Change Detection

Tuesday, September 3, 2024 7:43 PM

Change Detection Introduction

Change Detection: the process of determining if something has changed, often in time-series data.

Examples of change detection include:

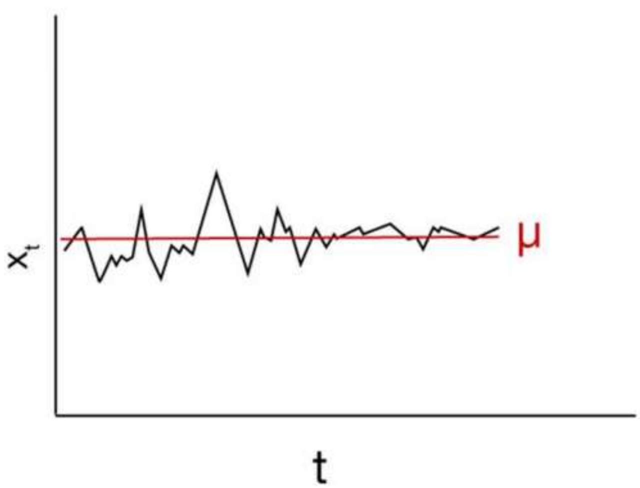
- Determining if an action is needed:
    - Has global temperature changed?
    - Have sales decreased due to competition?
    - When is the right time for preventative maintenance?
- Determining the impact of a past action:
    - Did business tax decrease cause more hiring?
    - Did price discounts increase sales?
- Determining changes to help plan ahead:
    - Have voting patterns changed?

Cumulative Sum (CUSUM) for Change Detection

CUSUM: an approach to change detection that can detect increases and decreases (or both).

Suppose the following:

$x_t = observed\ value\ at\ time\ t$
$\mu = mean\ of\ x\ if\ there\ is\ no\ change$



Then the CUSUM metric is as follows, with the goal to find if the metric grows larger than some threshold metric T.

$S_t = \max\{0, S_{t-1} + (x_t - \mu - C)\}$ Is $S_t \geq T$?

In layman's terms, this metric is cumulative; the S_t starts at 0. During the next check at time t, the difference between the observed value and the mean $(x_t - \mu)$ is added to the metric, which becomes a running total. If the metric becomes less than zero, it is reset to zero. Otherwise, it continues cumulatively summing until the metric grows larger than the threshold.

Resetting S_t to zero if the metric falls below zero allows CUSUM to detect changes quickly, since if the metric falls below zero that data isn't relevant for future change detection.

By pure random chance, an observed value may be higher than the mean, so a constant value C is included to pull the running total down a little bit. The bigger C is, the less sensitive the CUSUM metric is; the smaller C is, the more sensitive the CUSUM metric is (since it can grow larger faster).

The question then become: how do you find the right values for C and T so that the metric can find changes quickly but isn't so sensitive that it gives false positives?
There isn't a simple answer. It will depend on the model itself and what the costs associated with a very sensitive model versus a less sensitive model. There will be a tradeoff of the types of mistakes that can be made both ways; is it more costly to have false positives or to miss true positives? Finding the optimal trade-off between these types of mistakes is part of the process.

As an example of this tradeoff, observe the model output below with a higher threshold of 450. An actual change in observation values starts at t = 11, but the model doesn't detect it until a couple cycles later at t = 14. So the model didn't have any false positives, but it took longer to detect true change.

			T = 450, C = 0	
t	X_t	$X_t - \mu$	$X_t - \mu - C$	S_t
0				0
1	120	-15	-15	0
2	230	95	95	95
3	20	-115	-115	0
4	280	145	145	145
5	80	-55	-55	90
6	150	15	15	105
7	90	-45	-45	60
8	140	5	5	65
9	150	15	15	80
10	90	-45	-45	35
11	280	145	145	180
12	130	-5	-5	175
13	310	175	175	350
14	280	145	145	495
15	230	95	95	590
16	200	65	65	655
17	210	75	75	730
18	350	215	215	945
19	160	25	25	970
20	200	65	65	1035

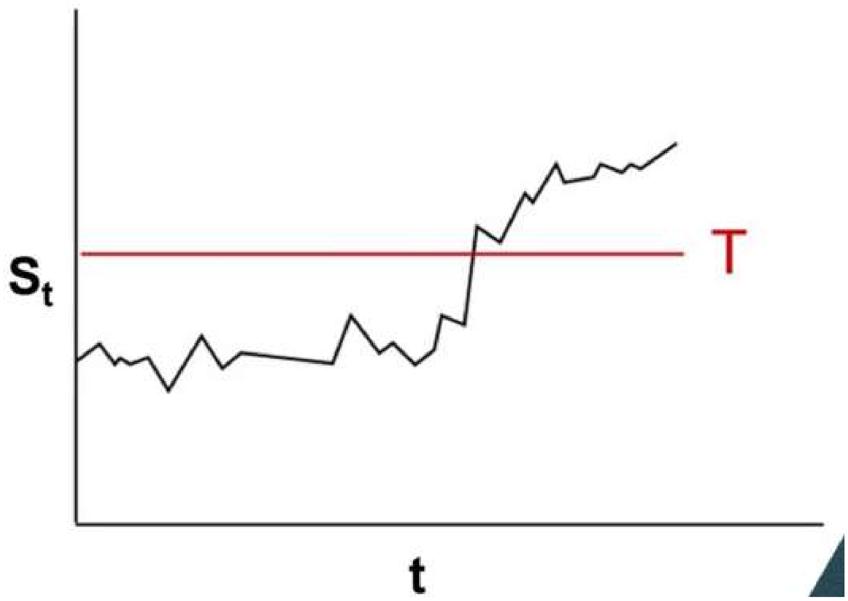
On the other hand, the same model was made more sensitive with a lower threshold of 150 and therefore detected the change at t = 11, but it also almost had a false positive at t = 4.

			T = 150, C = 0	
t	X_t	$X_t - \mu$	$X_t - \mu - C$	S_t
0				0
1	120	-15	-15	0
2	230	95	95	95
3	20	-115	-115	0
4	280	145	145	145
5	80	-55	-55	90
6	150	15	15	105
7	90	-45	-45	60
8	140	5	5	65
9	150	15	15	80
10	90	-45	-45	35
11	280	145	145	180
12	130	-5	-5	175
13	310	175	175	350
14	280	145	145	495
15	230	95	95	590
16	200	65	65	655
17	210	75	75	730
18	350	215	215	945
19	160	25	25	970
20	200	65	65	1035

Note: the metric for CUSUM changes depending on if you are trying to detect an increase or a decrease:

- Increase: $S_t = \max\{0, S_{t-1} + (x_t - \mu - C)\}$ Is $S_t \geq T$?
- Decrease: $S_t = \max\{0, S_{t-1} + (\mu - x_t - C)\}$ Is $S_t \geq T$?
- Both: Calculate both metrics at the same time.

CUSUM can be nicely graphed in what is known as a Control Chart by plotting the CUSUM metric over time; if the metric ever crosses the threshold, then the model has detected a change.


T
S_t
t