水準移動仮説に対する ノンパラメトリック検定

統計数理研究所　鈴　木　義一郎

(1977年9月　受付)

Nonparametric Test Procedures for a Change in a Location Parameter

Giitiro Suzuki
(The Institute of Statistical Mathematics)

Given independent sequence $X_1, X_2, \cdots, X_n$, each X_i being distributed as $F(x-\theta_i)$, where $\int t\,dF(t)=0$, $\int t^2\,dF(t)=\sigma^2$.

Consider the following testing situation:

$$\begin{cases} \text{Null Hypothesis } H_0: \theta_1=\theta_2=\cdots=\theta_n=\theta \\ \text{Alternative} \quad H_m: \theta_1=\theta_2=\cdots=\theta_m=\theta \\ \qquad\qquad \theta_{m+1}=\cdots=\theta_n=\theta' \\ \qquad\qquad (\theta \neq \theta') \end{cases}$$

The assumption that m is unknown for us is essential for the problem.

As test statistics, we consider the sum R_n, defined by (1), of record values (Rényi[5] called outstanding), the dispersion Q_n, defined by (2), of record values, the total number T_n, defined by (11), of inversions (defined by Feller [2]), and the linear statistic L_n^*, defined by (16), (18), which coincides with (22) proposed by Chernoff-Zacks [2]. Their mean values and variances under H_0 and H_m are derived as (3), (4), (5), (6), (7), (8), (9), (10), (12), (13), (14), (15), (19), (20), (21) and their distributions under H_0 are also considered. Historical developments are presented to some other related statistics.

1.　問題の設定と主要な結果

独立な観測値の系列 $X_1, X_2, \cdots, X_n$ が与えられていて，各 X_i は確率分布 $F(x-\theta_i)$ に従うものとする．但し

$$\int t\,dF(t)=0,\ \int t^2\,dF(t)=\sigma^2$$

とする．つまり，θ_i は X_i の平均を表わすパラメータである．そこで

$$\begin{cases} \text{帰無仮説} \quad H_0: \theta_1=\theta_2=\cdots=\theta_n=\theta \\ \text{対立仮説} \quad H_m: \theta_1=\theta_2=\cdots=\theta_m=\theta \\ \qquad\qquad \theta_{m+1}=\theta_{m+2}=\cdots=\theta_n=\theta' \end{cases} \quad (\theta \neq \theta')$$

という検定問題を考える．m の値を既知とすれば，通常の二標本問題になるが，m が未知でその探索が目的となると事情が異ってくる．対立仮説の方が真の場合，m はパラメータの変化する時点を示すことになるので，観測値の順序のもつ情報を有効に利用しなければならない．第2節でも述べるように，この種の問題は正規分布という仮定の下ではいろいろ議論されているが，ノンパラメトリックな立場からのものはほとんどみられない．そこでこの報告では，記録値に基づく統計量と反転の総数というノンパラメトリックな検定を考える．記録値とは，それ以前の観測値の中での最大値をとる値のことである．まず記録値の総数が考えられる．この値が

$$(1)\qquad R_n=\sum_{i=1}^{n}A_i,\quad A_i=\begin{cases}1, & X_i>\max(X_1,\ \cdots,\ X_{i-1})\\ 0, & \text{otherwise}\end{cases}$$

異常に大きすぎたり，逆に小さすぎたらパラメータが変動したものと考えられる．Rényi [10] は

$$(R_n-\log n)\ /\ \sqrt{\log n}$$

の極限分布が標準正規になることを示した．しかし我々がシミュレーションしてみた結果では収束が余り早くない．A_i の期待値が $1/i$ であることから

$$(2)\qquad Q_n=\sum_{i=1}^{n}\left(A_i-\frac{1}{i}\right)^2$$

という統計量を考えることもできる．

次の定理 1，2 の証明は第 3 節で与える．

定理 1.

$$(3)\qquad \mathrm{E}\,\{R_n|H_0\}=\alpha_n\equiv\sum_{i=1}^{n}\frac{1}{i}$$

$$(4)\qquad \mathrm{V}\,\{R_n|H_0\}=\beta_n\equiv\sum_{i=1}^{n}\frac{1}{i}\left(1-\frac{1}{i}\right)$$

$$(5)\qquad \mathrm{E}\,\{Q_n|H_0\}=\beta_n$$

$$(6)\qquad \mathrm{V}\,\{Q_n|H_0\}=\gamma_n\equiv\sum_{i=1}^{n}\frac{1}{i}\left(1-\frac{1}{i}\right)\left(1-\frac{2}{i}\right)^2$$

定理 2.

$$(7)\qquad \mathrm{E}\,\{R_n|H_m\}=\alpha_n+H$$

$$(8)\qquad \mathrm{V}\,\{R_n|H_m\}=\beta_n+H\,(1-H)-2\,(\alpha_n-\alpha_m)\,H+K$$

$$(9)\qquad \mathrm{E}\,\{Q_n|H_m\}=\beta_n+I$$

$$(10)\qquad \mathrm{V}\,\{Q_n|H_m\}=\gamma_n+J+L$$

ここで

$$H=H_{n,m}=\sum_{i=m+1}^{n}\varDelta_i$$

$$I=I_{n,m}=\sum_{i=m+1}^{n}\left(1-\frac{2}{i}\right)\varDelta_i$$

$$J=J_{n,m}=\sum_{i=m+1}^{n}\left(1-\frac{2}{i}\right)^2\varDelta_i\,(1-\varDelta_i)$$

$$K=K_{n,m}=2\sum_{i=m+1}^{n-1}\sum_{j=i+1}^{n}\varDelta_{i,j}$$

$$L=L_{n,m}=2\sum_{i=m+1}^{n-1}\sum_{j=i+1}^{n}\left(1-\frac{2}{i}\right)\left(1-\frac{i}{j}\right)\varDelta_{i,j}$$

$$\varDelta_\alpha\,(x)=\int_{-\infty}^{x}[F\,(t)^m-G\,(t)^m]\,G\,(t)^\alpha d\,G\,(t)\qquad(\alpha\geqq 0)$$

