林育研報 Bull. For. Tree. Breed. Center No.23, 2007

 $-433 \sim 450 -$ 

# 研究資料

# 正規確率プロットを作成するプログラムの開発

河崎 久男(1)

Hisao Kawasaki<sup>(1)</sup>

A Programming for Delineating Normalized Quantile Plots

要旨:各種の実験や調査で得られたデータについて、正規確率プロットを容易に作成するプログラムを開発した。正 規確率プロットは、所与のデータの正規性を確認する有効な方法の一つである。しかし、正規確率プロットを作成す る具体的な方法が必ずしも明確ではないために、この方法はあまり用いられていない。そこで、所与のデータがテキ ストファイルで準備されれば、正規確率プロットについての知識を持ちあわせなくても、簡単に描画できるようなプ ログラムを開発した。プログラムはマイクロソフト社のVisual Basic (Ver. 6.0)を用いて、ソースをコーディングし た。開発したプログラムでは、扱えるデータ数は180個までであり、また、データ変換も自然対数への変換ルーチン しか組み込んでいない。しかし、この報告を行ったことによってプログラムのソースも公開可能なので、今後、必要 であればソースの改良も行える。プログラムの具体的な使用方法についてもまとめ、解析事例もデータと結果とを併 せて紹介した。

 (1) 林木育種センター 〒319-1301 茨城県日立市十王町伊師3809-1 Forest Tree Breeding Center 3809-1 Ishi, Juo, Hitachi, Ibaraki 319-1301 Japan

# 1 はじめに

林木育種における事業や研究で扱う連続変量型データは、解析に際して、正規性が仮定される場合が多い。また、実験や調査で得られた一連のデータは、通常、経験的に正規分布に従う変量とみなして解析を進めている実 態がある。得られたデータが正規分布に従う変量ではないと想定される場合でも、データに対して適当な変数変 換や基準化を行うことによって、実用上、正規分布が仮定できることは実際によくある。

さて,データが正規分布に従う変量であるかどうか,すなわち,データの正規性を検定する方法<sup>4</sup>は存在する。 しかし,その検定方法は,かなり困難なため一般的ではない。

この問題に対する簡便な解決手段の一つに,所与のデータについて正規確率プロットを作成する方法がある。こ の方法は従来からあり,必ずしも新しい方法ではないが,実際にその方法を適用する具体的な手順は明確ではな い。正規確率紙に個々のデータをプロットするというアナログ的な方法<sup>1</sup>は,現在ではあまり用いられていない。

そこで、この正規確率プロットを所与のデータに容易に適用するプログラムを開発した。同様のプログラムは いたるところで開発されており<sup>3</sup>、ここで報告するプログラムに必ずしも新規性はない。しかし、この種のプロ グラムはデータ解析の際に頻繁に使用する、いわばツールであり、類似のプログラムが多数あったとしても意義 はある。また、この報告を行うことによって開発プログラムのソースの公開も可能なので、当該プログラムの利 用者はさらに使いやすくするための修正や変更もできる。

以上のことから,所与のデータの正規確率プロットを作成する開発プログラムについて,概略と使用方法をこ こに報告する。

# 2 正規確率プロットとは

## 2.1 正規確率プロットの意味

正規確率プロットとは,所与のデータの分布関数が一つの正規分布であると仮定したとき,個々のデータ値に ついて確率を媒介変数としてパーセント点をプロットする<sup>5</sup>ものである。

正規確率プロットの利点は、「母集団が正規分布に従う変量の集合体であるとき、得られた標本データに適当 な変換や基準化を行って正規確率プロットを作成すると、当該プロットは一直線を描くのでデータの正規性を視 覚的に確認することができる。」ことである。すなわち、正規確率プロットでは、異常な値は直線から明らかに ずれてプロットされるので、データの簡易なチェックが可能である。

