Generation of Chord Progressions using Guided Markov Chains

Steven Berger
Lehigh University
srb420@lehigh.edu
Orcid: 0009-0000-7577-9889

DOI: 10.46926/musmat.2026v10.13-53

Abstract: *In this paper, we study the generation of tonal chord progressions in the style of J.S. Bach chorales. While much of the previous research on stochastic music generation has employed Hidden Markov Models or Neural Networks, we introduce guided Markov chains (GMC) as a method for generating tonal chord progressions. The GMC model constructs an initial base chord progression by restricting the state space of allowable chords at each time step based on composer-specified information about the harmonic function of the next chord. Unlike standard Markov Chains, GMCs give composers local control over the structure and sound of the progression without requiring explicit specification of individual chords. Although knowledge of harmonic functions is well established in music theory, most existing stochastic models must learn this structure from data. We demonstrate that GMC models increase the likelihood of observing chord sequences consistent with the Baroque style relative to a standard Markov Chain and also provide comparisons of cumulative log-likelihoods between GMC and hidden Markov models. Additionally, we prove the Typical Set Cardinality Reduction Theorem and introduce two confidence intervals: a Monte Carlo confidence interval to quantify the difference in standard Markov chain and guided Markov chain entropy for finite-length chord sequences, and a permutation-based confidence interval to assess the statistical significance of a specific harmonic guide ordering. We also outline an extension for incorporating chord inversions.*

Keywords: *Markov chains. Entropy. Hidden Markov models. Typical sets.*

1. Introduction

Standard Markov chains (SMC) have many limitations which cause them to generate chord progressions that are unsatisfactory when compared against the progressions of a master composer such as J.S. Bach. First, they do not respect musical form. A composer using a standard Markov chain to generate a chord progression can only specify the total number of chords in the entire piece, not the number of chords that fall into each phrase. A quick inspection of any Bach chorale reveals that musical phrase length is governed by the number of syllables in each phrase of hymn text with each phrase ending with a standard cadence type. Since expected hitting times for a Markov chain state are determined by a system of equations defined by the

Received: January 6th, 2026
Approved: May 1st, 2026

transition matrix, it is possible to structure the transition probabilities so as to control, and in some cases fix, the expected number of steps required to reach a given state [3]. However, music where the phrase lengths are random is not appropriate for standard musical forms. Although extensions such as higher-order or semi-Markov models can partially address this issue, they do so at the cost of increased model complexity and reduced interpretability (see [7]). Second, while standard Markov chains can easily learn patterns of chord progressions from training data, merely producing a sequence of chords with allowable transitions does not capture the compositional intent of tonal music. Forgetting the previous history and only basing the choice of the next chord on the current chord without further context is not musically meaningful. Third, Bach uses text-painting in his chorales. This means that at a fixed musical time t during the chorale, the chord that is sounding must match the sentiment of the lyrics being sung at this same time. Standard Markov chains do not provide such local control over the generated chord sequences.

Deep learning methods such as neural networks are also becoming popular tools for music-related tasks as demonstrated by [1] and [2]. However, neural networks require large amounts of training data which is not always available. Furthermore, neural network models typically offer limited transparency and afford little direct control over harmonic function which makes them less suitable for composers seeking interpretable, rule-guided harmonic control. It is also difficult to adapt these networks to capture the stylistic evolution of a composer. While one could train separate models on data from different compositional periods of the same composer, in practice this may be limited by the relatively small amount of available data within each stylistic phase, as well as the lack of clear segmentation between styles.

This paper proposes a return to Markov-based models with the modification that we use an external variable to guide the models. By defining the state space of the Markov chain to consist of standard diatonic and chromatic Roman-numeral triads [1] and guiding the generation of chord sequences by the harmonic function of the chords, we create a system that gives the composer more local control of the generated sequences without requiring the composer to specify each individual chord. The contributions discussed here include (i) the formal definition of guided Markov Chains, (ii) log-likelihood-based comparisons between guided Markov chains and other Markov modles, and (iii) theoretical and empirical results quantifying the reduction in entropy provided by the harmonic guide.

2. Basic Definitions

In this section, we define standard Markov chains (SMC) and hidden Markov models (HMM) along with other important definitions and theorems relevant to these stochastic processes.

2.1. SMC Models

One commonly used stochastic model that can be used for modeling sequences of chords is a standard Markov chain (SMC) as defined in [3].

[1] Chordal sevenths and inversions are incorporated in a subsequent stage after a root-position chord progression has been generated. Thus, the triadic state space serves as an abstraction of underlying harmonic function rather than a restriction on what pitch classes appear in the final music score.

Definition 1. *Let S be a discrete set. A Markov chain is a sequence of random variables $X_0, X_1, \cdots$ taking values in S with the property that*

$$\boldsymbol{P}(X_{n+1} = j | X_0 = x_0, \cdots, X_{n-1} = x_{n-1}, X_n = i) = \boldsymbol{P}(X_{n+1} = j | X_n = i),$$

for all $x_0, \cdots, x_{n-1}, i, j, \in S$, and $n \geq 0$. The set S is the state space of the Markov chain.

By definition, Markov chains forget all states in the history except the current state when calculating the probability of transitioning to the next state. In a Markov chain, probabilities of transitioning between states are usually displayed in a transition probability matrix T with each row sum constrained to equal 1. Thus, entry T_{ij} would be the probability of transitioning to state j given that we are currently in state i. Markov chains provide many options for creating progressions of notes, chords, and rhythms but they have limitations that may affect the quality of music produced.

We now provide some important definitions and results relating to discrete Markov chains.

Definition 2. *A transition probability matrix T is said to be regular if some power of T contains only positive entries.*

Definition 3. *Let $X_0, X_1, \cdots$ be a Markov chain with transition probability matrix T. A limiting distribution for the Markov chain is a probability distribution $\vec{\lambda}$ with the property that for all i and j,*

$$\lim_{n \to \infty} T_{ij}^n = \lambda_j.$$

The value λ_j can be interpreted as the long-term probability the Markov chain visits state j. Since limits are unique, limiting distributions are unique (if a Markov chain has one).

Definition 4. *Let $X_0, X_1, \cdots$ be a Markov chain with transition matrix T. A stationary distribution is a probability distribution $\vec{\pi}$, which satisfies $\vec{\pi} = \vec{\pi}T$ or equivalently $\pi_j = \sum_i \pi_i T_{ij}$.*

Dobrow proves (see [3]) the following result called the Limit Theorem for Regular Markov Chains.

Theorem 1. *A Markov chain with a regular transition probability matrix has a limiting distribution which is the unique, positive, stationary distribution of the chain.*

This theorem allows us to study the long-term behavior of a Markov chain with a regular transition probability matrix simply by calculating its stationary distribution and interpreting the j-th entry of the distribution as the proportion of times the Markov chain would spend in state j over a long period of time.

Definition 5. *A state j is said to be accessible from state i if there exists an $n \geq 0$ such that $T_{ij}^n > 0$.*

Definition 6. *Two states of a Markov chain i and j communicate if i is accessible from j and j is accessible from i.*

The state space of any discrete time Markov chain can be partitioned into disjoint subsets (called communication classes) such that in each partition, all states communicate with each other but do not communicate with any states outside the partition.

Definition 7. *A Markov chain is irreducible if it contains only one communication class.*

Definition 8. *For any Markov chain with transition probability matrix T, a state i has period $d(i)$ which is the greatest common divisor of the set of all possible return times to state i. More specifically, $d(i) = gcd\{n > 0 : T^n_{ii} > 0\}$. Furthermore, if $d(i) = 1$, the state i is called aperiodic.*

Definition 9. *A Markov chain with finite state space is said to be ergodic if it is aperiodic and irreducible.*

Theorem 2. *For an ergodic Markov chain with finite state space, there exists a unique positive stationary distribution $\vec{\pi}$ which is the limiting distribution of the chain. Thus $\pi_j = \lim_{n\to\infty} T^n_{ij}$ for all i, j.*

In this paper, we will frequently be concerned with whether or not a given Markov chain on a discrete set has a unique stationary distribution.

2.2. HMM Models

One popular extension of the standard Markov chain is the hidden Markov chain (HMM).

Definition 10. *A pair of stochastic processes (X_t, C_t) is a Hidden Markov chain (HMM) if $\{C_t, t \in \mathbb{N}\}$ is an unobservable process satisfying the Markov property and $\{X_t, t \in \mathbb{N}\}$ is an observable process where X_t depends only on C_t. More specifically:*

$$\boldsymbol{P}(C_t|C_1, C_2, \cdots, C_{t-1}) = \boldsymbol{P}(C_t|C_{t-1})$$

and

$$\boldsymbol{P}(X_t|X_1, X_2, \cdots, X_{t-1}, C_1, C_2, \cdots, C_{t-1}) = \boldsymbol{P}(X_t|C_t).$$

Hidden Markov chains tend to be a more flexible approach to sequential data modeling than SMC models, especially when the observed data is a mixture of two distinct time series. Hidden Markov chains have historically been applied to bioinformatics, speech recognition, and natural language processing; see [4] for a thorough discussion of HMM applications. While HMMs introduce latent structure, they do not naturally encode functional harmonic constraints unless these are inferred implicitly from data.

3. Selected Examples of Stochastic Music Composition

From the musical dice games of Mozart to the modern world of AI-generated music, composers often turn to stochastic techniques for aid in the compositional process. In the 1950s, Hiller and Isaacson from the University of Illinois Urbana-Champaign used an early computer to compose a four-movement string quartet called the *Illiac Suite* (see [5]). Their work uses a weighted probability distribution created so that on average the melodic line oscillated between consonant and dissonant intervals. The work of Hiller and Isaacson is significant because it marks one of the first attempts to get computers to explicitly encode J.J. Fux's rules of counterpoint.

Also in the late 1950s, the Greek-French avant-garde composer Iannis Xenakis (1922-2001) was experimenting with Markov chains in music composition. His piece *Analogique A* (written for 9 string players) makes use of probabilistic transitions between pre-composed screens that contain information on the duration, frequency, and intensity of pitches. The full details of his model are described in Chapter 3 of his book *Formalized Music* [6].

More recently, Hugo Carvalho experimented with using Markov chains to compose a short melody in the style of J.S. Bach (see [7]). He uses data consisting of 1637 observations of notes all from chorale melodies in A major representing 13 unique pitches (E4 to G5). Carvalho's main point with this experiment was to evaluate the quality of the melodies resulting from simulation of notes from first and fourth order Markov chains. He points out that increasing the order of the Markov chain results in more frequent direct plagiarism of melodic sequences already in the training data.

Furthermore, prior work has applied tree-structured models to chord progression modeling and generation. Paiement, Eck, and Bengio [8] propose a graphical model that captures both local and meter-based global dependencies. Chords are represented as psychoacoustically motivated vectors in $\mathbb{R}^{12}$, and Euclidean distances between them are converted into substitution probabilities via a Gaussian kernel. Their empirical results show that the resulting tree-structured latent variable model outperforms a hidden Markov model in conditional likelihood on a corpus of jazz standards. Extending this line of work, Tsushima, Nakamura, and Yoshii [9] propose a Bayesian framework for melody harmonization that also leverages a tree-structured representation of harmony. Their approach formulates a unified hierarchical generative model consisting of a probabilistic context-free grammar (PCFG) that generates chord sequences with associated harmonic functions, a metrical Markov model governing chord onset positions, and a Markov model for melody conditioned on the chord sequence. By explicitly modeling hierarchical structure, harmonic function, and metrical timing, their method outperforms standard HMM-based approaches in chord prediction accuracy.

Previous research has also focused on attempts to control the output of Markov models. Pachet and Roy [10] propose a framework for steerable generation of Markov sequences by incorporating constraints into the sampling process. By representing constraints as finite automata and combining them with a Markov model, their method enables exact sampling from the distribution of sequences satisfying global constraints. This approach allows Markov models to maintain local statistical structure while enforcing long-range dependencies. Additionally, Eigenfeldt and Pasquier [11] propose a realtime chord sequence generation system with user influence over variable-order Markov transition tables.

This research develops a framework that can be used to guide a Markov model at each chord change by filtering out chords that do not have the desired harmonic function. The remaining chords have their probability masses rescaled to sum to 1 and then the next chord is randomly selected from this set. This provides a means for the composer to impose local constraint while still using the structure of a Markov-based model.

4. GMC Models

In this section, we introduce the definition and construction of Guided Markov Chain (GMC) models.

4.1. Harmonic Function Definitions

We begin by formalizing the notion of harmonic function for Roman numeral chords, which serves as the guiding variable in the GMC framework. In the rest of this paper, Roman numeral chords are understood in the conventional tonal context. Since the present focus is on probabilistic structure rather than pitch-class realization, explicit construction of chordal pitch content is omitted.

Let S denote the set of allowable Roman numeral chord symbols (or chordal units), and let $\mathcal{H}$ be a finite set of harmonic function labels. Define the mapping $h : S \to \mathcal{H}$, where $h(s)$ denotes the harmonic function associated with chord $s \in S$. Table 1 summarizes the values of $h(s)$ for Roman numeral chords in a major key. These assignments are intended to capture the most commonly observed harmonic roles in Bach chorales and closely related tonal repertoire.

s	$h(s)$
I	Tonic (T)
ii	Predominant (p)
IV	Predominant (p)
I^6_4-ii	Strong Predominant (P)
ii-I^6_4	Strong Predominant (P)
V	Dominant (d)
vii^{o}	Dominant (d)
V/V-V	Dominant (d)
vii^{o}/V-V	Dominant (d)
I^6_4-V	Strong Dominant (D)
iii	Elongation (e)
vi	Elongation (e)
IVp	Elongation (e)
V/vi-vi	Color (c)
V/IV-IV	Color (c)
vii^{o}/ii-ii	Tension (f)

Table 1: *The Harmonic Function $h(s)$ for Major Key Signatures*

We make several remarks concerning these harmonic function assignments.

1. The classifications in Table 1 are meant to be representative rather than exhaustive. They reflect the harmonic functions most frequently encountered in Bach chorales. This framework may be adapted to accommodate alternative compositional styles or historical periods.

2. The label *elongation* refers specifically to the prolongation of the tonic chord.

3. Since the IV chord may serve multiple harmonic roles, we distinguish its elongational use (denoted IVp) from its predominant function (denoted IV). In the final musical realization, these symbols will both contain the same pitch classes.

