Control Charts Based on Alternative Hypotheses

A. Di Bucchianico, M. Hušková (Prague),
P. Klášterecky (Prague), W.R. van Zwet (Leiden)

Dortmund, January 11, 2005

Goals of this talk

- introduce hypothesis testing framework for control charts in SPC

- develop monitoring procedures for practical out-of-control situations

Contents of talk

- Background:
  - statistical process control: control charts
  - change-point problems
  - sequential analysis
- Testing and control charts
- Some alternative hypotheses
- Some thoughts on performance measures of monitoring procedures
- Likelihood ratio tests
- Asymptotics for critical values
- Simulations
- Future work

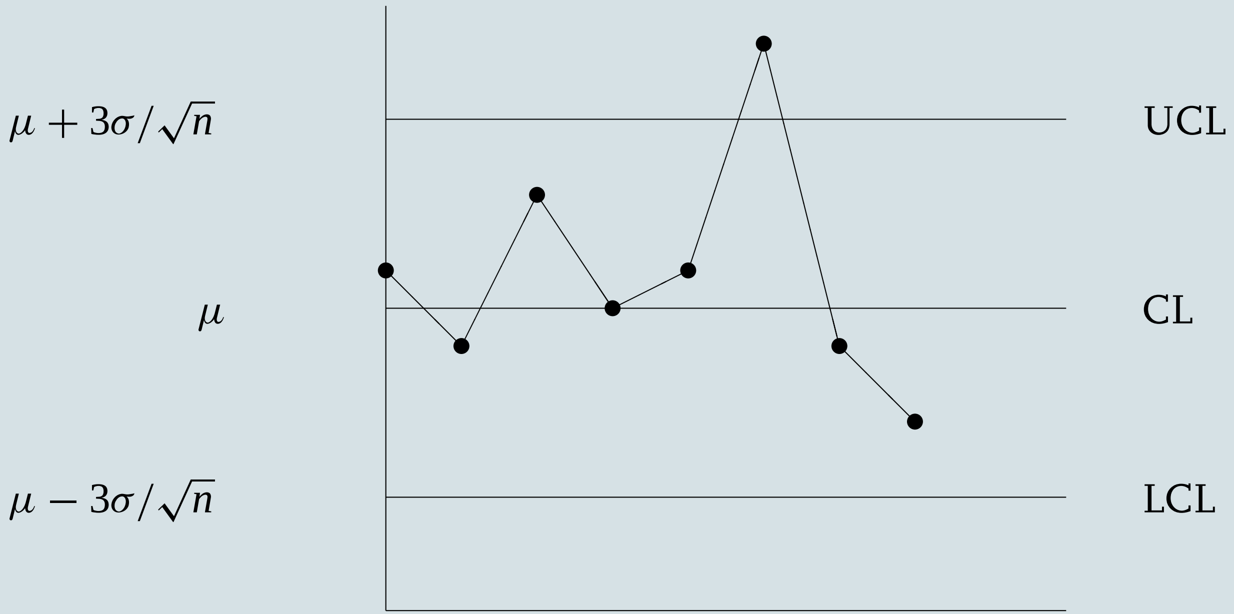
Background

1. SPC (Statistical Process Control)
   - background in industry (Shewhart 1924)
   - uses control charts as monitoring tools (detection of out-of-control situations)
   - emphasis on on-line monitoring (Phase II)
2. changepoint analysis (cf. Lai, J. Roy. Stat. Soc. B 57 (1995))
   - background in mathematical statistics
   - aims at estimation of changepoint
   - main emphasis on retrospective analysis
3. sequential analysis (cf. Lai, Stat. Sinica 11 (2001))
   - developed in military context (Wald, Wolfowitz 1940's)
   - initial emphasis on hypothesis testing

SPC: Shewhart charts

- introduced by Shewhart in 1924
- practical tool without theoretical background
- specific terminology: in-control (common causes), out-of-control (special causes), rational subgroups, ...
- chart signals if summary statistic of ith group is above or below $3\sigma_T$
- variants: $\overline{X}$, R, S, MR, attribute control charts, ...
- additions: VSR (Variable Sampling Rate), runs rules (Western Electric 1956), warning zones, ...

Shewhart $\overline{X}$-chart with control lines.



