

### Homework 7. Sample solution

1. The in- and out-of-control means are 10 and 13, so the allowance  $K$  is  $(13-10)/2 = 1.5$ . Since the sd is 2, this corresponds to  $k=1.52 = 0.75$ . For a two-sided IC ARL of 500, each of the upward and downward cusums needs an IC ARL of 1000. Going to the head-start table, this gives us  $h=3.556$ , leading to  $H=7.115$

2. The nearest for of the ARL table is that for  $h=3.5$ , with an IC ARL of 919. Copying out the numbers in this row gives

k	-0.25	0	0.25	0.5	0.75
ARL	7.87	16.56	47.6	185.9	918.7

We get the OOC ARL for a shift from  $\mu_0$  to  $\mu_t$  by changing  $k$  to the modified  $k-(\mu_t - \mu_0)/\sigma$ . Solving for  $\mu_t$  in this table then gives

Modified k	-0.25	0	0.25	0.5	0.75
Mean	12.0	11.5	11.0	10.5	10.0
ARL	7.87	16.56	47.6	185.9	918.7

3. The data, upward and downward cusums are:

n	1	X[n]	10	C+	2.06	C-	-2.06
n	2	X[n]	8	C+	0.00	C-	-2.56
n	3	X[n]	11	C+	0.00	C-	-0.06
n	4	X[n]	8	C+	0.00	C-	-0.56
n	5	X[n]	8	C+	0.00	C-	-1.06
n	6	X[n]	7	C+	0.00	C-	-2.56
n	7	X[n]	9	C+	0.00	C-	-2.06
n	8	X[n]	14	C+	2.50	C-	0.00
n	9	X[n]	9	C+	0.00	C-	0.00
n	10	X[n]	14	C+	2.50	C-	0.00
n	11	X[n]	13	C+	4.00	C-	0.00
n	12	X[n]	9	C+	1.50	C-	0.00
n	13	X[n]	13	C+	3.00	C-	0.00
n	14	X[n]	11	C+	2.50	C-	0.00
n	15	X[n]	14	C+	5.00	C-	0.00
n	16	X[n]	12	C+	5.50	C-	0.00
n	17	X[n]	12	C+	6.00	C-	0.00
n	18	X[n]	13	C+	7.50	C-	0.00
n	19	X[n]	12	C+	8.00	C-	0.00
n	20	X[n]	14	C+	10.50	C-	0.00

The downward cusum doesn't show anything striking. The upward cusum breaks through the control limit of 7.115 at observation number 18. The last time it was at zero was time 9, so we diagnose 9 as the last in-control time. The slope of the segment from (9,0) to (18,7.5) is 0.83, so we estimate the new mean as  $10 + 1.5 + 0.83 = 12.33$ .

## Lab portion

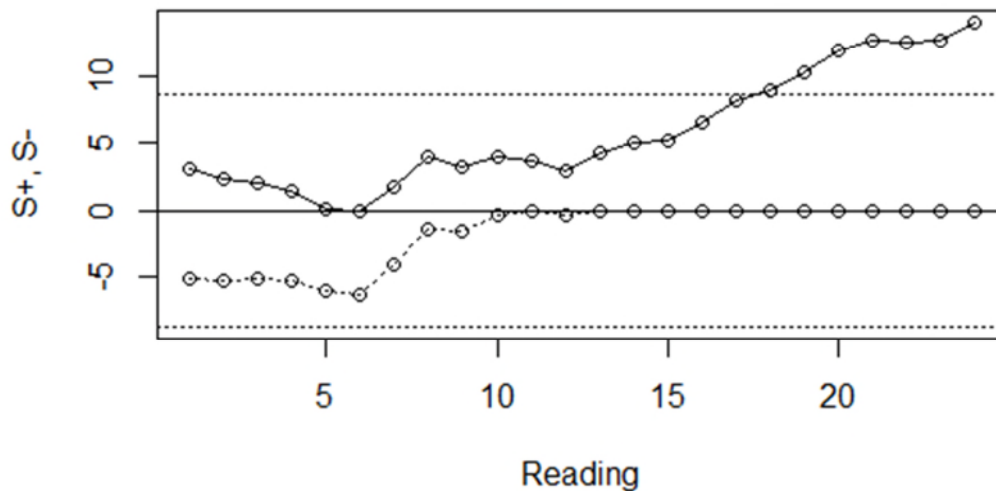
1. The cusums with the three parameter choices give these cusums:

Cusum with  $k=0.25, h=8.709$

n	1	U[n]	-1.00	C+	3.10	C-	-5.10
n	2	U[n]	-0.50	C+	2.35	C-	-5.35
n	3	U[n]	0.00	C+	2.10	C-	-5.10
n	4	U[n]	-0.50	C+	1.35	C-	-5.35
n	5	U[n]	-1.00	C+	0.10	C-	-6.10
n	6	U[n]	-0.50	C+	0.00	C-	-6.35
n	7	U[n]	2.00	C+	1.75	C-	-4.10
n	8	U[n]	2.50	C+	4.00	C-	-1.35
n	9	U[n]	-0.50	C+	3.25	C-	-1.60
n	10	U[n]	1.00	C+	4.00	C-	-0.35
n	11	U[n]	0.00	C+	3.75	C-	-0.10
n	12	U[n]	-0.50	C+	3.00	C-	-0.35
n	13	U[n]	1.50	C+	4.25	C-	0.00
n	14	U[n]	1.00	C+	5.00	C-	0.00
n	15	U[n]	0.50	C+	5.25	C-	0.00
n	16	U[n]	1.50	C+	6.50	C-	0.00
n	17	U[n]	2.00	C+	8.25	C-	0.00
n	18	U[n]	1.00	C+	9.00	C-	0.00
n	19	U[n]	1.50	C+	10.25	C-	0.00
n	20	U[n]	2.00	C+	12.00	C-	0.00
n	21	U[n]	1.00	C+	12.75	C-	0.00
n	22	U[n]	0.00	C+	12.50	C-	0.00
n	23	U[n]	0.50	C+	12.75	C-	0.00
n	24	U[n]	1.50	C+	14.00	C-	0.00