4. With the exception of the cadential six-four chord, which is widely understood not to function as a tonic chord, we primarily restrict attention to root-position chords. Inversions will be added later.

5. Certain conventional harmonic patterns, such as standard resolutions of secondary dominants, are incorporated directly into the model as chordal units.

6. The harmonic function *color* is not drawn from a formal theoretical framework. It is used here as a modeling abstraction to capture applied harmonic motion, specifically secondary-dominant progressions such as V/IV-IV and V/vi-vi. These progressions momentarily shift tonal focus and thus exhibit harmonic behavior that is not captured by the other harmonic function categories.

Table 2 provides the corresponding harmonic function assignments for Roman numeral chords in minor keys [2].

s	$h(s)$
i	Tonic (T)
ii^{ø}	Predominant (p)
iv	Predominant (p)
i$^{6}_{4}$-ii^{ø}	Strong Predominant (P)
ii^{ø}-i$^{6}_{4}$	Strong Predominant (P)
V	Dominant (d)
vii^{o}	Dominant (d)
V/V-V	Dominant (d)
vii^{o}/V-V	Dominant (d)
i$^{6}_{4}$-V	Strong Dominant (D)
VI	Elongation (e)
ivp	Elongation (e)
V/III-III	Color (c)
III-VI	Color (c)
VII	Color (c)
III^{+}	Tension (f)

Table 2: *The Harmonic Function $h(s)$ for Minor Key Signatures*

4.2. Transition Rules for Harmonic Functions

Having defined the harmonic functions associated with Roman numeral chords, we now specify the admissible transitions between harmonic functions. These transition rules encode standard constraints of tonal harmony and serve as a structural guide for the GMC model.

Figures 3 and 4 summarize the allowable transitions between harmonic functions in major and minor keys, respectively. Each diagram may be interpreted as a binary adjacency matrix on the set of harmonic functions: an entry marked `TRUE` indicates that a transition from the row function to the column function is permitted, while a `FALSE` entry indicates that the transition is not allowed.

	TO T	TO p	TO P	TO d	TO D	TO e	TO c	TO f
FROM T	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
FROM p	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
FROM P	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
FROM d	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
FROM D	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE
FROM e	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
FROM c	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
FROM f	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE

Table 3: *Allowable transitions between harmonic functions (major keys)*

[2] In the minor-mode transition table, half-diminished symbols are used without specifying chordal sevenths or inversions. The corresponding sevenths and inversions are introduced in a subsequent chord-realization stage (see Appendix 8).

	TO T	TO p	TO P	TO d	TO D	TO e	TO c	TO f
FROM T	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
FROM p	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
FROM P	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
FROM d	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
FROM D	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE
FROM e	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
FROM c	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
FROM f	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Table 4: *Allowable transitions between harmonic functions (minor keys)*

These transition constraints reflect common-practice tonal syntax. For example, tonic chords may progress to any harmonic function, whereas predominant functioning chords are not permitted to resolve directly to tonic-function chords.

In the GMC framework, these harmonic function transition rules are used to define a partition operation on the underlying Roman-numeral transition matrix, ensuring that only harmonically admissible chord transitions receive positive probability mass.

4.3. Expected Likelihood Matrices

Music composition involves balancing adherence to established stylistic norms with occasional deviation in order to maintain interest. To encode such stylistic expectations within a probabilistic framework, we introduce an *expected likelihood matrix* L, which represents a composer's prior beliefs about the relative frequency of transitions between Roman numeral chords in tonal music. Each entry of L quantifies the expected relative frequency that a chord of type i progresses to a chord of type j. While the specific values of this matrix may vary across composers and styles, composers seeking to emulate the harmonic practice of J.S. Bach should arrive at matrices with broadly similar structure. We first present the expected likelihood matrix for major keys in Figure 1.

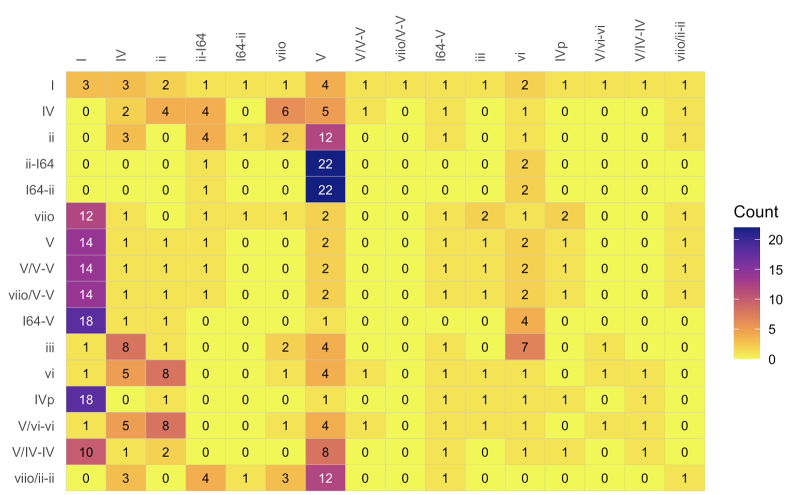

Figure 1: *Expected likelihood matrix for major keys*

The entry in row i and column j of L represents the expected number of times (out of $Z \in \mathbb{Z}^+$) that Roman numeral i transitions to Roman numeral j. In this work, $Z = 25$ but this could be changed based on preference of the composer. The matrix is constructed so that each row sums to Z, a normalization chosen for interpretability and motivated by the relative frequency of transitions from the tonic chord. The entries were selected using prior musical knowledge informed by common-practice harmonic syntax and empirical observation of Bach chorales.

We now make several remarks regarding the construction of this matrix.

1. Chord transitions that follow the circle of fifths are assigned at least 6 of the 25 available counts for a given row, reflecting their prominence in tonal harmony. Exceptions occur in the rows corresponding to I and IVp, due to specific stylistic constraints governing their usage.

2. Each Roman numeral is required to have transitions to at least two distinct harmonic functions in a single step, ensuring local harmonic flexibility. In some cases, one of these transitions may preserve the chord's own harmonic function.

3. Consistency is enforced among chords sharing the same harmonic function. For example, since IVp may resolve to I, the color-function chords iii and vi are also permitted to resolve to I. Although such progressions are less common in Bach chorales, they are musically valid within phrase interiors. The introduction of IVp specifically prevents all predominant chords from resolving directly to tonic while still accommodating the realized progression IV–I.

Composers wishing to modify the expected likelihood matrix for major keys should preserve these structural constraints.

There are several reasons why manual specification of the expected likelihood matrix is preferred over formal statistical estimation. First, maximum likelihood estimation would require a substantially larger corpus of chorales. Even then, rare chords such as iii in major keys would yield unstable estimates. For instance, the limited frequency of iii chords makes it unlikely that all transition probabilities in the corresponding row could be reliably estimated from data alone.

Second, the use of integer-valued expected likelihoods yields parameters that are easily interpretable and readily adjustable by composers. This flexibility is particularly advantageous when adapting the model to related stylistic contexts, such as Baroque dances, for which fewer training examples may be available. In such cases, modifying the entries of L is considerably more practical than re-estimating parameters from limited data.

Figure 2 presents the expected likelihood matrix for minor keys. This matrix is constructed using the same guiding principles as its major counterpart, with modifications reflecting the different set of color chords and changes in chord quality characteristic of minor tonalities.

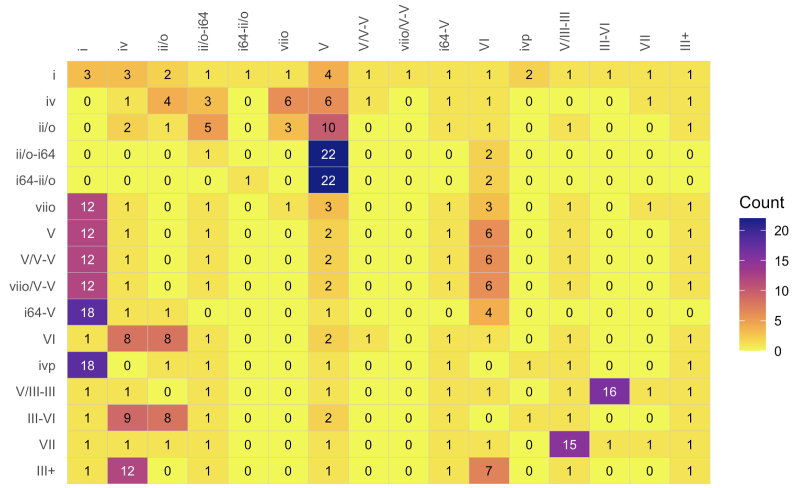

Figure 2: *Expected likelihood matrix for minor keys*

Each expected likelihood matrix can be converted into a transition probability matrix by dividing all entries by Z. Since the third power of each resulting transition matrix is regular, both chains have a unique stationary distribution.

4.4. Definition of GMC

We now formalize the definition of a guided Markov chain (GMC). First, assume the composer has chosen an expected likelihood matrix L (major or minor) from the previous section, along with its corresponding set of row labels S (the set of all possible Roman numerals).

Definition 11. *Let $H_1, H_2, \ldots, H_N$ be a given sequence of N harmonic functions, and define*

$$S_t = \{s \in S : h(s) = H_t\}, \quad t = 1, 2, \ldots, N.$$

A guided Markov chain *(GMC) is a sequence of random variables $X_1, X_2, \ldots, X_N$ such that $X_t \in S_t$ for all t, and*

$$\boldsymbol{P}(X_n = j \mid X_1 = x_1, \ldots, X_{n-1} = i; H_1, \ldots, H_N) = \boldsymbol{P}(X_n = j \mid X_{n-1} = i; H_n)$$

for all $x_t \in S_t$, all $i \in S_{n-1}$, all $j \in S_n$, and $2 \le n \le N$.

In practice, generating a chord progression using a GMC requires specifying the Roman numeral of the first chord and the full guiding sequence of harmonic functions. Intuitively, the GMC randomly selects the next chord from only those that match the prescribed harmonic function H_n, with probabilities rescaled so that they sum to one. Equivalently, chords not matching the guiding function are assigned zero probability, and the remaining probabilities are normalized.

This approach addresses a key limitation of standard Markov chains (SMCs), which lack respect for musical form. In SMCs, phrase lengths and cadence locations are random. However, a GMC allows a composer to impose structure directly; for example, a cadence can be added at a specific location by including the harmonic sequence P,d,T at the appropriate time steps. Thus, the GMC leverages harmonic function as a guiding variable to refine the sample space for each successive chord, producing progressions that are more musically coherent and textually sensitive.

Formally, a GMC is a time-inhomogeneous Markov chain whose state space at each step is conditioned on an external guiding variable. While the study of inhomogeneous Markov chains is well-established (see [12]), the application of such chains to music composition conditioned on the harmonic functions defined above and the impact of the guiding sequence on entropy calculations both have not been previously studied. In the remainder of this paper, we explore the theoretical properties, statistical implications, and practical applications of GMC models.

5. SMC and GMC Likelihoods

We will now prove that GMC models provide an improvement over SMC models in the sense that they increase the likelihood of observing chord progressions that Bach has actually written.

Theorem 3. *Let $C_1, C_2, \cdots, C_n$ be a sequence of $n \geq 2$ Roman numeral chords such that $L_{C_{i-1}C_i} > 0$ for all $i \in \{2, 3, \ldots, n\}$. Let the harmonic function of a chord C_i be $h(C_i) = H_i$.*

Then the probability of observing this sequence under the SMC is strictly less than the probability of observing the sequence under the GMC with harmonic sequence $H_1, \ldots, H_n$ assuming that the first chord of the sequence is fixed by the composer.

Proof. We will prove this result by induction. For the base case, consider a sequence of 2 Roman numerals C_1, C_2 such that $C_1 \in S$, $C_2 \in S$, and $L_{C_1C_2} > 0$ with associated harmonic functions $h(C_1) = H_1$ and $h(C_2) = H_2$. Let Y_1, Y_2 be observations from a SMC with probability transition matrix $\frac{L}{Z}$ where Z is the fixed sum of a row in matrix L. We see that:

$$\mathbf{P}_{\text{SMC}}(C_1C_2) = \mathbf{P}(Y_1 = C_1)\mathbf{P}(Y_2 = C_2 | Y_1 = C_1).$$

However, the first Roman numeral C_1 is always specified by the composer in order to provide the SMC with a place to start. Thus,

$$\begin{aligned}\mathbf{P}_{\text{SMC}}(C_1C_2) &= \mathbf{P}(Y_2 = C_2 | Y_1 = C_1) \\ &= \frac{L_{C_1C_2}}{Z}.\end{aligned}$$

Now let X_1, X_2 be observations from a GMC with transition matrix $\frac{L}{Z}$ and harmonic function $h(s)$ for all $s \in S$. Thus

$$\mathbf{P}_{\text{GMC}}(C_1C_2) = \mathbf{P}_{\text{GMC}}(X_1 = C_1)\mathbf{P}_{\text{GMC}}(X_2 = C_2 | X_1 = C_1).$$

Once again, the composers specifies C_1 and thus $h(C_1) = H_1$ is known. Therefore,

$$\begin{aligned}\mathbf{P}_{\text{GMC}}(C_1C_2) &= \mathbf{P}_{\text{GMC}}(X_2 = C_2 | X_1 = C_1) \\ &= \frac{L_{C_1C_2}}{\sum\limits_{s \in S: h(s) = H_2} L_{C_1 s}}.\end{aligned}$$

Next, we note that for any Roman numeral in either the major or the minor probability transition matrix we constructed earlier, there exists at least 2 classes of harmonic functions that Roman numeral has non-zero probability of reaching in one step. This means that

$$Z > \sum_{s \in S: h(s) = H_2} L_{C_1 s}$$

and therefore

$$\frac{L_{C_1C_2}}{Z} < \frac{L_{C_1C_2}}{\sum\limits_{s\in S:h(s)=H_2} L_{C_1s}}.$$

Thus we conclude that

$$\mathbf{P}_{\text{GMC}}(C_1C_2) > \mathbf{P}_{\text{SMC}}(C_1C_2).$$

Now assume that for a chord sequence of length $n \in \mathbb{N} \setminus \{1\}$,

$$\mathbf{P}_{\text{GMC}}(C_1C_2\cdots C_n) > \mathbf{P}_{\text{SMC}}(C_1C_2\cdots C_n). \tag{1}$$