SPC: CUSUM charts

- introduced in 1954 by Page
- cumulative sums enable to detect small changes of the mean ($< 1.5\sigma$)
- recursive practical form: threshold on cumulative sums:
  $Q_i = \max\{0, Q_{i-1} + X_i - k)\}$
- optimality with respect to ARL proved by Moustakides and Ritov
- performance quickly deteriorates away from optimal alternative (fine-tuning of k and decision threshold)
- additions: FIR (Fast Initial Response) by Lucas
- monograph Hawkins and Olwell

SPC: EWMA chart

- based on ideas from Girshick, Rubin, Roberts and Shiryaev (early 1960's)
- inspired by Bayesian analysis (prior distribution on changepoint)
- $V_i = \lambda(X_i) + (1 - \lambda)V_{i-1}, \quad V_0 = 0$
- $\lambda \to 0$: CUSUM, $\lambda = 1$: Shewhart
- practical choice for λ: $0.1 < \lambda < 0.3$
- performance nearly as good as CUSUM, but less sensitive to non-normality
- decision threshold changes with i

SPC: other charts

- control charts based on robust statistics
- combined control charts (*e.g.*, Shewhart-CUSUM)
- cuscore charts (Box; based on Fisher's efficient score statistics)
- nonparametric control charts
  - linear rank statistics (Wilcoxon, . . .)
  - precedence statistics (Chakraborti and V.d. Laan)
  - regression-type control charts (monitoring linear profiles)
- . . .

SPC: Phase I and II

Phase I

- retrospective (usually pilot study of new production process)
- determination of in-control parameter values

Phase II

- on-line (full scale process)
- uses in-control parameter values from Phase I

Detection performance of control charts in Phase II may heavily deteriorate when using estimated parameters in the control limits (see *e.g.*, Chakraborti, Comm. Stat. Simul. 29 (2000)). Robust estimation of parameters is required in noisy environments (see *e.g.*, Gather et al. , Estadistica 53 (2001)).

Hawkins et al. (J. Qual. Techn. 35 (2003)) argue that application likelihood ratio methods from changepoint analysis in SPC context makes Phase I/II distinction superfluous.

Neyman-Pearson Lemma

sample $X_1, \ldots, X_n$

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1$$

$$\text{reject } H_0 \text{ if } \frac{P_{H_1}(x_1, \ldots, x_n)}{P_{H_0}(x_1, \ldots, x_n)} > c.$$

This test has maximal power under all tests with the same type I error.

Frisén and De Maré, Biometrika 78 (1991), have version of Neyman-Pearson for detection of critical events at given time point.

Tests for composite hypotheses $H_0 : \mu \le \mu_0$ against $H_1 : \mu > \mu_0$ may be treated similarly if likelihood ratio is monotone.

Wald SPRT (Sequential Probability Ratio Test)

sample $X_1, X_2, \ldots$

$H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$

$$\begin{array}{lll} \text{accept } H_0 & \text{if} & \dfrac{P_{H_1}(x_1, \ldots, x_n)}{P_{H_0}(x_1, \ldots, x_n)} < a \\ \text{accept } H_1 & \text{if} & \dfrac{P_{H_1}(x_1, \ldots, x_n)}{P_{H_0}(x_1, \ldots, x_n)} > b \\ \text{continue testing} & \text{if} & a \leq \dfrac{P_{H_1}(x_1, \ldots, x_n)}{P_{H_0}(x_1, \ldots, x_n)} \leq b \end{array}$$

This procedure simultaneously minimizes the ASN's (Average Sample Number) under H_0 and H_1, given $P_{H_0}(\text{reject } H_0) \leq \alpha$ and $P_{H_1}(\text{reject } H_0) \leq \beta$.

Summary of existing procedures

Existing procedures either are not based on optimality criteria or are optimal w.r.t. to ARL (Average Run Length).

However, run length distributions are highly skewed (often close to geometric distribution). Performance should be judged on other features of run length distributions (*e.g.*, quantiles).

Cf. PSD (Probability of Successful Detection) in medical applications (Frisén, Stat. Medicine 11 (1992) and Frisé and Wessmann, Comm. Stat. Simul. 28 (1999)). In case of active surveillance: predictive value of alarm.

It would be nice to reconcile classical hypothesis testing with sequential detection. This would allow to develop optimal monitoring procedures for specific alternative hypotheses.

Examples of alternative hypotheses

Usually theoretical studies of control charts focus on persistent changes of the mean (mathematical convenience?!).

Examples of other alternative hypotheses (cf. Gitlow et al., *Quality Management: Tools and Methods for Improvement*, Chapter 8):

- persistent threshold crossing (sea and river levels)
- persistent monotone threshold crossing (tool wear, *e.g.*, chisels)
- persistent shift or drift in variance (wear of bearing)
- epidemic alternatives (*e.g.*, joint SPC APC scheme: feedback controller removes special cause)
- ...

One may also think of monitoring cycles (*e.g.*, business cycles or medical cycles (Frisén)) or profiles (cf. Kim et al., J. Qual. Technology 35 (2003)).

