Change Detect (3)

Ronglong Fang

1 Change Definition

A sequence of real value $x_1, x_2, x_3, \ldots, x_t, \ldots$. The value of x_t is available only at time t. Each x_t is generated according to some distribution D_t with mean μ_t and variance σ_t^2.

The following describe is from the book [1]

> If there is no change in the system, and the model is correct, then the residuals are so-called white noise, that is a sequence of independent stochastic variable with zero mean and known variance. After the change either the mean or variance or both changes, that is, the residuals becomes 'larges' in some sense. The main problem in statistical change detection is to decide what 'large' is.

Figure 1 and 2 shows the change detect based on mean and variance, which are obtain [1].

Definition 1. We call change happens at x_t if the mean μ_t is different from μ_{t-1} or the variance σ_t^2 is different from σ_{t-1}^2.

Here are some remark :

1. If no change happens, the mean and variance will be a constant. It does not mean the sequence $x_1, x_2, \ldots$ is a constant.
2. In this stage, we only focus on the mean μ_t.
3. If change happens suddenly, then the mean will change suddenly. If change happens gradually, then the mean will change gradually.
4. Instead of using the sequence directly, we should use the mean of the sequence. Since the mean is unknow, we can use the local sample mean, which is defined as

$$\bar{x}_t = \frac{1}{2m+1} \sum_{k=-m}^{m} x_{t+k} \tag{1}$$

   where the parameter m is giving by the Hoeffding inequality.

Lemma 2. (Hoeffding's Inequality). Let $Z_1, \ldots, Z_m$ be a sequence of i.i. d. random variables and let $\bar{Z} = \frac{1}{m}\sum_{i=1}^{m} Z_i$. Assume that $\mathbb{E}[\bar{Z}] = \mu$ and $\mathbb{P}[a \le Z_i \le b] = 1$ for every i. Then, for any $\epsilon > 0$

$$\mathbb{P}\left[\left|\frac{1}{m}\sum_{i=1}^{m} Z_i - \mu\right| > \epsilon\right] \le 2\exp\left(-2m\epsilon^2/(b-a)^2\right)$$

