Europa-Universität Viadrina Frankfurt (Oder)

The art of evaluating monitoring schemes – how to measure the performance of control charts?

Sven Knoth

Dep. of Statistics
European University Viadrina

June 2004

Synopsis

1. Notational Preliminaries.
2. Some history/drawbacks of performance measuring in SPC.
3. Concepts in detail.
4. Computational remarks.
5. Résumé.

The change-point model

Modeling of a stochastic process with a possible distributional change

Sequence of random variables $X_1, X_2, \ldots$ with cdf $\{F_{(i)}\}$ and a certain (unknown) time point m = **change-point** with

$$F_{(i)} = \begin{cases} F_0 & , \; i < m \\ F_1 & , \; i \geq m \end{cases} .$$

Example: $F_0 = \mathcal{N}(\mu_0, 1)\,,\; F_1 = \mathcal{N}(\mu_1, 1)$ + independence

Notation:

$\{X_i\}_{i=1}^{m-1}$ – process **in control**,

$\{X_i\}_{i=m}^{\infty}$ – process **out of control**.

Control charts – SPC at work

different names, same concepts:
change point detection, continuous inspection, surveillance, monitoring ...

Aim: Detect rapidly and reliably, whether there appeared change-point m!

- Transformation $\left\{X_i\right\}_{i=1,2,\ldots,n} \to Z_n$ and
- Stopping time $L = \min\left\{n \in \mathbb{N}:\ Z_n \notin \mathcal{O} = [c_l^*, c_u^*]\right\}$,

At time point L observation is stopped & the scheme signals an **alarm**.

L is a random value on $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Measuring control chart performance

1. Shewhart (192x,193x) similar to tests: error probabilities,
2. Aroian/Levene (1950) average spacing number and average efficiency number,
3. Girshick/Rubin (1952) Bayesian framework,
4. Page (1954) introduced term ARL *as the average number of articles inspected between to successive occasions when rectifying action is taken.*
5. Barnard (1959) *If it were thought worthwile one could use methods analogous to these given by Page (1954) and estimate the average run length as a function of the departure from the target value. However, as I have already indicated, such computations could be regarded as having the function merely of avoiding unemployment amongst mathematicians.*

Average Run Length (ARL)

Notation: $E_m(.)$ expectation for given change-point m.

Definition:

$$ARL = \begin{cases} E_\infty(L) & \text{, process in control} \\ E_1(L) & \text{, process out of control} \end{cases} .$$

Note that for dealing with the ARL, the sequence $\{X_i\}$ is (strong) stationary with the same probability law for all i. Thus, e. g.,

$$ARL = E_\mu(L) =: \mathcal{L}_\mu \, .$$

Measuring control chart performance II

6. SHIRYAEV (1961/3) random change-point model

$$P(M = m) = \begin{cases} \pi & , \; m = 0 \\ (1-\pi)\,(1-p)^{m-1} p & , \; m > 0 \end{cases} \; , \; \pi \in [0,1)\,, \; p \in (0,1)$$

and minimize

$$\begin{cases} P_{\pi,p}(L < M) + c\, E_{\pi,p}(L-M)^+ & \text{for all s. t. } L \\ E_{\pi,p}(L - M | L \geq M) & \text{for all s. t. } L \text{ with } P_{\pi,p}(L < M) \leq \alpha \end{cases}$$

7. ... 9. $\boxed{E_\infty(L) \geq A}$

7. ROBERTS (1966) $\mathcal{D} := \lim_{m\to\infty} E_m\big(L - m + 1 \,|\, L \geq m\big)$

("steady-state ARL", R. "replaced" ∞ by 9)

8. LORDEN (1971) $\mathcal{W} := \sup_{m\geq 1} \operatorname{ess\,sup} E_m\big((L - m + 1)^+ \,|\, \mathcal{F}_{m-1}\big)$

9. POLLAK/SIEGMUND (1975) $\mathcal{D}_{\text{PS}} := \sup_{m\geq 1} E_m\big(L - m + 1 \,|\, L \geq m\big)$

Anyway,

the ARL is the dominating measure!