## 2.2 正規確率の計算

正規確率プロットの作成では、プロットの具体的な方法を詳細に記した文献を見つけることは難しく、多くの 文献では未だに正規確率紙を適用する方法が記述されている。

パソコンを使用する場合では、標準正規分布をあてはめた場合の個々のデータについて、パーセント点を計算 する<sup>3</sup>必要がある。パーセント点の計算では、厳密には非線形方程式を解く必要があるが、近似値を用いたいく つかの計算方法が知られており<sup>2</sup>、実用上は、これらの方法で十分だと言われている。

ここでは、「戸田の近似式」と呼ばれる方法<sup>3</sup>を採用したが、この方法は「計算機を用いる実用上の目的から は十分な精度が得られる。」とされている。

## 2.3 基準化の方法

正規確率プロットを作成する際,所与のデータについては、いわゆる基準化を行う必要がある。 通常の基準化は、よく知られているように(式-1)で定義され、「個々のデータXiから平均値Xを差し引き、 それを標準偏差σで除した値とする。」ものである。

しかし、データの属していた母集団の分布型が不明であるときや、異常値の混入が予想される場合には、(式 – 2)に示す方法もある。この方法は、「個々のデータXiからメディアン $\widehat{X}$ を差し引き、それを四分位偏差 $\delta$ の 0.741倍した値で除した値とする。」ような、いわゆるロバストな推定量を用いて基準化する<sup>50</sup>ものである。

開発したプログラムでは、上記の基準化のいずれか一方を選択できるようになっている。なお、後者による基 準化では、メディアンX が描画グラフの必ず中央になるように正規確率プロットが作成される。

 $X_i \longrightarrow (X_i - \overline{X}) \neq \sigma$  ……… (式-1) ただし、  $\overline{X}$ は平均値で、  $\sigma$ は下式で定義される標準偏差である。  $\sigma = \sqrt{\sum (X_i - \overline{X})^2 \neq (n-1)}$  $X_i \longrightarrow (X_i - \widetilde{X}) \neq (0.741\delta)$  ……… (式-2) ただし、  $\widetilde{X}$ はメディアン(中央値)で、  $\delta$ は下式で定義される四分位偏差である。  $\delta = X(c) - X(n+1-c), c = 3(n+1) \neq 4$ 

# 3 プログラムの使用方法

## 3.1 プログラムソースのコーディング

正規確率プロットを作成するためのプログラムは、一つのツールとして使用するため、きわめて容易に扱える ようにソースをコーディングした。すなわち、予備的な知識がなくても、3.2節に記述するデータファイルだ けを準備すれば、正規確率プロットが描画できるようにプログラムを開発した。

なお、プログラムソースのコーディングは、WindowsのOS上で稼動するVisual BASIC (Ver.6.0) で行い、コ ンパイル後の実行プログラムの名称は「qqPlot.exe」とした。当該プログラムは、特殊なリンクなどは使用して いない。したがって、通常はWindowsに付属のシステムファイル「MSVBVM60.DLL」が、当該パソコンのハー ドディスク(多くの場合、C: ¥ Windows ¥ System) に存在していれば作動する。

## 3.2 データファイルの準備

開発プログラム「qqPlot.exe」を使用するためには、正規確率プロットを作成しようとするデータをテキスト ファイルで事前に準備する必要がある。データファイルに記述する文字は、原則的に半角の英数字である。

データファイルは,所与のデータをExcelやエディタなどを用いて,表-1,表-2および表-3などに例示 するように,一行に一つの数値,あるいは二つ以上の数値を記述してテキスト形式で作成する。記述する数値の 各行の順序は定めていないので,順不同でよい。結果の出力時には,所与のデータは昇順(小→大)に整列され る。

