Open Questions on the Bernoulli Factory Problem

This version of the document is dated 2023-05-14.

Peter Occil

1 Background

Suppose there is a coin that shows heads with an unknown probability, λ. The goal is to use that coin (and possibly also a fair coin) to build a "new" coin that shows heads with a probability that depends on λ, call it $f(\lambda)$. This is the *Bernoulli factory problem*, and it can be solved only for certain functions f. (For example, flipping the coin twice and taking heads only if exactly one coin shows heads, the probability $2\lambda(1-\lambda)$ can be simulated.)

Specifically, the only functions that can be simulated this way **are continuous and polynomially bounded on their domain, and map $[0, 1]$ or a subset thereof to $[0, 1]$**, as well as $f=0$ and $f=1$. These functions are called *factory functions* in this page. (A function $f(x)$ is *polynomially bounded* if both f and $1-f$ are greater than or equal to min(x^n, $(1-x)^n$) for some integer n (Keane and O'Brien 1994). This implies that f admits no roots on (0, 1) and can't take on the value 0 or 1 except possibly at 0, 1, or both.)

This page contains several questions about the **Bernoulli factory** problem. Answers to them will greatly improve my pages on this site about Bernoulli factories. If you can answer any of them, post an issue in the **GitHub issues page**.

2 Contents

3 Key Problems

The following summarizes most of the problems raised by these open questions.

1. **Suppose $f:[0,1]\to [0,1]$ is continuous and belongs to a large class of functions (e.g., the k-th derivative, $k\ge 0$, is continuous, Lipschitz, concave, strictly increasing, bounded variation, or Zygmund, or f is real analytic).**
    - *Exact Bernoulli factory*: **Assuming $0\lt f(\lambda)\lt 1$, compute the Bernstein coefficients of a sequence of polynomials (g_n) of degree 2, 4, 8, ..., 2^i, ... that converge to f from below and satisfy: $(g_{2n}-g_{n})$ is a polynomial with non-negative Bernstein coefficients**

    once it's rewritten to a polynomial in Bernstein form of degree exactly $2n$.

    - *Approximate Bernoulli factory*: **Given $\epsilon > 0$, compute the Bernstein coefficients of a polynomial or rational function (of some degree n) that is within ϵ of f.**
    - *Series expansion of simple functions*: **Find a random variable X and a non-trivial series $f(\lambda)=\sum_{a\ge 0}\gamma_a(\lambda)$ such that $\gamma_a(\lambda)/\mathbb{P}(X=a)$ is a polynomial or rational function with Bernstein coefficients lying in [0, 1].**

   The convergence rate must be $O(1/n^{r/2})$ if the class has only functions with Lipschitz-continuous $(r-1)$-th derivative. The method may not introduce transcendental or trigonometric functions (as with Chebyshev interpolants).

2. **Characterize the following three classes of factory functions $f(\lambda)$:**

    - **Can be simulated using nothing but the biased coin, when the biased coin can show heads every time, tails every time, or both.**
    - **Have a Bernoulli factory that can come arbitrarily close to the entropy limit if it produces multiple f-coin flips at a time, rather than just one.**
    - **Are algebraic and can be simulated by a finite-state machine with an unbounded stack.**

4 Polynomials that approach a factory function "fast"

https://math.stackexchange.com/questions/3904732/what-are-ways-to-compute-polynomials-that-converge-from-above-and-below-to-a-con

https://mathoverflow.net/questions/442057/explicit-and-fast-error-bounds-for-approximating-continuous-functions

https://mathoverflow.net/questions/427595/a-conjecture-on-consistent-monotone-sequences-of-polynomials-in-bernstein-form

https://mathoverflow.net/questions/429037/bounds-on-the-expectation-of-a-function-of-a-hypergeometric-random-variable

This question involves solving the Bernoulli factory problem with polynomials.

In this question, a polynomial $P(x)$ is written in *Bernstein form of degree n* if it is written as— $$P(x)=\sum_{k=0}^n a_k {n \choose k} x^k (1-x)^{n-k},$$ where $a_0, ..., a_n$ are the polynomial's *Bernstein coefficients*.

The degree-n *Bernstein polynomial* of an arbitrary function $f(x)$ has Bernstein coefficients $a_k = f(k/n)$. In general, this Bernstein polynomial differs from f even if f is a polynomial.

4.1 Main Question

Suppose $f:[0,1]\to [0,1]$ is continuous and belongs to a large class of functions (for example, the k-th derivative, $k\ge 0$, is continuous, Lipschitz continuous, concave, strictly increasing, bounded variation, or in the Zygmund class, or f is real analytic or in Gevrey's hierarchy) (**see note 4 in "End Notes"**).

1. (*Exact Bernoulli factory*): Compute the Bernstein coefficients of a sequence of polynomials (g_n) of degree 2, 4, 8, ..., 2^i, ... that converge to f from below and satisfy: $(g_{2n}-g_{n})$ is a polynomial with non-negative Bernstein coefficients once it's rewritten to a polynomial in Bernstein form of degree exactly $2n$. (**See note 5 in "End Notes".**) Assume $0\lt f(\lambda)\lt 1$ or f is polynomially bounded.
2. (*Approximate Bernoulli factory*): Given $\epsilon > 0$, compute the Bernstein coefficients of a polynomial or rational function (of some degree n) that is within ϵ of f.

The convergence rate must be $O(1/n^{r/2})$ if the class has only functions with Lipschitz-continuous $(r-1)$-th derivative. The method may not introduce transcendental or trigonometric functions (as with Chebyshev interpolants).

