หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

ระเบียบวิธีวิจัย

ในงานวิจัยนี้ต้องการศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณ 4 วิธีคือวิธี R วิธี B วิธี U และวิธี DS โดยจะศึกษาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณในแต่ละวิธีเมื่อประชากรมีการแจกแจงแบบพหุปกติทั้ง 2 กลุ่มประชากรโดยมีค่า Mahalanobis Square Distance , Δ^2 เป็นตัวแยกประชากรทั้ง 2 กลุ่มออกจากกันโดยให้มีค่าของ Square Root of Mahalanobis Distance, Δ แตกต่างกันไปดังนี้ 0.0 0.5 1.0 1.5 2.0 2.5 3.0 ขนาดตัวอย่างที่ใช้ในการศึกษาครั้งนี้มี 5 ขนาดคือ 10 20 25 50 และ 100 ขนาดของตัวแปรอิสระมีอิทธิพลต่อวิธีการต่าง ๆ ขนาด 3 5 7 และ 9 ทั้งนี้เทคนิคที่ใช้ในการหาค่าความคลาดเคลื่อนคือวิธีมอนติคาร์โลซึ่งเป็นเทคนิคอย่างหนึ่งที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์โดยเฉพาะกรณีที่ไม่สามารถหาคำตอบโดยวิธีทางทฤษฎีได้

เนื่องจากวิธีมอนติคาร์โลเป็นเทคนิคที่ใช้ในการวิจัยนี้ดังนั้นในตอนแรกของบทจะกล่าวถึงวิธีมอนติคาร์โลก่อนแล้วจึงแสดงรายละเอียดของแผนการทดลอง ขั้นตอนการวิจัยและโปรแกรมที่ใช้ในการวิจัยตามลำดับ ซึ่งรายละเอียดต่างๆ มีดังนี้

3.1 วิธีมอนติคาร์โล (Monte Carlo Method)

เทคนิคที่ใช้แก้ปัญหาในการคำนวนทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีมอนติคาร์โลเป็นวิธีหนึ่งที่จะใช้แก้ปัญหาได้และเป็นวิธีที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน แฮมเมอร์เลย์และแฮนสโคมบ์ (Hammersley and Handscomb 1964) กล่าวว่าวิธีมอนติคาร์โลเป็นสาขาหนึ่งของคณิตศาสตร์เชิงทดลอง ซึ่งหลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่ม (Random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ในการวิจัยครั้งนี้จะใช้เทคนิคมอนติคาร์โลดังกล่าวในการสร้างข้อมูลที่มีสภาพการแจกแจงตามต้องการ ซึ่งขั้นตอนที่ใช้กันอยู่ในปัจจุบันแบ่งออกได้เป็น 3 ขั้นตอน ดังนี้

3.1.1 การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งสำคัญมากในวิธีมอนติคาร์โลทั้งนี้เพราะว่าหลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา ลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบ ยูนิฟอร์ม (Uniform) ในช่วง (0,1) สำหรับวิธีการสร้างเลขสุ่มนั้นมีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของตัวเลขสุ่มที่เกิดขึ้นจะต้องมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) และเป็นอิสระกัน

3.1.2 การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรงแต่อาจมีขั้นตอนอื่นอีกหลายขั้นตอนซึ่งขั้นตอนเหล่านี้มีบางขั้นตอนต้องใช้ตัวเลขสุ่ม

3.1.3 การทดลองกระทำ เมื่อประยุกต์ปัญหาให้ใช้กับตัวเลขสุ่มได้แล้วขั้นตอนต่อไปคือการทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำๆ กันเพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

3.2 แผนการทดลอง

ในงานวิจัยนี้กำหนดสถานการณ์ต่างๆสำหรับศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณโดยสุ่มตัวอย่างจากประชากร 2 กลุ่มที่มีการแจกแจงแบบเดียวกันคือการแจกแจงแบบพหุปกติโดยกลุ่มแรกให้มีการแจกแจงแบบปกติคือ $X_1 \sim N(0, I)$ กลุ่มที่สองให้มีการแจกแจงแบบพหุปกติคือ $X_2 \sim N(\Delta , 0 , 0 , \ldots , 0 , I)$ สถานการณ์ต่างๆ ที่กำหนดขึ้นนั้นเกี่ยวข้องกับปัจจัยต่างๆ ดังนี้ ค่า Square Root of Mahalanobis Distance, Δ ขนาดของตัวอย่างและขนาดของตัวแปรอิสระ แสดงด้วยตารางดังนี้

Δ	0.0	0.5	1.0	1.5	2.0	2.5	3.0
ขนาดตัวอย่าง $n_1=n_2$	10	20	25	50	100		
ขนาดตัวแปรอิสระ	3	5	7	9			

I17848890

3.3 ขั้นตอนในการวิจัย

ขั้นตอนการวิจัย แบ่งออกเป็น 8 ขั้นตอน ดังนี้

1. การสร้าง (Generate) ข้อมูล 2 กลุ่ม ให้มีการแจกแจงปกติ โดยกลุ่มแรกมีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น 1 ดังนี้

$X_1 = (x_1 , x_2 , x_3 , ... , x_p) \sim N (0, I)$

กลุ่มที่ 2 มีค่าเฉลี่ยของกลุ่มตัวแปรอิสระกลุ่มแรกเป็น Δ นอกนั้นเป็น 0 มีค่าความแปรปรวนเป็น 1 ดังนี้

$X_2 = (x_1, x_2, x_3, ... ,x_p) \sim N (\Delta, 0, 0, ... , 0, I)$

2. การหาค่าอัตราความผิดพลาดที่มีเงื่อนไข (Conditional Error Rate) ในการจำแนกกลุ่มผิด

