COMMUN. STATIST.-THEOR. METH., A9(15), 1625-1639 (1980)

SEQUENTIAL PROCEDURES IN IDENTIFICATION

J. K. Ghosh
Indian Statistical Institute

N. Mukhopadhyay
Oklahoma State University

Key Words and Phrases: tests for composite hypothesis; invariance; sequential probability ratio test; maximal invariant; discrimination; simulation.

ABSTRACT

In this paper we study the problem of identifying a population with one of the two populations, with an aim to control both types of errors. We assume that the populations are normal with unknown means, but with unit variance. We have cited examples from anthropological studies where our formulation of the problem fits in quite nicely. We observe that SPRT's based on the maximal invariant may not terminate with probability one. Simulation studies reported here show a substantial saving in the average number of samples compared to the best invariant fixed sample test.

1. INTRODUCTION

Suppose there are three unknown populations Π_0, Π_1 and Π_2, where it is known that Π_0 is either Π_1 or Π_2. Based on information obtained from samples on these populations, we wish to identify Π_0 with Π_1 or Π_2 by controlling both types of errors committed in the decision-making.

Let X, Y and Z with suffixes be the random variables associated with Π_0, Π_1 and Π_2 respectively, where each is normally distributed with common variance σ^2 and (unknown) means μ, μ_1 and μ_2 respectively. We assume that X, Y and Z are mutually independent, and independent observations are available from these populations. Now, we want to decide whether $\mu = \mu_1$ or $\mu = \mu_2$, having both errors at preassigned levels. Towards this, we assume $\sigma^2 = 1$ and

$$\mu_1 \geq \mu_2 + \delta \tag{A}$$

where δ is a known positive constant, that is, the two populations Π_1 and Π_2 are separated by δ units. One is referred to section 2.4 for some discussions regarding our assumptions. Since any reasonable procedure of identification in the case μ_1 $\mu_2 + \delta$ is expected to perform in a still better way for $\mu_1 > \mu_2 + \delta$, we confine our attention to procedures for identifying Π_0 with Π_1 or Π_2 for the configuration $\mu_1 = \mu_2 + \delta$, which is referred to as the least favorable configuration (LFC).

Since we wish to control both the errors with savings in the sample sizes, we take resort to sequential sampling. We shall consider mainly the case where a sample of fixed size is given from Π_0 and no further sampling from it is feasible although unlimited sequential sampling is permitted from Π_1, Π_2. Such a restriction leads to some novel theoretical points, e.g. Hall, Wijsman and Ghosh (1965) SPRT's based on the maximal invariant do not terminate with probability one. Moreover, it has some practical interest where the sample from Π_0 refers to, say, anthropological specimens at a site where excavation has stopped and Π_1, Π_2 refer to sites where excavation is currently going on. For another application, one may wish to identify the Todas (Π_0) as originating from the Nairs of Malabar (Π_1) or Nairs of Nambutiris (Π_2). One may refer to Rivers (1906, p. 708). Since the total number of living Todas (Π_0) is very small (about seventy families) whereas the other two communities considered are fairly large, we have a situation that is close to our present formulation of the problem. In this second example, the variables X, Y or Z may be any one of the variables listed on p. 708 of Rivers (1906), e.g. stature, nasal length or a suitable linear combination thereof. (We may use an estimate of σ from the Toda data for the true value.)

We will assume that both Π_1 and Π_2 will be sampled at each stage until we stop sampling. Of course it would be more efficient to sample one population at each stage; since the variances are equal it seems natural that each of Π_1 and Π_2 should be sampled as nearly equally as possible. For example, we may start by sampling Π_1 and then, until we stop, we sample Π_2 and Π_1 alternately. A modification of this sort can be easily incorporated in our rules, but the consequent reduction in sample size is less than one.

