

特定の条件下におけるアルファベットの生成とその生起確率等について

芝浦工業大学 数理科学研究会
BP16082 堀井雷生

平成 29 年 5 月 20 日

1 はじめに

最近の情報系の授業で情報分野における生起確率やエントロピーについて少しだけ学習したのでそれらについて、アルゴリズムやプログラミングを作成してみたいと思い、研究を始めた。現段階では情報分野やプログラミングに関する知識不足から、自分で決定したテーマに基づいて作成した簡易的なプログラムを動かし、実行結果や既存のほかのデータなどと比較して自分なりに考察を行っている。

2 前提知識

2.1 生起確率

ある情報集合の中で、ある情報が発生すると予測される確率のこと。基本的に高校数学 1 で習った確率と同様の意味であると考えるとよく、以下のように表す

$$(\text{生起確率}) = \frac{(\text{ある事象が起こった回数})}{(\text{全ての事象の回数})}$$

2.2 エントロピー

生起確率が $P(x)$ で表されるとき、以下のように定義される値 $I(x)$ のことを自己情報量といい、単位は [bit] を用いる。

$$I(x) = -\log_2 P(x)$$

また、複数の自己情報量がある場合、その情報量の平均 (期待値) \bar{I} は以下のように表され、これをエントロピーという

$$\bar{I} = -\sum P(x) \log_2 P(x)$$

2.3 モールス信号

Samuel F.B Morse が考案した、各英文字などを短時間の通電状態の「トン」、長期間の通電状態の「ツー」、絶縁状態の空白の組み合わせで表した電通用の文字符号のことである。各文字に対応する符号語を記してあるモールス符号表についてはここでは割愛する。

3 作成したプログラム

3.1 プログラム 1

モールス符号表の出現頻度をもとに各アルファベットの出現確率を計算し、その確率に基づいてアルファベットを出力する。

利点: 1 回の試行で 100%アルファベットが生成出来る

欠点: 確率が完全に固定されているのでアルファベットの散らかり具合がある程度一定になりやすい

3.2 プログラム 2

0 と 1 を使って、13 桁の文字列を作成し、それが 0 と 1 で表したモールス符号表に対応していれば出力する。さらに、以下の条件を付け加える

1 奇数番目の要素には 1 を代入する

2 文字列の長さが奇数の時、ランダムで文字列への代入を終了する

利点: ランダムに出力するのでいい感じにばらけてくれそう

欠点: 必ずしもアルファベットが生成されないので、試行回数にわりにあまり短い文章が出来そう

3.3 プログラム 3

プログラム 2 の条件 1 をなくした版、基本プログラム 2 と同じ

4 今後の課題

とっさに思いついて始めた研究であるが、あまり思考や下調べしていなかったために、目的であるはずのアルファベットの生成があまり出来ないものが出来上がってしまった。今後は出来上がったプログラミングと目的をより注意深く吟味しながら作成していきたい。

参考文献

- [1] 稲井 寛, はじめての情報理論, 森北出版株式会社, 2011
- [2] rand 関数による乱数生成について: <http://mkubara.com/wp/%e4%b9%b1%e6%95%b0%e3%81%ae%e7%94%9f%e6%88%90%ef%bc%88rand%e9%96%a2%e6%95%b0-srand%e9%96%a2%e6%95%b0%ef%bc%89/>
- [3] エントロピーの説明 <http://dic.nicovideo.jp/a/エントロピー>