Risk-Based Metrics for Performance Evaluation of Control Charts

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß

Department of Mathematics & Statistics,

Helmut Schmidt University, Hamburg

Murat Caner Testik

Department of Industrial Engineering,

Hacettepe University, Ankara

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Performance Evaluation of Control Charts

Introduction

Performance Evaluation of Control Charts

Performance of control chart often evaluated by **run length** L, i.e., number of observations plotted on chart until alarm.

Performance metric: mean $E[L]$, **average run length** (ARL).

If process in control, L and ARL desired "sufficiently large" (e.g., ARL $\approx$ 370.4) since any alarm false signal.

In contrast, alarm for out-of-control process (= true signal) desired as early as possible, small delay in detection.

(Knoth, 2006; Montgomery, 2009)

Although most common metric, chart design only relying on ARL also criticized (Kenett & Pollak, 2012; Montgomery, 2009).

For example, if (i.i.d.) Shewhart chart, then L geometric.
Geometric distribution very skewed and large std. dev.,
so ARL very limited impression about typical RL behaviour.

Also often large discrepancy between ARL and median.

Therefore, also different performance metrics in literature,
like standard deviation of L, quantiles of L $(\ldots)$

(. . .) approach considered here:

Evaluate chart performance using **risk metrics**.

Risk associated with possibility of undesired event (e.g., loss).

Distribution of risk helps to prepare for case when undesired event occurs (Klüppelberg et al., 2014; McNeil et al., 2015).

Most basic risk metric: **Value at Risk (VaR)**,

i.e., quantile of risk's loss distribution.

$(1-\alpha)$-quantile of loss distribution as threshold value,

which only exceeded in worst $\alpha \cdot 100\,\%$ of cases.

But if undesirable event occurs, then

actual loss $\geq$ threshold value (= VaR).

Therefore, better to average VaRs or exceedances thereof to express "typical loss" if exceedance of VaR occurs.

Idea: Risk metrics for assessing chart performance.

If control chart evaluated for

- in-control performance, then risk of early occurrence of false alarm, so risk of low run lengths;
- out-of-control performance, then risk of delayed detection, so risk of high run lengths.

Outline:

- computation of risk metrics for run length distributions;
- application to common types of control chart;
- sketch of further application area:

  risk metrics to analyze estimation uncertainty

  in evaluating chart performance (Phase-I analysis).

HELMUT SCHMIDT UNIVERSITÄT

Universität der Bundeswehr Hamburg

MATH STAT

Risk Metrics for Run Length Distributions

Definition & Computation

Considered risk metrics: (Göb, 2011)

- discrete α-quantile of L in in-control situation:

$$\text{VaR}_\alpha^{\text{ic}} := \min\{x \in \mathbb{N} \mid P(L \leq x) \geq \alpha\};$$

- discrete $(1-\alpha)$-quantile of L in out-of-control situation:

$$\text{VaR}_\alpha^{\text{ooc}} := \min\{x \in \mathbb{N} \mid P(L \leq x) \geq 1-\alpha\};$$

we have $P(L < \text{VaR}_\alpha^{\text{ic}}) < \alpha$ and $P(L > \text{VaR}_\alpha^{\text{ooc}}) \leq \alpha$,

i.e., bound for L in worst $\alpha \cdot 100\,\%$ of all cases.

Risk Metrics for Run Length Distributions

Considered risk metrics: (Göb, 2011)

Conditional mean run length in case of violating VaR, i.e.,

- **tail conditional expectation (TCE)** in in-control situation:

$$\mathsf{TCE}_\alpha^{\mathsf{ic}} \;:=\; E\big[L \;\big|\; L < \mathsf{VaR}_\alpha^{\mathsf{ic}}\big];$$

- TCE in out-of-control situation:

$$\mathsf{TCE}_\alpha^{\mathsf{ooc}} \;:=\; E\big[L \;\big|\; L > \mathsf{VaR}_\alpha^{\mathsf{ooc}}\big];$$

"typical" run length in worst $\alpha \cdot 100\,\%$ of all cases.

