examples of events having probabilities given by transcendental functions like e^{-p}.

We represent the possible outcomes of the coin flipping process by the set Ω of all sequences $X = (X_1, X_2, X_3, \dots)$ such that $X_i \in \{0, 1\}$ for every i. A specific outcome of the process is represented by the sequence X such that $X_i = 1$ if the i:th toss results in "heads", and $X_i = 0$ if the i:th toss results in "tails".

The product topology on Ω is defined by taking the sets $U(a_1, \dots, a_k) = \{X \in \Omega \ : \ X_i = a_i, 1 \le i \le k\}$ of all 0-1 sequences beginning by $a_1, \dots, a_k$, as a basis for the open sets.

Every value of p, $0 < p < 1$, determines a probability measure P_p on Ω by the requirement that the X_i's are independent, and $P_p(X_i = 1) = p$ for every i. We make a few remarks on this construction.

We require that

$$P_p(U(a_1, \dots, a_k)) = \prod_{i=1}^{k} p^{a_i}(1 - p)^{1-a_i}.$$

Since every open set is a disjoint countable union of basic open sets, this determines $P_p(E)$ for every open set E. The *outer* measure of any subset S of Ω is now defined by

$$P_p(S) = \inf\{P_p(E){:}E \text{ is open and } S \subseteq E\}.$$

S is *measurable* with respect to P_p, if $P_p(S)+P_p(S^c) = 1$. It may happen that a set is measurable with respect to P_p for one value of p, and non-measurable for another. When we speak of an *event*, we mean a subset of Ω which is measurable with respect to P_p for every p.

For an introduction to probability spaces, as well as a discussion of coin flipping, we refer to [2].

Since we are not actually tossing the coin an infinite number of times, we are primarily interested in events which can be decided by observing the outcome of a finite number of tosses. This motivates the following definition.

Definition 1. An event E is *observable* if there are two sets A and B of finite 0-1 strings such that

(i) for every p in $(0,1)$, with probability 1 one and only one of the strings in A and B will occur as an initial segment of $X_1, X_2, X_3 \dots$, and

(ii) if this string is in A, E occurs, while if it is in B, E does not occur. (In the unlikely event that no string in the two lists occurs as an initial segment of $X_1, X_2, X_3 \dots$, E may or may not occur).

This notion is related to the topology of the probability space Ω. To say that an event E is open is to say that if $X = (X_1, X_2, X_3, \dots)$ belongs to E, then there is some neighborhood of X, which can be taken to be $U(X_1, \dots, X_k)$ for some k, which lies entirely in E. In terms of coin flipping, this means that if the event E occurs, then we will know this after a finite number of tosses.

In general, the interior of an event E, Int$E$, consists of those outcomes for which we can decide after some finite number of coin tosses that E occurs. The boundary BdE of E consists of those infinite 0-1-sequences for which we never know whether E occurs or not.

Proposition 1.1. *An event E is observable if and only if BdE is a null set for every p in* $(0,1)$.

Proof. Suppose E is observable. Then BdE is the set of all X such that no initial segment of $(X_1, X_2, X_3, \dots)$ belongs to A or B. By (i) of Definition 1, this is a null set for every p.

Suppose on the other hand that BdE is a null set for every p in $(0,1)$. Then we let A be the set of all $(a_1, \dots, a_k)$ which are minimal with the property that $U(a_1, \dots, a_k) \subseteq E$, and let B be the set of all $(b_1, \dots, b_k)$ which are minimal with the property that $U(b_1, \dots, b_k) \cap E = \emptyset$. Since Definition 1 is satisfied, E is observable. □

Note that an observable event is not necessarily open, and that an open event is not necessarily observable. For an example of an open event which is not observable, Let $E = \cup_{n=1}^{\infty} E_n$, where E_n is the event that $X_1 =$

$X_{n+1}, X_2 = X_{n+2}, \dots, X_n = X_{2n}$. E is clearly open and dense. Since the events E_n are not mutually exclusive, we have the strict inequality $P_{1/2}(E) < \sum_{n=1}^{\infty} P_{1/2}(E_n) = 1/2 + 1/4 + 1/8 + \cdots = 1$. Hence $\mathrm{Bd}E$, which equals E^c, has positive probability (at least for $p = 1/2$), which means that E is not observable.

However, a function representing the probability of an observable event E also represents the probability of an open event, since by Proposition 1.1, $P_p(E) = P_p(\mathrm{Int}E)$.

Note also that if $\mathrm{Bd}E$ has measure 0 for every p, then E is automatically an event, since $\mathrm{Int}E$ is open and therefore measurable for every p.

We now formulate Lovász' question as follows:

For which functions $f : (0,1) \to [0,1]$ does there exist an observable event E, such that $P_p(E) = f(p)$ for every p in $(0,1)$?

Without any restriction on the event E, the question would be more or less trivial. As we shall see in a moment, if f is any function $(0,1) \to [0,1]$, then there is an event E such that $P_p(E) = f(p)$.

We remark that if f is the probability of an open event, and $f(p) = 0$ for some $p \in (0,1)$, then f is identically 0, since any $U(a_1, \dots, a_k)$ has positive probability for every p. Similarly, if f is the probability of an observable event, and $f(p)$ is 0 or 1 for some p, then f is identically 0 or identically 1 respectively. Hence the restrictions of the range of f in Theorems 1.2 and 1.3 below.

In Section 2, we will discuss several methods for constructing observable events with given probability functions. As an example, we construct an event having probability $\exp\left(-\sqrt{\cos p}\right)$. In Section 3, we prove the following theorem, which answers Lovász' question.

