Asymptotic normality for sums along data-dependent sampling schemes

Takuhisa Shikimi

Abstract

Let $X_{(1)} = \{X_{1,t^1}, t^1 \in \mathbf{N}\},..., X_{(d)} = \{X_{d,t^d}, t^d \in \mathbf{N}\}$ be independent sequences of i.i.d. real-valued random variables and let $S_t = S_{1,t^1} + \cdots + S_{d,t^d}$, where $t = (t^1,...,t^d)$ and $S_{i,t^i} = \sum_{s^i \leqslant t^i}((X_{i,s^i} - \mu_i)/\sigma_i)$, $i = 1,...,d$. A sequential sampling plan determines the way of taking one observation from one of the processes $X_{(1)},...,X_{(d)}$, according to the previous sampled data. We show that the random sum of observations under any sequential sampling scheme is asymptotically normal. An easy application of the normality yields the classical result of sequential interval estimation of means of two polulations.

Key words: asymptotic normality; sequential sampling plan; sequential interval estimation.

1 Introduction

This note intends to prove asymptotic normality for random sums of random variables in which the summands are selected in a predictable manner from two or more independent sequences of i.i.d. real-valued random variables. Let $X_{(1)} = \{X_{1,t^1}, t^1 \in \mathbf{N}\},..., X_{(d)} = \{X_{d,t^d}, t^d \in \mathbf{N}\}$ be independent sequences of i.i.d random variables with $\mathbf{E}X_{i,1} = \mu_i$, $\mathbf{V}X_{i,1} = \sigma_i^2 \in (0, \infty)$, $i = 1,...,d$. For $t = (t^1, ...,t^d)$, let $S_t = S_{1,t^1} + \cdots + S_{d,t^d}$, where $S_{i,t^i} = \sum_{s^i \leqslant t^i}((X_{i,s^i} - \mu_i)/\sigma_i)$ $(i = 1,...,d)$. A sequential sampling scheme determines which population to be observed based on the previous observations, and therefore is represented by a sequence of $\mathbf{N}^d$-valued stopping points (γ_n) satisfying a predictable condition. Asymptotic normality of S_{γ_n} is shown by a straightforward application of a martingale central limit theorem (Brown(1971), McLeish(1974), Chow and Teicher (1988) or Billingsley (1995)). This result will be applied to sequential estimation of the defference of means of two populations.

2 Notation

The set $\mathbf{N}$ is the set of positive integers, $\mathbf{N}^*$ the set $\mathbf{N} \cup \{0\}$ and d is a positive integer. Throughout, the set $\mathbf{N}^{*d}$ will be denoted by $\mathbf{I}^*$, the set $\mathbf{N}^{*d} - \{0\}$ by $\mathbf{I}$, where $0 = (0,...,0)$. Ele-

ments of $\mathbf{I}^*$ are denoted by the letters $s, t, \ldots$ If $s = (s^1, \ldots, s^d)$ and $t = (t^1, \ldots, t^d)$ are elements in $\mathbf{I}^*$ and $s^i \leqslant t^i$ for all $i = 1, \ldots, d$, then we say that s is less than or equal to t (or t is greater than or equal to s). This relation between s and t is denoted by $s \leqslant t$, and this relation forms an partial order in the set $\mathbf{I}^*$. If $s \leqslant t$ and $s \neq t$, we write $s < t$. For $t = (t^1, \ldots, t^d) \in \mathbf{I}^*$, we set $|t| = t^1 + \cdots + t^d$. The direct successors of $t \in \mathbf{I}^*$ are the elements $u \in \mathbf{I}^*$ such that $t < u$ and $|u - t| = 1$. The set of direct successors of t is denoted by $\mathbf{D}_t$. The direct predecessors of $t \in \mathbf{I}$ are the elements $s \in \mathbf{I}^*$ such that $s < t$ and $|t - s| = 1$. If t_n are elements in $\mathbf{I}$, we will write $t_n \to \bowtie$ to express that $|t_n| \to \infty$, and we use the symbol $\bar{\mathbf{I}}^*$ to denote the set $\mathbf{I} \cup \{\bowtie\}$. The order structure of $\mathbf{I}^*$ is extended to $\bar{\mathbf{I}}^*$ by setting $t < \bowtie$. We agree by convention that $|\bowtie| = \infty$.

Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probalility space. A filtration $\{\mathscr{F}_t, t \in \mathbf{I}^*\}$ is a family of sub-σ-algebras of $\mathscr{F}$ such that $\mathscr{F}_s \subset \mathscr{F}_t$ for $s < t$. A function $T : \Omega \to \bar{\mathbf{I}}^*$ is called a stopping point if $\{T = t\} \in \mathscr{F}_t$ for all $t \in \mathbf{I}^*$. For a stopping point T relative to $\{\mathscr{F}_t, t \in \mathbf{I}^*\}$, $\mathscr{F}_T$ is the class of subsets F in $\mathscr{F}_{\bowtie} = \sigma(\mathscr{F}_t, t \in \mathbf{I}^*)$ such that $F \cap \{T = t\} \in \mathscr{F}_t$ for all $t \in \mathbf{I}^*$. A sequence of stopping points $\gamma = \{\gamma_n, n \in \mathbf{N}^*\}$ is said to be a predictable increasing path if

(i) $\gamma_0 = 0$ a.s.,

(ii) $\gamma_{n+1} \in \mathbf{D}_{\gamma_n}$ a.s.,

(iii) γ_{n+1} is $\mathscr{F}_{\gamma_n}$-measurable.

A predictable increasing path can be thought of as a sampling strategy. Namely, it provides at each stage of the observation a way of deciding which population to be observed, depending on the previous data. A predictable increasing path naturally arises in sequential sampling or multiarmed bandits problems (Cairoli and Dalang (1996)). Clearly, in dimension 1 $(d = 1)$, the concept of a predictable increasing path is of little interest since in this case there is only one predictable increasing path, namely, the deterministic increasing path defined by $\gamma_n = n$ for all $n \in \mathbf{N}^*$. For all predictable increasing paths γ, we agree by convention that $\gamma_\infty = \bowtie$. We will write $\mathscr{F}_n^\gamma$ instead of $\mathscr{F}_{\gamma_n}$. For precesses X indexed by $\mathbf{I}$ and predictable increasing paths $\gamma = \{\gamma_n\}$, X_n^γ denotes X_{γ_n}. If $\{X_t, t \in \mathbf{I}\}$ is a martingale relative to $\{\mathscr{F}_t, t \in \mathbf{I}\}$ and γ is a predictable increasing path, then $\{X_n^\gamma, t \in \mathbf{N}\}$ forms a martingale relative to the filtration $\{\mathscr{F}_n^\gamma, n \in \mathbf{N}\}$.

3　The result

Let $X_{(i)} = \{X_{i, t^i}, t^i \in \mathbf{N}\}$ $(i = 1, \ldots, d)$ be a sequence of independent and identically distributed real-valued random variables. We assume that $X_{(1)}, \ldots, X_{(d)}$ are independent sequences. Further symbols and assumptions are as follows: for $i = 1, \ldots, d$,

$$\mu_i = \mathbf{E} X_{i,1}, \quad \sigma_i^2 = \mathbf{V} X_{i,1} \in (0, \infty), \quad S_{i, t^i} = \sum_{s^i \leqslant t^i} \left(\frac{X_{i, s^i} - \mu_i}{\sigma_i} \right),$$

$$S_{i,0} = 0, \quad \mathscr{F}_{i, t^i} = \sigma(X_{i, s^i}, s^i \leqslant t^i), \quad \mathscr{F}_{i,0} = \{\varnothing, \Omega\}.$$

For $t = (t^1, \ldots, t^d) \in \mathbf{I}^*$, let

$$S_t = S_{1,t^1} + \cdots + S_{d,t^d}, \tag{3.1}$$

$$\mathscr{F}_t = \mathscr{F}_{1,t^1} \vee \cdots \vee \mathscr{F}_{d,t^d}. \tag{3.2}$$

