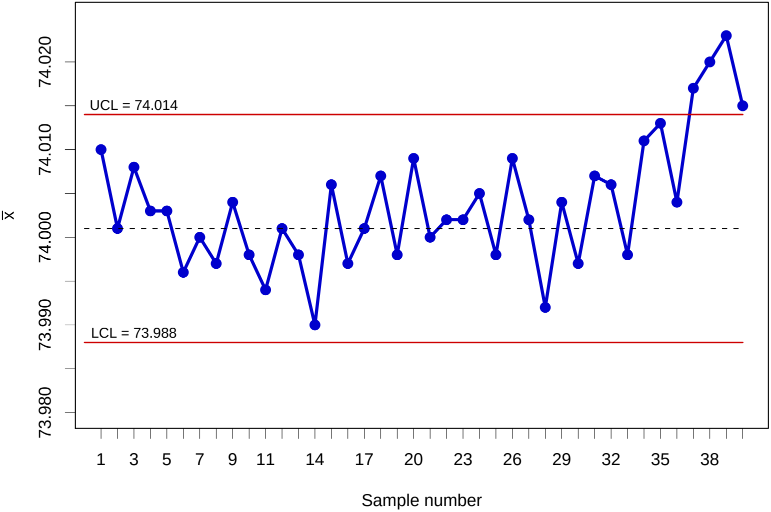
Monitoring Non-Stationary Processes

WOLFGANG SCHMID

European University Viadrina
Frankfurt (Oder), Germany

1. Introduction

Beispiel: Inside diameter (mm) of automobile engine piston rings (cf. Montgomery (2005))



target process $\longrightarrow$ **Changes?** (mean shifts, scale deviations, outliers, turning points, etc.) $\longrightarrow$ actual process

sequential problem $\rightsquigarrow$ statistical process control, change point analysis

here: At each time point exactly one observation is present.

change point model

$$X_t = \begin{cases} \mu_0 + a\sigma + \Delta(Y_t - \mu_0) & \text{for} \quad t \geq \tau \\ Y_t & \text{for} \quad t < \tau \end{cases} \quad , \quad t \geq 1$$

Let $\mu_0 = E(Y_t)$ and $\sigma^2 = Var(Y_t)$ be known, $(a, \Delta) \neq (0, 1)$ and $\tau \in I\!N \cup \{\infty\}$ be unknown.

$\{X_t\}$ is called to be in control if $\tau = \infty$. Else, it is out of control.

Assumption: The distribution of $\{Y_t\}$ is known. Frequently it is assumed to be i.i.d., $Y_t \sim N(\mu_0, \sigma^2)$. $\rightsquigarrow$ Shewhart, EWMA, CUSUM chart,...

Problem: In many applications the underlying data are dependent.

Figure: Facebook share price with estimated trend on the out-of-control period

2. Modeling

here: monitoring of non-stationary processes

change point model

$$\boldsymbol{X}_t = \begin{cases} \boldsymbol{Y}_t & \text{for} \quad 1 \leq t < \tau \\ \boldsymbol{Y}_t + \boldsymbol{D}_{t,\tau}\, \boldsymbol{a} & \text{for} \quad t \geq \tau \end{cases},$$

with

- $\tau \in I\!N \cup \{\infty\}$ and $\boldsymbol{a} \in \mathbb{R}^p - \{\boldsymbol{0}\}$ unknown,
  $\tau = \infty \rightsquigarrow$ no change point is present
  $\tau < \infty \rightsquigarrow$ observed process is out of control
- $\boldsymbol{D}_{t,\tau}$ is a known sequence in τ, e.g.
  mean change: $\boldsymbol{D}_{t,\tau} = diag(\sqrt{Var(Y_{t1})}, ..., \sqrt{Var(Y_{tp})})$,
  linear drift: $\boldsymbol{D}_{t,\tau} = (t - \tau + 1)\boldsymbol{I}$
- $E(\boldsymbol{Y}_t) = \boldsymbol{\mu}_t$ (known)

Examples:

- random walk:
  $\boldsymbol{Y}_t = \boldsymbol{Y}_{t-1} + \varepsilon_t$ for $t \geq 1$, $\boldsymbol{Y}_0 = \mathbf{0}$, $\{\varepsilon_t\}$ i.i.d., $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$
  here: $\boldsymbol{\mu}_t = \mathbf{0}$, $Cov(\boldsymbol{Y}_t) = t\,\mathbf{\Sigma}$
  mean change: $\boldsymbol{D}^2_{t,\tau} = t\,diag(\mathbf{\Sigma})$
- multiple linear regression:
  $\boldsymbol{\mu}_t = \boldsymbol{Z}_t \boldsymbol{\beta}$
- VARMA processes

Problem: How to get charts for time series?

