THE MARKOV MELODY ENGINE: GENERATING RANDOM MELODIES WITH TWO-STEP MARKOV CHAINS

DAVID STEINSALTZ AND DAVID WESSEL

1. Introduction: What is the Markov melody engine?

Stochastic algorithms have had a consistent, if somewhat disreputable, role in western musical composition at least since the 18th century. W. Mozart's *Musikalisches Würfelspiel* is perhaps the best known, but other composers, including J. Haydn and C. P. E. Bach dabbled in this domain. As the name suggests, this method of recombining carefully composed musical elements in random orders according to a throw of the dice, was seen as an amusing novelty, not as a serious compositional tool. Peter Welcker published in 1775, in London, a "Tabular System Whereby Any Person without the Least Knowledge of Musick May Compose Ten Thousand Different Minuets in the Most Pleasing and Correct Manner". (For a more extensive account of the history of stochastic music, see [Loy89] or [Pot71].)

In the twentieth century, the inclinations toward indeterminate notation on the one hand, and formalized (serial) compositional strategies on the other, came together in the application of formal probability theory to music. [Bri75][chapters 8 and 12]. This has gone in two directions: One, indeterminate composition, in which the precise choice of notes is turned over to dice, coins, or a computer pseudorandom number generator; two, compositions more or less formally inspired and structured by probability theory. The first mode has been advocated perhaps most prominently by John Cage, the second by Iannis Xenakis, who after some initial reticence plunged into the first as well.

Chance procedures can act on any element of music: pitch, timbre, rhythm, choice of musicians, time and location of the performance, choice of repertoire, etc. Most popular are random pitches, perhaps because of the superordinate role played by melody in Western music, and the early appearance of a tractable theory of musical pitches. The systems for generating random sequences of pitches have been generally of two types. [Find examples?] The simplest approach is pitch-centered, most commonly a Markov chain, where each pitch has a distribution given for its successors. Thus, if the range is thirteen chromatic notes of a single octave span, the model would be represented by $13^2 = 169$ numbers, each giving the probability that, say, $C\sharp$ is followed by another $C\sharp$, by D, or $E\flat$, or E, and so on. This sort of process generates melodies which are perfectly adequate for video games, but are on the whole drearily static. The reason is not hard to find: There is no sense of direction, no continuity, since every $C\sharp$ has the same consequences, no matter what preceded it. (Since each distribution has to add to 1, there are really only $13 \cdot 12 = 156$ parameters which specify the model.)

A bit more continuity is achieved by interval-based programs. Here one might define, say, 15 possible intervals, ranging in chromatic steps from a perfect fifth down, to a perfect fifth up, and give a stochastic matrix which defines the probability of one interval following another. For instance, it might be that a perfect fourth down is likely right after a halfstep up, but unlikely after another perfect fourth down. Such a model can have a minimal sense of direction, but all sense of key is lost. The melody will wander willy-nilly over the available pitches, without significantly emphasizing the notes of any scale.

We would like to combine these approaches, allowing some control over the pitch statistics, together with some memory for the direction. The obvious thing to do is to define a Markov chain whose states are several pitches in a row. For larger values of ell, this will allow quite good approximations to the statistics of genuine melodies, allowing us to distinguish between the likelihood of, for example, $C - D - E - F$ and $F - D - E - F$. On the other hand, as ℓ increases the complexity of the model grows exponentially. If there are n pitches, and the memory is ℓ notes long, then we need $n^\ell(n-1)$ numbers to specify the model. Furthermore, there is a danger of overspecifying. If we imitate order-10 probabilities from a fund of prior melodies, we are likely to end up largely imitating them piecewise note for note, while excluding some possibilities that were musically reasonable, but which happen to be absent from the data set.

