【土木学会論文集　第357号／II-3(ノート)　1985年5月】

確率分布の適合度の図式判定法について

A NEW GRAPHICAL METHOD OF TESTING THE GOODNESS OF FIT OF DATA TO PROBABILITY DISTRIBUTIONS

上田年比古*・河村　明**
By Toshihiko UEDA and Akira KAWAMURA

In this paper a new graphical method of judging whether or not a particular distribution adequately describes a set of observation is proposed. This method needs only an arithmetic-scale paper and a visual judgment of goodness of fit. The test makes a comparison between the plots of the actual data and the straight line having a gradient of one through the origin on the arithmetic-scale paper. To show the effectiveness of this method, it is applied to random numbers and actual precipitation data. Results show that it is easier to judge visually the goodness of fit and to apply to any continuous or discrete distribution than with the usual probability-paper method. The properties behind this visual test are discussed in detail.

1. まえがき

自然現象における確率変数の分布は，特に理論的に定まったものはなく，その形状は千差万別といってよい．現在，この分布形を表わすのに，既知の確率分布式のうちから適当に選んで用いているようであるが，この場合，分布式の実際資料への適合度をみて，分布式が適切であるかどうかの判定が必要となる．これには χ^2 検定などの統計的仮説検定による数量的判定法と，確率紙による図式判定法とがある[1]．確率紙による判定法は，χ^2 検定のような数量的な正確さはないが，その確率分布に対する確率紙があれば，簡単であり視覚的にわかりやすく，また小標本に対しても有効で，よく用いられている．

本報は，これまでの確率紙とは別な図式判定法を提案し，これを，模擬発生の乱数資料と実際の降水資料に適用して，本法の有用性を検討したものである．

2. 確率分布の適合度の図式判定法

この方法は横軸，縦軸ともに普通目盛の方眼紙を用いる．いま，判定すべき連続型確率分布の確率変数を x，確率密度関数を $f(x)$，累積分布関数を $F_I(x)$ として，横軸には，次式に示す $x=x_i$ に対する非超過確率 $F_I(x_i)$ をとる．

$$F_I(x_i)=\int_{-\infty}^{x_i} f(x)dx \quad \cdots\cdots(1)$$

次に，縦軸には，資料の各値から得られる x_i の非超過確率 $F_D(x_i)$ をとる．ここに $F_D(x_i)$ は従来の確率紙の縦軸にも用いられている plotting position の諸公式のうちで，代表的な次の3つの式のいずれかを用いることにする．

$F_D(x_i)=n(x_i)/N$　　California plot ……(2)

$F_D(x_i)=|2\,n(x_i)-1|/2N$　　Hazen plot ……(3)

$F_D(x_i)=n(x_i)/(N+1)$　　Weibull(Thomas)plot ……(4)

$F_D(x_i)$ は，N 個の標本を，小さい方から順に $x_1 \leqq x_2 \leqq \cdots\cdots \leqq x_i \leqq \cdots\cdots \leqq x_N$ と並べた順序統計量の i 番目の値 x_i の非超過確率の推定値であり，式中の $n(x_i)$ は x_i 以下の標本値の個数すなわち $n(x_i)=i$ である．なお，Cunnane は各種の分布形に対して，種々の plotting 公式の精度やそれぞれに適合する不偏的な plotting 公式について概説し，次いでパラメーターを導入して，不偏的な plotting 公式の一般形を示している[2]．

* 正会員　工博　九州大学教授　工学部水工土木学科
(〒812　福岡市東区箱崎6-10-1)
** 学生会員　工修　九州大学大学院工学研究科水工土木学専攻博士後期課程 (同上)

さて，いま考えている $f(x)$ が適切であれば，

$$F_t(x_i) \fallingdotseq F_b(x_i) \cdots\cdots (5)$$

となり，したがって，資料の各値 x_i に対して，方眼紙にプロットされる点は原点を通り傾き1の直線に並ぶことになる．すなわちこの傾き1の直線（判定直線）への並び具合によって，いま考えている確率密度関数 $f(x)$ が適切であるかどうかが判定できる．