Theorem 1.2. *A function $f : (0,1) \to (0,1)$ represents the probability of an observable event if and only if f is continuous, and for some k,*

$$p^k(1-p)^k < f(p) < 1 - p^k(1-p)^k.$$

Note that although f has to be continuous on the *open* interval $(0,1)$, it does not have to tend to a limit as p tends to 0 or 1. For example, the function

$$\frac{1}{2} + \left(p - \frac{1}{2}\right) \sin \frac{1}{p(1-p)},$$

which oscillates wildly near $p = 0$ and $p = 1$, still satisfies the condition of Theorem 1.2.

We then prove the analogous theorem for open events. We remind the reader that a function f is *lower semicontinuous* if $f(a) \leq \lim\inf_{x \to a} f(x)$ for every a in its domain.

Theorem 1.3. *A function $f : (0, 1) \to (0, 1]$ represents the probability of an open event if and only if f is lower semicontinuous, and for some k,*

$$f(p) > p^k(1 - p)^k$$

In particular, this shows that there is an open event with probability given by

$$f(p) = \begin{cases} 1, & \text{if } p \text{ is irrational} \\ 1 - 1/n, & \text{if } p \text{ is rational with minimal denominator } n \end{cases}$$

2 Some methods for constructing an event with a given probability

Our first nontrivial observation is the following, which seems to be "folklore".

Theorem 2.1. *There is an observable event which has constant probability 1/2.*

In other words, we can use an unfair coin to simulate a fair coin, even without knowing the probability p. There is a simple way of doing this, described in [6]: Flip the coin twice. If the outcome is heads-tails, we answer "heads". If the outcome is tails-heads, we answer "tails". If the two tosses give the same result, we repeat the procedure.

Formally, we introduce a new random variable $Y : \Omega \to \{0, 1\}$, by letting $Y = X_{2n}$, where n is minimal with $X_{2n-1} \neq X_{2n}$, if there is such an n, and letting $Y = 0$ or 1 arbitrarily if $X_{2n-1} = X_{2n}$ for every n. We claim that the event $E = Y^{-1}\{1\}$, that is, the event "$Y = 1$", is an observable event with constant probability 1/2.

To see that E is observable, note that

$$\text{Bd}(Y^{-1}\{1\}) = \{X \in \Omega : X_{2n-1} = X_{2n} \text{ for every } n\},$$

which is clearly a null set for every p in $(0,1)$. To see that $Y^{-1}\{1\}$ has measure $1/2$, it now suffices to show that $P(\mathrm{Int}E) = P(\mathrm{Int}E^c)$, which is clear by symmetry, since the outcomes heads-tails and tails-heads are equally probable.

This principle is used in tie-breaks in racket games like tennis and table tennis. In general, it is an advantage to serve. It is still reasonable to demand that if both players are equally skilled, that is, if the probability p that the server wins a given point is independent of which player has the serve, then both players should have a 50% chance of winning the game, regardless of the value of p. A simple solution to this problem is to let the serve alternate between the players, and declare the winner to be the first player to be two points ahead.

One can see that this corresponds exactly to the construction above. If the two first points are distributed one point to the server and one point to the receiver, then the player who won both these points is the winner. If the server wins both, or if the receiver wins both, the score is even, and the procedure is repeated.

More efficient ways of simulating a fair coin using a biased coin are discussed in [7].

It is interesting to note that Theorem 2.1 already gives us a method for constructing an event which has probability $\sqrt{p}$. This was an idea of S. Vempala. The Taylor expansion of $\sqrt{p}$ around $p = 1$ can be written as

$$\sqrt{p} = 1 - \sum_{n=0}^{\infty} \frac{C_n}{2^{2n+1}}(1-p)^{n+1}, \tag{1}$$

where $C_n = \binom{2n}{n}/(n+1)$ is the n:th Catalan number.

It is well known that C_n is equal to the number of parenthetically well formed expressions with n pairs of parentheses, or equivalently, the number of 0-1-strings with n zeros and n ones, such that every initial segment contains at least as many ones as zeros.

Suppose that we flip a fair coin until, for the first time, the total number of tails is greater than the total number of heads. Then for a given n, the probability that this happens after $2n+1$ steps is equal to $C_n/2^{2n+1}$, since there are C_n 0-1-strings of length $2n+1$ with n ones and $n+1$ zeros, such that no proper initial segment contains more zeros than ones, and each of these occurs with probability $1/2^{2n+1}$.

We now start by flipping a (simulated) fair coin until, for the first time, the total number of tails is greater than the number of heads (which of course happens at some point with probability 1). If this happens after $2n+1$ steps, we continue by flipping the coin (in the usual way) $n+1$ times. Then the event E that at least one of these $n+1$ tosses results in "heads" is an observable event, and the probability of E is given by (1), so that $P_p(E) = \sqrt{p}$.

We can make the following refinement of Theorem 2.1.

Theorem 2.2. *For every $c \in [0, 1]$, there is an observable event which has constant probability c.*

In order to prove this, we introduce a few concepts which will also be useful later. We define a sequence $(Y_k)_{k=1}^{\infty} : \Omega \to \Omega$, by letting $Y_k = X_{2n}$, where n is the k:th number with the property that $X_{2n-1} \neq X_{2n}$. We let Y_k be 0 or 1 arbitrarily if there are at most $k-1$ such values of n, but the probability of this is 0 for every p. Since the outcomes heads-tails and tails-heads are equally probable, and the X_i's are independent, it follows that the Y_k's are independent, and that $P_p(Y_k = 0) = P_p(Y_k = 1) = 1/2$.