$$\varDelta_i=\varDelta_{i-m-1}\,(\infty)\qquad(i\geqq m+1)$$

$$\varDelta_{i,i+1}=\int\varDelta_{i-m-1}\,(x)\,d\,G\,(x)\qquad(i\geqq m+1)$$

$$\varDelta_{i,j}=\int\varDelta_{j-m-2}\,(x)\,d\,G\,(x)$$

$$+\int dG(x)\int^{x}(j-i-1)\,G(y)^{j-i-2}\Delta_{i-m-1}(y)\,dG(y)$$
$$(j-2\geqq i\geqq m+1)$$

$$\Delta_i=\Delta_{i,j}=0 \qquad \text{(otherwise)}$$

ここで我々は，$\theta=0$ と仮定し $F(x-\theta')=G(x)$ と置いている．

次に我々は，反転の総数

$$(11)\qquad T_n=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Delta(X_i,\ X_j)$$

$$\Delta(x,\ y)=\begin{cases}1 & x>y\\ 0 & x\leqq y\end{cases}$$

について考える．次の定理 3, 4 の証明は第 4 節で与える．なお，定理 3 は Feller [2] p. 241 によっても与えられている．

定理 3.

$$(12)\qquad \mathrm{E}\{T_n|H_0\}=\frac{n(n-1)}{4}$$

$$(13)\qquad \mathrm{V}\{T_n|H_0\}=\frac{n(n-1)(2n+5)}{72}$$

定理 4.

$$(14)\qquad \mathrm{E}\{T_n|H_m\}=\frac{n(n-1)}{4}-m(n-m)\,\varepsilon$$

$$(15)\qquad \mathrm{V}\{T_n|H_m\}=\frac{n(n-1)(2n+5)}{72}$$
$$+m(n-m)\{(-\varepsilon+\varepsilon^2+2\beta)\,n+(3\varepsilon-3\alpha-4\beta+2\gamma)\,m$$
$$+(-\varepsilon-\varepsilon^2+3\alpha-2\gamma)\,n\}$$

ここで

$$\int G(x)\,F(x)\,dF(x)=\frac{1}{3}-\alpha$$
$$\int G(x)^2\,dF(x)=\frac{1}{3}-2\beta$$
$$\int F(x)\,G(x)\,dG(x)=\frac{1}{3}+\gamma$$
$$\int F(x)^2\,dG(x)=\frac{1}{3}+2\delta$$
$$\int[F(x)-G(x)]\,dG(x)=\varepsilon$$

確率分布が分布形 F に関係するが解析の最も容易な統計量は

$$(16)\qquad L_n=L_n(\boldsymbol{a})=\sum_{i=1}^{n}a_iX_i$$

という線形のものである．明らかに

$$\mathrm{E}\{L_n|H_0\}=\theta\sum_{i=1}^{n}a_i$$

$$\mathrm{E}\{L_n|H_m\}=\theta\sum_{i=1}^{n}a_i+(\theta'-\theta)\sum_{i=m+1}^{n}a_i$$

$$\mathrm{V}\{L_n|H_0\}=\mathrm{V}\{L_n|H_m\}=\sigma^2\sum_{i=1}^{n}a_i^2 \quad (\sigma^2=\mathrm{V}\{x_i\})$$

そこで

$$a_i\geqq 0,\ \sum_{i=1}^{n}a_i^2=1$$

という条件の下で

$$(17) \qquad \sum_{i=m+1}^{n}a_i$$

を最大にするよう a_i を定めると効率がよい．つまり

$$a_i=0 \qquad i=1,\ \cdots,\ m$$

$$a_j=1/\sqrt{n-m} \qquad j=m+1,\ \cdots,\ n$$

と置いたときに (17) は最大値 $\sqrt{n-m}$ をとる．ところがここで考えている問題では m の値が未知で，どの辺の値をとり易いかの情報もないと仮定している．そこで，m のとり得るすべての値に同じ重みをつけて，平均的な意味での効率をよくするよう定める．つまり

$$\sum_{m=1}^{n-1}\left\{\sqrt{n-m}-\sum_{j=m+1}^{n}a_j\right\}$$

を最小にするよう，結局

$$\sum_{m=1}^{n-1}\sum_{j=m+1}^{n}a_j$$

を最大にするように定めればよい．即ち

$$(18) \qquad a_i^*=\frac{i-1}{\sqrt{n(n-1)(2n-1)/6}}$$

と置いた統計量 $L_n^*=L_n(\boldsymbol{a}^*)$ が求めるものとなる．

証明は簡単なので省略して，L_n^* のモーメントについての結果のみを示しておく．

定理 5.

$$(19) \qquad \mathrm{E}\{L_n^*|H_0\}=\frac{n(n-1)/2}{\sqrt{n(n-1)(2n-1)/6}}\theta$$

$$(20) \qquad \mathrm{E}\{L_n^*|H_m\}=\frac{\theta n(n-1)/2+(\theta'-\theta)(n-m)(n+m-1)/2}{\sqrt{n(n-1)(2n-1)/6}}$$

$$(21) \qquad V\{L_n^*|H_0\}=\mathrm{V}\{L_n^*|H_m\}=\sigma^2$$

4種の統計量 R_n, Q_n, T_n, L_n^* に基づく検定の検定力を $n=100$ で正規分布の場合について検討してみた結果，総じて L_n^* に基づく検定の検定力が高かった．次いで検定力が高いのは T_n による検定で，記録値に関連した検定の検定力はかなり低いことが明らかになった．これは，$n=100$ 程度では記録値がそう多くでることがないことからも，当然予想される事実である．また記録値や反転の総数に基づく統計量は，水準が移動するというタイプの対立仮説 H_m より

$$H_m':\ \theta_i=\theta+(i-1)\delta,\ i=1,\ 2,\ \cdots,\ n$$

といった傾向変動の対立仮説に対してのほうがむしろ感度を示すであろう．

2. 歴史的考察

F が正規分布で m 以外のパラメータがすべて既知とすると，$\theta<\theta'$ の場合の最尤推定は

$$\sum_{i=m+1}^{n}\left(X_i-\frac{\theta+\theta'}{2}\right)$$

を最大にする $\hat{m}$ として与えられるから

$$M=\max_{0<m<n}\sum_{i=m+1}^{n}\left(X_i-\frac{\theta+\theta'}{2}\right)$$

という統計量を考えればよい．$\hat{m}$ や M の漸近分布が Hinkley [6] によって与えられている．このM と類似の

$$\tilde{M}=\max_{0<m\leq n}\{S_m-\min_{0<i<m}S_i\},\quad S_m=\sum_{i=1}^{m}(x_i-\theta)$$