次に確率変数が離散型の場合を考える．この変数を s で表わし，s の各値を s_j（$j=1, 2, \cdots\cdots, m$）とし，その度数を $g(s_j)$ とすれば，標本の大きさ N は次式で表わされる．

$$N=\sum_{j=1}^{m} g(s_j) \cdots\cdots (6)$$

この場合も，前述の連続型確率変数の場合と同様にして次のように行う．離散型確率変数 s の確率関数[3]を $p(s)$ とすれば，各離散値 s_j の非超過確率 $F_t(s_j)$ は次式で表わされ，これを横軸にとる．

$$F_t(s_j)=\sum_{j=1}^{j} p(s_j) \cdots\cdots (7)$$

次に s_j を越えない標本数すなわち s_j 以下の標本個数を $n(s_j)$ とすれば，資料からの s_j の非超過確率 $F_b(s_j)$ は，式（2）～（4）で $n(x_i)$ の代わりに

$$n(s_j)=\sum_{j=1}^{j} g(s_j) \cdots\cdots (8)$$

を代入した次式となり，この値を縦軸にとる．

$$F_b(s_j)=n(s_j)/N \quad \text{California plot} \cdots\cdots (9)$$

$$F_b(s_j)=|2n(s_j)-1|/2N \quad \text{Hazen plot} \cdots\cdots (10)$$

$$F_b(s_j)=n(s_j)/(N+1) \quad \text{Weibull plot} \cdots\cdots (11)$$

こうして，標本値を方眼紙にプロットすれば，連続型確率変数の場合と同様に，原点を通り傾き1の直線にのるかどうかで，考えている離散型の確率関数が適切であるかどうかを判定できる．なおこの場合は方眼紙上の点は標本全部についてプロットされず，離散値 s_j についてのプロットのため，点の数は連続型確率変数の場合に比べるとかなり減少する．

3. 適 用 例

（1） 模擬発生の乱数資料に対する適用

ここでは連続型確率変数として指数乱数，離散型確率変数としてポアソン乱数を模擬発生させた．次にこの乱数資料に本法を適用して，はたして，指数分布およびポアソン分布が適合するという結果が得られるかどうかを検討してみよう．

指数分布の確率変数 x に対する確率密度関数 $f(x)$ および $x=x_i$ に対する非超過確率 $F_t(x_i)$ はそれぞれ次式で表わされる．

$$f(x)=\alpha e^{-\alpha x} \quad (x \geqq 0) \cdots\cdots (12)$$

$$F_t(x_i)=\int_0^{x_i} f(x)\,dx=1-e^{-\alpha x_i} \cdots\cdots (13)$$

いま母数 $\alpha=0.50$ の指数乱数を100個模擬発生させ（範囲は $x_i=11$ 未満であった），この資料に対し横軸に式（13）による $F_t(x_i)$，縦軸に式（3）による $F_b(x_i)$ をとり，プロットしたものが Fig. 1 の○印である．次に比較のため，同じ乱数資料に対して $x=0$ で $f(x)=0.5$，$x=4$ で $f(x)=0$ と直線状に確率密度が減少する三角形分布をあてはめた場合を Fig. 1 の×印に示している．

次に離散型のポアソン分布の確率関数 $p(s)$ と $s=s_j$ に対する非超過確率 $F_t(s_j)$ を以下に示す．

$$p(s)=\lambda^s e^{-\lambda}/s! \quad (s=0, 1, 2, \cdots\cdots) \cdots\cdots (14)$$

$$F_t(s_j)=\sum_{s=0}^{s_j} p(s)=\sum_{s=0}^{s_j} \frac{\lambda^s}{s!} e^{-\lambda} \cdots\cdots (15)$$

