「正規母集団であることの検定について」

山田　彌

一

統計的推論において、標本X_1、X_2、……、X_nが正規分布に従うと仮定して議論をすすめることは、普通におこなわれていることである。

一般に、線型回帰モデル

$$y=X\beta+u \tag{1}$$

における誤差項は、

$$E(uu')=\sigma^2 I_n$$

という条件があれば、その分布型を特定化しなくても、最小二乗推定量は最良線型不偏推定量になる。しかし、最小二乗推定量が最尤推定量であり且つ一様最小分散不偏推定量となり、また、種々の有意性の検定を行なうためには、誤差項の確率分布が正規分布である、と仮定されなければならない。[0]

現実の誤差項の確率分布を、正規分布で近似しようという考え方は、従って、何よりも実用的な観点から生じたであろうことは明らかである。しかし、実際において、標本X_iなり、誤差項uなりが確かに正規分布をすることが知られている、あるいは正規分布をすることが証明されている場合は必ずしも多くはない。それにもかかわらず、観測値が正規分布をすると仮定して議論をすすめることができる積極的な理由は何であろうか。

観測値なるものが概して様々な多くの要因の複合体であると考えることができるならば、中心極限定理は正規分布の仮定を正当づける有力なよりどころとなるであろう。[1] あるいは又、自然科学における数多くの観測の蓄積が、それらの頻度

分布が多くの場合、正規分布によって近似しうることを経験的に示しているということもあろう。しかし、だからといって、個々の観測値が正規分布に従うことが、これらの事情によって証明された訳ではない。

従って、観測値が正規分布に従っているか否かが疑わしい場合には、統計的にしらべてみる必要があろう。直観的には、正規確率紙に点を打つことによって、推測統計的には、仮説検定によって確かめることになる。

正規分布に従うという仮説が棄却されたとすれば、その場合になお、正規分布を仮定して得られた手法を適用した場合に、その妥当性がどの程度保たれるのか、誤りはどの程度であるかが具体的に明らかにされる必要がある。この、手法の「頑健性」が十分なものでない場合には、その分布型の推定と、それにみあったより望ましい手法の開発が必要となる訳である。

ここでは、観測値が正規分布に従うという仮説の検定について考える。

二

x_1、x_2、……、x_nが同一の母集団から独立に抽出された標本である場合に、母集団分布が正規分布であるかどうかを検定するのに、普通考えられるのは適合度のχ^2検定である。適合度の検定とは、k個の排反事象E_iの生起確率がそれぞれP_iである、という帰無仮説H_0を、それぞれの事象がn_i回起ったという標本によって検定するものである。仮説H_0が正しければ、統計量

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-nP_i)^2}{nP_i} \qquad \text{但し} \sum n_i=n \qquad (2)$$

は近似的に自由度$k-1$のχ^2分布に従う。(2)

これを用いれば、x_1……x_nが正規母集団から抽出された標本であるという仮説H_0が正しいとすれば、統計量

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-n\hat{P}_i)^2}{n\hat{P}_i} \qquad (3)$$

は、自由度$k-3$のχ^2分布に近似的に従う訳であるから、(3) χ^2分布の右側検定によって、仮説H_0を検定することができる。ここで、f_iは標本x_iを小さいものから順にk組の級に分けた場合

に、第i階級に属する標本の個数であり、又$\hat{P}_i$は、母集団分布の母数が未知のために、それぞれ標本平均$\bar{x}$と標本分散S^2で代用して母集団が$N(\bar{x}, S^2)$であると考えた場合の、各階級の確率である。即ち、

$$
\left.
\begin{aligned}
&\sum^{k} f_i = n \\
&\bar{x}=\frac{1}{n}\sum^{n} x_i,\quad S^2=\frac{1}{n}\sum^{n}(x_i-\bar{x}^2) \\
&\hat{P}_1 = Prob\left(-\infty < y \leqq y_1+\frac{h}{2}\right) \\
&\hat{P}_i = Prob\left(y_i-\frac{h}{2} < y \leqq y_i+\frac{h}{2}\right) \\
&\qquad\qquad (i=2, \ldots\ldots, k-1), \\
&\hat{P}_p = Prob\left(y_k-\frac{h}{2} \leqq y < +\infty\right)
\end{aligned}
\right\} \quad (4)
$$

