外国語教育メディア学会 (LET) 関西支部メソドロジー研究部会 2016 年度第 9 号報告論集
草薙邦広 (pp. 1–19)

外国語教育研究における数値シミュレーションの基礎
—さまざまな分布の下での乱数生成によるデータの復元—

草薙 邦広
広島大学

概要

本稿の目的は，外国語教育をその応用分野として念頭に置いたシミュレーション研究の基礎を概観することである。特に，ある分布における母数の推定値が所与である状況下において，その分布の確率密度関数にしたがう乱数を生成することで，元データと遜色ないデータを復元する手続きについて紹介する。本稿は，連続確率分布に視野を留めて，（a）正規分布，（b）ガンマ分布，（c）ワイブル分布，（d）指数正規合成分布，（e）多変量正規分布，さらに非常に有益な手法である（f）カーネル密度推定，（g）ピアソン分布システム，（h）ジョンソン分布システムによる乱数生成を取り扱う。

Keywords: シミュレーション，乱数の生成，最尤推定，再現可能性，確率分布

1. 背景

外国語教育研究においては，従来より記述統計の報告を徹底することが重要であるとされてきた。これは，外国語教育研究に限らず，いうまでもなくその重要性が明らかな指針ではあるが，実際のところ，記述統計の報告に問題があった過去の研究例は多い。

記述統計の報告を欠くことの根本的な問題点は，単純にその研究の再現可能性を大幅に低下させることである。逆に，記述統計が適切に報告されている場合，記述統計の値より解析的に分析結果を再現することができる場合もある。また，ある検定の統計量や有意確率がもとめられない場合においても，シミュレーションによって元データ自体を復元（data recovery）することができれば，その研究はある点において，再現可能なものとなりうる。

再現可能性の問題は，統計科学を応用するどのような分野にも共通することであるが，昨今，特に心理学分野において注目を浴びる話題となった。これに合わせ，研究成果のみならず，研究過程をもオープン化することによって，再現可能性を向上させようとする動きも急速な広がりを見せている。もちろん，このような動きが外国語教育研究において例外となるような理由はない。再現可能性の問題は今後ますます重要となってくるといえよう。

再現可能性の問題は，「再現可能である」とはそもそもどういうことか，という抽象

的な問題を孕んでおり，これは本稿の限界を超える。しかしながら，再現可能性の中でも最も基本的な，「研究報告から遜色ないデータの復元が可能か」という点は，現在の外国語教育研究，そして当該分野の将来的発展の中で，重要な観点になりうる。そうとはいうものの，現状の外国語教育研究では，確率分布などに関する研究者の理解が十分であるとはいえず，さらにデータ復元ないし数値シミュレーションに関する技術の普及は，やや後進的であるといわざるを得ない。

本稿では，このような現在の状況を受けて，既存の分布，確率密度関数，そして任意の母数の下でのデータ復元について，その技術的概論を述べることとする。この手続きは，まさに数値シミュレーションの一種である。しかし本稿は，その大風呂敷のような題目に反して，シミュレーション研究全体の概論を述べるわけではない。また，具体的に取り扱う分布は，（a）正規分布，（b）ガンマ分布，（c）ワイブル分布，（d）指数正規合成分布，（e）多変量正規分布，そして関連する統計手法である，（f）カーネル密度推定，（g）ピアソン分布システム，（h）ジョンソン分布システムといった連続確率分布に関連するもののみである。これらは極めて限定的な内容であるものの，著者は，本稿がシミュレーション研究へのよい導入，そして研究対象に対して確率論的な見方を始める助けになることを密かに望んでもいる。

本稿は，外国語教育研究の入門者を対象としている。そのため，数学的原理の説明，そして得てして嫌われがちな数式のしつこい提示などを避けた。上記のような点については，より優れた統計学の入門書を参照されるとよい。また，全編に渡り，実用性も考慮して，R（R Core Team, 2016）の解析コードを付記した。