ここで母数 $\lambda=2.00$ としてポアソン乱数を100個模擬発生させた．この場合ポアソン分布の確率変数 s は0から7までが現われ8以上の度数は0となった．次にこの8個の s_j に対して，横軸に，式（15）による $F_t(s_j)$ をとり，縦軸に，式（10）による $F_b(s_j)$ をとり，各 s_j に対してプロットしたものが Fig. 2 である．

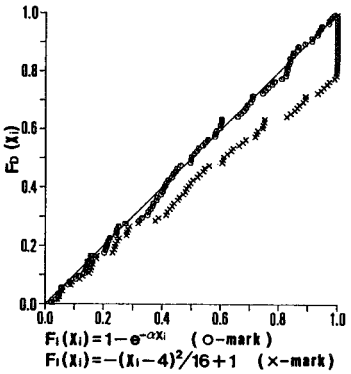

Fig. 1 The test of goodness of fit of exponential random numbers.

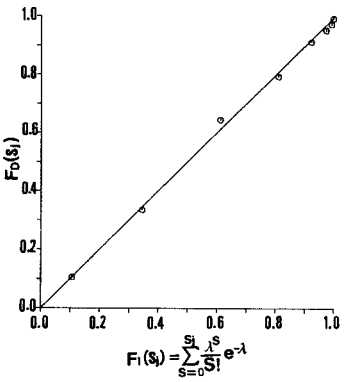

Fig. 2 The test of goodness of fit of Poisson random numbers.

（2） 降水資料に対する適用

福岡市の明治23年から昭和56年までの92年間の降水資料において，その1月上旬の旬平均降水量(mm/日)の分布に対して，連続型確率分布として指数分布とガンマ分布，離散型確率分布としてポアソン分布を用いて，その適合度を判定してみよう．なお，この資料の平均 $\bar{x}$ と不偏分散 V は次のようである．

$\bar{x}=2.19(\mathrm{mm/day})$，$V=2.70(\mathrm{mm/day})^2$ ……(16)

a) 指数確率分布の適用　式（13）による $F_I(x_i)$ を横軸に，式（3）による $F_D(x_i)$ を縦軸にとり，92個の資料をプロットしたものが Fig. 3 である．ここで母数 α は次式で定められる．

$$\alpha=1/\mu,\quad \alpha^2=1/\sigma^2 \cdots\cdots(17)$$

ここで μ，σ^2 は分布の母平均および母分散であって，これらの推定値として式（16）による資料の平均，不偏分散を用いる．このことは後述のガンマ分布，ポアソン分布でも同様である．

b) ガンマ分布の適用　この分布の $f(x)$ および $x=x_i$ に対する $F_I(x_i)$ をそれぞれ次式に示す．

$$f(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\quad (x\geqq 0)\cdots\cdots(18)$$

$$F_I(x_i)=\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^{x_i}x^{\alpha-1}e^{-\lambda x}ax\cdots\cdots(19)$$

ここに，$\Gamma(\alpha)$ はガンマ関数．α は形状母数，λ は尺度母数であり，これらは積率法より次式で計算される．

$$\alpha=\mu^2/\sigma^2,\quad \lambda=\mu/\sigma^2\cdots\cdots(20)$$

前述と同様にして，式（19）から $F_I(x_i)$，式（3）から $F_D(x_i)$ を求めて，それぞれ横軸，縦軸として92個の資料をプロットした結果を Fig. 4 に示している．

c) ポアソン分布の適用　この分布は離散型分布であり，離散化された資料が必要である．ここでは0.5(mm/日)を区分単位として離散化された降水量について適用することにする．式(15)による $F_I(s_j)$ と式(10)による $F_D(s_j)$ をそれぞれ横軸，縦軸として Fig. 5 を得る．なお母数 λ は次式によっている．

$$\lambda=\mu=\sigma^2\cdots\cdots(21)$$

（3） 適用結果の考察

指数分布およびポアソン分布の乱数資料に対する適用では，Fig. 1 の○印および Fig. 2 のように連続型確率変数，離散型確率変数ともに，プロットされた点は原点を通る傾き1の直線によくのっている．与えられた確率分布から発生した乱数資料では，この程度の適合度が期待されるといえる．次に，資料に対して，考えている分布式が相違する場合は，Fig. 1 の×印に示すように，原点を通る傾き1の直線から，はずれることがわかる．