Possible reasons:

- **Shewhart chart**

  $\mathcal{W} = \mathcal{D} = \mathcal{D}_{\text{PS}} = \mathcal{L} = E_1(L) = E_m(L - m + 1 | L \geq m)$.

- **CUSUM**

  $\mathcal{W} = \mathcal{D}_{\text{PS}} = \mathcal{L}$,

  modifications: $\mathcal{D} = \mathcal{D}_{\text{PS}} \neq \mathcal{L}$.

But:

- **EWMA**

  All measures provide different values.

- ...

Lai 1995

Sequential changepoint detection in quality control and dynamical systems, *J. R. Stat. Soc., Ser. B*, **57**, *613-658.*

- *The ARL constraint $E_{\theta_0}(T) \geq \gamma$ stipulates a long expected duration to false alarm. However, a large mean of T does not necessarily imply that the probability of having a false alarm before some specified time m is small. In fact, it is easy to construct positive integer-value random variables T with a large mean γ and also having a high probability that $T = 1$.*
- *In practice, the system only fails after a very long in-control period and we expect many false alarms before the first correct alarm. It is therefore much more relevant to consider*
  - *(a) the probability of no false alarm during a typical (steady state) segment of the base-line period and*
  - *(b) the expected delay in signaling a correct alarm,*

  instead of the ARL which is the mean duration to the first alarm assuming a constant in-control or out-of-control value.

Frisén 2003

Statistical Surveillance. Optimality and Methods.
International Statistical Review, **71**, *403-434.*

The whole subsection 3.5 (titled "ARL") is dedicated to a reckoning with "the criterion of minimal ARL".

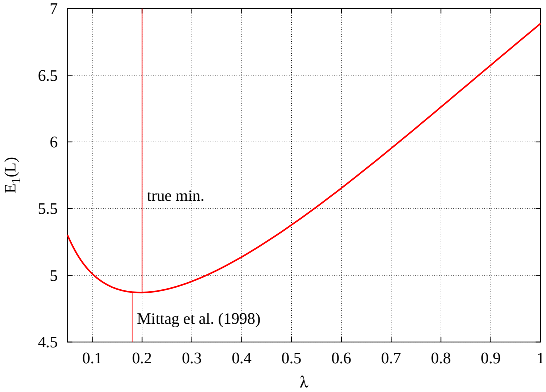
For demonstrating the possible drawbacks of utilizing "the criterion of minimal ARL", let us consider one-sided EWMA-S^2 control charts, which resemble good candidates for illuminating this phenomenon.

One-sided EWMA-S^2 chart for monitoring variance

$$Z_i = (1-\lambda) Z_{i-1} + \lambda S_i^2 \,,\ i \geq 1 \,,\ Z_0 = z_0 = \sigma_0^2 = 1 \,,\ S_i^2 = 1/4 \sum_{j=1}^{5} (X_{ij} - \bar{X}_i)^2 \,,$$

$$L = \inf \left\{ i \in \mathbb{N} : Z_i > \sigma_0^2 + c \sqrt{\lambda/(2-\lambda)} \sqrt{2/4}\, \sigma_0^2 \right\} \,,$$

$E_\infty(L) = 250 \,,\ \sigma_1^2 = 1.5^2$ (as in Mittag et al., 1998).

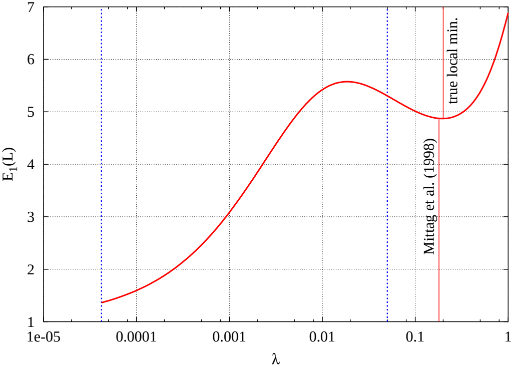


One-sided EWMA-S^2 chart for monitoring variance II

$$Z_i = (1-\lambda) Z_{i-1} + \lambda S_i^2\,,\; i \geq 1\,,\; Z_0 = z_0 = \sigma_0^2 = 1\,,\; S_i^2 = 1/4 \sum_{j=1}^{5} (X_{ij} - \bar{X}_i)^2\,,$$

$$L = \inf \left\{ i \in \mathbb{N} : Z_i > \sigma_0^2 + c \sqrt{\lambda/(2-\lambda)} \sqrt{2/4}\, \sigma_0^2 \right\}\,,$$