2. さまざまな分布の下での乱数生成

2.1 基本的用語

本稿の内容に入る前に，本稿が使用する基本的な統計用語の説明を行いたい。もちろん，ここでの説明は簡易的なものであり，より正確な定義については，各用語毎に統計学の入門書を参照されるとよい。

1. **確率変数**（random variable）: ある確率（確率分布）にしたがって，さまざまな値を取る変数。サイコロの目など。
2. **確率分布**（probability distribution）: 確率変数のある値についての起こりやすさを与えるもの。公正なコインで表と裏の目が出る値は 1/2 である，など。これを関数で表す。統計学上，正規分布など，さまざまな関数の種類が知られている。
3. **確率密度関数**（probability density function, PDF）: 連続的な値をとる確率変数（**連続確率分布**）についての確率分布を与える関数。この関数内の密度を積分することで，確率変数が当該の範囲となる確率を求めることができる。統計学

に関して訓練経験のない研究者にとって，大概解読不能な式で与えられる。離散型の確率変数の場合，**確率質量関数**などという。

4. **累積分布関数**（cumulative distribution function，CDF）: 確率変数がある任意の値以下を取る確率を表す関数。
5. **母数**（parameter）: ある確率分布を特徴づけるもの。確率密度関数の項として使用される。たとえば，正規分布の確率密度関数は，**平均**（μ），**標準偏差**（σ）という母数をもち，これらの母数が確率密度関数の項に入る。
6. **最尤推定**（maximum likelihood estimation，ML，MLE）: ある所与のデータから，任意の分布における母数を推定する方法のひとつ。一般に，大きなデータでないと正確な推定ができない場合が多い。
7. **モーメント法**（method of moments）: ある所与のデータから，ある分布における母数を推定する方法のひとつ。**モーメント**を計算にもちいる。場合によっては小さなデータに対して行うこともある。
8. **モーメント（積率）**: 統計量の一種。一次のモーメントを平均，二次のモーメントを**分散**（variance），三次のモーメントを**歪度**（skewness），四次のモーメントを**尖度**（kurtosis）という。
9. **乱数の生成**: ある確率密度関数，または確率質量関数にしたがう確率変数を人工的に生成すること。近年は計算機器の発展により，**メルセンヌ・ツイスタ法**など，かなり高精度な乱数生成が容易にできるようになってきている。なお，本稿でいう乱数とはすべて擬似乱数，ある意味，一種の確率変数のことである。
10. **モンテカルロ法**（monte carlo method）: 乱数の生成をもちいた数値解析手法のこと。現在は，さまざまな統計手法で**ブートストラップ法**や**マルコフ連鎖モンテカルロ法**が実装され，統計手法の主要な一部となっている。
11. **情報量基準**（information criterion）: **赤池情報量基準**（AIC），**ベイズ情報量基準**（BIC）などが知られる。データがモデルに適合しており，さらにモデルが複雑でない程度，つまり「モデリングのよさ」をあらわすと考えておけば問題ない。

2.2 正規分布

基本的な用語の確認ができたところで，まずは正規分布を題材として乱数生成の基礎について見てみる。ある研究論文において，確率変数 x の平均（M）と標準偏差（SD）が報告されていたとする。ここから，この確率変数 x に遜色ないデータを得たいと考える。

正規分布の確率密度関数は，平均と標準偏差のみを母数にもつ。一般に，確率変数の算術平均と標本によって推定された不偏母標準偏差推定値は，母数の推定値として望ましい性質をもつ。そのため，これらの値を母数としてみなし，正規分布の確率密度関数と所与の母数から確率変数を得ることができる。さまざまな統計解析環境で正規分布の確率密

度関数にしたがう乱数を生成できるが，たとえば，R ならば，デフォルトで搭載されている stats パッケージ（R Core Team, 2016）内の `rnorm` 関数を使うことで乱数を生成できる。

```
rnorm(100, 50, 10)
```

この `rnorm` 関数では，最初の項が生成する乱数の数，二項目が平均，三項目が標準偏差である。これで得られた変数の度数分布を確認するには，

```
dat<-rnorm(100, 50, 10)
hist(dat,col="lightblue")
```