次に実際の降水資料に対する適用結果をみる．指数分布では Fig. 3 のように各点は傾き1の直線にのっているとはみなされず，指数分布では表わされないと判断される．これを χ^2 検定によれば，有意水準1.5％で指数分布であるという仮説は棄却される．次にガンマ分布では，Fig. 4 のように各点はほぼ傾き1の直線上にのっていて，ガンマ分布が適合すると判断される．χ^2 検定によれば，χ^2 値の超過確率が36％以上となり，ガンマ分布であるという仮説を棄却できない結果が得られる．ま

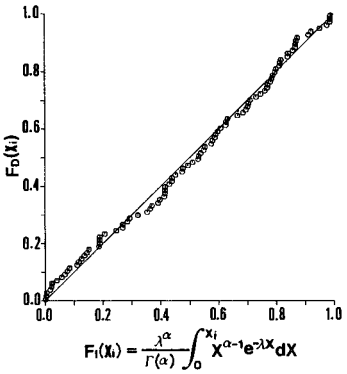

Fig. 4　The test of goodness of fit of actual precipitation data to gamma distribution.

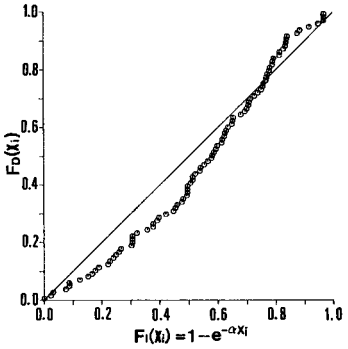

Fig. 3　The test of goodness of fit of actual precipitation data to exponential distribution.

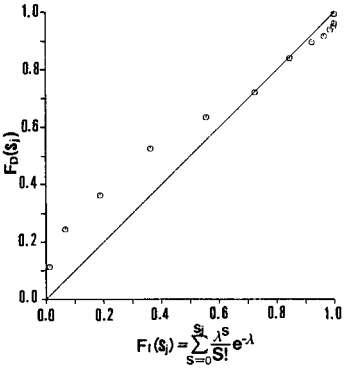

Fig. 5　The test of goodness of fit of actual precipitation data to Poisson distribution.

たポアソン分布では，Fig. 5 のように，各点は傾き 1 の直線にのっているとはみなされず，ポアソン分布では表わされないと判断されるが，χ^2 検定によれば有意水準 1 % でポアソン分布であることが棄却される．以上の各適用例からみて，本法は十分有用であるといえる．

4. 本法と確率紙による方法との比較

いま前述の模擬発生の 100 個の指数乱数（Fig. 1）を，式（3）により指数確率紙にプロットした結果を Fig. 6 に示している．資料の 96 番目以後の点では，式（2）と式（4）によるプロットも併記している．ここで式（2）の 100 番目の位置は 1 となりプロットできない．また図の直線は，指数分布に対する判定直線である．さて確率紙は，横軸に確率変数そのものがとられていて，各標本値のプロットが簡単に行われることは大きな利点である．ただし厳密な判定直線を引く場合には計算が必要でやや手間を要する．次に確率紙に標本値を，その非超過確率に従ってプロットすると，縦軸上の点の位置は連続型分布の場合確率密度の大きい部分では密に，小さい部分では粗に配列され，たとえば Fig. 6 では，縦軸の 0 に近い方が密に，1 に近づくにつれ粗に，また正規確率紙では，縦軸の 0.5 付近で密に，0 と 1 に近づくにつれ粗に配列される．Fig. 6 では $x_i=0\sim5$ の範囲で 90 点，$x_i=5\sim11$ の範囲では 10 点しかない．このような配列は，粗に配列された点の重みが視覚的に大きくなりがちで，各点に同じ重みをもたせた全体的立場からの適合度の判定には不利と考えられる．なお，縦軸上の 1 または 0 の近傍を重視する場合すなわち水文統計で極値の超過・非超過確率の評価が重要な場合には，この部分が拡大されている確率紙は有利と考えられる．この場合，Fig. 6 のように式（2）～（4）による点の位置が，縦軸方向にかなり相違し，また指数分布からの模擬発生の値