Let $\{\mathscr{F}_t, t \in \mathbf{I}^*\}$ denote the filtration defined by (3.2) and let γ denote *any* predictable increasing path relative to $\{\mathscr{F}_t, t \in \mathbf{I}^*\}$. Clearly $S = \{S_t, t \in \mathbf{I}\}$ is a martingale (note that the sum S_t has already been normalized), and hence so is $\{S_j^\gamma, j \in \mathbf{N}\}$ as was mentioned in the preceeding section. Define

$$Y_j^\gamma = S_j^\gamma - S_{j-1}^\gamma \quad (j \geq 1), \tag{3.3}$$

$$Z_{nj}^\gamma = Y_j^\gamma / \sqrt{n} \quad (j = 1, \ldots, n). \tag{3.4}$$

It is evident that the triangular array $\{Z_{nj}^\gamma, \mathscr{F}_j^\gamma, j = 1, \ldots, n\}, \ n \in \mathbf{N}\}$ is a martingale difference array. Thus applying the martingale central limit theorem yields the following result.

Theorem 3.1. *Let $S = \{S_t, t \in \mathbf{I}\}$ be the process defined by (3.1) and γ any predictable increasing path. Then we have*

$$\frac{S_n^\gamma}{\sqrt{n}} \xrightarrow{\mathscr{D}} N(0,1) \quad (1 \leq n \to \infty), \tag{3.5}$$

where $\xrightarrow{\mathscr{D}}$ denotes the convergence in distribution.

proof. By the martingale central limit theorem, it suffices to show that as $n \to \infty$

$$\sum_{j=1}^n \mathbf{E}[(Z_{nj}^\gamma)^2 \mid \mathscr{F}_{j-1}^\gamma] \xrightarrow{P} 1, \tag{3.6}$$

$$\sum_{j=1}^n \mathbf{E}[(Z_{nj}^\gamma)^2 1_{\{|Z_{nj}^\gamma| \geq \varepsilon\}}] \to 0, \tag{3.7}$$

where $\xrightarrow{P}$ denotes the convergence in probability and 1_A stands for the indicator function of the set $A \in \mathscr{F}$. The latter condition is called the Lindeberg condition. We assume without loss of generality that $\mu_i = 0$ and $\sigma_i = 1$, $i = 1, \ldots, d$. We shall denote by $\mathbf{P}_t$ the set of direct predecessors of $t \in \mathbf{I}$. For $j \geq 1$, we have

$$\mathbf{E}[(Y_j^\gamma)^2 \mid \mathscr{F}_{j-1}^\gamma] = \sum_{t: |t|=j} \int_{\{\gamma_j = t\}} \mathbf{E}[(S_j^\gamma - S_{j-1}^\gamma)^2 \mid \mathscr{F}_{j-1}^\gamma] d\mathbf{P}$$

$$= \sum_{t: |t|=j} \sum_{s \in \mathbf{P}_t} \int_{\{\gamma_j = t\} \cap \{\gamma_{j-1} = s\}} \mathbf{E}[(S_j^\gamma - S_{j-1}^\gamma)^2 \mid \mathscr{F}_s] d\mathbf{P}$$

It follows from the predictability of γ that $\{\gamma_j = t\} \cap \{\gamma_{j-1} = s\} \in \mathscr{F}_s$. We see therefore the last display is equal to

$$\sum_{t: |t|=j} \sum_{s \in \mathbf{P}_t} \int_{\{\gamma_j=t\} \cap \{\gamma_{j-1}=s\}} (S_j^\gamma - S_{j-1}^\gamma)^2 d\mathbf{P}$$

$$= \sum_{t: |t|=j} \sum_{s \in \mathbf{P}_t} \int_{\{\gamma_j=t\} \cap \{\gamma_{j-1}=s\}} (S_t - S_s)^2 d\mathbf{P}$$

$$= \sum_{t: |t|=j} \sum_{s \in \mathbf{P}_t} \mathbf{P}(\gamma_j=t, \gamma_{j-1}=s) \mathbf{E} X_{1,1}^2 = 1$$

The last equality follows from the independence of $1_{\{\gamma_j=t, \gamma_{j-1}=s\}}$ and $(S_t-S_s)^2$ (note that $\{\gamma_j=t\}$ $\cap \{\gamma_{j-1}=s\} \in \mathscr{F}_s$). Consequently, we obtain that

$$\sum_{j=1}^n \mathbf{E}[(Z_{nj}^\gamma)^2 | \mathscr{F}_{j-1}^\gamma] = \frac{1}{n} \sum_{j=1}^n \mathbf{E}[(Y_j^\gamma)^2 | \mathscr{F}_{j-1}^\gamma] = 1.$$