とでもすればよい。

もちろん，確率変数が先行する観測によって所与のとき，平均と標準偏差をもとめる関数を使い母数の推定値を得て，これらの母数をもとに，正規分布にしたがう乱数を再度生成することもできる。これは当初の確率変数と遜色ない乱数を再現すること，つまりデータの復元ともいえる。

```
mdat<-mean(dat)
sdat<-sd(dat)
dat2<-rnorm(100,mdat,sdat)
```

ちなみに，stats パッケージには，`dnorm` および `pnorm` という関数もある。前者は正規分布の確率密度関数を，後者は累積分布関数を返す。たとえば，

```
x<-seq(0,100,.01)
p<-dnorm(x,50,10)
plot(x,p,xlab="x",ylab="p",type="l")
```

とすれば，図 1 のような確率密度関数を描くことができる。

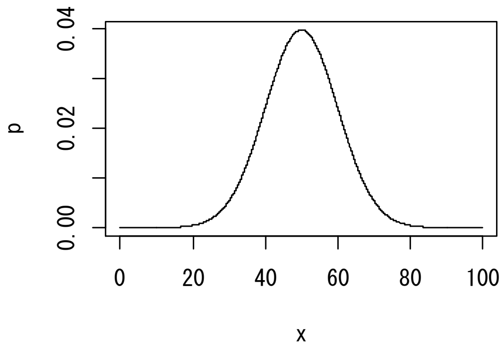

図 1. 正規分布の確率密度関数

また，

```
x<-seq(0,100,.01)
p<-pnorm(x,50,10)
plot(x,p,xlab="x",ylab="p",type="l")
```

とすれば，図 2 のように累積分布関数を描くことができる。

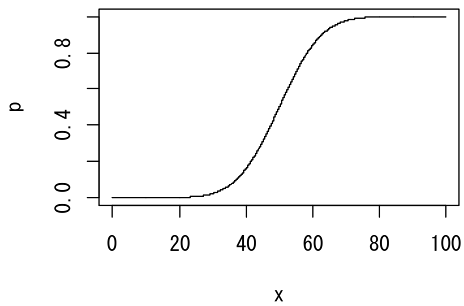

図 2. 正規分布の累積分布関数

もちろん，この手続きは任意の母数にしたがう確率変数を生成しているだけであり，この乱数セットの「標本平均」や「標本標準偏差」が設定した母数と必ず一致するわけではない。また，この手続きをするたびに異なる乱数セットが得られる。R 上で同じ乱数セットを生成するためには，`set.seed` 関数などをもちいるとよい。たとえば，

```
set.seed(100)
rnorm(100,50,10)
```

とすれば，同じセットの乱数が得られる。

正規分布は，中心極限定理などによって，さまざまな統計分析の中でも主要な役割を果たす。しかしながら，我々，外国語教育に携わるものが得る観測が，押しなべて正規分布にしたがう確率変数であるとは限らない。ある教材の学習時間，心理学実験における反応時間データ，読解時間データなどは，正規分布にしたがわない確率変数の代表である。

2.3 ガンマ分布

外国語教育に関わる時間データは，ガンマ分布として扱うほうが賢明な場合もある。ガンマ分布は，正規分布とは異なり，歪んだ，裾の重い分布を表現できるからである。そのため，ガンマ分布は時間のみならず，社会学や経済学においては所得の分布をモデル化するためにも使用される。ガンマ分布の確率密度関数は，形状母数 k，尺度母数 θ という 2 つの母数をもつ。また，主にベイズ統計による分析を行う場合には，形状母数を α，尺度母数の代わりに比率母数，ないし逆尺度母数を β と定義するときがある。実質的にこれらの扱いに大きな差はなく，（1）式のような関係にある。

$$\alpha = k \qquad (1.a)$$

$$\beta = \frac{1}{\theta} \qquad (1.b)$$

ガンマ分布には，統計上，便利な性質があり，形状母数が 1 である場合は指数分布に，整数である場合にはアーラン分布に帰着する。また，これらの母数は正でなければならない。

