数理解析研究所講究録
第 1909 巻 2014 年 147-156

統計教材の要素と作成ツールの評価

東邦大学・理学部　高遠　節夫 (Setsuo Takato)
Faculty of Sciences,
Toho University
阿南工業高専　小柴　俊彦 (Toshihiko Koshiba)
Anan National College of Technology
弓削商船高専　野町　俊文 (Toshifumi Nomachi)
Yuge National College of Maritime Technology

1　はじめに

高等学校の新指導要領では，記述統計の基礎が数学Iに「データの分析」として必修化された．また，数学Bでは従来は数学Cにあった内容をまとめて「確率分布と統計的推測」が扱われ，統計がより重視されることとなった．

高専・大学初年級（以下，カレッジ級）でも，多くの学校で記述統計・推測統計の基礎的な内容を扱う科目が必修科目として置かれている．ここで言う「統計の基礎的な内容」とは，概ね以下の通りである（大日本図書「新確率統計」[1] の目次より引用）．

2章．データの整理

(2-1) 1次元のデータ
　度数分布，代表値，散布度，四分位と箱ひげ図

(2-2) 2次元のデータ
　相関，回帰直線

3章．確率分布

(3-1) 確率変数と確率分布
　確率変数と確率分布，二項分布，ポアソン分布，連続型確率分布，
　連続型確率変数の平均と分散，正規分布，二項分布と正規分布の関係

(3-2) 統計量と標本分布
　確率変数の関数，母集団と標本，統計量と標本分布，いろいろな確率分布

4章．推定と検定

(4-1) 母数の推定
　点推定，母平均と区間推定 (1)，母平均と区間推定 (2)，母分散の区間推定，
　母比率の区間推定

(4-2) 統計的検定
　仮説と検定，母平均の検定 (1)，母平均の検定 (2)，母分散の検定，
　等分散の検定，母平均の差の検定，母比率の検定

なお，教科書[1]では，1章に確率，補章にいろいろな検定・いろいろな確率分布・回帰分析が入るが，ここでは省略する．

著者らは，いずれも教科書[1]の編集に執筆または執筆協力として参画した．本書の執筆にはすべてTEXを用い，図表についてはScilab（[4]またはR（[5]）上のKETpic（[3]）により図表ファイルを作成して挿入した．また，著者らは，それぞれ大学，専門学校，高専において，統計の授業を担当しており，随時授業を補完する教材を配付または提示しているが，それらの作成にもTEXとKETpicは使用頻度の高いツールである．

他の数学の場合と同様に，数式の表現はカレッジ級の教材において欠かすことができない要素である．Wordなどでも可能ではあるが，数式入力の効率性，出来上がった数式の美しさでは，TEXの比ではない．カレッジ級の数学教員の7割近くがTEXを教材作成に用いている事実も自然であろう．

しかしながら，カレッジ級の教材においては，数式と同程度以上に図表の使用が不可欠である．特に，統計教材では表の利用が重要になる．巻末におかれる確率分布表は勿論であるが，本文においても頻繁に用いられる．例えば，教科書[1]の各章の図表を数えると，下表のようになる．

章	節	ページ	図	表
2	2-1	14	5	10
	2-2	8	8	6
3	3-1	21	21	5
	3-2	13	6	1
4	4-1	11	6	0
	4-2	15	11	2
計		82	57	24

