

# ノンパラメトリック群逐次検定

北里大学大学院薬学研究科臨床統計部門 宇野 一 (Hajime Uno)  
Department of Biostatistics, Center for Clinical Pharmacy and Clinical Sciences,  
Kitasato University Graduate School

## 1 序

1977年に Pocock が提案した「群逐次検定」は、ある程度の例数が集積される毎に逐次検定を行い、データ集積途中であっても有意差がみられたらデータ収集を中止するとともに、帰無仮説を棄却する方法論である。「群逐次検定」の主な応用場面には、臨床試験における中間解析があり、「群逐次検定」における中止規準を試験の中止・継続の判断材料としている実例が多くみられる。「群逐次計画」に関する方法論のトピックの1つとしては、検定を繰り返し行うことに対する第一種の過誤確率の制御がある。各解析時点における棄却限界値の具体的な計算方法は、1) recursive integration, 2) 多変量正規分布の数値積分を用いる方法の2つに大別でき、検定統計量の列が独立増分構造を持つ場合には、1) の recursive integration を用いることで、また、検定統計量の列が独立増分構造を持たない場合であっても、多変量正規分布となる場合には2) の多変量正規分布の数値積分ルーチンが利用できる。これまでに、正規応答などの単純な状況をはじめとして、生存時間データや多重エンドポイントのデータを扱うような複雑な試験デザインに対し、検定統計量の列が独立増分構造を持つこと、あるいは検定統計量の列の同時分布が多変量正規分布に収束することが示されており、いろいろな試験デザイン、いろいろな検定統計量に対して、第一種の過誤が名目有意水準に制御できることが示されている。

分布関数についてのノンパラメトリック検定の1つであるコルモゴロフ・スミルノフ検定については、検定統計量の漸近分布が正規分布に収束しない、また独立増分が成立しているとも考えがたく、これまで提案されてきた方法では、棄却限界値の列は算出できない。従って、コルモゴロフ・スミルノフ検定を用いて群逐次検定を実施する場合には、新たな方法論が必要となる。

Hu and Lagakos (1999) は、観測データが有限次元の母数によって記述される確率変数ではなく、観測データが確率過程としてあらわされる場合に、その平均値関数の信頼帯を群逐次デザインにおいて求める方法、“繰り返し信頼帯”を提案した。繰り返し信頼帯は以下のように定義される、

$$\Pr\{\mu(\cdot) \in B_k, \text{ for } 1 \leq k \leq K\} \geq 1 - \alpha,$$

ここで  $\mu(\cdot)$  は平均値関数、 $K$  は予定している最大の解析回数、また、 $B_k$  は  $k$  回目の解析における信頼帯を表す。繰り返し信頼帯を用いて群逐次検定を行う手順は以下の通りである。

- $k$  回目の解析において
  - 信頼帯  $B_k$  が帰無の平均値関数,  $\mu_0(\cdot)$ , を含まない場合, 帰無仮説,  $H_0 : \mu(\cdot) = \mu_0(\cdot)$  を棄却し, データ観測を中止する.
  - 信頼帯  $B_k$  が帰無の平均値関数,  $\mu_0(\cdot)$ , を完全に含んでいる場合, データ観測を継続する.
- 最終解析 ( $K$  回目の解析) において
  - 信頼帯  $B_K$  が帰無の平均値関数,  $\mu_0(\cdot)$ , を含まない場合, 帰無仮説,  $H_0 : \mu(\cdot) = \mu_0(\cdot)$  を棄却する.
  - 信頼帯  $B_K$  が帰無の平均値関数,  $\mu_0(\cdot)$ , を完全に含んでいる場合, 帰無仮説,  $H_0 : \mu(\cdot) = \mu_0(\cdot)$  を受容する.

本稿では, このアプローチを分布関数に対して応用し, コルモゴロフ・スミルノフ検定を用いた群逐次検定方式および実データに対する適用例を示す.

## 2 解析方法

### 2.1 プロトタイプケース

$X_1, X_2, \dots, X_n$  を同一の分布関数  $F$  から得られた独立な確率変数とする. このとき, 経験分布関数は,

$$F_n(t) \equiv \frac{1}{n} \sum_{i=1}^n 1[X_i \leq t], \quad (1)$$

となり, 帰無仮説  $H_0 : F(\cdot) = F_0(\cdot)$  に対して, コルモゴロフ・スミルノフ検定の検定統計量は,

$$Z = \sup_t \sqrt{n} |F_n(t) - F_0(t)|, \quad (2)$$

であらわされる.

この検定統計量を用いた群逐次検定を考える. まず, 試験途中の解析も含め, 全部で  $K$  回の解析を行うとする. また, 各々の解析時点を  $T_1 < T_2 < \dots < T_K$  で表す. 群逐次検定においては, 終了時までには観測されるデータのうちのいくつかのデータは途中の解析時点でまだ観測されていない. 従って,  $k$  回目の解析時点  $T_k$  における経験分布関数を以下のように表す,

$$F_n(T_k, t) = \frac{\sum_{i=1}^n 1[S_i \leq T_k] \times 1[X_i \leq t]}{\sum_{i=1}^n 1[S_i \leq T_k]}. \quad (3)$$

ここで,  $S_i$  は  $X_i$  が実際に観測された時間を表す. なお,  $S_i$  と  $X_i$  は互いに独立とする.