- 435 -

表-1 ある次代検定林における家系別樹高の最小二乗推定値

: <beg< th=""><th>inniı</th><th>ng of</th><th>file&gt;&gt;</th></beg<>	inniı	ng of	file>>
:::LSE (	Color	Code	
13.4487	9	226	
13.1588	12	32	
13.7126	9	161	
13.4833	9	309	
13.3064	9	260	
14.0756	9	145	
14.0074	9	146	
13.3929	9	140	
13.3547	9	176	
13.5170	9	160	
13.3928	9	164	
13.3689	9	167	
13.5698	9	150	
13.9243	9	143	
13.2592	9	155	
13.8432	9	156	
13.6483	9	163	
13.7457	9	182	
13.6406	9	179	
13.5630	9	183	
13.6192	9	175	
14.0240	9	168	
13.3595	9	174	
13.5454	9	177	
13.7821	9	165	
13.6185	9	172	
12.8755	9	350	
12.7758	9	144	
13.5561	9	166	
13.1606	9	185	
13.1842	9	312	
13.5329	9	154	
: <bot< td=""><td>tom o</td><td>of fil</td><td>e&gt;&gt;</td></bot<>	tom o	of fil	e>>

::: <tc< th=""><th>op of</th><th>data</th><th>1&gt;&gt;</th><th>-</th></tc<>	op of	data	1>>	-
:::区画-1	色	行	列	家系
1.087	9	2	3	2
0.917	9	2	7	2
0.633	9	3	2	2
1.563	9	3	4	2
1.667	9	3	6	2
0.813	9	4	2	1
2.941	9	4	5	2
1.961	9	5	2	2
2.857	9	5	4.	2
· 黄金スギ	自殖			
1.724	10	5	5	0
1.923	9	5	7	1
4.651	9	6	3	2
0.568	9	7	2	2
1.538	9	7	4	2
4.348	9	8	1	2
4.762	9	9	4	2
ःःः	<b>4</b>	仁	Fil	安亚
::: 区画-2	巴	17	<i>9</i> 1	<u></u> 秋栄
1.053	12	1	2	2
1.111	12	1	3	1
2.128	12	1	4	2
1 667	12	1	2	2
5.66	12	2	5	2
1 887	12	2	7	2
11 111	12	2	, 6	2
3 636	12	3	8	2
6.122	12	4	3	2
40	12	4	5	2
5.66	12	4	7	2
4.054	12	5	1	1
28.571	12	5	2	2
25.926	12	5	4	2
<ul><li>: 黄金スギ</li></ul>	自殖			
4.938	8	5	5	0
3.226	12	5	8	2
12.5	12	6	1	2
19.048	12	6	5	2
1.01	12	6	7	2
7.143	12	7	2	2
4.444	12	8	1	2
1.818	12	8	9	2
4.348	12	9	4	2
::: <bo< td=""><td>ottom</td><td>of d</td><td>lata&gt;&gt;</td><td></td></bo<>	ottom	of d	lata>>	

表-2 中央に黄金スギを植栽したミニチュア採種園の植栽木から

採取した種子で分離した黄金スギ様苗の出現率

.

林木育種センター研究報告 第23号

表-3 他家受精, 自家受精による鉢別のトウモロコシにおける

個体の高さ(ダーウィンのデータ)

:::< <test data="" for="" qqplot.exe="">&gt;</test>	
:: 他家受精 自家受精	
::鉢 I	
23.5 . 17.375	
12.0 . 20.375	
21.0 . 20	
::鉢Ⅱ	
22.0 . 20	
19.125 . 18.375	
21.5 . 18.625	
::鉢Ⅲ	
22.125 . 18.625	
20.375 . 15.25	
18.25 . 16.5	
21.625 . 18	
23.25 . 16.25	
::鉢IV	
21.0 . 18	
22.125 . 12.75	
23.0 . 15.5	
12.0 . 18	
:::< <bottom data="" of="">&gt;</bottom>	

なお,正規確率プロットの対象となるのは一番目の数値であり,二番目の数値は表-4に示すようにプロット を描画する際の印字色の指定である。二番目の数値がない場合,あるいは表-4の範囲外の数値である場合には, プロットは「黒色」で描画される。

0 :黒	8:灰色
1:青	9:明るい青
2 :緑	10:明るい緑
3:シアン(水色)	1 1 :明るいシアン
4 :赤	1 2 : 明るい赤
5:マゼンタ(赤紫)	13:明るいマゼンタ
6:黄	1 4 : 明るい黄
7 :白	15:明るい白