3. การหาค่าประมาณอัตราความผิดพลาดในการจำแนกกลุ่มผิดโดยวิธี R

4. การหาค่าประมาณอัตราความผิดพลาดในการจำแนกกลุ่มผิด โดยวิธี U

5. การหาค่าประมาณอัตราความผิดพลาดในการจำแนกกลุ่มผิด โดยวิธี B

6. การหาค่าประมาณอัตราความผิดพลาดในการจำแนกกลุ่มผิด โดยวิธี DS

7. การหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณแต่ละวิธี

3.4 โปรแกรมที่ใช้ในการวิจัย

3.4.1 การสร้างข้อมูลแบบสุ่ม

ในงานวิจัยนี้สร้างโปรแกรมย่อย(Subprogram) เพื่อสร้างข้อมูลแบบสุ่ม (Random) เพื่อนำไปใช้ในการสร้างข้อมูลที่มีการแจกแจงปกติ ดังนี้

```
      SUBROUTINE RAND(IX,IY,YFL)
      IY = IX * 65539
      IF (IY) 5,6,6
    5 IY = TY + 2147483647 + 1
    6 YFL = IY
      YFL = YFL / 2147483647
      IX = IY
      RETURN
      END
```

เมื่อ IX = 973253

3.4.2 การสร้างโปรแกรมย่อยเพื่อสร้างข้อมูลให้มีการแจกแจงแบบปกติ

3.4.2.1 การแจกแจงแบบปกติ N (0,1)

ในงานวิจัยนี้สร้างโปรแกรมย่อยเพื่อใช้สร้างข้อมูลให้มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และความแปรปรวน 1 ดังนี้

```
      FUNCTION NORMAL( DMEAN , SIGMA )
      REAL NOMAL
      COMMON / SEED / IX / SELECT / kk
      PI = 3.1415926
      IF ( kk.EQ.1 ) GOTO 10
         CALL RAND( IX , IY , YFL )
         RONE = YFL
         CALL RAND( IX , IY , YFL )
         RTWO = YFL
       ZONE = SQRT( -2 * ALOG(RONE) * COS( 2 * PI * RTWO )
       ZTWO = SQRT( -2 * ALOG(RONE) * SIN( 2 * PI * RTWO )
       NORMAL = ZONE * SIGMA + DMEAN
       kk = 1
       RETURN
    10 NORMAL = ZTWO * SIGMA + DMEAN
       kk = 0
       RETURN
       END
```

3.4.2.2 การแจกแจงแบบปกติ N(Δ,1)

ในงานวิจัยนี้สร้างโปรแกรมย่อยเพื่อใช้สร้างข้อมูลให้มีการแจกแจงปกติที่มีค่าเฉลี่ย Δ และความแปรปรวน 1 ดังนี้

```
      SUBROUTINE SPECN( DMEAN2,SIGMA2,AX )
      COMMON / SEED / IX / SELECT / kk
      A = 0.0
      DO 51 I = 1 , 12
         CALL RAND(IX , IY , YFL )
         A = A + YFL
   51 CONTINUE
      AX = ( A - 6 ) * SIGMA2 + DMEAN2
      RETURN
      END
```

3.4.3 การสุ่มตัวอย่างแบบใส่คืน (Siampling with Replacement)

เป็นการสุ่มตัวอย่างที่ยอมให้มีหน่วยตัวอย่างซ้ำกันได้นั่นคือแต่ละหน่วยตัวอย่างมีโอกาส (Probability) ในการถูกสุ่มเท่ากันซึ่งเท่ากับ 1/N เมื่อ N เป็นขนาดของประชากร การวิจัยในครั้งนี้ใช้คอมพิวเตอร์เป็นเครื่องมือช่วยในการสุ่มตัวอย่างแบบใส่คืนโดยใช้ตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอมีค่าอยู่ในช่วง [0 , 1] เป็นตัวเปรียบเทียบกับค่าความน่าจะเป็นสะสม (Cumulative Probability) เพื่อกำหนดหน่วยตัวอย่างตามจำนวนที่ต้องการ ซึ่งขั้นตอนการสุ่มตัวอย่างแบบใส่คืนสรุปได้ดังนี้

1. คำนวณหาค่าความน่าจะเป็นของแต่ละหน่วยตัวอย่าง = 1/N
2. หาค่าความน่าจะเป็นสะสมแล้วจัดช่วง
3. สร้างตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง [0,1]
4. นำตัวเลขสุ่มในข้อ 3 มาเปรียบเทียบกับค่าความน่าจะเป็นสะสม ถ้าตกอยู่ในช่วงใดหน่วยนั้น ๆ จะถูกเลือกมาเป็นตัวอย่าง
5. ทำตามขั้นตอนในข้อ 3 - 4 n ครั้ง เมื่อ n คือขนาดตัวอย่างที่ต้องการ

<u>ตัวอย่าง</u> การสุ่มตัวอย่างแบบใส่คืน

เมื่อ N = 10
n = 3

คำนวณหาค่าความน่าจะเป็นของแต่ละหน่วยตัวอย่างได้ 1/10 ซึ่งเท่ากับ 0.10 ดังนั้นสามารถนำมาสร้างตารางได้ดังนี้

หน่วยตัวอย่าง	ความน่าจะเป็น	ความน่าจะเป็นสะสม	ช่วงความน่าจะเป็นสะสม
1	0.10	0.10	0.01 - 0.10
2	0.10	0.20	0.11 - 0.20
3	0.10	0.30	0.21 - 0.30
4	0.10	0.40	0.31 - 0.40
5	0.10	0.50	0.41 - 0.50
6	0.10	0.60	0.51 - 0.60
7	0.10	0.70	0.61 - 0.70
8	0.10	0.80	0.71 - 0.80
9	0.10	0.90	0.81 - 0.90
10	0.10	1.00	0.91 - 1.00

