Chopin's Nocturnes and Entropy

Benjamin Richards

Abstract

It is an oft-quoted fact that there is much in common between the fields of music and mathematics. This paper seeks to examine that commonality through the use of information theory techniques, specifically, entropy as defined by Shannon. The analysis will be presented through the process of creating an entropy table for a collection of Chopin's Nocturnes. The method for generating a random string of notes with respect to this entropy will also be briefly discussed.

1. Introduction

One of the most fascinating aspects of mathematics is its ability to describe so many different models with a relatively small number of techniques. In this vein of thinking I sought to further explore the application of these techniques to models from fields that have not commonly been thought of as scientific–models from the fine arts.

Given my own personal background in ballet, the original choices for this analysis were actually going to be the works of a choreographer. The mathematical approach was to be a statistical analysis yielding a probability distribution for each step (in the sense of a dance step) used in the ballet, and that would allow a rudimentary authentication of the choreographer's work to be approached. Choreography was soon abandoned in favor of musical compositions. The reason for this was the lack of well-defined discrete steps in any given piece of choreography. Though a reasonable argument could be made that choreography, particularly in the classical genre, is indeed composed of a finite number of discrete steps, the real problem lies in the breakdown of the larger work into these discrete steps. Such a process would be largely subjective and would introduce a considerable margin of error into the analysis. Thus musical compositions were chosen because they are readily available as scores consisting of a finite number of clearly defined components–music notes.

At the same time I decided that musical compositions would be the raw material for analysis, stronger and more specific mathematical techniques were adopted as well. Entropy, as defined by Shannon in "The Mathematical Theory of Communication", has already been used to great effect in authentication problems where works of literature were the subject of examination. It became clear at that point that the task at hand was to understand the relevant machinery of information theory and entropy, and then to adapt the method as needed to work with musical compositions.

2. Entropy

The definition of entropy is

$$H(X) = -\sum_{x \in \chi} p(x) \log p(x) \qquad (1.1)$$

where X is a discrete random variable taking on values from the set χ which depends on the information source being analyzed. For the purposes of this analysis, the elements of the set χ are all the musical notes used by a given composer in any of his compositions. If the analysis were to be one regarding the works of a famous literary author, the corresponding set would likely consist of all the words used by that author in his or her writings. The logarithm is base 2, which is important as it produces an entropy value with units of bits. This turns out to be a particularly useful because much of the work of information theory and entropy is relevant to the efficiency of coding schemes. Another important way of describing $H(X)$ is to say that it is the expected value of $\log p(x)$. There are many properties that arise from the manipulation of this definition that will not be treated here, but can be found in Cover and Thomas's *Elements of Information Theory*.

In the interest of adding an intuitive definition to the one above, we may think of entropy as the average number of yes or no questions needed to match a given X_i with an element from a set χ. This is a somewhat weak definition as it is important to note that we mean average number of questions needed in the most efficient scheme for asking yes or no questions if the answer is to be $H(X)$. In many cases the best possible scheme may not be known. Regardless, it is a useful definition for intuitive purposes. A particularly illuminating example is Shannon's finding of a value of 1.34 bits for the entropy of letters in the English language. What this means is that the bare minimum amount of information needed to identify each letter in a string of English text is 1.34 bits. Now this may seem counterintuitive because there are 26 letters in English and if we include the character "space" we have 27 symbols. It is quite obvious that any single symbol taken at random cannot definitely be identified by 2 yes or no questions, or 2 bits of information, much less 1.34 bits of information. Entropy though doesn't describe the uncertainty in a single independent symbol. It describes the uncertainty in the whole sample space in which the symbols occur, in this case, English. The entropy of English is as low as it is because the probability of a great many sequences of letters is zero. Likewise some letters are followed by all but one symbol with very low probability, "q" being a prime example as it is overwhelmingly followed by "u". That knowledge is due to the structure of English and so on average it takes less than 1.34 bits to identify the symbol following a "q". Also, after the first few letters of many words have been identified, the rest of the word can be guessed easily, again with much less information than 1.34 bits per symbol. As Cover and Thomas point out, "... a large number of guesses will usually be needed at the beginning of words or sentences." but that shouldn't be surprising considering those symbols are much more independent than are the symbols following them. Thus we may think of entropy as the average number of bits needed to identify a symbol that occurs with known probabilities and transition probabilities.

Though not made use of in my analysis due to time and labor constraints, it is worthwhile to mention the definition of conditional entropy. It is defined as

$$H(Y \mid X) = \sum_{x \in \chi} p(x) H(Y \mid X = x) \tag{1.2}$$

which shouldn't be too surprising considering it as the expected value of the entropy of Y given X. The reason I mention this definition as being important is because the strongest applications of entropy rely on the information given by a sequence of random variables as opposed to single random variables. Thinking back to the example using

English, it is the sequences of letters that we have the greatest intuition regarding the probability of occurrences.

