

『計量国語学』アーカイブ

<b>ID</b>	<b>KK300803</b>
<b>種別</b>	解説
<b>タイトル</b>	データの視覚化(8) —GnuplotとIllustratorを用いて—
<b>Title</b>	Data Visualization (8): Application of Gnuplot and Adobe Illustrator
<b>著者</b>	古橋 翔・早川 美德
<b>Author</b>	FURUHASHI Sho, HAYAKAWA Yoshinori
<b>掲載号</b>	30巻8号
<b>発行日</b>	2017年3月20日
<b>開始ページ</b>	506
<b>終了ページ</b>	521
<b>著作権者</b>	計量国語学会

## 解説

## データの視覚化 (8)

—Gnuplot と Illustrator を用いて—

古橋 翔 (東北大学)

早川 美徳 (東北大学)

## 要旨

多機能なグラフ作成ソフト Gnuplot と、イメージ編集ソフト Adobe Illustrator を組み合わせると、データの可視化において大変力を発揮する。本稿では、グラフの作成を通して、これらの基本的な使い方と連携方法を紹介する。具体的には、まず、用意した頻度データをもとに、Gnuplot によってグラフの作成とフィッティングを行い、作成したグラフを画像ファイルに保存する。次に、Illustrator を用いて、グラフの拡大、ラベルや線種等の変更を行い、さらに、ウェブアプリケーションである TeXclip で作成した数式画像を加えて、グラフをさらに見やすく加工する。

**キーワード** : Gnuplot, Adobe Illustrator, TeXclip

## 1. はじめに

本稿では、Gnuplot と Adobe Illustrator を用いた著者なりのグラフの作成方法を紹介する。具体的には、Gnuplot でグラフを作成し、作成したグラフをさらに見やすくするために、Illustrator で加工する。今までの連載では、統計処理機能を備えたソフトウェアが紹介されてきた。Gnuplot は、それらのソフトウェアに比べて統計処理機能は劣る一方で、豊富な描画機能を有する。そこで、今までの記事のように、具体的なデータ処理と、その可視化の研究上のメリットにはあまり立ち入らず、ソフトウェアの簡単な紹介と使い方に内容をとどめたい。なお、本稿で用いた環境は Windows 10, Gnuplot 5.0.4, Adobe Illustrator CS3 である。

## 2. Gnuplot によるグラフの作り方

## 2.1 Gnuplot について

Gnuplot はコマンド入力方式のグラフ作成ソフトであり、簡単なコマンドで二次元・三次元の多種多様なグラフを描けることから、学術分野で広く利用されている。また、Windows, Linux や Mac OS など多くの OS に対応しつつフリーソフトウェアである点も、利用者が多い一因と考えられる。

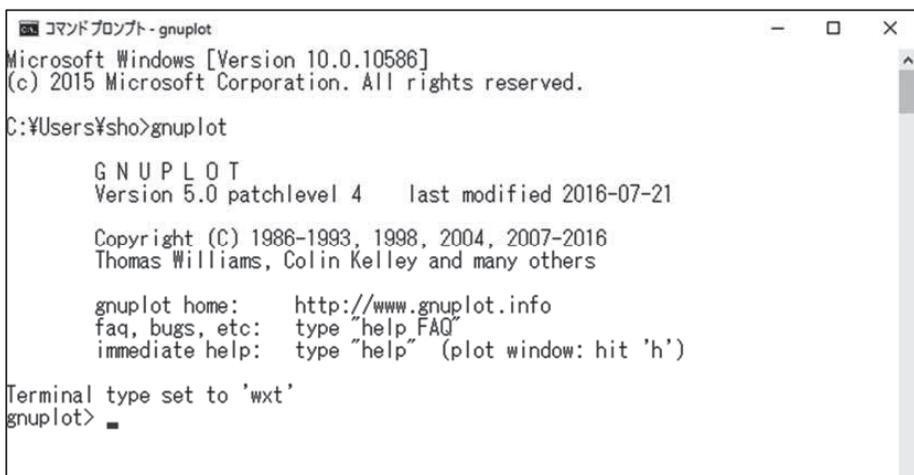
本稿では、Windows 上で Gnuplot を使用する場合を想定して説明する。Windows で初めて使用する場合、Gnuplot のウェブページ (<http://www.gnuplot.info/>) の Download から SourceForge に移動し、希望するバージョンのフォルダから zip ファイルをダウンロード