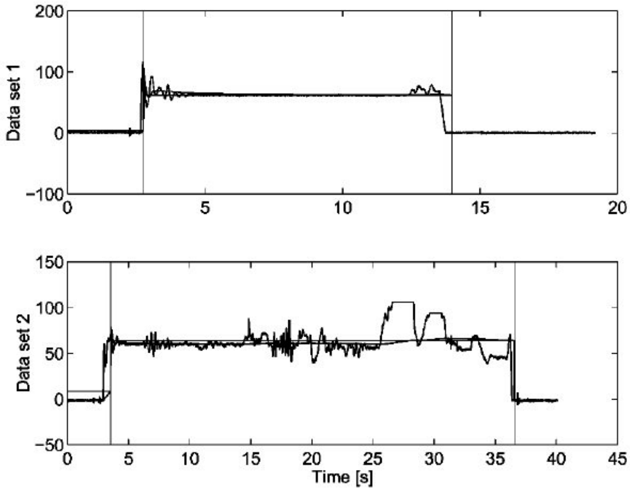


Figure 1: the change detection based on mean

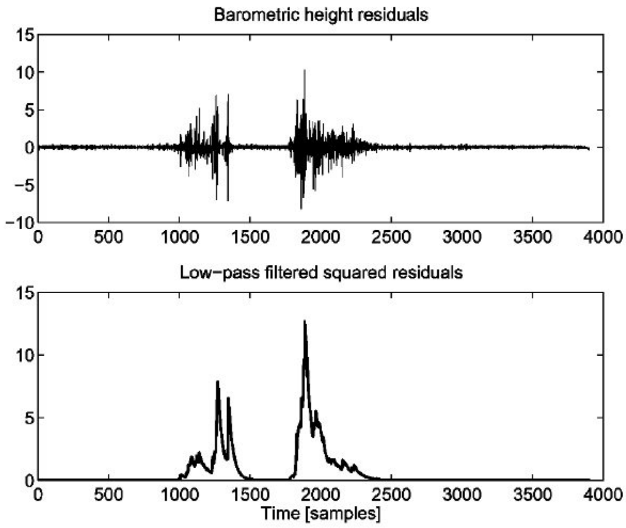


Figure 2: change detection based on variance

If no changes happens, the sequence $x_1, x_2, \ldots$ can be seen as the i.i. d. sequence. Giving confidence parameter $\delta \in [0,1]$, if we use local mean (1), then the Hoeffding's inequality

$$\mathbb{P}\left[\left|\frac{1}{2m+1}\sum_{i=-m}^{m} Z_i - \mu\right| > \epsilon\right] \leq 2\exp\left(-2(2m+1)\epsilon^2/(b-a)^2\right) \leqslant \delta$$

If we assume the sequence is the real number in $[0,1]$, we can give an lower estimate of m,

$$\boxed{m \geqslant -\frac{1}{6\epsilon^2}\text{In}(\delta/2)}$$

For the continuous function $g(x)$, we use the test function $f(x)$ to test the change. Usually, we assume $f(x)$ has the vanishing moment property. Vanishing moment property is key property we used in change detection. For continuous function $g(x)$, we check the inner product,

$$\langle g, f\rangle \tag{2}$$

For discrete data, we use summation instead of integration. For discrete data,

$$[d_0, d_1, d_2, \ldots, d_{n-1}, d_n]$$

we choose

$$[x_0, x_1, x_2, \ldots, x_{n-1}, x_n]$$

where $x_k = \frac{k}{n}$. We check the summation,

$$\sum_{k=0}^{n} d_k f(x_k) * \frac{1}{n} \tag{3}$$

Let $d_k = g(x_k)$, the summation (3) becomes

$$\sum_{k=0}^{n} g(x_k) f(x_k) * \frac{1}{n} \tag{4}$$

Actually, we use summation (4) to approximation integration (2).

Lemma 3. If f and g defined on $[0,1]$ satisfy the Lipschitz continuous with Lipschitz constant C, then the summation (4) convergence to the integration (2) and has the convergence rate $O(1/n)$.

Proof.

$$\begin{aligned}
&\Big|\sum_{k=0}^{n} g(x_k)f(x_k) * \frac{1}{n} - \int_0^1 f(x)g(x)dx\Big| \\
=&\Big|\sum_{k=1}^{n}\int_{x_{k-1}}^{x_k} g(x_k)f(x_k) - f(x)g(x)dx\Big| \\
\leqslant&\sum_{k=1}^{n}\Big|\int_{x_{k-1}}^{x_k} f(x_k)\big(g(x_k)-g(x)\big) + g(x)\big(f(x_k)-f(x)\big)\Big|dx \\
\leqslant&\sum_{k=1}^{n}\int_{x_{k=1}}^{x_k} |f(x_k)|C\frac{1}{n} + |g(x)|C\frac{1}{n}dx \\
\leqslant&\frac{1}{n}C(\|f\|_\infty + \|g\|_\infty)
\end{aligned}$$