但しy_iは階級値、hは級幅である。

χ^2の算式から明らかな様に、適合度検定(4)は実現度数と期待度数の差の二乗を期待度数で計ったものの総和の大きさで、あてはまりのよさを計ろうとするものである。実現度数と期待度数が近ければ近いほど、χ^2の値は小さくなる。従って、このようなあてはまりの良さによる母集団分布型の検定は、母集団分布が正規分布でない場合であっても、それが左右対称の山形（例えばコーシー分布）をしておれば、仮説H_0に対して有意な判定を下すことが比較的少数の標本の場合にはかなり困難になることは容易に想像できる。あてはまりの良さによる検定は、他に問題がないとしても、この点であまり望ましいものではないと思われる。

加えて、この検定方式にはいくつかの制約条件がある。一つは、標本数nが小さいと、個々の標本の影響が大きくなりすぎて適用できない。nが50前後では、とても良い検定ができない。

第二はf_iが小さいと駄目なことである。これは、χ^2の導出においてスターリングの公式で近似したことによる。普通はf_iが5以上でないと単独では扱わない。

こうして、分布の型のあてはめによる検定があまり望ましくないとすれば、正規分布だけがもっている属性を利用した判定法を考えねばならない。

三

では、正規分布にのみ特徴的な属性として何が考えうるであろうか。その様な属性の一つに、次のものがある。

定理一

正規母集団 $N(\mu,\ \sigma^2)$ からの、大きさ n の標本を x_1、……、x_n、その標本平均を $\bar{x}$、標本分散を S^2 とすれば確率変数 $\bar{X}$ と S^2 とは互に独立である。[5]

逆に、

定理二

同一の分布 $F(x)$ に従う独立な確率変数 X_1……X_n において、確率変数 $\bar{X}$ と S^2 が独立であれば、$F(x)$ は正規分布函数である。

但し、$\overline{X}=\frac{1}{n}\sum^{n}X_i,\ \ S^2=\frac{1}{n}\sum^{n}(X_i-\overline{X})^2$ [6]

従って、母集団の正規性と、標本平均と標本分散の独立性とは同値であることがわかる。

だから、母集団の正規性の検定は、標本平均と標本分散の独立性の検定によって代替される訳である。

四

標本の一つ一つについて標本平均と標本分散を考えることは意味がないから、数個の標本のひとかたまりを、新たに、一個の標本と考えよう。即ち n 個の標本とは、

$$x_j=(x_{1j},\ x_{2j},\ \cdots\cdots,\ x_{kj})\ j=1,\cdots\cdots,\ n \qquad (5)$$

$$\sum_{j=1}^{n}k_j=N$$

である。k_j は j について異った数であっても同じであっても良いであろう。

そこで、各標本毎の標本平均と標本分散

$$\left.\begin{aligned}\bar{x}_j&=\frac{1}{k_j}\sum_{j=1}^{k_j}x_{ij}\\ s_j{}^2&=\frac{1}{k_j}\sum_{1}^{k_j}(x_{ij}-\bar{x}_j)^2\end{aligned}\right\} \qquad (6)$$

を標本のもつ二つの属性であると考えると、この二つの属性の独立性を分割表による検定によってしらべることができる。

周知のように、ピアソンによって与えられた、$r\times l$ 分割表における行と列の独立性の仮説を検定するための検定規準は次の様に定義される。

$$\chi^2=\sum_i\sum_j\frac{\left(n_{ij}-\frac{n_i.n._j}{n}\right)^2}{\frac{n_i.n._j}{n}} \qquad (7)$$