データを並べたり度数分布表を作成するときなどに表を，確率分布（密度）関数の説明などに図を多用していることの現れである．

これから見られるように，統計の教科書や補助教材の作成では

数式，図，表

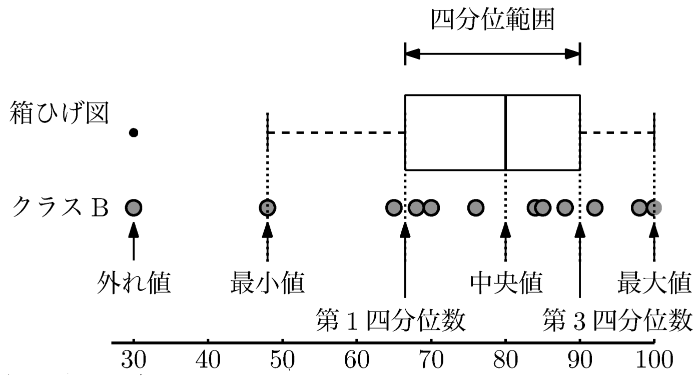
が必要不可欠な要素である．さらに，これらを効果的に配置するレイアウトも大切な要素となるが，TEXだけでは不十分である．すなわち，TEXの様々なパッケージは統計に必要な関数のすべてサポートされているとはいえず，また，TEXのtabular, array環境は多様な表を自由に作成する柔軟さ，容易さに欠ける．とりわけ，思い通りのレイアウトをTEXで実現するのは容易い作業ではない．KETpicは，これらのすべてを補完し，TEXとKETpicの組み合せにより初めて教員の求める教材の作成が可能になる．

以下，2節では，教科書[1]に現れる教材例とKETpicによる作図の概略，および他の方法との比較を述べる．3節では，配付・提示のための教材例を挙げることにする．

2　教科書からの教材例

本節の図表はすべて教科書 [1] からの引用であることを最初に断っておく．

視覚化としての図利用の 1 つは図式であろう．次は，箱ひげ図の説明図である．



R では関数`boxplot`により簡単に箱ひげ図を描くことができるが，教材のためには説明のための図式を付加することが不可欠である．

R には，確率分布の関数も整備されており，確率分布を表す語の前に，p, q, d, r を付加することにより，それぞれ分布関数，逆分布関数，密度関数，乱数生成が得られる．例えば，F 分布の密度関数

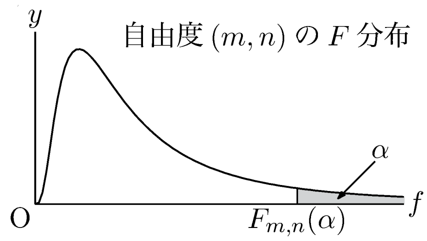
$$f(x)=\begin{cases}\dfrac{m^{\frac{1}{2}m}n^{\frac{1}{2}n}}{B\left(\dfrac{m}{2},\ \dfrac{n}{2}\right)}\dfrac{x^{\frac{1}{2}m-1}}{(mx+n)^{\frac{1}{2}(m+n)}} & (x>0\text{ のとき})\\ 0 & (x\leqq 0\text{ のとき})\end{cases}$$

を求める R の関数は単に rf である．

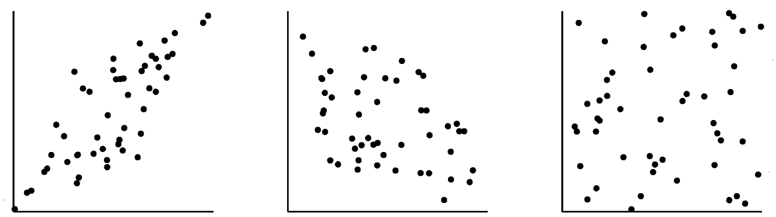
さらに，R 上の KETpic でグラフを描くためには

```
G<- Plotdata("rf(x,m,n)","x")
```

として，図データ G を TEX ファイルに書き出すだけでよい．



次の図は，相関係数の説明のために用いられたもので，一様乱数を発生させ，相関係数を表示している．



I　$(r=0.80)$　II　$(r=-0.40)$　III　$(r=0.10)$

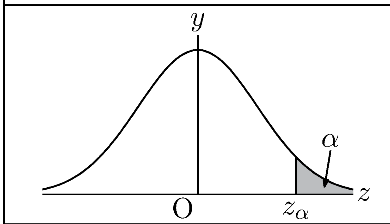
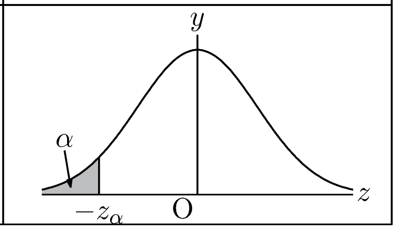
一様乱数の発生自体は TEX 上の多くのパッケージでもサポートされているが，ここでは，Scilab（R）のプログラミング機能を用いて，相関係数が小数第 1 位までの値になるようにしている．KETpic が CAS のマクロパッケージであることの利点の 1 つである．