Signal 18 last estimated IC 6 new mean estimate 1.00

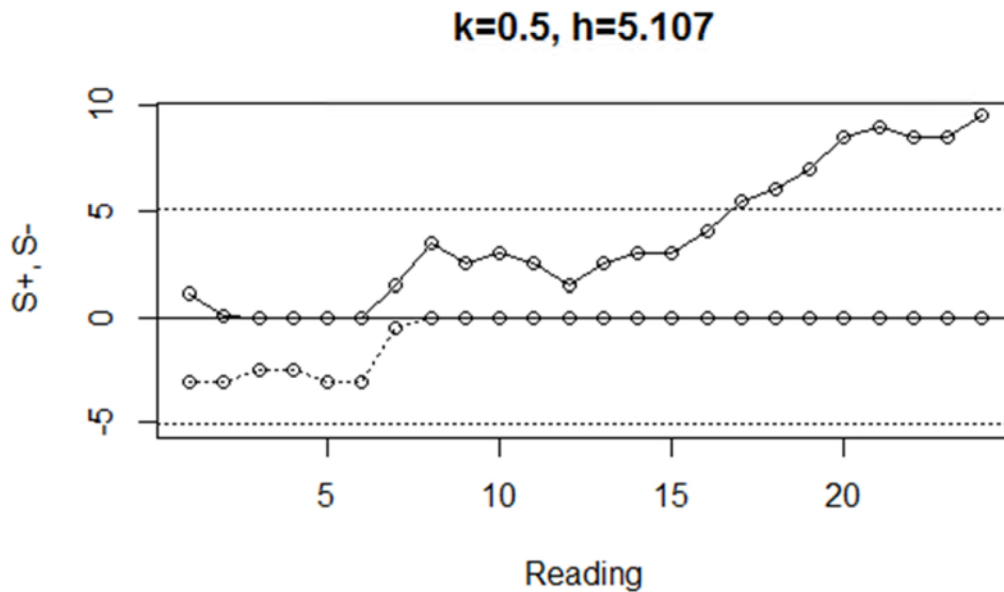
**$k=0.25, h=8.709$**



Cusum with  $k=0.5, h=5.107$

n	1	U[n]	-1.00	C+	1.05	C-	-3.05
n	2	U[n]	-0.50	C+	0.05	C-	-3.05
n	3	U[n]	0.00	C+	0.00	C-	-2.55
n	4	U[n]	-0.50	C+	0.00	C-	-2.55
n	5	U[n]	-1.00	C+	0.00	C-	-3.05
n	6	U[n]	-0.50	C+	0.00	C-	-3.05
n	7	U[n]	2.00	C+	1.50	C-	-0.55
n	8	U[n]	2.50	C+	3.50	C-	0.00
n	9	U[n]	-0.50	C+	2.50	C-	0.00
n	10	U[n]	1.00	C+	3.00	C-	0.00
n	11	U[n]	0.00	C+	2.50	C-	0.00
n	12	U[n]	-0.50	C+	1.50	C-	0.00
n	13	U[n]	1.50	C+	2.50	C-	0.00
n	14	U[n]	1.00	C+	3.00	C-	0.00
n	15	U[n]	0.50	C+	3.00	C-	0.00
n	16	U[n]	1.50	C+	4.00	C-	0.00
n	17	U[n]	2.00	C+	5.50	C-	0.00
n	18	U[n]	1.00	C+	6.00	C-	0.00
n	19	U[n]	1.50	C+	7.00	C-	0.00
n	20	U[n]	2.00	C+	8.50	C-	0.00
n	21	U[n]	1.00	C+	9.00	C-	0.00
n	22	U[n]	0.00	C+	8.50	C-	0.00
n	23	U[n]	0.50	C+	8.50	C-	0.00
n	24	U[n]	1.50	C+	9.50	C-	0.00

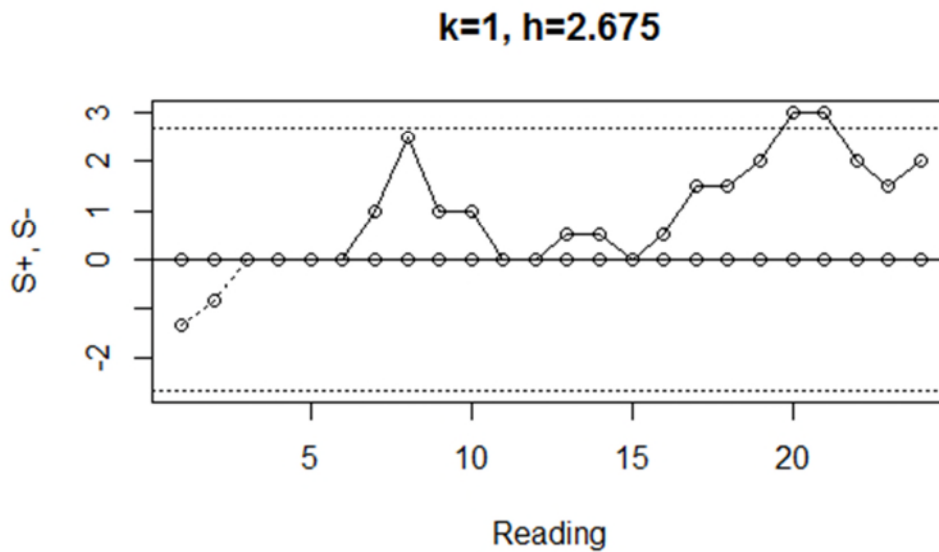
Signal 17 last estimated IC 6 new mean estimate 1.00