このとき,  $K$ -次元確率過程,

$$\begin{pmatrix} G_n(T_1, t) \\ G_n(T_2, t) \\ \vdots \\ G_n(T_K, t) \end{pmatrix} \equiv \begin{pmatrix} \sqrt{n}(F_n(T_1, t) - F(t)) \\ \sqrt{n}(F_n(T_2, t) - F(t)) \\ \vdots \\ \sqrt{n}(F_n(T_K, t) - F(t)) \end{pmatrix} \quad (4)$$

は, 平均 0, 共分散関数,

$$\frac{n \sum_{i=1}^n 1[S_i \leq T_k] 1[S_i \leq T_l]}{\sum_{i=1}^n 1[S_i \leq T_k] \sum_{i=1}^n 1[S_i \leq T_l]} F(s)[1 - F(t)], \text{ for } s < t. \quad (5)$$

の  $K$ -次元ガウシアンプロセスに弱収束する.

群逐次コルモゴロフ・スミルノフ検定の検定統計量の列を,

$$(\sup_t |G_n(T_1, t)|, \sup_t |G_n(T_2, t)|, \dots, \sup_t |G_n(T_K, t)|). \quad (6)$$

であらわす. このとき検定の第一種の過誤確率を  $\alpha$  に制御するためには, 以下を満足する定数の列,  $c_1, \dots, c_K$ , を定め, それを各解析時点における棄却限界値とすればよい.

$$\Pr[\sup_t |G_n(T_l, t)| \leq c_l \text{ for } 1 \leq l \leq k-1 \text{ and } \sup_t |G_n(T_k, t)| > c_k] = \pi_k \text{ for } k = 1, \dots, K, \quad (7)$$

ここで  $\sum_{k=1}^K \pi_k = \alpha$  である.

しかしながら, 検定統計量の列は独立増分構造をもつとは考えがたく, また,  $\sup_t |G_n(T_k, t)|$  の極限分布は正規分布とならないため, recursive integration (Armitage et. al (1969)) や, 多次元正規分布の多重積分ルーチン (Schervish (1984)) は利用できない. 従って, 各解析時点における棄却限界値を解析的に算出するのは非常に困難となる. そこで, Hu and Lagakos (1999) の方法を応用して, Lin, Wei and Ying (1993) で示された考え方を分布関数に対して適用し, 棄却限界値の列をシミュレーションにより求める方法を以下に示す.

#### 命題 1

$k = 1, \dots, K$ , に対し,

$$G_n^L(T_k, t) \equiv \sqrt{n} \left[ \frac{\sum_{i=1}^n 1\{S_i \leq T_k\} [1\{X_i \leq t\} - F_n(T_k, t)] Z_i}{\sum_{i=1}^n 1\{S_i \leq T_k\}} \right], \quad (8)$$

とする. ここで,  $Z_i$ , ( $i = 1, \dots, n$ ) は標準正規分布から得られる互いに独立な乱数である. このとき, データ及びデータが観測された時間  $\{X_i, S_i\}$  の条件付の  $K$ -次元確率過程,  $(G_n^L(T_1, t), \dots, G_n^L(T_K, t))'$ , は,  $K$ -次元ガウシアンプロセスに弱収束し, その平均値関数, 共分散関数は  $(G_n(T_1, t), \dots, G_n(T_K, t))'$  の極限と同一となる. 証明は付録に示した.

$c_1, \dots, c_K$  の具体的な算出アルゴリズムは以下の通り.

1. 各解析時に消費する第一種の過誤確率  $\pi_1, \dots, \pi_K$  の列及びサンプルサイズ,  $n$ , を遅くとも 1 回目の解析を実施する前までに定める.
2. 標準正規分布からの乱数,  $Z_1, \dots, Z_n$ , をそれぞれ独立に全部で  $M$  組発生させる.
3. 第 1 回目の解析,  $k = 1$ , において, 乱数にもとづいて発生させた  $M$  本のサンプルパス,  $G_n^L(T_1, t)$ , それぞれについて,  $\sup_t |G_n^L(T_1, t)|$  の値を計算し, その  $1 - \pi_1$  % 点を得る. それが 1 回目の解析の棄却限界値,  $c_1$ , となる. このとき, シミュレーションにより発生したサンプルパス,  $G_n^L(T_1, t)$ , がどこかで  $c_1$  を超える頻度は, 丁度  $\pi_1 M$  となる.