The function $R : \Omega \to [0, 1]$ is defined by $R(X) = \sum_{k=1}^{\infty} 2^{-k} Y_k$. Let λ denote Lebesgue measure on $\mathbf{R}$.

Lemma 2.3. *For every measurable subset S of $[0, 1]$, and every p,*

$$P_p(R^{-1}(S)) = \lambda(S).$$

In other words, R is uniformly distributed in $[0, 1]$, regardless of p.

For a proof of this, we refer to [3, p. 35]. We remark that the sequence $Y_1, Y_2, Y_3, \ldots$ is the binary representation of the number $R(X)$. In particular, if S is Lebesgue measurable, then $R^{-1}(S)$ is measurable for every p.

Let $\Omega' = \{X \in \Omega : X_{2n-1} \neq X_{2n} \text{ for infinitely many } n\}$. Clearly for every p, $P_p(\Omega') = 1$.

Lemma 2.4. *If $X \in \Omega'$, then R is continuous at X.*

Proof. If $X \in \Omega'$, then for every n there is a number k such that $X_1, \ldots X_k$ determine $Y_1, \ldots, Y_n$. This implies that R maps the open neighborhood $U(X_1, \ldots, X_k)$ of X into the interval $[y, y + 2^{-n}]$, where $y = \sum_{i=1}^{n} 2^{-i} Y_i$. □

We say that a set S of real numbers is observable if $\lambda(\mathrm{Bd} S) = 0$.

Lemma 2.5. *If S is an observable subset of $[0,1]$, then $R^{-1}(S)$ is observable.*

Proof. Since Ω' has measure 1 for every p, it suffices to show that the intersection of $\mathrm{Bd}(R^{-1}(S))$ with Ω' is a null set for every p. By Lemma 2.4, $\mathrm{Bd}(R^{-1}(S)) \cap \Omega' \subseteq R^{-1}(\mathrm{Bd}S)$. Since $\mathrm{Bd}S$ has Lebesgue measure 0, it follows from Lemma 2.3 that $R^{-1}(\mathrm{Bd}S)$ has measure 0 for every p. □

If I is a subinterval of $[0,1]$, then, since the boundary of I has Lebesgue measure 0, $R^{-1}(I)$ is an observable event. Hence the event $R(X) < c$ is an observable event which has probability c, and Theorem 2.2 is proved.

D. E. Knuth and A. C. Yao [4] give an optimal method for simulating a discrete random variable using a fair coin.

We now show that any function $f : (0,1) \to [0,1]$ represents the probability of some event $E \subseteq \Omega$. Let E be the event "There exists a number q such that $\{i : X_i = 1\}$ has density q, that is, such that

$$\frac{|\{i : X_i = 1\} \cap \{1, \dots, n\}|}{n} \to q \qquad \text{as } n \to \infty,$$

and $R(X) < f(q)$". For every p, $P_p(\{i : X_i = 1\}$ has density $p) = 1$. Hence $P_p(E) = P_p(R(X) < f(p)) = f(p)$.

The idea of introducing new random variables distributed on $\{0,1\}$ according to a given probability function f, and then use these as "input" for an observable event with another probability function g can be used to show the following:

Theorem 2.6. *If f and g are functions $(0,1) \to (0,1)$ representing the probability of some observable events, then so is $g \circ f$.*

Proof. Let E_f and E_g be observable events with probabilities f and g respectively. Given an element $X = (X_1, X_2, X_3, \dots)$ of Ω, we can almost certainly (for every p) divide this sequence into finite blocks $b_1 = (X_1, \dots, X_{k_1}), b_2 = (X_{k_1+1}, \dots, X_{k_2}), \dots$ such that for every i, $U(b_i)$ is a subset of either E_f or E_f^c.

We now assume that the blocks b_i have been chosen with minimal length with this property. Define a sequence $(Y_i)_{i=1}^{\infty}$ by letting $Y_i = 0$ or 1 as $U(b_i)$ is a subset of E_f or E_f^c respectively. $(Y_i)_{i=1}^{\infty}$ gives a map $\Omega \to \Omega$ which is well defined except on a null set, on which we can define it arbitrarily. Since the numbers $k_1, k_2, \dots$ are stopping times, the Y_i's are independent, and for every i, $P(Y_i = 1) = f(p)$ (for a discussion of stopping times, we refer to any textbook on probability, such as [2]). Hence the probability that $(Y_i)_{i=1}^{\infty} \in E_g$ is $g(f(p))$. The boundary of this event is the union of the set of strings for which one of the blocks b_i does not terminate, and the inverse image of the boundary of E_g. The first of these is a countable union of null sets, hence a null set. That the second is a null set follows from the fact that the Y_i's are independent and that $P_{f(p)}(\mathrm{Bd}E_g) = 0$. Hence the inverse image of E_g under $(Y_i)_{i=1}^{\infty}$ is an observable event with probability function $g \circ f$.

□

There is another way in which Theorem 2.2 can be generalized:

Theorem 2.7. *If $f_1, f_2, f_3, \ldots$ is a sequence, finite or infinite, of functions, each representing the probability of an observable event, then any convex combination of $f_1, f_2, f_3, \ldots$ is the probability of some observable event.*

Proof. Let $f = \alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3 + \ldots$, where $0 \leq \alpha_i \leq 1$, and $\alpha_1 + \alpha_2 + \alpha_3 + \cdots = 1$. Let E_i be an observable event with probability $f_i(p)$.

