Quantum Bernoulli factory

Suzie Brown

August 8, 2018

1 Introduction

2 Quantum information

From the course by Noah Linden (University of Bristol) and discussion with Thomas Hebdige and David Jennings (Imperial College London). For a comprehensive introduction see for example Nielsen and Chuang (2002) or Wilde (2013).

Quantum mechanics is essentially linear algebra in a Hilbert space, with a few additional properties. The definitions in Sections 2.2 to 2.4 are exactly what you would find in a linear algebra course, apart from the notation.

2.1 Dirac notation

Dirac notation is a convenient way of denoting vectors such that it is easy to visually identify inner and outer products, and thus quickly recognise scalars, vectors and matrices:

- $|v\rangle$ denotes a column vector

- $\langle v|$ denotes a row vector

- $\langle u|v\rangle$ denotes an inner product (resulting in a scalar)

- $|u\rangle\langle v|$ denotes an outer product (resulting in a matrix)

Additionally, $\overline{\alpha}$ denotes the complex conjugate of a scalar α, and $U^\dagger$ denotes the adjoint (conjugate transpose) of an operator U.

2.2 Hilbert space

A *Hilbert space* is a vector space with an inner product $\langle\cdot|\cdot\rangle$ satisfying the following:

- $\langle u|(\alpha|v\rangle + \beta|w\rangle) = \alpha\langle u|v\rangle + \beta\langle u|w\rangle$

- $\langle u|v\rangle = \overline{\langle v|u\rangle}$

- $\langle v|v\rangle \geq 0$ with equality iff $|v\rangle$ is the zero vector.

2.3 Orthonormal bases

An *orthonormal basis* of a Hilbert space $\mathcal{H}$ is a set of vectors $\{v_1, \ldots, v_n\}$ in $\mathcal{H}$ such that:

- $\mathrm{span}(\{v_1, \ldots, v_n\}) = \mathcal{H}$

- $\langle v_i|v_j\rangle = \delta_{ij}$

Restricting to the space $\mathbb{C}^2$, which is all that is needed to understand the quantum Bernoulli factory, we have the *computational basis* $\{|0\rangle, |1\rangle\} = \{(1,0)^T, (0,1)^T\}$. This is the canonical basis and is henceforth used wherever not specified otherwise. Since it is an orthonormal basis, every vector $|v\rangle$ in $\mathbb{C}^2$ has a unique representation

$$|v\rangle = \alpha|0\rangle + \beta|1\rangle \equiv (\alpha, \beta)^T$$

for some $\alpha, \beta \in \mathbb{C}$. For reasons which will probably not become apparent in this treatment, we will restrict to *normalised* vectors, requiring also $|\alpha|^2 + |\beta|^2 = 1$. We will also consider two vectors equivalent if they differ only by an overall phase, i.e. $|u\rangle \equiv |v\rangle$ if $|u\rangle = e^{i\theta}|v\rangle$ for some θ. This is because it is impossible to distinguish between two such vectors with any measurement.

To ensure coherency with the properties of the inner product, we have that

$$\langle v| = \overline{\alpha}\langle 0| + \overline{\beta}\langle 1|.$$

The inner product of $|u\rangle = u_0|0\rangle + u_1|1\rangle$ with $|v\rangle = v_0|0\rangle + v_1|1\rangle$ is therefore computed as

$$\begin{aligned}\langle u|v\rangle &= (\overline{u_0}\langle 0| + \overline{u_1}\langle 1|)(v_0|0\rangle + v_1|1\rangle) \\ &= \overline{u_0}v_0 \langle 0|0\rangle + \overline{u_0}v_1 \langle 0|1\rangle + \overline{u_1}v_0 \langle 1|0\rangle + \overline{u_1}v_1 \langle 1|1\rangle \\ &= \overline{u_0}v_0 + \overline{u_1}v_1.\end{aligned}$$

One alternative choice of orthonormal basis which is worth mentioning is given by $\{|+\rangle, |-\rangle\}$, consisting of the states

$$|+\rangle := \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

$$|-\rangle := \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

2.4 Linear operators

A linear operator is an operator with the property

$$A(\alpha|u\rangle + \beta|v\rangle) = \alpha A|u\rangle + \beta A|v\rangle.$$

It is therefore fully defined according to its action on an orthonormal basis. For instance, the quantum NOT operator (usually denoted X) is defined by

$$\begin{aligned}X|0\rangle &= |1\rangle \\ X|1\rangle &= |0\rangle\end{aligned}$$