即ち、二つの属性は独立であるという仮説 H_0 が正しいとき、

第1表

$\bar{x}$ ＼ S^2	$\sim S^2{}_1$	……	$S^2{}_l\sim$	計
$\sim \bar{x}_1$	n_{11}	……	n_{1l}	$n_{1\cdot}$
…… …… ……	……	……	……	……
$\bar{x}_r\sim$	n_{r1}	……	n_{rl}	$n_{r\cdot}$
計	$n_{\cdot 1}$	……	$n_{\cdot l}$	n

χ^2は漸近的に自由度$(r-1)(l-1)$のχ^2分布に従うことが知られている。[7]

尚、$\sum_j^r n_{ij}=n_{\cdot j}\,(j=1,\cdots,l)$

$\sum_j^l n_{ij}=n_{i\cdot}\,(i=1,\cdots,r)$

この、χ^2分布による独立性の検定の場合には、前と同様に標本数nや組度数n_{ij}が一定程度（n、n_{ij}はそれぞれ、およそ100、5以上）なければ検出力の高い検定にはならないが、nやn_{ij}が小なるときに、H_0が正しいときにその様な分割表の標本が抽出される確率を直接算出する方法をフィッシャーは示している。[8]

それによれば、2×2の分割表において、両属性が独立であり、且つ周辺度数$n_{1\cdot}$、$n_{2\cdot}$、$n_{\cdot 1}$、$n_{\cdot 2}$である場合の、n_{11}、n_{12}、n_{21}、n_{22}の条件付同時確率分布は、

$$\left.\begin{aligned}&P(n_{11},\ n_{12},\ n_{21},\ n_{22}|n_{1\cdot},\ n_{2\cdot},\ n_{\cdot 1},\ n_{\cdot 2})\\&\quad\equiv P(n_{ij}|n_{i\cdot},\ n_{\cdot j})\\&\quad=\frac{n_{1\cdot}!\,n_{2\cdot}!\,n_{\cdot 1}!\,n_{\cdot 2}!}{n!\,n_{11}!\,n_{12}!\,n_{21}!\,n_{22}!}\end{aligned}\right\}\quad(8)$$

となる。フィッシャーが与えたのは2×2分割表についてであるが、一般に、$r\times l$分割表では、

$$P(n_{ij}|n_{i\cdot},\ n_{\cdot j})=\frac{(\prod n_{i\cdot}!)(\prod n_{\cdot j}!)}{n!\prod n_{ij}!}\qquad(9)$$

で示される。[9]

従って、実現度数n_{ij}によって、仮説H_0が棄却されるのは、n_{ij}及び、n_{ij}よりも偏った事象の集合をAで示せば、

$$P=\int_A p(n_{ij}|n_{i\cdot}n_{\cdot j})dnij\qquad(10)$$

の値が有意水準よりも小さくなった場合であるということになる。

五

標本平均と標本分散の組（$\bar{x}_j$, $s_j{}^2$）を、二次元の結合分布$F(\bar{X},\ S^2)$をもつ母集団から得た標本であると考えれば、$\bar{X}$とS^2の独立性は$\bar{X}$とS^2の周辺分布をそれぞれ$G(\bar{X})$, $H(S^2)$とすれば、

$$F(\bar{X},S^2)=G(\bar{X})\cdot H(S^2)$$

で示されるのであるから、独立性の仮説H_0は次の様にあらわせる。

即ちΩを二次元ユークリッド空間R^2における連続な分布函数の集合、ωを、それらが周辺分布函数の積で表わされるようなΩの要素からなるΩの部分集合とすれば、

$$H_0 : F \in \omega$$

$$H_1 : F \in \Omega - \omega$$

が検定すべき仮説である。

今、S_nを$\overline{X}$、S^2の標本分布函数であるとすれば、S_nは次の様にあらわすことができる。

$$S_n(r_1,\ r_2) = \frac{1}{n}\sum_{j=1}^{n} \phi_{r1}(\overline{X}_j)\phi_{r2}(S_j{}^2) \tag{11}$$

$$\phi_{ri}(x) = \begin{cases} 1 & if \quad x \leqq r \\ 0 & if \quad x > r \end{cases}$$