という統計量が，この問題の創始者 Page [9] によって提示されている．彼は $\tilde{M}$ の確率分布の解析が厄介なことから，F を2項分布で近似して（つまり各 X_i が θ より大きいか小さいかだけを調べて）検定する簡便法を解析した．この種の統計量は逐次和 S_m の振巾に依存することから，θ, θ' の差に比較して分散が大きい場合には感度がにぶくなる．

Chernoff-Zacks [2] は，m と $\theta-\theta'$ にある種の先験分布を仮定したときのベイズ解による推定と検定の問題を考え，θ が既知の場合には

$$T=\sum_{i=1}^{n}(i-1)\,X_i \tag{22}$$

θ が未知の場合には

$$\tilde{T}=\sum_{i=1}^{n}(i-1)\,(X_i-\bar{X})$$

という検定統計量を導出し，(22) に基づく検定と Page の検定法との比較を行い，変動の時点がそれほど早めには起らない限り T による検定法の検定力のほうが高いことを確めている．なお (22) の統計量は (16) と (18) で定まる統計量 L_n^* と同じになる．

Kander-Zacks [7] は分布形が必ずしも正規でなく，一般の指数型分布族の場合のベイズ解も

$$T^*=\sum_{i=1}^{n}(i-1)\,U(X_i)$$

というタイプの統計量になることを示した．T^* に基づく検定の有意水準や検定力を求めるための手順（正規分布による近似法も）が与えられているが，その計算が結構厄介なことや U の形を定めるのに分布形 F の知識を必要とすることなどから考えて，実用性はあまり認められない．

Sen-Srivastava [11] は，F が分散既知の正規分布に従う場合の m に対する最尤推定の考えから

$$N=\max_{0<m<n}D_m\Big/\sqrt{\frac{n}{m\,(n-m)}}\ ,$$

$$D_m=\frac{1}{n-m}\sum_{i=m+1}^{n}X_i-\frac{1}{m}\sum_{i=1}^{m}X_i$$

という検定統計量を考え，$\tilde{T}$ による検定法との比較をモンテカルロ法で行った．変動の時点が端の場合には $\tilde{T}$ の検定力のほうが高くなるが，かなり多くの場合 N による検定法のほうが勝ることを確めた．また分散の値も未知の場合に対応する統計量として Sen-Srivastava [12] は

$$P=\sum_{i=1}^{n}(i-1)\ (X_i-\bar{X})/s$$

$$\tilde{P}=\sum_{i=1}^{n}(i-1)(X_i-\bar{X})/\tilde{s}$$

$$S=\max_{0<m<n} D_m\Big/\sqrt{\frac{n}{m\ (n-m)}S_m}$$

$$s=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$\tilde{s}=\frac{1}{2(n-1)}\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$$

$$S_m=\frac{1}{n-2}\left\{\sum_{i=1}^{m}\left(X_i-\frac{1}{m}\sum_{i=1}^{m}X_i\right)^2\right.$$

$$\left.+\sum_{i=m+1}^{n}\left(X_i-\frac{1}{n-m}\sum_{i=m+1}^{n}X_i\right)^2\right\}$$

を比較し，θ と θ' との差が小さい場合を除けば，P による検定の検定力が高くなることを計算機実験で確めた．また帰無仮説の下での P^2 の分布が F-分布に従う事実から，P による検定の有意水準も簡単に定められる．

$\theta-\theta'$ の符号が正負どちらであるか判らない両側仮説の場合には $\tilde{M}$ に類似した統計量として

$$V=\max_{0<m<n} S_m-\min_{0<m<n} S_m$$

が考えられる．Nadler-Robbins [8] は，V の帰無仮説及び対立仮説の下での極限分布がある種のガウス過程に関連した確率の計算に帰着できることを証明し，極限分布による近似がある程度の大きさの n で結構間に合うことを計算機実験により示している．

Gardner [4] は極限分布がスミルノフの ω^2- 統計量の分布と一致する2次形式の統計量

$$W=\sum_{m=2}^{n}\left[\sum_{i=m+1}^{n}(X_i-\bar{X})\right]^2$$

について扱い，有意水準や検定力についての考察を行った．彼の与えた方法はまた，任意の複数個の時点で勝手な値だけ変動するかもしれないという，より一般の対立仮説の場合にも適用できる．

Hawkins [5] は

$$U=\max_{0<m<n} E_m,\qquad E_m{}^2=\frac{n}{m\ (n-m)}\sum_{i=1}^{m}(X_i-\bar{X})^2$$

という検定統計量について考えた．E_m は

$$E_m{}^2=\sum_{i=1}^{n}(X_i-\bar{X})^2-(n-2)\ S_m$$

のように表現される（S_m は先に統計量 S を定義する際に用いたもの）から，S_m を最小にする

ことと同じになる．彼は $\{E_1, E_2, \cdots, E_{n-1}\}$ がマルコフ性を持つことから，U の帰無仮説の下での確率分布をある種のガウス過程の確率の計算に帰着させて近似する方法を提示している．