By the definition of conditional probability, we can decompose the left-hand side of inequality 1 into

$$\mathbf{P}_{\text{GMC}}(C_{n+1}|C_1C_2\cdots C_n)\cdot\mathbf{P}_{\text{GMC}}(C_1C_2\cdots C_n) \tag{2}$$

$$= \mathbf{P}_{\text{GMC}}(C_{n+1}|C_n)\cdot\mathbf{P}_{\text{GMC}}(C_1C_2\cdots C_n) \tag{3}$$

$$= \frac{L_{C_nC_{n+1}}}{\sum\limits_{s\in S:h(s)=H_{n+1}} L_{C_ns}}\cdot\mathbf{P}_{\text{GMC}}(C_1C_2\cdots C_n). \tag{4}$$

By a similar argument, we can decompose the right side of inequality 1 into

$$\mathbf{P}_{\text{SMC}}(C_1C_2\cdots C_{n+1}) = \mathbf{P}_{\text{SMC}}(C_{n+1}|C_n)\cdot\mathbf{P}_{\text{SMC}}(C_1C_2\cdots C_n) \tag{5}$$

$$= \frac{L_{C_nC_{n+1}}}{Z}\cdot\mathbf{P}_{\text{SMC}}(C_1C_2\cdots C_n). \tag{6}$$

Since

$$Z > \sum_{s\in S:h(s)=H_{n+1}} L_{C_ns}$$

this means that

$$\frac{L_{C_nC_{n+1}}}{Z} < \frac{L_{C_nC_{n+1}}}{\sum\limits_{s\in S:h(s)=H_{n+1}} L_{C_ns}}. \tag{7}$$

Therefore,

$$\mathbf{P}_{\text{SMC}}(C_{n+1}|C_n) < \mathbf{P}_{\text{GMC}}(C_{n+1}|C_n). \tag{8}$$

By multiplying inequality 1 by inequality 8, we obtain the following result.

$$\mathbf{P}_{\text{GMC}}(C_1C_2\cdots C_nC_{n+1}) > \mathbf{P}_{\text{SMC}}(C_1C_2\cdots C_nC_{n+1}). \tag{9}$$

By induction, we have shown that an allowable sequence of Roman numerals must have a higher likelihood under the GMC. ■

Remark: The theorem applies only to sequences where every transition has non-zero probability according to L. For Bach chorales, this condition is satisfied because each chord can transition to multiple harmonic function classes, ensuring the GMC probabilities are well-defined.

This increase in likelihood is not achieved by altering the underlying transition structure, but solely by reallocating probability mass to transitions consistent with the appropriate harmonic function. This provides the probabilistic mechanism underlying entropy reduction and typical-set restriction, both of which will be investigated later.

5.1. Numerical Confirmation of Likelihood Theorem

We can verify our likelihood theorem from the previous section by computing the likelihood of a chord sequence Bach uses in an actual chorale. We select the first 6 Roman numerals [3] from the opening phrase of *Komm, Heiliger Geist, Herre Gott* (BWV 59.3). The sequence we wish to evaluate is shown in Figure 3. Under the SMC, this sequence has likelihood 0.000012 while under the GMC it has harmonic function sequence T-p-d-e-T-d and likelihood 0.071428. As the theorem promises, the likelihood is higher under the GMC. In order to make the comparisons fair, both the SMC and GMC used the same transition matrix.

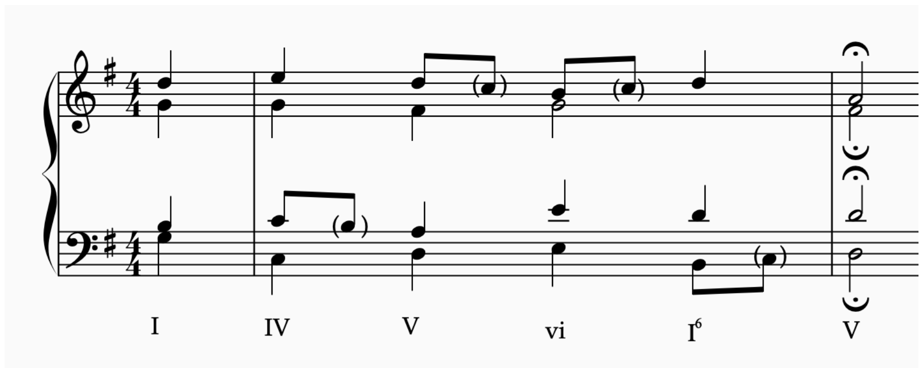

Figure 3: *Roman numeral analysis of BWV 59.3 in G major*

It is also interesting to study the average and variance of cumulative likelihoods under both models. Using the same sequence of harmonic functions (T-p-d-e-T-d), we ask the GMC to compose 200 chord progressions. For each of these 200 chord sequences, we calculate its cumulative likelihood under both models for each possible sequence length. Algorithm 1 (Appendix A) demonstrates how a sequence of Roman numeral chords of length n is generated from an SMC and Algorithm 2 (Appendix A) demonstrates how a sequence of Roman numeral chords of length n is generated from a GMC. Algorithms 3 and 4 (Appendix B) show how to compute the log-likelihood function of a sequence under the SMC and GMC respectively.

In Figure 4, we plot the average of the cumulative log-likelihoods under each model as a function of the phrase length (2 to 6). We see that on average, the GMC even manages to have a higher average cumulative log-likelihood for this short sequence of harmonic functions.

We can also study the variance of the cumulative log-likelihood under each model as shown in Figure 5.

Notice that initially, the variance in the cumulative log-likelihood is roughly the same for both models until the variance of the SMC spikes in the second half of the plot. This demonstrates the impact of the guiding sequence. Initially, the guide is forcing us to follow a sequence of chords that the SMC would likely have followed without the guide. However, the divergence in the average cumulative log-likelihoods of the models comes when the guide mandated that the fourth chord in the sequence have harmonic function elongation and the fifth chord have

[3]Note that both the SMC and GMC model do not distinguish between root position or first inversion tonic chords. Inversions are added later (see Appendix G).

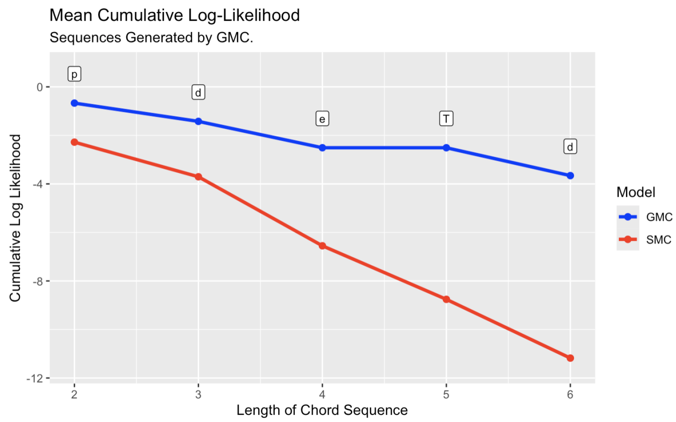

Figure 4: *Cumulative Average Log-Likelihood*

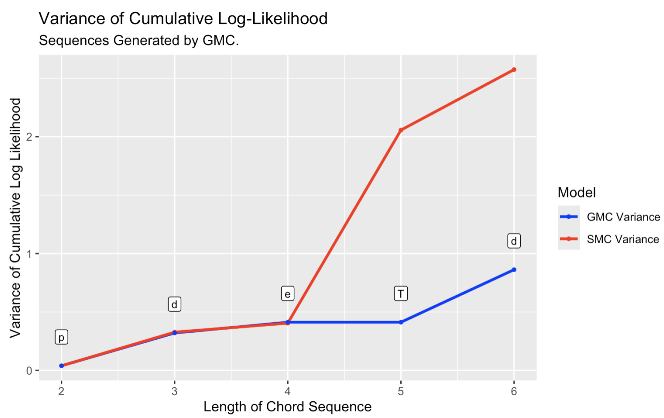

Figure 5: *Variance of the Cumulative Log-Likelihood*

harmonic function tonic. From tonal harmony, we know that elongation chords (IVp, iii, vi) are most likely to go to some kind of predominant or elongation chord. While it is certainly not wrong to follow an elongation chord with a tonic chord, it is not the most common option so the SMC is somewhat surprised by this sequence and the variance of the cumulative log-likelihood increases. However, notice that the GMC experiences constant variance of the cumulative log-likelihood at this point. It knows that for some musically significant reason, Bach chose to follow an elongation chord with a tonic chord and since there is only one tonic chord, the variance of the cumulative log-likelihood of every chord sequence stays the same. It makes sense from a musical perspective that we do not want to decrease the likelihood function for sounding a tonic chord.

Now we will examine the average and variance of the cumulative log-likelihood for a longer guiding sequence (T-p-d-T-f-P-d-T) that is repeated 3 times. The average over 200 simulations of the cumulative log-likelihood is shown in Figure 6.

Once again, we see that even for a longer sequence of chords, the GMC continues to have a higher average cumulative log-likelihood. Figure 7 shows the variance of the cumulative log-likelihood for these same 200 simulations on the longer sequence.

As this sequence is fairly typical of Bach's usage of tonal harmony, we see many plateaus or small increases in the plots (especially when the sequence wraps around on itself (i.e. T to T).

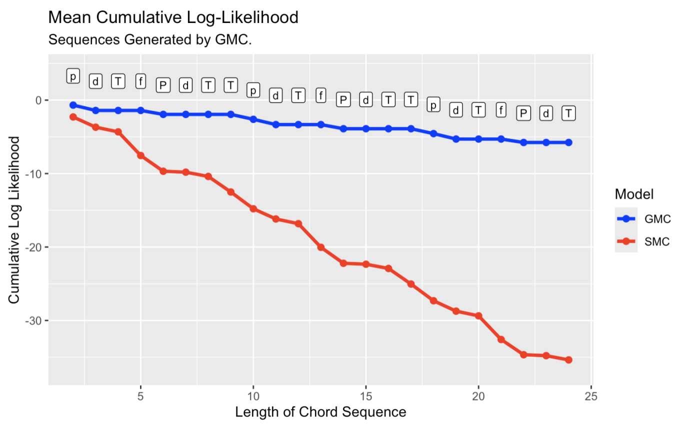

Figure 6: *Cumulative Average Log-Likelihood of a Longer Sequence*

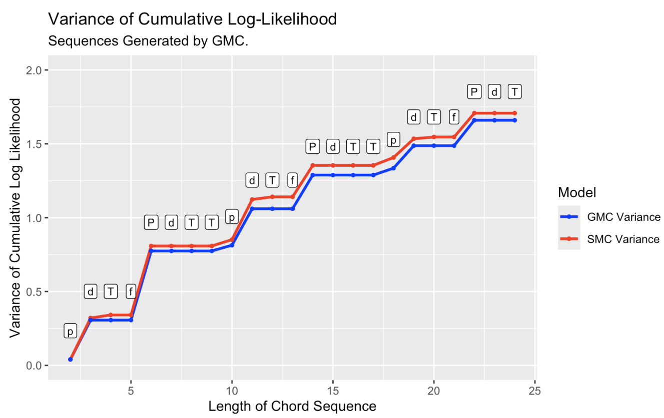

Figure 7: *Variance of Cumulative Log-Likelihood of a Longer Sequence*

When the guide commands a tonic chord followed by a tonic chord, both the SMC and GMC will have variance of cumulative log-likelihood equal to 0 because all of the 200 simulations must generate the progression I-I and thus there is no variability in the log-likelihood at this step. There is also a plateau in the variance when a chord with harmonic function P goes to a chord with harmonic function d because both strong predominant chords must go to V since that is the only dominant functioning chord they are allowed to transition to. It seems that the SMC and GMC both have similar variance in their cumulative log-likelihoods but the SMC tends to have a lower average cumulative log-likelihood. This provides strong numerical evidence that the GMC better captures functional tonal structure than the SMC.

5.2. Comparison with the HMM

It is clear that the GMC outperforms the simple SMC model because the GMC knows more about the structure of the music (i.e. the harmonic function of each chord). However, we now investigate whether the GMC outperforms the Hidden Markov model (HMM) that can be trained using the harmonic functions as hidden states. Algorithm 5 (Appendix C) demonstrates how to create a sequence of length n from an HMM. Furthermore, Algorithm 6 (Appendix C) demonstrates how the HMM scores a sequence of chords. This is the well-known Forward Algorithm.

In order to investigate the effectiveness of the GMC compared with an HMM, we train an HMM on Bach chorales and score the log-likelihood of the sequences it generates. In order to be included in the training data, chorales must only use the chords defined in the major key state space (i.e. no modulations to minor keys). If a chorale modulated to a minor key, all chords in that modulation were removed and a new training sequence was started to avoid creating a false chord transition. Chord progressions from 21 chorales were included in the training data along with 4 sequences composed (in the style of Bach) by the author in order for the HMM to observe theoretical transitions for rarely observed Roman numerals.

Once the training data has been collected, we need to fit the HMM to this data. Initial attempts to train the Hidden Markov Model using the standard Baum-Welch (Expectation-Maximization) algorithm proved unsuitable for this specific application due to three primary factors:

1. **Data Imbalance:** The distribution of chords in the Bach chorales is highly skewed, with tonic and dominant chords appearing with much greater frequency than other harmonic function categories. Thus, the likelihood-maximizing objective of the Baum-Welch algorithm only converged when the model collapsed the complex 7-state harmonic function definitions into a simplified binary system (tension-resolution). This ignores important distinctions between harmonic function categories required by music theory.
2. **Harmonic Function Ambiguity:** The model struggled to distinguish between chords that exhibit similar local transition behavior but play distinct roles in larger harmonic contexts. For example, consider the progressions V–I–IV and V–vi–IV. While both are stylistically valid in Bach chorales, the I chord represents harmonic resolution, whereas the vi chord in the second progression functions as a deceptive continuation that delays harmonic closure. We acknowledge that in some theoretical frameworks, I and vi may be treated as functionally related or partially interchangeable. However, for the purposes of this model, it is essential to distinguish between resolution and deferral of resolution. These behaviors have different structural implications. In practice, the Baum–Welch algorithm tended to group such chords into a single latent state due to their similar local transition patterns, thereby obscuring this distinction and reducing the interpretability of the learned harmonic structure.
3. **Numerical Instability with Priors:** To enforce music-theoretic definitions, prior weights were applied (later normalized to probabilities) to the emission matrix (effectively freezing the relationship between functions and chords). However, these extreme weights (10^{10} and 10^{-10}) caused numerical underflow during the summation steps of the standard Forward-Backward algorithm. This led to computational failure.