Consider a two-step memory. If we have a range of 20 pitches — rather small — the model will require 8000 parameters. How can we coherently supply so many numbers? The approach we explore here is to select from all the possible models a relatively few which have certain musical features in common. One musical feature is the key, which is specified by a distribution on the pitches. Another is a stochastic matrix of the second type described above, indicating the succession of intervals. To simplify even further, we may reduce the intervals to a few classes; for example: big jump down, small jump down, no move, small jump up, big jump up. With twenty pitches and five interval classes, the input data will be only $20 + 5 \cdot 4 = 40$ parameters, instead of 8000. What is more, the 40 parameters all have strong intuitive meanings.

Obviously these 40 parameters will not in themselves fully specify the whole 8000-parameter model. What we seek is a choice of the 8000 parameters which matches the 40 that we input (in the sense that the long-term fraction of time choosing any pitch is the given distribution, and the long-term fractions for successive intervals are those given), and which is otherwise well balanced, a notion which is hard to define precisely. In section 3 we describe one algorithm which more or less meets these conditions.

2. Engineering solutions to aesthetic challenges: An apologia

Algorithms have an essential place in the creative process, and are inseparable from randomness. This is the credo on which such experiments as we describe in this paper stand or fall.

Consider a flautist, playing a simple tune. The breath is directed over the sharp leading edge to create a turbulent, noisy flow. The noise is filtered by the form of the flute, modified by the choice of fingering, generating what we call a tone. The breath can be modified, the fingers know their place, but the foundation is the inchoate and unpredictable stream of air. The performer does not, and could not, manipulate the soundwave millisecond by millisecond, but instead, like a patient parent, nudges it this way or that, in fact only shaping the statistical features of the sound, such as are realized in the frequency spectrum. Not coincidentally, this suits the statistical nature of the listener's auditory apparatus so well, as to seem perfectly controlled.

Random generation of material, filtered by choice: this schema is ubiquitous.[1] Unfortunately, this tells us almost nothing, since neither element of this dichotomy, randomness or choice, is defined or understood. Even choices as simple as moving a finger are at the moment intractable to psychology, to say nothing of the status of aesthetic choices. Randomness is even worse off, being defined only by the absence of this epistemically troubled "choice" (or of "determinism", a philosophical swamp nonpareil, into which we shall not venture). In a universe of infinite complexity, our finite minds can only choose a very few elements of any process. What happens to the rest? Chance is the concept which structures the hidden, the undetermined: once you have fixed the range of possibilities, beyond any possibility of further distinction, what else can you imagine but that the possibilities are cards in a box, into which you reach your hand and pull one out at random. Probability theory is the science which formalizes this intuition of chance.

It is important to distinguish general stochastic composition from haphazardness on the one hand, and the strictures that might be called "mere randomness" on the other. "Merely random" is in itself an incoherent concept. People often have a notion, encouraged by some experiments in the middle of the 20th century, that random composition must mean choosing from a given set of pitches with equal probabilities, and performing the output without any further human intervention. In fact, to call a procedure stochastic does not limit it at all, as it includes deterministic procedures as an extreme case. A stochastic process is defined by a predetermined set of possibilities (the "sample space") and a probability distribution which determines the relative likelihoods. The sample space could contain whole phrases, melodic figures, or more abstract elements like crescendi and cadences, and the distribution could relate the choices to one another in arbitrarily complex ways. The procedure which chooses pitches from those available with equal probabilities is a particular compositional choice, which views pitches as the fundamental elements, and weights them equally. The sample space can be structured, complementing the structure of musical thought.

To transform large-scale structural models into a single random realization of music demands, in many cases, enormous amounts of computation. It is only beginning to become feasible, with modern computers, to manipulate these models within the timeframe of a musical performance. This allows the composer/performer to improvise on a scale heretofore impossible, beginning to fulfill Xenakis' fantasy from the early Apollo years:

"le compositeur devient à l'aide des cerveaux électroniques une sorte de pilote appuyant des boutons, introduisants des coordonées et surveillant les cadrans d'un vaisseau cosmique naviguant dans l'espace des sons à travers des constellations et des galaxies sonores que seulement par le réve lointain il pouvait entrevoir jadis."[Xen63]