ドする必要がある。ダウンロードして任意の場所で解凍した後、若干の設定作業が必要であるが、Gnuplot の使い方によって、その方法は 2 通りある。

方法 1 は、環境変数の PATH に実行ファイル (wgnuplot.exe) の絶対パスを追加して、コマンドプロンプトで Gnuplot を使用するための設定である。「コントロールパネル」→「すべてのコントロールパネル項目」→「システム」→「システムの詳細設定」から、「システムのプロパティ」のウィンドウが表示されるので、「詳細設定」のタブを選択し「環境変数」を押す。「ユーザーの環境変数」の「PATH」を選択し、「編集」を押すと、「環境変数名の編集」のウィンドウが表示されるので、「新規」を押して、wgnuplot.exe が置かれている bin フォルダの絶対パスを記入し、「OK」を押す。以降は、コマンドプロンプト内で gnuplot と入力すればソフトが起動する (図 1)。

方法 2 では、方法 1 のような設定は行わない。wgnuplot.exe を右クリックして、「ショートカットの作成」を選択し、データファイルがある場所にコピーする。ショートカットをダブルクリックすれば Gnuplot が起動する (図 2)。なお、終了する場合は、ともに exit と入力すれば終了する。

これらには、使用時にパスに関わる違いがある点に注意が必要である。方法 1 では、cd コマンドを使いデータファイルがあるフォルダに移動し、そこで Gnuplot を起動すれば、コマンド入力時にファイル名のみを指定すれば済み、画像ファイルを作成した際もそのフォルダに保存される。その一方で、方法 2 では、wgnuplot.exe が置かれている bin フォルダが作業フォルダになるため、各ファイルの名前の前にパスを指定する必要がある。bin フォルダに、ファイルを持ち込んだり作成したりすればパスの指定は不要だが、bin フォルダには設定ファイルがあり、そこで作業を行うと問題を生じる恐れがあるため、あまりお勧めできない。



```

コマンドプロンプト - gnuplot
Microsoft Windows [Version 10.0.10586]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\Users\%sho>gnuplot

  GNU PLOT
  Version 5.0 patchlevel 4   last modified 2016-07-21

  Copyright (C) 1986-1993, 1998, 2004, 2007-2016
  Thomas Williams, Colin Kelley and many others

  gnuplot home:      http://www.gnuplot.info
  faq, bugs, etc:   type "help FAQ"
  immediate help:   type "help" (plot window: hit 'h')

Terminal type set to 'wxt'
gnuplot>

```

図 1 : Gnuplot の起動画面 (方法 1)



図 2 : Gnuplot の起動画面 (方法 2)

## 2.2 データ様式

Gnuplot では、データをあらかじめファイルで用意しなければならない。フォーマットは空白区切りがデフォルトであるが、タブ区切り及びカンマ区切りに変更も可能である。変更方法は、カンマ区切りの場合、

```
>set datafile separator ","
```

と入力する。区切り文字を変更する場合は、ダブルクォーテーションで囲まれた文字を変更する。本稿では、図 3 のような空白区切りのデータ (sentence\_length.txt) を使用する。なお、一列目が文の長さ、二列目が相対頻度を表しているとする。

```
1_0.0339351160580969
2_0.104248676530474
3_0.197638115922356
4_0.227501018053482
5_0.197230894529659
6_0.127460295914212
7_0.0637980181892222
8_0.0293199402741957
9_0.0134383059590064
10_0.0032577711415773
11_0.00135740464232388
12_0.000542961856929551
13_0.000271480928464775
```

図 3 : データファイルの例

## 2.3 作図

Gnuplot では、plot コマンドでグラフを表示する。横軸と縦軸にラベルを付けた折れ線グラフを作成する場合は、以下のコマンドを入力する。

```
>set xlabel "Sentence length"
>set ylabel "Relative frequency"
>plot "sentence_length.txt" with linespoints title "Sample"
```

set xlabel コマンドで、横軸ラベル名をダブルクォーテーションで囲んで設定し、縦軸ラベル名も set ylabel コマンドで同様に設定する。そして、plot コマンドで、データファイルをダブルクォーテーションで囲んで指定し、with でグラフスタイル (linespoints: 折れ線と点) を、title で凡例をそれぞれ指定してグラフを表示する (図 4)。なお、折れ線グラフ以外も、グラフスタイルを変更することで表示可能である。例えば、points でデータポイント表示、lines はライン表示、boxes で棒グラフとなる。