Therefore, we adopted the supervised Viterbi training method (see [13]). By calculating transitions based solely on the single optimal path decoded via the Viterbi algorithm, we ensured numerical stability while strictly enforcing the a priori functional definitions derived from tonal music theory. Algorithm 7 (Appendix D) details the process of Viterbi training for an HMM.

Our HMM was trained to have 8 hidden states which each correspond to a distinct harmonic function. Recall that no Roman numeral chord is allowed to have more than one harmonic function. Furthermore, we freeze the emission probabilities for each harmonic function class. For example, the second harmonic function class (dominant chords) was assigned emission weights proportional to 10^{10} for V, V/V-V, vii^{o}/V-V, and vii^{o} chords. All other chords were assigned emission weights in this category of 10^{-10}. The emission weights for each harmonic function class are then normalized to be a probability distribution. By freezing the emission probabilities, we force the HMM to construct a model that will respect our defined harmonic function classes and that is of comparable quality to the GMC. Thus, the HMM only had to estimate transition probabilities between the hidden states (using the provided data).

Now, we ask the HMM to simulate 200 chord progressions of length 20 and we plot the average and variance of the cumulative log-likelihood under all the models (the SMC, GMC, and HMM). These plots are shown in Figures 8 and 9 respectively.

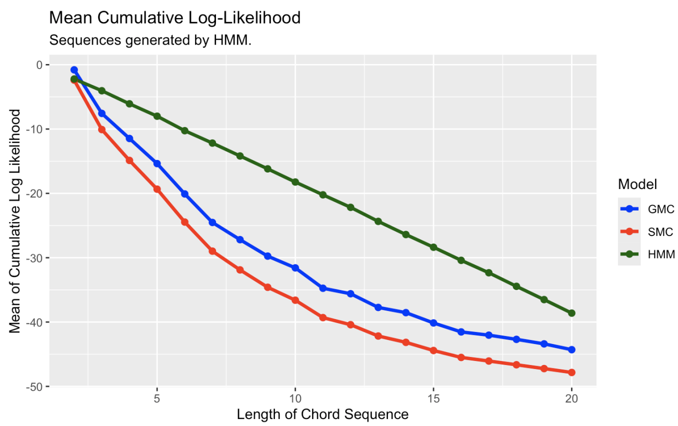

Figure 8: *Average Cumulative Log-Likelihood*

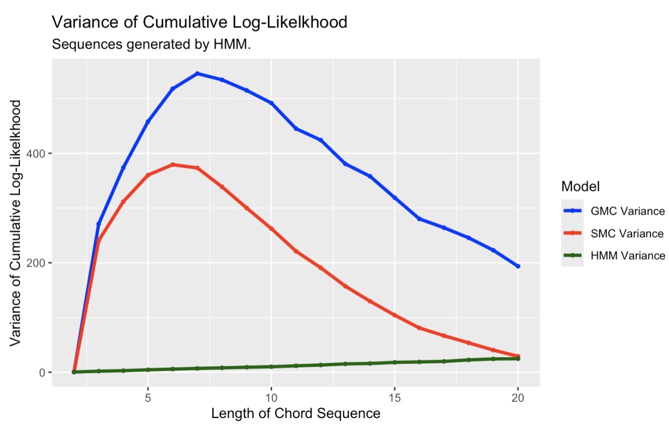

Figure 9: *Variance of Cumulative Log-Likelihood*

Notice how the average cumulative log-likelihood of both the GMC and SMC quickly dip below the average cumulative log-likelihood of the HMM. This is evidence that the HMM is generating chord sequences that both the GMC and SMC think are impossible. For example, the HMM could generate the sequence V-vii^{o} if it decides to sample two consecutive Roman numerals each with dominant harmonic function. However, the transition probability matrix in the SMC assigns this transition probability 0. Since the GMC also uses the same transition probability matrix as the SMC, it too finds this transition impossible. To avoid having to compute the mean and variance non-finite log-likelihoods, all non-finite values were replaced with a default floor value of -50. This is what causes the dip in the GMC and SMC curves. In the variance plot, this causes an initial spike in the variance of the SMC and GMC curves as a few of the 200 simulation runs encounter one of these invalid chord progressions. Eventually, most simulation runs encounter an invalid chord progression and the variance begins to decrease. While this simulation demonstrates that the HMM is getting creative with its chord progressions, this is not a desirable trait for a model that is trying to emulate J.S. Bach's style. This does not imply that the HMM is intrinsically flawed, but rather that unconstrained probabilistic models are misaligned with the strongly rule-governed nature of functional tonal harmony. In the harmonic analyses used in this study, V–vii^{o} does not occur as a functional progression at the phrase level, and its appearance typically reflects local voice-leading rather than harmonic succession. Tonal harmony tells us that the proper resolution of a V chord is I (or vi if we want to create a deceptive cadence). The progression V-vii^{o} does not give us the harmonic resolution that is expected of tonal music. This leads us to believe this transition probability should be 0 (as used by the SMC and GMC). Even if one were to find instances of this progression in the vocal parts of a Bach chorale, it is unlikely this is the correct overall harmonic analysis when considering the presence of the instrumental parts. In order to have a model that accurately replicates the typical style of Bach, that model must only use chord progressions that make use of functional tonal harmony. Instances where the composer wishes to add non-functional harmony should be identified after a functional harmonic pattern has been established.

However, the progression V-vii^{o} is not the only invalid progression the HMM has non-zero probability of generating. A mistake that is even more serious is the progression IVp to IV. The HMM could generate this if it samples an elongation chord followed by a predominant chord. While it is certainly possible to have two consecutive subdominant chords, this specific progression (IVp to IV) represents a blatant destruction of the harmonic system we established. IVp was designed for the times when the subdominant chord is serving as an elongation chord whereas IV is reserved for the times when the subdominant chord is serving as preparation for a dominant chord. There are no instances in Bach chorales where such functions occur in direct sequence.

Furthermore, Figure 10 shows that when invalid sequences generated by the HMM are filtered out, the GMC still has a lower variance of cumulative log-likelihood.

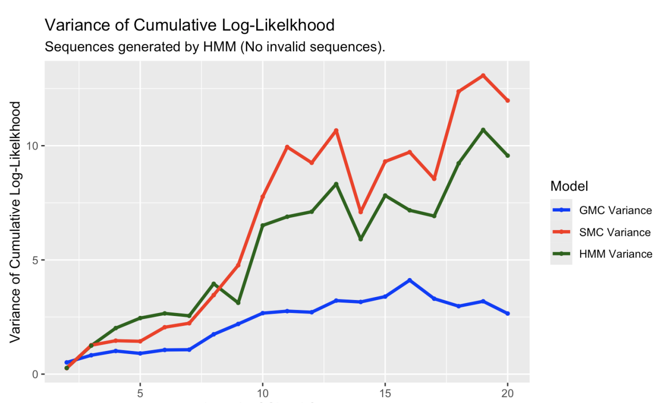

Figure 10: *Variance of Cumulative Log-Likelihood (Valid HMM sequences)*

5.3. Further Three Model Comparisons

In the previous section, we explained why the HMM may generate progressions the SMC and GMC find invalid. However, the GMC has superior performance (meaning lower variance of cumulative log-likelihood) over both the SMC and HMM when the HMM does not generate the chord progressions for which the log-likelihoods are computed. Figures 11 and 12 show the plots of the cumulative mean and variance when the SMC generates the chord progressions. In the plot of the mean cumulative log-likelihood we see the GMC curve is highest and the HMM has the lowest average cumulative log-likelihood. The HMM performs worst here because it has learned information about the harmonic functions which the SMC does not have. Since the SMC is just randomly moving through chords without any regard for musical form, the HMM scores these poorly. The GMC scores the SMC generated sequences with the highest log-likelihood out of all the models since it allows for exploratory harmonic motion consistent with the provided guide. This reflects the fact that the GMC conditions on harmonic intent rather than assuming strict cadential structure. This makes it robust to exploratory or wandering harmonic motion.

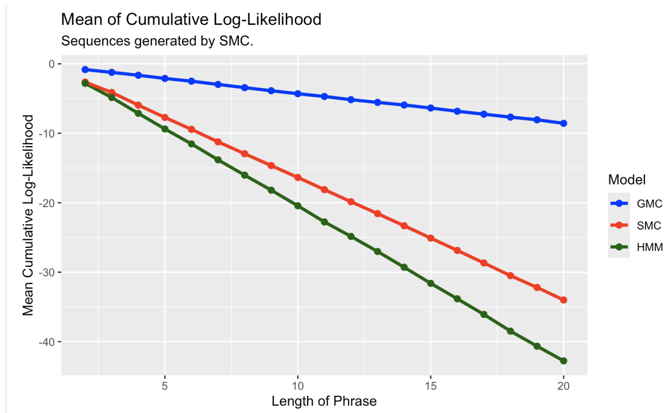

Figure 11: *Mean of Cumulative Log-Likelihood*

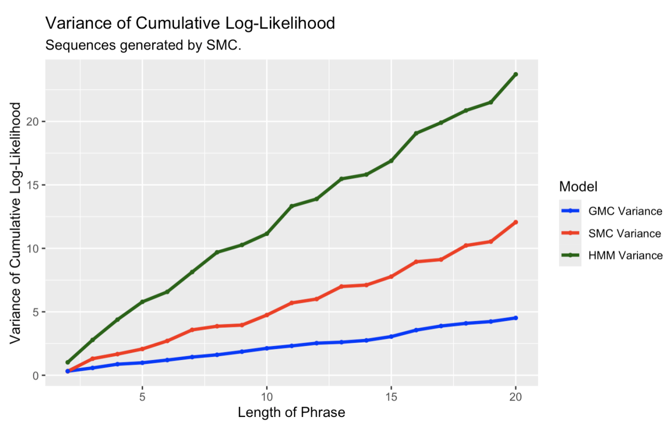

Figure 12: *Variance of Cumulative Log-Likelihood*

On the variance plot, we see the HMM has the highest variance. This model sees the most variance in the sequences generated by the SMC because by random chance the SMC generates sequences that the HMM scores with very high log-likelihood as well as sequences scored with very low log-likelihood.

6. GMC Stationary Distributions

It is well known that every finite irreducible and aperiodic Markov chain admits a unique stationary distribution. GMC models however, are generally time-inhomogeneous because their transition probabilities depend on an externally specified guide sequence. Consequently, a GMC need not converge to a single stationary distribution. When the guide sequence is periodic, the appropriate notion of stationarity becomes periodic as well.

In this section, we analyze the long-term behavior of a Guided Markov Chain whose transition probabilities depend on a periodic sequence of harmonic-function classes. Although the GMC itself is time-inhomogeneous, the process observed every K steps is a homogeneous Markov chain with transition matrix

$$P_{\text{cycle}} = P_1 P_2 \cdots P_K.$$

We show that under natural structural conditions, the restriction of this cycle matrix to the valid states at the cycle boundary is ergodic, and consequently the GMC admits a unique periodic stationary sequence of distributions.

Lemma 1 (Ergodicity of the Cycle Matrix)**.** *Let S be the collection of all Roman numeral chords. Assume the following musically natural conditions:*

1. ***Every chord has a clear harmonic function.*** *Each chord belongs to exactly one harmonic function class. This ensures that the guide can meaningfully restrict chord choices based on harmonic function.*

2. ***The harmonic language is fully connected for internal phrase transitions.*** *Consider the directed graph whose edges represent all chord transitions that occur with positive probability in the baseline (unguided) transition matrix P_0. Assume this graph is strongly connected over the entire state space S. This means that starting from any chord in S, it is possible to reach any other chord in S through some sequence of musically valid transitions. In the language of Markov chains, this is equivalent to irreducibility of the unguided chain.*

3. ***Cycle-level connectivity and aperiodicity on the cycle boundary.*** *Let $S_K = \{C \in S : h(C) = H_K\}$ be the set of all chords whose harmonic function matches the final step of the guide. Assume the guide $G = (H_1, \ldots, H_K)$ is such that for any pair of chords $C_1, C_2 \in S_K$, there exists an integer $m \geq 1$ and a sequence of chords consistent with the guide over m full periods for which the transition probability from C_1 to C_2 is strictly positive. Furthermore, to ensure the chain is not perfectly cyclic, assume there exists at least one chord $C \in S_K$ that permits return paths to itself of lengths m_1 and m_2 (measured in full periods) where m_1 and m_2 are relatively prime.*

Let the guide repeat with period K, and let

$$P_{cycle} = P_1 P_2 \cdots P_K.$$

Then the restriction of P_{cycle} to the state space S_K, denoted $P_{cycle}|_{S_K}$, is an ergodic transition matrix (irreducible and aperiodic).

Proof. Assumption 2 ensures that the baseline transition matrix P_0 is irreducible, so the full harmonic language is connected when no guide is applied. The introduction of a guide may render individual matrices P_t reducible, since transitions are restricted to chords whose harmonic function matches the guide at time t. However, we are concerned with the long-term behavior of the chain over full harmonic periods, which is governed by the cycle matrix P_{cycle}. By definition, the final transition in P_{cycle} is governed by P_K, so any state outside of

$$S_K = \{C \in S : h(C) = H_K\}$$

cannot occur at the end of a full cycle and therefore receives zero probability mass under P_{cycle}. Therefore, we evaluate the Markov chain restricted to the state space S_K. Assumption 3 guarantees cycle-level connectivity within this boundary: for any pair of chords $C_1, C_2 \in S_K$, there exists an integer $m \geq 1$ such that the composition of guided transitions over m full periods permits a path from C_1 to C_2 with strictly positive probability. Equivalently, some power of the restricted cycle matrix satisfies

$$(P_{\text{cycle}}|_{S_K})^m(C_1, C_2) > 0.$$

Therefore, $P_{\text{cycle}}|_{S_K}$ is irreducible.