2. FORMULATION AND PROCEDURES

The problem of identification described in the case of LFC reduces to testing between the following composite hypotheses:

$$H_1: (\mu_1 = \mu, \quad \mu_2 = \mu - \delta),$$

$$H_2: (\mu_1 = \mu + \delta, \quad \mu_2 = \mu).$$

We will use the notations P_θ and E_θ for probability and expectation respectively when $\theta = (\mu_1 - \mu, \quad \mu_2 - \mu)$ obtains. Let $\theta_1 = (0, -\delta)$ and $\theta_2 = (\delta, 0)$. In the sequel, there will be two types of errors, viz.

$$\alpha = P_{\theta_1}(\text{accept } H_2), \quad \beta = P_{\theta_2}(\text{accept } H_1).$$

Our object is to propose statistical methods so as to keep these errors α and β at desirable preassigned levels. Towards this end, we present two procedures in this section called I and II. As stated earlier, $k = \ell$ samples from Π_0 is given, and also we wish to take a sample each from Π_1 and Π_2 at every stage.

2.1 Fixed Sample Test

Suppose we have random variables $X_1, \ldots, X_k$ from Π_0, $Y_1, \ldots, Y_n$ from Π_1 and $Z_1, \ldots, Z_n$ from Π_2. Under the location shifts, i.e. $X_i \to X_i + c$, $Y_j \to Y_j + c$, $Z_j \to Z_j + c$ ($i = 1, \ldots, k$; $j = 1, \ldots, n$), $-\infty < c < \infty$, the invariant sufficient statistic is $(\bar{X}_k - \bar{Y}_n, \bar{X}_k - \bar{Z}_n)$ which follows from Stein's theorem as in Hall et al (1965). Since we require P_{θ_1} (reject H_1) $= \alpha$, and P_{θ_2} (accept H_1) $\leq \beta$, the best fixed sample-size invariant test is to find a constant $c(\alpha, \beta)$, depending on α, β, such that:

Reject H_1 if $\ell n\, W_n < c(\alpha, \beta)$

where W_n = ratio of the likelihoods of $(\bar{X}_k - \bar{Y}_n, \bar{X}_k - \bar{Z}_n)$ under H_1 and H_2 respectively. It is easy to see that

$$\ell n\, W_n = \frac{\delta k}{(2n + k)} \sum_{i=1}^{n} (2\bar{X}_k - Y_i - Z_i).$$

Let $\Phi(\tau_\alpha) = 1 - \alpha$ where Φ is the cdf of $N(0,1)$ distribution. Suppose, as usual, that $\alpha + \beta < 1$ (if $\alpha + \beta \geq 1$, one need not experiment at all) and thus $\tau_\alpha + \tau_\beta > 0$. In this case a fixed sample size test with given requirements exists if and only if

$$\delta \geq (\tau_\alpha + \tau_\beta)\left(\frac{1}{k} + \frac{1}{2n}\right)^{\frac{1}{2}}. \tag{1}$$

Now, for a given δ, equation (1) will have a solution in n if and only if

$$\delta > (\tau_\alpha + \tau_\beta)k^{-\frac{1}{2}}. \tag{2}$$

If (2) holds, the best choice of the sample size (ignoring its fractional part) is $M_1 = k[\{\delta^2 k/(\tau_\alpha + \tau_\beta)^2\} - 1]^{-1}/2$. Let $M = [M_1]$, the smallest integer bigger than or equal to M_1.

2.2 Procedure I

Conditional on $\bar{X}_k = k^{-1}\sum_{i=1}^{k} X_i$, we consider the auxiliary problem of deciding between the two simple hypotheses:

$$H_1^*: (\mu_1 = \bar{X}_k, \quad \mu_2 = \bar{X}_k - \delta)$$

$$H_2^*: (\mu_1 = \bar{X}_k + \delta, \quad \mu_2 = \bar{X}_k)$$

with preassigned errors α and β. Intuitively, if k is large we hope that H_1^*, H_2^* will not deviate much from H_1, H_2 respectively. At the same time, in case k is very small, the information about Π_0 itself will be very poor and it seems pointless to get more and more information about Π_1, Π_2 and then try to see which one of Π_1 or Π_2 looks more like Π_0.