$E_\infty(L) = 250\,,\; \sigma_1^2 = 1.5^2$ (as in Mittag et al., 1998).



$\lambda = 0.000\,042$,

$c = 0.000\,064\,375\,308$,

$\widehat{E_\infty(L)} = 250.103 \pm 0.091$,

$\widehat{E_1(L)} = 1.3628 \pm 0.0000$,

10^9 rep.

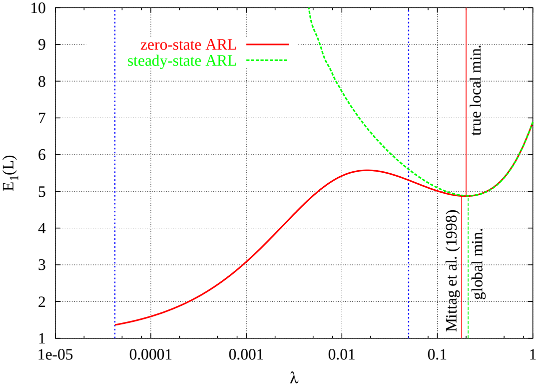
$P_\infty(L = 1) \approx 0.4!$

One-sided EWMA-S^2 chart for monitoring variance III

$$Z_i = (1-\lambda) Z_{i-1} + \lambda S_i^2 \,,\ i \geq 1 \,,\ Z_0 = z_0 = \sigma_0^2 = 1 \,,\ S_i^2 = 1/4 \sum_{j=1}^{5} (X_{ij} - \bar{X}_i)^2 \,,$$

$$L = \inf \left\{ i \in \mathbb{N} : Z_i > \sigma_0^2 + c\sqrt{\lambda/(2-\lambda)}\sqrt{2/4}\,\sigma_0^2 \right\} \,,$$

$E_\infty(L) = 250 \,,\ \sigma_1^2 = 1.5^2$ (as in Mittag et al., 1998).



$\lambda = 0.000\,042$,

$c = 0.000\,064\,375\,308$,

$\widehat{E_\infty(L)} = 250.103 \pm 0.091$,

$\widehat{E_1(L)} = 1.3628 \pm 0.0000$,

10^9 rep.

$P_\infty(L=1) \approx 0.4!$

Assessing the steady-state

$$\mathcal{F}_{m-1} = \sigma\big(X_1,\ldots,X_{m-1}\big)\,,\ \mathcal{F}^*_{m-1} = \sigma\big(\mathcal{F}_{m-1}\cap\{L > m-1\}\big)\,,$$

$$D^*_{(m)} = E_m\big(L-m+1\,|\,\mathcal{F}^*_{m-1}\big) = \mathcal{L}_{\mu_1}(Z^*_{m-1}) \quad \text{for Markovian schemes}\,,$$

$$Z^*_{m-1} \sim f^*_{m-1}(\cdot)\,,$$

$$\psi_{\mu_0}(z) = \lim_{m\to\infty} f^*_{m-1}(z) \qquad \text{(cf. to Madson/Conn, 1973)}\,,$$

$$D^* = \mathcal{L}_{\mu_1}(Z^*)\,,\ Z^* \sim \psi_{\mu_0}(\cdot) \qquad \boxed{\text{name } D^* \text{ "steady-state delay"}}\,,$$

$$E(D^*) = \int_{\mathcal{O}} \psi_{\mu_0}(z)\,\mathcal{L}_{\mu_1}(z)\,d\mathcal{M}(z)$$

$$= \lim_{m\to\infty}\int_{\mathcal{O}} f^*_{m-1}(z)\,\mathcal{L}_{\mu_1}(z)\,d\mathcal{M}(z) = \mathcal{D} \qquad \text{(steady-state ARL)}$$

$$\approx E(D^*_{(m)}) = E_m\big(L-m+1\,|\,L\ge m\big) \quad \text{already for small } m.$$