4. 第2回目の解析,  $k = 2$ , において, 第1回目と同じ手順で発生した各々のサンプルパス,  $G_n^L(T_1, t)$  および  $G_n^L(T_2, t)$ , から  $\sup_t |G_n^L(T_2, t)|$  の値を計算し,  $\sup_t |G_n^L(T_1, t)|$  が  $c_1$  を超えず,  $\sup_t |G_n^L(T_2, t)|$  が  $c_2$  を超える頻度が丁度  $\pi_2 M$  となる値,  $c_2$  を求める.  $c_2$  が2回目の解析における棄却限界値となる.
5. 3回目以降についても同様の手順で, 乱数により発生したサンプルパス  $G_n^L(T_1, t), G_n^L(T_2, t), G_n^L(T_3, t)$ , から順次算出する.

上記の手順で得られた棄却限界値の列を使用し, ある  $k$  で  $G_n(T_k, t)$  の実現値が  $c_k$  を超えたら,  $H_0$  は棄却されるとする. こうして, 群逐次コルモゴロフ・スミルノフ検定の第一種の過誤が  $\alpha$  に制御される.

ここでは, 検定に関心があるので, 信頼帯には主要な関心はないが, 分布関数  $F(\cdot)$  について,  $1 - \alpha$  水準の繰り返し信頼帯は, 以下で表される.

$$B_k \equiv \left\{ \nu(\cdot) : \nu(t) \in \left[ \max\left\{0, \hat{F}_n(T_k, t) - \frac{c_k}{\sqrt{n}}\right\}, \min\left\{\hat{F}_n(T_k, t) + \frac{c_k}{\sqrt{n}}, 1\right\} \right] \right\}, \quad k = 1, \dots, K. \quad (9)$$

## 2.2 2標本の場合

2群の場合の分布関数の比較を考える.  $X_1, \dots, X_m$  を分布関数  $F$  に従う互いに独立な確率変数,  $Y_1, \dots, Y_n$  を分布関数  $G$  に従う互いに独立な確率変数とする. このとき, コルモゴロフ・スミルノフ検定の検定統計量は,

$$D_{FG} \equiv \sup_t \left( \frac{mn}{m+n} \right)^{1/2} |F_m(t) - G_n(t)|,$$

ここで  $F_m(t) = \frac{1}{m} \sum_{i=1}^m 1[X_i \leq t]$ ,  $G_n(t) = \frac{1}{n} \sum_{i=1}^n 1[Y_i \leq t]$  である.

プロトタイプケースと同様に, 全部で  $K$  回の解析を行うこととし, 以下で表される  $K$  次元確率過程を考える.

$$\begin{pmatrix} M(T_1, t) \\ M(T_2, t) \\ \vdots \\ M(T_K, t) \end{pmatrix} \equiv \begin{pmatrix} \sqrt{\frac{n}{m+n}} G_m^X(T_1, t) - \sqrt{\frac{m}{m+n}} G_n^Y(T_1, t) \\ \sqrt{\frac{n}{m+n}} G_m^X(T_2, t) - \sqrt{\frac{m}{m+n}} G_n^Y(T_2, t) \\ \vdots \\ \sqrt{\frac{n}{m+n}} G_m^X(T_K, t) - \sqrt{\frac{m}{m+n}} G_n^Y(T_K, t) \end{pmatrix} \quad (10)$$

ここで,  $G_m^X(T_k, t) = \sqrt{m}(F_m(T_k, t) - F(t))$ ,  $G_n^Y(T_k, t) = \sqrt{n}(G_n(T_k, t) - G(t))$  であり,  $F_m(T_k, t)$  及び  $G_n(T_k, t)$  は, それぞれ  $k$  回目の解析時における経験分布関数である.

$G_m^X(T_k, t)$  と  $G_n^Y(T_k, t)$  はともに, 平均0のガウシアンプロセスに弱収束し, その共分散関数はそれぞれ,  $F(s)[1-F(t)]$ ,  $G(s)[1-G(t)]$  となる ( $s < t$ ). 従って, 帰無仮説  $H_0 : F(\cdot) = G(\cdot)$  の下では,  $(M(T_1, t), M(T_2, t), \dots, M(T_K, t))'$  も平均0の  $K$ -次元ガウシアンプロセスとなり, その共分散関数は,

$$\frac{n \text{Cov}(G_m^X(T_k, s), G_m^X(T_l, t)) + m \text{Cov}(G_n^Y(T_k, s), G_n^Y(T_l, t))}{m+n}, \quad (11)$$

となる。ここで,

$$\text{Cov}(\mathbf{G}_m^X(T_k, s), \mathbf{G}_m^X(T_l, t)) = \frac{m \sum_{i=1}^m 1\{S_i \leq T_k\} 1\{S_i \leq T_l\}}{\sum_{i=1}^m 1\{S_i \leq T_k\} \sum_{i=1}^m 1\{S_i \leq T_l\}} F(s)[1 - F(t)], \text{ for } s < t,$$