Equivalently, X can be expressed as a matrix with respect to the computational basis:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

In Dirac notation, X can be written in terms of outer products of basis states:

$$X = |0\rangle\langle 1| + |1\rangle\langle 0|$$

Then if X acts on the state $|v\rangle = \alpha|0\rangle + \beta|1\rangle$, we have

$$\begin{aligned}X|v\rangle &= (|0\rangle\langle 1| + |1\rangle\langle 0|)(\alpha|0\rangle + \beta|1\rangle) \\ &= \alpha|0\rangle\langle 1|0\rangle + \alpha|1\rangle\langle 0|0\rangle + \beta|0\rangle\langle 1|1\rangle + \beta|1\rangle\langle 0|1\rangle \\ &= \alpha|1\rangle + \beta|0\rangle\end{aligned}$$

as desired.

- A linear operator U is said to be *self-adjoint* (or *hermitian*) if $U^\dagger = U$

- A linear operator U is said to be *unitary* if $UU^\dagger = U^\dagger U = I$

For example, the operator X is self-adjoint and unitary. Crucially, untiary operators preserve normalisation, so they map states to states. It is also easy to see that unitary transformations are always reversible (i.e. the inverse operator exists).

2.5 Rules of Quantum Mechanics

1. *States* of a quantum mechanical system correspond to normalised vectors in Hilbert space, up to an overall phase.

2. *Evolutions* of the system correspond to unitary operators.

3. *Measurements* on quantum states correspond to self-adjoint operators — see below.

2.5.1 Spectral theorem and measurement

The outcome of a measurement depends on the state of the system and the type of measurement performed. The spectral theorem states that every self-adjoint operator A can be represented by its spectral decomposition

$$A = \sum_i \lambda_i P_i \tag{1}$$

where $\{\lambda_1, \ldots, \lambda_k\}$ is the set of *distinct* eigenvalues of A, and P_i is the projection operator onto the eigenspace corresponding to eigenvalue λ_i.

When we measure a state $|x\rangle$ using operator A, the measurement outcome we observe is one of the eigenvalues of A. In particular, we observe λ_i with probability $\langle x|P_i|x\rangle$. Making a measurement causes the system to collapse into the eigenstate corresponding to the observed eigenvalue; that is, the state after measurement is proportional to $P_i|x\rangle$.

Example 1. To give a concrete example, let us consider again the operator X. This operator has eigenvalues ± 1 corresponding to eigenvectors (known as *eigenstates* in quantum mechanics) $|+\rangle$ and $|-\rangle$ respectively. Therefore X admits the diagonal representation

$$X = |+\rangle\langle+| - |-\rangle\langle-|$$

which is of the form (1). Now suppose we make a measurement on the state $|v\rangle = |0\rangle$, using X. The outcome of the measurement will be an eigenvalue of X: either $+1$ or -1. We observe the outcome $+1$ with probability

$$\langle v|P_{+1}|v\rangle = \langle 0|+\rangle\,\langle+|0\rangle = 1/2$$

in which case the state after measurement is

$$P_{+1}|v\rangle = |+\rangle\langle+|0\rangle \propto |+\rangle.$$

Alternatively, with probability $1/2$ we observe the outcome -1 and the state after measurement is $|-\rangle$.

In this example, the two outcomes are equally likely because the state is 'equidistant' from the two eigenstates of X. In general, outcomes that leave the state after measurement closer to the original state are more likely. We will see what is meant by distance between states in Section 2.8.

A particularly important type of measurement is measurement in the computational basis. In this case the self-adjoint operator used is $0 \cdot |0\rangle\langle0| + 1 \cdot |1\rangle\langle1| = |1\rangle\langle1|$. Measuring a state $|v\rangle = \alpha|0\rangle + \beta|1\rangle$ in the computational basis returns 0 with probability $|\alpha|^2$ or 1 with probability $|\beta|^2$. This is particularly useful in quantum computing algorithms because it deterministically converts the basis states $|0\rangle$ and $|1\rangle$ to their classical equivalents 0 and 1, ready for classical output.