表-4 描画の色指定のパラメータ

また, 描画の背景色は「白」なので, 表-4に示したプロットの描画色のうち, パラメータ「7」および「15」 ではプロットされた印が目視できないので注意する。「黒」色との対比で目視しやすい色は,「9:明るい青」, 「12:明るい赤」,「13:明るいマゼンタ」の三色で, それ以外の色は明暗がいずれかに偏りすぎるためか, 見栄 えはよくない。

一行に記述される数値が複数のとき、各値の間は一つ以上の半角スペースで区切られていればよい。三番目以 上の値は、解析には影響を及ぼさない。ただし、一行に記述されている項目数が30を超えると、当該行そのも のがデータに無関係とみなされ、データ解析の対象にならないことがある。

行頭の最初の1文字が半角のコロン「:」である場合,当該行はコメント行とみなされ解析に影響しないので, ここには全角の日本語を使用することもできる。

解析用のファイル名は任意であるが、半角英数によることが望ましく、拡張子は「\*.dat」、「\*.prn」、「\*.txt」などを推奨する。ただし、拡張子「\*.dat」は、最近はWindowsシステムで使用する場合が多く、あまり 好ましくない。

なお、現在のバージョンでは、扱えるデータ数は最大180個である。

## 3.3 開発プログラムの使用方法

プログラム「qqPlot.exe」の使用方法について、解析ファイルの選択やボタンのクリック、オプション選択な どを具体的な項目の順に沿って以下に記載する。

0) プログラム「qqPlot.exe」の起動

プログラム「qqPlot.exe」は適当なフォルダに置くことができ、通常のソフトと同様に、アイコンをダブルク リックすることによって起動する。

プログラム「qqPlot.exe」を起動すると、画面-1が開く。

1)解析ファイルの存在するドライブ、フォルダの選択

解析ファイルの存在するドライブとフォルダの選択は、画面-1の左上方で、それぞれの欄でマウスをクリックして行う。

ドライブは、ドライブ選択用ボックスの右側の「▼」をクリックして、選択する。

フォルダは、フォルダ選択用ボックスで、ファイルの存在するフォルダをダブルクリックして選択する。

なお、フォルダ名が日本語である場合には、正しく表示されない。

#### 2)解析ファイルの選択

前項1)で解析ファイルの存在するドライブとフォルダを選択すると,画面-1の右上のファイル表示欄に,解 析対象ファイルの拡張子が「\*.dat」、「\*.dta」であるファイル名が表示される。拡張子が「\*.dat」、「\*.dta」以外 の場合は,解析対象ファイル名の上段にある「Input file (\*.dat; \*.dta)」をクリックすると,その部分が「Input file (\*.prn; \*.txt)」に変化し、下段のファイル表示覧に拡張子が「\*.prn」、「\*.txt」のファイルが表示される。さ らに、再度「Input file (\*.prn; \*.txt)」をクリックすると、その部分が「Input file (\*.\*)」、に変化し、当該フォル ダのすべてのファイルが表示される。

解析対象ファイルは、一覧表示されたファイルをクリックして選択し、上段のファイル表示欄に表示されることによって確定する。

-439-

林木育種センター研究報告 第23号

rive & folder for input	Input file (*.dat;*.dta)
<b>d</b> :	Table1.dta
D:\ VB_vr6 Analyses	Table1.dta Table2.dta Table3.dta
Data	QUIT Begin
+1+2.	····+
···+····1.···+···.2.	····+····4···
+1+2.	
+1+2. 	+3+4 
+1+2. 	+3+4 

画面-1 「qqPlot.exe」の起動画面

なお、フォルダ名の場合と同様に、ファイル名が日本語である場合には、正しく表示されない。

3) 解析ファイルからのデータ読み込み

前項2)で解析ファイルを確定した場合は、[Begin]ボタンをクリックする。

画面-1の中段に,選択した当該ファイルの内容について最初の6行分が表示される。[Begin]ボタンの代わりに,[QUIT]ボタンをクリックすると画面-1を閉じ、プログラム「qqPlot.exe」を終了する。