T や T^* の統計量で，観測値のランクだけの情報を用いたものとして

$$B = \sum_{i=1}^{n} (i-1)\ \rho(X_i)$$

$$\tilde{B} = \sum_{i=1}^{n} (i-1)\ \tau(X_i)$$

$$\rho(X_i) = X_i \text{ のランク}$$

$$\tau(X_i) = \begin{cases} \rho(X_i) - \frac{n}{2}, & \rho(X_i) - \frac{n}{2} > 0 \\ 0 & \text{otherwise} \end{cases}$$

といった統計量が考えられる．これは帰無仮説の近くでの平均検出力を最大にするノンパラメトリックな検定として，Bhattacharya-Johnson [1] によって提示されたものである．さらに Sen-Srivastava [11] は

$$A = \max_{0<m<n} \left(\sum_{i=m+1}^{n} \rho(X_i) - \mu_m \right) \Big/ \sigma_m$$

$$\tilde{A} = \max_{0<m<n} \left(\sum_{i=m+1}^{n} \tau(X_i) - \tilde{\mu}_m \right) \Big/ \tilde{\sigma}_m$$

$$\mu_m = \mathrm{E}\left\{ \sum_{i=m+1}^{n} \rho(X_i) \right\},\quad \sigma_m{}^2 = \mathrm{V}\left\{ \sum_{i=m+1}^{n} \rho(X_i) \right\}$$

$$\tilde{\mu}_m = \mathrm{E}\left\{ \sum_{i=m+1}^{n} \tau(X_i) \right\},\quad \tilde{\sigma}_m{}^2 = \mathrm{V}\left\{ \sum_{i=m+1}^{n} \tau(X_i) \right\}$$

という統計量による検定を考え，B や $\tilde{B}$ に基づく検定よりも検定力が勝ることを計算機実験により確かめている．

3. 記録値に基づく統計量

まず帰無仮説 H_0 の下では，A_i, A_j が独立であることを示そう．$i<j$ として

$$P_{ij}(l,\ h) = \Pr\{A_i = l,\ A_j = h\}$$

なる確率を求めてみる．まず $A_j=1$ であるためには，X_j のランク $r(X_j)$ が $r(X_j)=j$ でなければならない．更に $A_i=1$ であるためには，$i \leqq r(X_i) \leqq j-1$ でなければならない．

ところで $r(X_i)=k$ であるような順列の組合せ数は

$$\binom{k-1}{i-1}(i-1)!\,(j-i-1)!$$

であるから，結局 $A_i=1$, $A_j=1$ となるような順列の総数は

$$\sum_{k=i}^{j-1} \binom{k-1}{i-1}(i-1)!\,(j-i-1)! = \frac{(j-1)!}{i}$$

と算出され，これを $j!$ で割算して

$$P_{ij}(1,\ 1) = \frac{1}{i \cdot j}$$

なる関係が得られる．又

$$P_{ij}(0,\ 1)=\Pr\{A_j=1\}-P_{ij}(1,\ 1)$$
$$=\frac{1}{j}-\frac{1}{i\cdot j}=\left(1-\frac{1}{i}\right)\frac{1}{j}$$
$$P_{ij}(1,\ 0)=\Pr\{A_i=1\}-P_{ij}(1,\ 1)$$
$$=\frac{1}{i}-\frac{1}{i\cdot j}=\frac{1}{i}\left(1-\frac{1}{j}\right)$$
$$P_{ij}(0,\ 0)=1-P_{ij}(1,\ 1)-P_{ij}(0,\ 1)-P_{ij}(1,\ 0)$$
$$=\left(1-\frac{1}{i}\right)\left(1-\frac{1}{j}\right)$$

これで

$$\Pr\{A_i=l,\ A_j=h\}=\Pr\{A_i=l\}\Pr\{A_j=h\}\qquad l,\ h=0.1$$

という関係が成立することがわかった．

そこで A_i のモーメントを求めてみると

$$\mathrm{E}\{A_i\}=\mathrm{E}\{A_i^2\}=\mathrm{E}\{A_i^4\}=\frac{1}{i}$$
$$\mathrm{V}\{A_i\}=\mathrm{E}\left\{\left(A_i-\frac{1}{i}\right)^2\right\}=\frac{1}{i}\left(1-\frac{1}{i}\right)$$

さらに

$$\mathrm{E}\left\{\left(A_i-\frac{1}{i}\right)^4\right\}=\frac{1}{i}\left(1-\frac{1}{i}\right)\left[1-\frac{3}{i}+\frac{3}{i^2}\right]$$

であるから

$$\mathrm{V}\left\{\left(A_i-\frac{1}{i}\right)^2\right\}=\frac{1}{i}\left(1-\frac{1}{i}\right)\left[1-\frac{4}{i}+\frac{4}{i^2}\right]$$
$$=\frac{1}{i}\left(1-\frac{1}{i}\right)\left(1-\frac{2}{i}\right)^2$$

これより，定理1の成りたつことが証明できた．また A_i の独立性，A_i 及び $\left(A_i-\frac{1}{i}\right)^2$ の有界性より中心極限定理が成立するから，R_n 及び Q_n の漸近正規性も証明される．

定理 2. を証明するために，まず次の補題1を証明しよう．この補題から，対立仮説 H_m の下では $m+1\leqq i<j\leqq n$ であるような A_i と A_j とは独立でないことがわかる．

補題 1.