Our Model

Sequential observations $X_1, \ldots, X_k$ ($k \leq n$)

n fixed beforehand (batch industry, medical event,...)

$$X_i = \mu_i + e_i, \quad i = 1, \ldots, n,$$

$e_1, \ldots, e_n$ are i.i.d. error terms with density f symmetric around 0

$\mu_1, \ldots, \mu_n$ are unknown parameters.

Test statistic based on likelihood ratios (details later).

Critical value of test statistics based on false alarm rate (details later).

Possible alternative hypotheses

Persistent change of mean (μ_0 and δ known)

$$\mathrm{H}_0 : \mu_1 = \ldots = \mu_n$$

$$\mathrm{H}_1 : \begin{cases} \mu_i = \mu_0, & i = 1, \ldots, m, \\ \mu_i = \mu_0 + \delta, & i = m+1, \ldots, n \end{cases}$$

Epidemic alternative (μ_0 and δ known)

$$\mathrm{H}_0 : \mu_1 = \ldots = \mu_n$$

$$\mathrm{H}_1 : \begin{cases} \mu_i = \mu_0, & i = 1, \ldots, \ell, \\ \mu_i = \mu_0 + \delta, & i = \ell+1, \ldots, m, \\ \mu_i = \mu_0, & i = m, \ldots, n \end{cases}$$

Possible alternative hypotheses: continued

Persistent non-monotone threshold crossing (δ known)

$$\mathrm{H}_0 : \mu_i \leq \delta, \quad i = 1, \ldots, n,$$
$$\mathrm{H}_1 : \begin{cases} \mu_i \leq \delta, & i = 1, \ldots, m, \\ \mu_i > \delta, & i = m+1, \ldots, n, \end{cases}$$

Persistent monotone threshold crossing (δ known)

$$H_0 : \mu_1 \leq \ldots \leq \mu_n \leq \delta$$
$$H_1 : \begin{cases} \mu_1 \leq \ldots \leq \mu_m < \delta & i = 1, \ldots, m \\ \delta < \mu_{m+1} \leq \mu_{m+2} \leq \ldots \leq \mu_n & i = m+1, \ldots, n \end{cases}$$

Similar alternative hypotheses for the variance are also important (cf. philosophy SPC).

Toy example: Shewhart chart + persistent threshold crossing alternative

X_i independent with densities $f(x-\mu_i)$, where f is a symmetric density that is nonincreasing for $x>0$.
Hypotheses with known δ_0:

$$H_{0i}:\ E\,X_i=\mu_i\le\delta_0$$

against

$$H_{1i}:\ E\,X_i=\mu_i>\delta_0$$

The ML-estimators $\widehat{\mu}_{i0}$ and $\widehat{\mu}_{i1}$ based on X_i under H_{0i} and H_{1i}, respectively, are given by:

$$\widehat{\mu}_{i0}=\begin{cases}X_i & \text{if } X_i\le\delta_0\\ \delta_0 & \text{if } X_i>\delta_0\end{cases}\qquad\qquad \widehat{\mu}_{i1}=\begin{cases}\delta_0 & \text{if } X_i\le\delta_0\\ X_i & \text{if } X_i>\delta_0\end{cases}$$

The loglikelihood ratio based on X_i thus equals

$$\log\left(f(X_i-\widehat{\mu}_{i1})/f(X_i-\widehat{\mu}_{0i})\right)$$

Under normality with common known variance σ^2 this reduces up to a multiplicative constant 1/2 to:

$$Z_i(\delta_0)=\frac{(X_i-\delta_0)^2}{\sigma^2}\,\mathrm{sign}\,(X_i-\delta_0).$$

Since $(t-\delta_0)^2\,\mathrm{sign}\,(\delta_0-t)$ is a monotone function of t, we restrict ourselves to the least favourable situation of the null hypothesis H_0: the restricted null hypothesis H_0^*.

$$H_0^*:\quad \mu_1=\mu_2=\ldots=\delta_0.$$

$$N_S=\begin{cases}\min\{1\le k\le n;\ Z_k(\delta_0)\ge c\} & \text{if } \max_{1\le k\le n}\ Z_k\ge c\\ \infty & \text{if } \max_{1\le k\le n}\ Z_k<c\end{cases}.$$

Critical values Z_i

Standard choice of 3σ control limits yields in-control ARL of $1/0.027 \approx 370$.

$$E_{H_0^*} N_S = \frac{1}{1 - H(c)},$$

where H is distribution function of Z_i. Note that $E_{H_0^*} N_S$ is not meaningful because there is no good way to incorporate the event $\{\max_{1 \le k \le n} Z_k < c\}$.