4) 出力ファイルの指定

解析結果をファイルにも出力したい場合には、画面中央のやや下左方にある[Output file]のオプションボタン をクリックする。次に、その右側にある出力ファイル指定欄で出力ドライブ、フォルダ、ファイル名称などを必 要に応じて変更する。名称等を変更しない場合は、解析対象ファイルの存在するドライブ、フォルダに、拡張子 だけが「\*.res」と変更されたファイルが出力ファイル名として選定される。

なお、プログラムの規定値では、出力ファイルを作成しないようになっている。

5)解析データの分布型と基準化の指定

項目3)で[Begin]ボタンをクリックすると、項目2)で指定したファイルの内容が読み込まれ、画面-2のようになる。

-440 -

🗐 qqPlot Quantile data plots control.	x ing with a N.D. 2006-08-02
Drive & folder for input	Input file (*.dat;*.dta) Table1.dta
D:\ D:\ VB_vr6 Analyses Analyses	Table2.dta Table3.dta
	QUIT <u>B</u> egin
:<-Beginning of fil :::LSE Color Code 13.4487 9 226 13.1588 12 32 13.7126 9 161 13.4833 9 309 	e>>
Output file; [D:IVB_vr6VAn <u>Model type</u> Standard Normal Distrit Logarithm Normal Distri	alysestopplot(Data)Table1.res Standardization © Best linear © Robust
Start for analyzing	<u>Cancel</u> Start (b)

画面-2 「qqPlot.exe」のオプション指定画面

ここで、画面中段下の「Model type」と「Standardization」のボックスで、それぞれオプションを選択する。 「Model type」ボックスでは、所与のデータの分布型について選択する。

オプション「Standard Normal Distribution」は解析対象データが通常の正規分布が想定できる場合に、オプ ション「Logarithm Normal Distribution」は対数正規分布が想定できる場合に、それぞれ選択する。後者の場 合、所与のデータはプログラム内部で自然対数の値に変換されてから解析される。したがって、データ値に「0」 以下の値が含まれるときは、エラーを起こさないように当該データを除外して解析を進めるが、そうした処理の 正誤は利用者が判断する必要がある。

なお、通常の正規分布を選択した場合は、当然、そのような処理は行われない。

「Standardization」ボックスでは、データの基準化の方法を選択する。

オプション「Best linear」は(式-1)に示す方法によって、オプション「Robust」は(式-2)に示す方法によって、それぞれ所与のデータを基準化するものである。

先の「Model type」の選択と、この「Standardization」の選択とによって、同じデータであっても、その都 度、プログラム「qqPlot.exe」を用いるプロセスを繰り返すと、都合4通りの正規確率プロットを作成すること ができる。 - 442 -

## 6)解析の開始と結果の出力

前項で分布型および基準化のオプションの選択を終えた場合は、画面 – 2の最下段右側の[Start (b)]ボタンを クリックする。

この操作によって,所与のデータの解析が行われ,正規確率プロットを作成するための情報が得られる。当該 情報は,画面-2の一部に重なるように新たに開いた画面-3の左側に出力される。



画面-3「qqPlot.exe」の出力画面

出力情報は,選択したモデル,基準化の方法,平均値と標準偏差(あるいは,メディアンと四分位偏差)と, 個々のデータの解析結果である。個々のデータの解析結果は一行ごとに,①解析対象データを昇順(小→大)に ソートした場合の番号(1から解析データの個数まで),②ファイルから読み込まれた解析対象データ,③基準 化されたデータの値,④確率密度関数をあてはめたときの確率(ただし,0.00~0.50の範囲),⑤確率プロット の値,の5種の情報で構成される。一つのデータが一行に出力されるので,この情報は解析データの個数分の行 数によって表示される。

項目4)でファイルへの出力を指定した場合は、指定したファイル名とともに、これらの情報がテキスト形式で ファイルに出力されることになる。

データの数が多い場合は,解析データ表示欄の右側のスクロールバーを下方へドラッグすると,隠れていた データも表示できる。 なお、この項目の最初、すなわち画面-2の最下段で右側の[Begin]ボタンの代わりに左側の[Cancel]ボタン をクリックすると、画面-3は開かず、項目1)に戻ることができる。

