Working Notes Concerning Lester Lipsky's

A Markov Model of Maxwell Boltzmann & Bose Einstein Statistics

Nicholas Wheeler
November 2016

Introduction. My initial objective will be to recall to mind some standard material, to establish notation and terminology. I expect to draw fairly heavily from the *Mathematica* slide show "Some Miscellaneous Adventures in Experimental Mathematical Physics," (notes form a Reed College Physics Seminar presented on 9 November 2011), and also from other Wheeler notes, both ancient and modern.

Basics of discrete Markov processes. Let the stochastic vector

$$\mathsf{p}_0 = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}_0 \quad : \quad \text{all } p_i \geqslant 0,\ \sum_{i=1}^{n} p_i = 1$$

describe what is known about the initial state of an n-state system. We are interested in situations in which the state evolves by discrete steps, where the evolution is achieved by linear transformations that remain unchanged from step to step:

$$\mathsf{p}_i \longrightarrow \mathsf{p}_{i+1} = \mathbb{T}\,\mathsf{p}_i \quad : \quad \mathbb{T} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{pmatrix}$$

Hwere the elements of the "transition matrix" $\mathbb{T}$ are by nature conditional probabilities:

$$t_{jk} = \text{prob}(j|k) = \text{probability of transition } j \leftarrow k$$

The requirement that

$$\mathsf{p} \text{ stochastic} \quad \Longrightarrow \quad \mathbb{T}\mathsf{p} \text{ stochastic (all } \mathsf{p}\text{)}$$

gives

$$\sum_j \sum_k t_{jk} p_k = \sum_k p_k \sum_j t_{jk} = 1 \quad : \quad \text{all } \mathbf{p}_0$$

In short: the individual columns of the transition matrix $\mathbb{T}$ must be themselves stochastic.

Many (but by no means all) cases of interest are subject to the PRINCIPLE OF DETAILED BALANCE

$$\text{probability } j \leftarrow k = \text{probability } k \leftarrow j$$

which enforces the symmetry of $\mathbb{T}$; such real symmetric transition matrices are said to be "balanced."

The upshot of the Chapman-Kolmogorov equation (which at the moment I am content to justify by random numerical simulation) is that products of distinct (generally unbalanced) transition matrices are again transition matrices:

The set of all $\mathbb{T}$-matrices is multiplicatively closed

But the same cannot be said of the subset of all balanced transition matrices: products of symmetric matrices are symmetric if and only if they commute.

This subject is illuminated by the PERRON-FROBENIUS THEOREM, which asserts that if the real square matrix $\mathbb{M}$ is "positive" (meaning all elements greater than 0), then

- There is always a non-degenerate real eigenvalue (called the "leading" eigenvalue)...
- ...which is always greater than moduli of the other eigenvalues (real or complex);
- The elements of the associated eigenvector are all real and of the same sign (so it is always stochastic to within a multiplicative factor), and...
- ...there are no other such eigenvectors; all others have either complex elements or elements of opposite sign.

Those properties pertain also—subject to certain frequently satisfied conditions —to "non-negative" (some elements 0, but all positive) real square matrices.

Random simulations—which of course conform to the assertions of the Perron-Frobenius theorem—drive home the fact that the eigenvalues of unbalanced $\mathbb{T}$-matrices (which *Mathematica* presents in order of descending absolute value) very frequently occur in conjugate complex pairs (as also do their associated eigenvectors). The eigenvalues of balanced $\mathbb{T}$-matrices are, of course, all real, and the (invariably real) eigenvectors associated with distinct eigenvalues are invariably orthogonal. Simulations reveal that for all $\mathbb{T}$-matrices —balanced or unbalanced—the leading eigenvalue is invariably unity. And that for balanced (but not for unbalanced) $\mathbb{T}$-matrices the elements of the leading unitized eigenvector are invariably identical, and given by 1/dimension.

Familiarly, powerful conclusions can be drawn from the fact that every symmetric real matrix (more generally: every hermitian complex matrix) admits of "spectral resolution"

$$\mathbb{M} = \sum_i \lambda_i \mathbb{P}_i \tag{1}$$

where the $\{\lambda_i\}$ are the real eigenvalues of $\mathbb{M}$ and the $\{\mathbb{P}_i\}$ are a complete set of orthogonal projection matrices that project onto the eigenspaces of $\mathbb{M}$. Less familiar is the

GENERALIZED SPECTRAL RESOLUTION

that pertains to non-symmetric real (more generally: non-hermitian complex) matrices. The standard theory is as it stands of direct relevance to the theory of balanced Markov processes, but the theory of unbalanced processes requires the more general conception of spectral resolution that I digress now to develop. It hinges on the notion of "biorthogonality."

Let $\{|a_i)\}$ refer to a set of (generally complex) eigenvectors of the (real or complex) $n \times n$ matrix $\mathbb{M}$, and $\{\lambda_i\}$ to the associated (generally complex) eigenvalues. Following Dirac, we write $(a_i|$ to denote the conjugate transpose (adjoint) of $|a_i)$. When $\mathbb{M}$ is real and symmetric the $\{|a_i)\}$ are (or, in cases of spectral degeneracy, can be contrived to be) orthogonal, but in general they provide a non-orthogonal basis in the (generally complex) vector space $\mathcal{V}_n$. Since neither orthogonality nor normality are assumed, we have

$$(a_i|a_j) = g_{ij}$$

where $\|g_{ij}\|$ is hermitian and (by linear independence) non-singular. The generic element $|x) \in \mathcal{V}_n$ can be developed

$$|x) = |a_k)\, x^k$$

which gives

$$(a_j|x) = g_{jk} x^k$$

Writing $\|g_{ij}\|^{-1} = \|g^{ij}\|$ we have

$$g^{ij}(a_j|x) = g^{ij} g_{jk} x^k = \delta^i{}_k x^k = x^i$$

giving

$$|x) = |a_i) g^{ij} (a_j|x) \quad : \quad \text{all } |x)$$

from which we conclude that

$$|a_i) g^{ij} (a_j| = \mathbb{I}$$

Introduce now into $\mathcal{V}_n$ a second non-orthogonal basis with elements

$$|A^j) = |a_i) g^{ij} \quad \text{similarly} \quad (A^i| = g^{ij}(a_j|$$

which supply these alternative constructions of the unit matrix:

$$|A^i)(a_i| = |A^i)g_{ij}(A^j| = |a_j)(A^j| = \mathbb{I}$$

Moreover

$$(A^i|a_j) = g^{ik}(a_k|a_j) = g^{ik}g_{kj} = \delta^i{}_j$$

which is to say:

$$|A^i) \perp \text{all } |a_j) : j \neq i$$

The non-orthonormal bases $\{|a_i)\}$ and $\{|A^j)\}$ are said to be "biorthogonal" (or "reciprocal").[1]

Look now to the matrices

$$\mathbb{P}_i = |a_i)(A^i| \quad : \quad \text{no summation on } i$$

where the index placement on $\mathbb{P}_i$ is merely conventional (intended to convey no transformation-theoretic meaning). Those are seen to be orthogonal projection matrices

$$\mathbb{P}_i\mathbb{P}_j = |a_i)(A^i|a_j)(A^j| = |a_i)\delta^i{}_j(A^j| = \begin{cases} \mathbb{P}_i & : \quad i = j \\ \mathbb{O} & : \quad i \neq j \end{cases}$$

and have already been seen to be complete: $\sum_i \mathbb{P}_i = \mathbb{I}$. They project onto 1-spaces (rays); specifically

$$\left.\begin{array}{rl} \text{right action:} & \mathbb{P}_i|x) = |a_i)\,x^i \\ \text{left action:} & (x|\mathbb{P}_i = x_i\,(A^i| \end{array}\right\} \quad : \quad \text{no summation on } i$$