To establish aperiodicity, the second part of Assumption 3 guarantees the existence of return paths from at least one chord $C \in S_K$ to itself over coprime numbers of periods (m_1 and m_2). Consequently, the set of possible return times

$$\{n \geq 1 : (P_{\text{cycle}}|_{S_K})^n(C, C) > 0\}$$

contains m_1 and m_2. Since they are relatively prime, the greatest common divisor of this set is 1. In an irreducible Markov chain, aperiodicity of a single state implies aperiodicity of the entire chain. Since irreducibility and aperiodicity together imply ergodicity, the restricted cycle matrix $P_{\text{cycle}}|_{S_K}$ is ergodic. ■

Theorem 4 (Periodic Stationary Behavior of a Guided Markov Chain)**.** *Let $\{C_t\}$ be a Guided Markov Chain with a guide that repeats every K steps. Let P_t be the transition matrix at step t, and let $P_{cycle} = P_1 P_2 \cdots P_K$. Let S_K be the set of valid chords at step K. If the restricted cycle matrix $P_{cycle}|_{S_K}$ is ergodic, then the Guided Markov Chain has a unique sequence of probability distributions*

$$\pi^{(1)}, \pi^{(2)}, \ldots, \pi^{(K)},$$

such that

$$\pi^{(t+1)} = \pi^{(t)} P_t \quad \text{for } t = 1, \ldots, K-1, \qquad \text{and} \qquad \pi^{(1)} = \pi^{(K)} P_K$$

and P_t is the transition matrix from phase t to phase $t+1$, whose transition probabilities are dictated by the target harmonic function H_{t+1}. Thus the chain does not admit a single stationary distribution in the usual time-homogeneous sense. Instead, its stationary behavior is described by a periodic cycle of K distributions that repeat every K steps in synchrony with the guide.

Proof. By the previous lemma, the restricted matrix

$$P_{\text{cycle}}|_{S_K}$$

is ergodic. By standard Markov chain theory, every finite ergodic transition matrix has a unique stationary distribution. Let $\pi^{(1)}$ denote this stationary distribution on S_K, extended to the full state space by assigning zero probability to states outside S_K. Under this extension,

$$\pi^{(1)} P_{\text{cycle}} = \pi^{(1)}.$$

Now define the remaining phase distributions recursively using the guided transitions:

$$\pi^{(t+1)} = \pi^{(t)} P_t, \quad t = 1, \ldots, K-1.$$

Then

$$\pi^{(K)} = \pi^{(1)} P_1 P_2 \cdots P_{K-1}.$$

Multiplying by P_K gives

$$\pi^{(K)} P_K = \pi^{(1)} P_{\text{cycle}} = \pi^{(1)},$$

which establishes the periodicity.

Uniqueness follows because $\pi^{(1)}$ is unique, and each $\pi^{(t)}$ is uniquely determined from $\pi^{(1)}$ by forward iteration. ■

7. The Impact of a Guiding Variable on Typical Set Cardinality

After defining an appropriate notion of entropy for a Guided Markov Chain, we examine conditions under which the presence of a guiding variable reduces the cardinality of Shannon-type typical sets. We also introduce two confidence intervals: one to detect reductions in entropy for finite-length chord sequences, and another to assess whether a specific guide sequence is structurally valid.

7.1. Entropy for Standard and Guided Markov Chains

Entropy measures the average uncertainty or information content of a random variable, and is interpreted as the average number of bits required to encode typical outcomes drawn from its distribution.

Definition 12. *The entropy of a discrete random variable X with probability mass function $p(x)$ is*

$$H(X) = -\sum_{x} p(x) \log_2 p(x).$$

Let (C_n) be a time-homogeneous standard Markov chain with transition matrix P and stationary distribution π. The uncertainty in the next chord, given the current chord, is summarized by the entropy rate.

Definition 13 (Entropy rate of the SMC)**.** *The entropy rate of the standard Markov chain is*

$$H_{\text{SMC}} = H(C_{n+1} \mid C_n) = \sum_{i \in S} \pi_i \, H(P_i),$$

where P_i denotes the i-th row of P, and equivalently

$$H_{\text{SMC}} = -\sum_{i \in S} \pi_i \sum_{j \in S} \boldsymbol{P}(j \mid i) \, \log_2 \boldsymbol{P}(j \mid i).$$

Thus, H_{SMC} gives the average harmonic uncertainty of the next chord when only the current chord is known.

In the guided model, the next chord is chosen from a restricted set of possibilities consistent with the harmonic function supplied by the guide. Let $\mu^{(1)}$ denote the initial distribution of C_1. For each harmonic function H_t specified by the guide, the underlying transition probabilities are renormalized to form the guided transition matrix $P^{(H_t)}$ defined by

$$P_{ij}^{(H_t)} = \mathbf{P}_{\text{GMC}}(j \mid i; \, H_t).$$

Because the Guided Markov Chain (GMC) is generally time-inhomogeneous, the time-n marginal distribution of the chain, $\mu^{(n)}$, evolves according to the sequence of guided transition matrices:

$$\mu^{(n)} = \mu^{(1)} P^{(H_2)} P^{(H_3)} \cdots P^{(H_n)}.$$

However, when the guiding sequence is periodic with period K, the chain converges to a periodic sequence of distributions $\pi^{(1)}, \dots, \pi^{(K)}$. In this case, the entropy rate can be defined over one full cycle of the guide. Here $\pi^{(t)}$ denotes the stationary distribution of the chain conditioned on being at phase t, i.e. immediately before applying the transition matrix $P^{(H_{t+1})}$.

Definition 14 (Entropy rate of the GMC)**.** *Assume the guiding sequence $\{H_n\}$ is periodic with period K, and that the GMC converges to a corresponding periodic sequence of distributions $\pi^{(1)}, \dots, \pi^{(K)}$. The entropy rate of the Guided Markov Chain is defined as*

$$H_{\text{GMC}} = \frac{1}{K} \sum_{t=1}^{K} \left[-\sum_{i \in S} \pi_i^{(t)} \sum_{j \in S} \mathbf{P}_{\text{GMC}}(j \mid i;\, H_{t+1}) \, \log_2 \mathbf{P}_{\text{GMC}}(j \mid i;\, H_{t+1}) \right].$$

H_{GMC} quantifies the long-term average uncertainty of the guided process, incorporating both the restrictions imposed by the harmonic function at each step and the periodic stationary behavior of the chain. While H_{GMC} can be computed when the full transition structure and periodic stationary distributions are known, it cannot be consistently estimated from a single finite musical realization (except in overly simplistic harmonic systems or under highly restrictive guides).

Given a chord sequence $\tilde{C}^{(n)}$ generated under a fixed guide G, we therefore consider the normalized negative log-likelihood

$$\hat{H}_n^{\text{GMC}} = -\frac{1}{n} \log_2 \mathbf{P}_{\text{GMC}}(\tilde{C}^{(n)} \mid G)$$

as an empirical estimate of the entropy rate. By the Shannon–McMillan–Breiman theorem for periodic Markov chains, $\hat{H}_n^{\text{GMC}}$ converges in probability to H_{GMC} as $n \to \infty$ [15].

Once we actually apply the guide and renormalize the probabilities to form the GMC, the resulting stochastic process may exhibit either higher or lower long-term entropy than the SMC. In other words, the guide tells us more information about the SMC's next move but the new guided process created from that information need not be more predictable than the SMC. In Appendices E and F, we present some simple numerical examples that illustrate this point.

8. Typical Set Cardinality Reduction

Entropy provides a local and asymptotic measure of uncertainty in a stochastic process, but it does not directly describe the size of the set of sequences that the process is likely to produce. In applications to music generation, this distinction is particularly important since a model may have high entropy at individual steps while still concentrating its probability mass on a comparatively small collection of musically plausible chord progressions. The concept of a typical set makes this idea precise by characterizing the collection of sequences that carry almost all of the probability mass of a stochastic process. Cover and Thomas [14] describe the following definition of a typical set.

Definition 15. *Let $\{X_n\}$ be a stationary stochastic process with entropy rate H. The ϵ-typical set $A_\epsilon^{(n)}$ is defined as the set of sequences $(x_1, \dots, x_n)$ satisfying*

$$2^{-n(H+\epsilon)} \le p(x_1, \dots, x_n) \le 2^{-n(H-\epsilon)}.$$

We now present an important result, the Typical Set Cardinality Reduction Theorem:

Theorem 5 (Typical Set Cardinality Reduction)**.** *Let $\{C_n\}$ be a stationary Markov chain modeling chord progressions with entropy rate H_{SMC}, and let $\{C_n\}$ be guided by a harmonic guide to form a guided Markov chain with entropy rate H_{GMC}. If*

$$H_{\text{GMC}} < H_{\text{SMC}},$$

then for every $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \frac{\left|A^{(n)}_{\epsilon,\text{GMC}}\right|}{\left|A^{(n)}_{\epsilon,\text{SMC}}\right|} = H_{\text{GMC}} - H_{\text{SMC}} < 0.$$

Consequently, the harmonic guide exponentially reduces the number of typical musical progressions produced by the model.

Proof. Assume the Standard Markov Chain (SMC) is stationary and ergodic. By the Shannon-McMillan-Breiman (SMB) theorem for time-homogeneous Markov chains [14], the probability of a sequence $C^{(n)} = (C_1, \ldots, C_n)$ satisfies, for almost every realization:

$$-\frac{1}{n} \log_2 \mathbf{P}_{\text{SMC}}(C^{(n)}) \xrightarrow[n\to\infty]{\text{a.s.}} H_{\text{SMC}},$$

where H_{SMC} is the entropy rate of the SMC. Consequently, by the Asymptotic Equipartition Property (AEP), the typical set $A^{(n)}_{\epsilon,\text{SMC}}$ has a cardinality satisfying

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \left|A^{(n)}_{\epsilon,\text{SMC}}\right| = H_{\text{SMC}}.$$

We now establish the corresponding result for the GMC. Let the guiding sequence $\{H_n\}$ be periodic with period K. Because the GMC is time-inhomogeneous, we cannot apply the SMB theorem directly to the step-by-step transitions. Instead, we group the process into blocks of length K.

Define the augmented state process $X_n = (C_n, T_n)$, where $T_n = n \bmod K$ tracks the phase of the guide, and define the block process

$$Y_m = \left(X_{mK+1}, X_{mK+2}, \ldots, X_{mK+K}\right), \quad m \geq 0.$$

Here, Y_m represents the evolution of the system over the m-th cycle. As established previously, the restricted cycle-level transition matrix $P_{\text{cycle}}|_{S_K}$ is ergodic. Consequently, the block process $\{Y_m\}$ is a finite-state, time-homogeneous Markov chain governed by the ergodic cycle-level transition matrix $P_{\text{cycle}}|_{S_K}$. When initialized in its unique stationary distribution, the block process is stationary and ergodic.

Applying the SMB theorem to the restricted block process $\{Y_m\}$ and dividing by K to scale back to individual time steps, we obtain that for almost every realization drawn from the stationary periodic regime:

$$-\frac{1}{n} \log_2 \mathbf{P}_{\text{GMC}}(C^{(n)}) \xrightarrow[n\to\infty]{\text{a.s.}} H_{\text{GMC}},$$

where the entropy rate is evaluated over the periodic stationary distributions $\{\pi^{(t)}\}$:

$$H_{\text{GMC}} = \frac{1}{K} \sum_{t=1}^{K} \left[-\sum_{i\in S} \pi_i^{(t)} \sum_{j\in S} \mathbf{P}_{\text{GMC}}(j \mid i;\, H_{t+1}) \log_2 \mathbf{P}_{\text{GMC}}(j \mid i;\, H_{t+1}) \right].$$

Consequently, the typical set $A_{\epsilon,\text{GMC}}^{(n)}$ satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \left| A_{\epsilon,\text{GMC}}^{(n)} \right| = H_{\text{GMC}}.$$

We now compare the sizes of the typical sets associated with the SMC and the GMC. By the AEP, for any $\epsilon > 0$ and sufficiently large n, the typical sets are bounded by:

$$2^{n(H_{\text{SMC}}-\epsilon)} \leq \left| A_{\epsilon,\text{SMC}}^{(n)} \right| \leq 2^{n(H_{\text{SMC}}+\epsilon)}$$

and

$$2^{n(H_{\text{GMC}}-\epsilon)} \leq \left| A_{\epsilon,\text{GMC}}^{(n)} \right| \leq 2^{n(H_{\text{GMC}}+\epsilon)}.$$

Taking the ratio of these bounds yields

$$2^{-n(H_{\text{SMC}}-H_{\text{GMC}}+2\epsilon)} \leq \frac{\left| A_{\epsilon,\text{GMC}}^{(n)} \right|}{\left| A_{\epsilon,\text{SMC}}^{(n)} \right|} \leq 2^{-n(H_{\text{SMC}}-H_{\text{GMC}}-2\epsilon)}.$$

Letting $n \to \infty$ and then $\epsilon \to 0$ gives

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \frac{\left| A_{\epsilon,\text{GMC}}^{(n)} \right|}{\left| A_{\epsilon,\text{SMC}}^{(n)} \right|} = H_{\text{GMC}} - H_{\text{SMC}}.$$

In particular, whenever $H_{\text{GMC}} < H_{\text{SMC}}$, the ratio $\left| A_{\epsilon,\text{GMC}}^{(n)} \right| / \left| A_{\epsilon,\text{SMC}}^{(n)} \right|$ decays exponentially fast in n. This shows that the harmonic guide exponentially reduces the cardinality of the typical set of chord progressions. ■

Remark 1. *The preceding theorem concerns Shannon typical sets and therefore requires a comparison of entropy rates (including the periodic entropy rate of the GMC). Independently of this result, the construction of a non-trivial guided Markov chain necessarily reduces the combinatorial space of admissible chord progressions, since the guide removes at least one feasible transition at each step. This combinatorial restriction holds regardless of whether the entropy rate of the guided process is greater or less than that of the original Markov chain.*

This Typical Set Cardinality Reduction Theorem is significant because it shows us that the guide does not merely bias individual chord choices, but dramatically narrows the space of long-term harmonic trajectories that the model is likely to realize. The exponential reduction in typical-set cardinality implies that as sequence length increases, the guided model concentrates probability mass on a dramatically smaller family of harmonic trajectories.

From the perspective of music generation, this provides a theoretical explanation for why guided models produce more stylistically coherent results: rather than wandering through a large space of loosely constrained chord sequences, the model is effectively restricted to a tightly controlled subset of progressions consistent with the composer's harmonic plan. This leads to generated sequences that more reliably exhibit features such as cadential structure and phrase-level organization. Furthermore, this theorem shows that guide sequences providing larger entropy reductions correspond to stronger stylistic constraints. This allows us to compare different harmonic frameworks in terms of how tightly they restrict the space of possible progressions.