たとえば形状母数 3，尺度母数 0.1 のガンマ分布から乱数を生成するためには，`rgamma` 関数をもちいるとよい。これも stats パッケージに実装されている。

```
dat<-rgamma(100,3,1/10)
hist(dat)
```

同様に，確率密度関数，累積分布関数を以下のような方法で得ることもできる。これを描くと図 3 のようになる。

```
#確率密度関数
x<-seq(0,1500,1/10)
p<-dgamma(x,3,.1/10)



#累積分布関数
x2<-seq(0,1500,1/10)
p2<-pgamma(x,3,1/10)

#描画
par(mfrow=c(1,2))
plot(x,p,xlab="x",ylab="p",type="l",main="PDF")
plot(x2,p2,xlab="x",ylab="p",type="l",main="CDF")
```

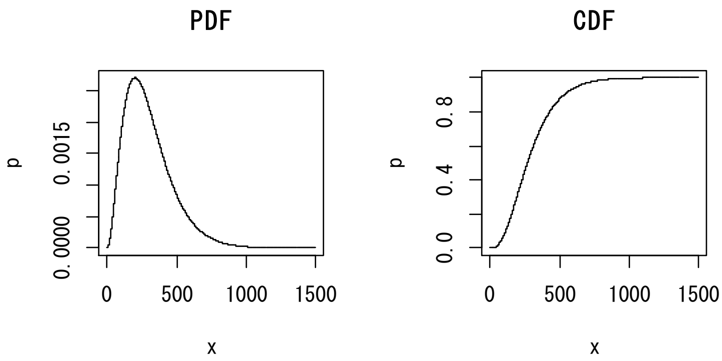

図 3. ガンマ分布の確率密度関数と累積分布関数の例

所与のデータから，ガンマ分布の母数を得るためには，最尤推定が最も一般的な方法である。最尤推定によって，ガンマ分布にデータをフィットさせ，母数を推定するには，MASS パッケージ（Venables & Ripley, 2002）の汎用確率分布フィット関数，`fitdistr` 関数をもちいるとよい。

ここでは，`rgamma` 関数で生成した乱数をもとに，最尤推定によって母数を推定する手続きをしてみる。

```
#パッケージのロード
library(MASS)

#乱数の生成
dat<-rgamma(100,3,1/10)

#最尤推定
model<-fitdistr(dat,densfun="gamma")
model
```

おそらく，およそ遜色ない値が得られるはずである。最尤推定を行ったときには，当該の分布のモデルがどの程度望ましいか，情報量基準によって報告するとよい。

```
AIC(model)
BIC(model)
```

とすれば，それぞれ赤池情報量基準，ベイズ情報量基準の値を返す。なお，誤差の大きさは，

```
model$sd
```

で知ることができる。

2.4 ワイブル分布

ワイブル分布は工学研究において頻繁に使用される分布であり，反応時間の解析などにも援用されることがある。ワイブル分布は，尺度母数 λ，形状母数 k をもつ。R において，ワイブル分布は，ガンマ分布と同じく，以下のように手続きで分析することができる。

```
#乱数の生成
dat<-rweibull(100,1,10)

#確率密度関数
x<-seq(0,100,1/10)
p<-dweibull(x,1,10)

#累積分布関数
x2<-seq(0,100,1/10)
p2<-pweibull(x,1,10)

#最尤推定
model<-fitdistr(dat,densfun="weibull")
model
AIC(model)
BIC(model)
model$sd
```

この他にも，本稿は取り上げないが，`fitdistr` 関数はさまざまな分布を取り扱っている。対数正規分布やポアソン分布などがそうである。

2.5 指数正規合成分布

指数正規合成分布ないし，ex-Gaussian 分布は，近年になって，国内の外国語教育研究にて頻繁に使用されるようになった。これは，反応時間のモデリングをするためのものとして，心理学においては古くから知られている。指数正規合成分布は，その名が示すように，指数分布と正規分布両方の特性を引き継いでいる。母数として，正規分布の平均（μ），標準偏差（σ），そして指数分布由来の τ をもつ。