The proof of (3.7) will be done similarly. Since for each $\varepsilon > 0$

$$\mathbf{E}[(Y_j^\gamma)^2 1_{\{|Y_j^\gamma| \geqslant \varepsilon \sqrt{n}\}}] \to \int (S_j^\gamma - S_{j-1}^\gamma)^2 1_{\{|S_j^\gamma - S_{j-1}^\gamma| \geqslant \varepsilon \sqrt{n}\}} d\mathbf{P}$$

$$= \sum_{t: |t|=j} \sum_{s \in \mathbf{P}_t} \int_{\{\gamma_j=t\} \cap \{\gamma_{j-1}=s\}} (S_t-S_s)^2 1_{\{|S_t-S_s| \geqslant \varepsilon \sqrt{n}\}} d\mathbf{P}$$

$$= \sum_{t: |t|=j} \sum_{s \in \mathbf{P}_t} \mathbf{P}\{\gamma_j=t\} \cap \{\gamma_{j-1}=s\} \int (S_t-S_s)^2 1_{\{|S_t-S_s| \geqslant \varepsilon \sqrt{n}\}} d\mathbf{P}$$

Since the integral of the last display converges to 0 as n $\to \infty$,

$$\sum_{j=1}^n \mathbf{E}\left[(Z_{nj}^\gamma)^2 1_{\{|Z_{nj}^\gamma| \geqslant \varepsilon\}}\right]$$

$$= \frac{1}{n} \sum_{j=1}^n \sum_{t: |t|=j} \sum_{s \in \mathbf{P}_t} \mathbf{P}(\{\gamma_j=t\} \cap \{\gamma_{j-1}=s\}) \int (S_t-S_s)^2 1_{\{|S_t-S_s| \geqslant \varepsilon \sqrt{n}\}} d\mathbf{P} \to 0 \quad (n \to \infty).$$

This completes the proof of the theorem.

4 An application to sequential estimation

In this section we will consider an application of the result in the previous section to sequential estimation for the difference of means of two independent populations, with the specified sampling plan introduced by Robbins, Simons and Starr (1967). Before proceeding to discussing the application, we need to prepare a simple extension of Anscombe's theorem to martingale cases. Anscombe's theorem asserts that randomly stopped sums are asymptotically normally distributed. The central idea used in the proof of Anscombe's theorem is the notion of uniform continuity in probability. A sequence $\{Y_n, n \in \mathbf{N}\}$ of random variables is said to be uniformly continuous in probability iff for all $\varepsilon > 0$, there exists an $\delta > 0$ for which

$$\sup_n P(\max_{\{0 \leqslant k \leqslant n\delta\}} |Y_{n+k} - Y_n| \geqslant \varepsilon) < \varepsilon. \tag{4.1}$$

Anscombe's theorem is obtained by using the following lemma, which will be also applied to prove its generalization to the martingale case stated below. The proof of the lemma can be found in Ghosh, Mukhopadhyay and Sen (1997).

Lemma 4.1. *Suppose that* $\{Y_n, n \in \mathbf{N}\}$ *is uniformly continuous in probability. Let* $\tau_b(b > 0)$ *be a positive integer-valued random variable such that* τ_b/b *converges in probability to a constant* $c \in (0, \infty)$. *If* Y_n *converges in distribution to a random variable* Y *as* $n \to \infty$, *then* Y_{τ_b} *converges in distribution to* Y *as* $b \to \infty$.

An adapted process $\{Y_n\}$ is called an L^2-martingale if it is a martingale and $\mathbf{E}Y_n^2 < \infty$ for each n.