Let $I_1, I_2, I_3, \ldots$ be disjoint subintervals of $[0, 1]$ such that $|I_i| = \alpha_i$. It follows from Lemma 2.3 that, with probability 1 for every p, $R(X) \in \mathrm{Int}I_i$ for some i, and $X \in \Omega'$. In this case by Lemma 2.4 there is a k such that $R(U(X_1, \ldots, X_k))$ is contained in I_i.

Assuming that k is minimal with this property, we let E be the event that $(X_{k+1}, X_{k+2}, X_{k+3}, \ldots) \in E_i$. For a certain i, the probability that $R(U(X_1, \ldots, X_k))$ is contained in I_i, and $(X_{k+1}, X_{k+2}, X_{k+3}, \ldots) \in E_i$, is $\alpha_i f_i$. Since these events are mutually inconsistent, the probability of E is given by $\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3 + \ldots$, which proves the theorem. □

Using Theorem 2.7, together with the fact that powers of p and $1 - p$ are probabilities of observable events, we can explicitly construct observable events with probability functions like e^{-p}, $\cos p$ and $\sin p$. For example, the expansion

$$e^{-p} = \frac{e^{1-p}}{e} = \frac{1}{e} + \frac{1-p}{e} + \frac{(1-p)^2}{2e} + \ldots$$

shows that e^{-p}, being a convex combination of powers of $1 - p$, is the probability of an observable event.

The usual Taylor series expansions of $\cos p$ and $\sin p$ can be rearranged as

$$\cos p = 1 - \sum_{k=0}^{\infty} \frac{1}{(4k+2)!} \left(p^{4k+2} \left(1 - \frac{p^2}{(4k+3)(4k+4)} \right) \right),$$

$$\sin p = p\left(1 - \sum_{k=0}^{\infty} \frac{1}{(4k+3)!}\left(p^{4k+2}\left(1 - \frac{p^2}{(4k+4)(4k+5)}\right)\right)\right),$$

showing that $\cos p$ and $\sin p$ are the probabilities of observable events. By Theorem 2.6, it follows that for example the function $\exp\left(-\sqrt{\cos p}\right)$ is the probability of an observable event.

3 Proof of Theorems 1.2 and 1.3

We begin by establishing the necessity of the conditions given in Theorems 1.2 and 1.3.

Theorem 3.1. *If $f : (0,1) \to (0,1)$ represents the probability of an observable event, then there is a positive integer k such that*

$$p^k(1-p)^k < f(p) < 1 - p^k(1-p)^k \tag{2}$$

This will follow from the analogous theorem for open events:

Theorem 3.2. *If $f : (0,1) \to (0,1]$ represents the probability of an open event, then there is a positive integer k such that*

$$p^k(1-p)^k < f(p) \tag{3}$$

Proof. It is enough to show that there is an m such that, for $p < 1/2$, $f(p) > p^m$. Let $f : (0,1) \to (0,1]$ be the probability of an open event E. Since $f(p) > 0$, there is some finite 0-1 string $a_1, \dots, a_m$ such that $U(a_1, \dots, a_m) \subseteq E$. For $p < 1/2$, the probability that a given 0-1-string of length m occurs as an initial segment of $(X_1, X_2, X_3, \dots)$ is at least p^m. Hence $f(p) > p^m$.

The same argument shows that there is an n such that, for $p > 1/2$, $f(p) > (1-p)^n$. We now take $k = \max(m,n)$ in (3), and the theorem is proved. □

Now Theorem 3.1 follows by noting that if $f : (0,1) \to (0,1)$ represents the probability of an observable event, then, since $P_p(\text{Int}E) = f(p)$, and $P_p(\text{Int}(E^c)) = 1 - f(p)$, both f and $1-f$ are probabilities of open events. By Theorem 3.2, we can find k_1 and k_2 such that $f(p) > p^{k_1}(1-p)^{k_1}$ and $1 - f(p) > p^{k_2}(1-p)^{k_2}$. If we take $k = \max(k_1, k_2)$, then (2) holds.

Theorem 3.3. *If $f : (0,1) \to [0,1]$ is the probability of an open event, then f is lower semicontinuous. Hence if f is the probability of an observable event, then f is continuous.*

Proof. Let E be an open event with probability $f(p)$. Let E_n be the event that the outcomes of $X_1, \dots, X_n$ determine that E occurs, that is,

$$E_n = \{X \in \Omega : U(X_1, \dots, X_n) \subseteq E\}.$$

Since E is open, $P_p(E_n) \to f(p)$ pointwise as $n \to \infty$. Since $P_p(E_n)$ is continuous (being a polynomial in p), and $P_p(E_n)$ is increasing with n, it follows that f is lower semicontinuous. The second statement of the theorem now follows as above by noting that if f represents the probability of an observable event, then both f and $1 - f$ are probabilities of open events. Hence both f and $1 - f$ are lower semicontinuous, which implies that f is continuous. □

We now turn to the proof of the sufficiency of the conditions given in Theorem 1.2. We first need to establish a couple of lemmas.

Definition 2. A function $\alpha : \Omega \to \mathbf{N}$ is *observable*, if $\alpha^{-1}(n)$ is observable for every n.

In other words, with probability 1 for every p, we will know the value of α after a finite number of tosses.

Lemma 3.4. *If $(0,1)$ is covered by countably many open subintervals U_i, $i \in \mathbf{N}$, and k is a positive integer, then there is an observable function $\alpha : \Omega \to \mathbf{N}$ such that for every $p \in (0,1)$,*

$$P_p\left(\alpha^{-1}\{n : p \notin U_n\}\right) < p^k(1-p)^k. \tag{4}$$

This means that there is a method by which we can stop after a finite number of tosses, and make a guess of the form "$p \in U_i$", in such a way that, for any given p, the probability that the guess is incorrect is bounded by $p^k(1-p)^k$.

