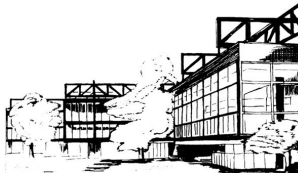


CUSUM, EWMA, and Shiryaev-Roberts under drift

... with attention to numerical algorithms

Sven Knoth

June 2009



Outline

- 1 Introduction.
- 2 Drift models and schemes.
- 3 Numerical Algorithms.
- 4 Comparison of various schemes.
- 5 Conclusions.

Introduction

- Dominating model in change point detection/surveillance/SPC is the step change.
- In the ads language of practical SPC trend and step change are often confused.
- At least one type of runs rule was created especially for detecting trends: Some number of successive points decrease or increase.
- Statistical analysis of control charts under trend started in the 1980s: BISSELL (1984), ASBAGH (1985).
- Usually, linear trend (in observation number) is deployed.



Some history

- BISSELL (1984): 1-sided, Shewhart, Shewhart with 2-of-2 runs rule, CUSUM, Markov chain.
- ASBAGH (1985): similar to BISSELL, more elaborated, better implemented, Shewhart-CUSUM added.
- SWEET (1988): coupled EWMA series (for mean and slope), none SPC performance measures calculated.
- DAVIS & WOODALL (1988): illustrate that runs trend rule is useless, Monte-Carlo.
- GAN (1991,2): CUSUM, EWMA, accurate results.
- AERNE, CHAMP & RIGDON (1991): extensive comparison study, Markov chain.



Some more history

- CHANG & FRICKER (1999): monotonically increasing means, detect asap level crossing and not the drift itself, CUSUM, EWMA, special GLR, Monte-Carlo.
- REYNOLDS & STOUMBOS (2001): simultaneous EWMA (mean and variance), Monte-Carlo.
- FAHMY & ELSAYED (2006): local regression vs. 1-sided Shewhart, CUSUM, EWMA, drift GLR, Monte-Carlo.
- ZOU, LIU & WANG (2008): 1-sided, EWMA, CUSUM, GEWMA, step change & drift GLR, Monte-Carlo.



Change point models for drift

- X_1, X_2, \dots – sequence of independent random variables.
- Change point τ , drift coefficient Δ :

$$E(X_t) = \begin{cases} \mu_0 = 0 & , t < \tau \\ \mu_{t-\tau} = (t - \tau + 1)\Delta & , t \geq \tau \end{cases} .$$

(GAN preferred $\mu_{t-\tau} = (t - \tau)\Delta \rightarrow E(X_\tau) = 0 = \mu_0$)

$$\text{Var}(X_t) = \sigma_0^2 = 1 .$$

- SWEET discussed treatment of non-equidistant sampling.
- CHANG & FRICKER considered $\mu_1 \leq \mu_2 \leq \dots$ with some threshold δ .



① Step change charts under drift

- Shewhart,
- Shewhart with runs rules (2 of 3, 4 of 5, ...)*,
- CUSUM,
- EWMA,
- Shiryayev-Roberts (GRSR),
- GEWMA,
- step change GLR.

② Special drift charts

- coupled EWMA (slope smoothing included),
- local regression,
- drift GLR,
- step change charts to $\text{diff}(X)$.

* DIVOKY & TAYLOR (1995) studied 613 different trend rules

Schemes/charts

Some formulas for 1-sided setups – step change charts

Shewhart: $L = \inf\{n \in \mathbb{N} : X_n > c_S\},$

CUSUM: $S_n = \max\{0, S_{n-1} + X_n - k\}, S_0 = s_0 = 0,$
 $L = \inf\{n \in \mathbb{N} : S_n > h\},$

EWMA: $Z_n = \max\{z_{\text{reflect}}^*, (1 - \lambda)Z_{n-1} + \lambda X_n\}, \lambda \in (0, 1], Z_0 = \mu_0 = 0,$
 $L = \inf\left\{n \in \mathbb{N} : Z_n > c^* = c\sqrt{\lambda/(2 - \lambda)}\right\},$