Let $f(H_i^*, n)$ denote the joint density of Y_j, Z_j, $j = 1, \ldots, n$ conditional on $\bar{X}_k$, under the hypothesis H_i^*, $i = 1, 2$. We write $\ln W_n^* = \ln\{f(H_1^*, n)/f(H_2^*, n)\} = \delta\sum_{i=1}^{n}(2\bar{X}_k - Y_i - Z_i)$. Like Wald's (1947) SPRT, we choose two constants A, B with $0 < A < 1 < B < \infty$. We now propose the following rule:

R_1: At the n^{th} stage,

decide $\Pi_0 = \Pi_1$ if $\delta\sum_{i=1}^{n}(2\bar{X}_k - Y_i - Z_i) \geq b$,

decide $\Pi_0 = \Pi_2$ if $\delta\sum_{i=1}^{n}(2\bar{X}_k - Y_i - Z_i) \leq a$,

and continue sampling by taking one observation on both Π_1, Π_2 if $a < \delta\sum_{i=1}^{n}(2\bar{X}_k - Y_i - Z_i) < b$ where $a = \ln A$, $b = \ln B$.

Now the problem is to find a and b for given α and β. Given $\bar{X}_k$, we find the expression for $L(\bar{X}_k)$, the probability of accepting H_1, by using Wald's fundamental identity. It is easy to see that

$$L(\bar{X}_k) \approx \{1 - \exp(t_0 a\delta^{-1})\}\{\exp(t_0 b\delta^{-1}) - \exp(t_0 a\delta^{-1})\}^{-1} \quad (3)$$

where $t_0 = \mu_1 + \mu_2 - 2\bar{X}_k$ which reduces to $2\mu - \delta - 2\bar{X}_k$ under H_1

and to $2\mu + \delta - 2\overline{X}_k$ under H_2. Our error requirements suggest that a and b be chosen so that

$$E_{\theta_1}(L(\overline{X}_k)) = 1 - \alpha \quad \text{and} \quad E_{\theta_2}(L(\overline{X}_k)) = \beta . \tag{4}$$

So, one has to solve these two integral equations to obtain a and b.

Here it may be noted that the usual justification of Wald's (1947) approximations consists of two arguments. One involves the use of Wald's inequalities to prove that the approximations are likely to be conservative, but this cannot be used in the present context. The other argument notes that if the mean and variance of the summand are small, the excess over the boundaries may be expected to be small (of course this is not always true). This applies to the present set-up also, provided δ is small compared with min(-a, b). Our Monte Carlo results of table II seem to be quite favorable.

Let N_1 be the random sample size given by the rule R_1. If $S_n = \sum_{i=1}^{n} (2\overline{X}_k - Y_i - Z_i)$, given $\overline{X}_k$, S_n is the sum of n iid random variables (with finite mean) and hence $P_\theta(N_1 < \infty | \overline{X}_k) = 1$ and so $P_\theta(N_1 < \infty) = 1$ under H_1 and H_2. In fact, we have the following stronger assertion.

<u>Theorem 1.</u> <u>For the rule</u> R_1, $E_\theta(N_1) < \infty$ <u>for all fixed</u> θ.

<u>Proof.</u> If we proceed as in Stein (1946), we get

$$P_\theta(N_1 > n|\overline{X}_k) \leq \{\rho(\overline{X}_k)\}^{n-1}$$

where

$$\rho(\overline{X}_k) = \min\{1 - g(\overline{X}_k), \ 1 - h(\overline{X}_k)\},$$

$$g(\overline{X}_k) = P_\theta\{\delta(2\overline{X}_k - Y_1 - Z_1) > 2(b - a)|\overline{X}_k\},$$

$$h(\overline{X}_k) = P_\theta\{\delta(2\overline{X}_k - Y_1 - Z_1) < -2(b - a)|\overline{X}_k\} .$$

Now $\rho(\overline{X}_k) < 1$ for all $\overline{X}_k$, and $\rho(\overline{X}_k)$ is also continuous. Also, $\rho(X_k) \to 0$ as $\overline{X}_k \to \pm\infty$. Hence, $\sup_{\overline{X}_k} \rho(\overline{X}_k) < 1$, which implies the required result.