2.6 Quantum randomness

It is worth remarking at this point on the difference between quantum and classical randomness. As statisticians we are used to dealing with randomness, but we accept that randomness is always part of a model, and is not purported to exist in nature. We artificially introduce random variables into our models to account for a lack of information, either about the state of the system or about the (presumably deterministic) processes that govern certain phenomena.

On the contrary, our uncertainty about the outcome of a measurement on a quantum system is of a different kind. This uncertainty is not a symptom of our lack of knowledge: even when we know exactly the state of the system, as in Example 1, we are still unable to predict with certainty the outcome of a measurement on the system. The randomness here is intrinsic; it really does exist in nature.
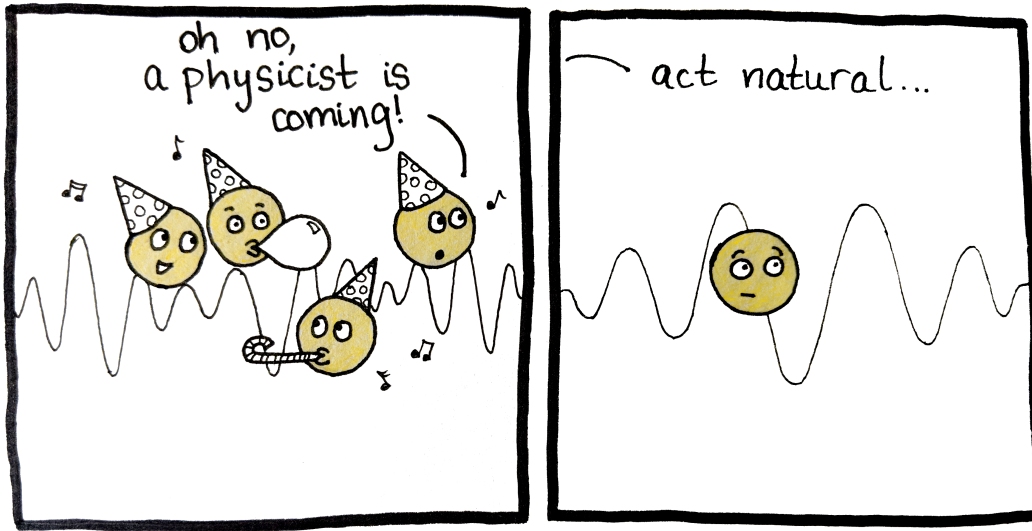
Figure 1: When a quantum system is in a superposition of states, making a measurement causes the system to collapse into the eigenstate corresponding to the observed eigenvalue.

2.7 Pure and mixed states

So far we have only encountered *pure states*, like $|0\rangle$ or $|+\rangle$. But we can also consider probabilistic mixtures of pure states, known as *mixed states*, like

$$|v\rangle = \sum_i p_i |v_i\rangle$$

where $\{p_i\}$ is a probability distribution and the $|v_i\rangle$ are pure states. You can think of this as a 'source' which emits a state $|v_i\rangle$ with probability p_i. A classical analogue is a bit which takes the value 1 with probability p or 0 with probability $1 - p$ (a p-coin). In this case, we can say that the expected value of the bit is p.

There is no full ordering on the set of quantum states, so it is not possible to define the expected value of a mixed state. However, we can calculate the expected value of a measurement on a mixed state, using the state's *density operator*.

2.7.1 Density operators

The density operator of a mixed state $|v\rangle = \sum_i p_i |v_i\rangle$ is given by

$$\rho = \sum_i p_i |v_i\rangle\langle v_i|.$$

Different mixtures of states may have the same density operator, but in this case is is impossible to distinguish between them by any measurement. For example, a source which emits $|0\rangle$ or $|1\rangle$ each with probability 1/2 has the same density operator $\rho = I/2$ as a source which emits $|+\rangle$ or $|-\rangle$ each with probability 1/2, and the two sources are indistinguishable.