Links between D^* and previous measures

$$
\begin{aligned}
D^*_{(m)} &= E_m\big(L - m + 1 \,|\, \mathcal{F}^*_{m-1}\big)\,, \\
E_m\big([L - m + 1]^+ \,|\, \mathcal{F}_{m-1}\big) &= D^*_{(m)} \cdot I_{\{L > m-1\}} + 0 \cdot I_{\{L \le m-1\}}\,, \\
\mathcal{W} &= \sup_{m \ge 1} \operatorname{ess\,sup}\big(D^*_{(m)}\big) = \operatorname{ess\,sup} D^*\,, \\
\mathcal{D}_{\mathrm{PS}} &= \sup_{m \ge 1}\; E\big(D^*_{(m)}\big)\,.
\end{aligned}
$$

Z^* is random variable on $\mathcal{O}$ with $z_0 \in \mathcal{O}$ and density $\psi_{\mu_0}(\cdot)$:

$$
\begin{aligned}
\mathrm{ARL} &= \mathcal{L}_{\mu_1}(z_0)\,, \\
\mathcal{W} &= \operatorname{ess\,sup} \mathcal{L}_{\mu_1}(Z^*)\,, \\
\mathcal{D} &= E\big(\mathcal{L}_{\mu_1}(Z^*)\big)\,, \\
\mathcal{D}_{\mathrm{PS}} &= \max\{\mathcal{L}_{\mu_1}, \mathcal{D}\}\,.
\end{aligned}
$$

One-sided Schemes

- CUSUM: Page (1954)

$$Z_n = \max\left\{0, Z_{n-1} + X_n - k\right\}, \; Z_0 = z_0 ,$$
$$L = \inf\left\{n \in \mathbb{N} : Z_n > h\right\} \qquad \big(k = (\mu_0 + \mu_1)/2\big)$$

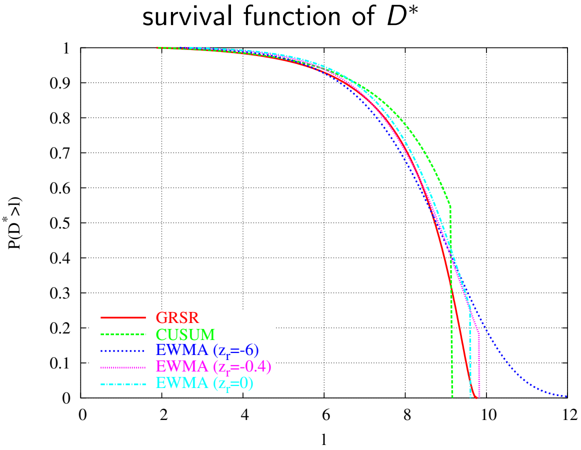
- EWMA: Roberts (1959) (reflecting barrier – Waldmann (1986), Gan (1993))

$$Z_n = \max\left\{z^*_{\text{reflect}}, (1-\lambda)\, Z_{n-1} + \lambda\, X_n\right\}, \; Z_0 = z_0 ,$$
$$L = \inf\left\{n \in \mathbb{N} : Z_n > c\,\sqrt{\lambda/(2-\lambda)}\right\}, \; z^*_{\text{reflect}} = z_r\,\sqrt{\lambda/(2-\lambda)}$$

- GRSR: Girshick/Rubin (1952), Shiryaev (1963/76), Roberts (1966)

$$Z_n = (1 + Z_{n-1})\exp(X_n - k)\,, \; Z_0 = z_0 ,$$
$$L = \inf\left\{n \in \mathbb{N} : Z_n > g\right\} \qquad \big(\exp\big[(\mu_1 - \mu_0)\, X_n - (\mu_1^2 - \mu_0^2)/2\big]\big)$$