สมมติเลขสุ่มตัวที่ 1 มีค่า = 0.18 หน่วยที่ 2 จะถูกเลือกมาเป็นตัวอย่าง
ตัวที่ 2 มีค่า = 0.55 หน่วยที่ 6 จะถูกเลือกมาเป็นตัวอย่าง
ตัวที่ 3 มีค่า = 0.12 หน่วยที่ 2 จะถูกเลือกมาเป็นตัวอย่าง

จะเห็นได้ว่าแต่ละหน่วยตัวอย่างมีโอกาสถูกเลือกมากกว่า 1 ครั้ง ทั้งนี้ขึ้นกับค่าของตัวเลขสุ่มว่าจะตกอยู่ในช่วงใดของค่าความน่าจะเป็นสะสม

ดังนั้นการสร้างโปรแกรมย่อยที่ใช้ในการสุ่มตัวอย่างแบบใส่คืน แสดงได้ดังนี้

```
      SUBROUTINE  REP( NN,ZE1,ZE2,ZE3 )
      COMMON / SEED / IX / SELECT / kk
      DIMENTION ZE1(NN) , ZE2(NN+1) , ZE3(NN)
      DO 5008 I08  =  1,NN
         ZE3(I08)  = 0.0
 5008 CONTINUE
      DO 5000 J = 1,NN
         CALL RAND( IX , IY , YFL )
              DO 5005 I = 1 , NN
              IF  ((  YFL . GT . ZE2(I)) .AND. ( YFL .
LE . ZE2(I+1)) THEN
                   ZE3(J)  = ZE1(I)
                   GOTO 5000
               END IF
 5005 CONTINUE
 5000 CONTINUE
      RETURN
      END
```

เมื่อ NN เป็นขนาดตัวอย่าง
ZE1 เป็นค่าของประชากรที่จะทำการสุ่มขึ้นมา
ZE2 เป็นค่าความน่าจะเป็นสะสม
ZE3 เป็นค่าของตัวอย่างที่ได้จากการสุ่มแบบใส่คืน

3.4.4 การสร้างโปรแกรมย่อยหาค่าของฟังก์ชันการแจกแจงแบบปกติ (Standard Normal Distribution Function)