GRSR: $R_n = (1 + R_{n-1}) \exp(X_n - k), R_0 = r_0 = 0,$
 $L = \inf\{n \in \mathbb{N} : R_n > g\},$

GEWMA[†]: $\tilde{Z}_n(\lambda) = \sqrt{\frac{2 - \lambda}{\lambda[1 - (1 - \lambda)^{2n}]}} \sum_{i=1}^n \lambda(1 - \lambda)^{n-i} X_i,$
 $L = \inf\{n \in \mathbb{N} : \max_{1 \leq k \leq n} \tilde{Z}_n(1/k) > c_g\}.$

[†]HAN & TSUNG (2004)



Schemes/charts

GLR – generalized likelihood ratio

step change:

$$LR_n(\tau, \mu) = \prod_{i=\tau}^n \frac{e^{-[X_i - \mu]^2/2}}{e^{-X_i^2/2}} \rightarrow \max_{1 \leq \tau \leq n, \mu},$$

$$T_n = \max_{1 \leq \tau \leq n, \mu} LR_n(\tau, \mu) = \max_{1 \leq \tau \leq n} \max_{\mu} LR_n(\tau, \mu),$$

$$L = \inf\{n \in \mathbb{N} : T_n > h_S\},$$

drift:

$$LR_n(\tau, \Delta) = \prod_{i=\tau}^n \frac{e^{-[X_i - (i - \tau + 1)\Delta]^2/2}}{e^{-X_i^2/2}} \rightarrow \max_{1 \leq \tau \leq n, \Delta},$$

$\tilde{T}_n =$ similar to step change,

$$L = \inf\{n \in \mathbb{N} : \tilde{T}_n > h_D\}.$$



Schemes/charts

Local regression

- moving window of size w .
- OLS fit at the end of window n : $\hat{\mu}_{wn} = \hat{\alpha}_n + \hat{\beta}_j t_n$.
- $M_n = \frac{(\mu_0 - \hat{\mu}_{wn})^2}{1/w + (t_w - \bar{t})^2 / S_{tt}}$ with $S_{tt} = \sum_{i=1}^w (t_i - \bar{t})^2$.

- w/o drift $M_n \sim \chi_1^2$.
- neglect autocorrelation of M_n and treat it like a Shewhart chart:

$$L = \inf\{n \in \mathbb{N} : M_n > c_R\}.$$

- For details see FAHMY & ELSAYED (2006).



Schemes/charts

Local regression

- Two coupled EWMA series (S_t for mean, B_t for slope):

$$S_0 = \mu_0 = 0,$$

$$S_t = (1 - \lambda_S)(S_{t-1} + B_{t-1}) + \lambda_S X_t,$$

$$B_0 = \Delta_0 = 0,$$

$$B_t = (1 - \lambda_B)B_{t-1} + \lambda_B(S_t - S_{t-1}).$$

- No usual performance study done so far.
- For details see SWEET (1988). He calculated adopted “control chart constants” to attain the usual chart behavior w/o checking it.



Control chart performance measures

Here (and there):

$$\text{zero-state ARL} \quad \mathcal{L} = \begin{cases} E_1(L) & , \tau = 1 \text{ ("early" change)} \\ E_\infty(L) & , \tau = \infty \text{ (no change)} \end{cases},$$

$$\text{steady-state ARL} \quad \mathcal{D} = \lim_{\tau \rightarrow \infty} E_\tau(L - \tau + 1 \mid L \geq \tau).$$



Calculation

Notes

- Inhomogeneous transition kernel \rightarrow most of the classics fail.
- Most of the papers deploy Monte-Carlo studies.
- BISSELL, ASBAGH, AERNE ET AL. utilize Markov chain.
- GAN developed an iteration rule based on the classical (Nyström) solution of the ARL integral equation.
- Here also usage of KNOTH (2003): yields \mathcal{L} and \mathcal{D} .