The number of tosses after which we stop and make a guess need not be specified in advance, but may depend on the outcome of the tosses.

To do this, we begin by tossing the coin until we have observed at least $2k+1$ heads and at least $2k+1$ tails. Suppose that this happens after m

tosses. With a very rough estimate, since there are 2^m binary strings of length m, and since each string containing $2k+1$ heads has probability at most p^{2k+1} of occurring, the probability of observing at least $2k+1$ heads in m tosses is at most

$$2^m p^{2k+1} < p^{2k} < p^k(1-p)^k,$$

if $p < 2^{-m}$. Similarly, the probability of observing at least $2k+1$ tails in m tosses is smaller than $p^k(1-p)^k$ if $p > 1-2^{-m}$. We can therefore safely decide to guess on an interval intersecting $[2^{-m}, 1-2^{-m}]$.

By compactness, we can choose a finite number of U_i's, say $U_1, \ldots, U_n$, which cover $[2^{-m}, 1-2^{-m}]$. We can also find an $\epsilon > 0$ such that $p^k(1-p)^k > \epsilon$ throughout $[2^{-m}, 1-2^{-m}]$. For positive integers N, we let

$$p_N = \frac{X_1 + \cdots + X_N}{N}.$$

It is now sufficient to note that, if $p \in U_i$, then as $N \to \infty$, $P_p(p_N \in U_i) \to 1$. This is a consequence of the Law of Large Numbers (see for example [2]).

We now choose a closed subinterval F_i of U_i, for $1 \leq i \leq n$, in such a way that the F_i's cover $[2^{-m}, 1-2^{-m}]$. We choose N large enough that for any $p \in [2^{-m}, 1-2^{-m}]$, the probability (with respect to P_p) that, for some U_i ,$1 \leq i \leq n$, $p_N \in F_i$ and $p \notin U_i$, is smaller than ϵ. We flip the coin N times, and let $\alpha = \alpha(X)$ be such that $p_N \in F_\alpha$, if $p_N \in [2^{-m}, 1-2^{-m}]$. If $p_N \notin [2^{-m}, 1-2^{-m}]$, we define α arbitrarily.

We have now shown that if we condition on the value of m, then the inequality (4) holds for every p. This proves Lemma 3.4.

The next lemma shows that any function satisfying the conditions of Theorem 1.2 can be approximated well by a function representing the probability of an observable event.

Lemma 3.5. *If f is any continuous function $(0,1) \to (0,1)$, and l is any positive integer, then there is a function g representing the probability of an observable event, such that for every p in $(0,1)$,*

$$|f(p) - g(p)| < p^l(1-p)^l.$$

Proof. Since f is continuous, we can choose countably many open subintervals $U_1, U_2, U_3, \ldots$ covering $(0,1)$, and numbers $q_1, q_2, q_3, \ldots$ so that for every $p \in U_i$,

$$|f(p) - q_i| < p^{l+1}(1-p)^{l+1}.$$

By Lemma 3.4, there is an observable function $\alpha : \Omega \to \mathbf{N}$ such that for every p, $P_p(\alpha \ : \ p \notin U_\alpha) < p^{l+1}(1-p)^{l+1}$. Since α is observable, with probability 1 there exists an r such that $X_1, \ldots, X_r$ determine the value of $\alpha = \alpha(X)$. We assume that r is minimal with this property.

Now let E be the event

$$R(X_{r+1}, X_{r+2}, X_{r+3}, \ldots) < q_\alpha,$$

By Lemma 2.5, E is observable. We let $g(p) = P_p(E)$. Then we have

$$g(p) = \sum_{i=1}^{\infty} P_p(\alpha^{-1}(i))q_i.$$

Hence for every p, the error $|f(p) - g(p)|$ is bounded by

$$P_p\,(p \notin U_\alpha) + \sup\{|f(p) - q_i| : p \in U_i\},$$

which is smaller than

$$2p^{l+1}(1-p)^{l+1} < p^l(1-p)^l.$$

□

We notice that it is possible to deduce the Weierstrass approximation theorem from Lemma 3.5 (or from Theorem 1.2). If $[a,b]$ is a compact subinterval of $(0,1)$, then, with the notation of Lemma 3.5, there exists an observable event E with probability given by a function g approximating f to within ϵ throughout $[a,b]$. Let $g_n(p)$ be the probability that after tossing a coin n times, we know that the event E occurs. Then $g_n(p)$ is a monotone sequence of polynomials converging pointwise, and hence uniformly, to $g(p)$ on $[a,b]$. In particular, there is a polynomial approximating f to within 2ϵ on $[a,b]$.

A proof of the Weierstrass approximation theorem using probabilistic arguments was given by S. N. Bernstein in [1].

We now conclude the proof of Theorem 1.2. Let $f : (0,1) \to (0,1)$ be any continuous function satisfying (2). If we apply Lemma 3.5, with $l = k+1$, we find that there exists a function f_1, representing the probability of an observable event, such that

$$|f_1(p) - f(p)| < p^{k+1}(1-p)^{k+1},$$

for every p. We then have

$$f_1(p) > f(p) - p^{k+1}(1-p)^{k+1}.$$

Since $f(p) < 1 - p^k(1-p)^k$, we see that

$$f_1(p) > 2f(p) - p^{k+1}(1-p)^{k+1} - 1 + p^k(1-p)^k > 2f(p) - 1 + p^{k+1}(1-p)^{k+1}.$$