また,

$$\text{Cov}(\mathbf{G}_n^Y(T_k, s), \mathbf{G}_n^Y(T_l, t)) = \frac{n \sum_{i=1}^n 1\{S_i \leq T_k\} 1\{S_i \leq T_l\}}{\sum_{i=1}^n 1\{S_i \leq T_k\} \sum_{i=1}^n 1\{S_i \leq T_l\}} G(s)[1 - G(t)], \text{ for } s < t.$$

である。

プロトタイプケースと同様のシミュレーションの方法を用いて, 各  $k = 1, \dots, K$  に対して,  $\mathbf{G}_m^X(T_k, \cdot)$ ,  $\mathbf{G}_n^Y(T_k, \cdot)$  のサンプルパスを発生することが可能となる。従って, 帰無仮説下における  $\mathbf{M}(T_k, \cdot)$  のサンプルパスを発生できる。

$$\mathbf{M}^L(T_k, t) \equiv \left( \frac{mn}{m+n} \right)^{1/2} \left[ \left\{ \frac{\sum_{i=1}^m 1\{S_i \leq T_k\} [1\{X_i \leq t\} - F_m(T_k, t)] Z_i^X}{\sum_{i=1}^m 1\{S_i \leq T_k\}} \right\} - \left\{ \frac{\sum_{j=1}^n 1\{S_j \leq T_k\} [1\{Y_j \leq t\} - G_n(T_k, t)] Z_j^Y}{\sum_{j=1}^n 1\{S_j \leq T_k\}} \right\} \right],$$

とする。ここで,  $Z_i^X$  と  $Z_j^Y$  はともに, 標準正規分布から発生される互いに独立な乱数である。このとき, 帰無仮説下において,  $K$ -次元確率過程,  $(\mathbf{M}^L(T_1, t), \dots, \mathbf{M}^L(T_K, t))'$  は, 各々のデータおよびそのデータの観測時間  $\{X_i, S_i\}, i = 1, \dots, m$  and  $\{Y_j, S_j\}, j = 1, \dots, n$ , の条件付で,  $(\mathbf{M}(T_1, t), \dots, \mathbf{M}(T_K, t))'$  と同一の極限に弱収束する ( $m, n \rightarrow \infty$ )。

群逐次コルモゴロフ・スミルノフ検定統計量は,

$$D_{FG} = \sup_t \left( \frac{mn}{m+n} \right)^{1/2} |F_m(t) - G_n(t)| = \sup_t |\mathbf{M}(T_k, t)|. \quad (12)$$

のようにあらわされるので,  $(\mathbf{M}^L(T_1, t), \dots, \mathbf{M}^L(T_K, t))'$  を用いて, シミュレーションにより発生されたサンプルパスから, 群逐次コルモゴロフ・スミルノフ検定統計量の帰無仮説下,  $H_0: F(\cdot) = G(\cdot)$ , での分布が得られる。

従って, プロトタイプケースと同様のアルゴリズムを用いて, 群逐次コルモゴロフ・スミルノフ検定について, 以下の条件を満足する棄却限界値の列を得ることができる。

$$\Pr[\sup_t |\mathbf{M}(T_l, t)| \leq c_l \text{ for } 1 \leq l \leq k-1 \text{ and } \sup_t |\mathbf{M}(T_k, t)| > c_k] = \pi_k \text{ for } k = 1, \dots, K, \quad (13)$$

ここで,  $\sum_{k=1}^K \pi_k = \alpha$ . もし, 検定統計量の実現値,  $\hat{D}_{FG}(T_k)$ , がある解析時点においてその時点における棄却限界値を超えれば, 帰無仮説を棄却する。

ここでの主な関心は検定であるが,  $F(\cdot) - G(\cdot)$  について,  $(1 - \alpha)$  水準の信頼帯も, 上記の棄却限界値を用いて得られる。  $k$  回目の解析時における信頼帯は以下のように表される。

$$B_k = \left\{ \nu(\cdot) : \nu(t) \in \left[ \hat{F}_m(T_k, t) - \hat{G}_n(T_k, t) - c_k \sqrt{\frac{mn}{m+n}}, \hat{F}_m(T_k, t) - \hat{G}_n(T_k, t) + c_k \sqrt{\frac{mn}{m+n}} \right] \right\}$$

### 2.3 群逐次検定実施後（試験終了後）の解析

帰無仮説下における群逐次の検定統計量の分布が得られるので、群逐次コルモゴロフ・スミルノフ検定についても、 $p$  値は容易に得ることができる。群逐次検定を実施する場合には、終了時の統計量として2つの確率変数が存在することとなる。1つは終了時点の解析回数 (stopping time,  $T^*$ )、もう1つは終了時点での検定統計量 ( $Z_{T^*}^*$ ) である。通常これら2つの確率変数 ( $T^*, Z_{T^*}^*$ ) を terminal statistics と呼ぶ。従って、 $p$  値を定義するためには、この確率変数の空間について順序付けを行う必要がある。順序付けに基づく terminal statistics の大小関係を記号  $\succeq$  であらわすと、 $p$  値は以下のように定義される。

$$p\text{-value} \equiv \Pr\{(T^*, Z_{T^*}^*) \succeq (t^*, z_{t^*}^*) | H_0\}, \quad (15)$$