(i) $\mathrm{E}\{A_i|H_m\}\equiv\mu_i=\frac{1}{i}+\varDelta_i$

(ii) $\mathrm{E}\left\{\left(A_i-\frac{1}{i}\right)^k\,\middle|\,H_m\right\}\equiv\nu_i^k=\left(1-\frac{1}{i}\right)^k\mu_i+\frac{1}{i^k}(1-\mu_i)$

(iii) $\mathrm{E}\{A_iA_j|H_m\}\equiv\gamma_{i,j}=\frac{1}{ij}+\varDelta_{i,j}\quad(j>i>m)$

$=\mu_i\mu_j$ （他の場合）

(iv) $\mathrm{E}\left\{\left(A_i-\frac{1}{i}\right)^2\left(A_j-\frac{1}{j}\right)^2\,\middle|\,H_m\right\}\equiv\delta_{i,j}$

$$=\left(1-\frac{2}{i}\right)\left(1-\frac{2}{j}\right)\gamma_{i,j}+\frac{1}{j^2}\left(1-\frac{2}{i}\right)\mu_i$$

$$+\frac{1}{i^2}\left(1-\frac{2}{j}\right)\mu_j+\frac{1}{i^2j^2}$$

証　明

初めの2つの関係は明らかであるから，(iii) の関係について考える．まず $1\leqq i\leqq m<j\leqq n$ の場合

$$U=\max(X_1,\ \cdots,\ X_{i-1})$$
$$V=\max(X_{i+1},\ \cdots,\ X_{j-1})$$

と置けば，U, V の分布関数はそれぞれ F^{i-1}, $F^{m-i}G^{j-m-1}$ であるから

$$\begin{aligned}
\mathrm{E}\{A_iA_j|H_m\}&=\Pr\{A_i=1,\ A_j=1|H_m\}\\
&=\Pr\{U<X_i<V<X_j\}+\Pr\{\max(U,\ V)<X_i<X_j\}\\
&=\int dG(x)\int^x d[F(y)^{m-i}G(y)^{j-m-1}]\int^y F(z)^{i-1}dF(z)\\
&\quad+\int dG(x)\int^x F(y)^{m-1}G(y)^{j-m-1}dG(y)
\end{aligned}$$

ここで

$$d[F^\alpha G^\beta]=F^{\alpha-1}G^{\beta-1}[\alpha G\,dF+\beta F\,dG]$$

といった関係を用いれば

$$\begin{aligned}
&=\frac{1}{i}\int F(x)^mG(x)^{j-m-1}dG(x)\\
&=\frac{1}{i}\int G(x)^{j-1}dG(x)+\frac{1}{i}\int[F(x)^m-G(x)^m]G(x)^{j-m-1}dG(x)\\
&=\frac{1}{i}\left(\frac{1}{j}+\Delta_j\right)=\mu_i\mu_j
\end{aligned}$$

次に $m+1\leqq i<j\leqq n$ の場合で $j-i\geqq2$ のケースを考える．前と同様

$$U=\max(X_1,\ \cdots\cdot,\ X_{i-1})$$
$$V=\max(X_{i+1},\ \cdots,\ X_{j-1})$$

と置く．U, V の分布関数はそれぞれ F^mG^{i-m-1}, G^{j-i-1} であるから

$$\begin{aligned}
\mathrm{E}\{A_iA_j|H_m\}&=\Pr\{A_i=1,\ A_j=1|H_m\}\\
&=\Pr\{U<X_i<V<X_j\}+\Pr\{\max(U,\ V)<X_i<X_j\}\\
&=\int dG(x)\int^x(j-i-1)G(y)^{j-i-2}dG(y)\int^y F(z)^mG(z)^{i-m-1}dG(z)\\
&\quad+\int dG(x)\int^x F(y)^mG(y)^{j-m-2}dG(y)\\
&=\int dG(x)\int^x(j-i-1)G(y)^{j-i-2}dG(y)\int^y G(z)^{i-1}dG(z)\\
&\quad+\int dG(x)\int^x G(y)^{j-2}dG(y)+\Delta_{i,j}=\frac{1}{ij}+\Delta_{i,j}
\end{aligned}$$

他のケースも同様にして証明される．

最後に (iv) の関係式を示そう.

$$\left(A_i-\frac{1}{i}\right)^2\left(A_j-\frac{1}{j}\right)^2=A_i^2A_j^2-\frac{2}{j}A_i^2A_j-\frac{2}{i}A_iA_j^2+\frac{4}{ij}A_iA_j$$
$$+\left(A_i^2-\frac{2}{i}A_i\right)\frac{1}{j^2}+\left(A_j^2-\frac{2}{j}A_j\right)\frac{1}{i^2}+\frac{1}{i^2j^2}$$

であるから，両辺の期待値をとることにより

$$\delta_{i,j}=\left(1-\frac{2}{i}\right)\left(1-\frac{2}{j}\right)\gamma_{i,j}$$
$$+\frac{1}{j^2}\left(1-\frac{2}{i}\right)\mu_i+\frac{1}{i^2}\left(1-\frac{2}{j}\right)\mu_j+\frac{1}{i^2j^2}$$

かくて，補題1が証明された.

定理 2. の証明

(7) の関係についての表現は，補題 1. (i) を用いて直ちに導かれる. (8) については

$$\begin{aligned}
\mathrm{V}(R_n)&=\mathrm{E}\{R_n^2|H_m\}-(\mathrm{E}\{R_n|H_m\})^2\\
&=\sum_{i=1}^n\mu_i+2\sum_{i=1}^{n-1}\sum_{j=i+1}^n\gamma_{i,j}-\left(\sum_{i=1}^n\mu_i\right)^2\\
&=\sum_{i=1}^n\mu_i(1-\mu_i)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^n(\gamma_{i,j}-\mu_i\mu_j)\\
&=\sum_{i=1}^n\frac{1}{i}\left(1-\frac{1}{i}\right)+\sum_{i=m+1}^n\left[\left(1-\frac{2}{i}\right)\varDelta_i-\varDelta_i^2\right]\\
&\quad+2\sum_{i=1}^{n-1}\sum_{j=i+1}^n\left[\varDelta_{i,j}-\frac{1}{j}\varDelta_i-\frac{1}{i}\varDelta_j-\varDelta_i\varDelta_j\right]\\
&=\sum_{i=1}^n\frac{1}{i}\left(1-\frac{1}{i}\right)+\sum_{i=m+1}^n\varDelta_i-\sum_{i=m+1}^n\left(\frac{1}{i}+\varDelta_i\right)^2\\
&\quad+2\sum_{i=m+1}^{n-1}\sum_{j=i+1}^n\varDelta_{i,j}+\left(\sum_{i=m+1}^n\frac{1}{i}\right)^2\\
&\quad-2\sum_{i=m+1}^{n-1}\sum_{j=i+1}^n\left(\frac{1}{i}+\varDelta_i\right)\left(\frac{1}{j}+\varDelta_j\right)\\
&=\sum_{i=1}^n\frac{1}{i}\left(1-\frac{1}{i}\right)+\sum_{i=m+1}^n\varDelta_i\\
&\quad+\left(\sum_{i=m+1}^n\frac{1}{i}\right)^2-\left[\sum_{i=m+1}^n\left(\frac{1}{i}+\varDelta_i\right)\right]^2+2\sum_{i=m+1}^{n-1}\sum_{j=i+1}^n\varDelta_{i,j}\\
&=\beta_n+H-2(\alpha_n-\alpha_m)H-H^2+K
\end{aligned}$$