前述したように，TeX の作表機能は貧弱であり，次のような単純な表の作成も容易ではない．

判定＼H_0 の真偽	H_0 が真	H_0 が偽 (H_1 が真)
H_0 を受容	正しい判断	**第 2 種の誤り**
H_0 を棄却	**第 1 種の誤り**	正しい判断

KETpic には，Tabledata を初めとする一連の作表コマンドが実装されており，大きくて複雑な表も作成可能である．その 1 つとして，各種の分布表は統計の教科書および教材では必須であるが，KETpic では，R（Scilab）のもつ確率密度関数とプログラミング（繰り返し文）により，必要な分布表がいつでも容易に得られる．

また，次のように字句，数式，図の入った表の作成も容易である．

<table>
<tr><th>右側検定</th><th>左側検定</th></tr>
<tr><td>$H_1 : \mu > \mu_0$，棄却域　$Z \geqq z_\alpha$</td><td>$H_1 : \mu < \mu_0$，棄却域　$Z \leqq -z_\alpha$</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

統計では，実データから表を作成することも多く，データの個数が大きいときは，TeX に直接書き込むのは手間がかかる．Scilab や R にはファイル入出力のコマンドが用意されており，KETpic でもそれらを利用したマクロが実装されている．

一例として，右の表を作成する R 上の KETpic のプログラムについて，その主要部分を以下に掲げておく．ここで，sampledata.csv は元の CSV ファイルである．

i	x_{1i}	x_{2i}	y_i
1	51	16	3.0
2	38	4	3.2
3	57	16	3.3
4	51	11	3.9
5	53	4	4.4
6	77	22	4.5
7	63	5	4.5
8	69	5	5.4
9	72	2	5.4
10	73	1	6.0

```
Dt<- Readtextdata("sampledata.csv")
Row<- nrow(Dt); Clm<- ncol(Dt)
Sv<- as.list(rep(8,Clm)); Sh<- as.list(rep(4,Row+1))
Tb<- Tabledata(Sv,Sh)
     ...
Putrowexpr(Tb,1,"c","i","x_{1i}","x_{2i}","y_i")
for(J in 1:Row){
  for(K in 1:(Clm-1)){
    Putcell(Tb,K,J+1,"c",Dt[J,K])
  }
  ...
}
```

3　配付・提示のための教材例

本節では，著者の一人（小柴）が作成した配付・提示教材を紹介する．

3.1　二項分布の正規分布による近似

二項分布 $B(n,p)$ の確率分布折れ線と，正規分布 $N(np,np(1-p))$ の確率密度関数のグラフを重ねて描くことは，二項分布 $B(n,p)$ が正規分布 $N(np,np(1-p))$ で近似されることを理解させるための教育的なよい方法である．ただそれは見た目に近づいていくことを伝える定性的な見方であるともいえる．そこで，KETpic を用いて定量的な実験観察を行なう教材を作成した．

$f_B(r)$: 二項分布 $B(n,p)$ の確率関数
$F_N(r)$: 正規分布 $N(np,\ np(1-p))$ の分布関数
$F = F_N(r+0.5) - F_N(r-0.5)$: 正規分布の分布関数による確率
$P(r-0.5 < X < r+0.5)$
正規分布による近似値との誤差：$G = F - f_B(r)$

とおくとき，誤差 G の絶対値の総和 $\sum|G|$ を各 $(n,\ p)$ について求めたものが下の表であり，よく知られた次の性質を読み取ることができる．