4.2 Solving the Bernoulli factory problem with polynomials

An **algorithm** simulates a factory function $f(\lambda)$ via two sequences of polynomials that converge from above and below to that function. Roughly speaking, the algorithm works as follows:

1. Generate U, a uniform random variate in $[0, 1]$.
2. Flip the input coin (with a probability of heads of λ), then build an upper and lower bound for $f(\lambda)$, based on the outcomes of the flips so far. In this case, these bounds come from two degree-n polynomials that approach f as n gets large, where n is the number of coin flips so far in the algorithm.
3. If U is less than or equal to the lower bound, return 1. If U is greater than the upper bound, return 0. Otherwise, go to step 2.

The result of the algorithm is 1 with probability *exactly* equal to $f(\lambda)$, or 0 otherwise.

However, the algorithm requires the polynomial sequences to meet certain requirements; among them, the sequences must be of Bernstein-form polynomials that converge from above and below to a factory function. Specifically:

For $f(\lambda)$ there must be a sequence of polynomials (g_n) *in Bernstein form of degree 1, 2, 3, ... that converge to f from below and satisfy:* $(g_{n+1}-g_{n})$ *is a polynomial with non-negative Bernstein coefficients once it's rewritten to a polynomial in Bernstein form of degree exactly $n+1$ (**see note 5 in "End Notes"**; Nacu and Peres 2005; Holtz et al. 2011). For $f(\lambda)=1-f(\lambda)$ there must likewise be a sequence of this kind.*

4.3 A Matter of Efficiency

However, ordinary Bernstein polynomials converge to a function at the rate $\Omega(1/n)$ in general, a result known since Voronovskaya (1932) and a rate that will lead to an **infinite expected number of coin flips in general**. (See also my **supplemental notes**.)

But Lorentz (1966) showed that if the function is positive and has a continuous k-th derivative, there are polynomials with nonnegative Bernstein coefficients that converge at the rate $O(1/n^{k/2})$ (and thus can enable a **finite expected number of coin flips** if the function is "smooth" enough).

Thus, people have developed alternatives, including linear combinations and iterated Boolean sums of Bernstein polynomials, to improve the convergence rate. These include Micchelli (1973), Guan (2009), Güntürk and Li (2021a, 2021b), the "Lorentz operator" in Holtz et al. (2011) (see also "**New coins from old, smoothly**"), Draganov (2014), and Tachev (2022).

These alternative polynomials usually include results where the error bound is the desired $O(1/n^{k/2})$, but nearly all those results (e.g., Theorem 4.4 in Micchelli; Theorem 5 in Güntürk and Li) have hidden constants with no upper bounds given, making them unimplementable (that is, it can't be known beforehand whether a given polynomial will come close to the target function within a user-specified error tolerance). (**See note 4 in "End Notes".**)

4.4 A Conjecture on Polynomial Approximation

The following is a **conjecture** that could help reduce this problem to the problem of finding explicit error bounds when approximating a function by polynomials.

Let $f(\lambda):[0,1]\to(0,1)$ have $r\ge 1$ continuous derivatives, let M be the maximum of the absolute value of f and its derivatives up to the r-th derivative, and denote the Bernstein polynomial of degree n of a function g as $B_n(g)$. Let $W_{2^0}(\lambda), W_{2^1}(\lambda), ..., W_{2^i}(\lambda),...$ be a sequence of functions on [0, 1] that converge uniformly to f.

For each integer $n\ge 1$ that's a power of 2, suppose that there is $D>0$ such that— $$|f(\lambda)-B_n(W_n(\lambda))| \le DM/n^{r/2},$$ whenever $0\le \lambda\le 1$. Then there is $C_0\ge D$ such that for every $C\ge C_0$, the polynomials (g_n) in Bernstein form of degree 2, 4, 8, ..., 2^i, ..., defined as $g_n=B_n(W_n(\lambda) - CM/n^{r/2})$, converge from below to f and satisfy: $(g_{2n}-g_{n})$ is a polynomial with non-negative Bernstein coefficients once it's rewritten to a polynomial in Bernstein form of degree exactly $2n$. (**See note 5 in "End Notes".**)

Equivalently (see also Nacu and Peres 2005), there is $C_1>0$ such that the inequality (PB) (see below) holds true for each integer $n\ge 1$ that's a power of 2 (see "Strategies", below).

My goal is to see not just whether this conjecture is true, but also which value of C_0 (or C_1) suffices for the conjecture, especially for any combination of the special cases mentioned at the end of "**Main Question**", above.

4.5 Strategies

The following are some strategies for answering these questions:

- For iterated Boolean sums (linear combinations of iterates) of Bernstein polynomials ($U_{n,k}$ in **Micchelli 1973**; see also **Güntürk and Li**), find an explicit bound, with no hidden constants, on the approximation error for functions with continuous r-th derivative, or verify my **proof of those error bounds**.
- For linear combinations of Bernstein polynomials (Butzer 1953, **Tachev 2022**), verify my proof of those error bounds in **my Proposition B10**.
- For the "**Lorentz operator**", find an explicit bound, with no hidden constants, on the approximation error for the operator $Q_{n,r}$ and for the polynomials (f_n) and (g_n) formed with it, and find the hidden constants θ_α, s, and D as well as those in Lemmas 15, 17 to 22, and 24 in the paper. Or verify my proof of the order-2 operator's error bounds in **my Proposition B10A**.
- Let $f:[-1,1]\to [0,1]$ be continuous. Find explicit bounds, with no hidden constants, on the error in approximating f with the following polynomials: The polynomials are similar to Chebyshev interpolants, but evaluate f at *rational* values of λ that converge to Chebyshev points (that is, converging to $\cos(j\pi/n)$ with increasing n). The error bounds must be close to those of Chebyshev interpolants (see, e.g., chapters 7, 8, and 12 of Trefethen, *Approximation Theory and Approximation Practice*, 2013).
- Find other polynomial operators meeting the requirements of the main question (see "Main Question", above) and having explicit error bounds, with no hidden constants, especially operators that