Suppose we measure a mixed state using the self-adjoint operator $A = \sum_k \lambda_k P_k$, where $\{\lambda_k\}$ are the distinct eigenvalues of A and P_k the projection onto the corresponding eigenspace. Recall (from Section 2.5.1) that the outcome of the measurement will be an eigenvalue of A. The probability of observing a particular outcome is

$$\mathbb{P}(\text{observe } \lambda_j) = \sum_i p_i \langle v_i | P_j | v_i \rangle = tr(\rho P_j)$$

where $tr(\cdot)$ denotes the trace of an operator. Note that $tr(\rho) = 1$ because the states are normalised and $\{p_i\}$ is a probability distribution.

The state $\rho = I/2$ is called the *maximally mixed* state; it expresses complete ignorance, like a classical $(1/2)$-coin.
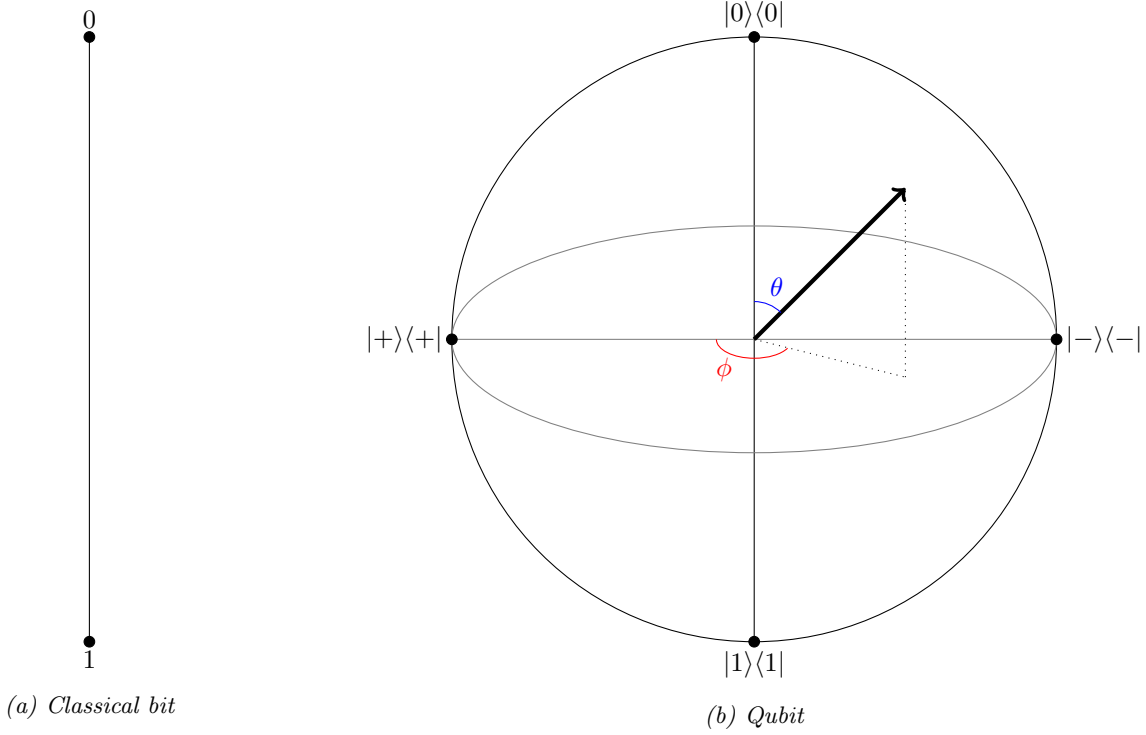
| (a) Classical bit | (b) Qubit |

Figure 2: State space of classical bits versus qubits. The state space of a classical bit is the line segment $[0,1]$. The 'pure' states 0 and 1 are at the endpoints, and probabilistic mixtures of the two lie inbetween. The state space of a qubit is the Bloch sphere. Pure states are on the surface of the sphere, and mixed states lie in the interior. The centre of the sphere is the maximally mixed state $\rho = I/2$.

2.8   The Bloch sphere

Density operators have a neat geometrical representation as points in a unit sphere (Figure 2b). This is called the *Bloch sphere*, and can be thought of as the state space of a quantum bit (qubit).