(1) pが0.5 に近いとき n が十分大きければ二項分布は正規分布により，よく近似される．
(2) $np > 5$ かつ $n(1-p) > 5$ のとき，二項分布は正規分布により，よく近似される．
実験例では誤差の絶対値の総和がすべて 0.1 より小である．
（太線の枠内が $np > 5$ かつ $n(1-p) > 5$ の場合）

n \ p	0.01	0.05	0.1	0.2	0.3	0.4	0.45
2	0.0390	0.0301	0.1668	0.2238	0.1642	0.1000	0.0813
3	0.0520	0.1253	0.2552	0.2090	0.1213	0.0510	0.0337
4	0.0547	0.1996	0.2845	0.1844	0.0901	0.0493	0.0339
5	0.0482	0.2523	0.2759	0.1483	0.0926	0.0416	0.0263
6	0.0352	0.2861	0.2603	0.1266	0.0824	0.0375	0.0278
10	0.0492	0.2964	0.1827	0.1005	0.0658	0.0330	0.0204
20	0.2241	0.1981	0.1293	0.0783	0.0497	0.0234	0.0128
30	0.3087	0.1817	0.1071	0.0679	0.0407	0.0192	0.0099
40	0.3290	0.1425	0.0995	0.0601	0.0346	0.0165	0.0084
50	0.3113	0.1308	0.0921	0.0531	0.0314	0.0147	0.0073
55	0.3095	0.1223	0.0909	0.0509	0.0301	0.0140	0.0070
80	0.2680	0.1090	0.0752	0.0423	0.0246	0.0116	0.0057
100	0.2097	0.0996	0.0671	0.0376	0.0219	0.0104	0.0051
500	0.1056	0.0467	0.0301	0.0169	0.0098	0.0046	0.0023

3.2 分布表における値の引き方

前述したように，カレッジ級の統計教育においては分布表がよく用いられ，KETpic により比較的容易に作成されるが，値の引き方を教えるのは，意外に手間取ることが多い．そこで，プロジェクタ提示用に，以下の教材を作成した．そのためには，KETpic のメタコマンド（TeX マクロを作るコマンド）によって作られた次のパッケージが重要である．

(1) `ketlayer.sty`

(2) `ketslide.sty`

(2) はプレゼンテーション作成用のマクロで，Beamer ほど高機能ではないが，提示教材作成用には手頃で，初めての教員にも使いやすいものである．また，(1) の中にある layer 環境は，1 ページの中の他の要素の位置を変えずに，新しい要素を付け加えることができて，教材作成には非常に有用である．特に，`ketslide` を利用してスライドを作成するとき，`layer` によって連続するページの全く同じ位置に少しずつ変化させた図表を入れることにより，アニメーション効果を作り出すことができる．正規分布表の引き方についての提示教材では，未知の値が入る枠内が空白の状態から出発して，例として，$P(z \leqq 1.15)$ の値を次の流れに従ってスライドを進めるようにした．

(1) 表の 13 行と 7 列の値を青色で表示する．

(2) 表の行と列の交差する部分にあるのが求める値であることを示すためにその値を赤色で表示する．同時に空白であった枠内に得られた値を表示する．

(3) 求まった値 0.8749 を枠内に表示する．

例 $P(z \leqq 1.15) = \boxed{}$

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

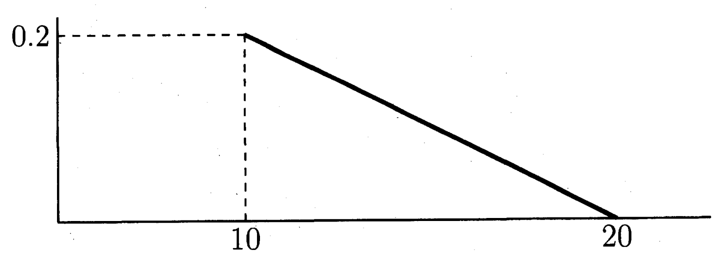
例 $P(z \leqq 1.15) = \boxed{0.8749}$

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

3.3 中心極限定理のシミュレーション

母集団がどんな分布であっても標本の大きさnを増やすと標本平均が正規分布に近づくという中心極限定理は，カレッジ級の統計教育においては重要な学習項目である．そこで，この中心極限定理を学生に実感させることを目的として，以下のシミュレーション教材を作成した．すなわち，任意の分布からn個の乱数を発生させ，標本平均の分布が変化する様子をアニメーションとして表示することにした．