ここで、記号  $\succeq$  は、定義された順序付けに基づく大小関係を表し、 $(t^*, z_{t^*}^*)$  は、terminal statistics の実現値を表す。

これまでにいろいろな順序付けが提案されているが、ここでは、Tsiatis, Rosner and Mehta (1984) により提案された stage-wise ordering を適用することとする。Stage-wise ordering は以下のように定義される。

$$\begin{aligned} T_1 > T_2 &\rightarrow (T_1, Z_{T_1}) \succeq (T_2, Z_{T_2}) \\ T_1 = T_2 \text{ and } Z_{T_1} > Z_{T_2} &\rightarrow (T_1, Z_{T_1}) \succeq (T_2, Z_{T_2}) \\ T_1 = T_2 \text{ and } Z_{T_1} < Z_{T_2} &\rightarrow (T_1, Z_{T_1}) \preceq (T_2, Z_{T_2}) \\ T_1 < T_2 &\rightarrow (T_1, Z_{T_1}) \preceq (T_2, Z_{T_2}) \end{aligned}$$

## 3 解析例

具体的な解析例として、痛みの尺度として用いられる Visual Analog Scale (VAS) データに対して群逐次コルモゴロフ・スミルノフ検定を適用した結果を示す。VAS は、10cm の直線の左端に、“no pain”、右端に“worst imaginable pain”と記載された痛みを測定する尺度として用いられる。患者は、痛みの強度に応じて、直線上の1箇所に印をつける。データとしては、通常左端から印がつけられたところまでの距離 (cm)、あるいはその距離の10cm に対する割合 (%) で表される。データの特徴としては、VAS のとり得る値が、0cm~10cm、あるいは0%~100% というように、一定の区間に限られる連続データとなる。

データを表1に示した (Akhtar-Darnesh and Appleton (2000))。抜歯後に、placebo あるいは paracetamol が投与され、痛みの強さが、0, 15, 30, 45, 60, 90 分後に測定された。痛みの経時的プロファイルについて関心がある場合、これらのデータを経時観察データとして解析すべきであるが、方法論の具体例を示すことを目的として、ここでは、抜歯後90分の測定値のみに着目して解析した。また、この試験自体は中間解析が実際に行われたわけではないが、途中2回 (全3回) の解析を行うことと仮定し、群逐次コルモゴロフ・スミルノフ検定方式を適用した。

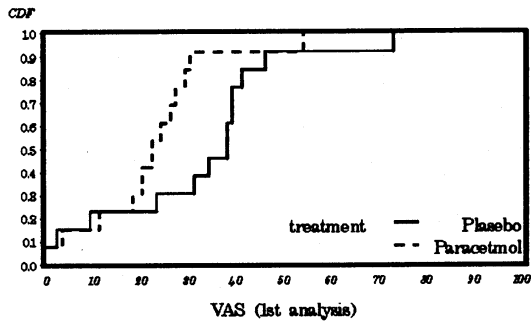
各解析時点における、paracetamol および placebo の経験分布関数を図1に、またそれらの差を図2に示した。また、シミュレーションメソッドにより生成したサンプルパスのいくつかを図3に例示した。総解析回数を3回、有意水準両側5%ととし、 $(\pi_1, \pi_2, \pi_3)$  を (0.07%, 1.57%, 3.36%) と設定した。paracetamol, placebo の分布関数が等しいという帰無仮説の下、10,000本のパス

を生成し、検定統計量の分布を求めた。各解析時点における棄却限界値は、(4.84, 2.61, 1.89) となった。最終解析における検定統計量の実現値が 2.097 であったことから、この例では最終解析において帰無仮説が棄却されることとなる(表 2)。