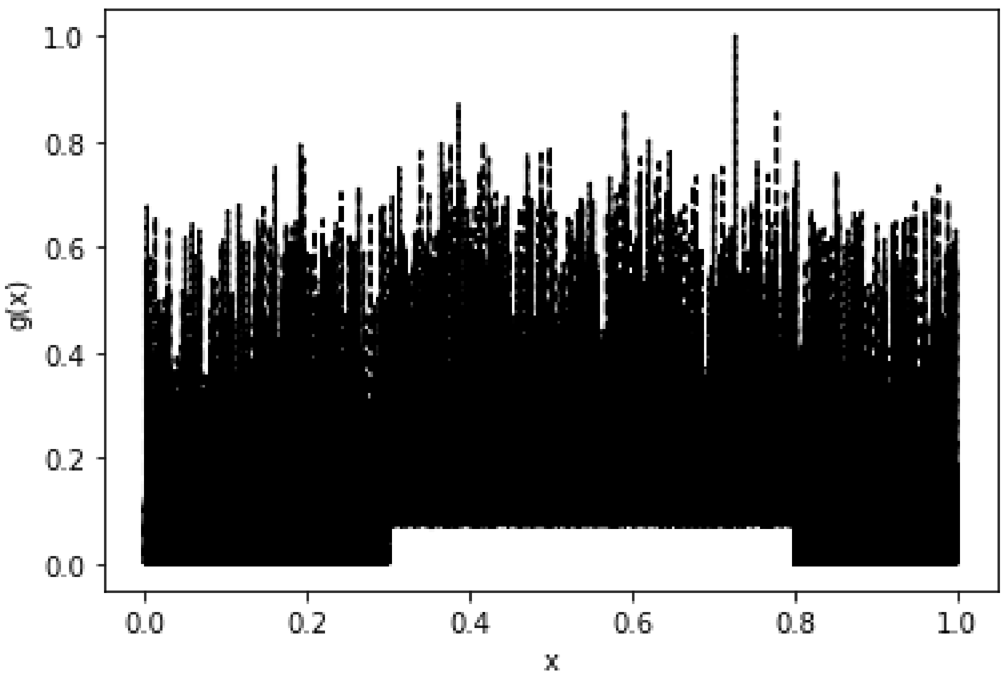


Figure 3: Sudden change data

□

Thus from Lemma 3, for giving an error $\epsilon \geqslant 0$,

$$\left|\sum_{k=0}^{n} g(x_k)f(x_k) * \frac{1}{n} - \int_0^1 f(x)g(x)dx\right| \leqslant \frac{1}{n}C(\|f\|_\infty + \|g\|_\infty) \leqslant \epsilon$$

then n satisfy

$$\boxed{n \geqslant \frac{C(\|f\|_\infty + \|g\|_\infty)}{\epsilon}}$$

2 Adwin2 algorithm and Multiscale algorithm

2.1 Suddenly Change Detect

data : 100000 change start : 30000 change end : 80000. The data is shown in Figure (3).
For the Adwin2 algorithm, we choose the parameter:

threshold : 2.0
delta : 0.01
change value : 2000
min time : 5000

For Multiscale algorithm, we choose parameter :

change level : N = 8
shrink level : k = 2
loacal mean length : K = 0, we doesn't need to smooth for the sudden change
integrate : I = 2000

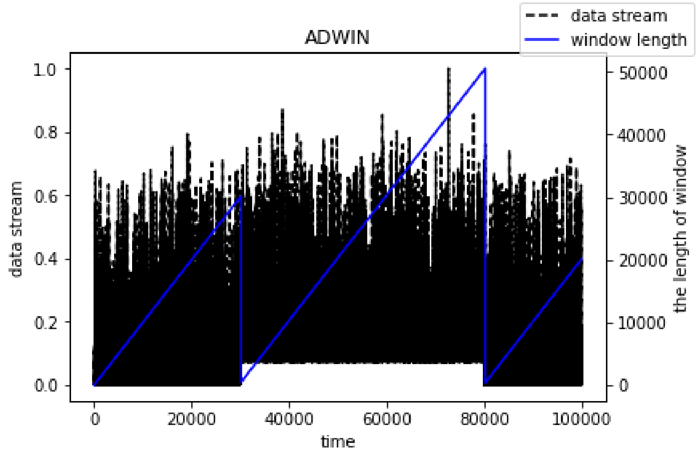


Figure 4: The result of Adwin2 for sudden change

extend : Ex = 10
threshold : epsilon = 0.7
The result of Adwin2 and Multiscale is shown in Table (1). Figure (4) shows the result of Adwin2 for sudden change.