An arbitrary pure state can be written in the form

$$|v\rangle = e^{i\psi} \left[\cos\left(\frac{\theta}{2}\right) |0\rangle + e^{i\phi} \sin\left(\frac{\theta}{2}\right) |1\rangle \right].$$

Since the overall phase $e^{i\psi}$ can be ignored, the state is parametrised by two angles θ and ϕ. Taking these as spherical coordinates with unit radius, we can think of each pure state as a point on the unit sphere (or equivalently, a unit vector). Orthogonal states lie opposite each other on the Bloch sphere.

Mixed states can also be represented in the Bloch sphere. The Bloch vector for a mixed state is found by taking the weighted average of the Bloch vectors of its component pure states. Due to the triangle inequality, mixed states therefore lie strictly inside the unit sphere. This illustrates that several different mixtures of states can have the same density operator. Qubit states are distinguishable if and only if they correspond to different points in the Bloch sphere.

There is an analogy here with classical information. The state space of a classical bit is the unit interval: its value could be 0 or 1, or we could take a probabilistic mixture of the two. The 'pure' states then are 0 and 1, which lie on the 'surface' of the interval, while 'mixed' states lie strictly inside the interval (Figure 2a).

Finally, unitary operators are rotations of the Bloch sphere. This illustrates that unitary transformations are reversible.

2.8.1   Geometric view of measurement

Suppose we are going to measure a pure state $|v\rangle = \cos\left(\frac{\theta}{2}\right) |0\rangle + e^{i\phi} \sin\left(\frac{\theta}{2}\right) |1\rangle$ in the computational basis. There is a simple relationship between the probability of each outcome and the projection of $|v\rangle$ onto the line segment between $|0\rangle$ and $|1\rangle$ on the Bloch sphere.
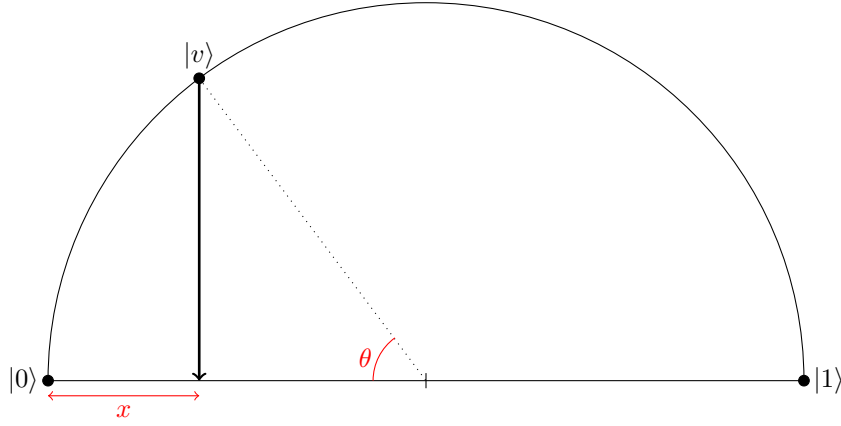
Figure 3: When measuring a state in the computational basis, the probability of observing a certain eigenvalue is related to the projected distance x from the corresponding eigenstate.

Figure 3 shows the semicircular slice of the Bloch sphere on which $|v\rangle$ lies. The value of ϕ determines which slice this is, and conditional on that the state only depends on θ — the following calculations are independent of ϕ and apply to an arbitrary pure state. The state $|v\rangle$ is projected onto the line segment between $|0\rangle$ and $|1\rangle$ as shown. Since the sphere has unit radius, the distance labelled x on Figure 3 is

$$x = 1 - \cos\theta.$$

Using the expansion of $|v\rangle$ in the computational basis, we can calculate the probability of observing measurement outcome 0:

$$\mathbb{P}(\text{observe } 0) = \langle v|0\rangle \langle 0|v\rangle = \cos^2\left(\frac{\theta}{2}\right) = \frac{\cos\theta + 1}{2} = 1 - \frac{x}{2}.$$

This relationship between the projected distance and measurement probabilities also applies to mixed states. If we take any mixture of states with the same value of θ (but different values of ϕ), the result will hold, because x and $\mathbb{P}(\text{observe } 0)$ will be the same for every component state, and hence for the mixed state too. But the result holds for general mixed states as well. Any point in the sphere whose projection onto the line segment between $|0\rangle$ and $|1\rangle$ lies a distance x from $|0\rangle$ will measure 0 with probability $1 - x/2$.