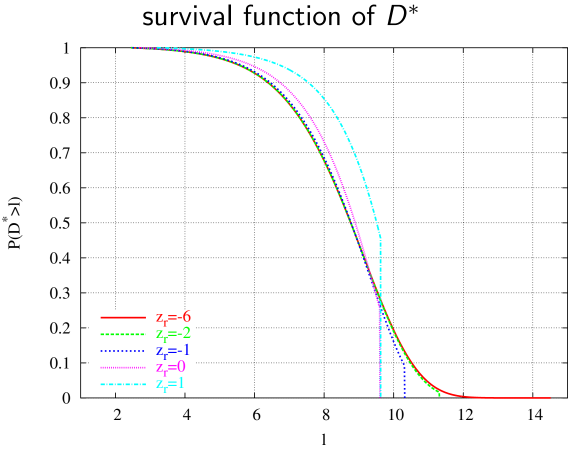
Dealing with D^* for one-sided schemes



$\mu_0 = 0,\ \mu_1 = 1,$
$\lambda = 0.155,$
$E_\infty(L) = 500,$

scheme	$\mathcal{L}$	$\mathcal{W}$	$\mathcal{D}$
GRSR	9.78	$\mathcal{L}$	8.32
CUSUM	9.16	$\mathcal{L}$	8.47
$EWMA_1$	8.73	14.52	8.59
$EWMA_2$	9.26	9.82	8.48
$EWMA_3$	9.60	$\mathcal{L}$	8.51

Dealing with D^* for one-sided EWMA schemes



$\mu_0 = 0,\ \mu_1 = 1,$
$\lambda = .155,$
$E_\infty(L) = 500,$

z_r	$\mathcal{L}$	$\mathcal{W}$	$\mathcal{D}$
-6	8.73	14.52	8.59
-5	8.73	13.86	8.59
-4	8.73	13.12	8.59
-3	8.73	12.28	8.59
-2	8.75	11.32	8.57
-1	8.93	10.32	8.51
0	9.60	$\mathcal{L}$	8.51
1	9.63	$\mathcal{L}$	8.96

Two-sided Schemes

- CUSUM: coupling of 1-sided schemes
- Crosier-CUSUM (1986)

$$Z_n = \begin{cases} 0 & , \; C_n \le k \\ (Z_{n-1} + X_n) \cdot \left(1 - \dfrac{k}{C_n}\right) & , \; C_n > k \end{cases} , \; C_n = |\, Z_{n-1} + X_n \,| \text{ for } n \ge 1 ,$$

$$L = \inf \{ n \in \mathbb{N} : |Z_n| > h \}$$

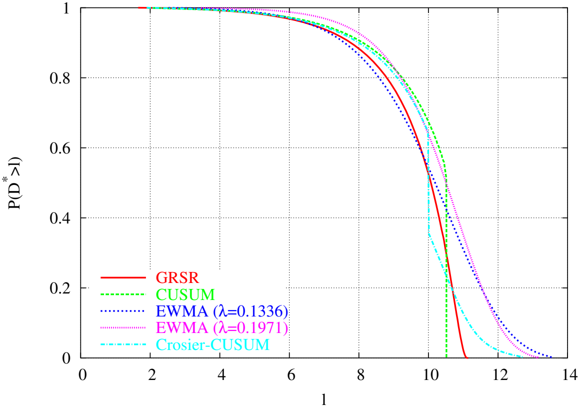
- EWMA: Roberts (1959)

$$Z_n = (1 - \lambda)\, Z_{n-1} + \lambda\, X_n \,, \; Z_0 = z_0 \,, \; L = \inf \left\{ n \in \mathbb{N} : |Z_n| > c\, \sqrt{\lambda/(2-\lambda)} \right\}$$

- GRSR: Pollak/Siegmund (1985) – $Z_n = \left(Z_n^+ + Z_n^-\right)/2$
  performs like coupling of 1-sided schemes $Z_n = \max\{Z_n^+, Z_n^-\} \rightsquigarrow$ coupling
  (PS: $\mathcal{L}_{\mu_1} = 11.142$, $E(D^*) = 9.644$, coupling: $\mathcal{L}_{\mu_1} = 11.142$, $E(D^*) = 9.630$, both: $\mathcal{L}_{\mu_0} = 500$)