Let $\mathbb{W}$ be an *arbitrary* $n \times n$ square matrix. We are in position now to write

$$\begin{aligned} \mathbb{W} &= \mathbb{I}\,\mathbb{W}\,\mathbb{I} \\ &= \textstyle\sum_{ij} \mathbb{P}_i\,\mathbb{W}\,\mathbb{P}_j \\ &= \textstyle\sum_{ij} |a_i)(A^i|\mathbb{W}|a_j)(A^j| \qquad (2) \\ &= \textstyle\sum_{ij} w^i{}_j|a_i)(A^j| \quad \text{where} \quad w^i{}_j = (A^i|\mathbb{W}|a_j) \qquad (3) \end{aligned}$$

Here $\mathbb{W}$ is displayed as a weighted linear combination of the n^2-population of matrices

$$\mathbb{F}_i{}^j = |a_i)(A^j| \quad : \quad \mathbb{F}_i{}^i = \mathbb{P}_i$$

(these provide a "basis in the space of matrices") and $\|m^i{}_j\|$ provides, with respect to the non-orthogonal $\{|a_i)\}$-basis, the matrix representation of $\mathbb{W}$; it permits $|x) \to |\tilde{x}) = \mathbb{W}|x)$ to be represented

$$x^i \to \tilde{x}^i = m^i{}_j x^j$$

[1] When $\{|a_i)\}$ is in fact orthonormal ($g^{ij} = \delta^{ij}$) the distinction between $\{|a_i)\}$ and $\{|A^i)\}$—as also between $\{(a_i|\}$ and $\{(A^i|\}$—evaporates.

We note in passing that the $\mathbb{F}$-matrices are tracewise orthogonal

$$\mathrm{tr}(\mathbb{F}_i{}^j\,\mathbb{F}_k{}^l) = \delta_{ik}{}^{jl}$$

so (3) can be written as a "Fourier identity"

$$\mathbb{W} = \textstyle\sum_{ij} w^i{}_j \mathbb{F}_i{}^j \quad \text{with} \quad w^i{}_j = \mathrm{tr}(\mathbb{F}_i{}^j\,\mathbb{W})$$

All of which simplifies very greatly when $\mathbb{W}$ is taken to be the matrix $\mathbb{M}$ that *gave birth* to the biorthogonal bases $\{|a_i)\}$ and $\{|A_i)\}$. For (2) then becomes

$$\begin{aligned}\mathbb{M} &= \textstyle\sum_{ij} |a_i)(A^i|\mathbb{M}|a_j)(A^j| \\ &= \textstyle\sum_{ij} |a_i)\lambda_j \delta^i{}_j (A^j| \\ &= \textstyle\sum_i \lambda_i \mathbb{P}_i \end{aligned} \tag{4}$$

which provides the "generalized spectral resolution" of $\mathbb{M}$. When $\mathbb{M}$ is hermitian (or real symmetric) the eigenvalues λ_i and projectors $\mathbb{P}_i$ are real and (4) gives back the standard spectral resolution (1). But when hermiticity/real symmetry are abandoned they may become complex, though the structure ($4 \equiv 1$) survives. When $\mathbb{M}$ is real but non-symmetric (as unbalanced transition matrices are) they—even though sometimes complex—still conspire to reproduce the reality of $\mathbb{M}$. **End of digression**.

The generalized spectral representation (4) permits some standard tricks to be performed in the non-standard context provided by balanced/unbalanced $\mathbb{T}$-matrices, tricks that hinge on the circumstances that the $\mathbb{P}_i$ comprise a complete set of orthogonal projectors with

$$\mathrm{tr}\,\mathbb{P}_i = \text{dimension } d_i \text{ of the subspace onto which } \mathbb{P}_i \text{ projects} \tag{5}$$

For example—working from

$$\mathbb{T} = \textstyle\sum_i \lambda_i \mathbb{P}_i \tag{6}$$

—we have

$$\begin{aligned}\mathbb{T}^n &= \textstyle\sum_i (\lambda_i)^n \mathbb{P}_i \quad : \quad n = 0, 1, 2, \ldots \\ \mathbb{T}^{-1} &= \textstyle\sum_i (\lambda_i)^{-1} \mathbb{P}_i \\ &= \infty \text{ if any of the eigenvalues vanish} \\ \mathrm{tr}\,\mathbb{T} &= \textstyle\sum_i \text{real parts of the eigenvalues}\end{aligned} \tag{7}$$

—the latter because the eigenvalues of $\mathbb{T}$ (real) occur in conjugate pairs. As previously noted, the leading eigenvalue of $\mathbb{T}$-matrices is invariably $\lambda_1 = 1$, while from the reality of such matrices it follows by the Perron-Frobenius theorem that for the other eigenvalues we have $1 > |\lambda_i| > 0 : i = 2, 3, \ldots, d$. It follows therefore from (7) that

$$\lim_{n\to\infty} \mathbb{T}^n = \mathbb{P}_1 = |a_1)(A^1| \tag{8}$$

which has this important consequence: all initial $|x)$ are sent asymptotically to the multiples of $|a_1)$; all stochastic states $\mathbf{p}_0$ evolve to a $\mathbb{T}$-dependent stochastic state $\mathbf{p}_\infty$ that retains no recollection of where it came from.

Markov interpolation & continuous Markov processes. If all λ_i are non-zero (*i.e.*, if $\mathbb{M}$ is nonsingular) we can by (4) introduce

$$\mathbb{L} = \sum_i (\log \lambda_i)\, \mathbb{P}_i \tag{9}$$

and obtain[2]

$$\mathbb{M} = \exp(\mathbb{L})$$

whence $\mathbb{M}^n = \exp(n\mathbb{L})$ or—reverting to $\mathbb{T}$-notation underscore the fact that we will henceforth have interest in the preceding material only as it relates to Markov processes—

$$\mathbb{T}^n = \exp(n\mathbb{L}) \tag{10}$$

It might, on this basis, seem natural to propose

$$\mathbb{T}(t) = \exp(t\mathbb{L}) \quad : \quad t \text{ a continuous parameter, called "time"} \tag{11}$$

as a t-parameterized family of Markov matrices that *interpolates between* the elements of the discrete set $\{\mathbb{T}^0, \mathbb{T}^1, \mathbb{T}^2, \mathbb{T}^3, \ldots\}$. That proposal is, however, marred by a fatal flaw: the eigenvalues (and therefore also the associated eigenvectors) of unbalanced $\mathbb{T}$-matrices are frequently *complex*, and the eigenvalues of balanced $\mathbb{T}$-matrices are—though always real—frequently *negative*. In both cases, $\log \lambda$ is complex (and multivalued). So T(t) is complex except at $t = 0, \pm 1, \pm 2, \ldots$, though simulations[3] indicate that its columns still sum to $1 + 0i$. The elements of

$$\mathbf{p}_t = \mathbb{T}(t)\, \mathbf{p}_0$$

are found similarly to sum to $1 + 0i$, but $\mathbf{p}_t$ when complex cannot possibly refer

[2] From (9), we note in passing, it follows by

$$\begin{aligned} \operatorname{tr} \mathbb{P}_i &= \text{dimension } d_i \text{ of the subpace onto which } \mathbb{P}_i \text{ projects} \\ &= \text{degeneracy } d_i \text{ of the eigenvalue } \lambda_i \end{aligned}$$

that

$$\operatorname{tr} \mathbb{L} = \sum_i d_i (\log \lambda_i) = \prod_i (\lambda_i)^{d_i} = \det \mathbb{M}$$

which gives the elegant (but nameless) equation

$$\det \mathbb{M} = \operatorname{tr} \log \mathbb{M}$$

The usual derivation of this equation proceeds from the strong assumption that $\mathbb{M}$ can be brought by similarity transformation to diagonal form: $\mathbb{M} = \mathbb{S}^{-1} \mathbb{D}\, \mathbb{S}$.

[3] The *Mathematica* command `MatrixPower[ M, n]` works even when `n` is not an integer. It produces results that conform to the principle

`MatrixPower[ M, p].MatrixPower[ M, q] = MatrixPower[ M, p+q]`

for all real or complex $\mathbb{M}$, `p` and `q`.

to the stochastic state of a system. Evidently, the interpolation proposal (11) fails except in cases were all eigenvalues of $\mathbb{T}$ are positive (or so it might appear; be prepared, however, for a surprise!).