Alternative 1: jointly considering $H(c)^n$, the probability of no alarm at all, and the conditional ARL:

$$E_{H_0^*}(N_S \mid N_S \le n) = \frac{1}{1 - H(c)} - \frac{n(H(c))^n}{1 - H(c)^n}.$$

Alternative 2: $100\alpha\%$ quantile of the distribution of N_S.

$$1 - H(c)^{n_S(\alpha)} = P_{H_0^*}\left(\max_{1 \le i \le n_S(\alpha)} Z_i(\delta_0) \ge c\right) = P_{H_0^*}(N_S \le n_S(\alpha)) \le \alpha$$

Critical values Z_i: continued

Alternative 2: $100\alpha\%$ quantile of the distribution of N_S. The lower bound $n_S(\alpha)$ for the $100\alpha\%$ quantile of the distribution of N_S satisfies

$$P_{H_0^*}\left(\max_{1\le i\le n_S(\alpha)} Z_i(\delta_0) \ge c\right) = P_{H_0^*}(N_S \le n_S(\alpha)) \le \alpha$$

In other words, the test H_0 against H_1 based on $n_S(\alpha)$ observations has significance level α. Hence, this choice of the critical value can be easily connected to the Neyman-Pearson framework.

Thinking in terms of hypothesis testing with level α it will be typically chosen small, while thinking in terms of the median of the run length distribution we think more about $\alpha = 1/2$.

The next slides present tables with numerical values.

n	$E_{H_0^*}(N_S \mid N_S \le n)$	$P_{H_0^*}(N_S > n)$	quantile of N_S				
			1%	5%	50%	80%	90%
100	49	0.87	7	38	100	100	100
200	96	0.76	7	38	200	200	200
300	140	0.67	7	38	300	300	300
400	183	0.58	7	38	400	400	400
500	223	0.51	7	38	500	500	500
600	260	0.44	7	38	513	600	600
700	296	0.39	7	38	513	700	700
800	330	0.34	7	38	513	800	800
900	361	0.30	7	38	513	900	900
1000	391	0.26	7	38	513	1000	1000
1500	513	0.13	7	38	513	1191	1500
2000	597	0.07	7	38	513	1191	1705

Influence of n on the run length distribution: $c = 9$.

n	$\beta = 0.8$ 90%	95%	99%	$\beta = 0.9$ 90%	95%	99%	$\beta = 1$ $P_{H_0^*}(N_S > n) = 0.05$
100	9.1	10.4	13.4	9.26	10.6	13.6	10.8
200	10.3	11.6	14.7	10.5	11.9	14.9	12.1
300	11.1	12.4	15.5	11.3	12.6	15.7	12.8
400	11.6	12.9	16.0	11.8	13.2	16.2	13.4
500	12.0	13.4	16.4	12.2	13.6	16.7	13.8
600	12.4	13.7	16.8	12.6	13.9	17.0	14.1
700	12.6	14.0	17.1	12.9	14.2	17.3	14.4
800	12.9	14.2	17.3	13.1	14.5	17.6	14.7
900	13.1	14.5	17.6	13.3	14.7	17.8	14.9
1000	13.3	14.7	17.8	13.5	14.9	18.0	15.1
1500	14.1	15.4	18.5	14.3	15.7	18.8	15.9
2000	14.6	16.0	19.1	14.8	16.2	19.3	16.4

Critical values for Shewhart charts with prescribed lower bound for quantiles of run length distribution and for prescribed false alarm rate: $n_s = \beta\, n$.

Detection performance Z_i

distribution function H of $Z_i(\delta_0)$ if $E\ X_i = \delta$ has the form:

$$H(x) = \begin{cases} \Phi\left(-\sqrt{|x|} - \frac{\delta-\delta_0}{\sigma}\right), & x < 0 \\ \Phi\left(\sqrt{x} - \frac{\delta-\delta_0}{\sigma}\right), & x \geq 0 \end{cases}$$

In the next table, we assume that no false alarm takes place before the changepoint, which gives a good impression of the performance, because the Shewhart chart is memoryless.

n	m	c	$P_{H_1}(N_S > n \mid N_s > m)$	H_1-quantile of $N_S - m \mid N_S > m$				
				1%	5%	50%	80%	90%
100	0	10.8	0.676	1	5	62	100	100
100	25	10.8	0.570	1	5	62	75	75
100	50	10.8	0.431	1	5	50	50	50
100	75	10.8	0.245	1	5	25	25	25
500	0	13.8	0.812	3	15	207	481	500
500	125	13.8	0.715	3	15	207	375	375
500	250	13.8	0.567	3	15	207	250	250
500	375	13.8	0.342	3	15	125	125	125
1000	0	15.1	0.860	5	26	353	820	1000
1000	250	15.1	0.771	5	26	353	750	750
1000	500	15.1	0.625	5	26	353	500	500
1000	750	15.1	0.388	5	26	250	250	250

Detection performance for Shewhart charts with prescribed false alarm probability 0.05 under jump alternative with $\delta_1 = \delta_0 + \sigma$: $E X_i = \delta_0 + \sigma$ for $i > m$.