7) 確率プロット作成のためのオプションの選択

画面-3の右側に,確率プロットを描画する際のオプションがある。

このオプションは、基準化したデータと確率プロットの値とを描画する際、縦軸と横軸とをどちらに配置する かを指定するものである。

ー部の文献には、「最近の欧米の文献では、横軸に確率値を、縦軸に基準化したデータをとる方が一般的である。」という記述があった<sup>50</sup>ので、当初は、その方針(オプションでは上のボタン)でプログラムの開発を進めた。

しかし、日本では市販の正規確率紙が以前からよく用いられていた<sup>1)</sup>ためか、横軸に基準化したデータを、縦 軸に確率をとる方が多く、しかも、縦軸は、通常、確率値(0.01~99.99)で表示されることが多い。こうした ことから、プログラムとしても従来型と違和感がない方が良いと考え、縦軸も確率値とすることを規定値(オプ ションでは下のボタン)とした。

8) 確率プロットの描画と描画画面のファイルへの保存

画面-3で前項7)のオプションを選択した後,画面左下にある[<u>Graph</u>]ボタンをクリックすると,出力画面 に換わって,画面-4のように確率プロットの描画画面が出現する。



画面-4 確率プロットの描画画面

- 444 -

この描画画面に何かを直接に書き込むことはできない。

しかし、画面-4の右下にある[SaveScreen]ボタンをクリックすることによって、この描画をファイルに出力 することは可能である。

出力ファイルはビットマップ形式で,解析データが存在するドライブのフォルダに保存される。ファイル名は, 解析ファイル名の拡張子が「\*.bmp」に変更されたものであるが,同じ名称のファイルが既存の場合は上書きさ れるので注意する。また,ファイルの大きさは概ね600KBを超えるので,例えば3.5インチディスクのように,出 力先の容量に余裕がないときには出力が行われない場合がある。

保存された出力ファイルは、他の市販のソフトで加工ができる。

なお, [<u>S</u>aveScreen]ボタンを一度クリックした場合,同じプロセスでは[<u>S</u>aveScreen]ボタンの二度目のクリックはできないようになる。

また、画面左下にある[Output]ボタンをクリックすると、項目7)の出力画面へ戻ることができる。

さらに、画面右下にある[Close(x)]ボタン、または画面右上隅の[×]ボタンをクリックすると画面-3を閉じ、 画面-5が表示される。この画面でクリックできるのは、画面右下にある[Finish]ボタンか、通常のソフトと同 様に右上隅の[×]ボタンだけである。前者の場合は画面-1に変わり、項目1)へ移動できるが、後者の場合は 画面-5を閉じた後、プログラム「qqPlot.exe」を終了する。

Shite a lotaol lot inpat	Input file (*.dat;*.dta)
🚍 d:	- Table1.dta
D:\ D:\ D-VB_vr6 Analyses apPlot	Table2.dta Table3.dta
ata Sata	QUIT Begin
.3.4487 9 226 .3.1588 12 32 .3.7126 9 161	
.3.4487 9 226 .3.1588 12 32 .3.7126 9 161 .3.4833 9 309 	••••••••••••••••••••••••••••••••••••••
.3.4487 9 226 .3.1588 12 32 .3.7126 9 161 .3.4833 9 309 	alvses\qqPlot\Data\Table1

画面-5 「qqPlot.exe」の終了画面

なお、画面-4(確率プロットの描画画面)の右側にあるスクロールバーを下方にドラッグすると、画面-3 が徐々に出現する。これは、プログラムの技法上、画面-4の描画画面は画面-3にカーテンを掛けるように作 成したものであることによる。