Differentially,

$$\mathsf{p}_t = \mathbb{T}(t)\,\mathsf{p}_0 = \exp(t\mathbb{K})\,\mathsf{p}_0$$

(the reason for the notational adjustment $\mathbb{L} \to \mathbb{K}$ will become clear in a moment) becomes

$$\mathsf{p}_{t+\tau} = \{\mathbb{I} + \tau\mathbb{K}\}\,\mathsf{p}_t$$

If p_t is stochastic then $\mathsf{p}_{t+\tau}$ will also be if and only if $\mathbb{I}+\tau\mathbb{K}$ is Markovian, which requires that $\mathbb{K}$ be a matrix the off-diagonal elements of which are non-negative and the columns of which sum to zero. Such matrices are called "Kirchhoff matrices" because they are central to S. Katutani's random walk reconstruction of Kirchhoff's laws.[4] With elements that are typically real numbers instead of integers, they generalize the essential structure of the "Laplace matrices" that enter into the classical/quantum theory of walks on graphs, which are in turn closely related to the "adjacency matrices" that describe the structure of graphs. The can be constructed by adding to a matrix of the "invertebrate" form

$$\mathbb{A} = \begin{pmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & 0 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & 0 & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 0 \end{pmatrix} \quad : \quad \text{non-negative real elements}$$

the diagonal matrix

$$\mathbb{D} = \begin{pmatrix} -A_1 & 0 & 0 & \cdots & 0 \\ 0 & -A_2 & 0 & \cdots & 0 \\ 0 & 0 & -A_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -A_n \end{pmatrix} \quad : \quad A_k = \textstyle\sum_j a_{jk}$$

The resulting Kirchhoff matrix

$$\mathbb{K} = \mathbb{D} + \mathbb{A}$$

has columns that sum to zero, and will be symmetric or not according as $\mathbb{A}$ is.

[4] "Markov processes and the Direchlet problem," Proc. Jap. Acad. **21**, 227–233 (1945). Katutani's fundamental work, as it applies specifically to circuit theory, was developed by J. A. Nash-Williams, "Random walk and electric currents in networks," Proc. Camb. Phil. Soc. **55**, 181–194 (1959) and is discussed in elaborate detail by Peter G. Doyle & J. Laurie Snell in *Random Walks and Electric Networks*, originally published as Carus Mathematical Monograph No. 22 by the American Mathematical Association and now available on the web as a free pdf download.

We note that if $\mathbb{K}$ is a Kirchhoff matrix then so is every non-negative real scalar multiple $t\mathbb{K}$ ($t \geqslant 0$) of $\mathbb{K}$. It follows that every matrix of which $t\mathbb{K}$ is the logarithm

$$\mathbb{T}(t) = \exp(t\mathbb{K})$$

—this can be construed to be the solution of

$$\tfrac{d}{dt}\mathbb{T}(t) = \mathbb{K}\mathbb{T}(t) \quad : \quad \mathbb{T}(0) = \mathbb{I}$$

—is Markovian. We conclude that *only when* $\log \mathbb{T}$ *is a Kirchhoff matrix* are we able to interpolate between the successive integral powers of $\mathbb{T} = \mathbb{T}(1)$; only in such cases can we speak of "continuous Markov processes." The resulting Markov matrices $\mathbb{T}(t)$ will balanced or unbalanced according as $\mathbb{K}$ is or is not symmetric.

A series of simulations has established—surprisingly—that in unbalanced cases the eigenvalues (ditto the eigenvectors) of $\mathbb{T} = \exp \mathbb{K}$ (and, more generally, of $\mathbb{T}(t) = \exp(t\mathbb{K})$) occur in conjugate complex pairs, though $\mathbb{T}(t)$ is in all cases real Markovian.[5] Balanced cases present no such mystery.

Graphic display of the unbalanced evolution of $\mathbf{p}_t$ reveals smooth curves that link elements of $\mathbf{p}_0$ in a one-to-one way to elements of $\mathbf{p}_\infty$. The linkage is, however, not invariably order-preserving; the curves sometimes cross. In balanced cases all curves converge to the same point (the elements of $\mathbf{p}_\infty$ are in such cases identical).

Irreversibility. There are $n!$ ways to arrange for every row/column of an $n \times n$ matrix to contain a solitary 1, all other elements being 0. The determinants of such "permutation matrices" $\mathbb{P}$ are ± 1 according as the permutational action of $\mathbb{P}$ on the elements of a vector $\mathbf{p}$ is even or odd. All such matrices are invertible and their inverses are again permutation matrices, as are all products of permutation matrices and (trivially) the identity matrix $\mathbb{I}$. All permutation matrices are Markovian, and the set of such all such matrices (unlike the set of all Markov matrices) possesses the group property. The subgroup structure of the permutation group is a rich field into which I need not enter.

Nothing precludes the possibility that Markov matrix might be singular. This simple unbalanced example

$$\mathbb{S} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}$$

serves to make the point; it is manifestly Markovian, manifestly singular. Its eigenvalues are $\{1, \frac{1}{2}, 0\}$ and its unitized leading eigenvector is $\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$. It sends

$$\mathbf{p}_0 \longrightarrow \mathbb{S}^\infty \mathbf{p}_0 = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} \quad : \quad \text{all } \mathbf{p}_0$$

[5] I do not linger to consider how this is possible.

in the asymptotic limit. All of which is standard Markovian stuff. Note, however, that because $\mathbb{S}$ is singular it is impossible to contemplate inversion of the process $\mathbf{p}_0 \longrightarrow \mathbf{p}_1 = \mathbb{S}\,\mathbf{p}_0$.

Look now to the manifestly *non*-singular example

$$\mathbb{N} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 1 & 0 \end{pmatrix}$$

The eigenvalues of $\mathbb{N}$ are $\{1, -\frac{1}{4}(\sqrt{5}+1), \frac{1}{4}(\sqrt{5}-1)\}$ and its unitized leading eigenvector is $\{\frac{2}{5}, \frac{1}{5}, \frac{2}{5}\}$. It sends

$$\mathbf{p}_0 \longrightarrow \mathbb{N}^\infty \mathbf{p}_0 = \begin{pmatrix} \frac{2}{5} \\ \frac{1}{5} \\ \frac{2}{5} \end{pmatrix} \quad : \quad \text{all } \mathbf{p}_0$$

—standard Markovian stuff again—and because $\mathbb{N}$ is non-singular it is possible at least to *contemplate* inversion of the process $\mathbf{p}_0 \longrightarrow \mathbf{p}_1 = \mathbb{N}\,\mathbf{p}_0$. But $\mathbb{N}^{-1}$, though it exists, is manifestly *non-Markovian*:

$$\mathbb{N}^{-1} = \begin{pmatrix} 2 & -2 & 0 \\ -1 & 1 & 1 \\ 0 & 2 & 0 \end{pmatrix}$$

contains elements that fall outside the interval $(0, 1)$ (though its columns do sum to unity). And that is the general situation: **the inverses of non-singular Markov matrices $\mathbb{T}$ are generally non-Markovian**; they contain elements that—either because of sign or magnitude—cannot be interpreted as transition probabilities (their columns generally sum to unity, but have not the structure of stochastic vectors). Permutation matrices are exceptions to the rule (which is the reason I brought them into the discussion). But generally speaking—if permutations are set aside (see below)—we can say that...

Markov processes proceed deterministically to asymptotic steady states, from which there is no Markovian means of retreat, no Markovian way to recover the initial state. In this respect they resemble the many mixing processes encountered in Nature, which lead to homogeneously mixed states that cannot by any natural process be unmixed.