Dealing with D^* for two-sided schemes

survival function of D^*



$\mu_0 = 0$, $\mu_1 = 1$,
$E_\infty(L) = 500$,

scheme	$\mathcal{L}$	$\mathcal{W}$	$\mathcal{D}$
GRSR	11.14	$\mathcal{L}$	9.63
CUSUM	10.52	$\mathcal{L}$	9.78
EWMA$_1$	10.19	13.54	9.98
EWMA$_2$	10.52	13.14	10.32
Crosier	10.01	12.73	9.76

α worst case ARL

Definition: $\mathcal{W}_\alpha = \inf\left\{w : P(D^* > w) \leq \alpha\right\}$

chart	shift Δ							measure
	0	0.5	1.0	1.5	2.0	2.5	3.0	
EWMA	689	39	10.0	5.1	3.4	2.6	2.2	$\mathcal{L}$
	*	42	12	6.3	4.3	3.4	2.8	$\mathcal{W}_{.05}$
	*	42	12	6.6	4.7	3.6	3.0	$\mathcal{W}_{.01}$
	*	43	13	7.0	5.0	3.9	3.3	$\mathcal{W}_{.001}$
	*	45	14	8.2	6.0	4.8	4.1	$\mathcal{W}_{<10^{-9}}$
CUSUM	724	34	9.9	5.6	3.9	3.1	2.6	$\mathcal{L}/\mathcal{W}$
GRSR	697	33	10.4	6.2	4.4	3.5	2.9	$\mathcal{L}/\mathcal{W}$

Roberts (1966) A comparison of some control chart procedures. *Technometrics*, **8**, *411-430.*

$\lambda = 0.25$, $c = 2.87$, one-sided, no reflexion (here reflected at $z_r = -6$)

Revisiting Mittag et al. (1998)

Their "winning" scheme is (for monitoring normal variance)

$$Z_i = (1-\lambda) Z_{i-1} + \lambda S_i^2 \,,\ i \geq 1 \,,\ Z_0 = z_0 = \sigma_0^2 = 1 \,,$$

$$S_i^2 = 1/4 \sum_{j=1}^{5} (X_{ij} - \bar{X}_i)^2 \,,$$

$$L = \inf \left\{ i \in \mathbb{N} : Z_i > \sigma_0^2 + c \sqrt{\lambda/(2-\lambda)} \sqrt{2/4}\, \sigma_0^2 \right\} \,,$$

$$E_\infty(L) = 250 \,,\ \lambda = .18 \,,\ c = 2.91 \,.$$

Revisiting Mittag et al. (1998) II

- Replace for EWMA-S^2 the original values λ and c by $\lambda = 0.197$ and $c = 2.971$.
- Remove the $\ln S^2$ based and the S based schemes, and deploy CUSUM-S^2:

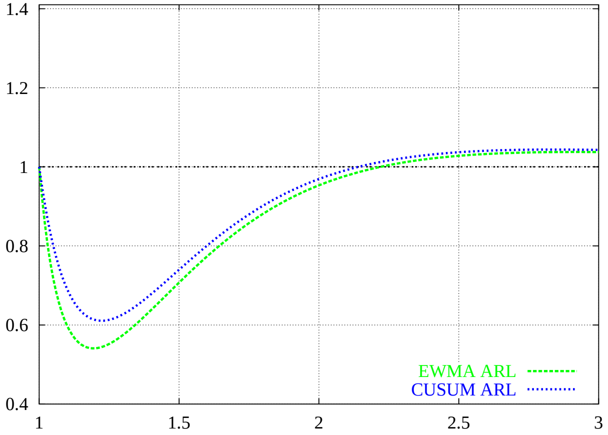
$$Z_i = \max\left\{0, Z_{i-1} + S_i^2 - k\right\}, \; i \geq 1 \,, \; Z_0 = z_0 = 0 \,,$$