この節に記述した一連の操作手順が理解できるように、フローチャートをAppendixに示した。

### 4 解析結果の事例解説

ここでは、プログラム開発時のテスト用として用いた3つのサンプルデータについて、出力結果を参考に、事 例を解説する。

## 4.1 次代検定林における家系別樹高の例

ある次代検定林の20年次データの分析を行い,対照1家系を含む32家系別の樹高について,表-1に示すよう に最小二乗推定値を求めた。表-1で,第1列目にある数値が各家系の樹高の最小二乗推定値であり,第3列目は 当該家系の精英樹コードで,「32」は対照家系である。

最小二乗推定値は、最小12.776m~最大14.076mの範囲に分布し、大差がほとんど認められないように見受けられる。

しかし、このデータに正規分布モデルで通常の基準化を行った正規確率プロットは、先に示した画面-4で描 画された。実際の描画はカラーで出力されるので、対照家系が下位から3番目に赤でプロットされ、他の精英樹 家系との対比は明確である。

分散分析の結果も家系間に有意な差を認めたが、表-1に示す最小二乗推定値から、その事実を認識すること は困難である。また、この事例のように、一次元である値を二次元に展開できることも、有効な方法の一つと言 える。

## 4.2 ミニチュア採種園の中央に黄金スギを植栽して花粉飛散を推定した例

表-2に示したデータは、黄金スギを中央に植栽したミニチュア採種園で、植栽木から採取した種子を播種し、 その種子から発芽した苗で分離した黄金スギ様苗の分離状況を植栽木ごとにまとめたものである。

ミニチュア採種園は9行×9列の大きさで2区画が造成され、いずれの区画も1本の黄金スギが中央に植栽され、 その周囲に80本のスギ(「出羽の雪1号」と「出羽の雪2号」)が植栽された。黄金スギの花粉を受粉してでき た種子からは、黄金スギ様苗が発芽するので、花粉の飛散状況や飛散距離を推測する実験にはよく使われる。

さて、表-2に示すように、種子を採取した植栽木ごとに、その種子から発芽した黄金スギ様苗の出現率(%) を1列目に、また、当該植栽木の植栽位置の座標(行No.と列No.)を3列目と4列目に、それぞれ記したファイル がある。黄金スギ様苗の出現率は、最小0.568%~最大40%に分布し、数値そのものからは分布型の特定は困難と 推察された。

実際,このデータに正規分布モデルで通常の基準化を行った正規確率プロットを作成すると,図-1が描画され,正規分布とは明らかに異なるデータであることが確認された。

- 445 -

林木育種センター研究報告 第23号



黄金スギ様苗の出現率の正規確率プロット(正規分布の適用)

しかし、同じデータに対数正規分布を適用し、ロバスト推定値で基準化を行った正規確率プロットは、図-2 のように、ほぼ直線にプロットされた。



図-2 ミニチュア採種園の植栽木から採取した種子から発芽した苗で分離した 黄金スギ様苗の出現率の正規確率プロット(対数正規分布の適用)

- 446 -

このことから、黄金スギ様苗の出現率は、対数正規分布に従う変量であることが理解できる。

なお,「3.2 データファイルの準備」に記したように,「qqPlot.exe」では,表-2の内容を持つファイルでも 第1列目と第2列目の数値が有効であり,第3列目以降の数値は解析には影響しない。

# 4.3 受精別のトウモロコシにおける個体の高さ

表-3に示すデータの原典は、「ゴルトンが解析したダーウィンのデータであり、その解析をフィッシャーが 批判したもの<sup>5</sup>。」である。データは、他家受精、自家受精別のトウモロコシにおける個体別の高さで、引用文 献<sup>5</sup>では他家受精のデータについて正規確率プロットを作成する事例が紹介されている。