preserve polynomials of a higher degree than linear functions.
- Find a sequence of functions $(W_n(f))$ and an explicit and tight upper bound on $C_1>0$ such that, for each integer $n\ge 1$ that's a power of 2— $$\left|\left(\sum_{i=0}^k W_n\left(\frac{i}{n}\right)\sigma_{n,k,i}\right)-W_{2n}\left(\frac{k}{2n}\right)\right|=|\mathbb{E}[W_n(X_k/n)] - W_{2n}(\mathbb{E}[X_k/n])|\le \frac{C_1 M}{n^{r/2}},\tag{PB}$$ whenever $0\le k\le 2n$, where $M = \max(L, \max|f^{(0)}|, ...,\max|f^{(r-1)}|)$, L is $\max|f^{(r)}|$ or the Lipschitz constant of $f^{(r-1)}$, X_k is a hypergeometric($2n$, k, n) random variable, and $\sigma_{n,k,i} = {n\choose i}{n\choose {k-i}}/{2n \choose k}=\mathbb{P}(X_k=i)$ is the probability that X_k equals i. (**See notes 5 and 6 in "**End Notes**" as well as "**Proofs for Polynomial-Building Schemes**.**)

5 Tossing Heads According to a Concave Function

https://mathoverflow.net/questions/409174/concave-functions-series-representation-and-converging-polynomials

Suppose $f:[0,1]\to[0,1]$ is continuous, polynomially bounded, and belongs to a large class of functions (for example, the k-th derivative, $k\ge 0$, is continuous, Lipschitz continuous, concave, strictly increasing, bounded variation, or in the Zygmund class, or f is real analytic or in Gevrey's hierarchy) (**see note 5 in "End Notes".**).

Then find a non-negative random variable X and a non-trivial series $f(\lambda)=\sum_{a\ge 0}\gamma_a(\lambda)$ such that $\gamma_a(\lambda)/\mathbb{P}(X=a)$ (letting 0/0 equal 0) has a simple **Bernoulli factory algorithm** (and is preferably a polynomial or rational function with rational Bernstein coefficients lying in $[0, 1]$).

- An example of X is $\mathbb{P}(X=a) = p (1-p)^a$ where $0 < p < 1$ is a known rational. That is, the probability of getting a is $p (1-p)^a$.
- The convergence rate must be $O(1/n^{r/2})$ if the class has only functions with Lipschitz-continuous $(r-1)$-th derivative. The

method may not introduce transcendental or trigonometric functions (as with Chebyshev interpolants).

See also Note 1 in "End Notes".

5.1 Special Cases

One special case of the question above is if—

- $f:[0,1]\to [0,1)$ is concave, and
- $\mathbb{P}(X=a)=p (1-p)^a$, where $0 < p < 1$ is rational, and
- $\gamma_a(\lambda) = B_{n_{a}}(f)(\lambda) - B_{n_{a-1}}(f)(\lambda)$ ($\gamma_a(0)=B_{n_{a}}(f)(\lambda)$), where $B_n(f)$ is the degree-$n$ Bernstein polynomial of $f$, and
- $(n_a)$ is an increasing sequence of positive integers, with $n_{-1} := 0$, and

However, using this technique for a given concave $f$ requires finding the appropriate sequence for $n_a$ (such as $2^{a+s}$ for some $s\ge 0$) and the appropriate value of $p$ so that the series expansion can be formed. Here is an example for $\min(\lambda, 1-\lambda)$ which *appears* to be correct, but finding it was far from rigorous: $n_a = 2^{a+1}$, $p = 0.27$.

Once the appropriate series and $X$ are found, an algorithm to toss heads with probability equal to $f$ would be:

1. Flip a coin that shows heads with probability $p$ until that coin shows heads. Set $a$ to the number of tails.
2. Write $\frac{\gamma_a(\lambda)}{\mathbb{P}(X=a)}$ (letting 0/0=0) as a polynomial in Bernstein form of degree $n_{a}$ (or a higher degree such that the Bernstein coefficients are all in [0, 1]). Flip the biased coin (with probability of heads $\lambda$) $n$ times, where $n$ is the polynomial's degree, and let $j$ be the number of heads.
3. Return 1 with probability equal to the polynomial's $j$th Bernstein coefficient ($j$ starts at 0), or 0 otherwise (see also Goyal and Sigman 2012 for an algorithm to simulate polynomials).

> **Note:** There is no general solution for all concave $f:[0,1]\to [0,1]$, not all of which are polynomially bounded (note the codomain is $[0,1]$, not $[0,1)$). Indeed, there are several counterexamples: $g(\lambda)=\lim_{t\to\lambda} (1-\exp(-2/t))$

> (which is smooth), or $h(\lambda)$ formed by taking $g(\lambda)$ at 0 and at all points of the form $1/n$, where $n\ge 1$ is an integer, and connecting them with linear functions (so that h is not even differentiable).
>
> **Note:** Another problem is to write $\gamma_a(\lambda)/\mathbb{P}(X=a)$ as a polynomial in Bernstein form with only 0 and 1 as coefficients. However, this can be reduced to rewriting the expression to polynomials with dyadic rational coefficients, or even with coefficients whose binary expansion is easy to calculate. In fact, the proof in Keane and O'Brien (1994) rewrites f as: $f(\lambda)=\sum_{a\ge 1}\mathbb{P}(X=a) Q_a(\lambda)$, where X is an integer-valued random variable 1 or greater and where each $Q_a(\lambda)$ is a polynomial in Bernstein form of degree k_a with only 0 and 1 as coefficients.

6 New coins from old, smoothly

https://mathoverflow.net/questions/407179/using-the-holtz-method-to-build-polynomials-that-converge-to-a-continuous-functi

Let $B_n(f)$ be the degree-n Bernstein polynomial of f.