つまり、S_nは

$$(\overline{X}_j,\ S_j{}^2) \leqq (r_1,\ r_2)$$

なる組の数の$1/n$倍である。

S_{nj}によってS_nの第j番目の要素に関する周辺分布函数を表わせば

$$[T_n(r)]^2 = [S_n(r) - \prod_{j=1}^{2} S_{nj}(r_j)]^2 \tag{12}$$

は、明らかに、H_0が正しければ、それだけ小さな値を示す。従ってこれを区分rのすべてについて積分した統計量で、H_0

の検定を行なうことができるであろう。

$$B_n = \int [T_n(r)^2] dS_n(r) \tag{13}$$

とおいたときのnB_nの漸近分布を、J. R. Blum, J. Kiefer, M. Rosenblatt が与えている。[10] nB_nは、H_0が正しいとき小さく、正しくないとき大きくなるから、右側検定である。

この検定法は、$T_n(r)$が、分割表の検定の場合の$(n_{ij} - n_{i\cdot}\, n_{\cdot j}/n)$と結局同じものを意味しているのであるから、分割表の検定と本質的にかわるものではない。nB_nの分布表が一般には簡単に手に入らない点は、この方法の大きな欠陥であろう。

六

さて、実際に非正規分布に従う乱数をとって、以上の諸検定を実際に行ってみる。正規分布に似た形の分布はある種のポアッソン分布やベータ分布などいくつか考えうるが、乱数を発生させる手続の比較的簡単な、コーシー乱数をここでは使用する。ここで想定するコーシー分布は、密度函数が、

$$f(x) = \frac{1}{\pi} \cdot \frac{3}{3^2 + x^2} \qquad -\infty < x < \infty \tag{14}$$

で与えられるものとする。

コーシー乱数xは、一様乱数yから次の様にして得ることができる。

$$y=\int_{-\infty}^{\infty}\frac{1}{\pi}\frac{3}{3^2+t^2}dt=\frac{1}{\pi}\tan^{-1}\frac{t}{3}\Big]_{-\infty}^{x}=\frac{1}{2}-\frac{1}{\pi}\tan^{-1}\frac{x}{3} \qquad (15)$$

第2表

区　　間	f_i	$\hat{p}_i$	$n\hat{p}_i$	$f_i-n\hat{p}_i$	$(f_i-n\hat{p}_i)^2$	$(f_i-n\hat{p}_i)^2/n\hat{p}_i$
$x\geqq 6$	21	0.18	22.5	−1.5	2.25	0.1
$6>x\geqq 3$	14	0.17	21.2	−7.2	51.84	2.44
$3>x\geqq 1$	22	0.14	17.5	4.5	20.25	1.15
$1>x\geqq -1$	19	0.14	17.5	1.5	2.25	0.12
$-1>x\geqq -3$	23	0.13	16.2	6.8	46.24	2.85
$-3>x\geqq -6$	14	0.12	15.0	−1	1	0.06
$-6>x$	12	0.11	13.7	−1.7	2.89	0.21
合　計	125	1.0				6.73

より、

$$x=3\tan\left(y-\frac{1}{2}\right)\pi \quad 0\leqq y\leqq 1 \qquad (16)$$

である。こうして得られた125個の標本[(11)]によって、適合度のχ^2検定を行なうと次の通りである。

$\hat{p}_i$は、この標本の標本平均と標本分散（それぞれ、0.88と34.1）をそれぞれ平均、分散とする正規分布の、各区間における理論値である。この計算結果によれば、

$$\sum_j\frac{(f_i-n\hat{p}_i)^2}{n\hat{p}_i}=6.73$$

となるが、自由度4のχ^2分布の10%水準$\chi^2_{0.1}$、5%水準$\chi^2_{0.05}$、1%水準$\chi^2_{0.01}$はそれぞれ

$$\chi^2_{0.1}=7.779$$
$$\chi^2_{0.05}=9.488$$
$$\chi^2_{0.01}=13.277$$

であるから、この検定によれば10%の有意水準でも、この標本が正規分布に従うという仮説H_0は棄却されないことになる。この場合、標本数が百余個でしかなかったことが大きく影響を与えていると思われる。