Considered risk metrics: (Göb, 2011)

Possible alternative to TCE:

- **expected shortfall (ES)** by averaging VaR:

$$\mathrm{ES}^*_\alpha \;:=\; \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}^*_u \,\mathrm{d}u.$$

How to compute these measures in practice?

Use L's **survival function** $S(x) = P(L > x)$.

Then (. . .)

- β-quantile of L as $\min\{x \in \mathbb{N} \mid S(x) \leq 1-\beta\}$,

- $\text{TCE}_\alpha^{\text{ic}} = \text{VaR}_\alpha^{\text{ic}} - \dfrac{\text{VaR}_\alpha^{\text{ic}} - \sum_{y=0}^{\text{VaR}_\alpha^{\text{ic}}-1} S(y)}{1 - S\big(\text{VaR}_\alpha^{\text{ic}} - 1\big)}$,

- $\text{TCE}_\alpha^{\text{ooc}} = 1 + \text{VaR}_\alpha^{\text{ooc}} + \dfrac{E[L] - \sum_{y=0}^{\text{VaR}_\alpha^{\text{ooc}}} S(y)}{S\big(\text{VaR}_\alpha^{\text{ooc}}\big)}$,

- $\text{ES}_\alpha^{\text{ic}} = \text{TCE}_\alpha^{\text{ic}} \dfrac{1 - S\big(\text{VaR}_\alpha^{\text{ic}} - 1\big)}{\alpha} + \text{VaR}_\alpha^{\text{ic}} \dfrac{S\big(\text{VaR}_\alpha^{\text{ic}} - 1\big) - (1-\alpha)}{\alpha}$,

- $\text{ES}_\alpha^{\text{occ}} = \text{TCE}_\alpha^{\text{ooc}} \dfrac{S\big(\text{VaR}_\alpha^{\text{ooc}}\big)}{\alpha} + \text{VaR}_\alpha^{\text{ooc}} \dfrac{\alpha - S\big(\text{VaR}_\alpha^{\text{ooc}}\big)}{\alpha}$,

also see Acerbi & Tasche (2002), Göb (2011).

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Risk Performance of Common Control Charts

Examples

Risk Performance of Common Control Charts

Crucial point for computation:

availability of **survival function**.

Often the case in practice:

- i.i.d. Shewhart chart, geometric distribution: formula;
- CUSUM and EWMA charts for i.i.d. Gaussian data: R-package “spc” of Knoth (2017);
- any chart that allows to (approximately) apply MC approach of Brook & Evans (1972).

Shewhart Charts for i. i. d. Data:

$L \sim$ shifted geometric with parameter p.

Then $S(x) = P(L > x) = (1-p)^x$ and ARL $E[L] = 1/p$.

$$\mathrm{VaR}_\alpha^{\mathrm{ic}} = \left\lceil \frac{\ln(1-\alpha)}{\ln(1-p)} \right\rceil, \qquad \mathrm{VaR}_\alpha^{\mathrm{ooc}} = \left\lceil \frac{\ln(\alpha)}{\ln(1-p)} \right\rceil,$$

$$\mathrm{TCE}_\alpha^{\mathrm{ic}} = \mathrm{VaR}_\alpha^{\mathrm{ic}} - 1 + \frac{1}{p} - \frac{\mathrm{VaR}_\alpha^{\mathrm{ic}} - 1}{1-(1-p)^{\mathrm{VaR}_\alpha^{\mathrm{ic}}-1}}, \qquad \mathrm{TCE}_\alpha^{\mathrm{ooc}} = \mathrm{VaR}_\alpha^{\mathrm{ooc}} + \frac{1}{p}.$$

$$\mathrm{ES}_\alpha^{\mathrm{ic}} = \frac{1-(1-p)^{\mathrm{VaR}_\alpha^{\mathrm{ic}}}}{p\,\alpha} - \frac{1-\alpha}{\alpha}\,\mathrm{VaR}_\alpha^{\mathrm{ic}},$$

$$\mathrm{ES}_\alpha^{\mathrm{occ}} = \frac{(1-p)^{\mathrm{VaR}_\alpha^{\mathrm{ooc}}}}{p\,\alpha} + \mathrm{VaR}_\alpha^{\mathrm{ooc}}.$$