Theorem 4.2. *Suppose that* $\{Y_n, \mathscr{F}_n, n \in \mathbf{N}\}$ *is an* L^2-martingale with mean 0. *Let* $Z_n = Y_n - Y_{n-1}(n \geqslant 2)$, $Z_1 = Y_1$ *and assume that* $\mathbf{E}Z_n^2$ *is bounded. Let* $\tau_b(b > 0)$ *be a positive integer-valued random variable such that* τ_b/b *converges in probability to a constant* $c \in (0, \infty)$. *If* $Y_n/\sqrt{n}$ *converges in distribution to* $N(0,1)$ *as* $n \to \infty$, *then as* $b \to \infty$

$$\frac{Y_{\tau_b}}{\sqrt{\tau_b}} \} \xrightarrow{\mathscr{D}} N(0,1), \tag{4.2}$$

$$\frac{Y_{\tau_b}}{\sqrt{bc}} \} \xrightarrow{\mathscr{D}} N(0,1), \tag{4.3}$$

Proof. It suffices to prove (4.2) because $Y_{\tau_b}/\sqrt{bc} = Y_{\tau_b}(\tau_b/bc)^{1/2}/\sqrt{\tau_b}$. In proving that $Y_{\tau_b}/\sqrt{\tau_b}$ converges in distribution to $N(0,1)$, by virtue of Lemma 4.1, we need only verify the uniform continuity in probability of $Y_n/\sqrt{n}$. Let us begin with showing that $\{Y_n/\sqrt{n}, n \in \mathbf{N}\}$ is uniformly integrable. Since $\{Z_n\}$ are uncorrelated and $\mathbf{E}Z_n = 0 (n \in \mathbf{N})$, it follows from the boundedness of $\mathbf{E}Z_n^2$ that

$$\mathbf{E}Y_n^2 = \mathbf{E}\left[\left(\sum_{j=1}^{n} Z_j \right)^2 \right] \leqslant nM \quad \text{(say)},$$

whence $\sup_n \mathbf{E}[(n^{-1/2}Y_n)^2] < \infty$. It follows therefore that $Y_n/\sqrt{n}$ is uniformly integralbe.

Writing $\rho(\delta) = 1 - (1/(1+\delta))^{1/2}$ for $\delta > 0$, we have

$$\sup_{n \geqslant 1} \mathbf{P}\left[\rho(\delta) \left| \frac{Y_n}{\sqrt{n}} \right| > \frac{\varepsilon}{2} \right] \to 0 \quad \text{as} \quad \delta \to 0. \tag{4.4}$$

For $n \geqslant 1$, $k \geqslant 1$, we have

$$\left| \frac{Y_{n+k}}{\sqrt{n+k}} - \frac{Y_n}{\sqrt{n}} \right| = \left| \frac{Y_{n+k}}{\sqrt{n+k}} - \frac{Y_n}{\sqrt{n+k}} + \frac{Y_n}{\sqrt{n+k}} - \frac{Y_n}{\sqrt{n}} \right|$$

$$\leqslant \frac{1}{\sqrt{n+k}} \left| Y_{n+k} - Y_n \right| + \left| \left(\frac{n}{n+k} \right)^{1/2} - 1 \right| \left| \frac{Y_n}{\sqrt{n}} \right|$$

$$\leqslant \frac{1}{\sqrt{n}} \left| Y_{n+k} - Y_n \right| + \left[1 - \left(\frac{n}{n+k} \right)^{1/2} \right] \left| \frac{Y_n}{\sqrt{n}} \right|.$$

If $\delta > 0$ and $k \leqslant n\delta$,

$$\left[1 - \left(\frac{n}{n+k} \right)^{1/2} \right] \left| \frac{Y_n}{\sqrt{n}} \right| \leqslant \left[1 - (1/(1+\delta))^{1/2} \right] \left| \frac{Y_n}{\sqrt{n}} \right|.$$

It follows thereby that

$$\sup_{n \geqslant 1} \mathbf{P} \left[\max_{0 \leqslant k \leqslant n\delta} \left| \frac{Y_{n+k}}{\sqrt{n+k}} - \frac{Y_n}{\sqrt{n}} \right| > \varepsilon \right]$$

$$\leqslant \sup_{n \geqslant 1} \mathbf{P} \left[\max_{0 \leqslant k \leqslant n\delta} \left| Y_{n+k} - Y_n \right| > \varepsilon\sqrt{n}/2 \right] + \sup_{n \geqslant 1} \mathbf{P} \left[\rho(\delta) \left| \frac{Y_n}{\sqrt{n}} \right| > \varepsilon/2 \right]. \tag{4.5}$$