1. Residual charts (e.g., Alwan and Roberts (1988),...)
   $\rightsquigarrow$ data transformation, studied for stationary processes,
2. Multivariate EWMA chart for time series (Kramer and Schmid (1997))
   $\rightsquigarrow$ simple, studied for stationary processes
3. LQ method

$$-2\, log\left(\frac{f_0(\boldsymbol{X}_1, ..., \boldsymbol{X}_n; \boldsymbol{\theta})}{\max\limits_{0 \le i \le n} f_{\boldsymbol{a},i}(\boldsymbol{X}_1, ..., \boldsymbol{X}_n; \boldsymbol{\theta})}\right) > c \rightsquigarrow \quad \text{signal}$$

- The approach is based on a fundamental statistical principle.
- Optimality property in the iid case.
- Done for stationary Gaussian processes (e.g., Yashchin (1993, 2015), Bodnar and Schmid (2007, 2011))

Remarks:

- Charts for time series are in general not directionally invariant (e.g., Bodnar and Schmid (2007))
- Mostly $\boldsymbol{\theta}$ is assumed to be known, else prerun (Phase I analysis), Bayesian approach,...

here: non-stationary in-control process

aim: flexible model for the in-control process which is quite general and easy to handle $\rightsquigarrow$ state-space model (e.g., Durbin and Koopman (2012))

State-Space Model

$$\boldsymbol{Y}_t = \boldsymbol{G}_t \boldsymbol{S}_t + \boldsymbol{W}_t, \quad t = 1, 2, ..., \qquad \text{where} \tag{1a}$$

$$\boldsymbol{S}_{t+1} = \boldsymbol{F}_t \boldsymbol{S}_t + \boldsymbol{V}_t, \quad t = 1, 2, ..., \tag{1b}$$

(1a) is called observation equation, (1b) is the state equation.

Note that the state-space model is applied in a different way as usual!

Examples:

- multivariate multiple regression model with VARMA errors
- multivariate random walk

Assumptions to be used in the following:

(A1) Let for all $t \geq 1$

$$E\begin{pmatrix} \boldsymbol{V}_t \\ \boldsymbol{W}_t \end{pmatrix} = \mathbf{0}, \quad E(\boldsymbol{V}_t \boldsymbol{V}_t') = \boldsymbol{Q}_t, E(\boldsymbol{W}_t \boldsymbol{W}_t') = \boldsymbol{R}_t, E(\boldsymbol{V}_t \boldsymbol{W}_t') = \boldsymbol{U}_t.$$

$\{\boldsymbol{Q}_t\}$, $\{\boldsymbol{R}_t\}$, and $\{\boldsymbol{U}_t\}$ are specified sequences of $q \times q, p \times p$ and $q \times p$ matrices, respectively.

(A2) Let $\boldsymbol{S}_1, (\boldsymbol{V}_1', \boldsymbol{W}_1')', (\boldsymbol{V}_2', \boldsymbol{W}_2')', \ldots$ be independent.

(A3) Let $E(\boldsymbol{Y}_0 \boldsymbol{V}_t') = \mathbf{0}$ and $E(\boldsymbol{Y}_0 \boldsymbol{W}_t') = \mathbf{0}$ for all $t \geq 1$.

(A4) Let Let $\boldsymbol{S}_1, (\boldsymbol{V}_1', \boldsymbol{W}_1')', (\boldsymbol{V}_2', \boldsymbol{W}_2')', \ldots$ be normally distributed.

(A5) Let $\boldsymbol{\Sigma}_t$ have a full rank for all $t \geq 1$.