Holtz et al. 2011, in the paper "New coins from old, smoothly", studied a family of polynomials $(Q_{n,r} f)$ (which they call the *Lorentz operators*) that approximate a continuous function f.

They used the Lorentz operators to build a family of polynomials (g_n) that converge from below to f and satisfy the following: $(g_{2n}−g_{n})$ is a polynomial with non-negative Bernstein coefficients, once it's rewritten to a polynomial in Bernstein form of degree exactly $2n$.

They proved, among other results, the following. A function $f(\lambda):[0,1]\to(0,1)$ admits a family (g_n) described above that converges at the rate—

- $O((\Delta_n(\lambda))^\beta)$ if and only if f is $\lfloor\beta\rfloor$ times differentiable and has a ($\beta-\lfloor\beta\rfloor$)-Hölder continuous $\lfloor\beta\rfloor$-th derivative, where $\beta>0$ is a non-integer and $\Delta_n(\lambda) = \max((\lambda(1-\lambda)/n)^{1/2}, 1/n)$. (Roughly speaking, the rate is $O((1/n)^{\beta})$ when $\lambda$ is close to 0 or 1, and $O((1/n)^{\beta/2})$ elsewhere.)

- $O((\Delta_n(\lambda))^{r+1})$ only if the $r$th derivative of $f$ is in the Zygmund class, where $r\ge 0$ is an integer.

The scheme is as follows:

Let $f:[0,1]\to (0,1)$ have a $\beta$-Hölder continuous $r$-th derivative, where $\beta$ is in (0, 1). Let $\alpha = r+\beta$; $b = 2^s$; $s\gt 0$ be an integer. Let $Q_{n, r}f$ be a degree $n+r$ approximating polynomial called a *Lorentz operator* as described in Holtz et al. 2011. Let $n_0$ be the smallest $n$, divisible by $b$, such that $Q_{n_0, r}f$ has coefficients within [0, 1]. Define the following for every integer $n \ge n_0$ divisible by $b$:

- $f_{n_0} = Q_{n_0, r}f$.
- $f_{n} = f_{n/b} + Q_{n, r}(f-f_{n/b})$ for each integer $n > n_0$.
- $\phi(n, \alpha, \lambda) = \frac{\theta_{\alpha}}{n^{\alpha}}+(\frac{\lambda(1-\lambda)}{n})^{\alpha/2}$.
- $BP(\lambda)$ is a polynomial defined as follows: Find the degree-$n$ Bernstein polynomial of $\phi(n, r+\beta, \lambda)$, then rewrite it as a degree-$n+r$ polynomial in Bernstein form.
- $g(n, r,\lambda) = f_{n}(\lambda) - D \cdot BP(\lambda).$

Thus, $\theta_\alpha$, $s$, and $D$ are hidden constants with no upper bounds given, making the Holtz method unimplementable. The same is true for the constants in Lemmas 15, 17 to 22, and 24 in the paper.

6.1 Questions

Let $f(\lambda):[0,1]\to (0,1)$ have a $\beta-\lfloor\beta\rfloor$)-Hölder continuous $\lfloor\beta\rfloor$-th derivative, where $\beta>0$ is a non-integer.

1. What is an explicit upper bound (with no hidden constants) on the error in approximating $f$ with the Lorentz operators $(Q_{n,r} f)$, described above, of the form $C\cdot M\cdot\max((\lambda(1-\lambda)/n)^{1/2}, 1/n)^r$, where $C=C(r)$ and $M=M(f,r)$ are constants?

2. Same question, but for the polynomial family (g_n) given in (1), above.
3. Same questions as 1 and 2, but f's $(r-1)$-th derivative is in the Zygmund class. (Note that the method of Holtz et al.'s paper as written doesn't apply to integer β; see also Conjecture 34 of that paper.)

7 Simulable and strongly simulable functions

https://mathoverflow.net/questions/404961/from-biased-coins-and-nothing-else-to-biased-coins

There are two kinds of Bernoulli factory functions:

- A function $f(\lambda)$ is *simulable* if an algorithm exists to toss heads with probability $f(\lambda)$ given a coin with probability of heads λ (the "biased coin") as well as a fair coin.
- A function $f(\lambda)$ is *strongly simulable* if an algorithm exists to toss heads with probability $f(\lambda)$ given **only** a coin with probability of heads λ.

Every strongly simulable function is simulable, but not vice versa.

In fact, Keane and O'Brien (1994) showed already that $f(\lambda)$ is strongly simulable if f is simulable and neither 0 nor 1 is in f's domain (that is, if the biased coin doesn't show heads every time or tails every time). And it's also easy to show that if f is strongly simulable, then $f(0)$ must be 0 or 1 if 0 is in f's domain and $f(1)$ must be 0 or 1 whenever 1 is in f's domain.

However, it's not so trivial to find the exact class of strongly simulable functions when f's domain includes 0, 1, or both.

As one illustration of this, the proof of Keane and O'Brien relies on generating a geometric random variate and using that variate to control which "part" of the target function $f(\lambda)$ to simulate. This obviously works on all of [0, 1] if the algorithm uses both the biased coin and a separate fair coin. However, if only the biased coin is used in the algorithm, the geometric random variate is generated using fair bits via the von Neumann method, which however will never terminate if λ is either 0 or 1. In addition, a **result I found** gives sufficient conditions for being strongly simulable when f's domain includes 0, 1, or both. Its proof proceeds by showing, among other things, that the Bernoulli factory for f must flip the input coin and get 0 and 1 before it simulates any fair coin flips via the von Neumann trick.