DIGRESSION: SOME PROPERTIES OF PERMUTATION MATRICES

Consider the permutation matrices

$$\mathbb{C}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbb{C}_3 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbb{C}_4 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \ldots$$

which, as non-negative (but not positive) matrices, lie somewhat beyond the reach of the Perrod-Frobenius theorem. All powers of such matrices are again permutation matrices. They are "cyclic" (describe cyclic permutations) in the

sense that

$$(\mathbb{C}_n)^p = (\mathbb{C}_n)^{p+n} \quad : \quad p = 0, 1, 2, 3, \ldots$$

from which it follows that

$$(\mathbb{C}_n)^n = \mathbb{I}, \quad \mathbb{C}_n^{-1} = (\mathbb{C}_n)^{n-1}$$

The eigenvalues of $\mathbb{C}_n$ are the complex n^{th} roots of unity

$$e^{2\pi i(k/n)} \quad : \quad k = 0, 1, 2, \ldots, (n-1)$$

which equi-partition the unit circle (one point invariably at $1 + 0i$); as $\mathbb{C}_n$ is raised to successively higher powers p those points chase each other around the circle, until at $p = n$ they have completed cycle. All eigenvalues have unit amplitude: none dies or blows up when raised to powers. Which is to say: $(\mathbb{C}_n)^p \mathbf{p}$ executes an endless series of cycles—does not approach a stable asymptote as $p \to \infty$.

Every $n \times n$ permutation matrix $\mathbb{P}$ can (by rearranging rows and columns) be brought to the canonical form

$$\mathbb{P} = \begin{pmatrix} \mathbb{C}_a & & & \\ & \mathbb{C}_b & & \\ & & \ddots & \\ & & & \mathbb{C}_z \end{pmatrix} \quad : \quad a + b + \cdots + z = n$$

and is cyclic with period ν

$$\mathbb{P}^{\nu} = \mathbb{P}, \text{ therefore } \mathbb{P}^{-1} = \mathbb{P}^{\nu-1}$$

where ν is the LCM of $\{a, b, \ldots, z\}$. To discover the value of $\nu_{\max}$ for given n, use `Partitions[n]` to list the partitions of n, use `DeleteDuplicates[ ]` to remove co-cyclic redundancies from the individual partitions, use `Times@@[ ]` to multiply the elements of each such non-redundant list, and search for the maximum among the list of numbers thus produced. In the case $n = 10$ one is led thus from $10 = 2+3+5$ to $\nu_{\max} = 30 \ll 10! = 3,628,800$. It is, I infer (this must be a classic result), generally impossible to exhibit a single $\mathbb{P}$ that cycles through all $n!$ permutations (though in the case $n = 2$ this is trivial).

REMARK: The periodicity of $\mathbb{P}$-matrices would appear to complicated the practice of card shuffling. Perhaps that is why the practice proceeds

$$\mathbf{p}_0 \to \mathbf{p} = \cdots \mathbb{P}_{\text{cut}} \mathbb{P}_{\text{suffle}} \mathbb{P}_{\text{cut}} \mathbb{P}_{\text{suffle}} \mathbf{p}_0$$

It would be interesting to study the degree of randomization that can be produced by such a procedure, but that would take me too far afield. I note, however, that balanced (non-permutational!) Markovian procedures achieve asymptotically perfect randomization (and that they disallow "unshuffling").
End of digression.

Entropic aspects of Markov processes. The entropy function

$$S(\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}) \equiv -\sum_{k=1}^{n} p_k \log p_k = \big\langle \log(1/p_k) \big\rangle$$

is used both in statistics and in physics (information theory) to provide a measure of the "randomness" of the elements of the stochastic vector $\mathbf{p}$. It assumes the minimal value 0 when all but one of the elements of $\mathbf{p}$ are 0, and the maximal value $\log n$ when all elements of $\mathbf{p}$ are equal (to $1/n$). The expression $-0 \log 0$ is (and is considered by *Mathematica* to be) indeterminate, but *Mathematica* knows that $\lim_{p \downarrow 0}(p \log p) = 0$. In computational work I make use of the function[6]

$$\mathrm{ent}(x) = \begin{cases} 0 & : \quad x = 0 \\ -x \log x & : \quad 0 < x \leqslant 1 \end{cases}$$

to avoid the disruption of awkward limiting procedures.

We begin by noting that multiplying $\mathbf{p}$ by $\mathbb{P}$—once or many times, or indeed by any sequence of permutation matrices—serves only to rearrange but not to alter the collective components of $\mathbf{p}$, so cannot change the value of $S(\mathbf{p})$. Entropy is a permutational invariant.[7] To illustrate the point, look to the case

$$\mathbb{P} = \mathbb{C}_2 \otimes \mathbb{C}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

which sends

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} \to \begin{pmatrix} f \\ d \\ e \\ c \\ a \\ b \end{pmatrix} \to \begin{pmatrix} b \\ c \\ a \\ e \\ f \\ d \end{pmatrix} \to \begin{pmatrix} d \\ e \\ f \\ a \\ b \\ c \end{pmatrix} \to \begin{pmatrix} c \\ a \\ b \\ f \\ d \\ e \end{pmatrix} \to \begin{pmatrix} e \\ f \\ d \\ b \\ c \\ a \end{pmatrix} \to \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} \to \cdots$$

[6] Entered by the command

```
ent[x_]:=Piecewise[{{0,x==0},{-xLog[x],0<x≤1}}]
```

[7] One must therefore use some other measure to quantify the randomizing effectiveness of various card shuffling procedures. Peri Diaconis, the Stanford magician/mathemaician, used "total variation distance" (one of several "statistical distance metrics") to establish the superiority of "riffling." This and related subjects are written about in elaborate detail in David A. Levin, Yuval Peres & Elizabeth L. Wilmer, *Markov Chains & Mixing Times* , a 364-page monograph that is available on the web as a free download. It is reported that in a well-shuffled deck of cards one can expect to guess an average of four cards, a claim that invites test by simulation.

each if which has entropy

$$S = -a\log a - b\log b - c\log c - d\log d - e\log e - f\log f$$

We have

$$\text{period}(\mathbb{P}) = \text{period}(\mathbb{C}_2)\cdot\text{period}(\mathbb{C}_3) = 2\cdot 3 = 6$$

—as illustrated.

> **NOTE**: *At this point I broke off (i) to construct figures showing the entropy evolution produced by various balanced/unbalanced discrete/continuous Markov processes—they reside in a notebook entitled* Lipsky Entropy Figures *and of course resemble some of the figures in Lipsky's paper—and (ii) to work out an idea that popped into my Markov-saturated head during a brunch hosted by the Turkish parents of a precocious young Reed student (November 20); that idea is developed in "Fractional Permutations." I turn now—finally—to Lipsky's paper, to an effort to comprehend what he sought and claims to have accomplished.*

The Poisson distribution, but not quite. The following remarks were inspired by Lipsky's "§3.2 Markov Assumption." I begin with an account of the assumptions that give rise to the Poisson distribution, which Lipsky's line of argument very closely resembles.[8] We posit a situation in which statistically independent events ("clicks") occur so sparcely that during any sufficiently brief interval of time Δt the probability that a click will occur can be described $\mu\Delta t$ and there is no probability that two (or more) clicks will occur. Let

$$P_0(t) = \left\{\begin{array}{l}\text{probability that there is NO click}\\ \text{during the (finite) interval } (0,t)\end{array}\right.$$

Then

$$P_0(t+\Delta t) = P_0(t)\cdot(1-\mu\Delta t)$$

gives

$$\frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\mu P_0(t)$$

In the limit $\Delta t \to 0$ we therefore have

$$\tfrac{d}{dt}P_0(t) = -\mu P_0(t) \quad\Longrightarrow\quad P_0(t) = ke^{-\mu t} \tag{12}$$

Now let

$$P_1(t) = \left\{\begin{array}{l}\text{probability that precisely ONE click}\\ \text{occurred the interval } (0,t)\end{array}\right.$$

$$P_n(t) = \left\{\begin{array}{l}\text{probability that precisely } n \text{ clicks}\\ \text{occurred the interval } (0,t)\end{array}\right.$$

[8] I follow the discussion that appears on pages 27–29 of Chapter I in my *Statistical Physics & Thermodynamics* (1969–70 and 1971–72), which follows closely the material that appears on pages 110–125 of W. Feller, *An Introduction to Probability Theory and its Applications* (1950)