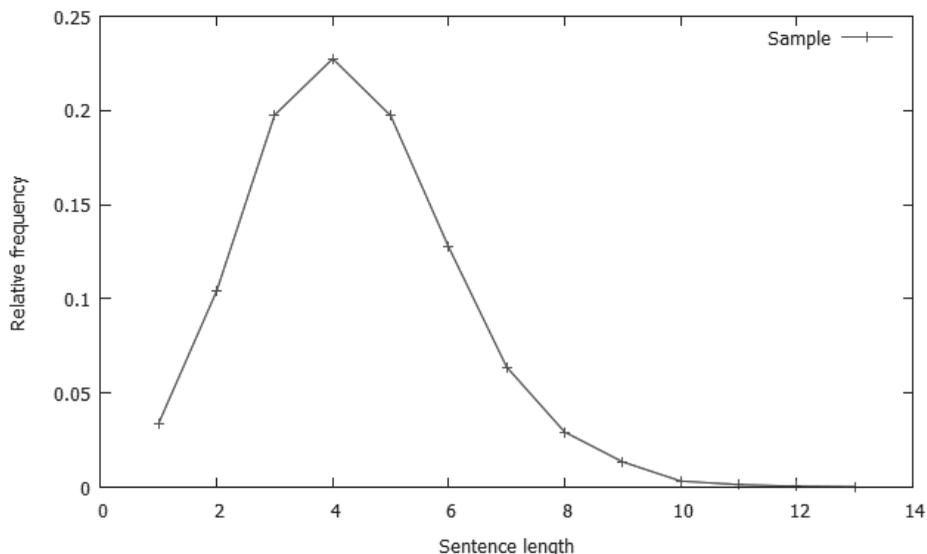


図 4: データの折れ線グラフ

## 2.4 フィッティング

データ点の分布形状がどのような関数でよく近似されるのかは、データの特徴づける際に有用な情報である。Gnuplot には、データ点を (非線形) 最小二乗法で近似する機能が備わっている。例として、データに二項分布を当てはめてみる。

```
>f(x)=gamma(n+1)/(gamma(x+1)*gamma(n-x+1))*p**x*(1-p)**(n-x)
>n=13
>fit f(x) "sentence_length.txt" via p
```

まず、データに当てはめる関数を定義する。本稿では、 $x$  の二項分布  $f(x) = Bi(n, p)$  を用いる。ここで、 $\gamma(x)$  はガンマ関数である。次に、文の長さの最大値が 13 なので、ここでは  $n$  の値を 13 と置いた。なお、パラメータのデフォルト値は 1 に設定されているが、パラメータの近似値をあらかじめ見積もれるならば、その値を設定した方が最適値により

確実に収束させることができる。準備が整ったところで、`fit` コマンドを使いパラメータの最適値を算出する。ダブルクォーテーションで囲まれた部分にデータファイル名、`via` の後にカンマ区切りで推定するパラメータを指定する。正常に終了すれば、パラメータ推定の結果が標準出力され (図 5) , フィットログがファイル (`fit.log`) に保存される。

Final set of parameters		Asymptotic Standard Error	
=====		=====	
p	= 0.325793	+/- 0.001847	(0.567%)

図 5 : パラメータの推定値

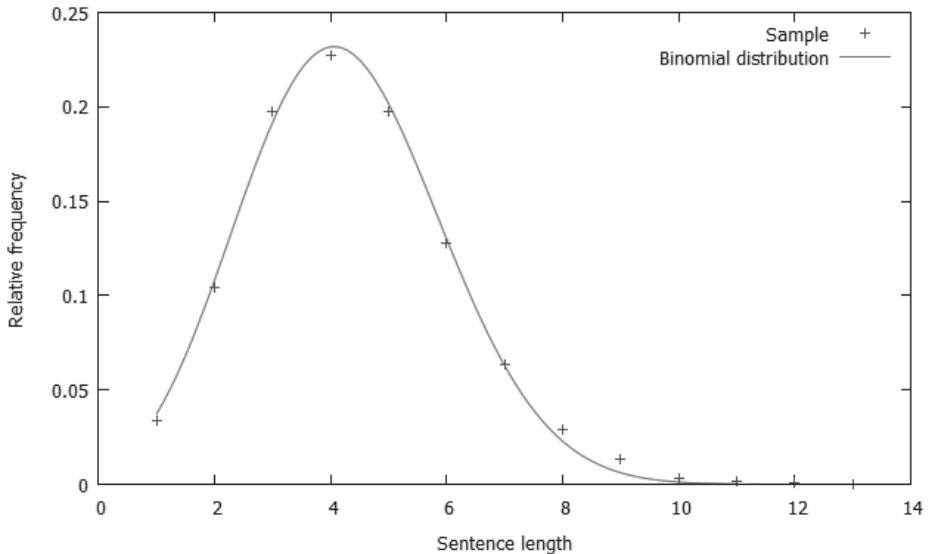


図 6 : データ点と二項分布をプロットしたグラフ

図 6 は、コマンド

```
>plot "sentence_length.txt" title "Sample",f(x) title "Binomial
distribution"
```

により作成したグラフである。パラメータの値は、`fit` コマンドで得られた推定値である。なお、複数のデータをプロットする場合、データファイル名をカンマ区切りで列挙すればよい。