Table 1: Sudden change detection : ADWIN2 and Multiscale

	30000	80000	time/s
ADWIN2	30154 30155 30156	80220 80221	52.01
Multiscale	29995	79995	10.76

2.2 Gradually Change Detect

data : 100000 change start : 30000 change end : 80000.
For the Adwin2 algorithm, we choose the parameter:
threshold : 2.0
delta : 0.01
change value : 2000
min time : 5000
For Multiscale algorithm, we choose parameter :
change level : N = 8
shrink level : k = 2
local mean length : K = 1001
integrate : I = 2000
extend : Ex = 10
threshold : epsilon = 0.8

The result is of Adwin2 and Multiscale is shown in Table (2). Figure (5) shows the result of Adwin2 for sudden change.

The advantage for Adwin2 algorithm are

1. Easy to detect the online data,

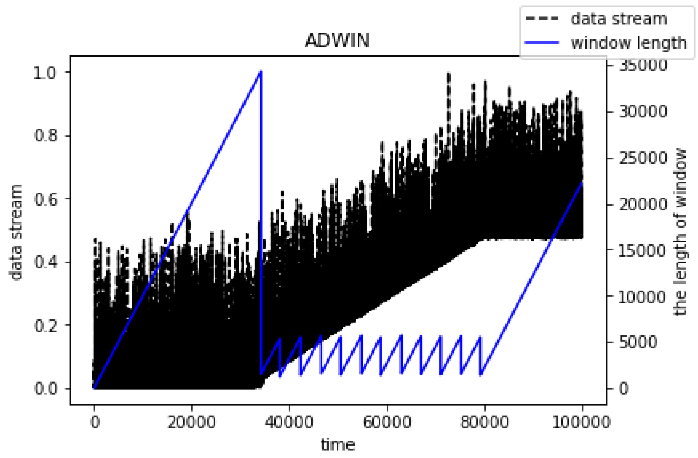


Figure 5: The result of Adwin2 for graduall change

Table 2: Gradual Change Detect : ADWIN2 and Multiscale

	30000	80000	time/s
ADWIN2	34261 38109 42363 46628 50577 54821	63036 58734 67024 71091 75248 79222	48.3125
Multiscale	28906 30320	80468	12.4155

2. Robust to the noise,
3. Efficient.

The disadvantage for Adwin2 algorithm are

1. Limited to the gradual change.

The advantage for the Multiscale algorithm

1. Can detect the gradually change
2. Efficient

The disadvantage for the Multiscale algorithm

1. Difficult to detect the online data,
2. Sensitive to the noise. We can use the mean of data instead of using data directly, we result would be much better.

3 Mixed Change Detect

In this section, we study the change happens both suddenly and gradually, we use the test function with order one vanishing moment to detect the sudden change and use the test

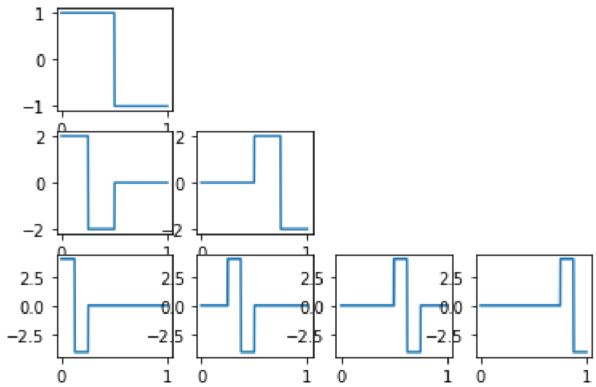


Figure 6: test function with order one vanishing moment

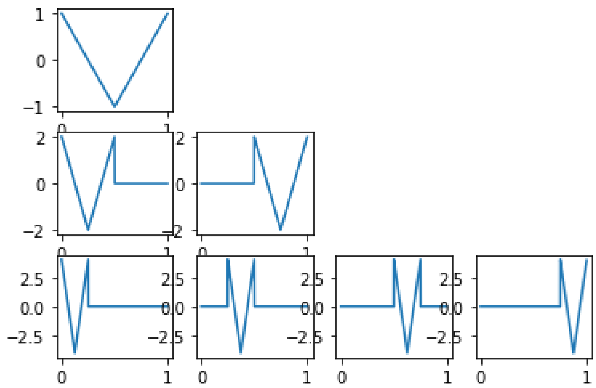


Figure 7: test function with order two vanishing moment

function with order two vanishing moment to detect the gradual change.