Influence of parameter estimation, e.g. Albers and Kallenberg (2004), Jensen et al. (2006), Capizzi (2015)

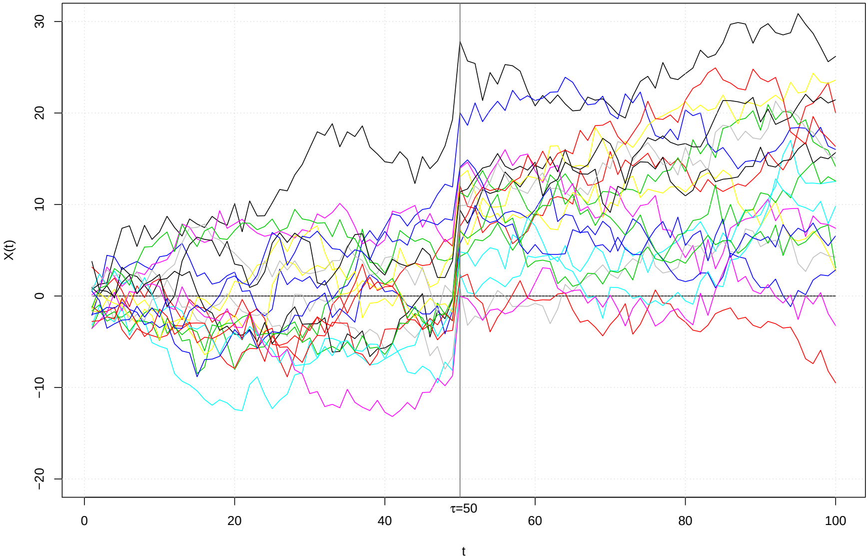


Figure: 20 realisations of an out-of-control process (here: random walk with AR(2) noise, with $\alpha_1 = 0.6$, $\alpha_2 = 0.3$, further: $\tau = 50$, $a = 1.0$)

3. Control Charts for State-Space Models

3.1 LR Approach

$$g_{n;LR}(\boldsymbol{a}) = -2\ log(\frac{f_0(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)}{\max\limits_{0 \leq i \leq n} f_{\boldsymbol{a},i}(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)})$$

best linear one-step predictor of $\boldsymbol{Y}_t$ given $\boldsymbol{Y}_0, ..., \boldsymbol{Y}_{t-1}$:

$$\hat{\boldsymbol{Y}}_t = \boldsymbol{b}_t(\boldsymbol{Y}_0) + \sum_{j=1}^{t-1} \boldsymbol{B}_{t,j} \boldsymbol{Y}_j, \quad t \geq 1$$

Let

$$\hat{\boldsymbol{X}}_t = \boldsymbol{b}_t(\boldsymbol{X}_0) + \sum_{j=1}^{t-1} \boldsymbol{B}_{t,j} \boldsymbol{X}_j, \quad t \geq 1.$$

It can be shown that

$$g_{n;LR}(\boldsymbol{a}) = \max\left\{0, \max_{1\le i\le n} -2\left(\sum_{t=i}^{n}\left((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t + \frac{1}{2}\boldsymbol{M}_{t,i}\boldsymbol{a})'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{M}_{t,i}\boldsymbol{a}\right)\right)\right.$$

run length of the LR chart

$$N_{LR}(c, \boldsymbol{a}^*) = \inf\{n \in \mathbb{N} : g_{n;LR}(\boldsymbol{a}^*) > c\}$$

3.2 SPRT Approach

$$-2\ log(\frac{f_0(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)}{\max\limits_{i=1} f_{\boldsymbol{a},i}(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)})$$

$$= 2\sum_{t=1}^{n}(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t + \tfrac{1}{2}\boldsymbol{M}_{t,1}\boldsymbol{a})'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{M}_{t,1}\boldsymbol{a} = g_{n;SPRT}(\boldsymbol{a})$$

Restart if negative.

$$N_{SPRT}(c, \boldsymbol{a}^*) = \inf\{n \in \mathbb{N} : \max_{0\leq i\leq n}\{g_{n;SPRT}(\boldsymbol{a}^*) - g_{i;SPRT}(\boldsymbol{a}^*)\} > c\}$$

3.3 Shiryaev-Roberts Approach

$$\sum_{i=1}^{n} \frac{f_i(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)}{f_0(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)}$$

$$= \sum_{i=1}^{n} \exp\left\{ -\sum_{t=i}^{n} (\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t + \tfrac{1}{2}\boldsymbol{M}_{t,i}\boldsymbol{a})' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{M}_{t,i}\boldsymbol{a} \right\} = g_{n;SR}(\boldsymbol{a}).$$

The run length of the SR chart is given by

$$N_{SR}(c, \boldsymbol{a}^{*}) = \inf\{n \in \mathbb{N} : g_{n;SR}(\boldsymbol{a}^{*}) > c\}.$$

4. Generalized Control Charts for State-Space Models

up to now: charts depend on a reference vector $\rightsquigarrow$ how to choose?
next: generalized control charts (e.g., Siegmund and Venkatraman (1995), Lai (2001), Reynolds and Wang (2013))