## 2.5 ファイルへの保存

次に、作成したグラフをファイルへ保存してみよう。Gnuplot では、作成したグラフを各種の形式でファイルに保存することができる。Illustrator で加工・編集するためには、EPS(Encapsulated PostScript)形式で保存する必要がある、その場合のコマンドは以下ようになる。

```
>set terminal postscript color eps
>set output "sentence_length.eps"
>set size 0.6
>replot
>set term wxt
```

ここでは、出力形式を設定する `set terminal` コマンドで、出力形式 (terminal type) に `postscript`、オプションに `eps` と指定している。なお、オプションに `color` を加えると、カラー出力になる。次に、`set output` コマンドで、出力先ファイル名をダブルクォーテーションで囲み指定する。このように出力形式とファイル名を指定した上で、グラフをプロットする。`replot` コマンドで、最後に行った `plot` コマンドを再実行し、プロットしたグラフは、`set output` コマンドで指定したファイルに出力される。最後に、グラフを再び画面上に出力するよう設定を戻すために、`set terminal` コマンド (`set term` は短縮形) で `wxt` に出力形式を変更しておく (図 1 で Terminal Type set to 'wxt' と表示されていることから初期設定が `wxt` であることが分かる)。なお、`set size` コマンドでグラフを縦横 0.6 倍に縮小しているが、この工程は、Gnuplot でグラフを縮小して Illustrator で拡大した方が美しい結果が得られるという著者の経験に基づいている。

Gnuplot では、`set terminal` コマンドで出力形式を設定すれば、EPS 以外の形式のファイルへの保存も可能である。また、プロットして作成したグラフは、画像ファイルに保存せずに「copy to clipboard」でクリップボードを経由して Word や PowerPoint に直接貼り付けることもできる。

## 2.6 スクリプトファイルの利用

Gnuplot では、一連のコマンドを一括して実行したり、コマンドラインに入力した一連のコマンドをファイルに保存することができる。コマンドが書かれたテキストファイルをスクリプトファイルと呼ぶ。

スクリプトファイルからコマンドを実行する場合は、`load` コマンドを使う。

```
>load "ファイル名"
```

コマンドラインに入力したコマンドを保存する場合は、`save` コマンドを使う。

```
>save "ファイル名"
```

最後に plot コマンドで作成したグラフの情報を save コマンドで保存しておけば, load コマンドで呼び出すことによって, コマンドラインに何度も同じコマンドを入力する手間が省ける.

### 3. グラフの加工方法

上記で作成したグラフは, 線の太さなどはデフォルトのままであった. デフォルトの設定では, 線が細く, またデータ点も小さいので見にくい. また, 場合によっては, 凡例が線と重なり見にくい場合もある. もちろん, Gnuplot では, 線の太さやデータ点のマークの種類, 凡例の位置などは変更可能で, コマンドやオプションを駆使すれば凝ったグラフも作成できる. しかしながら, コマンド方式のため, 数値を変えてはコマンドを入力しグラフを確認する作業を繰り返さなければならず大変である. そこで, 本節では, Gnuplot で作成したグラフを Illustrator を使ってさらに手を加える手法を紹介する.

#### 3.1 Illustrator の画面

初期設定の Illustrator の画面構成は, 基本的なメニューの他, 図 7 に示すように, 左端にツールパネル, 右端にパネルが表示されている. メニューバーの「ファイル」を選択し, 「開く」から, グラフを保存した EPS ファイルを選択して開くと, アートボードが立ち上がる. このときアートボード上にはグラフが表示されるが, グラフは Illustrator のオブジェクトとして変更・修正が可能である.

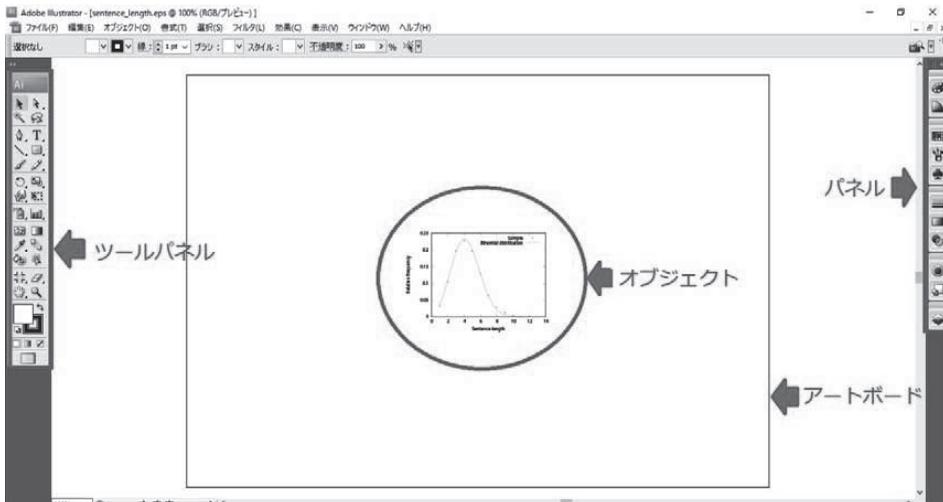


図 7 : Illustrator の画面

#### 3.2 図のサイズ変更

作成した図は, 使用目的に応じてサイズを変更する必要があるが, 著者は Gnuplot で作成した図のサイズを Illustrator で目的に応じて調整することが多い. 本節では Illustrator による図のサイズ変更の方法を紹介する.