$$f(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1/2 \\ -1, & 1/2 < x \leqslant 1 \end{cases} \tag{5}$$

$$f(x) = \begin{cases} 1 - 4x, & x \in [0, \frac{1}{2}] \\ 4x - 3, & x \in (\frac{1}{2}, 1] \end{cases} \tag{6}$$

To detect when and where the changes happens, we check the amplitude of the the summation (4). Actually, when the change level are the same, for sudden change and gradual change the amplitude of the the summation (4) are different. Based on this fact, we can tell when change happens suddenly and when change happens gradually.

First, we check

$$\sum_{k=1}^{2n} g(x_k) f(x_k) * \frac{1}{2n} \tag{7}$$

and f is given by (5).

$$g(x) = \begin{cases} C_1, & 0 \leq x \leq 1/2 \\ 2(C_2 - C_1)x - C_2 + 2C_1, & 1/2 < x \leq 1 \end{cases} \tag{8}$$

If f is given by (6), we have that There are some property about the summation (7).

the type of g	cons (C)	cons to grad (8)	grad with slope $(C_2 - C_1)$	sudden from C_1 to C_2
summation (7)	0	$\frac{C_2-C_1}{4} + \frac{C_2-C_1}{4n}$	$\frac{(C_2-C_1)}{4}$	$\frac{(C_2-C_1)}{2}$

Table 3: Test function f given by (5). The amplitude of summation with different cases

the type of g	cons (C)	cons to grad (8)	grad with slope $(C_2 - C_1)$	sudden from C_1 to C_2
summation (7)	0	$\frac{C_2-C_1}{12} - \frac{C_2-C_1}{4n} + \frac{C_2-C_1}{6n^2}$	$\frac{C_2-C_1}{4n}$	$\frac{C_2-C_1}{2n}$

Table 4: Test Function give by (6). The amplitude of summation with different cases

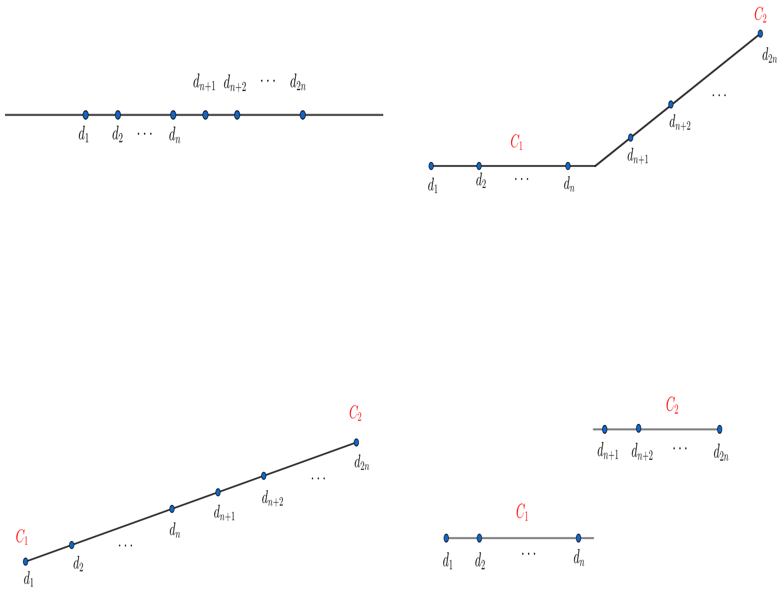


Figure 8: change types : constant, change from constant to gradual, gradual change, change from gradual to sudden

If change happens, we assume the initial state is C_1 and the final state C_2. For the test function (5), from Table (3), we observe that,

- when no change happens, the summation is 0. When gradual change happens the summation is approximately $\frac{C_2-C_1}{4}$, when sudden change happens the summation is $\frac{C_2-C_1}{2}$. Based on this fact, we change distinct the sudden change from the summation.
- When we detect from a large scale to a small scale, for the sudden change , the amplitude of the summation will change change, however, for the gradual change, the amplitude of the summation will becomes small and small since $C_2 - C_1$ will become small and small.