Calculation

Markov chain

- 1 Approximate the continuous chart statistic with a finite Markov chain.
- 2 Denote \mathbf{P}_n the transition matrix linked to transition $Z_{n-1} \rightarrow Z_n$, \mathbf{z}_0 the starting vector and $\mathbf{1}$ a vector of ones.
- 3 Then $\hat{P}(L > n) = \mathbf{z}_0' \left(\prod_{i=1}^n \mathbf{P}_i \right) \mathbf{1}$.
- 4 Thus, with N large (Cauchy Criterion like)

$$\hat{\mathcal{L}} = \sum_{n=0}^N \hat{P}(L > n),$$

BISSELL and others

$$\hat{\mathcal{L}} = \sum_{n=1}^N n [\hat{P}(L > n) - \hat{P}(L > n - 1)].$$



Calculation

Density recursion

For z within continuation region

$$f_1(z; z_0) = \phi_{\mu_1}(z_0 \rightarrow z),$$

$$f_n(z; z_0) = \int f_{n-1}(\tilde{z}; z_0) \phi_{\mu_n}(\tilde{z} \rightarrow z) d\tilde{z},$$

$$P(L > n)(z_0) = \int f_n(\tilde{z}; z_0) d\tilde{z},$$

+ quadrature for integral.

More Details in KNOTH (2003).



Calculation

GAN (1991/2)

Denote $\mathcal{L}_n(y, \mu_n)$ the final ARL for a control chart starting at y and mean sequence $\mu_n = n\Delta, \mu_{n+1}, \dots, \mu_N, \mu_N, \dots$

$$\mathcal{L}_n(y, \mu_n) \underset{n=(0,1,\dots,N-1)}{=} 1 + \int \mathcal{L}_{n+1}(y, \mu_{n+1}) \phi_{\mu_n}(y \rightarrow x) dx,$$

$$\mathcal{L}_N(y, \mu_N) = 1 + \int \mathcal{L}_N(y, \mu_N) \phi_{\mu_N}(y \rightarrow x) dx.$$

Procedure: Fix N large, solve integral equation for $\mathcal{L}_N(y, \mu_N)$ and proceed with recursion from $n = N - 1$ until $n = 1$ (GAN took $n = 0$).



Calculation

GAN (1991/2) vs. KNOTH (2003)

- For sufficiently large N (and same number of quadrature nodes) GAN and KNOTH provide same accuracy.
- Gradually increasing N in GAN's algorithm (his own approach) demands complete re-calculation of all $\mathcal{L}_j(y, \mu_j)$, while KNOTH's algorithm asks only for (some) new iterations. Thus, KNOTH's algorithm is faster.
- Additionally, KNOTH's algorithm allows computation of $E_\tau(L - \tau + 1 \mid L \geq \tau)$ for arbitrary τ (thus also for very large ones).
- Eventually, \mathcal{D} could be determined via $E_\tau(L - \tau + 1 \mid L \geq \tau)$ for $\tau \gg 1$ or the left eigenfunction of the transition kernel.



BISSELL (1984/6)

1-sided CUSUM, $k = 0.5$, $h = 5$ in-control $\mathcal{L} \approx 930$:

Δ	$\hat{\mathcal{L}}_{MC}$	s. e.	BISSELL	GAN/KNOTH	$\hat{\mathcal{L}}_{MC}$	s. e.
0.001	245	23.8	–	231	231	0.113
0.002	142	12.4	–	156	156	0.068
0.005	87	7.0	205	89	89	0.033
0.01	56.4	3.8	101	57.2	57.2	0.019
0.02	36.2	2.1	53	36.5	36.5	0.010
0.05	18.8	1.2	24	20.4	20.4	0.005
0.1	14.3	0.58	14.2	13.3	13.3	0.003
0.2	8.4	0.32	9.0	8.8	8.8	0.002
0.5	5.2	0.12	5.3	5.3	5.3	0.001
1.0	3.40	0.13	3.6	3.60	3.60	0.001
2.0	2.44	0.10	2.5	2.50	2.50	0.000
3.0	1.96	0.04	2.0	2.01	2.01	0.000
	25				10^6	



FAHMY & ELSAYED (2006)

2-sided CUSUM ($k = 0.25, h = 8$), EWMA ($\lambda = 0.1$)