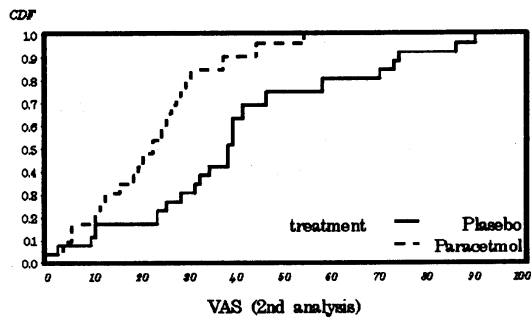
Placebo			Paracetamol		
患者 ID	解析	VAS (90min.)	患者 ID	解析	VAS (90min.)
X-01	1	34	Y-01	1	30
X-02	1	9	Y-02	1	29
X-03	1	31	Y-03	1	3
X-04	1	38	Y-04	1	18
X-05	1	46	Y-05	1	22
X-06	1	0	Y-06	1	26
X-07	1	23	Y-07	1	11
X-08	1	41	Y-08	1	20
X-09	1	38	Y-09	1	20
X-10	1	38	Y-10	1	24
X-11	1	2	Y-11	1	54
X-12	1	39	Y-12	1	27
X-13	1	73	Y-13	1	0
X-14	2	38	Y-14	2	24
X-15	2	25	Y-15	2	5
X-16	2	58	Y-16	2	37
X-17	2	10	Y-17	2	3
X-18	2	46	Y-18	2	28
X-19	2	90	Y-19	2	10
X-20	2	74	Y-20	2	12
X-21	2	32	Y-21	2	15
X-22	2	10	Y-22	2	44
X-23	2	70	Y-23	2	5
X-24	2	86	Y-24	2	19
X-25	2	28	Y-25	2	37
X-26	2	39	Y-26	2	25
X-27	3	16	Y-27	3	12
X-28	3	41	Y-28	3	21
X-29	3	63	Y-29	3	36
X-30	3	0	Y-30	3	36
X-31	3	34	Y-31	3	14
X-32	3	48	Y-32	3	45
X-33	3	42	Y-33	3	18
X-34	3	59	Y-34	3	0
X-35	3	38	Y-35	3	32
X-36	3	43	Y-36	3	60
X-37	3	45	Y-37	3	89
X-38	3	48	Y-38	3	46
			Y-39	3	19

表 1: VAS データ (Akhtar-Darnesh and Appleton (2000) から引用)。欠側値は省略した。また、列 "患者 ID" と "解析" は方法論の具体例を示す目的で追記した。

Stage 1 (patient: X-01 ~ X-13 and Y-01 ~ Y-13)



Stage 2 (patient: X-01 ~ X-26 and Y-01 ~ Y-26)



Stage 3 (patient: X-01 ~ X-38 and Y-01 ~ Y-39)

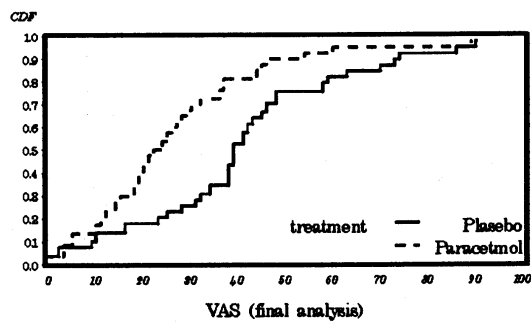
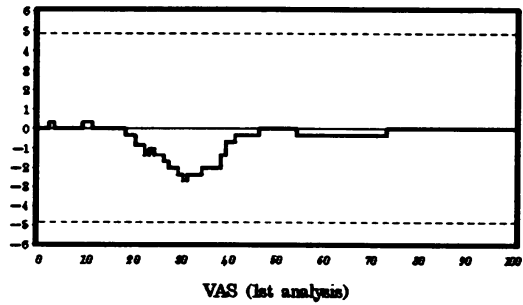


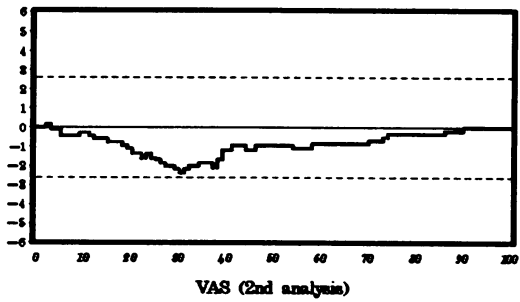
図 1: 各解析時点における経験分布関数



Stage 1 (patient: X-01 ~ X-12 and Y-01 ~ Y-12)



Stage 2 (patient: X-01 ~ X-24 and Y-01 ~ Y-25)



Stage 3 (patient: X-01 ~ X-38 and Y-01 ~ Y-39)

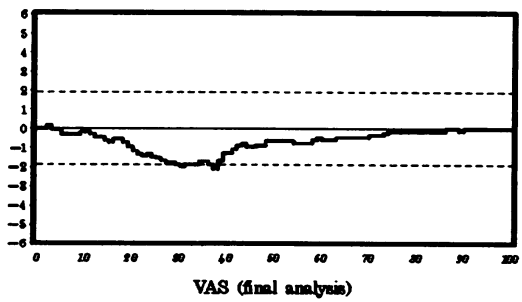
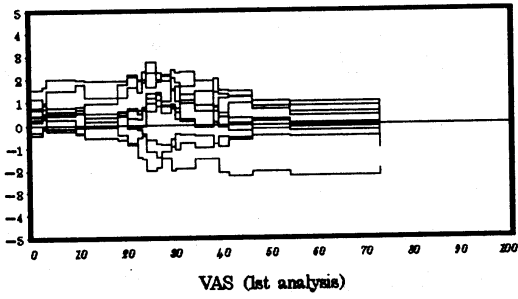
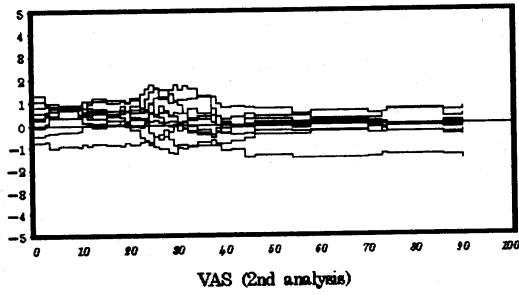