母集団は図の密度関数$f(x)$を持つ分布とした．

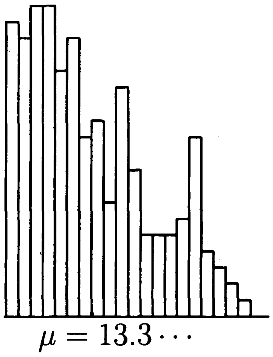


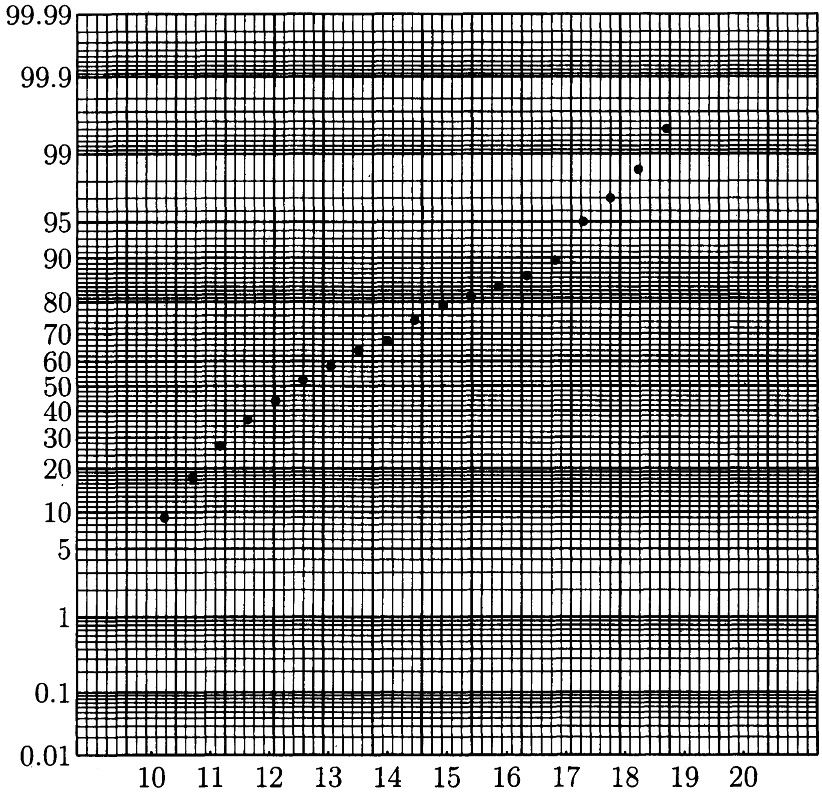
このような分布からの乱数発生の関数は一般にはサポートされていないが，分布関数の逆関数$x = F^{-1}(y)$を用いることにより容易に作成することができる．

$n = 1$のときは，元の分布を忠実に反映しているが，nを増やすにつれて，正規確率紙上のプロットが直線状になることがアニメーション効果により体感できる．

(**n** = 1の場合)