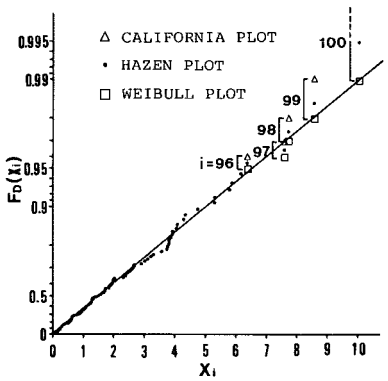

Fig. 6 The test of goodness of fit of exponential random numbers using exponential probability paper.

であるにもかかわらず，縦軸の 1 に近い部分でプロット方式によっては判定直線から，はずれていて，適合しているかどうかの判定がやや不明確となるが，これは極値の特性とプロット方式の問題であろう．

次に本法は，普通方眼紙を用いていること，判定直線が原点を通る傾き 1 の直線で，簡単に引くことができること，および本法では，横軸の位置を決めるのに式（1）の計算（積分できなければ数値積分）が必要で，点のプロットがやや面倒であるが，判定すべき確率分布式がどんな式であっても，また確率変数の連続型，離散型を問わず常に適用が可能であることは，その特長と考えられる．なおこのことは確率紙の作成が一般にはかなり面倒で，また確率分布によっては，それに含まれる母数が変わるごとに，確率紙が変わってくる場合もあり，このため確率紙が市販されていない場合（前述のガンマ分布，ポアソン分布など）の確率紙による方法の適用は一般には困難であることを考えると，1 つの利点と考えられる．また本法は縦軸が普通目盛であるため，各標本値の縦軸上の位置は連続型確率変数では等間隔に，また離散型確率変数では，確率増分に応じた間隔で並び，Fig. 1 の○印と Fig. 6 の各点の配置を比較してもわかるように，本法では各点に視覚的に同じ重みをもたせた適合度の判定ができると考えられる．また確率紙ではプロットできなかった，たとえば指数分布における非超過確率 1 の点（正規分布では 1 と 0 の点）もプロットできる．なお資料全体を等しい重みで判定しようとする思想は χ^2 検定の場合とほぼ同じと考えられる．次にこのようにして適合する分布式さえ決まれば，後は，たとえばある非超過確率に対する確率変数の値を求めたいときは式（1）から算定できるが，これは横軸の算定時にすでに得られている x_i と $F_i(x_i)$ との値から内挿して求められる．

5. むすび

以上のように本法は，確率紙と同様に，確率分布の適合度を図式的に判定するものであるが，普通方眼紙で行えること，確率分布のいかんにかかわらず適用可能であること，および標本の各点を等しい重みで適合度の判定が行えることなどに特徴をもつものといえる．

参考文献

1) たとえば鈴木栄一：気象統計学，地人書館，1981 年．
2) Cunnane, C.: Unbiased plotting positions——A review, Journal of Hydrology, Vol. 37, pp. 205～222, 1978.
3) I. ガットマン・S. S. ウィルス（石井恵一・堀 素夫共訳）：工科系のための統計概論，培風館，p. 20, 1968 年．

（1984. 6. 21・受付）