Лабораторна робота №5

$$Y_j = 1 + 2X_j + \varepsilon_j, \quad n = \overline{1, 1000}$$

$$X_j \sim N(1, 1), \qquad \varepsilon_j \sim N(0, (X_j - 1)^2)$$

<u>n = 1000</u>

```
> A1<-c(); B1<-c()
> A2<-c(); B2<-c()
> N = 1000
> for(i in 1:N) {
+   X<-rnorm(n,1,1)
+   eps<- c()
+   for(x in X) eps<-c(eps, rnorm(1, 0, gfunc(x)))
+   Y<-1+2*X + eps
+   D<-data.frame(Y,X)
+   colnames(D)<-c("Y","X")
+   mod1<-lm(Y~X, data = D)
+   A1<-c(A1, mod1$coefficients[1])
+   B1<-c(B1, mod1$coefficients[2])
+   vv<-(mod1$residuals)^2
+   Dres<-data.frame(vv, X^2, X)
+   colnames(Dres)<-c("v","Xsq", "X")
+   mod2<-lm(vv~Xsq+X, data = Dres)
+   ww<-1/pmax(mod2$fitted.values, 1e-5)      #"зріз" від'ємних та близьких до нуля значень
прогнозу дисперсії
+   lmW<-lm(Y~X, weights = ww, data = D)
+   A2<-c(A2, lmW$coefficients[1])
+   B2<-c(B2, lmW$coefficients[2])
+ }
```

Рисунки на наступних сторінках зображають діаграми розсіювання даних та діаграми прогноз-залишки для вибірки з 1000 згенерованих елементів на першому (звичайний МНК) та другому (двокроковий адаптивний навантажений МНК) відповідно.

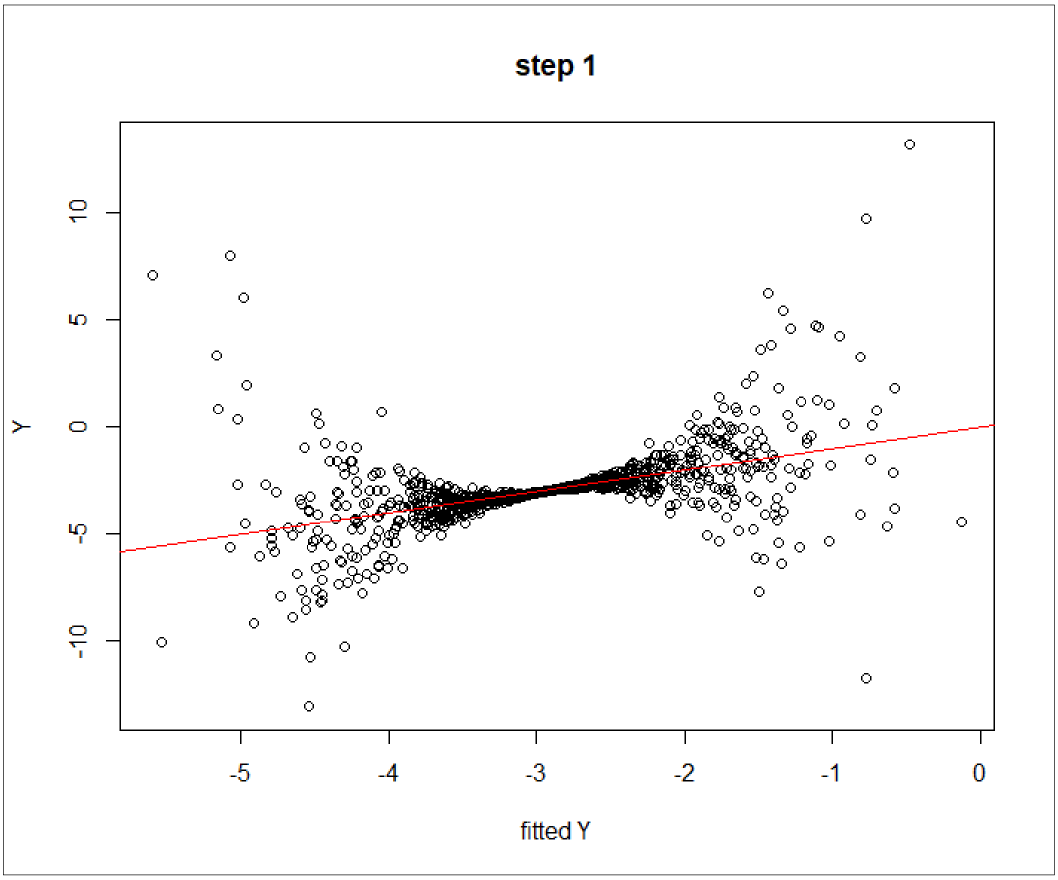
Далі – таблиці з даними оцінок для середнього та дисперсії розподілу коефіцієнтів a та b за допомогою 1000 повторів генерації даних та знаходження оцінок для цих коефіцієнтів на першому (A1, B1) та другому (A2, B2) кроці адаптивного МНК, а також гістограми відповідних розподілів.