4.1 Generalized Likelihood Ratio Charts

$$\max\{0, -2\, log(\frac{f_0(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)}{\max_{1 \le i \le n} \sup_{\boldsymbol{a} \neq \boldsymbol{0}} f_{\boldsymbol{a},i}(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)\}}$$

$$N_{GLR}(c) = \inf\left\{n \geq 1 : \max\left\{0, \max_{1 \le i \le n}\left(-\sum_{t=i}^{n}(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t + \frac{1}{2}\boldsymbol{M}_{t,i}\tilde{\boldsymbol{a}}_{i,n})'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{M}_{t,i}\tilde{\boldsymbol{a}}_{i,n}\right)\right\} > c$$

where $\tilde{\boldsymbol{a}}_{i,n}$ is any solution of

$$\left(\sum_{t=i}^{n}\boldsymbol{M}'_{t,i}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{M}_{t,i}\right)\tilde{\boldsymbol{a}}_{i,n} = \sum_{t=i}^{n}\boldsymbol{M}'_{t,i}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t).$$

4.2 The Generalized SPRT Approach

$$\sup_{\boldsymbol{a}\neq 0} \log \frac{f_{i=1}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)}{f_0(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)}$$

$$= -\sum_{t=1}^{n} (\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t + \frac{1}{2}\boldsymbol{M}_{t,1}\tilde{\boldsymbol{a}}_{1,n})' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{M}_{t,1}\tilde{\boldsymbol{a}}_{1,n} = g_{n;GSPRT}$$

$$N_{GSPRT}(c) = \inf \left\{ n \in \mathbb{N} : \max_{0\leq i\leq n} (g_{n;GSPRT} - g_{i;GSPRT}) > c \right\}$$

4.3 The Generalized SR Approach

problem: Since the approach leads to the analysis of exponential sums we consider the sum of the logarithms!

$$\begin{aligned} R_n(\boldsymbol{a}) &= \sum_{i=1}^{n} \log \frac{f_i(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)}{f_0(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)} \\ &= -\underbrace{\left(\sum_{i=1}^{n}\sum_{t=i}^{n}(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{M}_{t,i}\right)}_{=\dot{\boldsymbol{S}}_n'}\boldsymbol{a} - \frac{1}{2}\boldsymbol{a}'\underbrace{\left(\sum_{i=1}^{n}\sum_{t=i}^{n}\boldsymbol{M}_{t,i}'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{M}_{t,i}\right)}_{=\ddot{\boldsymbol{S}}_n}\boldsymbol{a} \\ &= -\dot{\boldsymbol{S}}_n'\boldsymbol{a} - \frac{1}{2}\boldsymbol{a}'\ddot{\boldsymbol{S}}_n\boldsymbol{a}. \end{aligned}$$

and

$$\sup_{\boldsymbol{a}\neq 0} R_n(\boldsymbol{a}) = R_n(\boldsymbol{a}_n^*) = -\dot{\boldsymbol{S}}_n'\boldsymbol{a}_n^* - \frac{1}{2}\boldsymbol{a}_n^{*\prime}\ddot{\boldsymbol{S}}_n\boldsymbol{a}_n^* = \frac{1}{2}\boldsymbol{a}_n^{*\prime}\ddot{\boldsymbol{S}}_n\boldsymbol{a}_n^*.$$

where $\boldsymbol{a}_n^*$ is any solution of the equation $\ddot{\boldsymbol{S}}_n\boldsymbol{a} = -\dot{\boldsymbol{S}}_n$.

$$N_{GMSR}(c) = \inf\left\{n\in\mathbb{N} : \boldsymbol{a}_n^{*\prime}\ddot{\boldsymbol{S}}_n\boldsymbol{a}_n^* > c\right\}$$

5.1 Behavior for Stationary Processes

here: $\{Y_t\}$ univariate AR(2) process with $\alpha_1 = 0.6$ and $\alpha_2 = 0.3$, $a^* \in \{0.5, 1.0, 1.5, 2.0\}$ and $\xi = 500$

a	LR	SPRT	SR	GLR	GSPRT	GMSR
0.5	179.33(0.5)	148.64(1.0)	172.64(1.0)	266.76	145.06	282.11
1.0	94.68(1.0)	56.58(1.0)	89.18(1.5)	119.53	43.30	168.29
1.5	45.25(2.0)	23.73(1.0)	42.12(2.0)	60.07	14.80	121.23
2.0	18.45(2.0)	9.80(1.0)	18.21(2.0)	29.76	5.35	95.62
2.5	7.38(2.0)	4.13(1.0)	7.61(2.0)	13.71	2.23	78.76
3.0	3.11(2.0)	2.07(1.0)	3.30(2.0)	5.90	1.34	67.05
3.5	1.69(2.0)	1.39(1.0)	1.77(2.0)	2.72	1.09	58.28
4.0	1.24(2.0)	1.15(1.0)	1.29(2.0)	1.59	1.03	51.39