Furthermore, this illustrates a more general concept, which applies not only to measurement in the computational basis but to any quantum measurement. Any valid measurement operator on one qubit (that is non-trivial in the sense of having more than one possible outcome) has two eigenstates that lie opposite each other on the Bloch sphere. Projecting a state $|v\rangle$ onto the line segment between the two eigenstates gives the probabilities of observing the associated eigenvalues.

3 Bernoulli factories

Roughly speaking, a Bernoulli factory is an algorithm that takes samples a Bernoulli(p) random variable as inputs and uses them to produce a Bernoulli($f(p)$) sample as output, where f is a known function but the parameter p is unknown (therefore the algorithm must not depend on p). The imaginary device for generating Bernoulli(p) random variables is sometimes referred to as a p-coin or black box; samples thereof may be called flips/tosses or queries.

Definition 1 (Bernoulli factory). Let $f : S \to [0,1]$ be a function with domain $S \subseteq [0,1]$, and suppose we have access to a sequence of Bernoulli random variables $X_1, X_2, \ldots \overset{iid}{\sim}$ Bernoulli(p) with unknown parameter $p \in [0,1]$. Let $U \in \mathcal{U}$ denote a set of auxiliary random variables with known distributions, independent of p. Let τ be a stopping time with respect to the natural filtration, such that $\mathbb{P}(\tau < \infty) = 1$.

Definition 2 (simulable). A function $f : S \to [0,1]$ is said to be *simulable* if there is a Bernoulli factory that simulates $f(p)$ for all $p \in S$. It is *strongly simulable* if there is a Bernoulli factory that simulates $f(p)$ without the need for auxiliary random variables U.
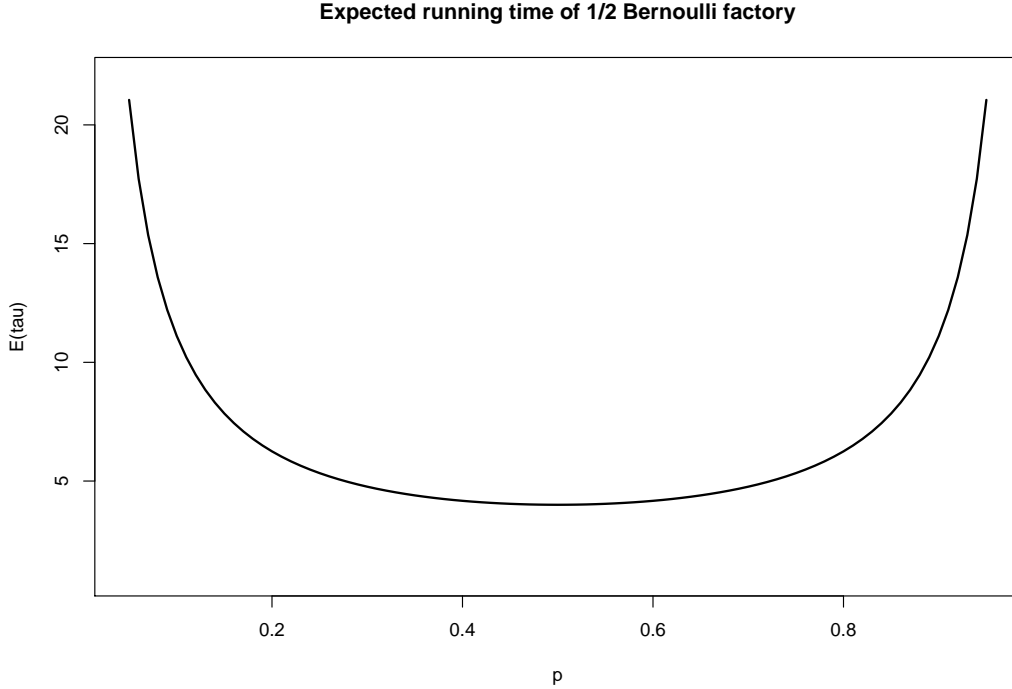
Expected running time of 1/2 Bernoulli factory