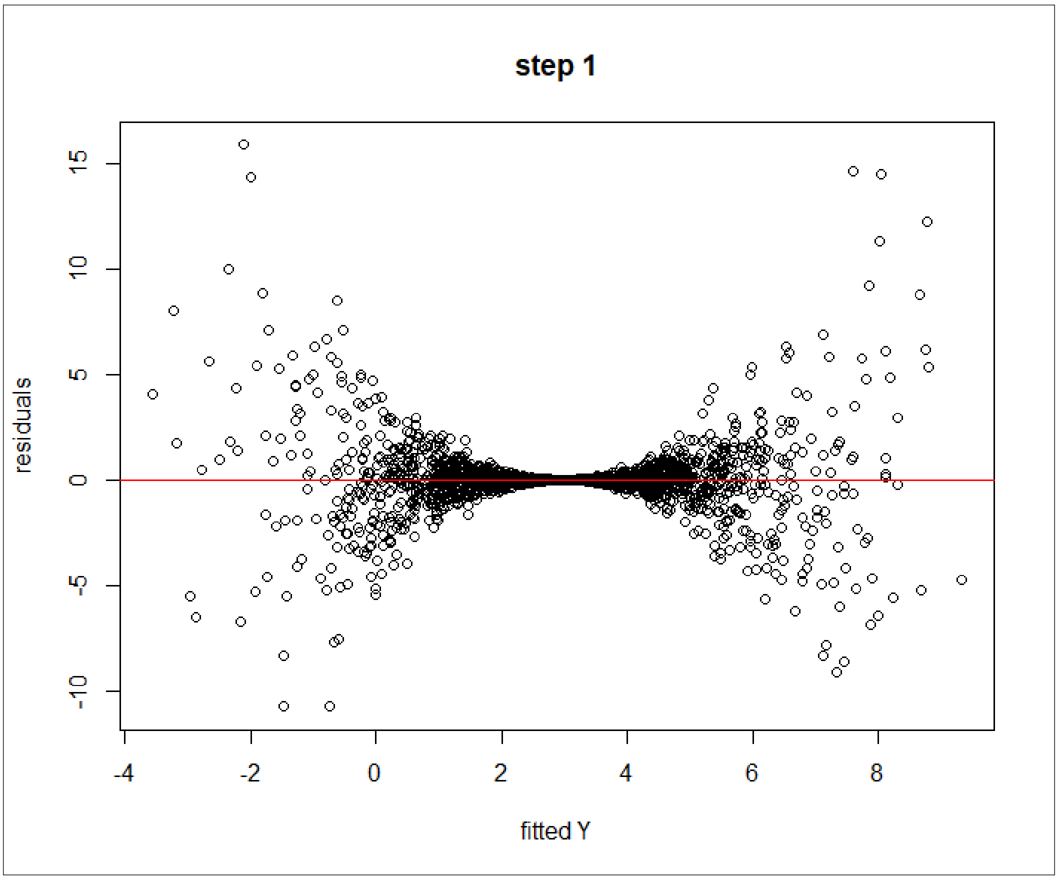
Зеленим кольором в таблицях позначено ближчі до справжніх значення коефіцієнтів a та b. При n = 25 та 50 звичайна МНК-модель дає кращі оцінки коефіцієнтів.
Але при збільшенні коефіцієнту «зрізу» (e.g. e-2 чи e-3) різниця між оцінками може стати менш значною. Наприклад, можна отримати таку таблицю:

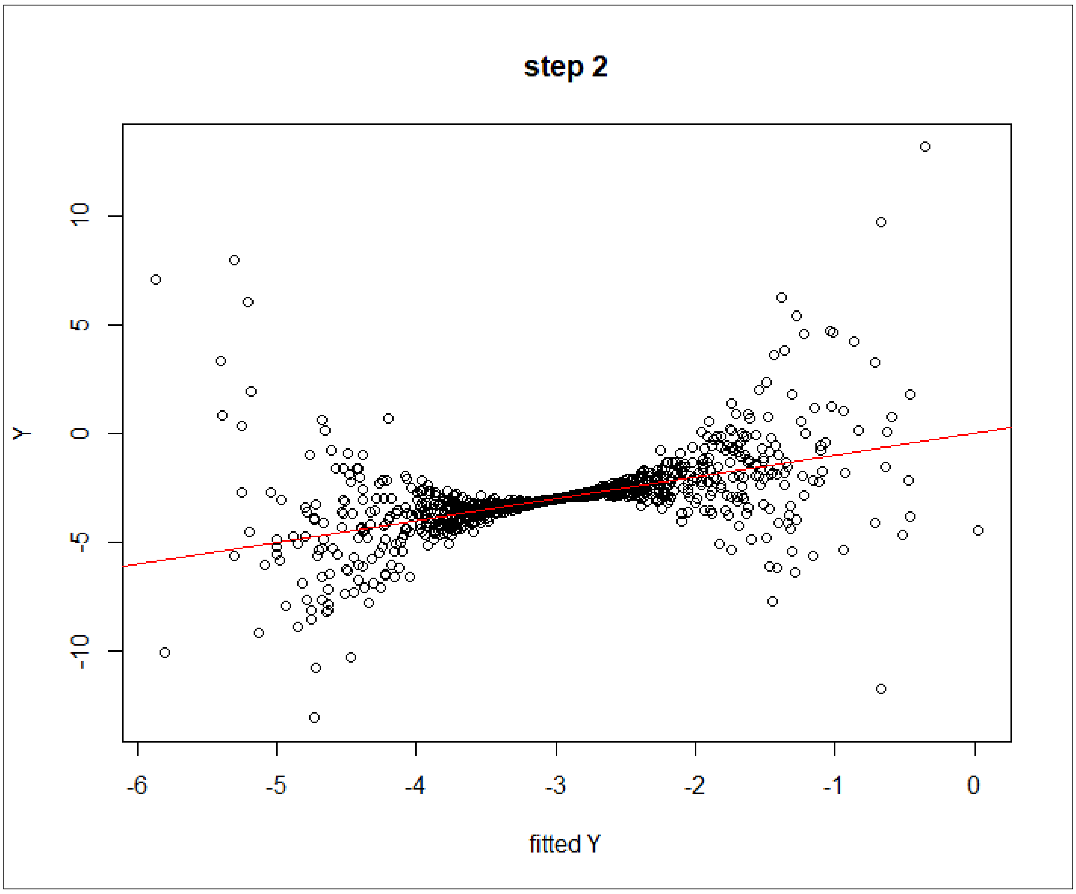
A	MEAN	SD
STEP 1	-2.000735	0.7423715
STEP 2	-1.993161	0.3607206