Table: Out-of-control ARLs

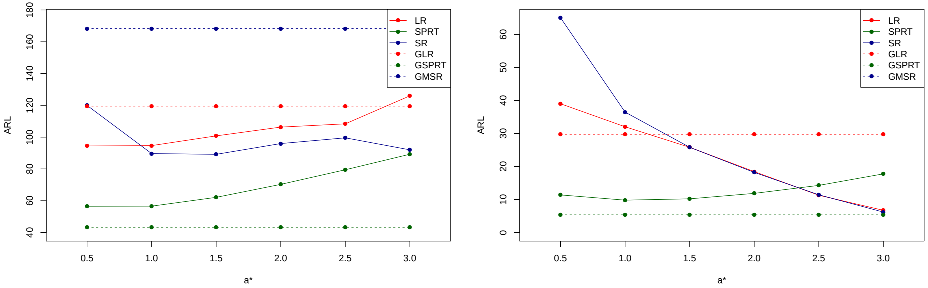


Figure: Out-of-control ARLs as a function of the reference value a^* for $a = 1.0$ (left) and for $a = 2.0$ (right)

	LR	SPRT	SR	GLR	GSPRT
a=1.0	94.51	56.76	89.87	119.80	43.09
	69.32	67.75	72.79	95.98	87.22
	60.65	59.12	56.19	88.06	88.68
a=2.0	18.26	9.77	17.91	29.69	5.32
	3.44	20.76	3.56	5.62	31.95
	3.33	16.59	3.18	5.19	32.65

Table: Average run length (above), worst average delay for $1 < \tau < 50$ (middle) and the value of the delay at position $\tau = 50$ (below)

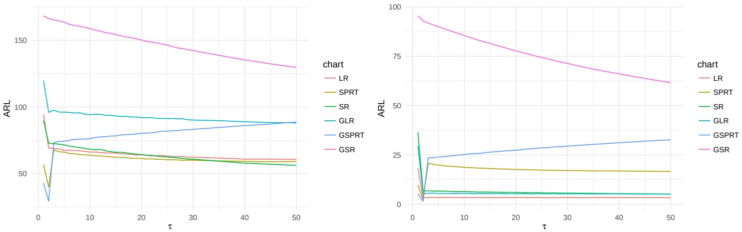


Figure: Average delay as a function of $\tau \in \{1, ..., 50\}$ for $a = 1.0$ (left) and for $a = 2.0$ (right)

5.2 Behaviour for Non-Stationary Processes

here: $Y_t = Y_{t-1} + \eta_t$, $\{\eta_t\}$ AR(2)

i	LR	SPRT	SR	GLR	GSPRT	GMSR
1	22.594	11.682	5.217	0.100	28.867	0.200
2	13.730	9.380	12.744	0.220	15.039	1.150
3	9.713	1.987	10.400	0.220	2.527	1.120
4	7.899	2.507	9.487	0.240	2.840	1.280
5	6.067	2.632	7.950	0.190	2.247	1.150
...	...	...	...	...	...	...
[1000, 2000]	0.097	2.273	0.114	11.660	1.993	3.280
[2000, 3000]	0.046	1.063	0.057	1.950	0.945	1.300
[3000, 4000]	0.045	0.614	0.036	0.230	0.636	0.600
[4000, 5000]	0.018	0.398	0.030	0.060	0.467	0.580
[5000, 6000]	0.020	0.310	0.021	0.000	0.339	0.360
[6000, 7000]	0.016	0.288	0.025	0.000	0.265	0.230
[7000, 8000]	0.005	0.224	0.011	0.000	0.250	0.180
[8000, 9000]	0.012	0.153	0.008	0.000	0.199	0.180
[9000,10000)	0.009	0.130	0.013	0.000	0.165	0.170
10000	4.832	2.622	4.852	0.000	2.738	1.770