Δ	CUSUM			EWMA		
	FE (10^4)	here (10^7)	FE (10^4)	here (10^7)	here (Gan)	
0	368.333 \pm 3.549	368.251 \pm 0.111	365.749 \pm 3.598	369.021 \pm 0.114	368.994	
0.10	13.986 \pm 0.026	14.086 \pm 0.001	12.971 \pm 0.029	12.986 \pm 0.001	12.986	
0.25	8.560 \pm 0.014	8.656 \pm 0.000	7.738 \pm 0.015	7.758 \pm 0.000	7.758	
0.50	5.946 \pm 0.008	6.033 \pm 0.000	5.312 \pm 0.009	5.318 \pm 0.000	5.318	
0.75	4.827 \pm 0.007	4.898 \pm 0.000	4.279 \pm 0.007	4.286 \pm 0.000	4.285	
1.00	4.156 \pm 0.006	4.224 \pm 0.000	3.680 \pm 0.006	3.688 \pm 0.000	3.688	
2.00	2.950 \pm 0.003	2.989 \pm 0.000	2.598 \pm 0.005	2.616 \pm 0.000	2.616	



FAHMY & ELSAYED (2006)

EWMA ($\lambda \in \{0.2, 0.3, 0.5\}$), FE's $\chi^2 (M_n)$, and trend GLR

Δ	FE χ^2			EWMA			tGLR $w = 50$
	$w^* = 3$	$w^* = 5$	$w^* = 20$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.5$	
0	379.138 \pm 3.790	370.048 \pm 3.682	373.458 \pm 3.546	370	370	370	367
0.10	17.445 \pm 0.056	16.047 \pm 0.049	12.860 \pm 0.035	12.747	13.041	14.136	13.229
0.25	8.537 \pm 0.032	8.127 \pm 0.023	7.623 \pm 0.018	7.304	7.231	7.497	7.380
0.50	5.027 \pm 0.021	4.869 \pm 0.013	5.260 \pm 0.012	4.881	4.722	4.706	4.758
0.75	3.672 \pm 0.017	3.673 \pm 0.009	4.250 \pm 0.009	3.886	3.715	3.620	3.665
1.00	2.939 \pm 0.014	3.055 \pm 0.007	3.660 \pm 0.008	3.318	3.149	3.023	3.055
2.00	1.816 \pm 0.003	2.042 \pm 0.000	2.579 \pm 0.000	2.254	2.124	2.005	1.964



2-sided CUSUM

Incidental remark

Already GAN mentioned that the idea of LUCAS/CROSIER (1982) to combine the ARL results of two one-sided CUSUM charts for getting the two-sided ARL does not work. Both GAN's and KNOTH's algorithms do not converge for the chart opposite to the change.

Monte-Carlo studies must be utilized for ARL analysis of 2-sided CUSUM under trend.



ZOU, LIU & WANG (2008)

1-sided schemes, EWMA, CUSUM, GRSR calculated with
GAN/KNOTH, others by Monte-Carlo (10^4 rep.) in ZOU ET AL.

Δ	EWMA			CUSUM			GRSR			GEWMA	GLR-S	GLR-L
	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$			
0	1750	1747	1733	1741	1742	1735	1730	1730	1730	-	-	-
0.0005	318	378	437	345	412	468	337	399	448	375	381	368
0.001	215	254	295	231	276	316	227	267	301	252	257	249
0.005	83.5	92.2	106	86.7	98.3	112	85.8	95.7	107	96.2	97.8	95.4
0.01	55.7	58.7	66.3	57.0	61.9	69.4	56.6	60.4	66.6	62.1	63.3	62.0
0.05	22.6	21.1	22.0	22.6	21.6	22.6	22.7	21.4	22.1	22.4	22.7	22.5
0.1	15.5	13.9	13.9	15.4	14.0	14.2	15.7	14.1	14.0	14.4	14.6	14.5
0.5	6.65	5.56	5.09	6.60	5.54	5.16	6.84	5.76	5.32	5.10	5.23	5.18
1.0	4.67	3.83	3.43	4.63	3.80	3.45	4.86	4.03	3.66	3.26	3.38	3.31
2.0	3.21	2.74	2.32	3.17	2.67	2.32	3.42	2.91	2.59	2.09	2.16	2.12
3.0	2.86	2.06	1.98	2.79	2.04	1.96	2.97	2.20	2.02	1.69	1.75	1.72
4.0	2.14	2.00	1.83	2.10	1.98	1.74	2.39	2.20	1.97	1.31	1.37	1.34
	$\lambda = 0.03$	$\lambda = 0.11$	$\lambda = 0.23$									