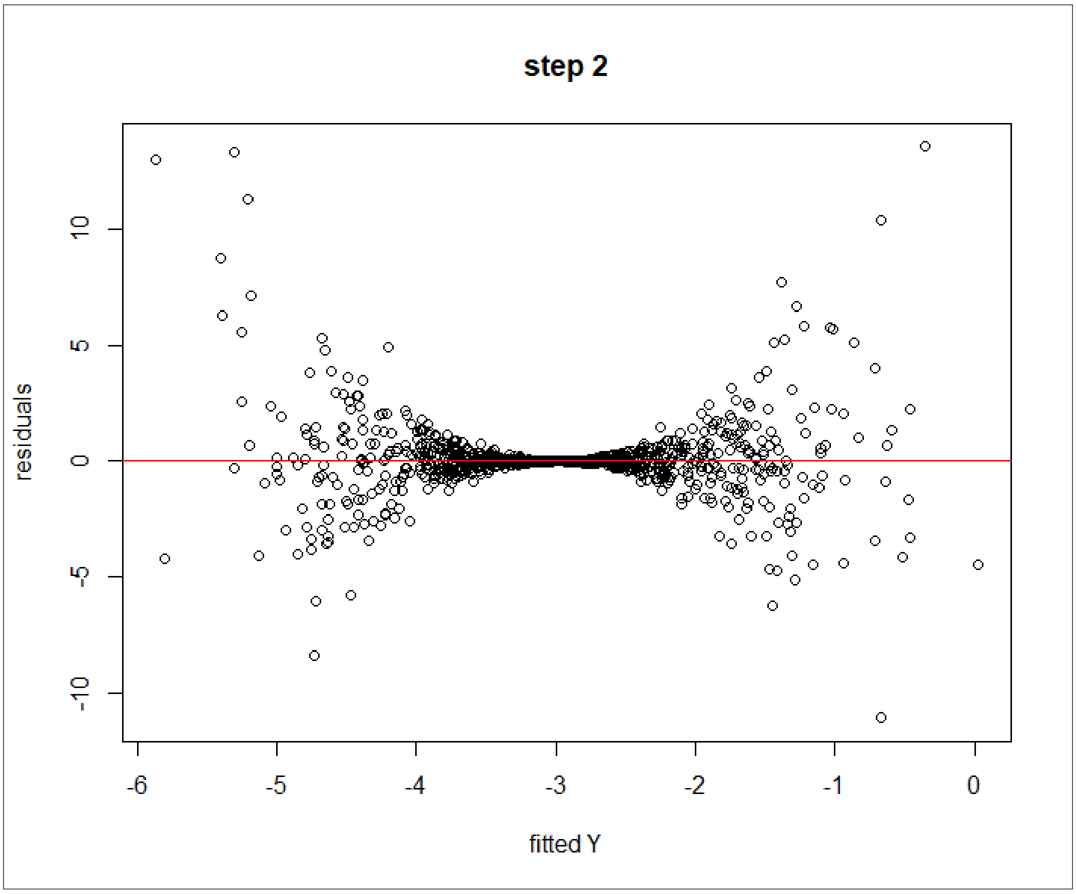
B	MEAN	SD
STEP 1	0.9892657	0.6914535
STEP 2	1.004716	0.3543287

Проте загалом результати роботи вказують на зростання доцільності використання саме адаптивного методу для знаходження коефіцієнтів при збільшенні об'єму вибірки.