Detection performance Z_i: uniform alternative

Another interesting alternative is that the mean of X_i after the changepoint is uniformly distributed on $[\delta_0, \delta_0 + \gamma\,\sigma]$ with $\gamma > 0$. We then have

$$P_{H_1}(Z_i \le c) = \frac{1}{\gamma\sigma}\int_0^{\gamma\sigma} P(Z_i \le c \mid \mu_i = \delta_0 + x)\, dx = \frac{1}{\gamma\sigma}\int_0^{\gamma\sigma} \Phi\left(\sqrt{c} - \frac{x}{\sigma}\right) dx$$

Writing Φ as an integral and interchanging the order of integration, we obtain that

$$\int_a^b \Phi(u)\, du = b\Phi(b) - a\Phi(a) + \varphi(b) - \varphi(a),$$

where φ is the density of the standard normal distribution. Combining everything, we obtain

$$P_{H_1}(Z_i \le c) = \sqrt{c}\Phi(\sqrt{c}) - (\sqrt{c} - \gamma)\Phi(\sqrt{c} - \gamma) + \varphi(\sqrt{c}) - \varphi(\sqrt{c} - \gamma).$$

General approach: likelihood ratio tests

$$\max_{1 \le m \le k} \frac{\max_{\theta \in \Theta} \prod_{i=1}^{m} f(x_i;\theta) \max_{\eta \in \Theta} \prod_{i=m+1}^{k} f_i(x_i;\eta)}{\max_{\theta \in \Theta} \prod_{i=1}^{k} f(x_i;\theta)}$$

"Persistent non-monotone threshold crossing" and "Persistent monotone threshold crossing" for normally distributed data with known variance:

$$Q_n = \max_{m<k} \frac{1}{2\sigma^2} \left\{ \sum_{i=m+1}^{k} (X_i - \delta)^2 \operatorname{sign}(X_i - \delta) \right\}, \quad k = 2, \ldots, n.$$

Standardization necessary to obtain distributional results (cf. critical values).

$$Q_{kn} = \frac{1}{\sqrt{n}} Q_k \qquad Q_{kk} = \frac{1}{\sqrt{k}} Q_k.$$

Q_{kn} may perform poorly for early changes.

Likelihood ratio: windowed versions (Willsky - Jones)

$$Q_k = \max_{m<k} \frac{1}{2\sigma^2} \left\{ \sum_{i=m+1}^{k} (X_i - \delta)^2 \operatorname{sign}(X_i - \delta) \right\}, \quad k = 2, \ldots, n.$$

The above expression (standardized or not) needs to be maximized once or twice with respect to m.
Problems:

- maximization may be time consuming
- asymptotical distributional results may be hard to obtain

$$\begin{aligned} Q_{k,G} &= \frac{1}{\sqrt{G}} \max_{k-G \le m < k} \frac{1}{2\sigma^2} \sum_{i=m+1}^{k} (X_i - \delta)^2 \operatorname{sign}(X_i - \delta), \\ Q_{k,G}^{\text{simp}} &= \frac{1}{\sqrt{G}} \frac{1}{2\sigma^2} \sum_{i=k-G+1}^{k} (X_i - \delta)^2 \operatorname{sign}(X_i - \delta). \end{aligned}$$

Isotonic regression

Chang and Fricker (J. Qual. Technology 31 (1999)) : isotonic regression method for "Persistent monotone threshold crossing" alternative

$$
\begin{aligned}
M_k &= \sum_{i=1}^{k} (X_i - Z_i)^2 - \sum_{i=1}^{k} (X_i - Y_i)^2 \\
&= \begin{cases} 0 & \text{for } Y_k \le \delta \\ \sum_{i=J}^{k} (X_i - \delta)^2 - \sum_{i=J}^{k} (X_i - Y_i)^2 & \text{for } Y_k > \delta, \end{cases}
\end{aligned}
$$

where

- $Z_1, \ldots, Z_k$ denotes an isotonic regression restricted to increase up to at most δ
- $Y_1, \ldots, Y_k$ is the corresponding unrestricted isotonic regression
- $J = \min\{i : Y_i > \delta\}$.