R において指数正規合成分布を扱うには，反応時間解析の専用パッケージである retimes パッケージ（Massidda, 2013）をもちいるとよい。反応時間の分布に関しては，莫大な研究数があるが，一般に，指数正規合成分布は反応時間によくフィットし，解析のツールとして必須のもののひとつとされている。指数正規合成分布についての R コードは以下のとおりである。

```
#パッケージのロード
library(retimes)

#乱数の生成
dat<-rexgauss(100,500,100,400)

#確率密度関数
x<-seq(0,3000,1/10)
p<-dexgauss(x,500,100,400)

#最尤推定
model<-timefit(dat)
model
AIC(model)
BIC(model)
        #誤差をもとめるためにブートストラップをもちいるがここでは省略

#モーメント法
mexgauss(dat)
        #標本サイズが小さい場合，モーメント法の方が好まれる
```

2.6 多変量正規分布

ここまでは，単変量の乱数生成と母数の推定について述べてきた。しかし，外国語教育研究は，本質的には潜在変数を対象にするものであり，実際に多変量データを扱う場合も多い。ここでは多変量データを対象とした乱数の生成について見ていく。

多変量正規分布は，因子分析や構造方程式モデリングなどにおいて条件となる分布であり，母数として，変数の数に等しい長さの平均ベクトル（μ），そして変数の数だけの大きさをもつ分散共分散行列（∑）をもつ。

平均ベクトルとは，n 個の変数のとき，

$$\mu = (\mu_1, \mu_2, \mu_3 \dots, \mu_n) \tag{2}$$

というようになる。

分散共分散行列の対角要素は分散であり，それ以外の要素は共分散である。分散共分散行列は必ず対称行列になる。まずは，（3）に 2 変量の分散共分散行列の例をあげる。

$$\Sigma = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix} \tag{3}$$

この例では，それぞれの変数の分散が 100，そして共分散が 0 である。これは対角行列である。

一方，（4）では，変数間の共分散行列が 50 であることを示している。この例では相関係数 r が.50 となる。

$$\Sigma = \begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix} \tag{4}$$

多変量正規分布にしたがう乱数を生成するためには，MASS パッケージの `mvrnorm` 関数を使うとよい。たとえば，平均ベクトルが，

$$\mu = (50, 60, 70) \tag{5}$$

であり，分散共分散行列が，

$$\Sigma = \begin{bmatrix} 100 & 80 & 30 \\ 80 & 120 & 60 \\ 30 & 60 & 140 \end{bmatrix} \tag{6}$$

であるような多変量正規分布の場合には，以下のようなコードを使う。

```
#平均ベクトルと分散共分散行列の作成
m<-c(50,60,70)
s<-matrix(c(100,80,30,80,120,60,30,60,140),3,3)

#乱数の生成
dat<-mvrnorm(100,m,s)
```

所与のデータから，平均ベクトルと分散共分散行列をもとめるには，以下のようにするとよい。

```
#平均ベクトル
colMeans(dat)

#分散共分散行列
cov(dat)
```

なお，多変量正規分布の確率密度関数などは，mvtnorm パッケージ（Genz et al., 2016）で実装されているが，やや複雑なので本稿では省略する。

もちろん，通常の正規分布同様，外国語教育に関するデータが多変量正規分布にうまくフィットするとは限らないことにも注意が必要である。

ちなみに，多変量正規分布における各値の中心からの距離には，マハラノビス距離を適用する。マハラノビス距離を R 上でもとめるには，以下のようにするとよい。

```
#マハラノビス距離
mahalanobis(dat,m,s)
```

3. より実践的な乱数の生成

3.1 カーネル密度推定

この節より，より柔軟で実践的な方法について紹介していく。まずは，カーネル密度推定である。この手法は，所与のデータから，確率密度関数を推定するもので，よくも悪くも，ノンパラメトリックな手法である。カーネル密度推定で得られた確率密度関数は，単純な母数を持たないが，その代わりにバンド幅（band width）の指定が必要で，これは分析者が設定するものである。なお，ここでのカーネル密度推定はガウス関数にもとづく。