However, since each new observation on Π_1, Π_2 supplies a rapidly diminishing amount of additional information, a truncation seems desirable. We check with R_1 at every stage until we reach $n = 2M$, M being defined in section 2.1. When $n = 2M$ and R_1 still dictates to go on sampling, we decide that $\Pi_0 = \Pi_1$ if $0 \leq \delta \sum_{i=1}^{n} (2\bar{X}_k - Y_i - Z_i) < b$, and that $\Pi_0 = \Pi_2$ if $a < \delta \sum_{i=1}^{n} (2\bar{X}_k - Y_i - Z_i) < 0$. (However, we may truncate R_1 at any stage and give a rule to decide between H_1 and H_2 quite analogously).

Table II on page 60 of Wald (1947) suggests that in the iid case the truncation point for the common values of α and β should lie between $2M'$ and $3M'$, where M' is the sample size for the best fixed sample-size test. In our context, because of diminishing information, an earlier truncation at 2M seems to be more reasonable.

2.3 Procedure II

We consider the same location shift of transformations (as discussed in section 2.1) and following the ideas of Hall et al (1965) and Mallows (1953) we consider SPRT's based on the invariantly sufficient sequence of random variables $U_n = (\bar{X}_k - \bar{Y}_n, \bar{X}_k - \bar{Z}_n)$, $n = 1, 2, \ldots$. where the distribution of U_n is bivariate normal with mean vector $(\mu - \mu_1, \mu - \mu_2)$ and the dispersion matrix $\Sigma_n = (\sigma_{ij})$ where $\sigma_{11} = \sigma_{22} = k^{-1} + n^{-1}$, $\sigma_{12} = k^{-1}$. Notice that the pdf of U_n is completely specified under H_1 and H_2. From section 2.1 we recall that the log-likelihood ratio under H_1 and H_2 is given by

$$\ell n W_n = \frac{\delta k}{(2n + k)} \sum_{i=1}^{n} (2\bar{X}_k - Y_i - Z_i).$$

Now, we proceed as in Mallows (1953) and we propose the following sequential test. Given two errors α and β, we choose constants $a = \ln(\beta/(1-\alpha))$ and $b = \ln((1-\beta)/\alpha)$, and stopping rule is as follows:

R_2: At the n^{th} stage,

decide that $\Pi_0 = \Pi_1$ if $\ln W_n \geq b$,

decide that $\Pi_0 = \Pi_2$ if $\ln W_n \leq a$,

and continue the experiment by taking one more observation on both Π_1 and Π_2 if $a < \ln W_n < b$. Let N_2 be the random sample size for this rule. Then we have

Theorem 2. $P_\theta(N_2 = \infty) > 0$ for all fixed θ.

A proof of Theorem 2 is provided in the appendix. Incidentally this gives an example of the fact that SPRT's based on the maximal invariants need not terminate with probability one. In this case, truncation of the rule R_2 is an absolute necessity.

We check with the rule R_2 at every stage until we reach $n = 2M$, M being defined in section 2.1. When $n = 2M$ and R_2 still needs more samples to stop, we decide $\Pi_0 = \Pi_1$ if $0 \leq \delta k \sum_{i=1}^{n} (2\bar{X}_k - Y_i - Z_i) < b(2n+k)$, and $\Pi_0 = \Pi_1$ if $a(2n+k) < \delta k \sum_{i=1}^{n} (2\bar{X}_k - Y_i - Z_i) < 0$.