[1] It is perhaps noteworthy that the word *aleatory*, now almost synonymous with chance-based compositional techniques, was first used in the 1950s, as a designation for the tiny fluctuations in a natural tone which lend a sense of warmth. [Sch99, p.99]

Randomness and automation, we claim, create genuinely new possibilities for musical composition, allowing the composer to work on larger scales with less effort, hopefully freeing the attention for heretofore neglected aspects of the music. For this hope to be realized, considerable work will be needed to produce a stochastic-music toolbox, and a theory comparable to the classical Western music theory that has structured composition over the past centuries. This paper is an attempt to provide one small wrench for the toolbox.

The implications go beyond music composition, though. As with music recording, which has generated hundreds of applications far from its original domain, stochastic music models open up new uses, many of which will only become apparent through use and familiarity. We mention here just two of the more obvious possibilities, applications to music pedagogy and psychology. In music pedagogy it is the production of so-called "intelligent instruments". It has long been possible to practice playing with recorded accompaniment, allowing individual musicians more flexibility. But aside from contributing to a breakdown in musical society, such recordings have the disadvantage of repetitiveness. Particularly in an improvisational music form, the exact notes of the accompaniment may be unimportant, the accompanist in practice need not be brilliant, but repetition can be fatally soporific. An automaton that can vary the accompaniment in moderately sensible ways, or, even better, that can respond to the musician, would be a great improvement. One impressive effort in this direction has been the BoB (Band out of the Box) software by Belinda Thom.[Tho01]

Stochastic music models could also be useful in studying the psychology of music. An analogy may be drawn to the famous Julesz conjecture in vision science. It had been conjectured that the human visual system could only discriminate between visual textures on the basis of their "second-order statistics". The conjecture was handily disproved in the mid-1970's, as soon as statisticians conceived of algorithms which could generate many patterns with identical second-order statistics.[DF81] One might similarly hope that the algorithms which generate melodies with fixed statistical features might be of use in testing hypotheses regarding the human perception of statistical features in music. One example, to which we are eager to apply the model in this paper, is the current controversy over the importance of "gap-fill".[vHH00]

3. Technical details

We begin with the following data: a probability distribution, π, on the pitches (represented as the integers 0 through $n-1$, and a stochastic matrix P on the set of k interval classes. Implicit is also a list of m allowable intervals, and an assignment of the intervals to classes. (The set of intervals in class ν will be denoted S_ν.) If there is more than one interval in a class, we will also need to determine the relative weight to be sought for each interval in the class; the weights will be denoted by an m-vector w, where $\sum_{r\in S_\nu} w_r = 1$ for each class ν. (The reduction of the intervals to a smaller number of interval classes is intended both to reduce the computational difficulty, and to spare the user from distractingly irrelevant choices.) Intuitively, π represents the target long-term pitch distribution, while P gives the relative probabilities of the available alternations of interval class. The goal is to realize these local data in a Markov chain whose states are pairs (r_k, X_k), where X_k is the

k-th pitch and r_k is the interval $X_k - X_{k-1}$. By a realization, we mean that

(1) $$\text{for all pitches } i,\ \frac{1}{K}\#\{1 \le k \le K\ :\ X_k = i\} \xrightarrow{K\to\infty} \pi_i, \text{ and}$$

(2)

$$\text{for all interval classes } \mu \text{ and } \nu,\ \frac{\#\{1 \le k \le K\ :\ r_k \in S_\mu \text{ and } r_{k+1} \in S_\nu\}}{\#\{1 \le k \le K\ :\ r_k \in S_\mu} \xrightarrow{K\to\infty} P_{\mu\nu}.$$

We represent the desired stochastic matrix by $P^*(r,i;j)$, meaning the probability that pitch i, following on the interval of type r, will be succeeded by pitch j. Except in pathological circumstances (which our conditions will inevitably exclude), the matrix will be ergodic, so that there will be a unique stationary distribution, which is to say, a probability distribution on pairs (r,i) such that for all intervals s and pitches j,