Intermezzo: scale alternatives

$$X_i = \mu + \sigma_i e_i, \quad i = 1, \ldots, n,$$

where μ is a known constant , σ_i, $i = 1, \ldots, n$ are unknown positive parameters, $e_1, \ldots, e_n$ are i.i.d. random variables with $N(0, 1)$ distribution.
Hypotheses:

$$H_0 : \sigma_i \le \delta, \quad i = 1, \ldots,$$

against

H_1 : there is an m such that $\quad \sigma_i \le \delta, \quad i = 1, \ldots, \quad \sigma_i > \delta \quad i = m+1, \ldots, n,$

The estimator $\widehat{\sigma}^2_{i0}(H_0)$ of σ_i^2 under the restriction that $\sigma_i^2 \le \delta^2$ is the solution of the minimization problem:

$$\min_{\sigma_i^2 \le \delta^2} \left\{ \log \sigma_i^2 + \frac{1}{\sigma_i^2}(X_i - \mu)^2 \right\}.$$

Scale alternatives: continued

$$\widehat{\sigma}_{i0}^2 = \min((X_i - \mu)^2, \delta^2)$$

The estimator $\widehat{\sigma}_{i1}^2$ of σ_i^2 under the restriction that $\sigma_i^2 \geq \delta^2$ is

$$\widehat{\sigma}_{i1}^2 = \max((X_i - \mu)^2, \delta^2).$$

Hence, the loglikelihood ratio for the i-th observation is

$$\left(\log\left(\frac{\delta^2}{(X_i - \mu)^2}\right) - 1 + \frac{(X_i - \mu)^2}{\delta^2}\right) \operatorname{sign}\left(\frac{(X_i - \mu)^2}{\delta^2} - 1\right).$$

$q(t) = (-\log(t^2) - 1 + t^2)\operatorname{sign}(t - 1), \quad t > 0$ is nondecreasing

$$Q_k^{\sigma^2} = \max_{0 \leq j \leq k} \sum_{i=j+1}^{k} \left(-\log\left(-\frac{(X_i - \mu)^2}{\delta^2}\right) - 1 + \frac{(X_i - \mu)^2}{\delta^2}\right) \operatorname{sign}\left((X_i - \mu)^2/\delta^2 - 1\right).$$

Clearly, the test statistics $Q_k^{\sigma^2}$ are similar to the Q_k's.

Critical values

The most common way to control false alarm in SPC is to use run lengths, in particular the in-control ARL (average run length). Sometimes also the SRL (standard deviation of the run length) is taken into account, or even better one uses quantiles.
We reject H_0 as soon as for some $k \leq n$

$$T_k > c_{n,\alpha},$$

where T_k denotes any of the statistics presented in this talk and $c_{N,\alpha}$ is chosen in such a way the significance level is α, i.e.

$$P_{H_0}\left(\max_{1 \leq k \leq n} T_k > c_{n,\alpha}\right) = \alpha.$$

Since the distribution of our statistics do not admit closed-form expression, these critical values have to be obtained by simulations or by asymptotics.

Asymptotics of critical values

$$P_{H_0^*}(N_C(c) > n_C(\alpha)) = (\geq) 1 - \alpha$$

for a prechosen α and a prechosen integer $n_C(\alpha)$. We write $c = c_C(\alpha)$ and rewrite the above relations as

$$P_{H_0^*}\left(\max_{1 \leq k \leq n_C(\alpha)} Q_k < c_C(\alpha)\right) \geq 1 - \alpha,$$

so the probability of false alarm based on $n_C(\alpha)$ observations is $\leq \alpha$.

$$\lim_{n\to\infty} P_{H_0^*}\left(\max_{1<k\leq n} Q_k < c\sqrt{n}(\mathrm{Var}_{H_0^*} Z_1(\delta_0))^{1/2}\right) = P\left(\sup_{0<s<t<1} W(t) - W(s) < c\right)$$

where $\{W(t), t \in (0,1)\}$ is a Wiener process. This implies that

$$\lim_{n\to\infty} c_C(\alpha)\sqrt{n} = c_\alpha\left(\mathrm{Var}_{H_0^*} Z_1(\delta_0)\right)^{1/2},$$

where c_α is such that $P\left(\sup_{0<s<t<1} W(t) - W(s) < c_\alpha\right) = 1 - \alpha$.

Asymptotics of critical values continued

Hence, $c_C(\alpha)$ can be approximated by $c_\alpha\sqrt{n}(\mathrm{Var}_{H_0^*}Z_1(\delta_0))^{1/2}$ and as soon as n is large so does $c_C(\alpha)$ and the procedure will perform poorly in detecting early changes.
Alternative: *weighted* $CUSUM$

$$N_{C,\beta}(c_n(\beta), n) = \min\{k \geq 1;\ (k/n)^{-\beta} Q_k \geq \sqrt{n} c_n(\beta)\}, \quad \beta \in [0, 1/2],$$

where $c_n(\beta)$ is determined in such a way that

$$P_{H_0^*}\left(\max_{1\leq k\leq n} (k/n)^{-\beta} Q_k < \sqrt{n} c_n(\beta)\right) \geq 1 - \alpha.$$

Hence, we reject H_0 as soon as there is a k such that

$$Q_k \geq \sqrt{n}\, c_n(\beta)(k/n)^\beta.$$

Obviously, choosing $\beta = 0$ we have back the original procedure but for $\beta \in (0, 1/2]$ larger weights are assigned to smaller k's.