次に (9), (10) の関係を証明する.

$$\mathrm{E}\{Q_n|H_m\}=\sum_{i=1}^n\mathrm{E}\left\{\left(A_i-\frac{1}{i}\right)^2\Big|H_m\right\}$$

$$= \sum_{i=1}^{n} \left[\left(1 - \frac{1}{i}\right)^2 \mu_i + \frac{1}{i^2} (1 - \mu_i) \right]$$

$$= \sum_{i=1}^{n} \left[\left(1 - \frac{2}{i}\right) \mu_i + \frac{1}{i^2} \right]$$

$$= \sum_{i=1}^{n} \frac{1}{i} \left(1 - \frac{1}{i}\right) + \sum_{i=m+1}^{n} \left(1 - \frac{2}{i}\right) \Delta_i = \beta_n + I$$

$$\mathrm{V}\,(Q_n) = \mathrm{E}\left\{ \left[\sum_{i=1}^{n} \left(A_i - \frac{1}{i}\right)^2 \right]^2 \middle| H_m \right\} - \left[\mathrm{E}\left\{ \sum_{i=1}^{n} \left(A_i - \frac{1}{i}\right)^2 \middle| H_m \right\} \right]^2$$

$$= \sum_{i=1}^{n} \nu_i^4 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta_{i,j} - \left(\sum_{i=1}^{n} \left[\left(1 - \frac{2}{i}\right) \mu_i + \frac{1}{i^2} \right] \right)^2$$

$$= \sum_{i=1}^{n} \left[\left(1 - \frac{1}{i}\right)^4 \mu_i + \frac{1}{i^4} (1 - \mu_i) \right]$$

$$+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\left(1 - \frac{2}{i}\right)\left(1 - \frac{2}{j}\right) \gamma_{i,j} + \frac{1}{{}^2}\left(1 - \frac{2}{i}\right) \mu_i \right.$$

$$\left. + \frac{1}{i^2}\left(1 - \frac{2}{j}\right) \mu_j + \frac{1}{i^2 j^2} \right] - \sum_{i=1}^{n} \left[\left(1 - \frac{2}{i}\right) \mu_i + \frac{1}{i^2} \right]^2$$

$$- 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\left(1 - \frac{2}{i}\right) \mu_i + \frac{1}{i^2} \right] \left[\left(1 - \frac{2}{j}\right) \mu_j + \frac{1}{j^2} \right]$$

$$= \sum_{i=1}^{n} \left[\left(1 - \frac{1}{i}\right)^4 \mu_i + \frac{1}{i^4} (1 - \mu_i) \right]$$

$$+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(1 - \frac{2}{i}\right)\left(1 - \frac{2}{j}\right) [\gamma_{i,j} - \mu_i \mu_j]$$

$$- \sum_{i=1}^{n} \left[\left(1 - \frac{2}{i}\right) \mu_i + \frac{1}{i^2} \right]^2$$

$$= \sum_{i=1}^{n} \left(1 - \frac{2}{i}\right)^2 \frac{1}{i}\left(1 - \frac{1}{i}\right) + \sum_{i=m+1}^{n} \left(1 - \frac{2}{i}\right)^2 \Delta_i (1 - \Delta_i)$$

$$+ 2 \sum_{i=m+1}^{n-1} \sum_{j=i+1}^{n} \left(1 - \frac{2}{i}\right)\left(1 - \frac{2}{j}\right) \Delta_{i,j} = \gamma_n + J + L$$

4. 反転の総数

まず反転の総数 T_n の，帰無仮説 H_0 の下での分布について考える．X の分布が連続型とすると，“タイ” の起る確率は0であるから

$$r = (r_1,\ r_2,\ \cdots,\ r_n)$$

$$r_i = X_i \text{ のランク}$$

は，1, 2, ⋯, n の任意の順列の集合 Π の各値を，確率 $1/n!$ でとるような確率変数である．更に k という元によって生ずる反転の数を示す確率変数を S_k とすると

(23) $\quad T_n = S_1 + S_2 + \cdots + S_n$

である．つまり，T_n の分布は H_0 の下では X の分布形には依存しない．

そこでS_kの値がl $(0\leqq l\leqq k-1)$ となるような組合せ数を求めてみる．まずj番目にkがあらわれて“$S_k=l$”となる場合の数は

$$\binom{k-1}{k-1-l}\binom{n-k}{j-k+l}(j-1)!\,(n-j)!$$
$$=(k-1)!\,(n-k)!\binom{j-1}{j-k+l}\binom{n-j}{n-j-l}$$

又“$S_k=l$”となるためには

$$k-l\leqq j\leqq n-l$$

でなければならない．Feller [2] による II の等式 (12.16) によれば

$$\sum_{j=k-l}^{n-l}\binom{j-1}{j-k+l}\binom{n-j}{n-j-l}$$
$$=\sum_{h=0}^{n-k}\binom{k-l+h-1}{h}\binom{n-k+l-h}{n-k-h}=\binom{n}{n-k}$$

即ち，“$S_k=l$”となる組合せの数は

$$(k-1)!\,(n-k)!\binom{n}{n-k}=\frac{n!}{k}$$

表 1 反転の総数T_nのパーセント点
(右側の ( ) 内の数字は正規近似により求めたもので，正確な値と異る場合のみ記してある.)