From (4.4) and (4.5) it remains only to show that the first term of (4.5) converges to 0 as $\delta \to 0$. It should be noted that for each $n \in \mathbf{N}$, $(Y_{n+k} - Y_n, \mathscr{F}_{n+k}, k \in \mathbf{N})$ is a martingale. Hence, using the maximal inequality for martingales, we obtain

$$\mathbf{P} \left[\max_{0 \leqslant k \leqslant n\delta} \left| Y_{n+k} - Y_n \right| > \varepsilon\sqrt{n}/2 \right] = \mathbf{P} \left[\max_{0 \leqslant k \leqslant [n\delta]} \left| Y_{n+k} - Y_n \right| > \varepsilon\sqrt{n}/2 \right]$$

$$\leqslant \frac{4}{n\varepsilon^2} \mathbf{E} \left| Y_{n+[n\delta]} - Y_n \right|^2 = \frac{4}{n\varepsilon^2} \mathbf{E} \left| \sum_{j=n+1}^{n+[n\delta]} Z_j \right|^2$$

$$\leqslant \frac{4}{n\varepsilon^2} [n\delta] M \leqslant \frac{4M\delta}{\varepsilon^2},$$

whence

$$\sup_{n \geqslant 1} \mathbf{P} \left[\max_{0 \leqslant k \leqslant n\delta} \left| Y_{n+k} - Y_n \right| > \varepsilon\sqrt{n}/2 \right] \leqslant \frac{4M\delta}{\varepsilon^2} \to 0 \qquad (\delta \to 0).$$

This completes the proof of the theorem.

By this theorem, if $\{S_n^r, n \in \mathbf{N}\}$ is the process defined by (3.1), then

$$\frac{S_{\tau_b}^r}{\sqrt{\tau_b}} \xrightarrow{\mathscr{D}} N(0,1),$$

$$\frac{S_{\tau_b}^r}{\sqrt{bc}} \xrightarrow{\mathscr{D}} N(0,1),$$

provided that τ_b/b converges in probability to a positive constant c.

Let us introduce two independent sequences of i.i.d. real-valued random variables, say $X_{(1)} = \{X_{1,t^1}, t^1 \in \mathbf{N}\}$ and $X_{(2)} = \{X_{2,t^2}, t^2 \in \mathbf{N}\}$, where $\mu_i \mathbf{E} X_{i,1}$ and $0 < \sigma_i^2 = \mathbf{V} X_{i,1} (i=1,2)$ are unknown. The problem is interval estimation of $\mu = \mu_1 - \mu_2$. Set $\mathscr{F}_t = \mathscr{F}_{1,t^1} \vee \mathscr{F}_{2,t^2}$, for $t = (t^1, t^2)$, where $\mathscr{F}_{i,t^i} = \sigma(X_{i,s^i}, s^i \leqslant t^i)$ and $\mathscr{F}_{i,0} = \{\varnothing, \Omega\}$. Define, for $t^i \geqslant 2$,

$$U_{t^1}^2 = \frac{1}{t^1 - 1} \sum_{s^1 \leqslant t^1} (X_{1,s^1} - \bar{X}_{1,t^1})^2,$$

$$V_{t^2}^2 = \frac{1}{t^2 - 1} \sum_{s^2 \leqslant t^2} (X_{2,s^2} - \bar{X}_{2,t^2})^2,$$

where $\bar{X}_{i,t^i} = (1/t^i) \sum_{s^i \leqslant t^i} X_{i,s^i} (i=1,2)$. Let γ be a predictable increasing path with $d=2$. We introduce a stopping rule relative to $\{\mathscr{F}_n^{\gamma}, n \in \mathbf{N}\}$, which is one of the stopping rules proposed by Robbins, Simons and Starr (1967): for $b > 0$

$$\tau_b = \inf\{n \geqslant 2m : n \geqslant b(U_{\gamma_n^1} + V_{\gamma_n^2})^2\}. \tag{4.6}$$