Pretty clearly[9]

$$\begin{aligned}
P_1(t) &= \int_0^t P_0(s)\mu ds P_0(t-s) \\
&= k^2\mu \int_0^t e^{-\mu s} e^{-\mu(t-s)} ds \\
&= k^2 \mu t\, e^{-\mu t} \\
P_2(t) &= \iint_{0<s_1<s_2<t} P_0(s_1)\mu ds_1 P_0(s_1-s_2)\mu ds_2 P_0(t-s_2) \\
&= k^3\mu^2 e^{-\mu t} \iint_{0<s_1<s_2<t} ds_1 ds_2 \\
&= k^3 \tfrac{1}{2}(\mu t)^2\, e^{-\mu t} \\
&\vdots \\
P_n(t) &= k^{n+1} \tfrac{1}{n!}(\mu t)^n\, e^{-\mu t}
\end{aligned}$$

From the requirement

$$\sum_{n=0}^{\infty} P_n(t) = ke^{-\mu t} \sum_{n=0}^{\infty} \tfrac{1}{n!}(k\mu t)^n = ke^{(k-1)\mu t} = 1 \quad : \quad \text{all } t$$

we obtain $k = 1$, whence finally the Poisson distribution

$$P_n(t) = \tfrac{1}{n!}(\mu t)^n\, e^{-\mu t} \tag{13}$$

which S. Poisson (1837) obtained as a discrete limiting case[8] of the binomial distribution.[10] The classic manifestation: given a radioactive sample, record the number of Geiger clicks suring each of a large number of t-second intervals. The data can be expected to conform (to with in experimental error) to (13) for some best estimate of μ, from which can recover the half-life of the radioactive atoms in question.

Lipsky has interest in a truncated version of the problem discussed above, one which involves only the function $P_0(t)$. To the argument that gave (12) he adjoins the initial condition $P_0(0) = 1$, and instead of speaking of the "probability

[9] The only tricky part of the argument lies in the observation that

$$\iint \cdots \int_{0<s_1<s_2<\cdots<s_n<t} ds_1 ds_2 \cdots ds_n = \tfrac{1}{n!} t^n$$

[10] The binomial distribution describes probable winnings after N Bernoulli trial flips of a loaded coin, probable location after N random steps with unequal left/right preferences. By another limiting process it gives rise to the continuous Gaussian (normal) distribution.

that no click has occurred" he speaks

$$e^{-\mu\, t} = \text{"probability that the particle } \textit{has not departed} \text{ from state } s_1\text{"}$$

Let

A signify the event "no departure during the interval $(0, t)$

B signify the event "no departure during the interval (t, τ)

We have P(A and B) = P(B given A)·P(A) whence

$$\text{P(B given A)} = \frac{\text{P(A and B)}}{\text{P(A)}} = \frac{e^{-\mu(t+\tau)}}{e^{-\mu t}} = e^{-\mu\tau}$$

It is the t-independence of this result that leads Lipsky to attach Markov's name to it, though it seems too primitive to be called a "Markov process." And it was not a "Markov assumption" that led him to it, but a "Poisson assumption." He remarks (following Maxwell) that the functional equation

$$f(x+y) = f(x)\cdot f(y) \quad \Longrightarrow \quad f(x) = e^{ax}$$

A simple master equation. Lipsky now (in his "§3.3 One Particle & Two States") posits that when his particle (under control of μ_1) disappears from state s_1 it promptly reappears in state s_2, from which (under control of μ_2) it can disappear only to reappear in state s_1. He looks to the dynamics of this duplex process; *i.e.*, to the evolution of the stochastic vector

$$\mathbf{p}(t) = \begin{pmatrix} p_1(t) \\ p_2(t) \end{pmatrix}$$

that describes the probability that the particle is in state s_i at time t.

Evidently

$$\mathbf{p}(t+dt) = \begin{pmatrix} 1-\mu_1 dt & \mu_2 dt \\ \mu_1 dt & 1-\mu_2 dt \end{pmatrix} \mathbf{p}(t) \tag{14}$$

which gives

$$\tfrac{d}{dt}\mathbf{p}(t) = \mathbb{K}\mathbf{p}(t) \quad : \quad \mathbb{K} = \begin{pmatrix} -\mu_1 & +\mu_2 \\ +\mu_1 & -\mu_2 \end{pmatrix} \tag{15}$$

where $\mathbb{K}$ (whose columns sum to zero) is a t-independent Kirchhoff matrix, the generator of the Markov matrix (elements are transition probabilities, columns sum to unity)

$$\mathbb{M}(t) = e^{t\mathbb{K}} = \frac{1}{\mu}\begin{pmatrix} \mu_2 + \mu_1 e^{-\mu t} & \mu_2(1-e^{-\mu t}) \\ \mu_1(1-e^{-\mu t}) & \mu_1 + \mu_2 e^{-\mu t} \end{pmatrix} \quad : \quad \mu = \mu_1 + \mu_2 \tag{16}$$

The implication is that $\mathbf{p}(t)$ evolves by a t-parameterized continuous Markov process

$$\mathbf{p}(t) = \mathbb{M}(t)\,\mathbf{p}(0) \tag{17}$$

Asymptotically we have

$$\mathbf{p}(\infty) = \mu^{-1} \begin{pmatrix} \mu_2 & \mu_2 \\ \mu_1 & \mu_1 \end{pmatrix} \mathbf{p}(0) = \begin{pmatrix} \mu_2/\mu \\ \mu_1/\mu \end{pmatrix} \quad : \quad \text{all } \mathbf{p}(0) \tag{18}$$

Lipsky remarks that $\mathbb{K}\mathbf{p}(\infty) = \mathbf{0}$, which is an expression simply of the stationarity of the asymptotic distribution.

Equation (14) derives from the simple proposition that

$$\text{differential growth} = \text{differential gain} - \text{differential loss}$$

so (14) is in effect the "master equation" for Lipsky's duplex process.

I used the *Mathematica* command `MatrixExp` to obtain the evaluation (16) of $e^{t\mathbb{K}}$. Lipsky remarks that ((16) can be obtained by appeal to the "spectral decomposition theorem." Since $\mathbb{K}$ is not symmetric, and its eigenvectors

$$\begin{pmatrix} \mu_2 \\ \mu_1 \end{pmatrix} \text{ and } \begin{pmatrix} -1 \\ +1 \end{pmatrix}, \text{ associated with eigenvalues 0 and } -\mu, \text{ respectively}$$

are not orthogonal, one is obliged actually to appeal to the *generalized* spectral decomposition, which—since he distinguishes between left and right eigenvectors —is evidently what he has in mind. Since the generalized decomposition is, in my experience, non-standard it would be interesting to know from what uncited source he learned of it.

For at least 350 years it has been conventional to assume that mathematical operators (think of $\frac{d}{dx}$) operate to the right, and that is a convention that for 17% of that time I have honored in my own work. But somewhere along the line Lipsky has acquired the habit of thinking of vectors preferentially as *row* vectors, with the consequence that his (transposed) matrices operate to the left, like the verbs that stand at the end of German sentences. I should look into his book, and work that he cites, to discover whether that is the general convention among queueing theorists (and engineers? I think not)... perhaps because clients with their demands stand patiently at the end of the que?