Table: Frequency table of the in-control run lengths

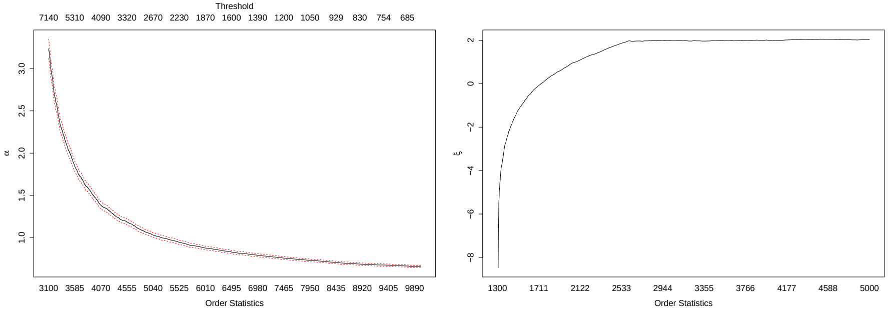


Figure: Hill plot (left) and Pickands plot (right)

of the SPRT chart

Modified Shewhart chart for the mean

Signal at time t if

$$|X_t - \mu_t| > c\sqrt{\mathsf{Var}(Y_t)}$$

where $c > 0$ denotes a specified constant.

Run length of the modified Shewhart chart

$$N = \inf\left\{t \in \mathbb{N} : \ |X_t - \mu_t| > c\sqrt{\mathsf{Var}(Y_t)}\ \right\}.$$

Theorem

Let $Y_t = Y_{t-1} + w_t = \sum_{v=1}^{t} w_v$ for $t \geq 1$, $A_n = P(\max_{1\leq t\leq n} \frac{|Y_t|}{\sqrt{t}} \leq c)$ and $c > 0$. Suppose that the random variables $\{w_v\}$ are independent and identically distributed with distribution function F. F is assumed to be absolutely continuous with existing second moment and symmetric around 0. Then it holds that the series $\sum_{n=1}^{\infty} A_n$ diverges to infinity.

probability of a successful detection:

$$PSD(\tau, d) = P(N(c) - \tau + 1 \leq d \mid N(c) \geq \tau), \quad \tau, d \in \mathbb{N}$$

a	LR	SPRT	SR	GLR	GSPRT
0.5	3.46(2.0)	4.91(0.5)	5.00(0.5)	2.14	2.63
1.0	9.72(2.0)	15.05(0.5)	10.61(0.5)	5.81	8.85
1.5	23.54(2.0)	33.53(0.5)	24.03(2.0)	15.54	22.74
2.0	44.21(2.0)	57.65(0.5)	45.37(2.0)	33.65	44.54
2.5	67.75(2.0)	79.63(0.5)	68.70(2.0)	57.47	68.41
3.0	85.75(2.0)	92.50(0.5)	86.49(2.0)	78.40	86.21
3.5	95.45(2.0)	98.01(0.5)	95.69(2.0)	91.89	95.67
4.0	98.91(2.0)	99.65(0.5)	98.92(2.0)	97.79	99.05

Table: Probabilities of successful detection (in percent, $\%$), $PSD(1, 5)$, $P(N(c) \leq 5) = 0.01$

a	LR	SPRT	SR	GLR	GSPRT
1.0	9.82(2.0)	2.32(0.5)	10.58(0.5)	5.92	0.35
2.0	44.57(2.0)	8.46(0.5)	45.54(2.0)	33.66	1.59

Table: Worst case scenario, $\min\limits_{1 \leq \tau \leq 50} PSD(\tau, 5)$

Conclusions

- If the change is expected at the beginning the SPRT and the GSPRT chart show the best performance.
- If the shift occurs at a later time point the LR and the SR chart should be preferred since they have the smallest delay.
- The LR and the SR chart dominate the GLR chart over a wide range of reference values. Only if the reference value is chosen far away from the optimal one the GLR chart turns out to be better.
- The GLR chart is the robustest scheme.

6. Summary

- Derivation of control charts for non-stationary processes using the LR, SPRT, SR, GLR, GSPRT, GMSR procedures
- Comparison study based on the ARL, average delay, and the probability of a successful detection
- Note that for non-stationary processes the ARL may not exist and thus other performance criteria based on probabilities and not on expectations must be used.
- The probability structure of the in-control process must be known.

References

Lazariv, T. and Schmid, W. (2018a): Surveillance of non-stationary processes. AStA - Advances in Statistical Analysis.

Lazariv, T. and Schmid, W. (2018b): Challenges in monitoring non-stationary processes. In S. Knoth and W. Schmid (eds.): Frontiers in Statistical Quality Control 12, pp, 257-275, Springer. .