Asymptotics of critical values: weighted version

$$\lim_{n\to\infty} c_n(\beta) = d_\alpha(\beta)\left(\mathrm{Var}_{H_0^*} Z_1(\delta_0)\right)^{1/2}, \quad \beta \in [0, 1/2),$$

where $d_\alpha(\beta)$ is determined as

$$P\left(\sup_{0<s<t<1} \frac{1}{t^\beta}(W(t) - W(s)) \le d_\alpha(\beta)\right) = 1 - \alpha,$$

where $\{W(t), t \in (0, 1)\}$ is a Wiener process. It implies that $c_n(\beta)(\mathrm{Var}_{H_0^*} Z_1(\delta_0)/n)^{-1/2}$ can be approximated by $d_\beta(\alpha)$.
For $\beta = 1/2$ we have for $c_n(1/2)$ the approximation $d_{\alpha,n}(1/2)(\mathrm{Var}_{H_0^*} Z_1(\delta_0))^{1/2}$, where

$$P\left(\sup_{1\le s\le t\le n} \frac{1}{\sqrt{t}}(W(t) - W(s)) \le d_{\alpha,n}(1/2)\right) = 1 - \alpha.$$

We can show that $\lim_{n\to\infty} c_n(\beta) = O(1), \quad \beta \in [0, 1/2)$, while $c_n(1/2) = O(\sqrt{\log\log n})$.

Background of asymptotics I

Komlós et al. (Z. Wahr. Verw. Gebiete 32 (1975) and 34 (1976))

$Y_1, \ldots, Y_n$ i.i.d. with zero mean, unit variance and $\mathrm{E}\,|X_i|^{2+\Delta} < \infty$ for some $\Delta > 0$. Then there exists a sequence of Wiener processes $W_n = \{W_n(t), t \geq 0\}$, $n \geq n_0$, such that, as $n \to \infty$,

$$P\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} Y_i - W_n(k)\right| > x\right) \leq C_1 n x^{-(2+\Delta)}, \quad x > 0$$

$$\max_{1\leq k\leq n} \frac{1}{k^{1/(2+\Delta)}}\left|\sum_{i=1}^{k} Y_i - W_n(k)\right| = O_P(1)$$

and

$$\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} X_i - W_n(k)\right| = o_P(n^{1/(2+\Delta)}).$$

Background of asymptotics II

Erdös-Darling theorems

$$P\left(d_1(\log n) \max_{1<k\le n} \frac{1}{\sqrt{k}} \sum_{i=1}^{k} Y_i \le y + d_2(\log n)\right) \to \exp\{-\exp\{-y\}\}.$$

and

$$P\left(d_1(\log n) \max_{1<k\le n} \frac{1}{\sqrt{k}} \left|\sum_{i=1}^{k} Y_i\right| \le y + d_2(\log n)\right) \to \exp\{-2\exp\{-y\}\},$$

where

$$d_1(t) = \sqrt{2\log t}, \quad t > 1,$$

$$d_2(t) = 2\log t + \frac{1}{2}\log\log t - \log(\pi), \quad \log t > 1.$$

Simulation of critical values

Sample quantiles ξ_α are asymptotically normal with variance $\frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}$ (Bahadur)

We estimated $f^2(\xi_\alpha)$ by applying a kernel density estimator to a pre-run of reasonable size.

Example: for the 90% quantile the standard deviation equals $\left(\frac{0.9*0.1}{0.15^2 n}\right)^{1/2}$, because the the density at the quantile approximately equals 0.15 . Hence, if $n = 10000$, then the standard deviation of this quantile approximately equals 0.02.

Performance simulation

Simulation performed using R (see `www.r-project.org`), version 1.8.1 on a Unix platform.

Its new built-in procedure `isoreg()` is much faster in computing the isotonic regression (needed for simulations of the M_n statistics) than the PAVA algorithm.

Source codes of all procedures we used are available at
`http://www.karlin.mff.cuni.cz/~klaster/compstat04`

Simulation improvement

Since $\mathrm{Var}\, N_S \approx \mathrm{E}\,(N_S)^2$, so increase in ARL leads to increase in variance of estimates.
Variance reduction techniques have been developed for these simulations.
Jun and Choi, Comm. Stat. A 22 (1993) present two techniques for CUSUM simulations:

1. hazard controlled estimator (following a suggestion of Ross to use the total hazard of a Markov chain)
2. ratio estimators using a cycle.

Simulation setup

$n = 100$, $\alpha = 0.05$, $G = 0.1n = 10$, $\sigma = 1$, 1000 repetitions.