Relationship between theory and observation. In his "§4 Interpretations of $\mathbf{p}(t)$" makes remarks concerning the design of observational procedures that might be used to confirm theoretical assertions about the evolved stochastic vector $\mathbf{p}(t)$. He labors, however, subject to this handicap: the Poisson theory purports to model systems of which one can present observable physical instances (samples of radioactive material); one collects data, and shows that it conforms (to within statistically determined limits) to a Poisson distribution with some optimally selected μ-value. But Lipsky can point to no specific physical system that behaves in the manner his idealized "one particle two state" theory prescribes. This circumstance lends to his remarks a certain disembodied abstract quality. Lipsky—proceding in the steps of Maxwell and Boltzmann—proposes two alternative procedures:

FIRST PROCEDURE Look to an N-particle ensemble of particles, all initially in state s_1, let them jump up/down for a time $T \gg t_r = 1/\mu$,[11] then record the number n_1 in state s_1, the number n_2 in state s_2. From

$$\begin{pmatrix} n_1/(n_1+n_2) \\ n_2/(n_1+n_2) \end{pmatrix} = \begin{pmatrix} \mu_2/(\mu_1+\mu_2) \\ \mu_1/(\mu_1+\mu_2) \end{pmatrix}$$

one obtains the estimate

$$\frac{\mu_1}{\mu_2} = \frac{n_2}{n_1} \quad : \quad \text{slightly different for every long run}$$

but the individual values of μ_1 and μ_2 remain undetermined.

SECOND PROCEDURE Sample the state of a single particle at times

$$t = \{T, 2T, 3T, \ldots, NT\}$$

and record the number n_1 of times it was found to be in state s_1, the number n_2 of times it was found to be in state s_2. The particle begins successive episodes in a variety of states, but by the time it is again observed its initial state has (by $T \gg t_r$) been forgotten. The resulting data is processed as before, and leads to the same (limited) conclusion.

Lipsky attributes the equivalence of the two procedures—informally (since no actual averaging is going on): ensemble average = time average) to "the ergodic theorem." Actually, such "ergodic theorems" as exist (such as the one devised by George Birkoff and John von Neumann in the early 1930s) pertain only to systems of specialized types, and hinge on restrictive assumptions. Ergodic theory began with Boltzmann's attempt to understand how reversible dynamics can give rise to irreversible physics (the H-theorem), and in a broadened has since spread into many parts of mathematics. As encountered by Lipsky (and in the theory of Markov processes generally, where loss of memory and irreversibility are characteristic) the justification of the ergodic hypothesis poses no great mystery.

The charge that "Lipsky can point to no specific physical system that behaves in the manner his idealized 'one particle two state' theory prescribes" is, I have now to admit, somewhat unfair. For he can point to *simulations* of such systems, which can be constructed by methods borrowed from the simulation of random walks. Imagine a situation in which a walker's next-step decision is determined by the flip of a loaded coin. HEADS mean "if on s_1 step to s_2, but if already on s_2 stay there." TAILS mean "if on s_2 step to s_1, but if already on s_1 stay there." Let his state (location) at any instant be represented (as

[11] t_r is the "relaxation" or "equilibration" time by which the exponentials in (16) have effectively died, and all memory of the initial state has been lost; Lipsky's assumption that all particles were initially in the same state is therefore superfluous.

previously) by

$$s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{else} \quad s_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The actions dictated by the loaded coin are accomplished by

$$\mathbb{H} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad \mathbb{T} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

both of which are manifestly Markovian (and singular) The simulate a sequence of 50 flips (assuming HEADS to be twice as likely as TAILS) command

```
flipsequence=RandomChoice[{2/3, 1/3} → {ℍ, 𝕋},50];
```

The command `netaction=Apply[Dot,flipsequence]` constructs the product of that sequence of matrices, and

```
finalstate=netaction.RandomChoice[{1/2, 1/2} → {s1, s2}]
```

constructs the state that results when `netaction` acts upon a randomly selected initial state. The following command

```
Tally[Sort[Table[Apply[Dot,RandomChoice[{2/3, 1/3} → {ℍ, 𝕋},50].
RandomChoice[{1/2, 1/2} → {s1, s2}]//MatrixForm,{k,1,100}]]]
```

(which runs in less than one second) conflates those actions, repeats the process 100 times, and tallys the results, producing a final result of which this

$$\left\{ \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, 40 \right\}, \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, 60 \right\} \right\}$$

is typical. Which is, however, a hard way to do a simple thing. For from

$$\mathbb{H}\,\mathbb{H} = \mathbb{H}\,\mathbb{T} = \mathbb{H}$$
$$\mathbb{T}\,\mathbb{T} = \mathbb{T}\,\mathbb{H} = \mathbb{T}$$

we see that

$$\mathbb{H} \cdot (\text{any product}) = \mathbb{H}$$
$$\mathbb{T} \cdot (\text{any product}) = \mathbb{T}$$

In short: the effect of sequenced flips is entirely determined by the *last* flip. So we might more simply have looked only to the randomized last flip, proceeding

```
Tally[Sort[Table[RandomChoice[{2/3, 1/3} → {ℍ, 𝕋}].
RandomChoice[{1/2, 1/2} → {s1, s2}]//MatrixForm,{k,1,100}]]]
```

Or still more simply

```
Tally[Sort[Table[
RandomChoice[{2/3, 1/3} → {s1, s2}]//MatrixForm,{k,1,100}]]]
```

We conclude that "Lipsky walks" are so simple that their simulation teaches us nothing that we did not already know.[12] Note that the simulations discussed above serve equally well to model *either* of Lipsky's procedures.

The simplest parallel process. In his "§5 Two Particles & Two States" Lipsky looks to the simultaneous description of two independent copies of his "One Particle & Two States" scenario; *i.e.*, to "duplex" the argument developed on pages 14–15. I break away from Lipsky's paper to indicate how I would approach that problem.

DIGRESSION: **The most natural way to duplex.** Let the stochastic vectors

$$\mathbf{p}_1 = \begin{pmatrix} p_{11} \\ p_{12} \end{pmatrix} \quad \text{and} \quad \mathbf{p}_2 = \begin{pmatrix} p_{21} \\ p_{22} \end{pmatrix}$$

refer to the distinguished particles #1 and #2, respectively, and let

$$\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2 = \begin{pmatrix} p_{11}p_{21} \\ p_{11}p_{22} \\ p_{12}p_{21} \\ p_{12}p_{22} \end{pmatrix} \equiv \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}$$

(which is stochastic if $\mathbf{p}_1$ and $\mathbf{p}_2$ are) refer to the composite pair. We make use of this property of the Kronecker product:[13]

$$(\mathbb{A} \otimes \mathbb{B})(\mathbb{C} \otimes \mathbb{D}) = \mathbb{A}\mathbb{C} \otimes \mathbb{B}\mathbb{D}$$

whenever the matrices are of such dimensions as to cause all of the products to make sense (= be defined).

Working from

$$\mathbf{p}(t+dt) = \mathbf{p}_1(t+dt) \otimes \mathbf{p}_2(t+dt)$$

we from (14) have

$$\begin{aligned} \mathbf{p}_1(t+dt) &= (\mathbb{I} + \mathbb{K}_1 dt)\,\mathbf{p}_1(t) \quad : \quad \mathbb{K}_1 = \begin{pmatrix} -\mu_1 & +\mu_2 \\ +\mu_1 & -\mu_2 \end{pmatrix} \\ \mathbf{p}_2(t+dt) &= (\mathbb{I} + \mathbb{K}_2 dt)\,\mathbf{p}_1(t) \quad : \quad \mathbb{K}_2 = \begin{pmatrix} -\nu_1 & +\nu_2 \\ +\nu_1 & -\nu_2 \end{pmatrix} \end{aligned}$$

[12] This in high contrast to the "Golden Walk" and to other "Pisot walks with shrinking steps" illustrated in "On some Borwein-inspired properties of random walks with shrinking steps" (November 2016), or the Lévy walks on the plane or random walks on graphene that I have illustrated in "Some miscellaneous adventures in experimental mathematical physics," notes (in the form of a *Mathematica* slide show) for a Reed College Physics Seminar (9 November 2011).