バンド幅の設定によって，推定の結果得られる確率密度関数が，大きく異なるものとなることもある。バンド幅の設定方法には，シルバーマンの経験則などをはじめとしてさまざまなものがあるが，よほどでない限り，最近のツールでは，それらのデフォルト設定で問題ないこともある。

R にはデフォルトの状態で density 関数というカーネル密度推定の関数があるが，著者は ks パッケージ（Duong, T., 2016）を使用している。当該のパッケージには，データを

指定し，そこから直接的に確率密度関数を得る kde 関数がある。この関数で得たオブジェクト（確率密度関数）は，そのままプロットしたり，乱数生成に使えたりする。

カーネル密度推定の手法上のメリットを見るために，図 4 のようなヒストグラムで表されるような奇妙な確率変数があるとする。これは，以下のようなコードで作成する。

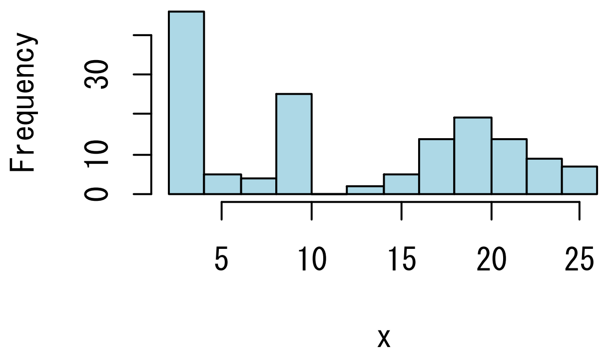

図 4. 奇妙な確率変数の例

```
dat<-c(rexp(50,1)+2,10-rexp(30,1),rnorm(70,20,3))
hist(dat,xlab="x",main="",col="lightblue")
```

このデータにカーネル密度推定を行って得られた確率密度関数は，図 5 のようなものになる。

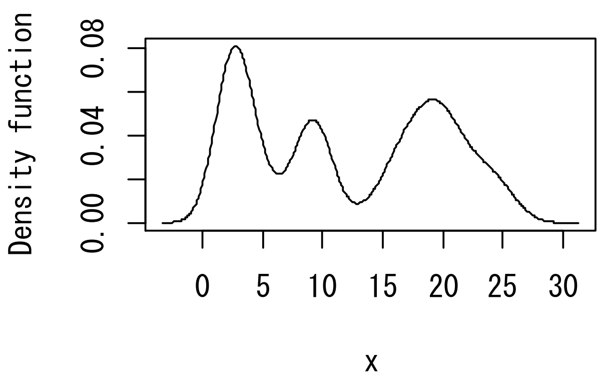

図 5. カーネル密度推定の例

この図に至るコードは以下の通りである。また，こうして得られた確率密度関数から，乱数を生成することができる。この乱数の度数分布は図 6 のようになり，ほぼ分布の特徴を正確に再現できていることがわかる。

```
#データの作成
dat<-c(rexp(50,1)+2,10-rexp(30,1),rnorm(70,20,3))

#カーネル密度推定（バンド幅に注意すること）
k<-kde(dat)
plot(k)

#乱数の生成とヒストグラムによる可視化
dat2<-rkde(fhat=k,150)

hist(dat2,xlab="x",main="",col="lightblue",breaks=10)
```