Cusum with  $k=1, h=2.675$

n	U[n]	C+	C-	new mean estimate
1	-1.00	0.00	-1.34	
2	-0.50	0.00	-0.84	
3	0.00	0.00	0.00	
4	-0.50	0.00	0.00	
5	-1.00	0.00	0.00	
6	-0.50	0.00	0.00	
7	2.00	1.00	0.00	
8	2.50	2.50	0.00	
9	-0.50	1.00	0.00	
10	1.00	1.00	0.00	
11	0.00	0.00	0.00	
12	-0.50	0.00	0.00	
13	1.50	0.50	0.00	
14	1.00	0.50	0.00	
15	0.50	0.00	0.00	
16	1.50	0.50	0.00	
17	2.00	1.50	0.00	
18	1.00	1.50	0.00	
19	1.50	2.00	0.00	
20	2.00	3.00	0.00	
21	1.00	3.00	0.00	
22	0.00	2.00	0.00	
23	0.50	1.50	0.00	
24	1.50	2.00	0.00	

Signal 20 last estimated IC 15 new mean estimate 1.60



The first two cusums paint almost the same picture; the second cusum signals one reading earlier than the first. They give the same estimate of the time of change and (to two decimals) of the new mean. The third cusum though signals a bit later, and estimates the time of the shift much later than the first two. Its estimate of the new mean is a lot higher than the first two. This gives an indication of the fact that cusums tuned for different shifts will give somewhat different analyses of the data.

While we should not read too much into this, the first two cusums estimate the shift as one standard deviation, and  $k=0.5$  gives the optimal cusum for this shift. So it may not be surprising that the  $k=0.5$  cusum is the first of the three to signal a shift.

## 2. The first of the three GETH runs gives

Program to calculate cusum decision intervals

Copyright 1997, D M Hawkins and D H Olwell

Which distribution do you want?

Normal location

Give the in-control mean and standard deviation

10.000 2.000

Give the out-of-control mean

12.000

The cusum offset is 11.000

What are the Winsorizing constants?

-999.00 999.00

Want zero-start (say Z), steady state (say S) or FIR (say F)?

F

Enter ARL

1000.0000

DI 10.214, IC ARL 1000.2, OOC ARL Zero start 10.6, FIR 6.5, SS 6.3

So to get the two-sided FIR cusum with an IC ARL of 500, we need  $K=1$  and  $H=10.214$ . The OOC ARL is 10.6 for the zero-start, 6.5 for the FIR and 6.3 for the steady state.

The other two designs give

OOC mean	cusum offset	H	Zero	FIR	SS
12	11	10.214	10.6	6.5	6.3
11	10.5	17.418	31.6	19.4	20.1
14	12	5.350	3.4	2.2	1.6

Note that all three FIR OOC ARLs are smaller than their zero-start counterparts, illustrating the benefit you get from giving the cusum a head start.

The first GETARL run produces the output

Program to calculate cusum average run lengths

Copyright 1997, D M Hawkins and D H Olwell

Which distribution do you want?

Normal location

Enter the mean and standard deviation

10.5000 2.0000

Enter the offset  $\mu_0+K$  and decision interval H:

11.000 10.214

The cusum is assumed upward

What are the Winsorizing constants?

-999.00 999.00

Standardized k and h values

0.2500 5.1070

zero start, FIR, steady state ARLs 150.84 133.28 142.26

from which we extract the FIR OOCARL of 133.28.

Doing all 21 GETARL runs and collating the results give the table

	Design assuming		
$\mu_1$	$\mu_1 = 12$	$\mu_1 = 11$	$\mu_1 = 14$
10.5	133.28	73.37	257.69
11.0	29.76	19.39	75.21
11.5	11.48	10.13	26.19
12.0	6.46	6.86	11.30
12.5	4.45	5.23	6.04
13.0	3.43	4.25	3.85
13.5	2.81	3.60	2.79

This table shows what we expect. For OOC means up to 11.5, the second cusum with its assumption of an OOC mean of 11, has the lowest OOC ARL. For OOC means of 12.0 through 13 the first cusum, with its assumption of an OOC mean of 12, has the lowest OOC ARL, and for an OOC mean of 13 the third cusum, which assumes an OOC mean of 14, is the best of the three.