さて次に、分割表による独立性の検定を試みることにする。

前と同じ標本を、この場合は、五個づつに区分して、合計二五個の標本、即ち、それぞれの標本平均と標本分散の組をつくる。[12] これらの組を、標本平均については0.88を境に二分し、標本分散については20を境に二分して2×2の分割表をつくると次の様になる。この場合、データの数が非常に少ないので、χ^2分布による検定は勿論つかえない。そこで、フィッシャーの直接確率計算法によって、$\bar{X}$とS^2が独立であるとするH_0が正しいとき、$\bar{X}$、S^2の周辺度数が第三表のように固定されているときに、第三表の実現度数も含めて、それよりも偏った組合せが生ずる確率を求める。第三表よりも偏った場合というのは、次の二つの場合であるから、

第3表

$\bar{x}$ ＼ S^2	$S^2 \leqq 20$	$S^2 > 20$	計
$\bar{x} \geqq 0088$	10	2	12
$0.88 > \bar{x}$	4	9	13
計	14	11	25

10	1
3	11

12	0
2	11

結局求める確率は、

$$P=\frac{14!13!12!11!}{25!}\left\{\frac{1}{10!9!4!2!}+\frac{1}{11!10!3!}+\frac{1}{12!11!2!}\right\}$$

$$\fallingdotseq 0.009$$

従って、有意水準0.01でも、$\bar{X}$とS^2が独立である、従って元の標本が正規母集団からのものであるという仮説は棄却されることになる。こうしてみるならば、同じ標本を同じ個数使用して試した結果は、この標本について云う限りでは、二つの方法の優劣をかなり明瞭に示している様に思われる。

しかしながら、標本平均と標本分散の独立性を用いて検定する方法は、標本平均や標本分散を計算しなければならないという手数を考慮に入れないとしても、標本数そのものが数分の1に減少（今の例では5分の1に）するという、本質的な特徴からくる欠陥をもっている様に思われる。今の例ではかなりうまくH_0の誤りを検出した訳であるが、標本数25個という小標本による検定は、やはり不安定な要素を残しているといわざるを得ない。

又、第一次標本を、どう組合せて標本の組をつくるかによって、検定の結果が大きく変化する可能性も否定できない。従って、二つの方法のいずれが優った検定法であるかは、上の例の結果から直ちに判断は出来ない。

しかし、この様な事情を考慮した上でなお、上の例は、分

割表による検定法の一定の有効性を示しているものと思われる。

さて、最後の検定法を同じデータについて実際にやってみると次の如し。r_iに関して次の様にとる。

$$r_1^{1}=1 \quad r_1^{2}=3$$

$$r_2^{1}=10 \quad r_2^{2}=20$$

このr_i^{j}についてnB_nを求めると、(13)

$$nB_n=25\times B_{25}=0.55$$

となるが、J.R. Blum らの数表によれば有意水準5%及び1%のnB_nの値はそれぞれ

$$nB_{n\cdot 0.05}\doteqdot 0.06$$

$$nB_{n\cdot 0.01}\doteqdot 0.11$$

であるから、この場合、H_0は有意水準1%でも棄却される訳である。ただし、この検定法も、標本数がある程度大きくないと信頼度はかなり落ちる様である。

六

以上、標本母集団の正規性の検定について、標本平均と標本分散の独立性の検定によって行なうことができることを提起し、従来の適合度検定と対比して若干の数値実験を試みた。行なった結果は、独立性の検定を用いる二つの方法とも、ほぼ満足できることがわかった。尤も、この「実験」で、直ちに、独立性の検定による方法の優位性が証明された訳ではもとより全くあり得ない。より多面的な検討が必要である。