Question: **Prove or disprove:** Let $f:(D\subseteq [0, 1])\to [0,1]$. Given a coin that shows heads with probability λ (which can be 0 or 1), it is possible to toss heads with probability $f(\lambda)$ using the coin and no other sources of randomness (and, thus, f is *strongly simulable*) **if and only if—**

- f is constant on its domain, or is continuous and polynomially bounded on its domain (*polynomially bounded* means, both f and $1-f$ are bounded below by min(x^n, $(1-x)^n$) for some integer n [Keane and O'Brien 1994]), and
- $f(0)$ is 0 or 1 if 0 is in f's domain and $f(1)$ is 0 or 1 whenever 1 is in f's domain, and
- if $f(0) = 0$ or $f(1) = 0$ or both, then there is a polynomial $g(x):[0,1]\to [0,1]$ with computable coefficients, such that $g(0) = f(0)$ and $g(1) = f(1)$ whenever 0 or 1, respectively, is in the domain of f, and such that $g(x)>f(x)$ for every x in the domain of f, except at 0 and 1, and
- if $f(0) = 1$ or $f(1) = 1$ or both, then there is a polynomial $h(x):[0,1]\to [0,1]$ with computable coefficients, such that $h(0) = f(0)$ and $h(1) = f(1)$ whenever 0 or 1, respectively, is in the domain of f, and such that $g(x)<f(x)$ for every x in the domain of f, except at 0 and 1.

A condition such as "0 is not in the domain of f, or f can be extended to a Lipschitz continuous function on $[0, \epsilon)$ for some $\epsilon>0$" does not work. A counterexample is $f(x)=(\sin(1/x)/4+1/2)\cdot(1-(1-x)^n)$ for $n\ge 1$ ($f(0)=0$), which is strongly simulable at 0 despite not being Lipschitz at 0. ($(1-x)^n$ is the probability of the biased coin showing zero n times in a row.)

8 Multiple-Output Bernoulli Factories

https://mathoverflow.net/questions/412772/from-biased-coins-to-biased-coins-as-efficiently-as-possible

Let J be a closed interval on $(0, 1)$, and let $f(\lambda):J \to (0, 1)$ be continuous.

Then by Keane and O'Brien, f admits an algorithm that solves the Bernoulli factory problem for f (using only the biased coin, in fact). A related problem is a Bernoulli factory that takes a coin with unknown probability of heads $\lambda \in J$ and produces *one or more* samples, at a time, of the probability $f(\lambda)$. This question calls it a *multiple-output Bernoulli factory*.

Obviously, any single-output Bernoulli factory can produce multiple outputs by running itself multiple times. But for some functions f, it may be that producing multiple outputs at a time may use fewer coin flips than producing one output multiple times.

Define the entropy bound as— $$h(f(\lambda))/h(\lambda),$$ where— $$h(x)=-x \ln(x)-(1-x) \ln(1-x),$$ is related to the Shannon entropy function.

8.1 Questions

1. Given that a function $f(\lambda)$ is continuous and maps a closed interval in (0, 1) to (0, 1), is there a multiple-output Bernoulli factory algorithm for f with an expected number of coin flips per sample that is arbitrarily close to the entropy bound, uniformly for every λ in f's domain? Call such a Bernoulli factory an *optimal factory*. (See Nacu and Peres 2005, Question 1.)
2. Does the answer to question 1 change if the algorithm can also use a fair coin in addition to the biased coin?

8.2 Functions with Optimal Factories

So far, the following functions do admit an optimal factory:

- The functions λ and $1-\lambda$.
- Constants in [0, 1]. As Nacu and Peres (2005) already showed, any such constant c admits an optimal factory: generate unbiased random bits using Peres's iterated von Neumann extractor (Peres 1992), then build a binary tree that generates 1 with probability c and 0 otherwise (Knuth and Yao 1976).

It is easy to see that if an optimal factory exists for $f(\lambda)$, then one also exists for $1-f(\lambda)$: simply change all ones returned by the $f(\lambda)$ factory into zeros and vice versa.

Also, as Yuval Peres (Jun. 24, 2021) told me, there is an efficient multiple-output Bernoulli factory for $f(\lambda) = \lambda/2$: the key is to flip the input coin enough times to produce unbiased random bits using his extractor (Peres 1992), then multiply each unbiased bit with another input coin flip to get a sample from $\lambda/2$. Given that the sample is equal to 0, there are three possibilities that can "be extracted to produce more fair bits": either the unbiased bit is 0, or the coin flip is 0, or both are 0. This algorithm, though, might not count as an *optimal factory*, and Peres described this algorithm only incompletely. Indeed, the correctness might depend on how the three possibilities are "extracted to produce more fair bits"; after all, the number of coin flips per sample, for every $\lambda$, must not surpass the entropy bound.

In any case, I believe that not all factory functions admit an optimal factory described here; especially because—

- the question may depend on $f$'s range, and
- the efficiency of even a *single-output* Bernoulli factory depends on $f$'s smoothness (e.g., $O(1/n^{(r+\alpha)/2})$ only if $f$'s $r$th derivative is Hölder continuous with Hölder exponent $\alpha$; Holtz et al. 2011).

See an **appendix in one of my articles** for more information on my progress on the problem.

9 From coin flips to algebraic functions via pushdown automata

https://cstheory.stackexchange.com/questions/50853/from-coin-flips-to-algebraic-functions-via-pushdown-automata

This section is about solving the Bernoulli factory problem on a restricted computing model, namely the model of *pushdown automata* (finite-state machines with a stack) that are driven by flips of a coin and produce new probabilities.