One of the things we learn in this lab is that, while each cusum is the optimal method of detecting the shift for which it is designed, a cusum with different design assumptions can give lower power and also a somewhat different interpretation of the data.

```
# R code for some of homework 7

X <- c(10, 8, 11, 8, 8, 7, 9, 14, 9, 14, 13, 9, 13, 11, 14, 12, 12, 13, 12, 14)
mu0 <- 10
sigma <- 2
K <- 1.5
k <- K / sigma
h <- 3.556
H <- sigma * h

stocuup <- NULL
stocudn <- NULL
cuup <- 0.5 * H
cudn <- -0.5 * H

lenof <- length(X)
for (n in 1:lenof) {
  cuup <- max(c(0, cuup + X[n] - mu0 - K))
  stocuup <- c(stocuup, cuup)
  cudn <- min(c(0, cudn + X[n] - mu0 + K))
  stocudn <- c(stocudn, cudn)
  cat(sprintf("n %4.0f X[n] %4.0f C+ %6.2f C- %6.2f \n", n, X[n], cuup, cudn))
}

eps <- 1.e-5

signal <- min((1:lenof)[stocuup > H])
lastIC <- max((1:signal)[abs(stocuup) < eps])
slope <- stocuup[signal] / (signal - lastIC)
newmu <- mu0 + K + slope
cat(sprintf("signal %4.0f last estimated IC %4.0f slope %7.2f new mean %7.2f \n",
```

```

    signal, lastIC, slope, newmu))

# Lab portion

oil <- read.csv("http://stat.umn.edu/hawkins/5031/Table_9E_2_Oil_additive.csv")
head(oil)

U      <- (oil$Fill - 8.02) / 0.02
lenof <- length(U)
k      <- c(0.25 , 0.5 , 1.0 )
h      <- c(8.709, 5.107, 2.675)

labls <- c("k=0.25, h=8.709", "k=0.5, h=5.107", "k=1, h=2.675")

for (cusm in 1:3) {
  cat(sprintf("\nCusum with %20s \n", labls[cusm]))
  stocuup <- NULL
  stocudn <- NULL
  cuup    <- 0.5 * h[cusm]
  cudn    <- -cuup
  for (n in 1:lenof) {
    cuup <- max(c(0, cuup + U[n] - k[cusm]))
    cudn <- min(c(0, cudn + U[n] + k[cusm]))
    stocuup <- c(stocuup, cuup)
    stocudn <- c(stocudn, cudn)
    cat(sprintf("n %4.0f U[n] %6.2f C+ %6.2f C- %6.2f \n",
      n, U[n], cuup, cudn))
  }
  upsignal <- min((1:(lenof+1))[c(stocuup,1e10)> h[cusm]])
  dnsignal <- min((1:(lenof+1))[c(stocudn,-1e10)< -h[cusm]])
  signal <- min(c(upsignal, dnsignal))
  if(signal <= lenof) { # have a signal. Diagnose
    if (upsignal < dnsignal) { # is upward
      lastIC <- max((1:signal)[abs(stocuup) < eps])
      newmean <- k[cusm] + stocuup[signal]/(signal-lastIC)
    } else {
      lastIC <- max((1:signal)[abs(stocudn) < eps])
      newmean <- k[cusm] + stocudn[signal]/(signal-lastIC)
    }
    cat(sprintf("Signal %4.0f last estimated IC %4.0f new mean estimate %6.2f \n",
      signal, lastIC, newmean))
  }
  ylim <- c(min(c(-h[cusm], stocudn)), max(c(h[cusm], stocuup)))
  dev.new()
  plot(1:lenof, stocuup, ylim=ylim, main=labls[cusm], xlab="Reading",
    ylab="S+, S-", type="o")
  lines(1:lenof, stocudn, type="o", lty=3)
  abline(h=0)
  abline(h=h[cusm], lty=3)
  abline(h=-h[cusm], lty=3)
}

```