図 2: 各解析時点における経験分布関数の差

Stage 1 (patient: X-01 ~ X-13 and Y-01 ~ Y-13)



Stage 2 (patient: X-01 ~ X-26 and Y-01 ~ Y-26)



Stage 3 (patient: X-01 ~ X-38 and Y-01 ~ Y-39)

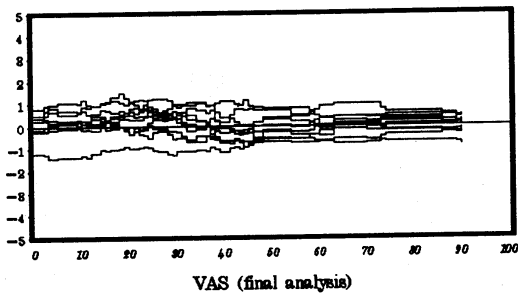


図 3: シミュレーションメソッドにより生成したサンプルパスの例

表 2: 各解析時点における棄却限界値と検定統計量の実現値

analysis stage	$\pi_k$	cum. $\pi_k$	critical value	test statistics
k=1	0.07%	0.07%	4.8384	2.6998
k=2	1.57%	2.64%	2.6094	2.3623
k=3	3.36%	5.00%	1.8950	2.0974

## 4 考察

本稿では、繰り返し信頼帯を応用して、コルモゴロフ・スミルノフ型の検定統計量による群逐次検定方式を導いた。また、分布関数を比較する検定として、Crámer-von Mienses 検定、Anderson-Darling 検定などがあるが、サンプルパスをシミュレーション的に発生させる手順であることから、これらについても同様の方法により、群逐次検定を構成できる。

臨床試験において、有効性を検証目的とする主たる解析方法としてプロトタイプケースにおいて示したようなコルモゴロフ・スミルノフ検定を採用する機会は少ないかもしれない。臨床試験のデータモニタリングにおいては、エンドポイントを含めた試験デザインの適切性を逐次吟味することも重要であることから、要約統計量や検定結果のみでなく、主要変数の分布関数の差を眺めることは有益な情報を提供するだろう。例えば、今回とりあげた例のような状況で、経験分関数が交叉するような場合など、試験デザインを見直すきっかけを与える具体的なガイドラインとしてモニタリング委員会にとって役に立つツールとなりえると考えられる。その他にも、例えば、対照薬と被験薬の同等性を示すことを目的とする試験において、位置母数だけでなく、分布関数そのものの類似性を議論する場合に、コルモゴロフ・スミルノフ検定が採用されることは考えられるし、その場合、適切な同等域を分布関数の差に対して設定することで本方法が適用できるだろう。また、コルモゴロフ・スミルノフ検定を途中で実施して有意差があるようなら同等性を示すことを諦めることに使用することに使用することもできるだろう。

## 付 録

### A 命題 1 の証明

まず、群逐次検定ではなく、検定を 1 回しか行わない単純な状況を考える。

#### A.1 表記法と基礎的な結果

独立同一分布に従う確率変数を  $X_1, X_2, \dots, X_n$  で表し、その分布関数を  $F$  とする。このとき、経験分布関数は、

$$F_n(t) \equiv \frac{1}{n} \sum_{i=1}^n 1[X_i \leq t],$$

で表される。

$F$  が連続分布のとき、経験分布関数について、以下のことが知られている。

$$\mathbb{G}_n(\cdot) = \sqrt{n}[F_n(\cdot) - F(\cdot)] \Rightarrow B^0(F(\cdot)), \text{ as } n \rightarrow \infty, \quad (16)$$

ここで  $B^0(\cdot)$  はブラウン橋を表す。  $B^0(F(\cdot))$  の平均値関数は0、共分散関数は、

$$\text{Cov}[B^0(F(s)), B^0(F(t))] = F(s)[1 - F(t)], \text{ for } s < t,$$

となる。

## A.2 シミュレーションメソッド

$$\mathbb{G}_n^L(t) \equiv \sqrt{n} \sum_{i=1}^n \frac{[1\{X_i \leq t\} - F_n(t)]Z_i}{n}, \quad (17)$$