9. A Monte Carlo Confidence Interval

While closed-form entropy expressions describe asymptotic behavior, they do not directly quantify how strongly a guide restricts the space of realizable musical sequences at finite lengths. To assess the practical impact of guidance on typical-set cardinality while accounting for stochasticity in both the unguided and guided processes, we construct a Monte Carlo confidence interval for the entropy reduction.

Fix a standard Markov chain (SMC), a non-trivial guide G, and a sequence length n. For a chord sequence

$$C^{(n)} = (C_1, C_2, \ldots, C_n),$$

we define the empirical entropy rate

$$\hat{H}_n(C^{(n)}) = -\frac{1}{n} \log_2 \mathbf{P}(C^{(n)}),$$

where $\mathbf{P}(\cdot)$ is evaluated using the transition probabilities of the model that generated the sequence. By the Shannon–McMillan–Breiman Theorem, $\hat{H}_n(C^{(n)})$ converges in probability to the entropy rate of the underlying process as $n \to \infty$, for both the SMC and the GMC .

Monte Carlo Sampling: We generate B independent SMC realizations

$$C_1^{(n)}, \ldots, C_B^{(n)},$$

each of length n, and compute their empirical entropy rates

$$\hat{H}_n^{\mathrm{SMC},(b)} = -\frac{1}{n} \log_2 \mathbf{P}_{\mathrm{SMC}}(C_b^{(n)}), \qquad b = 1, \ldots, B.$$

Simultaneously, we generate B independent GMC realizations using the same guide G:

$$\tilde{C}_1^{(n)}, \ldots, \tilde{C}_B^{(n)},$$

and compute their empirical entropy rates

$$\hat{H}_n^{\mathrm{GMC},(b)} = -\frac{1}{n} \log_2 \mathbf{P}_{\mathrm{GMC}}(\tilde{C}_b^{(n)}; G), \qquad b = 1, \ldots, B.$$

For each replication, define the empirical entropy reduction

$$\Delta_n^{(b)} = \hat{H}_n^{\mathrm{SMC},(b)} - \hat{H}_n^{\mathrm{GMC},(b)}.$$

Target Parameter:

$$\mu_{\Delta_n} := \mathrm{E}[\Delta_n]$$

where $\Delta_n = \hat{H}_n(C^{(n)}) - \hat{H}_n(\tilde{C}^{(n)})$. This parameter quantifies the difference in entropy between the guided and unguided processes. Note that the expectation is taken over independent chord realizations generated by the SMC and GMC.

The quantities $\{\Delta_n^{(b)}\}_{b=1}^{B}$ are independent realizations of the random variable Δ_n. The sample mean

$$\hat{\mu}_{\Delta_n} = \frac{1}{B} \sum_{b=1}^{B} \Delta_n^{(b)}$$

provides a Monte Carlo estimator of the target parameter $\mu_{\Delta_n} = \mathbb{E}[\Delta_n]$.

Monte Carlo Confidence Interval: Let $\{\Delta_n^{(1)}, \cdots, \Delta_n^{(B)}\}$ denote the empirical observed entropy differences and let the empirical distribution function be given by

$$\hat{F}_B(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}_{\{\Delta_n^{(b)} \leq x\}}.$$

For $\alpha \in (0, 1)$, define the empirical α-quantile as

$$q_\alpha = \inf\{x \in \mathbb{R} : \hat{F}_B(x) \geq \alpha\}.$$

Thus, we arrive at the $100(1-\alpha)\%$ one-sided Monte Carlo confidence interval for μ_Δ:

$$[q_\alpha, \infty).$$

Interpretation: If $q_\alpha > 0$, we have evidence that the GMC yields a strictly lower entropy rate than the unguided process for sequences of length n. This would mean that the guide causes a reduction in the cardinality of Shannon-type typical sets for finite-length chord sequences.

Remarks:

- By simulating both the SMC and GMC sequences, this confidence interval accounts for the stochastic variability of both the SMC and the GMC.
- This interval is constructed directly from the empirical distribution of entropy differences and does not rely on asymptotic normality.
- For this work, we used $B = 1000$.

10. Permutation-Based Confidence Interval for Guide Ordering

While the Monte Carlo confidence interval quantifies the magnitude of entropy reduction under a fixed guide, it does not address whether or not the ordering of harmonic functions contributes meaningful structural constraint. In order to assess this, we will construct a permutation-based 2-sided confidence interval.

Let

$$\vec{G}_{\text{true}} = (g_1, g_2, \ldots, g_L), \qquad g_t \in \mathcal{F},$$

denote a harmonic guide of length L, where $\mathcal{F}$ is the set of harmonic functions. We fix the initial function g_1 and consider permutations of the remaining positions.

For any guide $\vec{G}$, the empirical entropy rate of a chord sequence C generated by the Guided Markov Chain (GMC) under $\vec{G}$ is

$$\hat{H}_L(C; \vec{G}) = -\frac{1}{L} \sum_{t=2}^{L} \log_2 \mathbf{P}_{\text{GMC}}(c_t \mid c_{t-1}; g_t),$$

where $\mathbf{P}_{\text{GMC}}$ is the renormalized transition probability given the guide.

Monte Carlo Entropy Estimate: We generate M independent sequences

$$C_{\text{true}}^{(1)}, \ldots, C_{\text{true}}^{(M)}$$

from the GMC conditioned on $\vec{G}_{\text{true}}$, with the initial chord fixed at c_1. We observe the mean entropy across these realizations:

$$\bar{H}(\vec{G}_{\text{true}}) = \frac{1}{M} \sum_{m=1}^{M} \hat{H}_L(C_{\text{true}}^{(m)}; \vec{G}_{\text{true}}).$$

Permutation Sampling Next, we generate N_{perm} permuted guides $\vec{G}^{(k)}$ by randomly shuffling positions 2 through L of the original guide (keeping g_1 fixed). For each permuted guide $\vec{G}^{(k)}$, we attempt to generate M new independent sequences to compute the mean entropy

$$\bar{H}(\vec{G}^{(k)}) = \frac{1}{M} \sum_{m=1}^{M} \hat{H}_L(C^{(m,k)}; \vec{G}^{(k)}).$$

.

Handling Structurally Invalid Permutations: Because the GMC strictly follows the underlying transition matrix, a randomly permuted guide may frequently force a transition with zero probability. Such paths represent dead ends. In practice, if a permuted guide $\vec{G}^{(k)}$ fails to produce any valid sequences across the simulation attempts, it is deemed structurally non-viable. To correctly penalize this mathematical impossibility, we define its empirical entropy as infinite:

$$\bar{H}(\vec{G}^{(k)}) = \infty.$$

Empirical Distribution and Confidence Interval: Let $\{\bar{H}(\vec{G}^{(1)}, \cdots, \bar{H}(\vec{G}^{(N_{\text{perm}})}\}$ denote the empirical distribution of entropies for the permuted guides. We define the empirical distribution function

$$\hat{F}_{\text{perm}}(x) = \frac{1}{N_{\text{perm}}} \sum_{k=1}^{N_{\text{perm}}} \mathbf{1}_{\{\bar{H}(\vec{G}^{(k)}) \le x\}}.$$

For $\alpha \in (0, 1)$, define the empirical quantiles as

$$q_{\alpha/2} = \inf\{x : \hat{F}_{\text{perm}}(x) \ge \alpha/2\}, \qquad q_{1-\alpha/2} = \inf\{x : \hat{F}_{\text{perm}}(x) \ge 1 - \alpha/2\}.$$

A two-sided $100(1-\alpha)\%$ confidence interval for the entropy under random permutations is given by

$$\text{CI}_{1-\alpha}^{\text{perm}} = [\, q_{\alpha/2},\ q_{1-\alpha/2}\,].$$

Interpretation: This confidence interval quantifies the range of entropy values expected under random permutations of a given guide. If the true average entropy lies outside this interval,

$$\bar{H}(\vec{G}_{\text{true}}) \notin [q_{\alpha/2}, q_{1-\alpha/2}],$$

then the ordering of the original guide produces entropy that is atypical relative to random reorderings of the same harmonic material.

If the entropy lies above the upper bound, this indicates that the ordering produces unusually high entropy, suggesting that the arrangement does not efficiently organize the underlying harmonic functions. Conversely, if it lies below the lower bound, the ordering produces unusually low entropy, indicating a particularly strong structuring effect beyond what is typical among permutations.

Remark:

- For this work, we used $N_{\text{perm}} = 500$ and $M = 100$.

11. Empirical Comparison of Confidence Intervals

To illustrate the complementary roles of the Monte Carlo and permutation-based confidence intervals, we consider several representative harmonic guides and report their corresponding intervals under both procedures.

Due to the structural properties of the GMC, even relatively weak guides tend to reduce entropy compared to the SMC. To avoid overinterpreting such effects, we construct 99.9% one-sided Monte Carlo confidence intervals. Only those intervals whose lower confidence bound exceeds 0 provide evidence that the corresponding harmonic guide creates a nontrivial reduction in entropy.

In contrast, the permutation-based confidence interval evaluates a more subtle property: whether the ordering of harmonic functions produces entropy that is atypical relative to random permutations of the same collection of functions. Since this effect is inherently weaker and subject to Monte Carlo variability, we construct 90% two-sided confidence intervals.

Table 5 summarizes the results for three representative guides.

Guide	Monte Carlo C.I.	Permutation C.I.	$\bar{H}(\vec{G}_{\text{true}})$	Interpretation
$T-e-T-e-T-e$	$[0.6377, \infty)$	$[0.5020, 0.6689]$	0.7400	Heavily constrains the state space; the ordering does not provide additional structure relative to permutations and in fact yields higher entropy.
$T-d-T-d$	$[-0.0189, \infty)$	$[0.3989, 0.4965]$	0.7937	Fails to significantly reduce entropy since the lower bound is negative; reflects the behavior of the unguided chain, indicating that the ordering provides no distinct structural significance.
$T-e-p-P-d-T$	$[0.7009, \infty)$	$[0.3333, 0.7251]$	0.4227	Strong entropy reduction relative to the SMC; observed ordering yields entropy within the range expected under permutations of the same harmonic functions.

Table 5: *Comparison of Monte Carlo and permutation-based confidence intervals for representative harmonic guides.*

These examples highlight that the two confidence intervals capture distinct aspects of harmonic structure.

The Monte Carlo confidence interval measures the extent to which a guide reduces uncertainty relative to the SMC. If the lower bound is strictly above 0, this indicates that the guide restricts the space of realizable sequences, leading to more predictable outcomes.

In contrast, the permutation-based confidence interval evaluates whether the *specific ordering* of harmonic functions contributes explanatory power. If the true average entropy lies below the lower bound, this indicates that the observed ordering produces entropy that is unusually low relative to random permutations, and is therefore structurally meaningful in the sense of increasing predictability. If the true entropy lies within the interval, the ordering is consistent with random rearrangements of the same collection of harmonic functions. If the true average entropy lies above the upper bound, this indicates that the observed ordering produces entropy that is unusually high relative to random permutations, suggesting that the ordering disrupts structural regularity rather than reinforcing it.

Guide $T-e-p-P-d-T$ provides evidence for the effectiveness of both confidence intervals: it produces a substantial reduction in entropy and exhibits a non-random ordering. Figure 13 shows the Monte Carlo confidence intervals obtained from 1000 resamples. We observe that all lower bounds lie well above 0, indicating consistent entropy reduction relative to the SMC baseline.

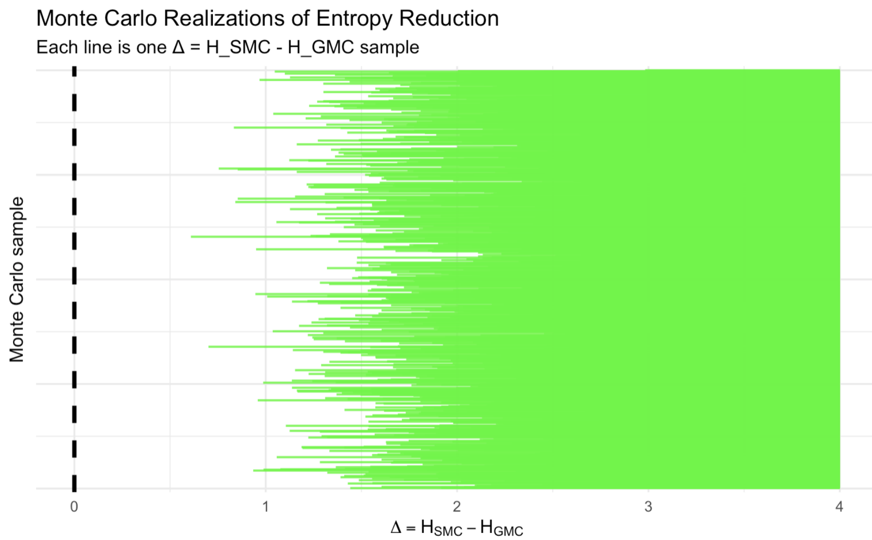

Figure 13: *1000 Monte Carlo Confidence Intervals for Guide $T-e-p-P-d-T$*

Additionally, Figure 14 displays the empirical CDF of entropy for this same guide. This plot shows the position of the observed entropy relative to the distribution of entropies for random permutations of the same harmonic functions.

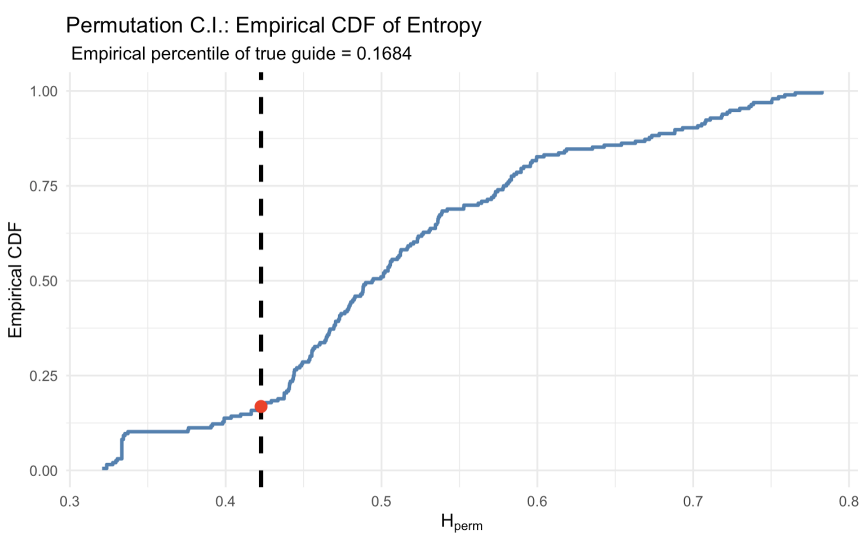

Figure 14: *Empirical CDF of Entropy for Guide $T-e-p-P-d-T$*

In contrast, the guide $T-e-T-e-T-e$ demonstrates that entropy reduction alone does not imply ordering significance. While this guide constrains the process, its entropy is much higher than that of random permutations of the same harmonic functions. Figure 15 shows the empirical CDF of entropy for this highly repetitive guide.