Similarly,

$$f_1(p) < f(p) + p^{k+1}(1-p)^{k+1}.$$

Since $f(p) > p^k(1-p)^k$, we have

$$f_1(p) < 2f(p) + p^{k+1}(1-p)^{k+1} - p^k(1-p)^k < 2f(p) - p^{k+1}(1-p)^{k+1}.$$

Combining the inequalities for f_1, we get

$$2f(p) - 1 + p^{k+1}(1-p)^{k+1} < f_1(p) < 2f(p) - p^{k+1}(1-p)^{k+1},$$

which can be rearranged as

$$p^{k+1}(1-p)^{k+1} < 2f(p) - f_1(p) < 1 - p^{k+1}(1-p)^{k+1}.$$

If we repeat this argument, replacing $f(p)$ by $2f(p) - f_1(p)$, and k by $k+1$, we find that there exists a function f_2 which is the probability of an observable event, such that

$$p^{k+2}(1-p)^{k+2} < 4f(p) - 2f_1(p) - f_2(p) < 1 - p^{k+2}(1-p)^{k+2}.$$

Continuing in this way, we can find functions $f_1, f_2, f_3, \ldots$, each representing the probability of an observable event, such that for every n,

$$p^{k+n}(1-p)^{k+n} < 2^n f(p) - 2^{n-1} f_1(p) - \cdots - f_n(p) < 1 - p^{k+n}(1-p)^{k+n}.$$

If we replace the lower and upper bounds by 0 and 1, and divide by 2^n, we get

$$0 < f(p) - \frac{1}{2} f_1(p) - \frac{1}{4} f_2(p) - \cdots - \frac{1}{2^n} f_n(p) < \frac{1}{2^n}.$$

Letting $n \to \infty$, we obtain

$$f(p) = \sum_{n=1}^{\infty} \frac{1}{2^n} f_n.$$

By Theorem 2.7, this shows that f is the probability of an observable event. This completes the proof of Theorem 1.2.

Once Theorem 1.2 is established, the proof of Theorem 1.3 is a relatively simple matter. We have already shown that the conditions of Theorem 1.3 are necessary for the existence of an open event with probability function f.

Lemma 3.6. *If $g_1, g_2, g_3, \dots$ are functions $(0,1) \to (0,1)$ representing the probabilities of observable events $E_1, E_2, E_3, \dots$, then there is an open event E such that*

$$P_p(E) = 1 - \prod_{n=1}^{\infty} g_n(p). \tag{5}$$

Proof. With probability 1, for every $p \in (0,1)$, $X = (X_1, X_2, X_3, \dots)$ can be divided into blocks $b_1 = (X_1, \dots, X_{k_1}), b_2 = (X_{k_1+1}, \dots, X_{k_2}), \dots$, such that for every i, $U(b_i) \subseteq E_i$, or $U(b_i) \subseteq E_i^c$.

We assume that every k_i be minimal with this property, given that all k_j for $j < i$ have already been chosen. This means that $k_1, k_2, k_3, \dots$ are stopping times, so that the events $U(b_i) \subseteq E_i^c$ are independent. It may happen that only finitely many k_i's can be chosen in this way, so that only finitely many blocks are well defined. The probability that the blocks $b_1, \dots, b_n$ are well-defined, and that $U(b_i) \subseteq E_i^c$ for some i, $1 \le i \le n$, is given by $1 - g_1 g_2 g_3 \dots$.

Let E be the event that for some i, the blocks $b_1, \dots, b_i$ are well defined, and $U(b_i) \subseteq E_i^c$. Then E is an observable event, with probability given by (5). □

Let f be any lower semicontinuous function $(0,1) \to (0,1]$ satisfying $f(p) > p^k(1-p)^k$ for some k. We extend the domain and range of f by letting $f(0) = f(1) = 0$, and define, for $n \ge 0$, functions $f_n : (0,1) \to [0,1]$ by

$$f_0(p) = 0,$$
$$f_n(p) = \inf_{x \in [0,1]} \{f(x) + n\,|x - p|\} - p^{k+n}(1-p)^{k+n}.$$

The function f_n is continuous for every n. Since f is lower semicontinuous, $f_n(p) \to f(p)$ as $n \to \infty$.

To be able to apply Lemma 3.6, we let

$$g_n = \frac{1 - f_n}{1 - f_{n-1}}, \quad n \ge 1,$$

so that $1 - f_n = g_1 g_2 g_3 \dots g_n$, and

$$f = 1 - \prod_{n=1}^{\infty} g_n.$$

Since for every p, $f_n(p) > f_{n-1}(p)$, g_n is a continuous function $(0,1) \to (0,1)$.

We have to show that, for some l,

$$p^l(1-p)^l < g_n(p) < 1 - p^l(1-p)^l. \tag{6}$$

¿From the definition we see that, for every n, $f_n(p) \to 0$ as $p \to 0$ or $p \to 1$ (this is why we defined $f(0)$ and $f(1)$ to be 0). Hence $g_n(p) \to 1$ as $p \to 0$ or $p \to 1$, so that the left inequality of (6) is satisfied. For the right inequality, we have

$$\begin{aligned} g_n(p) &= 1 - f_n(p) + f_{n-1}(p)g_n(p) < 1 - f_n(p) + f_{n-1}(p) \leq \\ &\leq 1 + p^{k+n}(1-p)^{k+n} - p^{k+n-1}(1-p)^{k+n-1} < 1 - p^{k+n}(1-p)^{k+n}. \end{aligned}$$

Hence by Theorem 1.2, every g_n represents the probability of an observable event, so that by Lemma 3.6, f is the probability of an open event. This proves Theorem 1.3.