step 1
Y
10
5
0
-5
-10
-5
-4
-3
-2
-1
0
fitted Y


step 1
residuals
15
10
5
0
-5
-10
-4
-2
0
2
4
6
8
fitted Y


step 2
Y
10
5
0
-5
-10
-6
-5
-4
-3
-2
-1
0
fitted Y


step 2
residuals
10
5
0
-5
-10
-6
-5
-4
-3
-2
-1
0
fitted Y

- n = 1000 ---

A	MEAN	SD
STEP 1	-2.006623	0.1392853
STEP 2	-2.001262	0.0375627

B	MEAN	SD
STEP 1	0.996739	0.1274625
STEP 2	0.9985423	0.03609452

- n = 25 ---

A	MEAN	SD
STEP 1	-1.991917	0.7457154
STEP 2	-2.01058	0.7859033

B	MEAN	SD
STEP 1	0.9986159	0.6625453
STEP 2	0.9617716	0.8999805

- n = 50 ---

A	MEAN	SD
STEP 1	-2.006501	0.5618257
STEP 2	-1.966082	0.9768374

B	MEAN	SD
STEP 1	0.9958356	0.5368918
STEP 2	0.9761962	0.8572907

- n = 100 ---

A	MEAN	SD
STEP 1	-2.000857	0.4059337
STEP 2	-2.005999	0.1556211

B	MEAN	SD
STEP 1	1.00169	0.3674382
STEP 2	0.9942298	0.1533314

- n = 500 ---

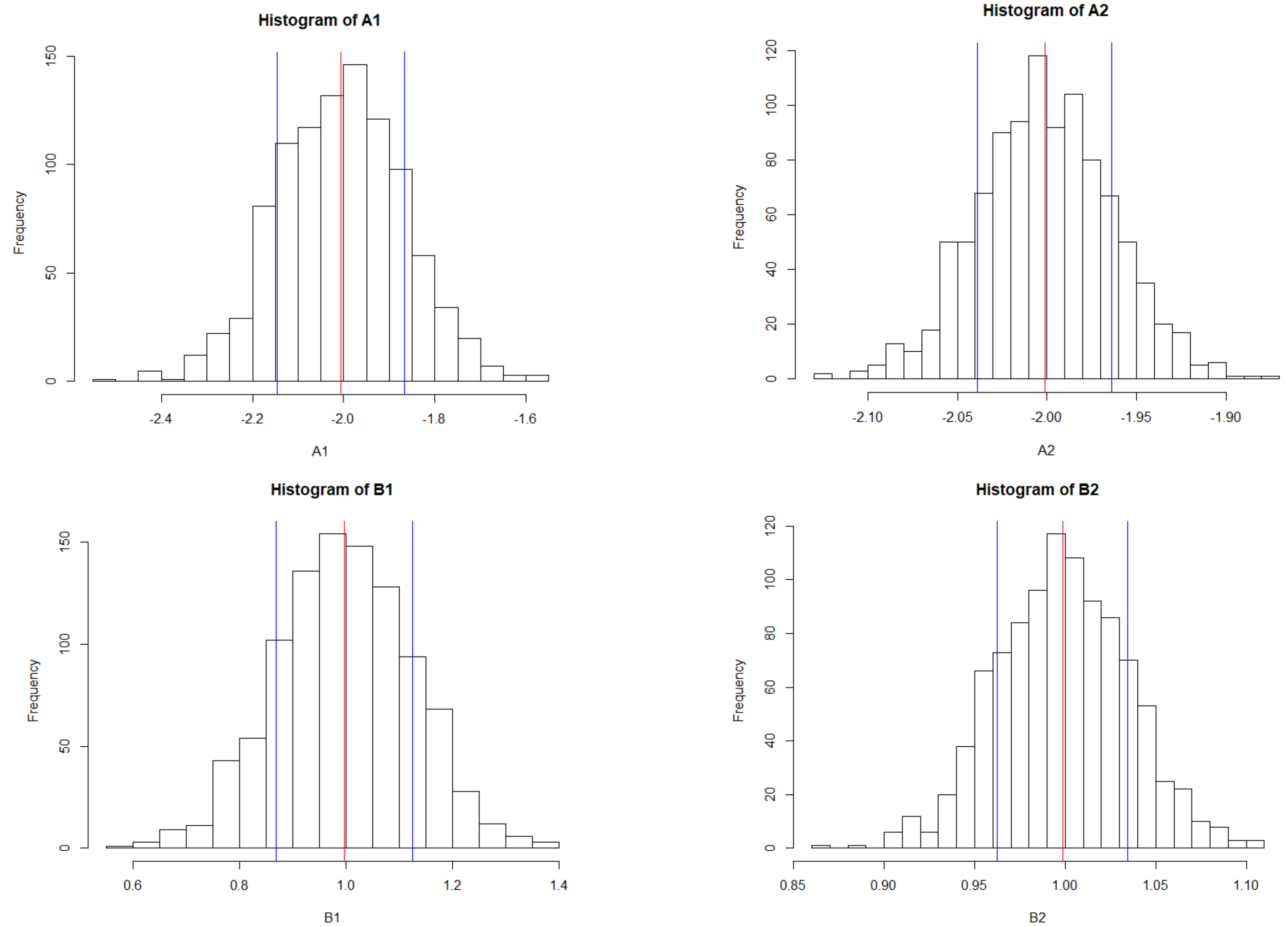
A	MEAN	SD
STEP 1	-1.995341	0.1850261
STEP 2	-2.000308	0.05290454