アートボードがオブジェクトを配置できる領域となるので、まず、アートボードのサイズを適宜変更する。メニューバーの「ファイル」から「ドキュメント設定」を選択して、アートボードの幅と高さを変更し、「OK」ボタンを押す(図8)。次に、ツールパネルから「選択ツール」を選んだのち、左クリックでドラッグしてグラフ全体を選択する。その後、メニューバーの下に現れる図9の「W」と「H」に、それぞれアートボードの「幅」と「高さ」より少し小さい値を入力すると、図10のように、グラフがアートボード全体に広がる。



図8：アートボードの設定画面



図9：オブジェクトの座標 (X,Y)，幅 (W) と高さ (H)

一番左のマークは、座標値がオブジェクトのどの位置に対応するかを表す基準点  
今回はオブジェクトの中心が基準点。座標の原点はアートボードの左下のコーナー

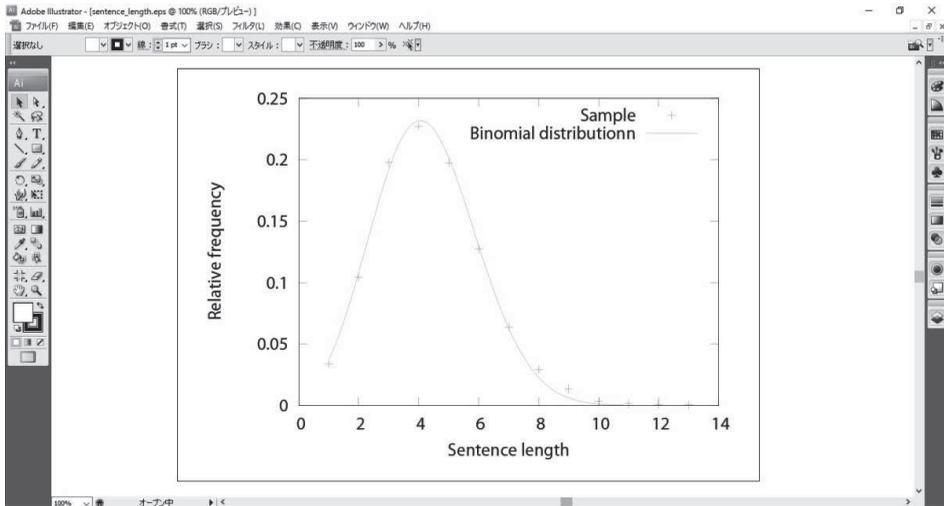


図 10：拡大後のグラフ。アートボードのサイズは、図 7 と同じ

### 3.3 線の加工

図 10 のままではグラフの線が細く見にくいので、次に、線を太くして明瞭にする。ツールパネルから「選択ツール」を選択し、左クリックでドラッグしてグラフ全体を選択する。

「線」パネル（図 11）を選択して、線幅を適切な値に変更する。図 12 は、図 10 のグラフの線の幅を 5pt へ太くしたものであり、見やすくなっていることが分かる。