โปรแกรมย่อยนี้ต้องการคำนวณหาค่าของ

$$y = P(x) = Prob(X \leqslant x)$$

เมื่อ X คือ ตัวแปรสุ่มที่มีการแจกแจงแบบปกติมีค่าเฉลี่ย 0 และความแปรปรวน 1

$$P(x) = 1/\sqrt{2\pi} \int_{-\alpha}^{x} \exp(-u^2/2)\, du$$

$$= \Phi(x)$$

$$\Phi(-x) = 1 - \Phi(x)$$

ดังนั้น แฮสติงส์ (Hastingns) ได้ประมาณค่าที่ได้ของ $\Phi(x)$ ดังนี้

$$P(x) = 1 - f(x) \sum_{i=1}^{5} a_i w^i \quad ; \quad x \geqslant 0$$

เมื่อ

$$W = 1/(1+px)$$

$$f(x) = \exp(-x^2/2)/\sqrt{2\pi}$$

$$p = 0.2316419$$

$$a_1 = 0.3193815$$

$$a_2 = -0.3565638$$

$$a_3 = 1.781478$$

$$a_4 = -1.821256$$

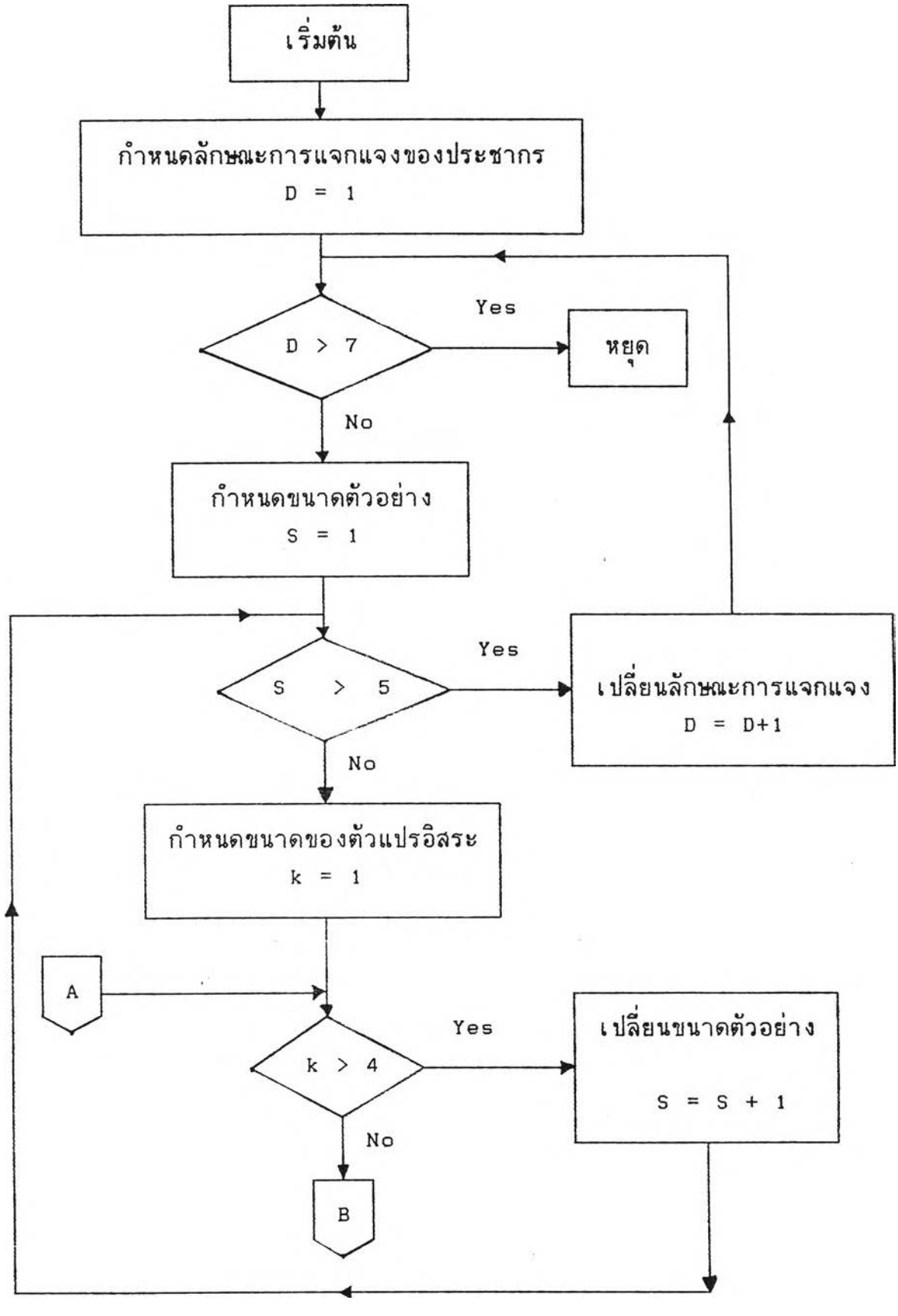
$$a_5 = 1.330274$$

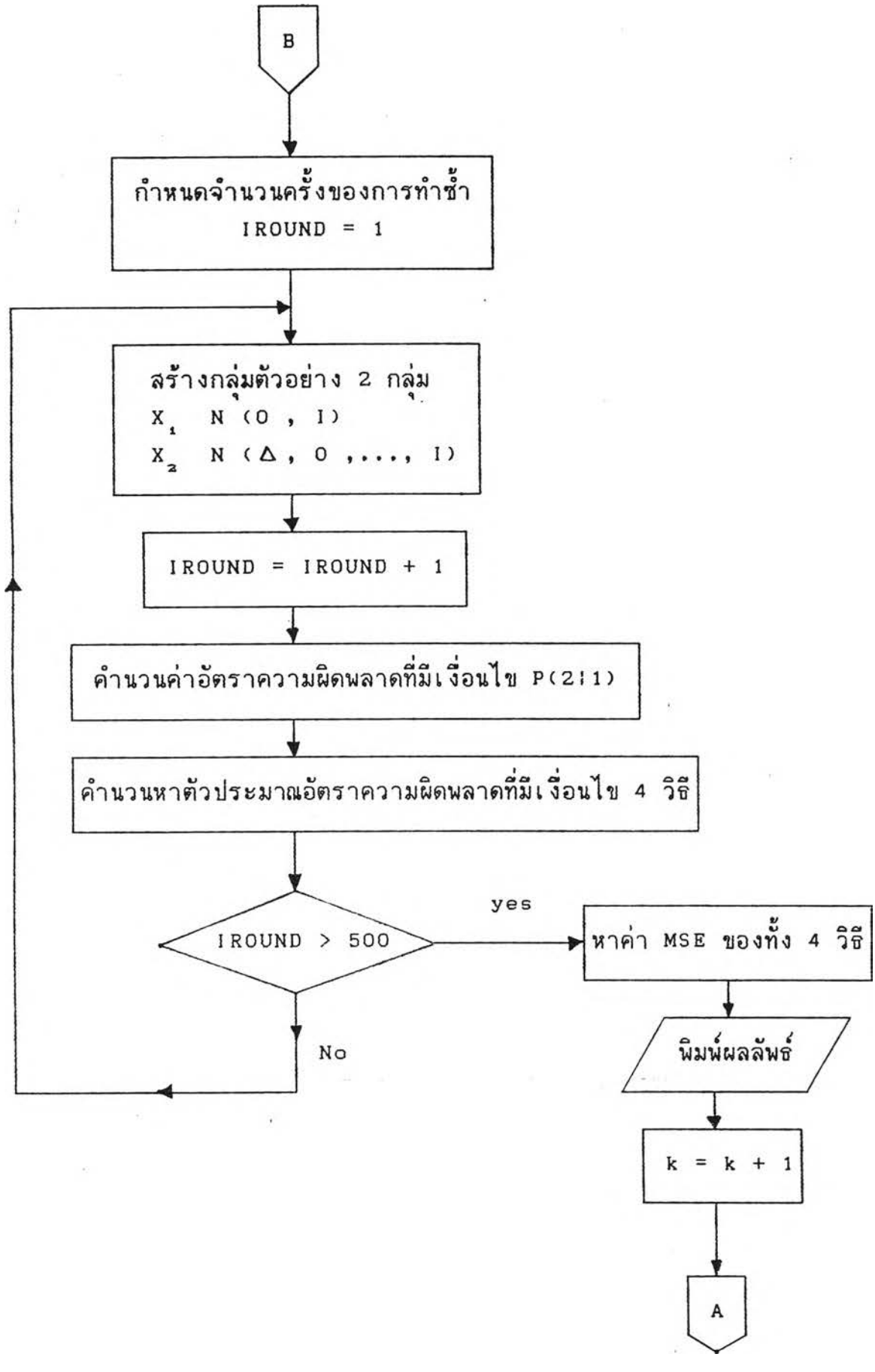
ค่าความผิดพลาดสูงสุด (Maxinum Error) คือ $7(10^{-7})$
เขียนโปรแกรมย่อยได้ดังนี้