Figure 4: ●

The first formal treatment of a Bernoulli factory problem was presented by Von Neumann (1951), who described a Bernoulli factory for the function $f(p) \equiv 1/2$:

> *To cite a human example, for simplicity, in tossing a coin it is probably easier to make two consecutive tosses independent than to toss heads with probability exactly one-half. If independence of successive tosses is assumed, we can reconstruct a 50-50 chance out of even a badly biased coin by tossing twice. If we get heads-heads or tails-tails, we reject the tosses and try again. If we get head-tails (or tails-heads), we accept the result as heads (or tails). The resulting process is rigorously unbiased, although the amended process is at most 25 percent as efficient as ordinary coin-tossing.*

Most authors now use the slightly more convenient convention with the outcomes inverted, so that once 10 or 01 is observed, the output of the Bernoulli factory is equal to the last coin flip (0 or 1 respectively). It is clear that, as claimed, this procedure produces a $(1/2)$-coin, because 01 and 10 appear with the same probability, so conditional on observing one or the other, both outcomes have probability $1/2$. The procedure is formalised in Example 2.

Example 2. $f(p) \equiv 1/2$.
This constant function is strongly simulable, using the procedure:

$$\tau = \min\{t \in \{2, 4, \dots\} : X_{t-1} \neq X_t\}$$
$$\mathcal{A}(X_1, X_2, \dots) = X_\tau.$$

The running time of this algorithm is random, and indeed unbounded (but it is finite almost surely, making it a valid Bernoulli factory). In particular, $\tau \stackrel{d}{=} 2 \times \mathrm{Geom}(2p(1-p))$, with expectation $\mathbb{E}(\tau) = \frac{1}{p(1-p)}$. Figure 4 shows the expected running time for various values of p. The expected running time achieves its minimum value of four when $p = 1/2$ (that is, the coin was already unbiased), hence von Neumann's claim that "the amended process is at most 25 percent as efficient as ordinary coin-tossing" (which has deterministic running time one). The expected running time approaches infinity when p is close to 0 or 1, but is moderate across a large range non-extreme values.

The algorithm given in Example 2 can be extended to simulate other constant functions of the form $f(p) \equiv 1/k$

for $k \in \{2, 3, \dots\}$. The generalised procedure is:

$$\tau = \min \left\{ t \in \{k, 2k, 3k, \dots\} : \sum_{i=t-k+1}^{t} X_i = 1 \right\}$$
$$\mathcal{A}(X_1, X_2, \dots) = X_\tau.$$

However, the running time increases badly with k, even under certain early stopping strategies. Of course, any constant $f(p) \equiv c$ is simulable by employing an auxiliary random variable $U \sim \text{Bernoulli}(c)$, but von Neumann's example shows that $c = 1/2$ is also strongly simulable.

Another important example is the case of $f(p) = p^2$, which is strongly simulable using exactly two queries, returning heads if and only if both inputs are heads. Example 3 presents a generalisation of this procedure.

Example 3. $f(p) = p^k, k \in \{1, 2, \dots\}$.
This problem is solved easily by flipping the p-coin k times and outputting heads if and only if all k flips return heads. No auxiliary random variable is required.

$$\tau = k$$
$$\mathcal{A}(X_1, \dots, X_T) = \mathbb{I}\{X_1 = 1, X_2 = 1, \dots, X_\tau = 1\}$$

In this case the running time is deterministically equal to k. It is clear that this procedure produces heads with probability p^k, since the flips are independent and each gives heads with probability p.

The running time of this algorithm can be decreased by using an early stopping strategy, whereby we stop the process as soon as a tails input appears (in which case we know immediately that the output will be tails). In this case the running time is random, but still bounded above by k:

$$\tau = k \wedge \min\{t \in \{1, 2, \dots\} : X_t = 0\} \overset{d}{=} \min\{k, \text{Geom}(1-p)\}.$$

Figure 5 shows the expected running time of this procedure for a few values of k. We see that early stopping can significantly reduce the expected running time when p is not close to 1, and that the gain in efficiency is greater for larger k. In the case $k = 1$, where f is the identity, the number of coin flips required is deterministically equal to one, for any p.