Shewhart Charts for i. i. d. Data:

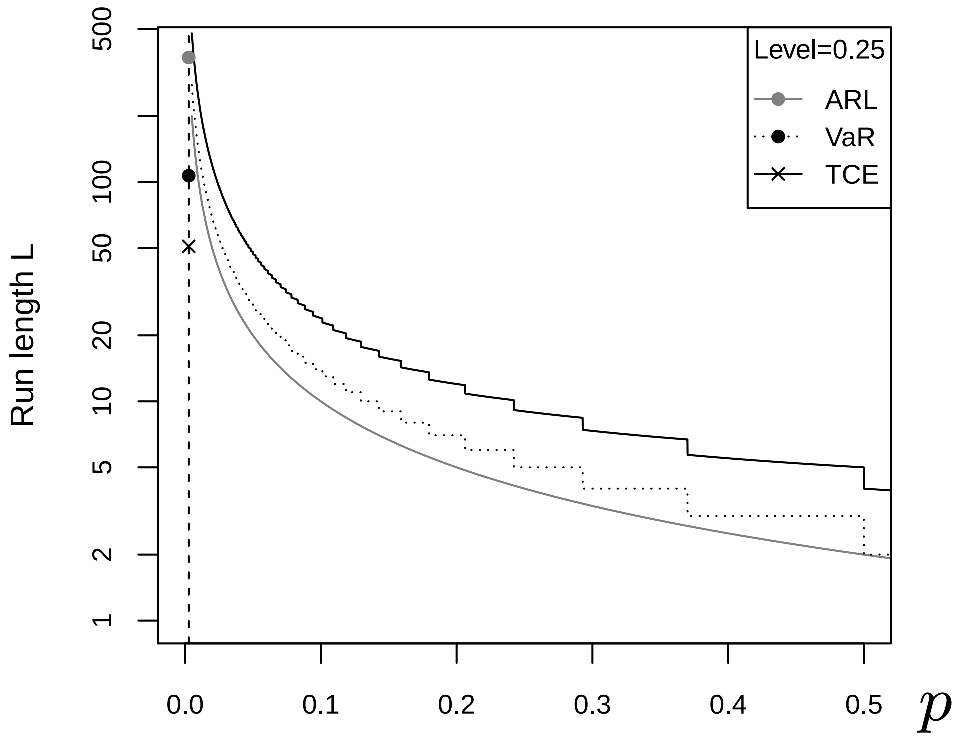
	in-control, $ARL_0 = 370.4$					out-of-control, $ARL_1 = 10.0$				
α	0.50	0.25	0.10	0.05	0.01	0.50	0.25	0.10	0.05	0.01
VaR_α	257	107	39	19	4	7	14	22	29	44
TCE_α	113.9	51.0	19.2	9.4	2.0	17.0	24.0	32.0	39.0	54.0
ES_α	114.0	51.2	19.6	9.9	2.4	16.6	23.2	31.8	38.4	53.7

Some conclusions:

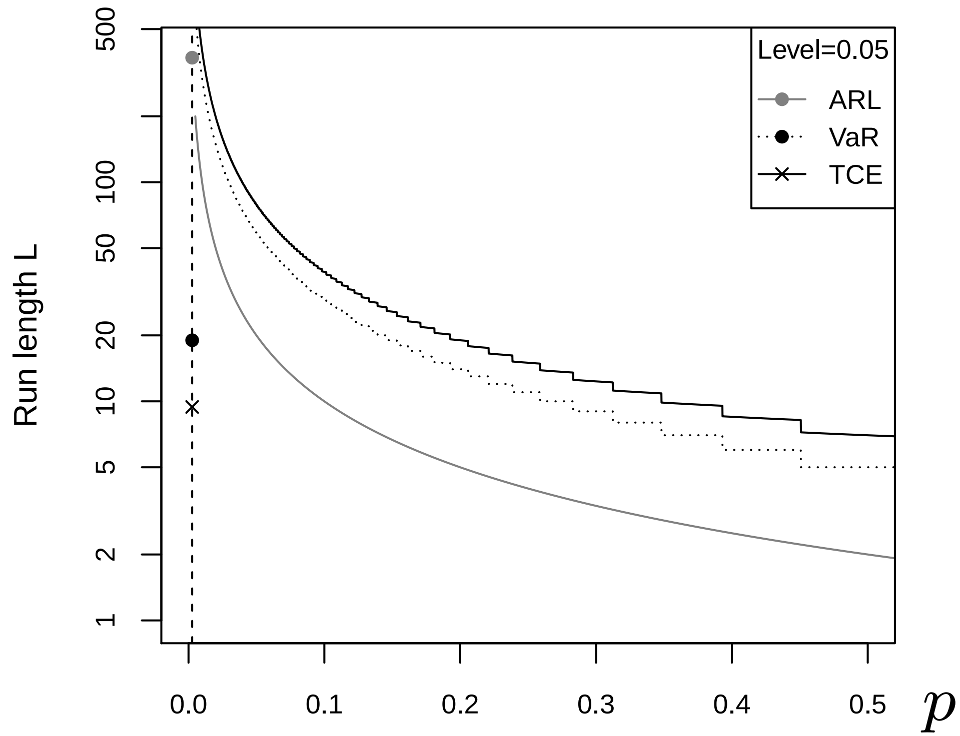
- in-control mean of L in worst 50 % less than 1/3 of ARL_0,
- out-of-control mean of L in worst 10 % larger than 3 times ARL_1,
- values of TCE and ES very close to each other.

Shewhart Charts for i. i. d. Data:

ARL, VaR_α and TCE_α for $\alpha = 0.25$:



ARL, VaR_α and TCE_α for $\alpha = 0.05$:



CUSUM and EWMA Charts for i. i. d. Gaussian Data:

In-control VaR_α and TCE_α:

	CUSUM				Shew-	EWMA (fixed limits)				
k, λ	0.1	0.5	1.0	2.0	hart	0.75	0.25	0.10	0.05	0.01
h, L	10.73	4.10	2.18	0.79		2.84	2.80	2.62	2.43	1.77
$\mathrm{VaR}_{0.05}$	44	24	21	20	19	20	23	27	32	50
$\mathrm{TCE}_{0.05}$	31.4	13.9	11.0	9.9	9.4	10.1	13.0	16.9	21.4	36.7
$\mathrm{VaR}_{0.25}$	126	110	108	107	107	107	109	113	117	133
$\mathrm{TCE}_{0.25}$	73.1	54.7	52.0	51.0	51.0	51.2	53.7	57.9	62.7	80.1
$\mathrm{VaR}_{0.50}$	265	258	257	257	257	257	258	259	261	268
$\mathrm{TCE}_{0.50}$	132.0	116.7	114.5	113.9	113.9	114.1	116.2	119.2	123.2	137.8

All chart designs for $\mathrm{ARL}_0 = 370.4$,

but their in-control performances are quite different!