（0）例えば W. Feller, An Introduetion to Probability theory and its Application, John Wiley, 1957. 238頁。

（1）例えば J. Johnston, Econometric Methods, McGraw-Hill. 1963. 第1章参照。

（2）仮説H_0が正しいとき、この様な結果が生じる確率は多項分布によって、

$$P=n!\,p_1^{n_1}p_2^{n_2}\cdots\cdots p_k^{n_k}/n_1!\,n_2!\cdots\cdots n_k!$$

n_iが大きいときは、スターリングの公式により

$$P\sim e^{-\frac{\chi^2}{2}}\Big/(2\pi n)^{\frac{k-2}{2}}(p_1p_2\cdots\cdots p_k)^{\frac{1}{2}}$$

但し　$\chi^2=\sum(n_i-np_i)^2/np_i$

（3）自由度が3だけ減るのは、この場合に三つの拘束式があるからである。即ち、

$$\sum f_i=n$$

$$n\bar{x}=\sum f_ix_i$$

$$nS^2=\sum f_i(x_i-\bar{x})^2$$

（4）累積分布を比較する適合度検定の方法として、ω^2検定法が

「正規母集団であることの検定について」（山田）

ある。

$$\omega^2=\int_{-\infty}^{\infty}\{F_n(x)-F(x)\}^2dx$$

ここで $F_n(x)$ は標本累積分布函数、$F(x)$ はあてはめようとする理論分布であるが、ω^2 の分布はわからないから、

$$E(\omega^2)=\frac{1}{\mathrm{n}}\int_{-\infty}^{\infty}F(x)\{1-F(x)\}dx$$

との比較で検定を行う。

S. S. Wilks, Mathematical Statistics, 1944. 小河原訳、第六章参照。

（5） 例えば S. S. Wilks 前掲書 一六二頁参照。

（6） T. Kawata and H. Sakamoto " On the Characterization of the normal popluation by the independence of the sample mean and sample variance, Jour. Math. Soc. Japan, vol. 1.

E. Lukacs " A characterization of the normal distribution, Ann. Math, Stat., vol. 13.

ルーカスによるこの定理の証明の要点は次の通りである。

$f(x)$ を母集団の密度函数とすれば特性函数は

$$\psi(t)=\int e^{itx}f(x)dx$$

一方 $\bar{x}$ と S^2 の同時分布の特性函数は、

$$\varphi(t_1,t_2)=\int\cdots\cdots\int e^{it_1\bar{x}+it_2S^2}f(x_1)\cdots\cdots f(x_n)dx_1\cdots\cdots dx_n$$

同様に $\bar{x}$ と S^2 それぞれの特性函数は、

$$\varphi_1(t_1)=\varphi(t_1,0)=\int\cdots\cdots\int e^{it_1\bar{x}}f(x_1)\cdots\cdots f(x_n)dx_1\cdots\cdots dx_n$$

$$\varphi_2(t_2)=\varphi(0,t_2)=\int\cdots\cdots\int e^{it_2S^2}f(x_1)\cdots\cdots f(x_n)dx_1\cdots\cdots dx_n$$

$\bar{x}$ と S^2 が独立とは

$$\varphi(t_1,t_2)=\varphi_1(t_1)\phi_2(t_2)$$

即ち

$$\partial\varphi(t_1,t_2)/\partial t_2\Big|_{t_2=0}=\varphi_1(t_1)\frac{\partial\varphi(t_2)}{\partial t_2}\Big|_{t_2=0}$$

又、

$$\varphi_1(t_1)=\prod_{\alpha=1}^{n}\int e^{it_1x_\alpha/n}f(x_\alpha)dx_\alpha=\Big[\int e^{it_1x/n}f(x)dx\Big]^n$$

$$=[\psi(t_1/n)]^n$$

だから

$$\frac{\partial\varphi}{\partial t_2}\Big|_{t_2=0}=[\psi(t_1/n)]^n\frac{\partial\varphi_2}{\partial t_2}\Big|_{t_2=0}$$