2.4 Discussion and Comments

It is more reasonable to assume the separation is $\delta\sigma$ and regard σ as unknown. In this case one can develop a truncated SPRT based on the student's t-statistic

$$t_n = (2\bar{X}_k - \bar{Y}_n - \bar{Z}_n)/S_n$$

where

$$S_n^2 = \sum_1^k (X_i - \bar{X}_k)^2 + \sum_1^n (Y_i - \bar{Y}_n)^2 + \sum_1^n (Z_i - \bar{Z}_n)^2 .$$

This is similar to our procedure II. A test similar to our procedure I can also be developed. We did not consider this extension partly to keep our Monte Carlo calculations simple and partly because we felt the properties of the tests for unknown σ are likely to be similar to those of the tests for known σ.

If, on the other hand, we assumed a two-sided alternative $|\mu_1 - \mu_2| \geq \delta$ in place of (A), the problem becomes much more difficult. One can again invoke invariance and sufficiency (vide Hall et al. (1965)) to develop a truncated SPRT for

$$H_1':(\mu = \mu_1, \quad \mu_2 = \mu \pm \delta)$$
$$H_2':(\mu = \mu_2, \quad \mu_1 = \mu \pm \delta)$$

but its performance for $|\mu_2 - \mu_1| > \delta$ is not clear. The difficulty remains even in the fixed sample size case. But a number of reasonable procedures have been studied [vide Khatri and Srivastava (1979), Chapter 8]. A different sequential formulation with sequential sampling from all the three populations is given in Srivastava (1973). For an elaborate review on theories and methods in classification, one may refer to DasGupta (1973).

3. NUMERICAL STUDIES

In this section we study the truncated procedures I and II as described in sections 2.2 and 2.3, and compare these with the best fixed sample-size invariant test given in section 2.1.

3.1 Procedure I of Section 2

For simplicity we take $\alpha = \beta$, so that $a = -b$ and $L(\overline{X}_k) \simeq \{1 + \exp(t_0 b\delta^{-1})\}^{-1}$. If we take $M_1 = 2k$, from section 2.1 it follows that

$$\delta \geq \left(\frac{5}{k}\right)^{\frac{1}{2}} \tau_\alpha = \delta_0, \text{ say.} \tag{5}$$

For computational purposes, we make $\alpha = \beta = .01$, $k = 1, 5, 20, 50, 100, 200$, $\delta = \delta_0$, $\delta_0 + 0.5$. The first value in table 1 (column for δ) is δ_0. Now, (4) reduces to

$$\begin{aligned} \alpha &= E\{1 + \exp(b(1 + 2\delta^{-1}\overline{X}_k))\}^{-1} \\ &\geq [1 + E\{\exp(b(1 + 2\delta^{-1}\overline{X}_k))\}]^{-1}, \end{aligned} \tag{6}$$

by Jensen's inequality, which implies that

TABLE I

Values of δ, b and the Integral in (6)

k	δ	b	Integral
1	5.2018	8.5770	.0114
	5.7018	7.4970	.0100
5	2.3263	8.5770	.0114
	2.8263	6.7001	.0092
20	1.1632	8.5770	.0114
	1.6632	5.9389	.0087
50	0.7356	8.5770	.0114
	1.2356	5.2895	.0101
100	0.5202	8.5770	.0114
	1.0202	5.2769	.0085
200	0.3678	8.5770	.0114
	0.8678	4.9785	.0096

$$b \geq \frac{k\delta^2}{4} + \left\{\frac{k^2\delta^4}{16} + \frac{k\delta^2}{2}\,\ell n\left(\frac{1-\alpha}{\alpha}\right)\right\}^{\frac{1}{2}} = b_0, \text{ say.}$$

So, for given α, δ, k, one can take b_0 as the first approximation to b and use numerical integration techniques to solve (6) for b. In table I, we present the b-values for different k, δ's.