Two-sided EWMA Charts for i. i. d. Gaussian Data:

VaR_α and TCE_α with fixed or varying limits; $\text{ARL}_0 = 370.4$:

				in-control, $\mu = 0$				out-of-control, $\mu = 0.5$					
				VaR_α		TCE_α		ARL		VaR_α		TCE_α	
λ	L_{fix}	L_{var}	α	fix	var	fix	var	fix	var	fix	var	fix	var
0.05	2.49	2.52	0.05	32	8	21.2	3.3	26.5	21.4	56	53	70.5	67.9
			0.25	116	94	61.7	38.3			33	29	47.7	44.4
			0.50	261	251	122.7	102.8			23	18	38.1	34.4
0.10	2.70	2.71	0.05	26	14	16.1	6.0	28.2	25.7	68	66	88.0	86.3
			0.25	112	103	57.1	46.5			36	34	56.0	54.3
			0.50	259	255	118.8	110.0			23	20	43.1	40.6
0.25	2.90	2.90	0.05	22	19	12.3	9.1	41.1	40.2	114	114	151.1	151.3
			0.25	109	106	53.5	50.1			55	55	92.1	92.3
			0.50	258	257	115.9	113.3			30	29	67.1	66.3

Much lower in-control VaR and TCE values if variable limits, whereas relatively small difference in out-of-control case!

c Chart for Markov Count Processes:

MC approach applicable (exactly), analogously also for types of CUSUM and EWMA charts (bivariate MC approach), see Weiß & Testik (2009); Weiß (2018).

Here: DGP Poisson INAR(1) $X_t = \rho \circ X_{t-1} + \epsilon_t$.

Design of upper-sided c chart
if erroneously treating counts as i.i.d.,
i.e., run length performance under disregarded autocorrelation!

c Chart for Markov Count Processes:

In-control, $\mu_0 = 2.844946$ corresp. to i. i. d.-$\text{ARL}_0 = 370.4$:

		$\text{VaR}_\alpha^{\text{ic}}$, where $\alpha =$					$\text{TCE}_\alpha^{\text{ic}}$, where $\alpha =$				
ρ	ARL	0.50	0.25	0.10	0.05	0.01	0.50	0.25	0.10	0.05	0.01
0.1	372.9	259	108	40	20	4	114.7	51.4	19.7	9.9	2.0
0.2	377.9	262	109	40	20	4	116.1	51.9	19.7	9.9	2.0
0.3	386.8	268	112	41	20	4	118.7	53.3	20.1	9.9	2.0
0.4	401.5	278	116	43	21	4	123.1	55.2	21.1	10.4	2.0

Out-of-control, $\mu_1 = 5.432468$ corresp. to i. i. d.-$\text{ARL}_1 = 10.0$:

		$\text{VaR}_\alpha^{\text{ooc}}$, where $\alpha =$					$\text{TCE}_\alpha^{\text{ooc}}$, where $\alpha =$				
ρ	ARL	0.50	0.25	0.10	0.05	0.01	0.50	0.25	0.10	0.05	0.01
0.1	10.4	7	14	23	30	46	17.5	24.5	33.5	40.5	56.5
0.2	11.0	8	15	25	32	49	19.1	26.1	36.1	43.1	60.1
0.3	11.8	8	16	27	35	53	20.0	28.0	39.0	47.0	65.0
0.4	12.8	9	18	29	38	58	22.2	31.2	42.2	51.2	71.2

Conclusions

- Very limited impression about chart's actual performance by looking solely at ARL.
- We proposed to use risk metrics for performance evaluation as supplements to ARL metric.
- Risk metrics computed numerically exactly for many common chart types, as illustrated with number of examples.
- **Further application scenario** (see paper for details): risk metrics for evaluating control chart's ARL performance under estimation uncertainty. (. . .)

Phase-I analysis: estimator $\widehat{\boldsymbol{\theta}}_0$ of in-control parameter value $\boldsymbol{\theta}_0$ (Montgomery, 2009; Jones-Farmer et al., 2014).