9.1 Pushdown Automata

A *pushdown automaton* has a finite set of *states* and a finite set of *stack symbols*, one of which is called EMPTY and takes a biased coin with an unknown probability of heads. It starts with a given state and its stack starts with EMPTY. On each iteration:

- The automaton flips the coin.
- Based on the coin flip (HEADS or TAILS), the current state, and the top stack symbol, it moves to a new state (or keeps it unchanged) and replaces the top stack symbol with zero, one, or two symbols. Thus, there are three kinds of *transition rules*:
    - (*state*, *flip*, *symbol*) → (*state2*, {*symbol2*}): move to *state2*, replace top stack symbol with same or different one.
    - (*state*, *flip*, *symbol*) → (*state2*, {*symbol2*, *new*}): move to *state2*, replace top stack symbol with *symbol2*, then *push* a new symbol (*new*) onto the stack.
    - (*state*, *flip*, *symbol*) → (*state2*, {}): move to *state2*, *pop* the top symbol from the stack.

When the stack is empty, the machine stops and returns either 0 or 1 depending on the state it ends up at. (For the questions below, let *flip* be HEADS, TAILS, or a rational number in [0, 1]; this likewise reduces to the definition above. The rest of this question assumes the pushdown automaton terminates with probability 1.)

9.2 Algebraic Functions

Let $f: (0, 1) \to (0, 1)$ be continuous. Mossel and Peres (2005) showed that a pushdown automaton can simulate f only if f is *algebraic over the rational numbers* (there is a nonzero polynomial $P(x, y)$ in two variables and whose coefficients are rational numbers, such that $P(x, f(x)) = 0$ for every x in the domain of f). The algebraic function generated by pushdown automata corresponds to a system of polynomial equations, as described by Mossel and Peres (2005) and Esparza et al. 2004.

Let $\mathcal{C}$ be the class of continuous functions that map (0, 1) to (0, 1) and are algebraic over rationals. The constants 0 and 1 are also in $\mathcal{C}$.

Let $\mathcal{D} \subseteq \mathcal{C}$ be the class of functions that a pushdown automaton can simulate.

I don't yet know whether $\mathcal{D}=\mathcal{C}$ (and that was also a question of Mossel and Peres).

The following section of my open-source page, **Pushdown Automata and Algebraic Functions**, contains information on the question. That section sets forth the following results about the class $\mathcal{D}$:

- $\sqrt{\lambda}$ is in $\mathcal{D}$, and so is every rational function in $\mathcal{C}$.
- If $f(\lambda)$ and $g(\lambda)$ are in $\mathcal{D}$, then so are their product and composition.
- If $f(\lambda)$ is in $\mathcal{D}$, then so is every Bernstein-form polynomial in the variable $f(\lambda)$ with coefficients in $\mathcal{D}$.
- If a pushdown automaton can generate a discrete distribution of *n*-letter words, then that distribution's probability generating function is in $\mathcal{D}$ (cf. Dughmi et al. 2021).
- If a pushdown automaton can generate a discrete distribution of *n*-letter words of the same letter, it can generate that distribution conditioned on a finite set of word lengths, or a periodic infinite set of word lengths (e.g., odd word lengths only).
- Every quadratic irrational in (0, 1) is in $\mathcal{D}$.

9.3 Questions

1. For every function in class $\mathcal{C}$, is there a pushdown automaton that can simulate that function? (In other words, is $\mathcal{D}=\mathcal{C}$?).
2. In particular, is min($\lambda$, $1-\lambda$) in class $\mathcal{D}$? What about $\lambda^{1/p}$ for some prime $p\ge 3$?

See also Notes 2 and 3.

10 Reverse-time martingales

This section is withdrawn. For the Bernoulli factory problem, rational functions are probably not much better than polynomials when approximating functions with low smoothness (e.g., those with only three continuous derivatives). This follows from Borwein (1979, theorem 29) and Holtz et al. (2011) (which disproved a theorem of Lorentz relied on by Borwein but maintained it with an extra assumption used in the Bernoulli factory setting).

11 Other Questions

- Given integer $m\ge 0$, rational number $0<k\le \exp(1)$, and unknown heads probability $0\le \lambda \le 1$, find a **Bernoulli factory** for— $$f(\lambda)=\exp(-(\exp(m+\lambda)-(k(m+\lambda)))) = \frac{\exp(-\exp(m+\lambda))}{\exp(-(k(m+\lambda)))},\tag{PD}$$ that, as much as possible, avoids calculating $h(\lambda) = \exp(m+\lambda)-k(m+\lambda)$; in this sense, the more implicitly the Bernoulli factory works with irrational or transcendental functions, the better. A solution is sought especially when k is 1 or 2. Note that the right-hand side of (PD) can be implemented by **ExpMinus** and division Bernoulli factories, but is inefficient and heavyweight due to the need to calculate ϵ for the division factory. In addition there is a Bernoulli factory that first calculates $h(\lambda)$ and $floor(h(\lambda))$ using constructive reals and then runs **ExpMinus**, but this is likewise far from lightweight. (Calculating exp(.) with floating-point operations is not acceptable for this question.)

- Special case of "**Tossing Heads According to a Concave Function**": Let $f(\lambda):[0,1]\to [0,1]$ be writable as $f(\lambda)=\sum_{n\ge 0} a_n \lambda^n,$ where $a_n\ge 0$ is rational, a_n is nonzero infinitely often, and $f(1)$ is irrational. Then what are simple criteria to determine whether there is $0\lt p\lt 1$ such that $0\le a_n\le p(1-p)^n$ and, if so, to find such p? Obviously, if (a_n) is nowhere increasing then $1\gt p\ge a_0$.

- **Simple simulation algorithms**: What simulations exist that are "relatively simple" and succeed with an irrational probability between 0 and 1? What about "relatively simple" Bernoulli factory algorithms for factory functions? Here, "relatively simple" means that the algorithm:

  - Should use only uniform random integers (or bits) and integer arithmetic.
  - Does not use floating-point arithmetic, make direct use of irrational or transcendental functions or constants, or calculate the *p*-adic digit expansion of an irrational or transcendental function, for any real *p*.
  - Should not use rational arithmetic or increasingly complex approximations, except as a last resort.