where $m \geqslant 2$ is the initial sample size of observations on X^1 and X^2. Defining three more or less equivalent stopping rules including (4.6), Robbins, Simons and Starr (1967) studied the sequential interval estimation of the difference of the means of two populations, where the variances are unknown. More precisely, with a sampling rule and several stopping rules Robbins, Simons and Starr (1967) found a approximate confidence interval for μ of fixed width and of preassigned coverage probality. Their procedure consists of (i) a sampling scheme that tells us at each stage whether the next observation, if needed, would be an $X_{(1)}$ or $X_{(2)}$ and (ii) a stopping rule, e.g., (4.6). Their sampling strategies is as follows: we take $m (\geqslant 2)$ observations on $X_{(1)}$ and $X_{(2)}$ to begin with. Then, inductively define γ by

$$\gamma_n = \gamma_{n-1} + e_1 I\left(\frac{\gamma_{n-1}^1}{\gamma_{n-1}^2} \leqslant \frac{U_{\gamma_{n-1}^1}}{V_{\gamma_{n-1}^2}}\right) + e_2 I\left(\frac{\gamma_{n-1}^1}{\gamma_{n-1}^2} > \frac{U_{\gamma_{n-1}^1}}{V_{\gamma_{n-1}^2}}\right), \tag{4.7}$$

where $e_1 = (1,0)'$ and $e_2 = (0,1)'$. Hereafter γ denotes the predictable increasing path defined (4.7). With the sampling scheme (4.7) and the stopping rule (4.6), we will show that $\bar{X}_{1,\gamma_{\tau_b}^1} - \bar{X}_{2,\gamma_{\tau_b}^2} - \mu$ has asymptotically normal distribution with mean 0 and variance b^{-1} for large b. This consequence implies the result of Robbins, Simons and Starr (1967).

For γ and τ_b, they showed that as $b \to \infty$

$$\frac{\gamma_{\tau_b}^1}{\gamma_{\tau_b}^2} \to \frac{\sigma_1}{\sigma_2} \quad \text{a.s.,} \tag{4.8}$$

and

$$\frac{\tau_b}{b(\sigma_1+\sigma_2)^2} \to 1 \quad \text{a.s.} \tag{4.9}$$

Set for $t=(t^1,t^2)$

$$D_t = \sum_{s^1 \leqslant t^1}\left(\frac{X_{1,s^1}-\mu_1}{\sigma_1}\right) - \sum_{s^2 \leqslant t^2}\left(\frac{X_{2,s^2}-\mu_2}{\sigma_2}\right).$$

Using Theorem 3.1, (4.9) and Theorem 4.2, we have

$$\sqrt{\tau_b}\left[\frac{\gamma^1_{\tau_b}}{\sigma_1\tau_b}\left(\bar{X}_{1,\gamma^1_{\tau_b}}-\mu_1\right) - \frac{\gamma^2_{\tau_b}}{\sigma_2\tau_b}\left(\bar{X}_{2,\gamma^2_{\tau_b}}-\mu_2\right)\right] = \frac{D_{\tau_b}}{\sqrt{\tau_b}} \overset{\mathscr{D}}{\to} N(0,1), \tag{4.10}$$

and so by (4.9) and (4.10) we have

$$\sqrt{b}(\sigma_1+\sigma_2)\left[\frac{\gamma^1_{\tau_b}}{\sigma_1\tau_b}\left(\bar{X}_{1,\gamma^1_{\tau_b}}-\mu_1\right) - \frac{\gamma^2_{\tau_b}}{\sigma_2\tau_b}\left(\bar{X}_{2,\gamma^2_{\tau_b}}-\mu_2\right)\right]$$

$$= \frac{\sqrt{b}(\sigma_1+\sigma_2)}{\sqrt{\tau_b}} \cdot \frac{D_{\tau_b}}{\sqrt{\tau_b}} \overset{\mathscr{D}}{\to} N(0,1), \tag{4.11}$$

By (4.8) and (4.9) we have

$$\frac{\gamma^1_{\tau_b}}{b} = \frac{\tau_b}{b(\sigma_1+\sigma_2)^2} \cdot \frac{\gamma^1_{\tau_b}(\sigma_1+\sigma_2)^2}{\tau_b} \to \sigma_1(\sigma_1+\sigma_2), \quad b \to \infty.$$