Examples 2 and 3 cast some light on the sort of manipulations that are possible when constructing Bernoulli factories. Lemma 1 provides some general properties, which are trivial to prove.

Lemma 1. *Suppose we have a Bernoulli factory $\mathcal{A}_f$ for the function $f : S_f \to [0, 1]$ and a Bernoulli factory $\mathcal{A}_g$ for $g : S_g \to R_g \subseteq [0, 1]$. Then*

(a) *There exists a Bernoulli factory for $1 - f : S_f \to [0, 1]$, given by inverting the output of $\mathcal{A}_f$.*

(b) *If $R_g \subseteq S_f$, then there exists a Bernoulli factory for the composition $f \circ g : S_g \to [0, 1]$, given by the composition $\mathcal{A}_f \circ \mathcal{A}_g$.*

(c) *There exists a Bernoulli factory for the product $fg : S_f \cap S_g \to [0, 1]$, given by $A_f \wedge A_g$.*

In the case of (b) and (c), the running time is bounded if and only if the running time of both A_f and A_g is bounded.

References

Nielsen, M. A. and Chuang, I. (2002), *Quantum computation and quantum information*, AAPT.

Von Neumann, J. (1951), 'Various techniques used in connection with random digits', *Appl. Math Ser* **12**(36-38), 3.

Wilde, M. M. (2013), *Quantum information theory*, Cambridge University Press.

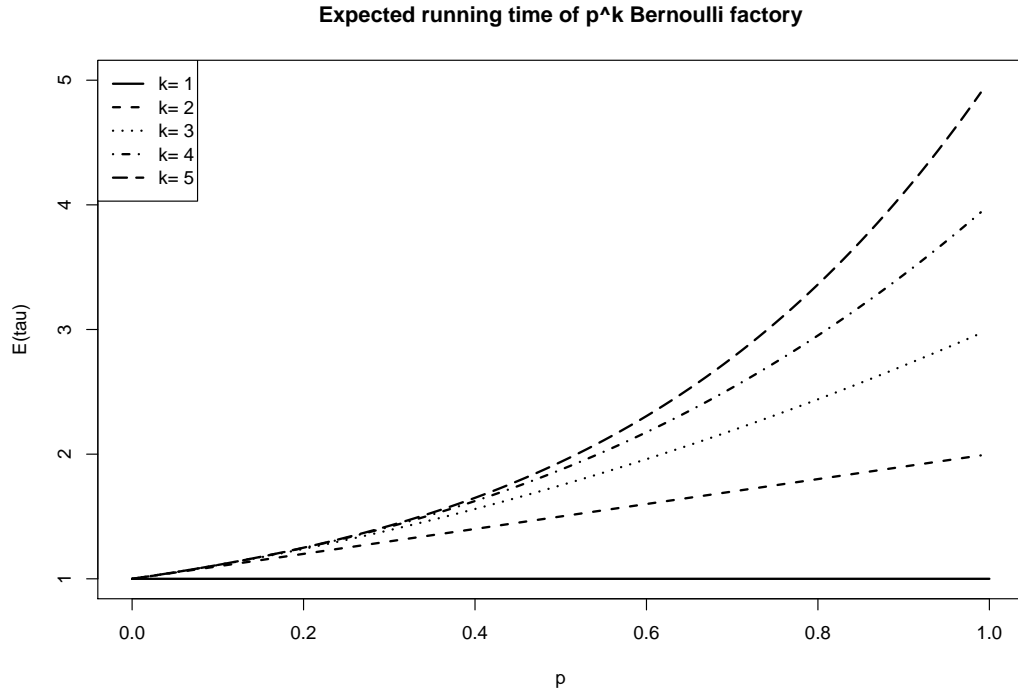**Expected running time of p^k Bernoulli factory**

Figure 5: Expected running time for the $f(p) = p^k$ Bernoulli factory described, with early stopping. Without early stopping, the running time is deterministically equal to k. The early stopping procedure significantly increases the efficiency, particularly when p is small or k is large. The minimum number of queries is one for all k and p, and of course the $k = 1$ (identity) Bernoulli factory requires exactly one query for any p.