โปรแกรมย่อยการหาค่าของฟังก์ชันการแจกแจงแบบปกติ

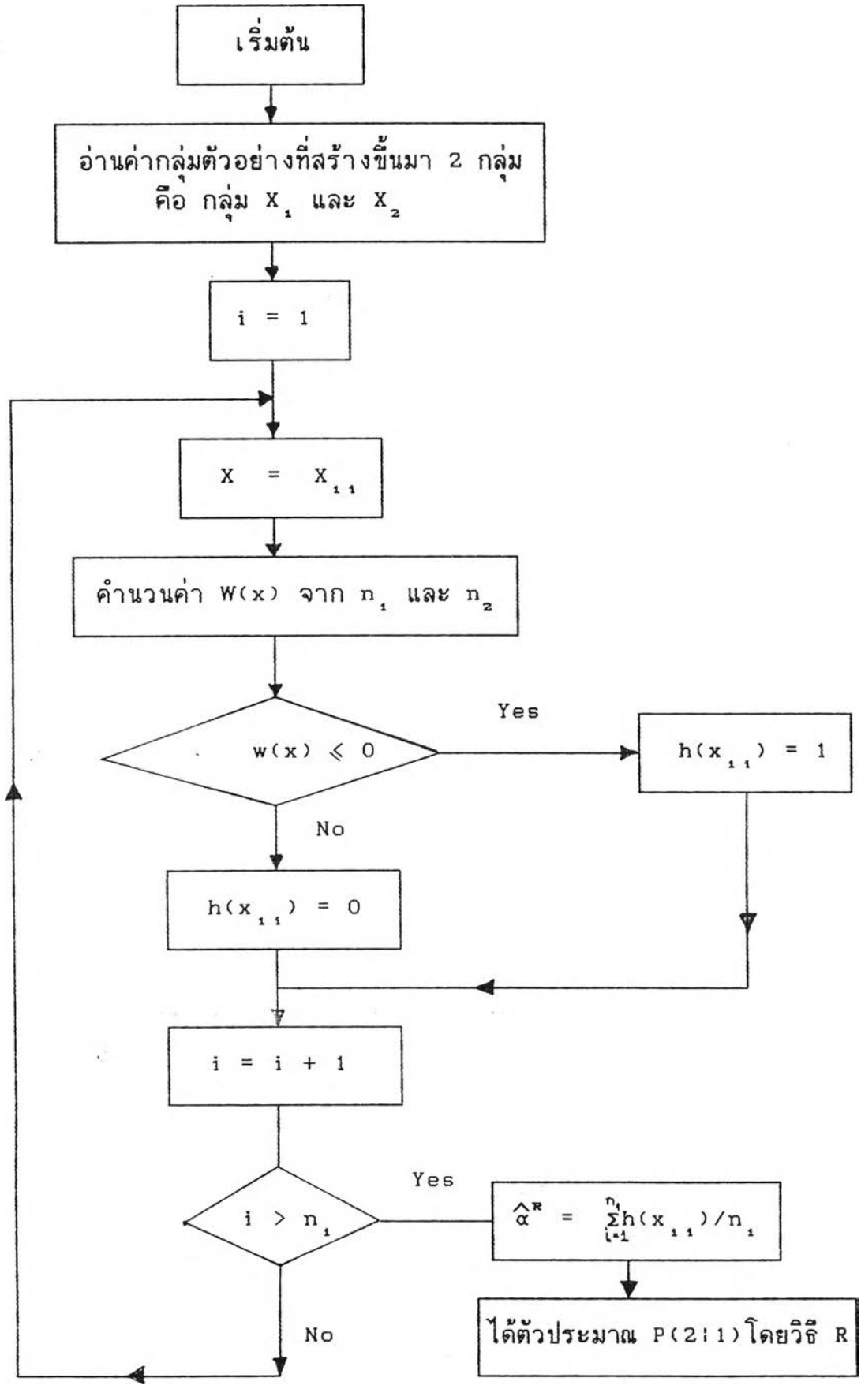
```
      SUBROUTINE NDTR(X,D)
      DOUBLE PRECISION A(5) , WPX(5)
      PI = 3.1415926
      P  = 0.2316419
      FX = 0.0
      XAB = ABS(x)
      D  = 0.0
      A(1) =  0.3193815
      A(2) = -0.3565638
      A(3) =  1.781478
      A(4) = -1.821256
      A(5) =  1.330274
      DO 1500 I15 = 1,5
         WPX(I15) = 0.0
1500  CONTINUE
      WPX(1) =  1/(1+P*XAB)
      WPX(2) =  WPX(1)**2
      WPX(3) =  WPX(1)**3
      WPX(4) =  WPX(1)**4
      WPX(5) =  WPX(1)**5
      FX =  (EXP(-(XAB*XAB)/2.0))/SQRT(2.0*PI)
      PX =  0.0
      DO 1501  I = 1,5
         PX =  PX + A(I) * WPX(I)
1501  CONTINUE
      IF (X) 12 , 13 , 13
   12 D =  FX * PX
      GOTO 14
   13 D =  1 - ( FX * PX )
   14 RETURN
      END
```

รูปที่ 3.1 แสดงแผนผังงาน สำหรับการหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ของตัวประมาณโดยวิธี R วิธี U วิธี B และวิธี DS