Likewise, as $b \to \infty$

$$\frac{\gamma^2_{\tau_b}}{b} \to \sigma_2(\sigma_1+\sigma_2).$$

Therefore applying the central limit theorem and Anscombe's theorem yield

$$\sqrt{b}(\bar{X}_{1,\gamma^1_{\tau_b}}-\mu_1) = \left(\frac{b}{\gamma^1_{\tau_b}}\right)^{1/2} \cdot \frac{1}{\sqrt{\tau_b}}\sum_{s^1 \leqslant \gamma^1_{\tau_b}}(X_{1,s^1}-\mu_1) \overset{\mathscr{D}}{\to} N(0,\sigma_1^2(\sigma_1+\sigma_2)), \tag{4.12}$$

$$\sqrt{b}(\bar{X}_{2,\gamma^2_{\tau_b}}-\mu_1) = \left(\frac{b}{\gamma^2_{\tau_b}}\right)^{1/2} \cdot \frac{1}{\sqrt{\tau_b}}\sum_{s^1 \leqslant \gamma^2_{\tau_b}}(X_{1,s^2}-\mu_1) \overset{\mathscr{D}}{\to} N(0,\sigma_2^2(\sigma_1+\sigma_2)). \tag{4.13}$$

Further, note that by (4.8) as $b \to \infty$

$$\frac{(\sigma_1+\sigma_2)\gamma^1_{\tau_b}}{\sigma_i\tau_b} \to 1. \tag{4.14}$$

Since by (4.12) and (4.13) $\sqrt{b}(\bar{X}_{i,\gamma^i_{\tau_b}}-\mu_1)$ converges in distribution, it follows from (4.14) that as $b \to \infty$

$$\sqrt{b}\,(\sigma_1+\sigma_2)\left[\frac{\gamma^1_{\tau_b}}{\sigma_1\tau_b}\left(\bar{X}_{1,\gamma^1_{\tau_b}}-\mu_1\right)-\frac{\gamma^2_{\tau_b}}{\sigma_2\tau_b}\left(\bar{X}_{2,\gamma^2_{\tau_b}}-\mu_2\right)\right]-\sqrt{b}\,(\bar{X}_{1,\gamma^1_{\tau_b}}-\bar{X}_{2,\gamma^2_{\tau_b}}-\mu)$$

$$=\left[\frac{(\sigma_1+\sigma_2)\gamma^1_{\tau_b}}{\sigma_1\tau_b}-1\right]\sqrt{b}\,(\bar{X}_{1,\gamma^1_{\tau_b}}-\mu_1)-\left[\frac{(\sigma_1+\sigma_2)\gamma^2_{\tau_b}}{\sigma_1\tau_b}-1\right]\sqrt{b}\,(\bar{X}_{2,\gamma^2_{\tau_b}}-\mu_1)\xrightarrow{P}0.$$

Consequently, by (4.11) we obtain the following asymptotic normality:

$$\sqrt{b}\,(\bar{X}_{1,\gamma^1_{\tau_b}}-\bar{X}_{2,\gamma^2_{\tau_b}}-\mu)\xrightarrow{\mathscr{D}}N(0,1)\ (b\to\infty).$$

This implies the result of Robbins, Simons and Starr (1967).

References

[1]Billingsley, P. (1995). *Probability and Measure*, 3rd ed., J. Wiley & Sons, New York.

[2]Brown, B. M. (1971). Martingale central limit theorems, *Ann. Math. Statist.*, **42**,59-66.

[3]Cairoli, R. and Dalang, R.C. (1996). *Sequential Stochastic Optimization*, J. Wiley & Sons, New York.

[4]Chow, Y. S. and Teicher, H. (1988). *Probability Theory: Independence, Interchangability, Martingales*, 2nd ed., Springer-Verlag, New York.

[5]item Ghosh, M., Mukhopadhyay, N., and Sen, P.K (1997). *Sequential Estimation*, J. Wiley & Sons, New York.

[6]McLeish, D. L. (1974). Dependent central limit theorems and invariance principles, *Ann. Prob*, **2**,620-628.

[7]Robbins, H., Simons, G., and Starr, N. (1967). A sequential analogue of the Behrens-Fisher problem, *Ann. Math. Statist.*, **38**,1384-1391.