Two types of changes, each of which has a "small" and a "large" version:

small version corresponds to change of mean of σ

large version corresponds to change of mean of 3σ

general change mean is uniformly distributed on $[\delta - r, \delta]$ before change-point, and uniformly distributed on $[\delta, \delta + r]$ after the change-point, where $r = 1$ (small change) or $r = 3$ (large change), respectively (so the expected mean increases by 1 and 3, respectively).

gradual change mean increases on 5 equally spaced intervals before change point from μ to δ and on 5 equally spaced intervals after the change-point from δ to $2\delta - \mu$, where $\delta = 1$ and $\mu = 0.5$ (small change) or $\delta = 3$ and $\mu = 1.5$ (large change).

Simulation interpretation

Detection delays must be interpreted with care.

Suppose change at x and number of observations equals n, then in our simulations the number $n - x$ indicates either

- a detection at time point n
- or no detection at all

Summary statistics of simulation include (run lengths are skewed!):

- mean
- standard deviation
- quantiles

Simulation results: general small change

m	Q_{kn} mean; sd	Q_{kk} mean; sd	$Q_{n,G}$ mean; sd	$Q_{n,G}^{\mathrm{simp}}$ mean; sd	M_n mean; sd
5	42.34; 15.07	37.08; 23.10	52.12; 31.69	51.54; 31.47	37.99; 20.81
15	42.95; 15.68	44.87; 23.05	48.79; 29.11	48.69; 28.99	38.67; 21.47
25	41.88; 14.38	47.51; 19.86	46.10; 25.10	45.68; 24.86	37.34; 19.10
40	41.35; 12.47	49.39; 13.89	41.80; 19.40	41.46; 19.32	36.45; 16.25
60	35.32; 6.56	38.16; 9.73	31.18; 11.96	31.06; 11.86	30.47; 10.88

Gradual small change

m	Q_{kn} mean; sd	Q_{kk} mean; sd	$Q_{k,G}$ mean; sd	$Q_{k,G}^{\text{simp}}$ mean; sd	M_k mean; sd
5	74.96; 16.29	80.98; 20.88	82.96; 20.38	82.96; 20.09	75.19; 20.06
15	69.49; 14.55	76.26; 15.85	75.88; 16.58	75.67; 16.78	69.72; 17.41
25	63.37; 12.53	69.48; 13.36	66.67; 16.24	66.60; 16.13	62.79; 15.68
40	54.29; 9.03	57.96; 9.32	55.21; 11.38	55.24; 11.20	53.45; 11.51
60	38.43; 4.20	39.01; 8.77	37.30; 8.24	37.36; 7.89	37.06; 7.87

General large change

m	Q_{kn} mean; sd	Q_{kk} mean; sd	$Q_{k,G}$ mean; sd	$Q_{k,G}^{\text{simp}}$ mean; sd	M_k mean; sd
5	10.45; 3.99	5.35; 3.33	6.22; 3.19	8.03; 2.52	6.32; 3.57
15	10.52; 3.90	7.42; 3.82	6.15; 3.32	8.15; 2.52	6.24; 3.53
25	10.60; 4.35	7.80; 3.22	5.70; 2.31	7.60; 1.71	5.30; 2.71
40	10.71; 3.93	10.64; 5.38	6.34; 3.20	8.18; 2.51	6.47; 3.67
60	10.57; 3.83	12.39; 5.58	6.16; 3.05	8.07; 2.38	6.29; 3.49

Gradual large change

m	Q_{kn} mean; sd	Q_{kk} mean; sd	$Q_{k,G}$ mean; sd	$Q_{k,G}^{\text{simp}}$ mean; sd	M_k mean; sd
5	42.34; 8.60	39.47; 13.24	44.50; 15.26	44.07; 15.24	37.78; 10.78
15	39.69; 8.47	40.49; 10.99	39.98; 14.86	39.57; 14.75	34.83; 10.61
25	37.45; 7.38	40.24; 9.25	37.45; 12.88	37.20; 12.63	33.18; 9.51
40	33.29; 7.04	37.63; 7.99	32.06; 11.09	31.90; 10.90	28.77; 8.59
60	27.37; 5.17	32.27; 5.90	24.89; 7.79	24.84; 7.62	22.89; 6.61

Conclusions

- Chang and Fricker's M_k has the smallest mean (both general and gradual case)
- Q_{kn} has a considerably smaller standard deviation.
- M_k has better low quantiles
- Q_{kn} has better median and higher quantiles

Summary: In most cases Q_{kn} performs better, but M_k sometimes detects extremely fast.

Future work

- other alternative hypotheses (in particular, changes of the variance)
- composite alternative hypotheses
- (asymptotical) distributional results
- combination of procedures
- application to real data
- …