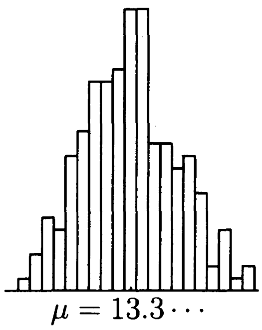
階級の幅
0.47

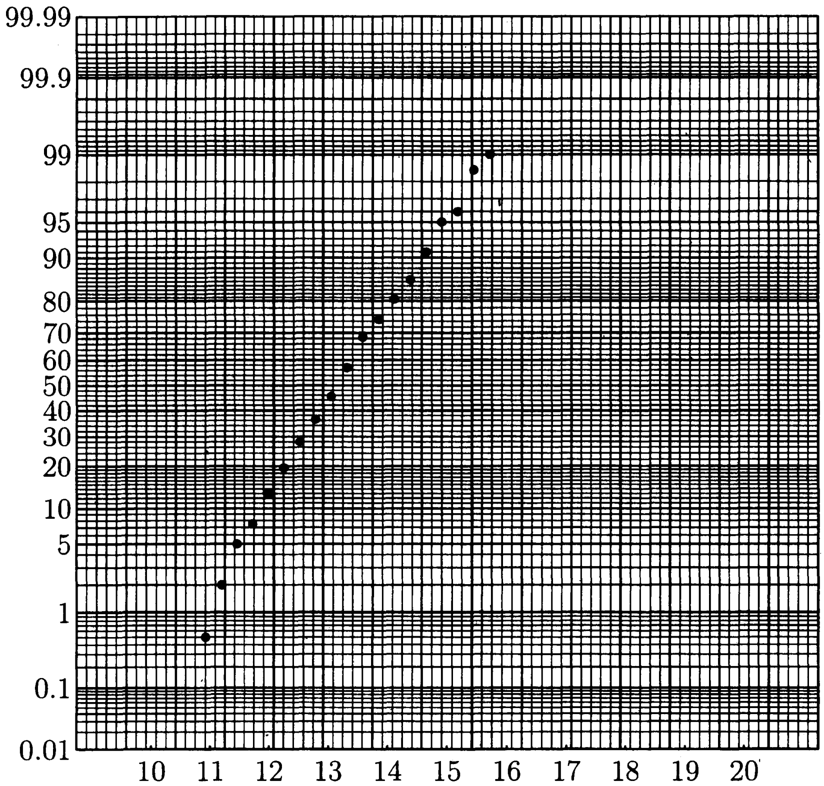




(**n** = 5の場合)

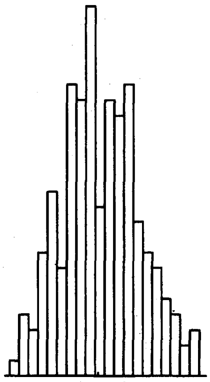
階級の幅
0.27



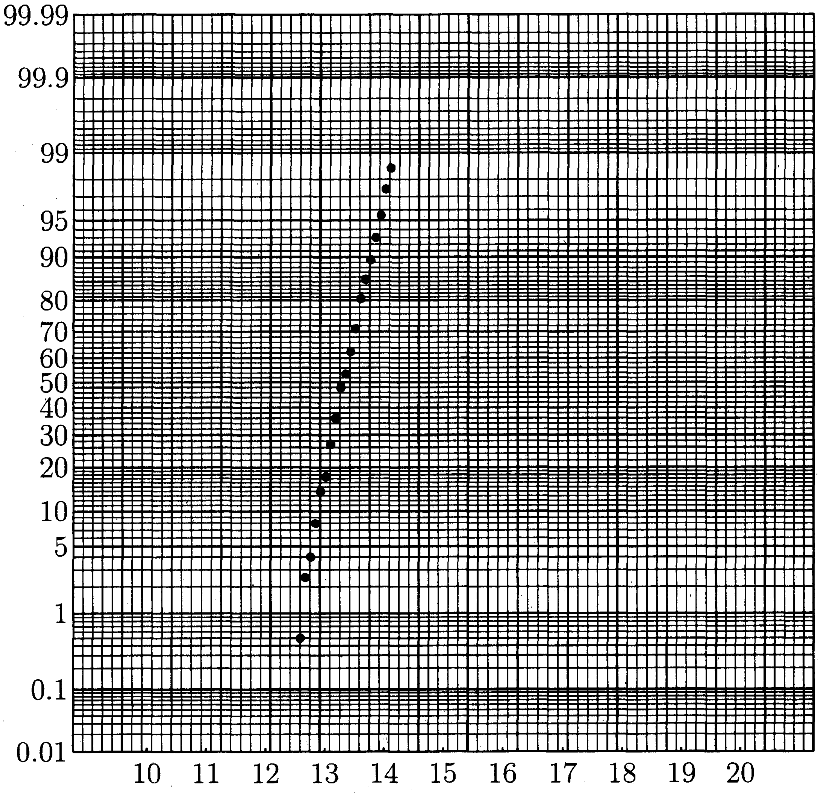


($\boldsymbol{n = 50}$ の場合)