とおく。ここで、  $Z_i$ , ( $i = 1, \dots, n$ ) は標準正規分布から得られる互いに独立な乱数である。このとき、

$\mathbb{G}_n^L$  は  $\{X_i\}$  の条件付で、  $\mathbb{G}_n$  と同一の極限に弱収束する。すなわち、

$$\mathbb{G}_n^L\{X_i\}(\cdot) \Rightarrow B^0(F(\cdot)),$$

である。

### 証明

Functional Central Limit Theorem (Pollard, Ch.10) より、  $\mathbb{G}_n^L(\cdot)\{X_i\}$  はガウシアンプロセスに弱収束する。また、  $\mathbb{G}_n^L(\cdot)\{X_i\}$  の平均値関数が0であることは明らか。また、共分散関数  $\text{Cov}[\mathbb{G}_n^L(s), \mathbb{G}_n^L(t)\{X_i\}]$  については、

$$\begin{aligned} & \text{Cov}[\mathbb{G}_n^L(s), \mathbb{G}_n^L(t)\{X_i\}] \\ &= \text{Cov} \left\{ \sqrt{n} \sum_{i=1}^n \left[ \frac{1\{X_i \leq s\} - F_n(s)}{n} \right] Z_i, \sqrt{n} \sum_{i=1}^n \left[ \frac{1\{X_i \leq t\} - F_n(t)}{n} \right] Z_i \right\} \\ &= \frac{1}{n} \mathbb{E} \left[ \left\{ \sum_{i=1}^n [1\{X_i \leq s\} - F_n(s)] Z_i \right\} \left\{ \sum_{i=1}^n [1\{X_i \leq t\} - F_n(t)] Z_i \right\} \right] \\ &= \frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^n [1\{X_i \leq s\} - F_n(s)][1\{X_i \leq t\} - F_n(t)] Z_i^2 \right] \\ &\quad + \frac{1}{n} \mathbb{E} \left[ \sum_{i \neq j} [1\{X_i \leq s\} - F_n(s)][1\{X_j \leq t\} - F_n(t)] Z_i Z_j \right] \\ &= \frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^n \{1\{X_i \leq s\} 1\{X_i \leq t\} - 1\{X_i \leq s\} F_n(t) - 1\{X_i \leq t\} F_n(s) + F_n(s) F_n(t)\} Z_i^2 \right] \\ &= \frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^n 1\{X_i \leq s\} 1\{X_i \leq t\} - n F_n(s) F_n(t) \right] \mathbb{E}(Z_i^2) \\ &\rightarrow F(s)[1 - F(t)] \text{ in prob. } (s < t) \end{aligned}$$

であり、これは、  $\text{Cov}[\mathbb{G}_n(s), \mathbb{G}_n(t)]$  の極限の共分散関数と同値となる。

### A.3 命題 1 の証明

上記の証明と同様、命題 1 についても、Functional Central Limit Theorem (Pollard, Ch.10) から、 $K$ -次元確率過程の各要素、 $G_n^L(T_k, t) | \{X_i, S_i\}(\cdot)$  は平均値関数 0 のガウシアンプロセスに弱収束する。また、共分散関数、 $\text{Cov}(G_n^L(T_k, s), G_n^L(T_l, t) | \{X_i, S_i\})$  は、 $s < t$  に対して、

$$\begin{aligned} & \text{Cov} \left\{ \sqrt{n} \left[ \frac{\sum_{i=1}^n 1\{S_i \leq T_k\} [1\{X_i \leq s\} - F_n(T_k, s)] Z_i}{\sum_{i=1}^n 1\{S_i \leq T_k\}} \right], \right. \\ & \quad \left. \sqrt{n} \left[ \frac{\sum_{i=1}^n 1\{S_i \leq T_l\} [1\{X_i \leq t\} - F_n(T_l, t)] Z_i}{\sum_{i=1}^n 1\{S_i \leq T_l\}} \right] \right\} \\ & \rightarrow \frac{n \sum_{i=1}^n 1\{S_i \leq T_k\} 1\{S_i \leq T_l\}}{\sum_{i=1}^n 1\{S_i \leq T_k\} \sum_{i=1}^n 1\{S_i \leq T_l\}} F(s)[1 - F(t)] \text{ in prob.} \end{aligned}$$

となり、 $G_n(T_k, \cdot)$  の極限の共分散関数に一致する。

### 参考文献

- [1] Armitage, P., McPherson, C. K. and Rowe, B. C. (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society A132*, 235–244.
- [2] Akhtar-Danesh, N and Appleton, D. R. (2000). Using an antedependence test to analyse post-operative pain measurements. *Statistics in Medicine* 19, 1889–1899.
- [3] Hu, X. J and Lagakos, S. W. (1999a). Interim analyses using repeated confidence bands. *Biometrika* 86, 517–529.
- [4] Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* 80, 557–572.
- [5] Pollard, D. (1990). *Empirical Processes: Theory and Applications*, Regional Conference Series in Probability and Statistics 2. Hayward, CA: Institute of Mathematical Statistics.
- [6] Schervish, M. J. (1984). Multivariate normal probabilities with error bound. *Applied Statistics* 33, 81–94.
- [7] Slud, E. and Wei, L. J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *Journal of the American Statistical Association* 77, 862–868.
- [8] Tsiatis, A. A., Rosner, G. L. and Mehta, C. R. (1984). Exact confidence intervals following a group sequential test. *Biometrics* 40, 797–803.