(3) $$\pi^*(s,j) = \sum_{\text{intervals } r} \pi^*(r, j-s)P^*(r, j-s; j).$$

It is an elementary fact of the theory of Markov chains that the empirical occupation measure converges to the stationary measure. This means that the condition (1) will automatically hold if for all pitches i,

(4) $$\sum_{\text{intervals } r} \pi^*(r,i) = \pi_i.$$

The asymptotic interval condition requires that

(5) $$\sum_{\text{pitches } i} \sum_{r\in S_\mu} \sum_{s\in S_\nu} \pi^*(r,i)P^*(r,i;i+s) = \rho_\mu P_{\mu\nu}.$$

On the left side we have the fraction of steps which follow an interval of class μ and then choose an interval of class ν; on the right side is the fraction of pairs of steps in the target interval-class chain, in which the first is μ and the second ν.

In addition, π^* must satisfy two consistency conditions. First, if the current interval and pitch are r and $i+r$ respectively, then the preceding pitch was i, which must have the same stationary pitch distribution π. Thus

(6) $$\sum_{\text{intervals } r} \pi^*(r, i+r) = \pi_i \qquad \text{for all pitches } i.$$

Also, the long-term distribution of the interval classes must be the same as the stationary distribution of P, which means that

(7) $$\sum_{\text{pitches } i} \sum_{r\in S_\mu} \pi^*(r,i) = \rho_\mu \qquad \text{for all classes } \mu.$$

Alternatively, we may preassign weights to the different intervals within a class, call them w, so $\sum_{r\in S_\nu} w(r) = 1$, with the larger class of equations

(8) $$\sum_{\text{pitches } i} \pi^*(r,i) = \rho_\mu w(r) \qquad \text{for all intervals } r \in S_\mu$$

in place of (7).

There is one more set of conditions which needs to be imposed: suppose i is the lowest pitch, and r is a positive interval. Then $\pi^*(r,i)$ is the fraction of notes which are tone i, and follow a move up of size r, which is impossible. In general, then, we must have

$$\pi^*(r,i) = 0 \quad \text{if } i-r \text{ is not a valid tone.} \tag{9}$$

We may alternately view this as simply reducing the number of variables.

The condition (5) is quadratic in the unknowns, which could make it difficult to solve. We evade this challenge by breaking it into two parts. We first find a qualified candidate for the joint stationary distribution π^* as a solution to the linear system of equations (4), (6), (9), and (7) (or (8)). On the basis of this choice of π^* we then find the transition probabilities P^* as a solution to the linear system (3), (5), and the condition for these to be transition probabilities, namely

$$\sum_{\text{pitches } j} P^*(r,i;j) = 1 \text{ for all pitches } i \text{ and intervals } r. \tag{10}$$

3.1. **Existence of solutions.** The first question we need to ask is whether these systems actually have a solution. We will consider only the larger system (with the equations (7)), since any solution to these also solves the smaller system, with equations (8). Unfortunately, we have at this point no general answer. Consider first the equations for π^*: there are $2n+k$ equations and mn unknowns, minus the zeroed-out variables, so that if the pitches are consecutive and the intervals pass from $-(m-1)/2$ consecutively up to $(m-1)/2$, there are $mn-(m^2-1)/2$ unknowns. Two of the equations are duplicates, but there are no contradictory equations. When $n>m>3$ there are more variables than equations, which guarantees a solution. What is not guaranteed is a positive solution.

There are many obvious problems that can obstruct a joint stationary distribution. For instance, if there are gaps of zero probability in the pitch distribution larger than the largest available interval, no positive solution will be possible. A solution may also be blocked by parity problems, such as when the pitches are $0,1,\dots,n-1$, where n is even, and the intervals are all even as well.