[13] A comprehensive list of such properties can be found on page 24 of Chapter I in my Quantum Mechanics (2000).

where—because it costs us nothing—I have for the moment assumed that the parameters $\{\nu_1, \nu_2\}$ that regulate the hopping of particle #2 are distinct from those $\{\mu_1, \mu_2\}$ that regulate the activity of #1. In first order (*i.e.*, neglecting the dt^2 term) we have

$$\begin{aligned}\mathbf{p}(t+dt) &= [(\mathbb{I}+\mathbb{K}_1 dt)\,\mathbf{p}_1(t)] \otimes [(\mathbb{I}+\mathbb{K}_2 dt)\,\mathbf{p}_2(t)] \\ &= [\mathbf{p}_1(t)\otimes\mathbf{p}_2(t)] + \big\{[\mathbb{K}_1\mathbf{p}_1(t)\otimes\mathbb{I}\mathbf{p}_2(t)] + [\mathbb{I}\mathbf{p}_1(t)\otimes\mathbb{K}_2\mathbf{p}_2(t)]\big\}dt \\ &= \mathbf{p}(t) + \big\{(\mathbb{K}_1\otimes\mathbb{I}) + (\mathbb{I}\otimes\mathbb{K}_2)\big\}\,\mathbf{p}_1(t)\,dt\end{aligned}$$

which gives

$$\tfrac{d}{dt}\,\mathbf{p}(t) = \mathbb{K}\mathbf{p}(t) \tag{19.1}$$

$$\mathbb{K} = (\mathbb{K}_1\otimes\mathbb{I}) + (\mathbb{I}\otimes\mathbb{K}_2) = \begin{pmatrix} -\mu_1-\nu_1 & \nu_2 & \mu_2 & 0 \\ \nu_1 & -\mu_1-\nu_2 & 0 & \mu_2 \\ \mu_1 & 0 & -\mu_2-\nu_1 & \nu_2 \\ 0 & \mu_1 & \nu_1 & -\mu_2-\nu_2 \end{pmatrix} \tag{19.2}$$

Note that $\mathbb{K}$ is a real asymmetric Kirchhoff matrix (generator of a Markov matrix): its off-diagonal elements are non-negative real numbers, and its columns sum to zero. The eigenvalues of $\mathbb{K}$ are[14]

$$\{\lambda_1,\lambda_2,\lambda_3,\lambda_4\} = \{0, -(\mu_1+\mu_2), -(\nu_1+\nu_2), -(\mu_1+\mu_2)-(\nu_1+\nu_2)\}$$

The generalized spectral representation of $\mathbb{K}$—worked out with the asssistance of *Mathematica*—reads

$$\mathbb{K} = \lambda_1\mathbb{P}_1 + \lambda_2\mathbb{P}_2 + \lambda_3\mathbb{P}_3 + \lambda_4\mathbb{P}_4 \tag{20}$$

with

$$\mathbb{P}_1 = \sigma^{-1}\begin{pmatrix} \mu_2\nu_2 & \mu_2\nu_2 & \mu_2\nu_2 & \mu_2\nu_2 \\ \mu_2\nu_1 & \mu_2\nu_1 & \mu_2\nu_1 & \mu_2\nu_1 \\ \mu_1\nu_2 & \mu_1\nu_2 & \mu_1\nu_2 & \mu_1\nu_2 \\ \mu_1\nu_1 & \mu_1\nu_1 & \mu_1\nu_1 & \mu_1\nu_1 \end{pmatrix} \quad : \quad \sigma \equiv (\mu_1+\mu_2)(\nu_1+\nu_2)$$

$$\mathbb{P}_2 = \sigma^{-1}\begin{pmatrix} \mu_1\nu_2 & \mu_1\nu_2 & -\mu_2\nu_2 & -\mu_2\nu_2 \\ \mu_1\nu_1 & \mu_1\nu_1 & -\mu_2\nu_1 & -\mu_2\nu_1 \\ -\mu_1\nu_2 & -\mu_1\nu_2 & \mu_2\nu_2 & \mu_2\nu_2 \\ -\mu_1\nu_1 & -\mu_1\nu_1 & \mu_2\nu_1 & \mu_2\nu_1 \end{pmatrix}$$

$$\mathbb{P}_3 = \sigma^{-1}\begin{pmatrix} \mu_2\nu_1 & -\mu_2\nu_2 & \nu_2\nu_1 & -\mu_2\nu_2 \\ -\mu_2\nu_1 & \mu_2\nu_2 & -\nu_2\nu_1 & \mu_2\nu_2 \\ \mu_1\nu_1 & -\mu_1\nu_2 & \mu_1\nu_1 & -\mu_1\nu_2 \\ -\mu_1\nu_1 & \mu_1\nu_2 & -\mu_1\nu_1 & \mu_1\nu_2 \end{pmatrix}$$

$$\mathbb{P}_4 = \sigma^{-1}\begin{pmatrix} \mu_1\nu_1 & -\mu_1\nu_2 & -\nu_2\nu_1 & \mu_2\nu_2 \\ -\mu_1\nu_1 & \mu_1\nu_2 & \nu_2\nu_1 & -\mu_2\nu_2 \\ -\mu_1\nu_1 & \mu_1\nu_2 & \mu_2\nu_1 & -\mu_2\nu_2 \\ \mu_1\nu_1 & -\mu_1\nu_2 & -\mu_2\nu_1 & \mu_2\nu_2 \end{pmatrix}$$

[14] Note in this connection that the eigenvalues of $\{\mathbb{K}_1, \mathbb{K}_2\}$ are $\{0, -(\mu_1+\mu_2)\}$ and $\{0, -(\nu_1+\nu_2)\}$, respectively.

Each of those matrices is, as it happens, asymmetric. The columns of $\{\mathbb{P}_2, \mathbb{P}_3, \mathbb{P}_4\}$ are seen to sum to zero. The columns of $\mathbb{P}_1$ sum to unity, but that term is killed by $\lambda_1 = 0$, so does not destroy the Kirchhoff property of $\mathbb{K}$.

If we set $\{\nu_1, \nu_2\} \to \{\mu_1, \mu_2\}$ the eigenvalues of $\mathbb{K}$ become $\{0, -\mu, -\mu, -2\mu\}$ (here $\mu \equiv \mu_1 + \mu_2$) and the $\mathbb{P}$-matrices become

$$\mathbb{Q}_1 = \sigma^{-1} \begin{pmatrix} \mu_2^2 & \mu_2^2 & \mu_2^2 & \mu_2^2 \\ \mu_1\mu_2 & \mu_1\mu_2 & \mu_1\mu_2 & \mu_1\mu_2 \\ \mu_1\mu_2 & \mu_1\mu_2 & \mu_1\mu_2 & \mu_1\mu_2 \\ \mu_1^2 & \mu_1^2 & \mu_1^2 & \mu_1^2 \end{pmatrix} \quad : \quad \sigma \equiv (\mu_1 + \mu_2)^2 = \mu^2$$

$$\mathbb{Q}_2 = \sigma^{-1} \begin{pmatrix} \mu_1\mu_2 & \mu_1\mu_2 & -\mu_2^2 & -\mu_2^2 \\ \mu_1^2 & \mu_1^2 & -\mu_1\mu_2 & -\mu_1\mu_2 \\ -\mu_1\mu_2 & -\mu_1\mu_2 & \mu_2^2 & \mu_2^2 \\ -\mu_1^2 & -\mu_1^2 & \mu_1\mu_2 & \mu_1\mu_2 \end{pmatrix}$$

$$\mathbb{Q}_3 = \sigma^{-1} \begin{pmatrix} \mu_1\mu_2 & -\mu_2^2 & \mu_1\mu_2 & -\mu_2^2 \\ -\mu_1\mu_2 & \mu_2^2 & -\mu_1\mu_2 & \mu_2^2 \\ \mu_1^2 & -\mu_1\mu_2 & \mu_1^2 & -\mu_1\mu_2 \\ -\mu_1^2 & \mu_1\mu_2 & -\mu_1^2 & \mu_1\mu_2 \end{pmatrix}$$

$$\mathbb{Q}_4 = \sigma^{-1} \begin{pmatrix} \mu_1^2 & -\mu_1\mu_2 & -\mu_1\mu_2 & \mu_2^2 \\ -\mu_1^2 & \mu_1\mu_2 & \mu_1\mu_2 & -\mu_2^2 \\ -\mu_1^2 & \mu_1\mu_2 & \mu_1\mu_2 & -\mu_2^2 \\ \mu_1^2 & -\mu_1\mu_2 & -\mu_1\mu_2 & \mu_2^2 \end{pmatrix}$$

to which similar remarks pertain.