4 Finite events

Even if an event E is observable, it may happen that for some 0-1-string X, we cannot tell from any finite number of outcomes $X_1, \dots, X_n$ whether E occurs or not. If we demand that this should never happen, that is, that $\mathrm{Bd}E = \emptyset$, then by compactness, there must be a number N such that $X_1, \dots, X_N$ always determine whether or not E occurs. Such an event will be called *finite*, and the minimal value of N such that E is determined by $X_1, \dots, X_N$ will be called the *degree* of E.

Lemma 4.1. *A function $f(x)$ represents the probability of a finite event if and only if for some N we can express it as:*

$$f(x) = \sum_{i=0}^{N} a_i x^i (1-x)^{N-i}, \tag{7}$$

with $0 \leq a_i \leq \binom{N}{i}$ for every i.

Proof. Suppose E is a finite event of degree N, and let a_i be the number of strings of length N containing exactly i heads, for which E occurs. Then

$$P_p(E) = \sum_{i=0}^{N} a_i p^i (1-p)^{N-i}. \tag{8}$$

In (8), for each coefficient a_i we have $0 \leq a_i \leq \binom{N}{i}$. Conversely, if a polynomial $f(p)$ can be expressed as (8), then it represents a finite event of degree at most N, since we can choose, for every i, a_i sequences of i heads and $N-i$ tails, and consider the event that one of these occurs. □

In particular, the probability of E is a polynomial with integer coefficients.

Example 4.2. The polynomial $2x - 2x^2$ represents a finite event, since it can be written $2x(1-x)$. It represents the probability that, if a coin is tossed twice, the two outcomes are different.

The polynomial $3x(1-x)$ does not represent an event of degree 2, since $3 > \binom{2}{1}$. However, it can be written $3x(1-x)^2 + 3x^2(1-x)$, from which it follows that it represents the probability that, if a coin is flipped three times, not all outcomes are the same.

We say that a polynomial representing the probability of a finite event is a *coin flipping polynomial.* We define the *coin flipping degree* of a coin flipping polynomial $f(x)$ to be the minimal degree of a finite event having probability represented by f. Clearly the coin flipping degree is larger than or equal to the degree of the polynomial, and we see from the example above that equality need not hold.

The following lemma shows that it is easy to determine, for a given number N, whether or not a specific polynomial $f(x)$ represents an event of degree at most N.

Lemma 4.3. *Every polynomial $f(x)$ of degree at most N with integer coefficients can be expressed as*

$$f(x) = \sum_{i=0}^{N} a_i x^i (1-x)^{N-i}, \tag{9}$$

with integers a_i, which are uniquely determined by f.

Proof. The term corresponding to $i = 0$ is the only one which has a nonzero constant. Hence a_i must be taken to be the constant term of f. By induction, the degree $N - 1$ polynomial

$$\frac{f(x) - a_i(1-x)^N}{x}$$

can be expressed in a unique way as

$$\sum_{i=0}^{N-1} b_i x^i (1-x)^{N-1-i},$$

with integers b_i. Now we can, and must, take $a_{i+1} = b_i$ for every i. □

In order to determine whether or not f is the probability of an event of degree at most N, we need only compute the coefficients a_i, and compare them with the binomial coefficients $\binom{N}{i}$.

Example 4.4. A computation shows that the degree 3 polynomial $1 - 8x + 20x^2 - 13x^3$ is a coin flipping polynomial of coin flipping degree 46.

Apparently, the coin flipping degree of a polynomial can be much larger than the degree. This suggests that the question whether a given polynomial is a coin flipping polynomial or not might be nontrivial. Although there seems to be no simple criterion in terms of the *coefficients* of the polynomial, it turns out that the answer is given by the direct analogue of Theorem 1.2.

By Theorem 1.2, if a coin flipping polynomial is not identically 0 or identically 1, then it must map the open unit interval $(0, 1)$ to itself. We show that this condition is also sufficient for a polynomial to be "coin flipping".

Theorem 4.5. *A function $f(x)$ is the probability of a finite event if and only if $f(x)$ is a polynomial with integer coefficients which is identically 0, identically 1, or maps $(0, 1)$ to itself.*

Proof. Let $f(x)$ be a polynomial of degree d with integer coefficients. Then f can be expressed as

$$f(x) = \sum_{i=0}^{d} a_{d,i} x^i (1-x)^{d-i}, \tag{10}$$

with integer coefficients $a_{d,i}$. Let N be an integer, and suppose $N \geq d$. Multiplying (10) with the factor

$$1 = \sum_{j=0}^{N-d} \binom{N-d}{j} x^j (1-x)^{N-d-j}, \tag{11}$$

and collecting terms, we obtain

$$f(x) = \sum_{j=0}^{N} a_{N,j} x^j (1-x)^{N-j}, \tag{12}$$

where

$$a_{N,j} = \sum_{i=0}^{d} a_{d,i} \binom{N-d}{j-i}. \tag{13}$$

We wish to compare $a_{N,j}$ to $\binom{N}{j}$ for large N. We have

$$\begin{aligned} \frac{a_{N,j}}{\binom{N}{j}} &= \sum_{i=0}^{d} a_{d,i} \frac{\binom{N-d}{j-i}}{\binom{N}{j}} = \sum_{i=0}^{d} a_{d,i} \frac{j!(N-j)!(N-d)!}{(j-i)!(N-d-j+i)!N!} = \\ &= \sum_{i=0}^{d} a_{d,i} \frac{j(j-1)\dots(j-i+1)(N-j)\dots(N-j-d+i+1)}{N(N-1)\dots(N-d+1)}. \end{aligned} \tag{14}$$