Another constraint is that $\sum_r r\rho_r = 0$. Here $\rho_r = w(r)\rho_\mu$ when r is an interval in class μ, so ρ_r is the long-term fraction of time that the interval is r. The constraint simply says that there is no average drift, which of course must be the case if this is to describe a Markov chain on a finite set. The constraint follows formally from (7) and (4), since

$$\begin{aligned}
\sum_{i,r} \pi^*(r,i)i = \sum_i \pi_i i &= \sum_{r,i} \pi^*(r,i-r)i \\
&= \sum_{r,i} \pi^*(r,i-r)(i-r) + \sum_{r,i} \pi^*(r,i-r)r \\
&= \sum_i \pi_i i + \sum_r \rho_r r.
\end{aligned}$$

These hindrances are easy enough to exclude, but others are harder to define. For instance, suppose there are three tones, called 0, 1, 2, with distribution $(.1,.8,.1)$. Then we have the relations

$$\begin{aligned}
\pi^*(-1,1)+\pi^*(+1,1)+\pi^*(0,1) = .8 = \pi^*(-1,0)&+\pi^*(+1,2)+\pi^*(0,1) \\
\pi^*(-1,0) &\le .1 \\
\pi^*(+1,2) &\le .1,
\end{aligned}$$

which together imply that $\pi^*(0,1)$ (the fraction of time that the process holds on tone 1) is at least .6. This cannot be fulfilled, of course, if $\rho_0 < .6$. More generally, consider a melody process on the n tones $0, 1, \ldots, n-1$, where the only possible motions are -1, 0, and $+1$. The no-drift condition implies that $\rho_{-1} = \rho_{+1}$. What is more, if we define $\tilde{\pi}_i = \pi_i - \pi^*(0,i)$, the equations imply that

$$\pi^*(-1,i) = \pi^*(+1,i+1) = \tilde{\pi}_i - \tilde{\pi}_{i-1} + \tilde{\pi}_{i-2} - \cdots \pm \tilde{\pi}_0$$
$$\pi^*(-1,i) = \pi^*(+1,i+1) = \tilde{\pi}_{i+1} - \tilde{\pi}_{i+2} + \cdots \pm \tilde{\pi}_{n-1}.$$

These collectively imply the parity condition,

$$0 = \tilde{\pi}_{n-1} - \tilde{\pi}_{n-2} + \tilde{\pi}_{n-3} - \cdots \pm \tilde{\pi}_0,$$

which says that the time spent moving into even-numbered tones must be the same as time spent out of odd-numbered ones.

References

[Bri75] Reginald Smith Brindle. *The new music: the avant-garde since 1945.* Oxford University Press, Oxford, 1 edition, 1975.

[DF81] Persi Diaconis and David Freedman. On the statistics of vision: the Julesz conjecture. *Journal of Mathematical Psychology*, 24:112–38, 1981.

[Loy89] Gareth Loy. Composing with computers — a survey of some compositional formalisms and music programming languages. In Max V. Mathews and John R. Pierce, editors, *Current directions in computer music research.* MIT University Press, 1989.

[Pot71] Gary Morton Potter. *The role of chance in contemporary music.* PhD thesis, Indiana University, 1971.

[Sch99] Holger Schulze. Das Modell der nichtintentionalen Werkgenese. über Werkgeneratoren zwischen Cage und *frontpage.* In Peter Gendolla and Thomas Kamphusmann, editors, *Die Künste des Zufalls*, pages 94–122. Suhrkamp, 1999.

[Tho01] Belinda Thom. Machine learning techniques for real-tim improvisational solo trading. Available at http://www-2.cs.cmu.edu/ bthom/pubs.html, 2001.

[vHH00] Paul von Hippel and David Huron. Why do skips precede reversals? the effect of tessitura on melodic structure. *Music Perception*, 18(1):59–85, 2000.

[Xen63] Iannis Xenakis. *Musique stochastique libre, à l'ordinateur*, chapter 5. Editions Richard-Masse, 1963.

Department of Statistics, 367 Evans Hall, University of California, Berkeley, CA 94720, [1cm] CNMAT, 1750 Arch Street, Berkeley, CA 94720

E-mail address: dstein@stat.berkeley.edu, wessel@cnmat.berkeley.edu