階級の幅
0.08



$\mu = 13.3\cdots$



本教材作成と KETpic について，要点を述べる．

まず，乱数発生については，CAS に組み込まれている数値積分の関数により

$$F(x) = \int_{10}^{x} f(x)\,dx \quad (10 \leqq x \leqq 20)$$

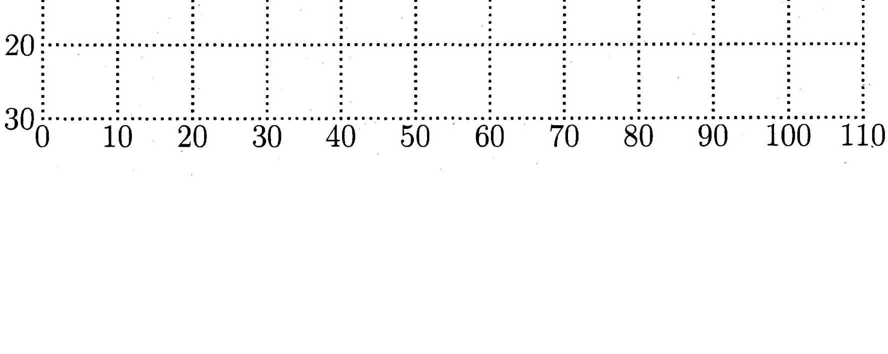
の数表を作り，2 分法のアルゴリズムによって $F(x) = y$ から x を求めた．

KETpic の作表コマンドは，このような正規確率紙を作るときにも応用できる．CAS のプログラミングを利用できることが大きな強みであり，TEX の tabular 環境では，ほとんど不可能である．また，ヒストグラムの作成にも繰り返し文が用いられている．

KETslide における図の配置は，layer 環境を用いて，例えば次のようにする．

```
\begin{layer}{110}{30}
\putnotese{20}{20}{\input{図ファイル}}
\end{layer}
```

これにより，複数のページにおける図の配置を正確に決定することができる．



layer の引数は，横縦グリッドの表示範囲である．グリッドを見ながら図を配置して，完成した後は，縦の引数を 0 にするとグリッドが消えて，図表などの要素だけが残る．

4 まとめ

KETpic は，当初は TEX 文書への図挿入のために開発されたが，その後，作表，メタコマンド，layer 環境，ketslide などが組み込まれて，総合的 TEX 支援ツールになった．

カレッジ級の統計教育における教材・教科書を作成するためには，これらの KETpic の機能が必要かつ有効である．また，R などで提供されている多くの統計関数や CAS のプログラミングをそのまま利用できることも大きな強みになっている．TEX とその周辺パッケージだけでは，統計の教材で必要とされる要素をすべて実現することは不可能であり，結局，他の統計ソフトを補助的に使うことになるが，KETpic は元々CAS のマクロパッケージであるため，CAS のプログラミングをすることが統計上必要な計算と図表の作成のいずれにもつながっている．すなわち，TEX と（R などの）1 つの CAS があれば，必要とされるすべての教材を作成できる点にも，他の教材作成ツールに対する KETpic の優位性があるといえる．

参考文献

[1] 高遠節夫ほか，新確率統計，大日本図書，2013

[2] 高遠節夫ほか，新確率統計問題集，大日本図書，2014.1（出版予定）

[3] KETpic ホームページ `http://ketpic.com`, `http://www55.atwiki.jp/ketpic/`

[4] Scilab ホームページ `http://www.scilab.org/`

[5] R ホームページ `http://www.r-project.org/`