For Monte Carlo experiments, we take Π_0 and Π_1 to be the N(0, 1) populations, and Π_2 as N($-\delta$, 1). Then we use our truncated decision rule R_1 (as described in section 2.2) 200 times for each entry, and the results are presented in table II. In tables II and III, P_{θ_1}(CI) stands for the estimated probability of correctly identifying Π_0 with Π_1.

From column 3 of tables II and III, we note that $P\theta_1$(Truncation) is quite small, and so we do not record the $P\theta_1$(CI) for the

TABLE II

Small and Moderate Sample Behavior of Procedure I

k	δ	Truncated Cases: No. of truncations	Truncated Cases: Truncation point	Untruncated Cases: Relative frequency $P_{\delta_1}(CI)$
1	5.2018	2	4.00	.9848
	5.7018	6	1.99	.9948
5	2.3263	-	20.00	1.0000
	2.8263	-	5.92	.9950
20	1.1632	-	80.00	1.0000
	1.6632	1	12.86	1.0000
50	0.7356	-	200.00	.9900
	1.2356	1	19.79	.9899
100	0.5202	-	400.00	.9950
	1.0202	1	26.26	1.0000
200	0.3678	-	800.00	.9900
	0.8678	1	33.57	.9950

truncated case. Also one may note that the sample size for the best fixed sample-size invariant test is about half of the truncation point -- and this remark applies to both the tables II and III.

3.2 Procedure II of section 2

As in (3.1), in this case, δ_0 is given by

$$\delta_0 = 2(\tau_\alpha + \tau_\beta)(5/k)^{\frac{1}{2}} .$$

We consider $\alpha = .05$, $\beta = .01$; $\delta = \delta_0$, $\delta_0 + 0.5$, and the same

TABLE III

Small and Moderate Sample Behavior of Procedure II

k	δ	Truncated Cases: No. of truncations	Truncated Cases: Truncation point	Untruncated Cases: Relative frequency $P_{\theta_1}(CI)$
1	4.4399	15	4.00	1.0000
	4.9399	23	1.83	1.0000
5	1.9856	12	20.00	1.0000
	2.4856	8	5.21	.9948
20	0.9928	10	80.00	.9895
	1.4928	3	10.95	.9797
50	0.6279	9	200.00	.9948
	1.1279	3	16.48	.9898
100	0.4434	8	400.00	.9899
	0.9434	1	21.50	.9949
200	0.3139	8	800.00	.9844
	0.8139	1	27.02	.9899

k-values as in section 3.1. We have $a = -4.5539$, $b = 2.9857$ to be used in the rule R_2. For Monte-Carlo experiments, we proceed as in section 3.1, and present our findings in table III.

It can be seen from tables II and III that if the truncation point is 2M, the attained α, β can be substantially lower than the desired values. We have not reported here the values of ASN, however our Monte Carlo studies also revealed that the average sample sizes of our tests are between one-fourth and one-half of the sample sizes of the corresponding UMP invariant fixed sample size tests.

ACKNOWLEDGEMENTS

We are grateful to Mr. T. Rajeev for his able assistance with the programming. One of the authors (NM) is grateful to Professor Somesh DasGupta for some discussions which he had with him at a very early stage. Professor C. G. Khatri pointed out the relevance of the work in Khatri and Srivastava (1979). We appreciate the comments of a very knowledgeable referee -- the suggestions made by the referee were invaluable for the improvements in this text.

DIBLIOGRAPHY

DasGupta S. (1973). Theories and methods in classification: a review. Discriminant Anaylsis and Applications, 77-137. New York: Academic Press.

Hall, W. J., Wijsman, R. A. & Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. Ann. Math. Statist., 36, 575-614.

Khatri, C. G. & Srivastava, M.S. (1979). An Introduction to Multivariate Analysis. North Holland Publishing Co.

Mallows, C. L. (1953). Sequential discrimination. Sankhyā, 12, Part 4, 321-338.

Rivers, W. H. R. (1906). The Todas. New York: MacMillan Co.