ZOU, LIU & WANG (2008)

steady-state ARL

1-sided schemes, EWMA, CUSUM, GRSR calculated with
GAN/KNOTH

Δ	EWMA			CUSUM			GRSR		
	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.0005	314	376	436	340	410	467	333	397	446
0.001	213	253	295	228	275	315	224	266	301
0.005	82.6	91.8	106	85.4	97.9	112	84.2	95.1	107
0.01	55.1	58.4	66.2	55.9	61.6	69.2	55.3	60.0	66.4
0.05	22.3	20.9	21.9	21.8	21.4	22.6	21.6	21.1	21.9
0.1	15.4	13.8	13.8	14.8	13.8	14.1	14.7	13.7	13.8
0.5	6.59	5.50	5.05	6.17	5.36	5.08	6.18	5.38	5.08
1.0	4.62	3.79	3.40	4.30	3.65	3.37	4.31	3.68	3.40
2.0	3.27	2.66	2.33	2.98	2.53	2.26	3.00	2.58	2.30
3.0	2.68	2.13	1.91	2.50	1.99	1.90	2.52	2.01	1.92
4.0	2.32	1.90	1.73	2.01	1.89	1.66	2.04	1.90	1.73



Exclusive drift monitoring

The naïve approach – $\text{diff}(X)$

If

$$E(X_t) = \begin{cases} \mu_0 = 0 & , t < \tau \\ \mu_{t-\tau} = (t - \tau + 1)\Delta & , t \geq \tau \end{cases} ,$$

then

$$D_t \stackrel{t > 1}{=} X_t - X_{t-1} \quad \text{and} \quad D_1 = X_1 - \mu_0 = X_1 ,$$

$$E(D_t) = \begin{cases} 0 & , t < \tau \\ \Delta & , t \geq \tau \end{cases} \quad \text{with} \quad \text{Var}(D_t) = 2\text{Var}(X_t) = 2\sigma_0^2$$

and

$$\text{Corr}(D_t, D_{t-1}) = -1/2 \quad , \text{ unfortunately.}$$



$diff(X)$ -EWMA

Some ARL results

2-sided, for $diff(X)$ -EWMA Monte-Carlo with 10^7 rep.

Δ	$diff(X)$ -EWMA				EWMA
	$\lambda = 0.001$	$\lambda = 0.005$	$\lambda = 0.01$	$\lambda = 0.05$	$\lambda = 0.1$
0.000	499.043 _{0.158}	499.941 _{0.158}	500.011 _{0.158}	499.909 _{0.158}	500
0.001	352.154 _{0.082}	446.749 _{0.134}	480.844 _{0.150}	498.781 _{0.157}	200.366
0.005	156.924 _{0.024}	195.202 _{0.037}	269.405 _{0.068}	476.528 _{0.150}	81.377
0.010	100.950 _{0.013}	116.058 _{0.017}	144.803 _{0.027}	417.469 _{0.130}	53.341
0.050	32.538 _{0.003}	33.973 _{0.005}	36.002 _{0.004}	79.414 _{0.018}	20.028
0.100	19.430 _{0.002}	19.922 _{0.002}	20.591 _{0.002}	29.565 _{0.004}	13.343
0.500	5.703 _{0.000}	5.742 _{0.000}	5.790 _{0.000}	6.227 _{0.001}	5.439
1.000	3.368 _{0.000}	3.380 _{0.000}	3.394 _{0.000}	3.520 _{0.000}	3.768
2.000	2.019 _{0.000}	2.022 _{0.000}	2.026 _{0.000}	2.060 _{0.000}	2.688
3.000	1.537 _{0.000}	1.538 _{0.000}	1.540 _{0.000}	1.553 _{0.000}	2.047
4.000	1.181 _{0.000}	1.182 _{0.000}	1.183 _{0.000}	1.192 _{0.000}	1.993
5.000	1.028 _{0.000}	1.028 _{0.000}	1.028 _{0.000}	1.031 _{0.000}	1.927