プログラム「qqPlot.exe」によって表-3のデータを解析し、結果をファイルに出力した例を表-5に示す。

Fitting model: Normal distribution									
Stand	lardizati	on; Robus	t method						
Media	an= 21.5,	Semiquar	tile range	e= 3					
		< <t o="" p<="" th=""><th>&gt;&gt;</th><th></th><th></th><th></th><th></th><th></th><th></th></t>	>>						
No.	Data	Scaling	(Norm.P)	z.plot	:				
1.	12.000	-4. 2735	(0. 033)	-1. 834					
2.	12.000	-4. 2735	(0. 100)	-1. 282					
3.	18. 250	-1. 4620	(0. 167)	-0.967					
4.	19.125	-1.0684	(0. 233)	-0. 728					
5.	20. 375	-0. 5061	(0.300)	-0. 524					
6.	21.000	-0. 2249	(0. 367)	-0. 341					
7.	21.000	-0. 2249	(0. 433)	-0.168					
8.	21. 500	0.0000	(0. 500)	0. 000					
9.	21. 625	0. 0562	(0. 433)	0. 168					
10.	22.000	0. 2249	(0.367)	0.341					
11.	22. 125	0. 2812	(0.300)	0. 524					
12.	22. 125	0. 2812	(0. 233)	0. 728					
13.	23. 000	0. 6748	(0. 167)	0.967					
14.	23. 250	0. 7872	(0. 100)	1. 282					
15.	23.500	0.8997	(0.033)	1.834					

表-5 解析結果の出力例(表-3のデータ)

- 447 -

また,引用文献<sup>5</sup>と同様,当該データに正規分布を適用し,ロバスト推定値で基準化し,さらに,確率プロットで縦軸に変換したデータの値,横軸に確率を設定して作成した確率プロットを図-3に示す。 詳細は,引用文献<sup>5</sup>を参照されたい。



正規確率プロット(正規分布の適用、横軸が確率目盛)

# 5 おわりに

パソコンの発達は、データ解析、論文作成、成果発表等、知的生産活動に大きな利便をもたらした。

しかし、ここで報告した正規確率プロットについては、未だに正規確率紙を使う方法が紹介されている実態が あり、具体的に所与のデータについて確率プロットを作成する手順は不明確な場合が多い。正規確率プロットを 作成するプログラム「qqPlot.exe」自体は、解析を行うためのソフトとは言えないが、「所与のデータの正規性 を視覚的に容易に検定するためのツールである。」とは言える。

現在のバージョンでは、ファイルからそのままの値を扱う正規分布と、それを自然対数値に変換した値を扱う 対数正規分布の二通りしか扱えない。しかし、データ変換などのルーチンを組み込み、それに応じてプログラム を修正すれば、2重指数分布やワイブル分布など、他の確率分布についても確率プロットを容易に作成できるソ フトに改良できる。

今後、そうした改良も行われ、有効なソフトの一つになることを期待して、この報告を終える。

- 448 -

# 謝 辞

この報告をとりまとめるにあたり、プログラム「qqPlot.exe」の作動確認および報告内容との整合性のチェックを遺伝資源部保存評価課長の久保田正裕氏にお願いした。同氏には、作動確認、報告内容のチェックのほか、 プログラムの改良点も指摘いただいた。

記して、謝意を表す。

# 補足

この報告のとりまとめの後、プログラム「qqPlot.exe」を改良した。改良は、所与のデータの確率プロットを 作成する際、描画する小円を塗りつぶせるようにしたことである。

塗りつぶした小円で描画を作成したい場合は、「3.2 データファイルの準備」に記載した解析用ファイルの作 成で、描画色を指定する二番目の数値の直後にスペースを入れずに「9 f」、「12 F」のように半角英字の「f」 や「F」を記述する。もちろん、「f」や「F」を記述した行と、そうでない行とが混在しても差し支えない。 なお、描画色の指定は、表-4のとおり変更はない。

# 引用文献

1) 中里博明・武田知己:二項確率紙の使い方(改訂版),日科技連,64pp. (1965)

2) 柴田義貞:正規分布 ——特性と応用——, 東京大学出版会, 307pp. (1981)

3) 新村秀一:パソコンによるデータ解析 ― 統計ソフトを使いこなす― , 講談社 (ブルーバックス), 318pp. (1995)

4) 竹内 啓:確率分布と統計解析,日本規格協会,314pp. (1977)

5) 竹内 啓·大橋靖雄:統計的推測——2標本問題, 日本評論社, 192pp. (1981)

- 449 -



プログラム「qqPlot.exe」の操作手順

- 450 -