n	0.5%		1%		1.5%		2.5%		3%		5%	
5	0		1	(0)	1		1		1		2	
6	1		2	(1)	2		2		3	(2)	3	
7	2		3		3		4		4		5	
8	4		5		5		6		6		7	
9	6		7		8		9		9		10	
10	9	(8)	10	(9)	11		12		12		13	
11	12	(11)	13		14		15		15		17	
12	15	(14)	16		17		19		19		21	
13	18		20		21		23		24		26	
14	23	(22)	25	(24)	26		28		28		30	
15	27	(26)	29		31		33		34	(33)	36	
16	32	(31)	35	(34)	36		38		39		42	
17	37		40		42		44		45		48	
18	43	(42)	46		48		51		52		55	
19	49		53	(52)	55		58	(57)	59		62	
20	56	(55)	60	(59)	62		65		66		70	
21	63	(62)	67	(66)	69		73		74		78	
22	71	(70)	75	(74)	77		81		82		86	
23	78		83	(82)	86		90		91		95	
24	87	(86)	92	(91)	95		99	(98)	100		105	
25	96	(95)	101		104		108		110		115	
26	105	(104)	110		114		118		120		125	
27	114		120		124		129	(128)	130		136	
28	125	(124)	131	(130)	134		140	(139)	141		147	
29	135	(134)	141		145		151		153		159	
30	146	(145)	153	(152)	157		163		165		171	

と算出されるから，結局

$$\Pr\{S_k = l\} = \frac{1}{k},\ l = 0,\ 1,\ \cdots,\ k-1$$

また，元 k によって生ずる反転の数が 1, 2, $\cdots$, $k-1$ 及び $k+1$, $\cdots$, n のグループ内での順序には無関係に定まることから，S_1, S_2, $\cdots$, S_n は互に独立な確率変数で

$$\mathrm{E}\{S_k\} = \frac{k-1}{2},\ \mathrm{V}\{S_k\} = \frac{k^2-1}{12}$$

と (23) の関係より定理3の成立することがわかる．

$n \leqq 30$ の場合についての T_n のパーセント点の計算結果が表1に示してある．正規近似による値でも結構代用できることがわかる．これは，記録値に基づく統計量 R_n, Q_n の分布が正規分布へ近づく速度がかなり遅いのと対称的である．

次に，対立仮説 H_m の下での T_n のモーメントを求めてみよう．Y_i を $F(x)$ に従う変量，Z_i を $G(x)$ に従う変量とし，各変量は互に独立とする．次の補題2が成立することは容易に判る．

補題 2.

$$\mathrm{E}\{\Delta(Y_1, Y_2)^2\} = \mathrm{E}\{\Delta(Y_1, Y_2)\} = \frac{1}{2}$$

$$\mathrm{E}\{\Delta(Y_1, Z_1)^2\} = \mathrm{E}\{\Delta(Y_1, Z_1)\} = \frac{1}{2} - \varepsilon$$

$$\mathrm{E}\{\Delta(Z_1, Z_2)^2\} = \mathrm{E}\{\Delta(Z_1, Z_2)\} = \frac{1}{2}$$

又，各種共分散に対しては次のような記号を用いることにする．

$$A_1 = \mathrm{E}\{\Delta(Y_1, Y_2)\,\Delta(Y_1, Y_3)\}$$
$$A_2 = \mathrm{E}\{\Delta(Y_1, Y_2)\,\Delta(Y_1, Z_1)\}$$
$$A_3 = \mathrm{E}\{\Delta(Y_1, Z_1)\,\Delta(Y_1, Z_2)\}$$
$$A_4 = \mathrm{E}\{\Delta(Z_1, Z_2)\,\Delta(Z_1, Z_3)\}$$

$$B_1 = \mathrm{E}\{\Delta(Y_1, Y_2)\,\Delta(Y_3, Y_2)\}$$
$$B_2 = \mathrm{E}\{\Delta(Y_1, Z_1)\,\Delta(Y_2, Z_1)\}$$
$$B_3 = \mathrm{E}\{\Delta(Y_1, Z_1)\,\Delta(Z_2, Z_1)\}$$
$$B_4 = \mathrm{E}\{\Delta(Z_1, Z_2)\,\Delta(Z_3, Z_2)\}$$

$$C_1 = \mathrm{E}\{\Delta(Y_1, Y_2)\,\Delta(Y_2, Y_3)\}$$
$$C_2 = \mathrm{E}\{\Delta(Y_1, Y_2)\,\Delta(Y_2, Z_1)\}$$
$$C_3 = \mathrm{E}\{\Delta(Y_1, Z_1)\,\Delta(Z_1, Z_2)\}$$
$$C_4 = \mathrm{E}\{\Delta(Z_1, Z_2)\,\Delta(Z_2, Z_3)\}$$

補題 3.

i	A_i	B_i	C_i
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$
2	$\frac{1}{3}-\alpha$	$\frac{1}{3}+2(\delta-\varepsilon)$	$\frac{1}{6}+\alpha-\varepsilon$
3	$\frac{1}{3}-2\beta$	$\frac{1}{3}+\gamma-\varepsilon$	$\frac{1}{6}-\gamma$
4	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

証　明

他の場合も同じようにして証明できるから B_3 の値だけを求めてみよう．

$$\begin{aligned} B_3 &= \mathrm{E}\{\Delta(Y_1, Z_1)\,\Delta(Z_2, Z_1)\} \\ &= \Pr\{\min(Y_1, Z_2) > Z_1\} \\ &= \int [1-F(x)][1-G(x)]\,dG(x) \\ &= \int dG(x) - \int [F(x)-G(x)]\,dG(x) \\ &\quad -2\int G(x)\,dG(x) + \int F(x)\,G(x)\,dG(x) \\ &= -\varepsilon + \frac{1}{3} + \gamma \end{aligned}$$

