3.4 The Entropy of Markov chains

The probability of a given chord progression X is simply [4]

$$P(X) = p_1^{p_1 N} p_2^{p_2 N} \cdots p_n^{p_n N}$$

where N is the length of the progression and $p_i N$ is the number of occurrences of the ith chord in the sequence X. We measure the **entropy rate of a Markov chain** X by quantifying the entropy per chord $X_t = c_i$ in X

$$\begin{aligned} H(X) &= \sum_{X_t \in X} p(X_t) H(X_t) = -\sum_{i=0}^{m} \sum_{j=0}^{m} p(c_i) p(c_j|c_i) \log p(c_j|c_i) \\ &= -\sum_{i=0}^{m} p(c_i) \sum_{j=0}^{m} p(c_j|c_i) \log p(c_j|c_i), \end{aligned}$$

c_i, $c_j \in C$, the state space of X, whose cardinality $|C| = m$. This is the entropy per chord[5] because it multiplies each of the inner sums by the probability of observing the initial chord, and each of those probabilities ($p(c_i) = p_i$) is found by dividing the number of its occurrences ($p_i N$) by the total number of chords observed, N. Hence, the entropy of an entire Markov chain is just NH.

3.4.1 How to interpret this measure

Since the entropy of an entire Markov chain is NH, where N is the length of the sequence and H is the entropy of each chord, systems with more observations (larger N) will tend to have more entropy than just as chaotic systems with a smaller N. Therefore, in characterizing a system by its entropy, it is clearer to use simply H to describe it.

The entropy of a Markov chain has the same form as the entropy of two conditional events, because by the Markov property, we only consider two states in the calculation of transition probabilities. We already knew that a Markov chain was simply a sequence of conditional probabilities, so it should be unsurprising that its entropy is modeled after this conditional character.

[5]In fact, our entropy rate is also the "entropy per time interval." Since there is no rhythm associated with our Markov chain, these time intervals are likely not uniform (though it is certainly possible for a song to uniformly change over time).