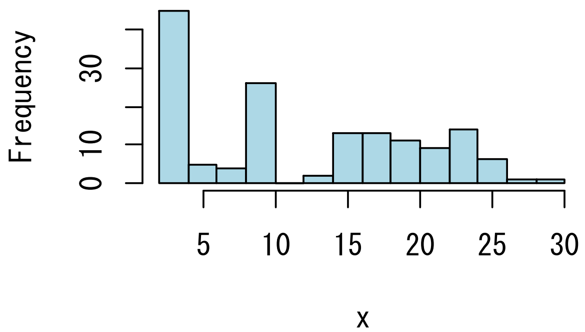

図 6. カーネル密度推定で再現された乱数の度数分布

カーネル密度推定は，多変量にも応用可能であり，非常に有益であるが，これについては省略する。

カーネル密度推定はこのように非常にパワフルな道具なのだが，最初に述べたように，よくも悪くもノンパラメトリックな手法であるということを忘れてはならない。このように推定された確率密度関数が，なにか実質科学的な知見を必ずしも直接的に示すわけではないのである。ある現象を表す確率変数が「どのような分布を取るか」,「そしてそれはどのような因果メカニズムで説明できるか」は，統計科学を応用する研究分野における，最も重要で，そしてしばしば忘れられてしまいがちなことである。このような面において，カーネル密度推定が常に優れた手法とはいいがたい。

また，パラメトリックな手法には，いくつかの母数の値のみから乱数を生成できる，そしてデータ復元できる，という非常に効率的で望ましい性質があることも常々意識した

いものである。

3.2 ピアソン分布システムとジョンソン分布システム

これらは統計学の中では，ともに非常に古典的なものであるが，近年になって注目を浴びている手法でもある。前者はもちろん，カール・ピアソン（Karl Pearson）の「歪んだ分布族」として知られる業績であり，後者は，ノーマン・ジョンソン（Norman Johnson）による手法である。まず，ピアソン分布システムの概要について説明するが，数学的に高度であるため，実用的な要点だけを述べるに留まることをご容赦願いたい。

ピアソン分布システムは，ある微分方程式を基盤とし，それを満たす係数の制約によって 7 ないし 12 種類の分布を分類する。その微分方程式の係数は 4 つあり（a, b, c, d，または位置母数，尺度母数，第一形状母数，第二形状母数），これらは積率の関数でもある。つまり，任意の平均，標準偏差，歪度，尖度を組み込んだ確率密度関数をもとめることができる。R で，ピアソン分布システムを扱うには，PearsonDS（Becker & Klosner, 2016）パッケージを使うとよい。コードは以下を参照されたい。

```
#パッケージのロード
library(PearsonDS)

#歪んだデータの作成
dat<-log(rnorm(100,50,10))

#最尤推定によるパラメターの取得
para<-pearsonFitML(dat)

#乱数の生成とヒストグラムによる可視化
dat2<-rpearson(100,para)


#確率密度関数
x<-seq(0,5,.01);p<-dpearson(x,para)

#累積分布関数
x<-seq(0,5,.01);p<-ppearson(x,para)
```

一方，ジョンソン分布システムも同様に使うことができるが，その設計思想はやや異なる。ジョンソン分布システムは，4 つの母数を持つ。

1. γ
2. ξ
3. δ
4. λ

R の SuppDists パッケージ（Wheeler, 2016）では，ジョンソン分布システムを扱っている。コードについては，ピアソン分布システムとほぼ同様のコードで分析できる。以下を参照されたい。

```
#パッケージのロード
library(SuppDists)

#歪んだデータの作成
dat<-log(rnorm(100,50,10))

#分位点法によるパラメターの取得
para<-JohnsonFit(dat)

#乱数の生成
dat2<-rJohnson(100,para)


#確率密度関数
x<-seq(0,5,.01);p<-dJohnson(x,para)

#累積分布関数
x<-seq(0,5,.01);p<-pJohnson(x,para)
```

これらピアソン分布システムおよびジョンソン分布システムの手法上の利点は，通常記述統計において報告されているはずのモーメントから乱数を生成できることである。た

とえば，ある研究論文において，以下のような表で記述統計が示されているとする。

表 1.
記述統計の例

	標本サイズ	平均	標準偏差	歪度	尖度
統制群	123	12.31	4.33	0.93	3.41
実験群	118	11.88	2.99	1.23	7.34

これらの推定値が妥当であれば，この記述統計から実際に当該の研究のデータを再現できる。通常の正規分布による乱数の生成であれば，歪度や尖度の情報は再現できないが，表 1 のように歪度と尖度が報告されてあれば，

```
control<-rpearson(123,moments=c(12.31,4.33,0.93,3.41))
experiment<-rpearson(118,moments=c(11.88,2.99,1.23,7.34))
```