Chart design for Phase-II monitoring based on $\widehat{\boldsymbol{\theta}}_0$, i.e., alarm triggered if

$$Z_t \notin \left[l\big(\widehat{\boldsymbol{\theta}}_0\big); u\big(\widehat{\boldsymbol{\theta}}_0\big)\right].$$

Therefore, performance metrics become function of $\widehat{\boldsymbol{\theta}}_0$.
E.g., $\mathrm{ARL}\big(\boldsymbol{\theta}\,\big|\,\widehat{\boldsymbol{\theta}}_0\big)$, where probability is zero that $\widehat{\boldsymbol{\theta}}_0$ equals $\boldsymbol{\theta}_0$.
So $\mathrm{ARL}\big(\boldsymbol{\theta}_0\,\big|\,\widehat{\boldsymbol{\theta}}_0\big)$ deviates from ARL_0 with certainty.

Risk Metrics for ARL under Estimation Uncertainty

One way to quantify this deviation:
mean of conditional $\text{ARL}\big(\boldsymbol{\theta}\,\big|\,\widehat{\boldsymbol{\theta}}_0\big)$ distribution over $\widehat{\boldsymbol{\theta}}_0$,
i.e., **marginal ARL**, see Jensen et al. (2006); Testik (2007).

We propose to also compute **risk metrics for ARL**, e.g.,

$$\text{AVaR}_\alpha^{\text{ic}} \;:=\; \inf\Big\{x \in (1;\infty)\;\Big|\; P\Big(\text{ARL}\big(\boldsymbol{\theta}_0\,\big|\,\widehat{\boldsymbol{\theta}}_0\big) \le x\Big) \ge \alpha\Big\},$$

$$\text{ATCE}_\alpha^{\text{ic}} \;:=\; E\Big[\text{ARL}\big(\boldsymbol{\theta}_0\,\big|\,\widehat{\boldsymbol{\theta}}_0\big)\;\Big|\;\text{ARL}\big(\boldsymbol{\theta}_0\,\big|\,\widehat{\boldsymbol{\theta}}_0\big) < \text{AVaR}_\alpha^{\text{ic}}\Big]$$

$$= \;\text{AVaR}_\alpha^{\text{ic}} \;-\; \frac{\text{AVaR}_\alpha^{\text{ic}} - 1 - \int_1^{\text{AVaR}_\alpha^{\text{ic}}} S_{\text{ARL}}(x)\,\text{d}x}{1 - S_{\text{ARL}}(\text{AVaR}_\alpha^{\text{ic}})}.$$

Details and examples in paper!

Thank You

for Your Interest!

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

`weissc@hsu-hh.de`

Literature

Acerbi & Tasche (2002) On the coherence . . . *J Banking Fin* **26**, 1487–1503.
Brook & Evans (1972) An approach to . . . *Biometrika* **59**, 539–549.
Göb (2011) Estimating Value at Risk . . . *QREI* **27**, 659–672.
Jensen et al. (2006) Effects of parameter estimation . . . *JQT* **32**, 395–409.
Jones-Farmer et al. (2014) An overview of Phase I . . . *JQT* **46**, 265–280.
Kenett & Pollak (2012) On assessing the perform. . . *QREI* **28**, 500–507.
Klüppelberg et al. (2014) *Risk – A Multidisciplinary Introduction*. Springer.
Knoth (2006) The art of evaluating . . . *Frontiers SQC 8*, 74–99.
Knoth (2017) The spc package (statistical process control), Version 0.5.4.
McNeil et al. (2015) *Quantitative Risk Management*. 2nd ed., Princeton UP.
Montgomery (2009) *Introd. to Statistical Quality Control*. 6th ed., Wiley.
Testik (2007) Conditional and marginal perf. . . *IJPR* **45**, 5621–5638.
Weiß & Testik (2009) CUSUM monitoring of first-. . . *JQT* **41**, 389–400.
Weiß (2018) *An Introduction to Discrete-Valued Time Series*. Wiley.