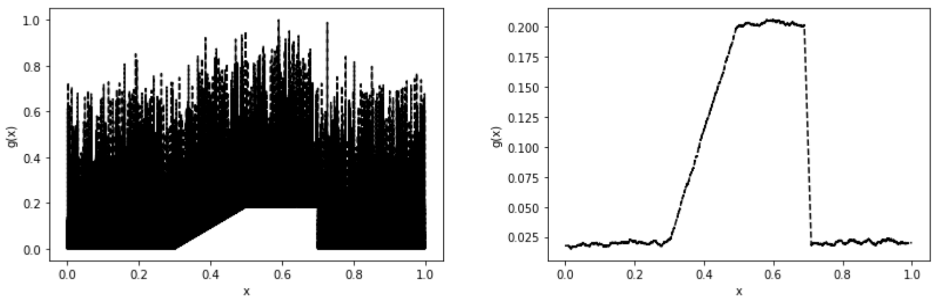
For the test function (6), from Table (4), we observe that,

- When no change happens, the summation is 0. When sudden change happens, if the change happens in the middle of the data, the summation is approximate to 0, however if the sudden change does not happens in the middle of the, we summation will change dramatically. Thus, for the sudden change, use the test function (5) is better.

- If the data in gradual change, the summation is approximation to 0. If the data changes from constant to gradual, the summation is approximate $\frac{C_2-C_1}{12}$.
- In our numerical example, we first use the function (5) to detect the sudden change. And the for gradual change, we use the function (6) to detect the gradual change. Since the summation for sudden change is difficult to control, we set the summation is 0.

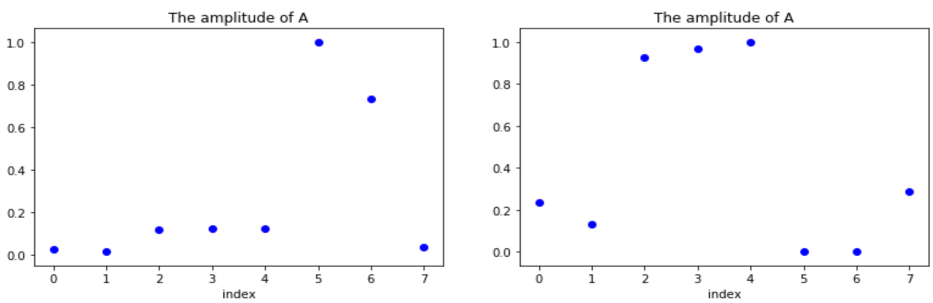
3.1 Numerical Result

data : 100000. gradual change start : 30000 gradual change end : 50000. Sudden change : 70000. Figure (9) show the data we use in this example.



(a) (b)

Figure 9: streaming data. (a) mixed change data; (b) smooth mixed change data with smooth parameter 2001.



(a) (b)

Figure 10: amplitude of the (4) for first level. (a) detect sudden change; (b) detect gradual change.

The parameter we use in this numerical example :

N = 2

I = 2000

Ex = 10

epsion = 0.7
k = 3
smooth parameter = 2001
The auto-detect result is :
****
the sudden change happens at : [71385.75]
****
the gradual change happens at : [31249.0, 49608.4375]

References

[1] F. GUSTAFSSON AND F. GUSTAFSSON, *Adaptive filtering and change detection*, vol. 1, Citeseer, 2000.