というように分布の歪みも含めて再現できる。

3.3 パラメトリック・ブートストラップ

シミュレーション研究では，本来，このような乱数の生成を十分な回数反復し，解析的には求められない統計量の経験分布をもとめたりする。これはパラメトリック・ブートストラップといわれる手法でもある。

たとえば，表 1 で示されているデータの平均差の分布を経験分布として得て，パーセンタイル法で 95%信頼区間を構成するとする。以下のようなコードで簡単にもとめられる。ここでは，ブートストラップ回数（*B*）を 1,000 とする。

この平均差の 95%ブートストラップ信頼区間は，およそ [-0.03, 0.90] 程度になるはずである。しかしながら，この例はあくまでも手続きの例示のためのものであり，この手続きがことさら統計上望ましいとか，優れている，という点は含意しないことをご理解いただきたい。

```
m.c<-numeric(1000)
m.e<-numeric(1000)
md<-numeric(1000)

for(i in 1:1000){
m.c[i]<-mean(rpearson(123,moments=c(12.31,4.33,0.93,3.41)))
m.e[i]<-mean(rpearson(118,moments=c(11.88,2.99,1.23,7.34)))
  md[i]<-m.c[i]-m.e[i]
}
quantile(md,c(0.025,.975))
```

4. 総括

本稿では，ある特定の分布と任意の母数にもとづいて乱数を生成する手続きについて概説した。一般的なシミュレーション研究の利点である母数の誤差の検討や，より複雑な統計モデル下において母数の経験分布を得ることなどについて，詳しく触れることはしなかった。これは別の機会としたい。

結語となるが，そして，これはもはや予定調和的な展開でもあるのだが，どのような統計手法も，その新しい手法を学ぶひとが期待するほどの成果を出すことは，元来あまりないのである。本稿では，シミュレーション研究が今後の研究の突破口であるとか，今後シミュレーション研究を積極的に行うべきである，といった類の論に触れるつもりはまったくない。むしろ本稿の本当の目的は，分布や関数，そして確率変数といった，統計学の基礎知識について理解を深める機会を増やすことである。

外国語教育研究においては，統計改革や，それに関連する近年のめまぐるしい動きの中，とかく統計手法上の発展に追いつくことがさも急務であるかのような，そんな風潮が見られるようになった。同時に質的研究への関心の高まりも顕著に見て取れる。

しかし，昨今の統計手法の高度化といっても，その内実は，確率論的な世界観への移行という大きな流れの一部である。確率論的な世界観とは，我々の学術対象を，誤差も許さないような決定論的な因果法則にもとめるのではなく，むしろ，教育という営為，言語習得という現象を，柔軟に確率としてとらえることで，より現実的な理解を目指す態度である。そしてこれは，背景の節で触れた再現可能性の問題とも強く関連する。ある統計モデル，ないし分布の下でデータを再現する，という手続きについての知識は，件の態度を強く表すものに相違ないし，実務的にも研究の再現可能性を高める第一歩となる。

そのような流れの中，各分布の特徴，確率変数，そして母数の推定，そういった概念

に慣れ親しむ題材のひとつとして，本稿が少しでも外国語教育研究者，ないし関連諸分野の研究者の助けになれば幸いである。

参考文献

Becker, M., & Klosner, S. (2016). *PearsonDS: Pearson Distribution System*. R package version 0.98. https://CRAN.R-project.org/package=PearsonDS

Duong, T. (2016). *ks: Kernel Smoothing*. R package version 1.10.4. https://CRAN.R-project.org/package=ks

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Hothorn, T. (2016). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. http://CRAN.R-project.org/package=mvtnorm

Massidda, D. (2013). *retimes: Reaction Time Analysis.* R package version 0.1-2. https://CRAN.R-project.org/package=retimes

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Wheeler, B. (2016). *SuppDists: Supplementary Distributions*. R package version 1.1-9.4. https://CRAN.R-project.org/package=SuppDists