図 11：「線」パネル

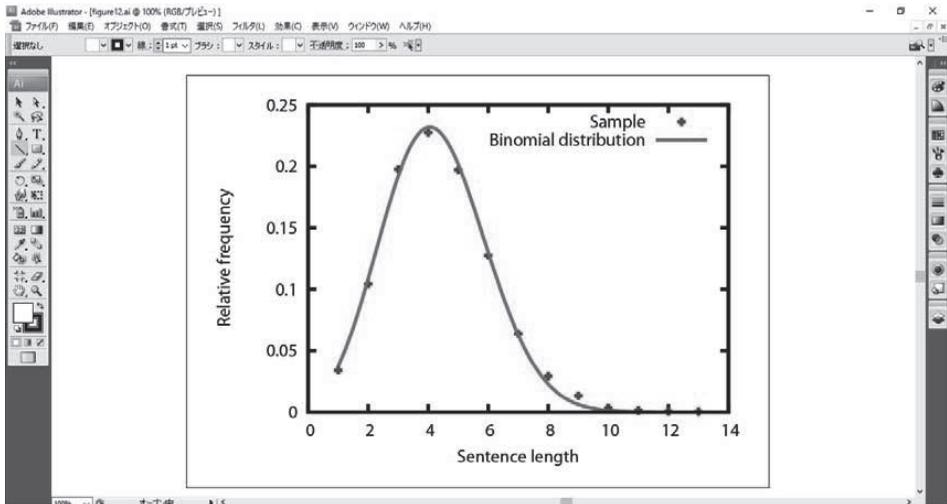


図 12：線の太さを 5pt に変更後のグラフ

グラフ中に複数の線が含まれる場合、それらを区別するため線の種類や色を変更することは有効である。「ダイレクト選択ツール」で、破線にしたい線を選択する。「線」パネルを選択して、「オプションを表示」を選択する。パネルの「破線」にチェックを入れると、選択した線が破線になる。線分と間隔の数値を変更することで、色々な破線ができる。

### 3.4 ラベルの変更

グラフのラベル等の変更は GnuPlot でも可能ではあるが、Illustrator を使えばより簡単に行うことができる。ここでは、グラフの軸のラベルと凡例を日本語に書き換える。図 13 のようにツールパネルから「文字ツール」を選択して、変更したい部分を左クリックしてテキストを入力する。

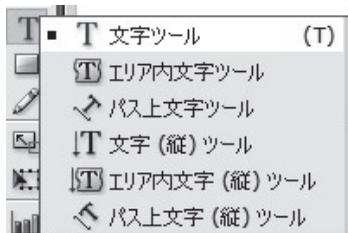


図 13：文字ツール

その際、テキストの長さが異なると、配置がずれてしまうので、ツールパネルにある「ダイレクト選択ツール」を選択して、該当するオブジェクトを左クリックして方向キーで位置をずらす。移動させる際、アートボードに定規とグリッドを表示すると便利である。

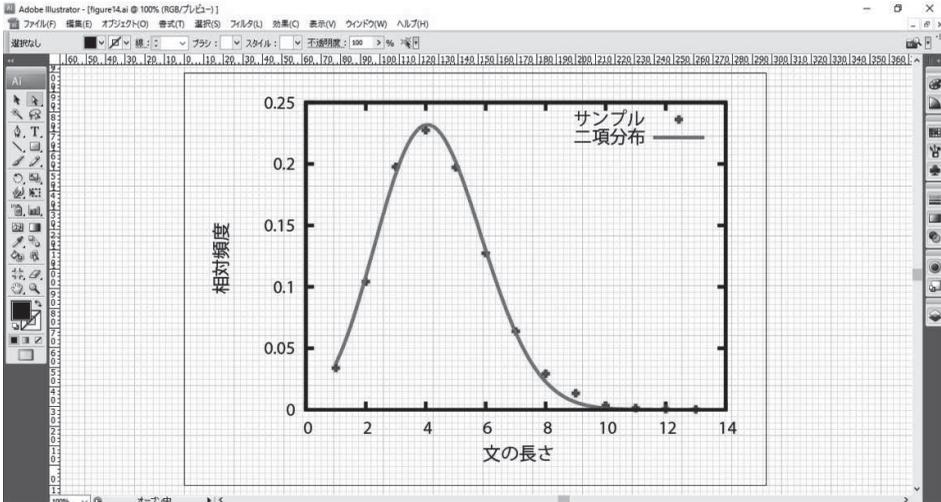


図 14 : ラベル変更後のグラフ

グリッドを表示するには、メニューバーにある「表示」から「定規を表示」と「グリッドを表示」を選択する。図 14 は、定規とグリッドを表示させた上で、縦軸・横軸のラベルと凡例を日本語に変換したグラフである。

### 3.5 数式の挿入

グラフ上に関数をプロットする際に、その数式を本文、凡例やキャプションなどに明記することになるが、グラフ中に数式を挿入したい場合もある。本節では、美しい数式のタイプセットが可能な TeX の数式を載せる方法を紹介する。

まず、Web サイト TeXclip (<https://texclip.marutank.net/>) で、LaTeX のコマンドで数式を記入してイメージを作成する。このイメージをコピーして、グラフのアートボードに直接貼り付けることができればよいのだが、残念ながらその機能は提供されていない。そこで、一旦、数式を画像ファイル (SVG, EPS) として保存したものを、Illustrator で開く必要がある。開いたアートボードで、「ダイレクト選択ツール」を選択し、左クリックでドラッグして数式を選択し、メニューバーの「編集」から「コピー」を選択する。そして、グラフのアートボードに移り、メニューバーの「編集」から「ペースト」を選択して数式のオブジェクトを貼り付けた後、左クリックのドラッグで配置したい位置に移動する。右クリックで「変形」から「拡大・縮小」を選択すれば大きさも変更できるので、適当な大きさに変更する (図 15)。