  See also Flajolet et al., "On Buffon machines and numbers", 2010. There are many ways to describe the irrational probability or factory function. References are sought to papers or books that describe irrational constants or factory functions in any of the following ways:

  - For irrational constants:
    - Simple **continued fraction** expansions.
    - Closed shapes inside the unit square whose area is an irrational number. (Includes algorithms that tell whether a box lies inside, outside, or partly inside or outside the shape.) **Example.**
    - Generate a uniform (*x*, *y*) point inside a closed shape, then return 1 with probability *x*. For what shapes is the expected value of *x* an irrational number? **Example.**
    - Functions that map [0, 1] to [0, 1] whose integral is an irrational number.
  - Bernoulli factory functions with any of the following series expansions, using rational arithmetic only:
    - Alternating power series (see "**Certain Power Series**").
    - Series with nonnegative terms and bounds on the truncation error (see "**Certain Converging Series**").

Prove or disprove:

1. Let $n≥1$ be an integer, and denote the degree-n Bernstein polynomial of g as $B_n(g)$. Suppose $f(λ):[0,1]\to [0,1]$ is non-negative, is concave, and has a continuous second derivative. Then $U_{n,2}(f)=B_n(2f-B_n(f))$ is within $3Mλ(1−λ)/(n^2)$ and within $0.75M/(n^2)$ of f, where M is the maximum of the absolute value of f's second derivative. (Note that Bustamante (2008)'s theorem 11 can't be used here since $U_{n,2}$, though linear, is not a *positive* operator for all continuous functions on the closed unit interval.)

2. Statement 1's conclusion is true if the condition that f is concave is replaced with the condition that $U_{n,2}(f)$ is non-negative for every $n\ge n_0$, for a fixed integer $n_0 \ge 1$.
3. Given that $f:[0,1]\to (0,1]$ is convex, the polynomials $(g_n) = (B_n(f) - \max_{0\le\lambda\le 1}|B_n(f)(\lambda)-f(\lambda)|)$ (where $n\ge 1$ is an integer power of 2) are in Bernstein form of degree n, converge to f from below, and satisfy: $(g_{2n}-g_{n})$ is a polynomial with non-negative Bernstein coefficients once it's rewritten to a polynomial in Bernstein form of degree exactly $2n$. The same is true for the polynomials $(g_n) = (B_n(f) - |B_n(f)(1/2)-f(1/2)|)$, if f is also symmetric about 1/2.

12 End Notes

Note 1: Besides the questions on concave functions given above, there is also the question of whether the solution terminates with a finite expected running time. In this sense, Nacu & Peres showed that a finite expected time is possible only if f is Lipschitz continuous, and I strongly suspect it's not possible either unless f has a Hölder continuous fourth derivative, in view of the results by Holtz given in the section "New coins from old", smoothly.

Note 2: On pushdown automata: Etessami and Yannakakis (2009) showed that pushdown automata with rational probabilities are equivalent to recursive Markov chains (with rational transition probabilities), and that for every recursive Markov chain, the system of polynomial equations has nonnegative coefficients. But this paper doesn't deal with the case of recursive Markov chains where the transition probabilities cannot just be rational, but can also be λ and $1-\lambda$ where λ is an unknown rational or irrational probability of heads.

Note 3: On pushdown automata: Banderier and Drmota (2014) showed the asymptotic behavior of power series solutions $f(\lambda)$ of a polynomial system, where both the series and the system have nonnegative real coefficients. Notably, functions of the form $\lambda^{1/p}$ where $p\ge 3$ is not a power of 2, are not possible solutions, because their so-called "critical exponent" is not dyadic. But the result seems not to apply to *piecewise* power series such as $\min(\lambda,1-\lambda)$, which are likewise algebraic functions.

Note 4: An exception is Chebyshev interpolants, but my implementation experience shows that Chebyshev interpolants are far from being readily convertible to Bernstein form without using transcendental functions or paying attention to the difference between first vs. second kind, Chebyshev points vs. coefficients, and the interval [-1, 1] vs. [0, 1]. For purposes of these open questions, Chebyshev interpolants are impractical, and so are other approximating functions that introduce transcendental functions. By contrast, other schemes (which are of greater interest to me) involve polynomials that are already in Bernstein form or that use only rational arithmetic to transform to Bernstein form (these include linear combinations and iterated Boolean sums of Bernstein polynomials). Indeed, unlike with rational arithmetic (where arbitrary precision is trivial), transcendental functions require special measures to support arbitrary accuracy, such as constructive/recursive reals — floating-point numbers won't do for purposes of these open questions.

$g(\lambda)$ is in the Zygmund class if there is $D>0$ such that $|g(x-h) + g(x+h) - 2g(x)|\le Dh$ wherever the left-hand side is defined and $0\lt h\le\epsilon$. $f(\lambda)$ is in *Gevrey's hierarchy* if there are $B\ge 1, l\ge 1, \gamma\ge 1$ such that $\max |f^{(n)}(\lambda)| \le Bl^n n^{\gamma n}$ for every $n\ge 0$) (see also Kawamura et al. 2015 which however relies on Chebyshev polynomials which are undesirable for my purposes).

Note 5: This condition is also known as a "consistency requirement"; it ensures that not only the polynomials "increase" to $f(\lambda)$, but also their Bernstein coefficients do as well. This condition is equivalent in practice to the following statement (Nacu & Peres 2005). For every integer $n\ge 1$ that's a power of 2, $a(2n, k)\ge\mathbb{E}[a(n, X_{n,k})]= \left(\sum_{i=0}^k a(n,i) {n\choose i}{n\choose {k-i}}/{2n\choose k}\right)$, where $a(n,k)$ is the degree-n polynomial's k-th Bernstein coefficient, where $0\le k\le 2n$ is an integer, and where $X_{n,k}$ is a hypergeometric($2n$, k, n) random variable. A hypergeometric($2n$, k, n) random variable is the number of "good" balls out of n balls taken uniformly at random, all at once, from a bag containing $2n$ balls, k of which are "good". See also my **MathOverflow question** on finding bounds for hypergeometric variables.