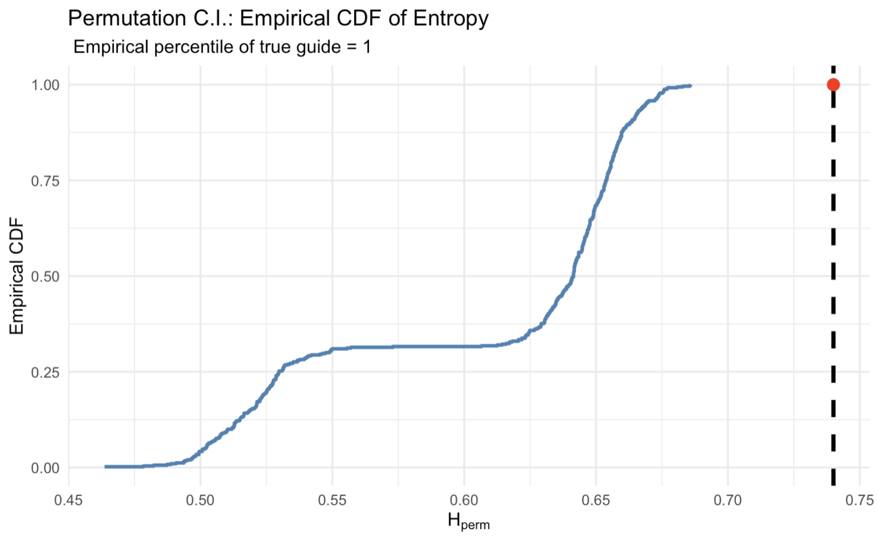

Figure 15: *Empirical CDF of Entropy for Guide* $T-e-T-e-T-e$

12. Impact of the Monte Carlo and Permutation-based Confidence Intervals

We are now in a position to isolate two complementary aspects of Bach's harmonic syntax.

First, Bach's harmonic syntax reduces the uncertainty of chord sequences relative to the SMC baseline, thereby restricting the space of realizable progressions and promoting stylistic coherence. This effect is quantified through a Monte Carlo confidence interval on the entropy reduction, which captures the variability of this constraint under resampling.

Second, Bach's syntax may or may not exhibit meaningful sequential structure beyond its harmonic content. This is evaluated using a two-sided permutation-based confidence interval, which compares the observed entropy to the distribution obtained from random permutations of the same harmonic functions. Deviations in either direction indicate that the ordering is unusual relative to this permutation collection.

Taken together, these procedures form a two-stage inference framework. The Monte Carlo confidence interval determines whether a guide imposes a statistically stable reduction in entropy relative to the SMC, while the permutation-based confidence interval determines whether the ordering of harmonic functions contributes additional structure beyond the unordered collection. This distinction allows us to separate mere constraint from genuine sequential organization in Bach chorales.

13. Chord Inversion Procedure

Now that we have discussed how GMC models generate a base chord progression (predominantly consisting of root position chords), we now describe the construction of musically practical chord progressions incorporating inversions. The first pre-processing step is to remove all dash characters that group Roman numerals and to replace all occurrences of IVp with IV. Note that removing dashes in a chord group such as V/V–V can produce a progression longer than the number of harmonic functions specified in the guiding sequence. Composers could, if desired, adjust the harmonic rhythm of specific chords to ensure that the extended progression fits within the guide's originally prescribed temporal structure. The assignment of inversions is probabilistic: for each chord, the inversion is sampled according to its observed relative frequency in a dataset of 90 Bach chorales, conditioned on the preceding chord and its inversion. Algorithm 8 (Appendix G) formalizes this procedure, showing how inversions are added to the base progression.

14. Conclusion and Future Research Directions

In summary, we have introduced the Guided Markov Chain (GMC) as a formal framework for incorporating an external guiding variable into a standard Markov chain. In the context of Baroque-style harmony, we demonstrated how the GMC can generate musically coherent chord progressions. We have showed that conditioning on a harmonic guide can reduce the cardinality of Shannon-type typical sets when the GMC has lower entropy than the underlying standard Markov chain (SMC). This provides a quantitative measure of how harmonic guidance narrows the space of likely chord sequences. Additionally, we developed confidence intervals to help us understand two complementary aspects of Bach's chord progressions.

Future research directions include:

- **Stylistic extensions (ongoing work):** Applying the GMC to generate chord progressions outside the Baroque style in order to evaluate how well the guiding variable adapts to new or expanded harmonic vocabularies.
- **Minor key analysis (ongoing work):** Assessing entropy reduction and ordering of harmonic functions in selected guides using the minor key expected likelihood matrix.
- **Empirical evaluation:** Conducting perceptual studies or user experiments to assess whether GMC sequences are judged as more coherent, stylistically appropriate, or aesthetically satisfying than unguided Markov chain sequences.
- **Sensitivity analysis:** Studying how sensitive the likelihood function of a Bach sequence is to changes in one of the expected likelihood matrices.

Overall, the GMC framework provides a versatile and stylistically grounded approach to the generation of chord progressions. By combining theory and practical applications, the GMC framework opens up a wide range of possibilities for algorithmic music composition and the quantitative study of musical structure.

15. Acknowledgments

The author would like to thank Wei-Min Huang, Darin Lewis, Taeho Kim, and Ping-Shi Wu for their guidance and feedback on this research.

A. Generation of SMC and GMC sequences

Algorithm 1 Standard Markov Chain (SMC) Sequence Generation

Require:

Transition Probability Matrix M (where $M_{ij} = \mathbf{P}(c_j \mid c_i)$)
Vocabulary of chords S
Starting chord $c_{\text{start}} \in S$
Desired sequence length n

Ensure:

Generated chord sequence $C = (c_1, c_2, \ldots, c_n)$

$c_1 \leftarrow c_{\text{start}}$ (Initialize the first chord)
for $t = 2$ to n **do**
  $c_{\text{prev}} \leftarrow c_{t-1}$ (Identify the previous chord)
  $\mathbf{p} \leftarrow M_{c_{\text{prev}},\cdot}$ (Extract the probability row vector for c_{prev})
  $c_{\text{next}} \leftarrow \text{Sample}(S, \mathbf{p})$ (Randomly sample next chord using $\mathbf{p}$)
  $c_t \leftarrow c_{\text{next}}$ (Append to sequence)
end for
return C

Algorithm 2 Guided Markov Chain (GMC) Sequence Generation

Require:

Transition Probability Matrix M (where $M_{ij} = \mathbf{P}(c_j \mid c_i)$)
Vocabulary of chords S
Set of harmonic functions $\mathcal{F}$
Harmonic Guide sequence $G = (g_1, g_2, \ldots, g_n)$ where $g_t \in \mathcal{F}$
Harmonic Function Map $h : S \to \mathcal{F}$ (maps each chord to its function)
Starting chord $c_{\text{start}} \in S$
Sequence length n

Ensure:

Generated chord sequence $C = (c_1, c_2, \ldots, c_n)$

$c_1 \leftarrow c_{\text{start}}$ (Initialize the first chord)
for $t = 2$ to n **do**
  $c_{\text{prev}} \leftarrow c_{t-1}$ (Identify the previous chord)
  $f_{\text{target}} \leftarrow g_t$ (Identify the target function from the Guide)
  $S_{\text{allowed}} \leftarrow \{s \in S \mid h(s) = f_{\text{target}}\}$ (Filter S for chords matching the target function)
  $\mathbf{p}_{\text{raw}} \leftarrow M_{c_{\text{prev}}, S_{\text{allowed}}}$ (Extract transition probabilities only for allowed chords)
  $Z \leftarrow \sum \mathbf{p}_{\text{raw}}$ (Calculate normalization constant (partition function))
  if $Z = 0$ **then**
    Error "Dead End: No valid transition to target function"
  end if
  $\mathbf{p}_{\text{norm}} \leftarrow \mathbf{p}_{\text{raw}} / Z$ (Renormalize probabilities to sum to 1)
  $c_{\text{next}} \leftarrow \text{Sample}(S_{\text{allowed}}, \mathbf{p}_{\text{norm}})$ (Randomly sample next chord)
  $c_t \leftarrow c_{\text{next}}$ (Append to sequence)
end for
return C

B. Algorithms for SMC and GMC Log-Likelihood Calculations

Algorithm 3 Standard Markov Chain (SMC) Log-Likelihood Calculation

Require:
  Transition Probability Matrix M (where $M_{ij} = \mathbf{P}(c_j \mid c_i)$)
  Observed chord sequence $C = (c_1, c_2, \ldots, c_n)$
Ensure:
  Log-Likelihood $\mathcal{L}$

$\mathcal{L} \leftarrow 0$ (Initialize total log-likelihood)
for $t = 2$ to n **do**
  $c_{\text{prev}} \leftarrow c_{t-1}$ (Identify the previous chord)
  $c_{\text{curr}} \leftarrow c_t$ (Identify the current chord)
  $p \leftarrow M_{c_{\text{prev}}, c_{\text{curr}}}$ (Lookup raw transition probability)
  if $p \leq 0$ **then**
    return $-\infty$ (Error: Transition impossible ($P = 0$))
  else
    $\mathcal{L} \leftarrow \mathcal{L} + \log(p)$ (Accumulate log probability)
  end if
end for
return $\mathcal{L}$

Algorithm 4 Guided Markov Chain (GMC) Log-Likelihood Calculation

Require:
  Transition Probability Matrix M (where $M_{ij} = \mathbf{P}(c_j \mid c_i)$)
  Vocabulary of chords S
  Set of harmonic functions $\mathcal{F}$
  Harmonic Guide sequence $G = (g_1, g_2, \ldots, g_n)$ where $g_t \in \mathcal{F}$
  Harmonic Function Map $h : S \to \mathcal{F}$ (maps each chord to its function)
  Observed chord sequence $C = (c_1, c_2, \ldots, c_n)$
Ensure:
  Log-Likelihood $\mathcal{L}$

$\mathcal{L} \leftarrow 0$ (Initialize total log-likelihood)
for $t = 2$ to n **do**
  $c_{\text{prev}} \leftarrow c_{t-1}$ (Identify the previous chord)
  $c_{\text{curr}} \leftarrow c_t$ (Identify the current chord)
  $f_{\text{target}} \leftarrow g_t$ (Identify the target function from the Guide)
  if $h(c_{\text{curr}}) \neq f_{\text{target}}$ **then**
    return $-\infty$ (Error: Observed chord violates the harmonic guide)
  end if
  $S_{\text{allowed}} \leftarrow \{s \in S \mid h(s) = f_{\text{target}}\}$ (Filter S for allowed chords)
  $Z \leftarrow \sum_{k \in S_{\text{allowed}}} M_{c_{\text{prev}}, k}$ (Calculate normalization constant)
  if $Z = 0$ **then**
    return $-\infty$ (Error: Dead End (No valid path to function))
  end if
  $p_{\text{raw}} \leftarrow M_{c_{\text{prev}}, c_{\text{curr}}}$ (Lookup raw transition probability)
  if $p_{\text{raw}} \leq 0$ **then**
    return $-\infty$ (Error: Transition impossible)
  end if
  $p_{\text{norm}} \leftarrow p_{\text{raw}} / Z$ (Renormalize probability)
  $\mathcal{L} \leftarrow \mathcal{L} + \log(p_{\text{norm}})$ (Accumulate log probability)
end for
return $\mathcal{L}$

C. HMM Sequence Generation and Log-Likelihood Calculation

Algorithm 5 Hidden Markov Model (HMM) Sequence Generation

Require:
Transition Matrix A (where $A_{ij} = \mathbf{P}(q_{next} = f_j \mid q_{prev} = f_i)$)
Emission Matrix B (where $B_{ik} = \mathbf{P}(c = s_k \mid q = f_i)$)
Vocabulary of chords S
Set of harmonic functions $\mathcal{F}$ (Hidden States)
Starting chord $c_{\text{start}} \in S$
Sequence length n
Ensure:
Generated chord sequence $C = (c_1, c_2, \ldots, c_n)$
$c_1 \leftarrow c_{\text{start}}$ (Initialize the first chord)
$\mathbf{e} \leftarrow B_{\cdot,c_1}$ (Get emission probabilities for the start chord across all states)
$q_1 \leftarrow \arg\max_{f \in \mathcal{F}} \mathbf{e}_f$ (Determine initial hidden state using emission matrix)
for $t = 2$ to n **do**
  $q_{\text{prev}} \leftarrow q_{t-1}$ (Identify the previous hidden state)
  $\mathbf{p}_{\text{trans}} \leftarrow A_{q_{\text{prev}},\cdot}$ (Extract transition row for previous state)
  $q_{\text{curr}} \leftarrow \text{Sample}(\mathcal{F}, \mathbf{p}_{\text{trans}})$ (Sample next hidden state)
  $\mathbf{p}_{\text{emit}} \leftarrow B_{q_{\text{curr}},\cdot}$ (Extract emission row for current state)
  $c_{\text{curr}} \leftarrow \text{Sample}(S, \mathbf{p}_{\text{emit}})$ (Sample next chord from emission distribution)
  $q_t \leftarrow q_{\text{curr}}$ (Update hidden state path)
  $c_t \leftarrow c_{\text{curr}}$ (Append to chord sequence)
end for
return C

Algorithm 6 Hidden Markov Model (HMM) Log-Likelihood Calculation (Forward Algorithm)