Conclusions

- Performance measures of step change control charts under drift could be calculated accurately (except 2-sided CUSUM and GRSR).
- It seems so that the additional (huge) efforts of the more sophisticated charts (GLR for step change and drift, GEWMA, local regression) are not worth the effort.
- The naïve approach does not work for very small drift coefficients such as $\Delta < 0.1$.
- Why do we consider such small drift coefficients (down to $\Delta = 0.0001$)? Nobody checks the detection performance for step change sizes smaller than $\delta = 0.1$.



Conclusions continued

- Clarification needed whether the drift itself is the target or the increasing mean (beyond a critical level)!
- Apply your favorite step change scheme and get a sufficiently powerful drift detection scheme.
- Re-consider SWEET's coupled schemes.



Miscellaneous

ARLs of step change schemes under drift

In R library `spc`

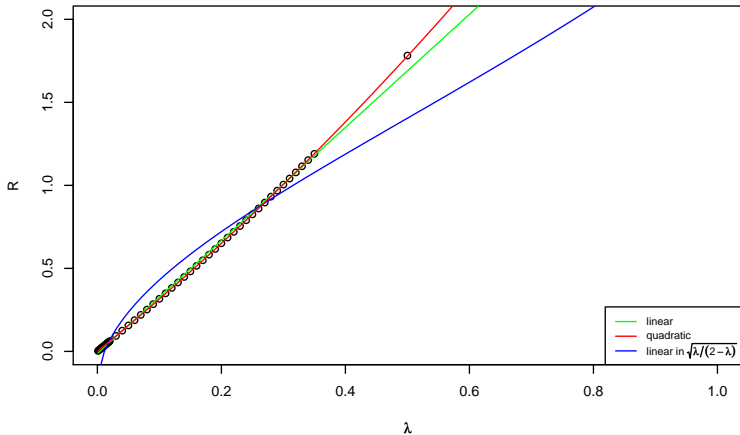
In the recent version of the R library `spc` the following new functions are implemented:

- `xDewma.arl(1, c, delta, zr = 0, hs = 0, sided = "one", ...`
- `xDcusum.arl(k, h, delta, hs = 0, sided = "one", ...`
- `xDgrsr.arl(k, g, delta, zr = 0, hs = NULL, sided = "one", ...`
- `xDshewhartrunrules.arl(delta, c = 1, m = NULL, type = "12")`



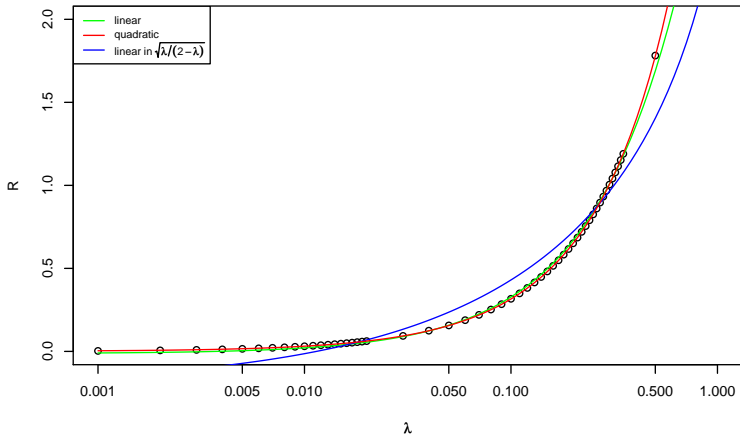
Calculating alarm threshold for $\text{diff}(X)$ -EWMA

interesting regression pattern



Calculating alarm threshold for $\text{diff}(X)$ -EWMA

interesting regression pattern



Calculating alarm threshold for $\text{diff}(X)$ -EWMA

interesting regression pattern