Note 6: If $W_n(0)=f(0)$ and $W_n(1)=f(1)$ for every n, then the inequality (PB) is automatically true when $k=0$ and $k=2n$, so that the statement has to be checked only for $0\lt k\lt 2n$. If, in addition, W_n is symmetric about 1/2, so that $W_n(\lambda)=W_n(1-\lambda)$ whenever $0\le \lambda\le 1$, then the statement has to be checked only for $0\lt k\le n$ (since the values $\sigma_{n,k,i} = {n\choose i}{n\choose {k-i}}/{2n \choose k}$ are symmetric in that they satisfy $\sigma_{n,k,i}=\sigma_{n,k,k-i}$). This question is a problem of finding the *Jensen gap* of W_n for certain kinds of hypergeometric random variables (**see Note 5**). Lee et al. (2021) deal with a problem very similar to this one and find results that take advantage of f's (here, W_n's) smoothness, but unfortunately assume the variable is supported on an *open* interval, rather than a *closed* one (namely $[0,1]$) as in this question.

13 References

- E. Voronovskaya, "Détermination de la forme asymptotique d'approximation des fonctions par les polynômes de M. Bernstein", 1932.
- Łatuszyński, K., Kosmidis, I., Papaspiliopoulos, O., Roberts, G.O., "**Simulating events of unknown probabilities via reverse time martingales**", arXiv:0907.4018v2 [stat.CO], 2009/2011.
- Keane, M. S., and O'Brien, G. L., "A Bernoulli factory", *ACM Transactions on Modeling and Computer Simulation* 4(2), 1994.
- Holtz, O., Nazarov, F., Peres, Y., "**New Coins from Old, Smoothly**", Constructive Approximation 33 (2011).
- Nacu, Şerban, and Yuval Peres. "Fast simulation of new coins from old", The Annals of Applied Probability 15, no. 1A (2005): 93-115.
- Knuth, Donald E. and Andrew Chi-Chih Yao. "The complexity of nonuniform random number generation", in *Algorithms and Complexity: New Directions and Recent Results*, 1976.
- Peres, Y., "**Iterating von Neumann's procedure for extracting random bits**", Annals of Statistics 1992,20,1, p. 590-597.
- Mossel, Elchanan, and Yuval Peres. New coins from old: computing with unknown bias. Combinatorica, 25(6), pp.707-724, 2005.
- Icard, Thomas F., "Calibrating generative models: The probabilistic Chomsky–Schützenberger hierarchy." Journal of Mathematical Psychology 95 (2020): 102308.
- Dughmi, Shaddin, Jason Hartline, Robert D. Kleinberg, and Rad Niazadeh. "Bernoulli Factories and Black-box Reductions in Mechanism Design." Journal of the ACM (JACM) 68, no. 2 (2021): 1-30.

- Etessami, K. And Yannakakis, M., "Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations", Journal of the ACM 56(1), pp.1-66, 2009.
- Banderier, C. And Drmota, M., 2015. Formulae and asymptotics for coefficients of algebraic functions. Combinatorics, Probability and Computing, 24(1), pp.1-53.
- Esparza, J., Kučera, A. and Mayr, R., 2004, July. Model checking probabilistic pushdown automata. In Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 2004. (pp. 12-21). IEEE.
- Flajolet, P., Pelletier, M., Soria, M., "**On Buffon machines and numbers**", arXiv:0906.5560v2 [math.PR], 2010.
- von Neumann, J., "Various techniques used in connection with random digits", 1951.
- G.G. Lorentz, "The degree of approximation by polynomials with positive coefficients", 1966.
- Micchelli, C. (1973). The saturation class and iterates of the Bernstein polynomials. Journal of Approximation Theory, 8(1), 1-18.
- Butzer, P.L., "Linear combinations of Bernstein polynomials", Canadian Journal of Mathematics 15 (1953).
- Guan, Zhong. "**Iterated Bernstein polynomial approximations**." arXiv preprint arXiv:0909.0684 (2009).
- Güntürk, C. Sinan, and Weilin Li. "**Approximation with one-bit polynomials in Bernstein form**", arXiv:2112.09183 (2021); Constructive Approximation, pp.1-30 (2022).
- Güntürk, C. Sinan, and Weilin Li. "**Approximation of functions with one-bit neural networks**", arXiv:2112.09181 (2021).
- Draganov, Borislav R. "On simultaneous approximation by iterated Boolean sums of Bernstein operators." Results in Mathematics 66, no. 1 (2014): 21-41.
- Kawamura, Akitoshi, Norbert Müller, Carsten Rösnick, and Martin Ziegler. "**Computational benefit of smoothness: Parameterized bit-complexity of numerical operators on analytic functions and Gevrey’s hierarchy**." Journal of Complexity 31, no. 5 (2015): 689-714.
- Borwein, P.B., "Restricted Uniform Rational Approximations", dissertation, University of British Columbia, 1979.
- Tachev, Gancho. "**Linear combinations of two Bernstein polynomials**", *Mathematical Foundations of Computing*, 2022.
- Lee, Sang Kyu, Jae Ho Chang, and Hyoung-Moon Kim. "Further

sharpening of Jensen's inequality." Statistics 55, no. 5 (2021): 1154-1168.

- Bustamante, J., "Estimates of positive linear operators in terms of second order moduli", J. Math. Anal. Appl. 345 (2008).
- S.N. Bernstein, "The asymptotic behavior of the approximation of functions by their Bernstein polynomials", 1932.
- X. Han, "**Multi-node higher order expansions of a function**", Journal of Approximation Theory, October 2003.