Looking to the Markov matrix generated by $\mathbb{K}$, we have

$$\mathbb{M}(t) = e^{t\,\mathbb{K}} = \mathbb{P}_1 + e^{\lambda_2 t}\mathbb{P}_2 + e^{\lambda_3 t}\mathbb{P}_3 + e^{\lambda_4 t}\mathbb{P}_4 \tag{21}$$

From established properties of the $\mathbb{P}$-matrices it follows that at all times all columns of $\mathbb{M}(t)$ sum to unity. The $\{\lambda_2, \lambda_3, \lambda_4\}$ are all < 0, so

$$\lim_{t\to\infty} \mathbb{M}(t) = \mathbb{P}_1$$

The eigenvalues of $\mathbb{P}_1$ are $\{0, 0, 0, \sigma\}$. The associated eigenvectors (as presented by *Mathematica*)

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \quad \sigma^{-1} \begin{pmatrix} \mu_2\nu_2 \\ \mu_2\nu_1 \\ \mu_1\nu_2 \\ \mu_1\nu_1 \end{pmatrix} \equiv \mathbf{p}_\infty$$

all have negative elements (so cannot be stochastic vectors) except for the last (the Peron-Frobenius theorem again), which I have presented in the unitized form required of stochastic vectors. Every initial state

$$\mathbf{p}_0 \to \mathbf{p}_\infty \text{ asymptotically}$$

We have been looking to the evolution of

$$\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2 = \begin{pmatrix} p_{11}p_{21} \\ p_{11}p_{22} \\ p_{12}p_{21} \\ p_{12}p_{22} \end{pmatrix}$$

where

$$p_{1\alpha}p_{2\beta} = \text{probability that} \begin{cases} \#1 \text{ is in state } s_\alpha \\ \text{and } \#2 \text{ is in state } s_\beta \end{cases}$$

and should understand that the analysis hinges on two key assumptions:

- The particles are *distinguished* (wear identifying names);
- Their hopping activity is *uncorrelated* (statistically independent).

If we abandon the former assumption (*i.e.*, if we assume the particles to be *indistinguishable*) then we (*i*) can no longer entertain the possibility that $\{\nu_1, \nu_2\}$ and $\{\mu_1, \mu_2\}$ are distinct, and (*ii*) must *symmetrize the state vector*, writing

$$\mathbf{q} = \tfrac{1}{2}(\mathbf{p}_1 \otimes \mathbf{p}_2) + \tfrac{1}{2}(\mathbf{p}_2 \otimes \mathbf{p}_1) = \tfrac{1}{2} \begin{pmatrix} p_{11}p_{21} + p_{21}p_{11} \\ p_{11}p_{22} + p_{21}p_{12} \\ p_{12}p_{21} + p_{22}p_{11} \\ p_{12}p_{22} + p_{22}p_{12} \end{pmatrix} = \begin{pmatrix} q_1 \\ q \\ q \\ q_4 \end{pmatrix}$$

where

$$\begin{aligned} q_1 &= \text{probability that both are in state } s_1 \\ q = q_2 = q_3 &= \text{probability that they are in different states} \\ q_4 &= \text{probability that both are in state } s_2 \end{aligned}$$

Looking now to the dynamics of $\mathbf{q}$, we note that the elements of $\mathbf{p}_2 \otimes \mathbf{p}_1$ differ only by a (cyclic) permutation from those of $\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2$:

$$\mathbb{C}(\mathbf{p}_1 \otimes \mathbf{p}_2) = \mathbf{p}_2 \otimes \mathbf{p}_1 \quad : \quad \mathbb{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbb{C}^2 = \mathbb{I}$$

So from (19)—which by $\{\nu_1, \nu_2\} \to \{\mu_1, \mu_2\}$ has become

$$\tfrac{d}{dt}(\mathbf{p}_1 \otimes \mathbf{p}_2) = \mathbb{L}(\mathbf{p}_1 \otimes \mathbf{p}_2) \ : \ \mathbb{L} = \begin{pmatrix} -2\mu_1 & \mu_2 & \mu_2 & 0 \\ \mu_1 & -\mu_1 - \mu_2 & 0 & \mu_2 \\ \mu_1 & 0 & -\mu_1 - \mu_2 & \mu_2 \\ 0 & \mu_1 & \mu_1 & -2\mu_2 \end{pmatrix}$$

—we obtain (after multiplication by $\mathbb{C}$)

$$\tfrac{d}{dt}(\mathbf{p}_2 \otimes \mathbf{p}_1) = \mathbb{C}\mathbb{L}\mathbb{C}(\mathbf{p}_2 \otimes \mathbf{p}_1) = \mathbb{L}(\mathbf{p}_2 \otimes \mathbf{p}_1) \quad \text{by } \mathbb{C}\mathbb{L}\mathbb{C} = \mathbb{L}$$

Adding and dividing by 2, we obtain finally

$$\tfrac{d}{dt}\mathbf{q} = \mathbb{L}\mathbf{q}$$

Results already in hand (see again page 20) supply the manifestly Kirchhoffian matrix

$$\mathbb{L} = 0\,\mathbb{Q}_1 - \mu(\mathbb{Q}_2 + \mathbb{Q}_3) - 2\mu\,\mathbb{Q}_4$$

and we note that vectors of the form $\mathbf{q}$ (2^{nd} and 3^{rd} components equal) are sent by $\mathbb{L}$ into vectors of that same form. We are in position now to write

$$\begin{aligned}\mathbf{q}(t) &= \mathbb{M}(t)\,\mathbf{q}(0)\\ \mathbb{M}(t) &= e^{t\mathbb{L}} = \mathbb{Q}_1 + e^{-\mu t}(\mathbb{Q}_2 + \mathbb{Q}_3) + e^{-2\mu t}\mathbb{Q}_4\end{aligned}$$

which (for the reason noted just above) preserves the characteristic form of symmetrized stochastic $\mathbf{q}$-vectors. Finally, we have

$$\lim_{t\to\infty} \mathbb{M}(t)\,\mathbf{q}_0 = \mathbf{q}_\infty \equiv \mu^{-2}\begin{pmatrix}\mu_2^2\\ \mu_1\mu_2\\ \mu_1\mu_2\\ \mu_1^2\end{pmatrix} \quad : \quad \text{all } \mathbf{q}_0$$

CONCLUDING REMARK: In quantum contexts symmetrization speaks of bosons; fermions, on the other hand, entail antisymmetrization of the wave function, which suggests that classical interest might attach to antisymmetrized two-particle stochastic vectors

$$\mathbf{p}_1 \otimes \mathbf{p}_2 - \mathbf{p}_2 \otimes \mathbf{p}_1 = \begin{pmatrix}0\\ +(p_{11}p_{22} - p_{12}p_{21})\\ -(p_{11}p_{22} - p_{12}p_{21})\\ 0\end{pmatrix}$$

But such vectors cannot be stochastic: their elements are of opposite sign (therefore cannot both be positive) and sum to zero. For antisymmetrized Kronecker products of three 2-vectors the problem becomes even more vivid:

$$\begin{aligned}&\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c} - \mathbf{a}\otimes\mathbf{c}\otimes\mathbf{b}\\ +&\,\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{a} - \mathbf{b}\otimes\mathbf{a}\otimes\mathbf{c}\\ +&\,\mathbf{c}\otimes\mathbf{a}\otimes\mathbf{b} - \mathbf{c}\otimes\mathbf{b}\otimes\mathbf{a} = \mathbf{0},\ \text{invariably}\end{aligned}$$

which is an instance of the Jacobi identity.[15] The short of it: in quantum theory minus signs pose no problem, since probabilities arise from the modulus, but in classical stochastic contexts they are intolerable, even when they do not interfere with unitization. **End of digression.**

I return now to Lipsky's §5.

[15] For higher dimensional vectors one gets a non-zero vector the elements of wich sum to zero.