B	MEAN	SD
STEP 1	1.003684	0.1671175
STEP 2	0.9990279	0.05150707

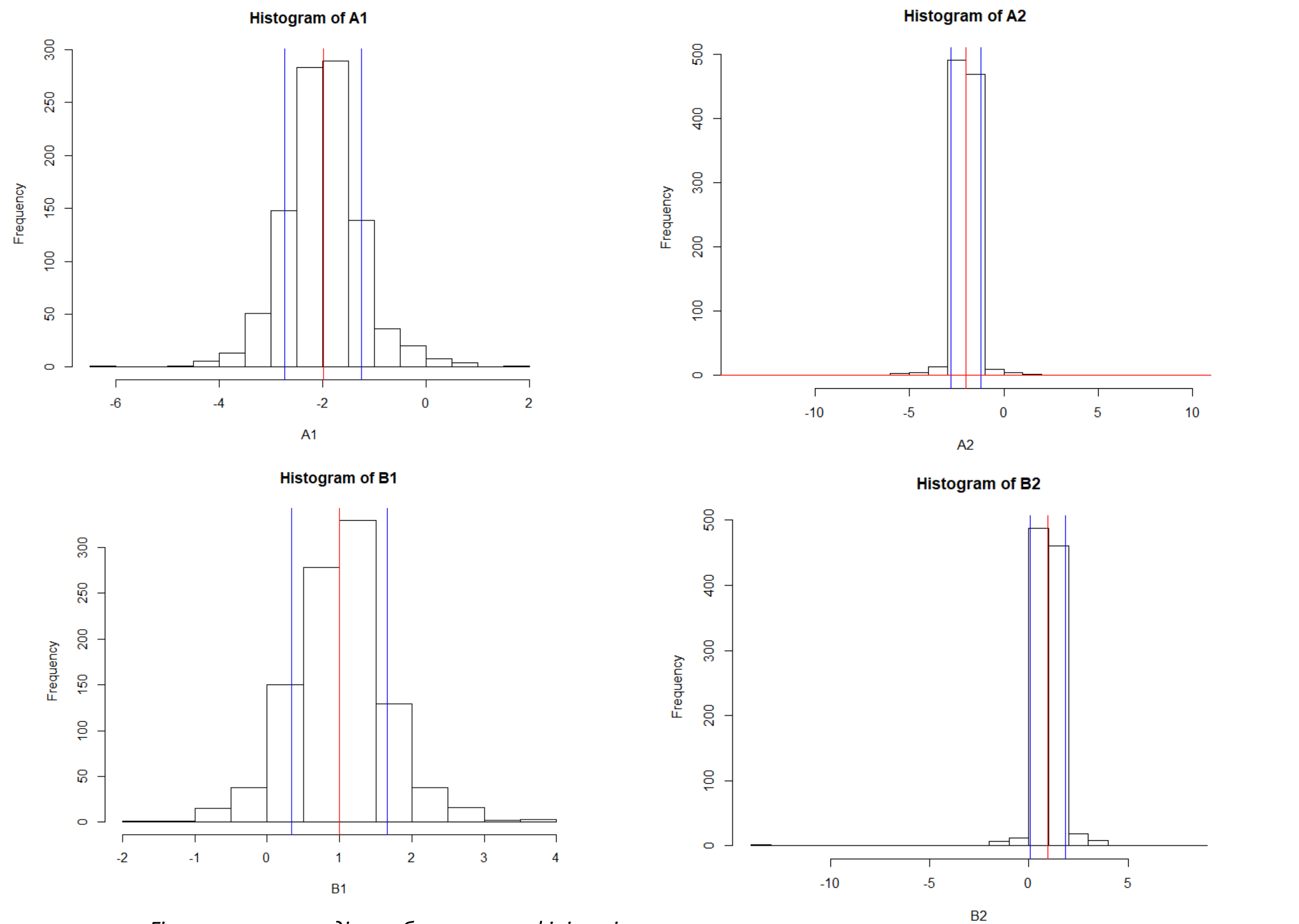
- n = 2000 ---

A	MEAN	SD
STEP 1	-2.00317	0.09441231
STEP 2	-2.000008	0.02582254

B	MEAN	SD
STEP 1	0.9969829	0.08613112
STEP 2	1.000239	0.02522969



Гістограми розподілу наближень коефіцієнтів a та b на 1-му та 2-му кроках при n = 1000



Гістограми розподілу наближень коефіцієнтів a т