### 3.6 矢印の作成

図 15 では実線が表す関数に数式が対応しているか明らかであるが、数式がどの曲線に

対応しているのか明示したい場合もある。そのような場合、矢印の使用が有効である。

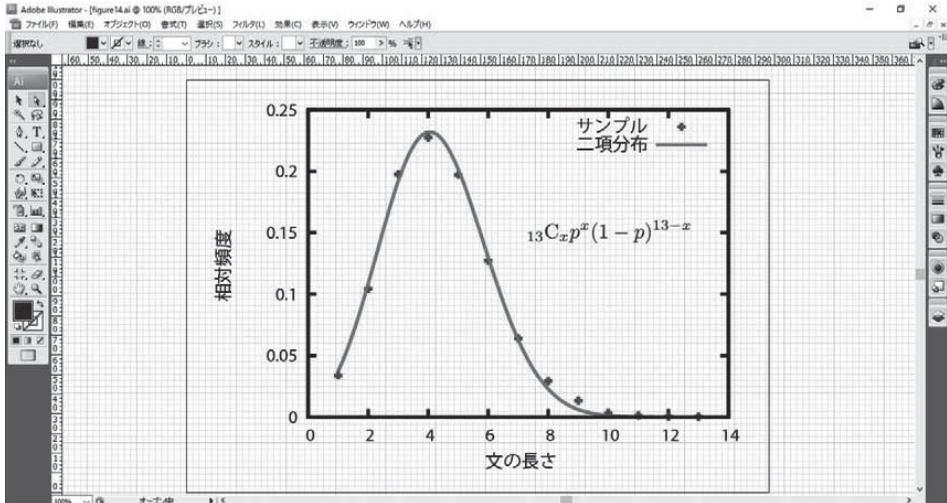


図 15 : 数式挿入後のグラフ

矢印を描くには、まず、ツールパネルから「直線ツール」(図 16) を選択し、始点から終点まで左クリックでドラッグして、「線」パネルで線幅を選択する。

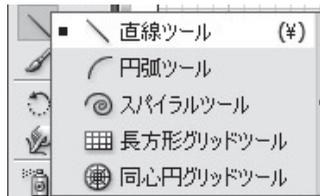


図 16 : 直線ツール

次に、メニューバーの「効果」から「スタイライズ」そして「矢印にする」を選択する。ダイアログ(図 17)が表示されるので、始点と終点のデザインを選択する。その際、「プレビュー」にチェックを入れると、アローヘッドの大きさが確認できる。図 18 のようにアローヘッドが大きい場合は、「拡大・縮小」の数値を変更して調整する。

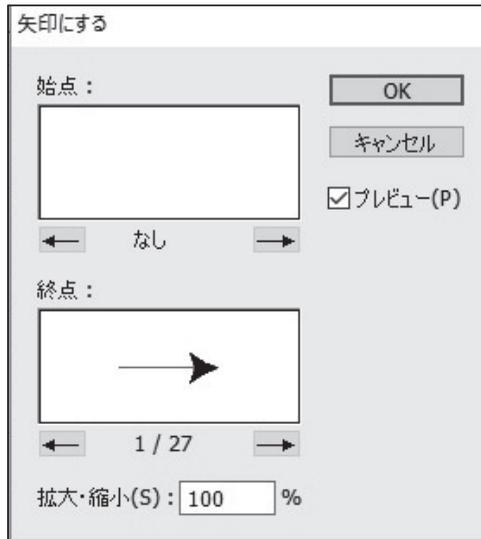


図 17 : ダイアログ

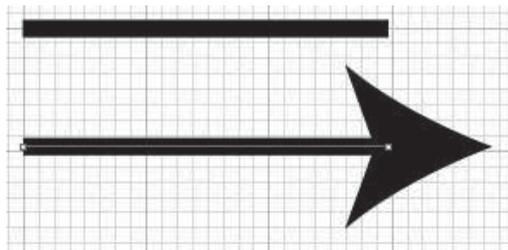


図 18 : 直線 (上). アローヘッドの追加後 (下)

以上の手順でグラフ上に矢印を作成したものが図 19, 図 20 である.