又、

$$\frac{\partial\varphi(t_1t_2)}{\partial t_2}=i\int\cdots\cdots\int S^2e^{it_1\bar{x}+it_2S^2}f(x_1)\cdots f(x_n)dx_1\cdots dx_n$$

に

$$\bar{x}=\frac{1}{n}\sum x_\alpha,\quad S^2=[(n-1)\sum x_\alpha{}^2-2\sum\sum x_\alpha x_{\beta+1}]/n^2$$

に代入すれば、

$$\frac{\partial\varphi(t_1,t_2)}{\partial t_2}\Big|_{t_2=0}=\frac{i(n-1)}{n}\Big\{[\psi(t_1/n)]^{n-1}\int x^2e^{it_1x/n}f(x)dx$$

$$-[\psi(t_1/n)]^{n-2}\Big[\int xe^{it_1x/n}f(x)dx\Big]^2\Big\}$$

同様に

$$\left.\frac{\partial\varphi_2(t_2)}{\partial t_2}\right|_{t_2=0}=\frac{i(n-1)}{n}\sigma^2$$ （但し σ^2 は母集団の分散）

となり $t=t_1/n$ とすれば結局

$$\psi(t)\int x^2e^{itx}\cdot f(x)dx-\left[\int xe^{itx}\cdot f(x)dx\right]^2=[\psi(t)]^2\sigma^2$$

特性函数については

$$\frac{d^l\psi(t)}{dt^l}=i^l\int x^le^{itx}f(x)dx$$

だから、これを代入すると結局 $\psi(t)$ に関する微分方程式

$$-\psi(t)\psi''+(\psi')^2=\sigma^2\psi^2$$

初期条件 $\psi(0)=1,\ \psi'(0)=i\mu$ で解けば

$$\psi(t)=e^{i\mu t-\frac{1}{2}\sigma^2t^2}$$

となり、これは正規分布の特性函数となる。

（7） 例えば 竹内 啓「数理統計学」東洋経済 第21章参照。

（8） R. A. Fisher, Statistical Methods for Research Workers, Oliver and Boyd 1963. 遠藤・鍋谷訳、第四章。

（9） $n_{ij}(i=1\cdots\cdots r,\ j=1\cdots\cdots,\ l)$ の同時分布は、独立の仮説が正しければ、

$$p(n_{11},\cdots\cdots,n_{rl})=\frac{n!}{\prod_{ij}n_{ij}!}(\prod_i p_i^{n_{i\cdot}})(\prod_j q_j^{n_{\cdot j}})$$

$p_i,\ q_j$ はそれぞれひとつの標本が属性 $A_i,\ B_j$ をもつ確率である。

$n_{i\cdot}$、$n_{\cdot j}$ の同時確率分布 $g(n_{i\cdot},\ n_{\cdot j})$ は、$n_{i\cdot}$、$n_{\cdot j}$ を固定して

$\sum_i n_{ij}=n_{\cdot j},\ \sum_j n_{ij}=n_{i\cdot}$

のすべての n_{ij} に対する $p(n_{11},\cdots\cdots n_{rl})$ の値を計算しなければならないが、

$$\prod_j^l(\sum_i^r x_i)^{n_{\cdot j}}=(\sum_i^r x_i)^n$$

の係数比較によって $\sum_i n_{ij}=n_{\cdot j}$ $\sum_j n_{ij}=n_{i\cdot}$ なるすべての n_{ij} についての和

$$\sum\prod_j^l\frac{n_{\cdot j}!}{\prod_i n_{ij}!}=\frac{n!}{\prod_i n_{i\cdot}!}$$

を得る。これより

$$\sum_{i,j}\frac{1}{\prod n_{ij}!}=\frac{n!}{(\prod_i n_{i\cdot}!)(\prod_j n_{\cdot j}!)}$$