定理 4. の証明

$$\begin{aligned} T_n &= \sum_{1\le i<j\le n} \Delta(X_i, X_j) \\ &= \sum_{1\le i<j\le m} \Delta(X_i, X_j) + \sum_{1\le i\le m<j\le n} \Delta(X_i, X_j) + \sum_{m<i<j\le n} \Delta(X_i, X_j) \\ &\equiv T_n^{(1)} + T_n^{(2)} + T_n^{(3)} \end{aligned}$$

補題2より

$$\mathrm{E}\{T_n^{(1)}\,|\,H_m\} = \frac{m(m-1)}{2}\cdot\frac{1}{2}$$

$$\mathrm{E}\{T_n^{(2)}\,|\,H_m\} = m(n-m)\left(\frac{1}{2}-\varepsilon\right)$$

$$\mathrm{E}\{T_m^{(3)}\,|\,H_m\} = \frac{(n-m)(n-m-1)}{2}\cdot\frac{1}{2}$$

従って

$$\mathrm{E}\{T_n\,|\,H_m\} = \frac{n(n-1)}{4} - m(n-m)\,\varepsilon$$

次に (15) の関係を証明するために次のように分解をする．

$$\begin{aligned} T_n^2 &= P + U + V + W + Q \\ P &= \sum_{1\le i<j\le n} \Delta(X_i, X_j)^2 \\ U &= 2\sum_{1\le i<j<l\le n} \Delta(X_i, X_j)\,\Delta(X_i, X_l) \\ V &= 2\sum_{1\le i<k<j\le n} \Delta(X_i, X_j)\,\Delta(X_k, X_j) \\ W &= 2\sum_{1\le i<j<l\le n} \Delta(X_i, X_j)\,\Delta(X_j, X_l) \\ Q &= 24\sum_{1\le i<j<k<l\le n} \Delta(X_i, X_j)\,\Delta(X_k, X_l) \end{aligned}$$

まず

$$\mathrm{E}\{P\,|\,H_n\} = \mathrm{E}\{T_n\,|\,H_n\} = \frac{n(n-1)}{4} - m(n-m)\,\varepsilon$$

次に U のモーメントは

$$\begin{aligned} U &= 2\sum_{1\le i<j<l\le m} \Delta(X_i, X_j)\,\Delta(X_i, X_l) \\ &\quad + 2\sum_{1\le i<j\le m<l\le n} \Delta(X_i, X_j)\,\Delta(X_i, X_l) \end{aligned}$$

$$+2\sum_{1\leq i\leq m<j<l\leq n}\Delta(X_i,X_j)\,\Delta(X_i,X_l)$$

$$+2\sum_{m<i<j<l\leq n}\Delta(X_i,X_j)\,\Delta(X_i,X_l)$$

のように分解して補題3を用いれば

$$\begin{aligned}\mathrm{E}\{U|H_m\}&=\frac{m(m-1)(m-2)}{3}A_1\\&\quad+m(m-1)(n-m)A_2\\&\quad+m(n-m)(n-m-1)A_3\\&\quad+\frac{(n-m)(n-m-1)(n-m-2)}{3}A_4\\&=\frac{n(n-1)(n-2)}{9}-m(n-m)\left[(m-1)\alpha+(n-m-1)2\beta\right]\end{aligned}$$

同様にして

$$\begin{aligned}\mathrm{E}\{V|H_m\}&=\frac{n(n-1)(n-2)}{9}\\&\quad-m(n-m)\left[(m-1)2(\varepsilon-\delta)+(n-m-1)(\varepsilon-\gamma)\right]\\\mathrm{E}\{W|H_m\}&=\frac{n(n-1)(n-2)(n-3)}{18}\\&\quad-m(n-m)\left[(m-1)(\varepsilon-\alpha)+(n-m-1)\gamma\right]\\\mathrm{E}\{Q|H_m\}&=\frac{n(n-1)(n-2)(n-3)}{16}\\&\quad-m(n-m)\left[\{(m-1)(m-2)+(n-m-1)(n-m-2)\}\frac{\varepsilon}{2}\right.\\&\quad\left.+(m-1)(n-m-1)\varepsilon(1-\varepsilon)\right]\end{aligned}$$

といった関係も得られる．これらの結果から (15) の関係が示される．

参 考 文 献

[1] Bhattacharya, G.K. and Johnson, R.A. (1968) Nonparametric tests for shift at unknown time point, *Ann. Math. Statist.* **39**, 1731–1743.

[2] Chernoff, H. and Zacks, S. (1964) Estimating the current mean of a normal distribution which is subjected to changes in time, *Ann. Math. Statist.* **35**, 999–1018.

[3] Feller, W. (1957) *An Introduction to Probability Theory and its Applications*, second edition, John Wiley.

[4] Gardner, L.A. (1969) On detecting changes in the mean of normal variates, *Ann. Math. Statist.* **40**, 116–126.

[5] Hawkins, D.M. (1977) Testing a sequence of observations for a shift in location, *J. Amer. Statist. Ass.* **72**, 180–186.

[6] Hinkley, D.V. (1970) Inference about the change point in a sequence of random variables, *Biometrika* **57**, 1–17.

[7] Kander, Z. and Zacks, S. (1966) Test procedures for possible changes in parameters of statistical distributions occuring at unknown time points, *Ann. Math. Statist.* **37**, 1196–1210.

[8] Nadler, J. and Robbins, N.B. (1971) Some characteristics of Page's two-sided procedure for detecting a change in a location parameter, *Ann. Math. Statist.* **42**, 538–551.

[9] Page, E.S. (1955) A test for a change in a parameter occuring at unknown point, *Biometrika*, **42**, 523–526.

[10] Rényi, A. (1962) Théorie des éléments saillants d'une suite d'observations, *Colloquium Aarhus*, August **1-10**, 104–117.
[11] Sen, A. and Srivastava, M.S. (1975) On tests for detecting change in mean, *Ann. Statist.* **3**, 98–108.
[12] Sen, A. and Srivastava, M.S. (1975) Some one-sided tests for change in level, *Technometrics* **17**, 61–64.