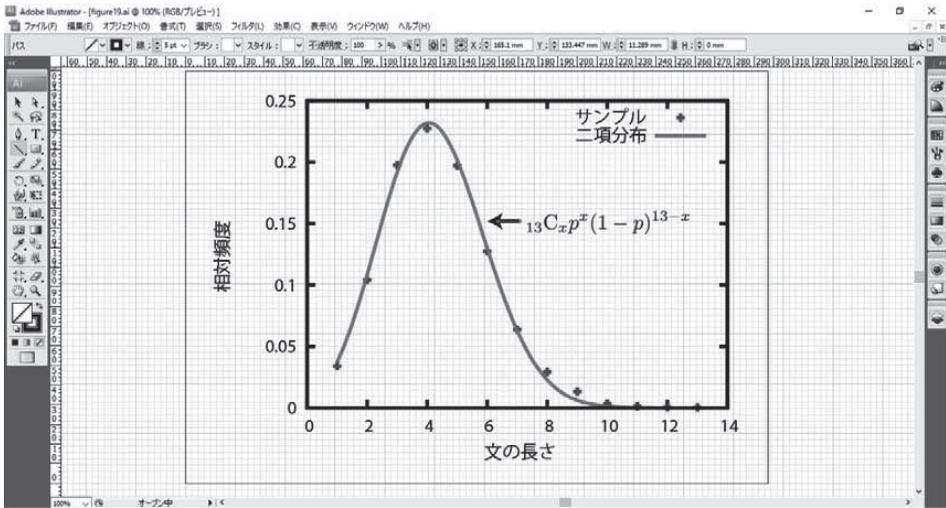


図 19：関数の曲線とその数式を矢印で対応付けたグラフ

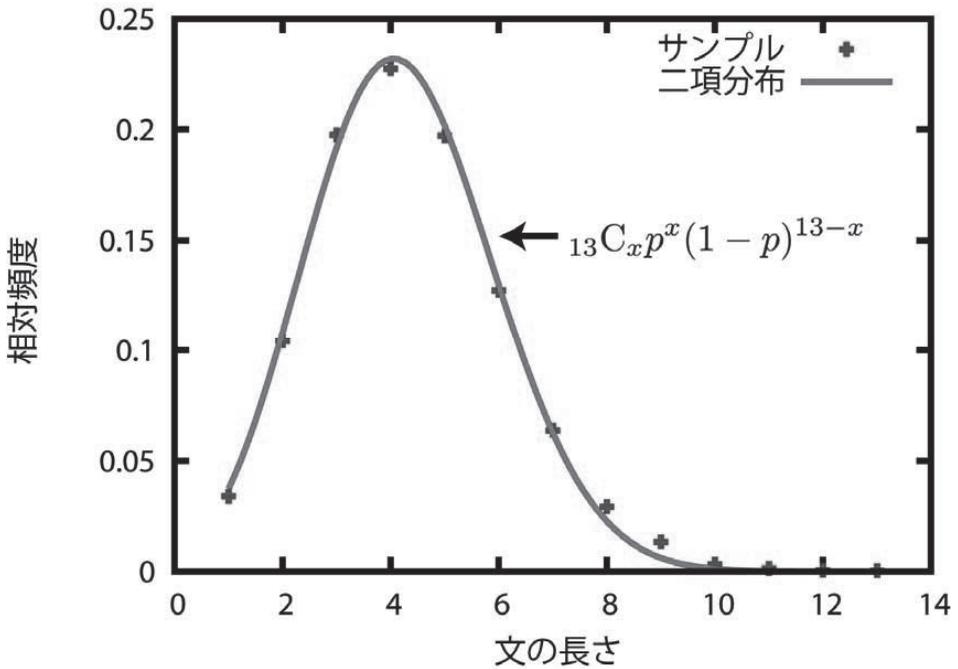


図 20：完成した図

#### 4. おわりに

本稿では、Gnuplot と Illustrator の使い方について簡単な説明をした。ここで紹介したのはどちらのソフトウェアでもごく基本的な機能であり、紹介されていないものの中にも、読者の研究に役立つ機能が多くあるはずである。巷には、これらのソフトウェアについて

詳説した書籍やウェブページが沢山ある。本稿をきっかけに、こうした情報にアクセスし、データの視覚化のスキルアップの一助となれば幸いである。

## 文献

- 生田信一・柘植ヒロポン・ヤマダジュンヤ・順井守 (2009) 『Illustrator 逆引きデザイン事典 [CS4/CS3 対応]』株式会社翔泳社
- 大竹敢 (2004) 矢吹道郎 (監) 『使いこなす gnuplot 改訂第 2 版 Version4.0 対応』テクノプレス

(2016 年 11 月 29 日受付)

*Tutorial*

## Data Visualization (8): Application of Gnuplot and Adobe Illustrator

FURUHASHI Sho (Tohoku University)

HAYAKAWA Yoshinori (Tohoku University)

Abstract:

We demonstrate an example of visualization process of linguistic data with a combination of a scientific plotting software (Gnuplot) and a graphics editor (Adobe Illustrator). This paper consists of two parts. In the first part, we explain how to use Gnuplot through constructing a line chart and fitting a binomial distribution to data points. In the second part, we use Adobe Illustrator to make a graph more visible. We can change the size of graph, the width of lines, and the text of axis labels and graph legends. Moreover, we insert a TeX-based expression image made by TeXclip into the graph with an arrow that points the solid line corresponding to the expression.

Keywords: Gnuplot, Adobe Illustrator, TeXclip