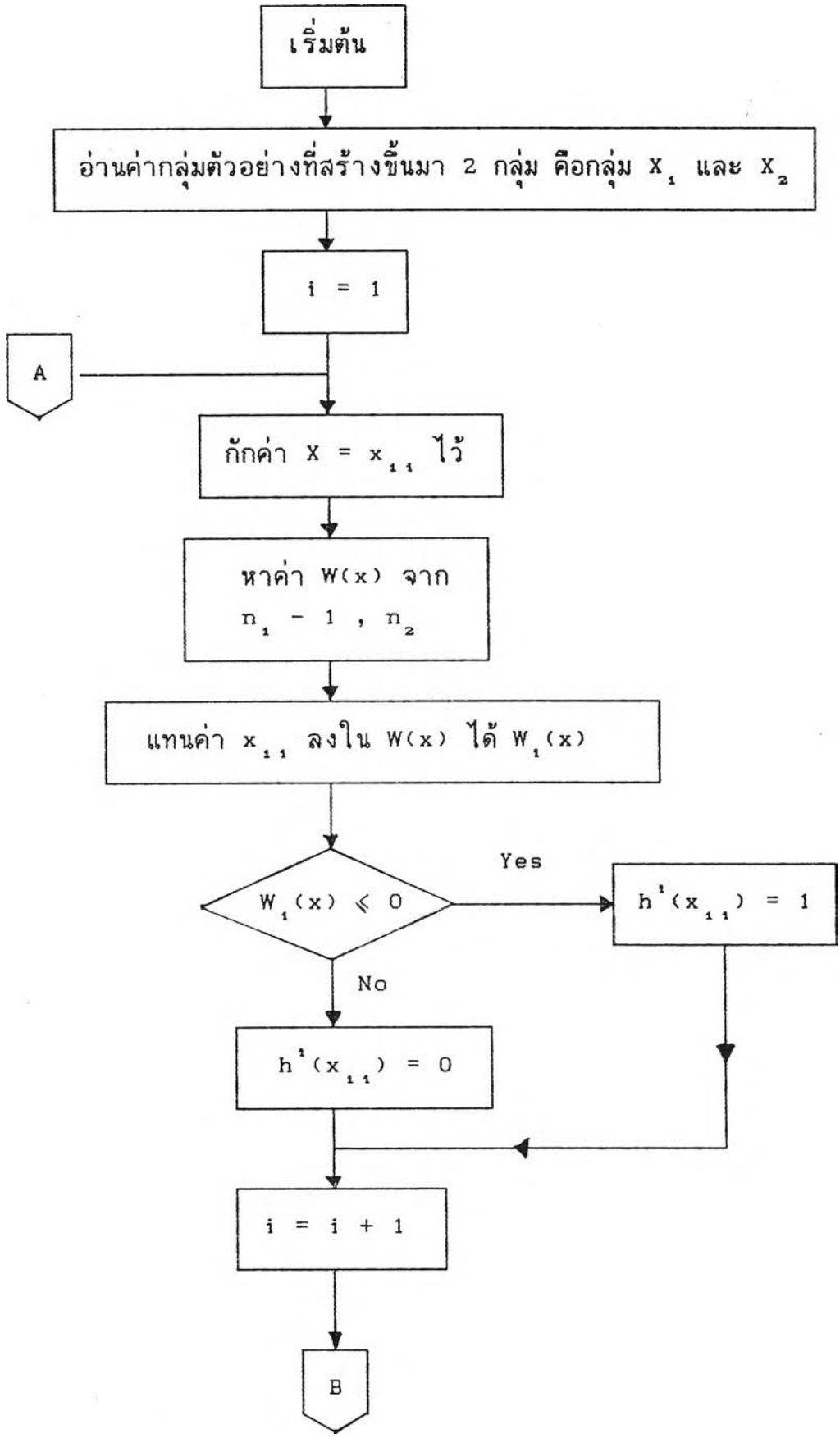


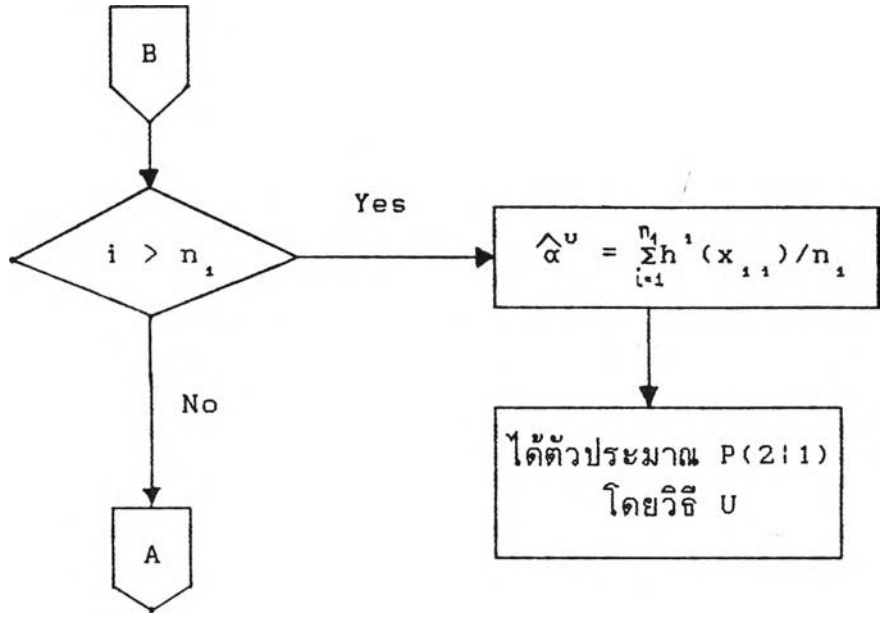

B
กำหนดจำนวนครั้งของการทำซ้ำ
IROUND = 1
สร้างกลุ่มตัวอย่าง 2 กลุ่ม
X_1 N (0 , I)
X_2 N (Δ , 0 ,... , I)
IROUND = IROUND + 1
คำนวนค่าอัตราความผิดพลาดที่มีเงื่อนไข P(2|1)
คำนวนหาตัวประมาณอัตราความผิดพลาดที่มีเงื่อนไข 4 วิธี
IROUND > 500
yes
No
หาค่า MSE ของทั้ง 4 วิธี
พิมพ์ผลลัพธ์
k = k + 1
A