$$E_\infty(L) = 250 \,, \; k = 1.4597 \,, \; h = 3.117 \,.$$

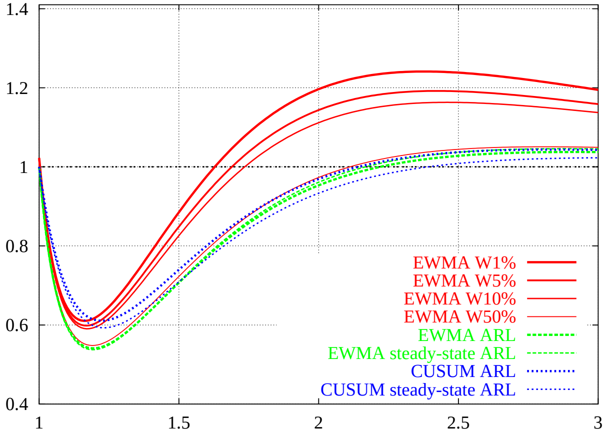
- Determine as in Mittag et al. the efficiency

$$\frac{\text{measure}_{\text{EWMA/CUSUM-}S^2}}{\text{measure}_{\text{Shewhart-}S^2}} \,.$$

Revisiting Mittag et al. (1998) III



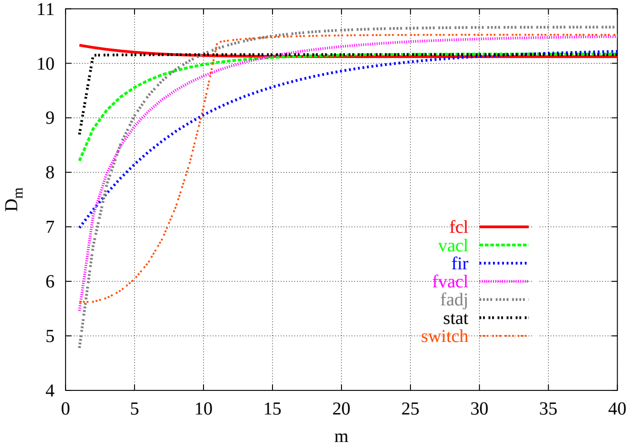
Revisiting Mittag et al. (1998) IV



Final advocacy for the steady-state ARL

$$D_m = E_m(L - m + 1 | L \geq m) \text{ vs. limit } E(D^*)$$

(different EWMA charts with $\lambda = 0.1$, $E_\infty(L) = 500$, $\mu_1 = 1$)



Computational remarks

- It seems so that fast accurate computation of

  $\mathcal{L}_\mu(z)$, $\psi_\mu(z)$, $P_\infty(L \le n)$, $P_m(L \le n+m-1 | L \ge m-1)$, $\mathcal{D}$

  is possible for univariate distributions which span to the whole real line (or at least they ensure smooth transition kernels over $\mathcal{O}$).
- For univariate distributions restricted to the half real line or similar, collocation (might) provide the same for

  $\mathcal{L}_\mu(z)$, $\psi_\mu(z)$, $\mathcal{D}$.
- Exact results are available in only some cases.
- Monte-Carlo simulations are always possible. They demand, however, elaborate cpu time. Eventually, they are useful for validating of results.
- Mostly, the universal Markov chain approach provides suitable results, while programming efforts remain on a reasonable level.
- All become worse for multivariate distributions or dependent data.

Résumé

- Be cautious in applying the "minimal ARL criterion", especially in the one-sided case.
- Use the "steady-state ARL criterion". GRSR is asymptotically optimal in the one-sided case. The performance differences between the common schemes are practically negligible. Thus, look for the "steady-state ARL" optimal mode of your favorite scheme/chart.
- The pessimistic user has to apply the CUSUM scheme, which operates with probabilities of 10 % and considerably more under worst case condition.
- Use the Markov chain approach in the first place as analysis tool. It provides all the information you need with sufficient accuracy and parsimonious expenditure of writing program code. Monte-Carlo simulations might complete your quantitative analysis.