We now assume that for every N there is some j such that $a_{N,j} < 0$ or $a_{N,j} > \binom{N}{j}$. Under this assumption, we wish to find a number x, $0 < x < 1$, such that $f(x) \leq 0$ or $f(x) \geq 1$. Without loss of generality, we can assume that for infinitely many values of N, there is a j such that $a_{N,j} < 0$. Otherwise there are infinitely many N such that $a_{N,j} > \binom{N}{j}$, but then we can replace $f(x)$ by

$$1 - f(x) = \sum_{j=0}^{N} \left(\binom{N}{j} - a_{N,j} \right) x^j (1-x)^{N-j}.$$

Since the interval $[0,1]$ is compact, there has to be a sequence of numbers $N_1, N_2, N_3, \dots$, and a corresponding sequence $j_1, j_2, j_3, \dots$ such that $a_{N_n,j_n} < 0$ for every n, and such that j_n/N_n converges to a number $\xi \in [0,1]$, as $n \to \infty$. We now consider two cases depending on whether ξ lies in the interior of $[0,1]$, or is one of the endpoints 0 or 1. Suppose first that $0 < \xi < 1$.

We can write (14) as

$$\sum_{i=0}^{d} a_{d,i} \cdot \frac{j}{N} \cdot \frac{j-1}{N-1} \cdots \frac{j-i+1}{N-i+1} \cdot \frac{N-j}{N-i} \cdots \frac{N-j-d+i+1}{N-d+1}. \tag{15}$$

If we let N and j tend to infinity through a sequence of values N_n and j_n in such a way that $j_n/N_n \to \xi$ as $n \to \infty$, then (15) becomes

$$\sum_{i=0}^{d} a_{d,i} \left(\frac{j_n}{N_n}\right)^{i} \left(\frac{N_n - j_n}{N_n}\right)^{d-i} + O(1/N), \tag{16}$$

which converges to

$$\sum_{i=0}^{d} a_{d,i} \xi^{i} (1-\xi)^{d-i} = f(\xi), \quad \text{as } n \to \infty. \tag{17}$$

Since we assumed that N_n and j_n run through values for which a_{N_n,j_n} is negative, it follows that $f(\xi) \leq 0$.

Suppose on the other hand that ξ equals 0 or 1. Without loss of generality, we may assume $\xi = 0$, since otherwise we can replace $f(x)$ by $f(1-x) = \sum_{j=0} N a_{N,N-j} x^j (1-x)^{N-j}$. In (14), the values of i for which $i \leq j$ and $a_{d,i} \neq 0$ will give nonzero terms. Of the nonzero terms, the one with the largest number of factors of order of magnitude N will dominate. This will be the nonzero term corresponding to the smallest value of i. If N is large, then the sign of the first nonzero coefficient $a_{d,i}$ will determine the sign of (14). Note that for this value of i, we must have $i \leq j$, since otherwise all the terms would be zero, and we have assumed that $a_{N,j} < 0$. This shows that the first nonzero coefficient $a_{d,i}$ is negative. For this value of i, we will have $f(x) = a_{d,i} x^i + O(x^{i+1})$, as $x \to 0$. This shows that for small positive values of x, $f(x) < 0$. □

Example 4.4 raises the question how large the coin flipping degree of a polynomial can be, in terms of the degree. As the following theorem shows, there are only finitely many coin flipping polynomials of a certain degree.

Theorem 4.6. *There are only finitely many coin flipping polynomials of a given degree.*

Proof. For a fixed d, we consider the set $\mathcal{C}$ of vectors $(a_0, \ldots, a_d)$ such that $\left|a_0 + a_1 x + \cdots + a_d x^d\right| \leq 1$ for every $x \in [0, 1]$. The function

$$\max_{x\in[0,1]} \left|a_0 + a_1 x + \cdots + a_d x^d\right| \tag{18}$$

defines a norm on $\mathbf{R}^{d+1}$. Hence $\mathcal{C}$, being the unit ball, is compact. It contains only finitely many points with integer coordinates, and the coefficients of every coin flipping polynomial of degree d must be among these points. □

This shows that it must be possible to give a bound on the coin flipping degree of a polynomial in terms of its degree. Let $N(d)$ be the maximal coin flipping degree among all coin flipping polynomials of degree at most d. Then it is easily verified that $N(1) = 1$ and $N(2) = 3$, and it seems that $N(3) = 46$. The example $f(x) = 13x(1-x)^3 - 21x^2(1-x)^2 + 14x^3(1-x)$ shows that $N(4) \geq 735$.

Acknowledgments. I thank László Lovász, Santosh Vempala and Anders Björner for stimulating discussions. I also wish to thank P. Diaconis for providing the references [7] and [4], and P. Winkler for some helpful remarks.

References

[1] Bernstein, S. N., *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités*, Comm. Soc. Math. Kharkov **13** (1912) 1–2.

[2] Breiman, L., *Probability*, Addison-Wesley Publishing Company, Inc., 1968.

[3] Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. 2, 2nd Ed., John Wiley & Sons, Inc., 1971.

[4] Knuth, D. E., Yao, A. C., *The complexity of nonuniform random number generation*, in *Algorithms and Complexity, New Directions and Recent Results*, ed. J. F. Traub, Academic Press 1976.

[5] Lovász, L., Personal communication, February 1997.

[6] von Neumann, J., Various Techniques Used in Connection With Random Digits. Reprinted in *John von Neumann Collected Works*, vol. V (Pergamon Press, 1963), p. 768–770.

[7] Stout, Q. F., Warren, B., *Tree algorithms for unbiased coin tossing with a biased coin*, Ann. Prob. **12** (1984), 212–222.