รูปที่ 3.2 แสดงแผนผังงาน สำหรับการหาค่าประมาณของอัตราความผิดพลาดที่มีเงื่อนไขโดยวิธี R

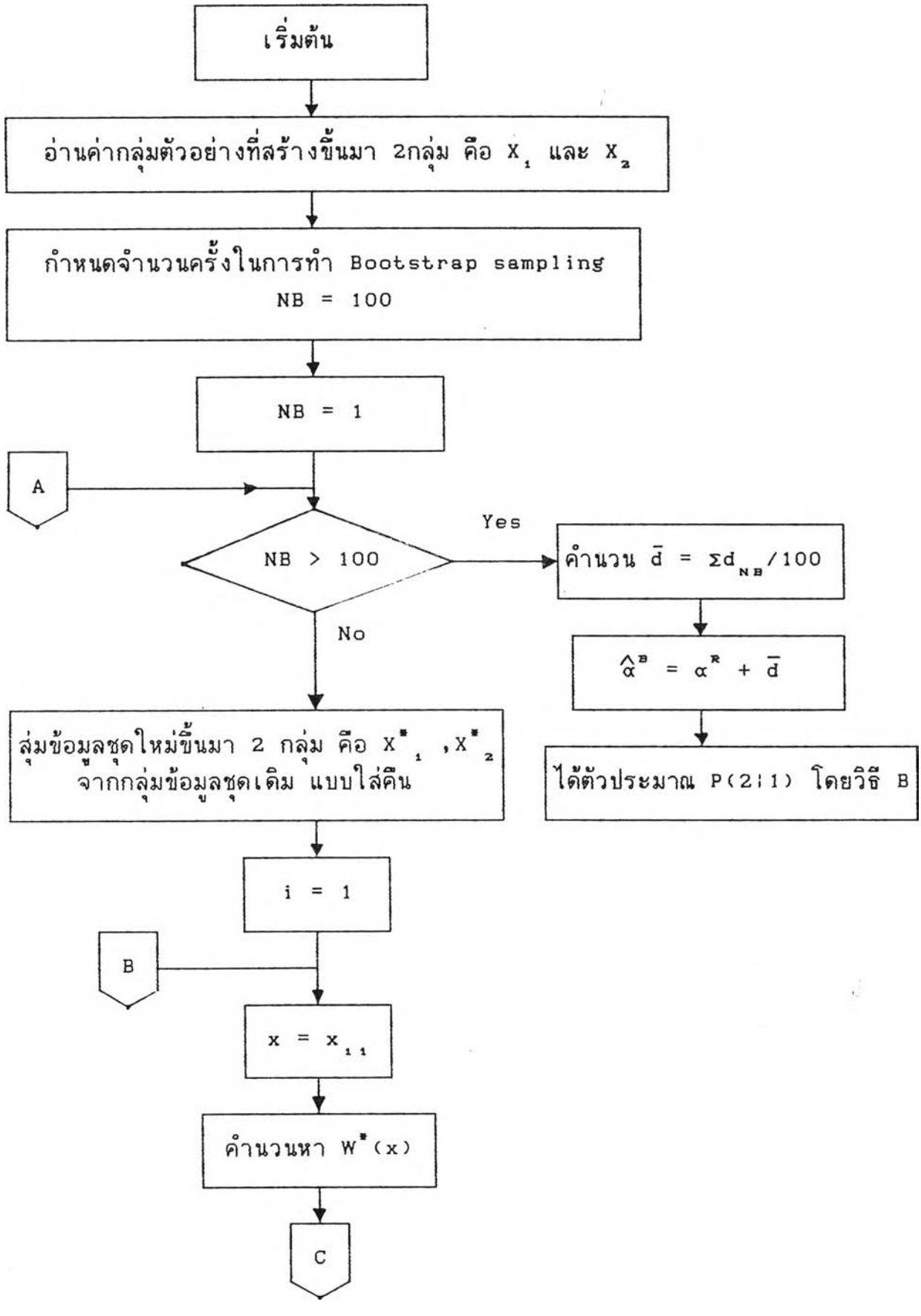


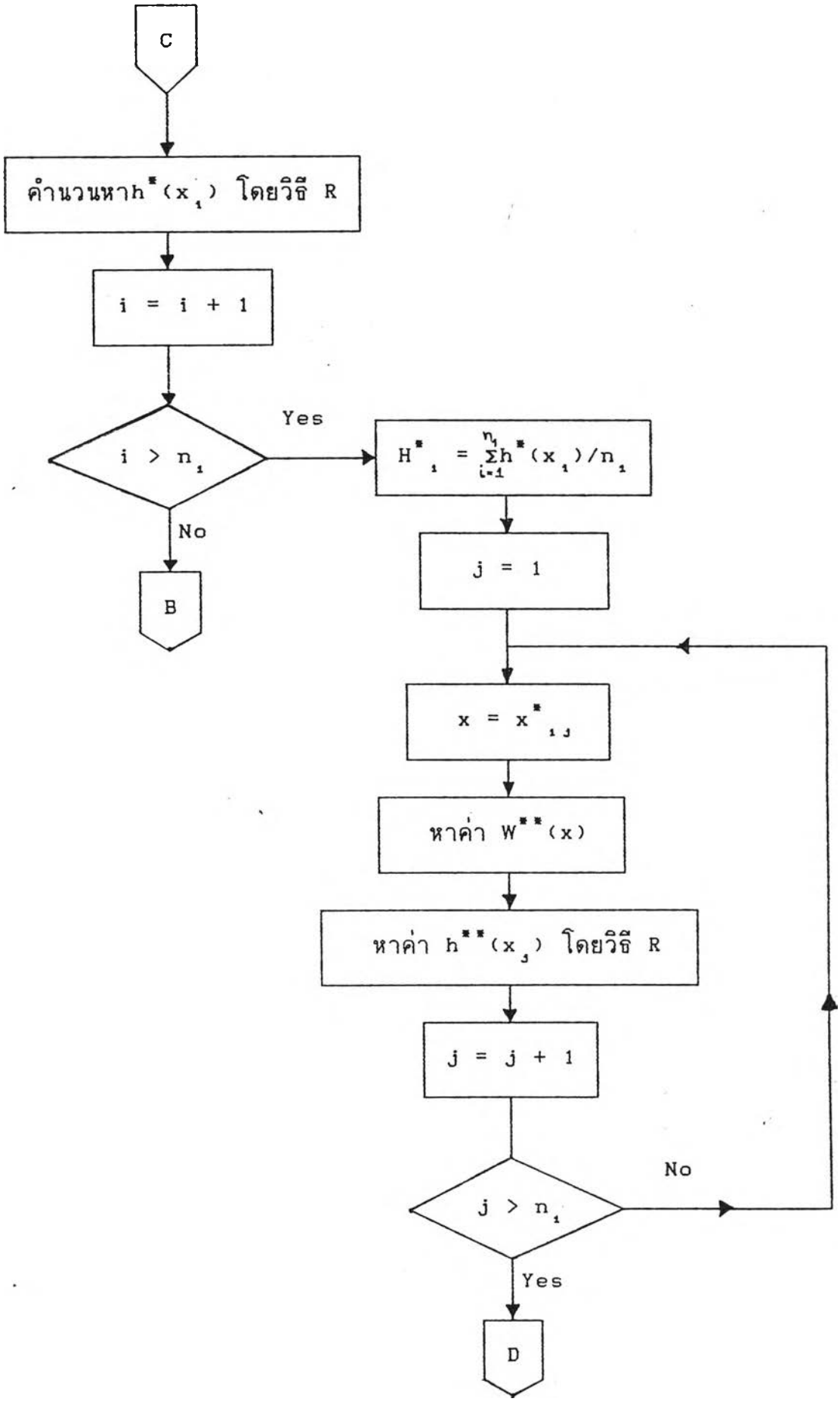
รูปที่ 3.3 แสดงแผนผังงาน สำหรับการหาค่าประมาณของอัตราความผิดพลาดที่มีเงื่อนไขโดยวิธี U

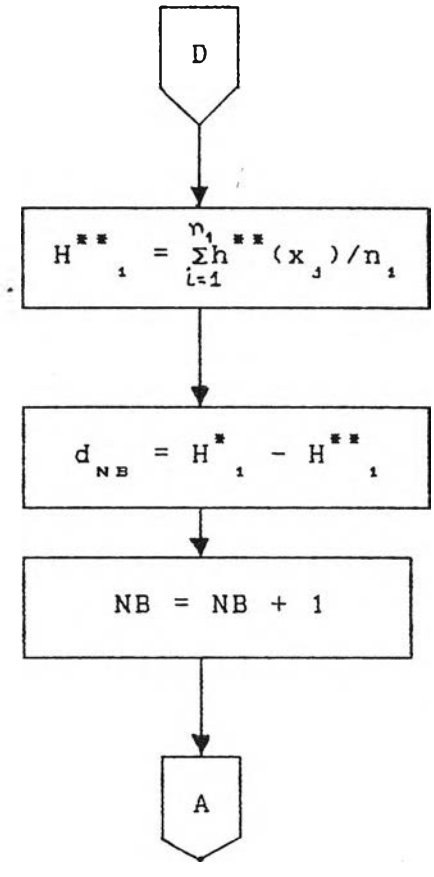



B
i > n1
Yes
No
A
ได้ตัวประมาณ P(2|1)
โดยวิธี U

รูปที่ 3.4 แสดงแผนผังงานสำหรับการหาค่าประมาณของอัตราความผิดพลาดที่มีเงื่อนไข โดยวิธี B




C
คำนวนหา $h^*(x_i)$ โดยวิธี R
i = i + 1
i > n_1
Yes
No
B
$H^*_1 = \sum_{i=1}^{n_1} h^*(x_i)/n_1$
j = 1
$x = x^*_{1j}$
หาค่า $W^{**}(x)$
หาค่า $h^{**}(x_j)$ โดยวิธี R
j = j + 1
j > n_1
No
Yes
D


D
$H^{**}_1 = \sum_{i=1}^{n_1} h^{**}(x_j)/n_1$
$d_{NB} = H^{*}_1 - H^{**}_1$
NB = NB + 1
A

รูปที่ 3.5 แสดงแผนผังงานสำหรับการหาค่าประมาณของอัตราความผิดพลาดที่มีเงื่อนไข โดยวิธี DS