Thus we see that there are only a few necessary pieces of information to apply the definition of entropy to a data set. In the case of Chopin's Nocturnes all that will be really needed is an alphabet of music notes and the probabilities with which they occur. We shall discuss the procedure in the following section.

3. The Musical Compositions

The first task regarding music was the selection of a composer. Some considerations had to be taken into account when identifying a composer with a plausible sampling of work. First and foremost, the composer needed to have a large number of works that are known to be their own. Without this stipulation, there would be no way to take a sampling large enough to discern the typical elements of that composer's work, thus an entropy from a small sampling would carry with it an uncertainty too large to be of much interest. Also we would hope for a composer of piano music or solo pieces for other instruments. The reason for this is again to provide a stronger entropy measure from the sampling. As more instruments are added to a score, the number of elements in χ increase substantially. This is because when one instrument is being played, fewer notes can be played in any time step. As more instruments are added many more combinations of notes can be played at any given time step. The problem caused by this is that it is much more likely that no single set of notes played within a time step will be repeated. We could of course choose to analyze only one instrument, but it is much more likely that the stylistic elements that make a composer unique are not broken up neatly into their treatment of each instrument but rather can be found in the instruments interactions and the melodic structure. Under that assumption we could of course try to follow only the melody of a score, but there would be instances in which that could in fact be a tall order. A prime example of this would be canonical structure in which multiple instruments all play the melody but are displaced in their timing much like the singing of "Row Row Row Your Boat" in a round where each singer starts the song after the first measure has passed. Thus we would prefer, at least in our first analysis, to tackle a set of composition with a small number of instruments. Chopin became a logical choice because he had a very prolific collection of works for solo piano. He was in fact so prolific that I was able to go a step further and use only his Nocturnes for a complete sample space.

Ninety-three pages of Chopin's Nocturnes for piano make up the sample space of my analysis. I analyzed only the right hand, or treble clef, of each stave considering it is the one that generally plays the melody. I chose this way as a means of simplifying the number of note combinations that would need to be recorded. The scheme for choosing notes at random is to chose a page at random, then a stave from that page, a measure from that stave, and a note from that measure, also all at random. All of this is done with a table of random numbers produced by random.org. In this analysis a note can be either a single tone with a corresponding length of time, or a chord with several tones and time values. The important factor in determining a note is where it begins. For instance a chord consisting of three tones, two held for a beat and one for two beats would be recorded as such if all three tones commence on the same beat. If however a chord with three tones were to begin with one tone and then have the other two added on a subsequent beat, they would be considered two different notes because they commenced at two different times. This of course is not musical theory or even notational custom; I merely needed to choose a method for handling such instances to be used systematically throughout the analysis. As I mentioned, it was not only the tone of notes that was recorded but also the length. Originally I had planned on recording whether each note was a whole, half, quarter, sixteenth note, et cetera. However, given that each of those notes would be held for different numbers of beats depending on the time signature of the music, I chose instead to record the number of beats each tone was held.

Concerning the actual set-up of the entropy table, it's little more than a spreadsheet with values for length and tone along the two axes. For the ease of recording I assigned the various possible tones numerical values by letting middle c equal 0 and adding or subtracting the number of half steps above or below to get every other note's respective numerical value. In other words, I used a chromatic scale and assigned it integer values with c being the center at 0.

4. Results

With a sample space of 439 randomly selected notes I had an alphabet of 274 distinct musical notes and a value of 7.71 bits for the entropy of the space. This shows that there is indeed identifiable structure in the music. An alphabet of this size would require 8.09 bits for entropy if all elements of the alphabet were equally probable. Likewise for equally probable elements 7.71 bits could only describe an alphabet of 209 elements.

5. Future Applications

With an entropy table now in hand the most exciting application immediately available would be to create randomly generated strings of notes according to the probabilities recorded during sampling. Such a task would be carried by partitioning a uniformly distributed interval from zero to one into 274 sub-intervals each with length equal to the probability of an element of the alphabet. It is then simply a matter of choosing random sequences of numbers along the original uniformly distributed interval and recording the elements they correspond to based on which partition they fell in. Such a string of elements would be a randomly generated sequence based upon the characteristics of Chopin's Nocturnes and should sound similar as compared to a sequence based on partitions of equal length.

References

[1] Cover, Thomas M. and Thomas, Joy A. *Elements of Information Theory*. 2nd Edition. Hoboken, New Jersey, John Wiley and Sons, 2006.
[2] Shannon, Claude E. and Weaver, Warren. *The Mathematical Theory of Communication.* Urbana, Chicago, London, University of Illinois Press, 1949.
[3] Ross, Sheldon. *A First Course in Probability.* 6th Edition. Upper Saddle River, New Jersey, Prentice-Hall, 2002.