Require:
Transition Matrix A (where $A_{ij} = \mathbf{P}(q_{next} = f_j \mid q_{prev} = f_i)$)
Emission Matrix B (where $B_{ik} = \mathbf{P}(c = s_k \mid q = f_i)$)
Initial State Probabilities π (where $\pi_i = \mathbf{P}(q_1 = f_i)$)
Set of harmonic functions $\mathcal{F}$ (Hidden States) of size N
Observed chord sequence $C = (c_1, c_2, \ldots, c_n)$
Ensure:
Log-Likelihood $\mathcal{L}$
Initialization ($t = 1$):
for $j = 1$ to N **do**
  $\alpha_1(j) \leftarrow \pi_j \cdot B_{j,c_1}$ (Probability of starting in state f_j and observing c_1)
end for
Recursion ($t = 2$ to n):
for $t = 2$ to n **do**
  for $j = 1$ to N **do**
    $\sigma \leftarrow \sum_{i=1}^{N} \alpha_{t-1}(i) \cdot A_{ij}$ (Sum probability of all paths to state f_j)
    $\alpha_t(j) \leftarrow \sigma \cdot B_{j,c_t}$ (Update with emission probability of c_t)
  end for
end for
Termination:
$P_{\text{total}} \leftarrow \sum_{j=1}^{N} \alpha_n(j)$ (Sum probabilities of all valid paths)
if $P_{\text{total}} \leq 0$ **then**
  return $-\infty$ (Error: Sequence impossible under HMM)
else
  return $\log(P_{\text{total}})$
end if

D. Algorithm for Viterbi Training

Algorithm 7 Supervised Viterbi Training (Transition Learning)

Require:
Set of harmonic functions $\mathcal{F}$ (Hidden States) of size N
Fixed Emission Matrix B (determined by music theory)
Fixed Initial State Probabilities π
Training Corpus $\mathcal{O} = \{O^{(1)}, O^{(2)}, \ldots, O^{(K)}\}$ (set of chord sequences)
Convergence threshold ϵ

Ensure:
Learned Transition Matrix $\hat{A}$

Initialization:
Initialize A such that $A_{ij} = \frac{1}{N}$ for all $1 \leq i, j \leq N$ (Initialize uniform transitions)
$\delta \leftarrow 1$ (Smoothing constant (Laplace))
repeat
  $C \leftarrow \mathbf{0}_{N \times N}$ (Initialize transition count matrix with zeros)
  for $k = 1$ to K **do** (Iterate through all training chorales)
    $O^{(k)} \leftarrow \mathcal{O}^{(k)}$ (Select the k-th sequence from the corpus)
    $\hat{Q} \leftarrow \arg\max_Q \mathbf{P}(Q \mid O^{(k)}, A, B, \pi)$ (Decode optimal path using current A)
    $T_k \leftarrow \text{Length}(O^{(k)})$ (Length of sequence k)
    for $t = 1$ to $T_k - 1$ **do**
      $i \leftarrow \hat{q}_t$ (Index of current hidden state)
      $j \leftarrow \hat{q}_{t+1}$ (Index of next hidden state)
      $C_{ij} \leftarrow C_{ij} + 1$ (Tally observed transition)
    end for
  end for
  Initialize A' as $N \times N$ matrix
  for $i = 1$ to N **do**
    $Z_i \leftarrow \sum_{j=1}^{N} (C_{ij} + \delta)$ (Calculate row sum with smoothing)
    for $j = 1$ to N **do**
      $A'_{ij} \leftarrow (C_{ij} + \delta)/Z_i$ (Normalize counts to probabilities)
    end for
  end for
  $\Delta \leftarrow \max_{i,j} |A'_{ij} - A_{ij}|$ (Calculate maximum element-wise change)
  $A \leftarrow A'$ (Update transition matrix)
until $\Delta < \epsilon$ (Repeat until convergence)
return A

E. Numerical Entropy Examples for 4 Chord Models

In this section, we present numerical examples illustrating the two long-term entropy rates defined above.

Consider the following transition matrix P for a Standard Markov Chain (SMC) on four Roman numeral chords:

$$\begin{array}{c} \\ \text{I} \\ \text{ii} \\ \text{IV} \\ \text{V} \end{array} \begin{array}{c} \begin{array}{cccc} \text{I} & \text{ii} & \text{IV} & \text{V} \end{array} \\ \begin{bmatrix} 0 & 0.05 & 0.05 & 0.90 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0 & 0.5 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

We define the harmonic function mapping by

$$h(\text{I}) = T, \quad h(\text{ii}) = P, \quad h(\text{IV}) = P, \quad h(\text{V}) = D.$$

A direct computation shows that P^5 has no zero entries, so P is regular. The stationary distribution π satisfies

$$\begin{cases} \pi_{\text{I}} = 0.5\pi_{\text{IV}} + \pi_{\text{V}}, \\ \pi_{\text{ii}} = 0.05\pi_{\text{I}}, \\ \pi_{\text{IV}} = 0.05\pi_{\text{I}}, \\ \pi_{\text{V}} = 0.9\pi_{\text{I}} + \pi_{\text{ii}} + 0.5\pi_{\text{IV}}, \\ \pi_{\text{I}} + \pi_{\text{ii}} + \pi_{\text{IV}} + \pi_{\text{V}} = 1. \end{cases}$$

Solving yields

$$\pi = \left(\frac{40}{83}, \frac{2}{83}, \frac{2}{83}, \frac{39}{83}\right).$$

We now compute the entropy rate H_{SMC}. Using the convention $\lim_{p\to 0^+} p \log_2 p = 0$, we obtain:

$$H(P_{\text{I}}) \approx 0.5689, \quad H(P_{\text{ii}}) = 0, \quad H(P_{\text{IV}}) = 1, \quad H(P_{\text{V}}) = 0.$$

Thus,

$$H_{\text{SMC}} = \sum_{i\in S} \pi_i H(P_i) \approx \frac{40}{83}(0.5689) + \frac{2}{83}(1) \approx 0.2982 \text{ bits}.$$

In order to compute H_{GMC}, we consider the periodic guide (T, P, D).

Step 1: Periodic GMC Distributions

Solving for the periodic distributions $\pi^{(t)}$, we obtain:

$$\pi^{(1)} = (1, 0, 0, 0), \quad \pi^{(2)} = (0, 0.5, 0.5, 0), \quad \pi^{(3)} = (0, 0, 0, 1).$$

Step 2: Conditional Entropy per Step

We evaluate the inner sum of the H_{GMC} definition, representing the expected conditional entropy at each step t:

$$H^{(t)} = \sum_{s\in S} \pi^{(t)}(s)\, H(C_{n+1} \mid C_n = s;\, H_{t+1}).$$

Case $t = 1$ (T $\to$ P): From I, transitions to $\{\text{ii}, \text{IV}\}$ are equally likely:

$$H^{(1)} = -\big(0.5 \log_2 0.5 + 0.5 \log_2 0.5\big) = 1.$$

Case $t = 2$ (P $\to$ D): Both ii and IV transition deterministically to V:

$$H^{(2)} = 0.$$

Case $t = 3$ (D $\to$ T): V transitions deterministically to I:

$$H^{(3)} = 0.$$

Step 3: GMC Entropy Rate

Averaging over the period $K = 3$,

$$H_{\text{GMC}} = \frac{1}{3} \sum_{t=1}^{3} H^{(t)} = \frac{1}{3}(1 + 0 + 0) = \frac{1}{3} \approx 0.333 \text{ bits.}$$

To summarize:

$$H_{\text{SMC}} \approx 0.2982, \quad H_{\text{GMC}} \approx 0.333.$$

This example demonstrates that it is possible to construct a Guided Markov Chain whose entropy rate exceeds that of the original SMC.

The increase in entropy arises from the way the guide reshapes the transition structure. In particular, when the guide requires a predominant chord to follow a tonic chord, it removes the most likely successor under the SMC (the dominant chord, which occurs with probability 0.9) and restricts attention to the remaining predominant chords. After renormalization, this produces a more uniform distribution over the allowed chords, thereby increasing uncertainty. The GMC entropy H_{GMC} reflects the behavior of a new stochastic process in which the transition probabilities have been globally modified by the guide. As this example illustrates, these global changes can increase the long-term entropy even though each individual step becomes more constrained.

A contrasting example shows that the opposite behavior can also occur. Consider the transition matrix Q:

$$\begin{array}{c} \\ \text{I} \\ \text{ii} \\ \text{IV} \\ \text{V} \end{array} \begin{array}{c} \begin{array}{cccc} \text{I} & \text{ii} & \text{IV} & \text{V} \end{array} \\ \begin{bmatrix} 0 & 0.1 & 0 & 0.90 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0 & 0.5 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Under the same periodic guide $H_t = (P, D, T)$, the guided process becomes deterministic, yielding

$$H_{\text{GMC}} = 0.$$

For this matrix, the SMC entropy is

$$H_{\text{SMC}} \approx 0.2233.$$

Thus, depending on the interaction between the transition matrix and the guide, the GMC may have either higher or lower entropy than the original SMC.

F. Numerical Entropy Examples for Major Chord Transition Matrix

For the major chord transition matrix $\frac{L}{Z}$ discussed in Section 4.3, we can compute $H_{\text{SMC}} \approx 2.6310$ bits. Table 6 presents a collection of 6 GMC guides and the approximate resulting entropy of the GMC when we use each of these guides as one period. This is not intended to be an exhaustive listing of all possible guides for which $H_{\text{GMC}} < H_{\text{SMC}}$. Instead, we demonstrate the musical variety among the many guides obeying this inequality for our choice of $\frac{L}{Z}$.

Guide	H_{GMC}
T-d-T-P-d-T	0.4441 bits
T-d-T-p-d-T	0.6112 bits
d-T-d-T-d-T	0.8322 bits
T-c-T-f-P-d	0.2870 bits
T-e-T-f-p-d	0.4708 bits
T-p-D-T-e-c-p-d	0.5457 bits

Table 6: *Selected GMC Guides and their Approximate Entropies*

G. Algorithm for Chord Inversion Procedure

Algorithm 8 Chord Inversion Assignment for GMC Progressions

Require: Base chord progression $C = (c_1, c_2, \ldots, c_N)$, dataset of historical transitions $\mathcal{D}$
Ensure: Harmonized progression with inversions $\tilde{C} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_N)$

Initialize chord list $L_C \leftarrow [c_1]$
Initialize inversion list $L_I \leftarrow [53]$ (root position for the first chord)
for $i = 2$ **to** N **do**
  $c_{\text{prev}} \leftarrow L_C[-1]$, $i_{\text{prev}} \leftarrow L_I[-1]$
  $c_{\text{next}} \leftarrow c_i$
  if $c_{\text{next}} \in \{\text{I64}, \text{i64}\}$ **then**
    Append c_{next} to L_C
    Append 53 to L_I
    continue
  end if
  Define $\mathcal{D}_{\text{pair}} = \{(c_1, i_1, c_2, i_2) \in \mathcal{D} : c_1 = c_{\text{prev}}, i_1 = i_{\text{prev}}, c_2 = c_{\text{next}}\}$
  if $\mathcal{D}_{\text{pair}} \neq \emptyset$ **then**
    Compute probability distribution $P(i_2 \mid c_{\text{prev}}, i_{\text{prev}}, c_{\text{next}})$ from $\mathcal{D}_{\text{pair}}$
    Sample $i_{\text{next}} \sim P(i_2 \mid c_{\text{prev}}, i_{\text{prev}}, c_{\text{next}})$
  else
    $i_{\text{next}} \leftarrow 53$
  end if
  Append c_{next} to L_C and i_{next} to L_I
end for
Initialize $\tilde{C} \leftarrow \emptyset$
for $j = 1$ **to** N **do**
  $c \leftarrow L_C[j]$, $i \leftarrow L_I[j]$
  if $i = 53$ **then**
    $\tilde{c}_j \leftarrow c$
  else if c contains a slash **then**
    Insert i before the slash in c, set $\tilde{c}_j$ to the result
  else
    Concatenate c and i, set $\tilde{c}_j \leftarrow ci$
  end if
  Append $\tilde{c}_j$ to $\tilde{C}$
end for
return $\tilde{C}$

The first chord is always retained in root position. This procedure guarantees that any half-diminished chord receives a seventh and may additionally introduce sevenths to other chords (such as V or ii) in accordance with the empirical distribution observed in the dataset. It was decided that chord inversions would be added after the GMC creates a base progression in order to reduce the size of the transition matrix.

References

[1] Franklin, J. (2006). Recurrent Neural Networks for Music Computation. *INFORMS Journal on Computing*, v.18, n.3, pp.321–338.

[2] Herremans, D.; Chuan, C. H.; & Chew, E. (2017). A Functional Taxonomy of Music Generation Systems. *ACM Computing Surveys (CSUR)*, v.50, n.5, pp.1–30.

[3] Dobrow, R. (2016). Introduction to Stochastic Processes with R. Wiley.

[4] Zucchini, W.; MacDonald, I.; Langrock, R.(2016). Hidden Markov Models for Time Series. CRC Press.

[5] Hiller, L.; Isaacson, L. (1959). Experimental Music Composition with an Electronic Computer. McGraw-Hill.

[6] Xenakis, I. (1992). Formalized Music: Thought and Mathematics in Composition. Pendragon Press.

[7] Carvalho, H.T. (2019). An Introduction to Markov Chains in Music Composition and Analysis. *Brazilian Journal of Music and Mathematics*, v.3, n.2, pp.18–43.

[8] Paiement, J. F.; Eck, D.; & Bengio, S. (2005). A Probabilistic Model for Chord Progressions. *Proceedings of the Sixth International Conference on Music Information Retrieval*, pp.312–319.

[9] Tsushima, H.; Nakamura, E.; & Yoshii, K. (2020). Bayesian Melody Harmonization Based on a Tree-Structured Generative Model of Chord Sequences and Melodies, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v.28, pp.1644–1655.

[10] Pachet, F.; Roy, P. (2011). Markov Constraints: Steerable Generation of Markov Sequences. *Constraints*, v.16, n.2, pp.148–172.

[11] Pasquier, P; Eigenfeldt, A. (2010). Realtime Generation of Harmonic Progressions Using Controlled Markov Selection. *Proceedings of ICCC-X Computational Creativity Conference*.

[12] Bielecki, T.; Jakubowski, J.; Niewęgłowski, M. (2017). Conditional Markov chains: Properties, construction and structured dependence. *Stochastic Processes and their Applications*, v.127, n.4, pp.1125–1170.

[13] Jelinek, F. (1976). Continuous Speech Recognition by Statistical Methods. *Proceedings of the IEEE*, v.64, n.4, pp.532–556.

[14] Cover, T.; Thomas, J. (1991). Elements of Information Theory. Wiley Series in Telecommunications. John Wiley and Sons.

[15] Algoet, P.; Cover, T. (1988). A Sandwich Proof of the Shannon-McMillan-Breiman Theorem. *The Annals of Probability*, v.16, n.2, pp.899–908.