Srivastava, M. S. (1973). A sequential approach to classification: Cost of not knowing the covariance matrix. Jour. Mult. Analysis, 3, 173-183.

Stein, C. (1946). A note on cumulative sums. Ann. Math. Statist., 17, 498-499.

Wald, A. (1947). Sequential Analysis. New York: John Wiley & Sons, Inc.

Received at the Editorial Office March, 1980.

Recommended by P. K. Sen, The University of North Carolina, Chapel Hill, NC.

Refereed by B. K. Ghosh, Virginia Polytechnic Institute and State University, Blacksburg, VA.

APPENDIX

Proof of Theorem 2

Let $\ell n\, W_n$ be as defined in section 2.1. First note that a.s.

$$\lim_{n\to\infty} \ell n\, W_n = \frac{\delta k}{2}\left\{2\overline{X}_k - \mu_1 - \mu_2\right\}$$

Hence $\operatorname{Sup}_n \ell n\, W_n < \infty$ a.s. and $\inf_n \ell n\, W_n < \infty$ a.s. This immediately implies we can choose $a < b$ such that

$$P_\theta\left\{a < \inf_n \ell n\, W_n\ ,\ \operatorname{Sup}_n \ell n\, W_n < b\right\} > 0\ . \tag{7}$$

To show that (7) holds for all $a < 0 < b$ requires a bit more calculation carried out in the next paragraph.

For given $a < 0 < b$ choose ε such that $0 < \varepsilon < (b - a)/k\delta$. Let A_ε be the event

$$2a/k + \delta(\mu_1 + \mu_2 + \varepsilon) < 2\delta\,\overline{X}_k < 2b/k + \delta(\mu_1 + \mu_2 - \varepsilon)\ .$$

Let B_ε be the event

$$|\overline{Y}_n + \overline{Z}_n - \mu_1 - \mu_2| < \varepsilon \text{ for all } n \geq 1\ .$$

We shall show below

$$P_\theta(B_\varepsilon) > 0\ . \tag{8}$$

Then $P_\theta(N_2 = \infty) \geq P_\theta(A_\varepsilon \cap B_\varepsilon) = P_\theta(A_\varepsilon)\, P_\theta(B_\varepsilon) > 0$.

It remains to prove (8). Let

$$S_{m,n} = \sum_1^n (Y_{m+i} + Z_{m+i} - \mu_1 - \mu_2)\ .$$

By the strong law of large numbers given $\varepsilon > 0$, $\eta > 0$, there exists n_0 such that

$$P_\theta\left|\left|\frac{S_{m,n}}{n}\right| < \frac{\varepsilon}{2}\ ,\quad n \geq n_0\right| \geq 1 - \eta$$

for all m which implies

$$P_\theta\left|\left|\frac{S_{m,n}}{m+n}\right| < \frac{\varepsilon}{2}\ ,\quad n \geq n_0\right| \geq 1 - \eta\ .$$

Choose now m so large that

$$P_\theta\left\{\left|\frac{S_{m,n}}{m+n}\right| < \frac{\varepsilon}{2}, \quad 1 \le n < n_0\right\} \ge 1 - \eta .$$

Then if η is sufficiently small and m sufficiently large

$$P_\theta\left\{\left|\frac{S_{m,n}}{m+n}\right| < \frac{\varepsilon}{2}, \quad 1 \le n < \infty\right\} > 0 . \tag{9}$$

For any fixed m,

$$P_\theta\left\{\left|\frac{S_{0,n}}{n}\right| < \frac{\varepsilon}{2}, \ n = 1, \ldots, m\right\} > 0 . \tag{10}$$

Combining (9) and (10) and noting (i) that the events in (9) and (10) are independent and (ii) that for $n \ge m$

$$\left|\frac{S_{0,n}}{n}\right| \le \frac{m}{n} \frac{|S_{0,m}|}{m} + \frac{|S_{0,n-m}|}{n} ,$$

we get (8).