だから

$$g(n_{i\cdot},n_{\cdot j})=\frac{(n!)^2(\prod_i p_i^{n_{i\cdot}})(\prod_j q_j^{n_{\cdot j}})}{(\prod_i n_{i\cdot}!)(\prod_j n_{\cdot j}!)}$$

従って

$$p(n_{ij}|n_{i\cdot},\ n_{\cdot j})=p(n_{ij})/g(n_{i\cdot},\ n_{\cdot j})=\frac{(\prod_i n_{i\cdot}!)(\prod_j n_{\cdot j}!)}{n!\prod n_{ij}!}$$

（10） J.R. Blum, J. Kiefer, and M. Rosenblatt, Distribution free tests of independence based on the sample distribution function, Ann. Math. Stat. vol. 32.

（11） 使用したコーシー乱数は次の通りである。

0.57	8.34	−10.32	− 1.41	− 2.19
1.53	0.18	− 0.57	4.41	− 8.34
7.59	0.0	1.77	7.59	− 7.59
1.02	0.39	3.63	− 4.74	− 1.08
31.74	− 3.39	2.82	− 2.34	−15.72
− 5.07	1.89	− 0.39	− 2.64	11.67
−18.93	5.88	5.07	− 5.04	6.39
− 2.19	− 0.15	− 8.34	− 9.24	1.89
− 0.18	3.00	2.49	0	2.49
13.41	1.53	− 2.34	− 2.43	− 2.64
− 3.39	2.64	11.67	− 2.04	− 1.41
− 1.77	− 1.2	− 0.48	− 2.49	− 3.18
1.65	− 0.27	1.2	− 1.29	2.19
− 4.41	4.41	− 1.41	− 1.20	9.24
3.39	− 3.87	− 3.0	9.24	− 0.48
6.39	− 2.19	6.39	9.24	− 6.39
6.69	5.07	2.34	2.19	− 3.18
− 3.18	− 0.18	0.39	1.53	0.96
3.18	1.53	2.04	−11.28	3.87
0.87	6.39	− 1.53	− 2.64	6.84
− 4.41	0.96	0.96	1.65	−13.41
1.41	3.87	3.87	5.46	− 3.63
− 0.27	− 2.64	−11.67	10.32	9.24
6.84	− 3.84	− 5.01	1.89	−11.67
6.39	− 2.04	1.08	9.24	3.87

(12) 25組の $\bar{x}_i$ と S_i^2 は次の通り。

$\bar{x}_i$	S_i^2	$\bar{x}_i$	S_i^2
8.37	14.37	1.92	6.93
−2.58	10.71	−2.16	32.05
−0.99	86.76	0.69	20.88
2.79	13.51	3.87	9.72
1.98	17.91	−0.66	19.35
11.1	15.03	0.18	45.27
2.43	3.96	−5.7	13.14
1.89	4.68	6.99	28.80
2.13	10.17	−3.96	23.04
−0.81	8.19	1.26	18.9
−0.54	25.92	0.42	22.59
−0.69	21.57	−3.12	76.14
1.59	27.0		

(13) 左の表から、

γ_2^1　　γ_2^3

		$0 \leqq S^2 < 10$	$10 \leqq S^2 < 20$	$20 \leqq S^2$	計
	$\bar{x} \geqq 3$	1	2	1	4
γ_1^2	$3 \geqq \bar{x} > 1$	3	4	1	8
γ_1^1	$\bar{x} < 1$	1	3	9	13
	計	5	9	11	25

$$B_n = \left\{\frac{1}{25} - \frac{13\times 5}{25^2}\right\}^2 + \left\{\frac{4}{25} - \frac{5\times 21}{25^2}\right\}^2 + \left\{\frac{4}{25} - \frac{13\times 14}{25^2}\right\}^2 + \left\{\frac{11}{25} - \frac{14\times 21}{25^2}\right\}^2$$