Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 40: 647–668, 2013
doi: 10.1002/sjos.12006

Guaranteed Conditional Performance of Control Charts via Bootstrap Methods

AXEL GANDY

Department of Mathematics, Imperial College London

JAN TERJE KVALØY

Department of Mathematics and Natural Sciences, University of Stavanger

ABSTRACT. To use control charts in practice, the in-control state usually has to be estimated. This estimation has a detrimental effect on the performance of control charts, which is often measured by the false alarm probability or the average run length. We suggest an adjustment of the monitoring schemes to overcome these problems. It guarantees, with a certain probability, a conditional performance given the estimated in-control state. The suggested method is based on bootstrapping the data used to estimate the in-control state. The method applies to different types of control charts, and also works with charts based on regression models. If a non-parametric bootstrap is used, the method is robust to model errors. We show large sample properties of the adjustment. The usefulness of our approach is demonstrated through simulation studies.

Key words: bootstrap, confidence interval, control chart, CUSUM, estimation error, guaranteed performance, monitoring

1. Introduction

Control charts such as the Shewhart chart (Shewhart, 1931) and the cumulative sum (CUSUM) chart (Page, 1954) have been valuable tools in many areas, including reliability (O'Connor, 2002; Xie *et al.*, 2002), medicine (Carey, 2003; Lawson & Ken, 2005; Woodall, 2006) and finance (Frisén, 2008). See Stoumbos *et al.* (2000) and the special issues of '*Sequential Analysis*' (2007, Volume 26, Issues 2,3) for an overview. Often, heterogeneity between observations is accounted for by using risk-adjusted charts based on fitted regression models (Grigg & Farewell, 2004; Horváth *et al.*, 2004; Gandy *et al.*, 2010).

A common convention in monitoring based on control charts is to assume the probability distribution of in-control data to be known. In practice, this usually means that the distribution is estimated from a sample of in-control data and the estimation error is ignored (e.g. Steiner *et al.*, 2000; Grigg & Farewell, 2004; Biswas & Kalbfleisch, 2008; Bottle & Aylin, 2008; Fouladirad *et al.*, 2008; Gandy *et al.*, 2010). However, the estimation error has a profound effect on the performance of control charts (see e.g. Stoumbos *et al.*, 2000; Albers & Kallenberg, 2004b; Jones *et al.*, 2004; Jensen *et al.*, 2006; Champ & Jones-Farmer, 2007).

To illustrate the effect of estimation, we consider a CUSUM chart (Page, 1954) with normal observations and estimated in-control mean. We observe a stream of independent random variables $X_1, X_2, \ldots$ with $X_i \sim N(\mu, 1)$ in-control and $X_i \sim N(\mu+\Delta, 1)$ out-of-control, where $\Delta > 0$ is the shift in the mean. The chart switches from the in-control state to the out-of-control state at an unknown time κ. The unknown in-control mean μ is estimated by the average $\hat{\mu}$ of n past in-control observations $X_{-n}, \ldots, X_{-1}$ (this is often called phase 1 of the monitoring; the running of the chart is called phase 2). We consider the CUSUM chart

$$S_t = \max(0, S_{t-1} + X_t - \hat{\mu} - \Delta/2), \quad S_0 = 0$$

with hitting time $\tau = \inf\{t > 0 : S_t \geq c\}$ for some threshold $c > 0$.

The in-control average run length, $\text{ARL}=\text{E}(\tau \mid \hat{\mu}, \kappa=\infty)$, depends on $\hat{\mu}$ and is thus a random quantity. The top part of Fig. 1 shows boxplots of its distributions with threshold $c=2.84$, $\Delta=1$ and various numbers of past observations. If $\hat{\mu}=\mu$, that is, μ was known, this would give an in-control ARL of 100. The estimation error is having a substantial effect on the attained ARL even for large samples such as $n=1000$. For further examples of the impact of estimation error see Jones *et al.* (2004) for CUSUM charts and Albers & Kallenberg (2004b) for Shewhart charts.

So far, no general approach for taking the estimation error into account has been developed, but there are many special constructions for specific situations. For instance, for some charts so called self-starting charts (Hawkins, 1987; Hawkins & Olwell, 1998; Sullivan & Jones, 2002), maximum likelihood surveillance statistics to eliminate parameters (e.g. Frisén & Andersson, 2009), correction factors for thresholds (Jones, 2002; Albers & Kallenberg, 2004b) modified thresholds (Zhang *et al.*, 2011) and threshold functions (Horváth *et al.*, 2004; Aue *et al.*, 2006) have been developed. Various bootstrap schemes for specific situations have also been suggested (see e.g. Kirch, 2008; Capizzi & Masarotto, 2009; Chatterjee & Qiu, 2009; Hušková & Kirch, 2010). Further, some non-parametric charts which account for the estimation error in past data have been proposed, see Chakraborti & Graham (2007) and references therein. Recently some modified charts for monitoring variance in the normal distribution with estimated parameters have been suggested by Maravelakis & Castagliola (2009) and Castagliola & Maravelakis (2011).

When addressing estimation error, the above methods mainly focus on the performance of the charts averaged over both the estimation of the in-control state as well as running the chart once. In the middle part of Fig. 1, the threshold has been chosen such that, averaged over both the estimation of the in-control state and running the chart once, the average run length is 100 (this results in a different threshold for each n). It turns out that only a small change in the threshold is needed and that the distribution of the conditional $\text{ARL}=\text{E}(\tau \mid \hat{\mu})$ is only changed slightly. This bias correction for the ARL actually goes in the wrong direction in the sense of implying more short ARLs. This is due to the ARL being substantially influenced by the right tail of the run length distribution, see also Albers & Kallenberg (2006, section 2).

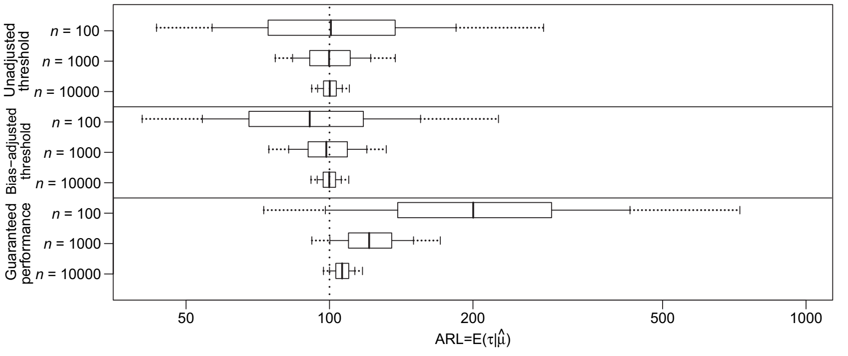

Fig. 1. In-control distribution of $\text{ARL} = E(\tau|\hat{\mu})$ for CUSUMs for standard normally distributed data. The mean $\hat{\mu}$ used in the monitoring is estimated based on n past observations. Log-scale on the horizontal axis. A Markov chain approximation is used to calculate the ARLs, and the distributions of ARLs is estimated based on 1000 replications. The boxplots show the 2.5%, 10%, 25%, 50%, 75%, 90% and 97.5% quantiles of the ARL distribution. The top part of the plot shows the situation when estimation error is ignored. In the middle part the threshold has been chosen to give an unconditional ARL of 100 (averaging out the parameter estimation). In the bottom part the threshold is adjusted to guarantee with 90% probability an in-control ARL of at least 100.

However, after the in-control distribution has been estimated, the chart is usually run for some time without any reestimation of the in-control state even if the chart signals. Moreover, in some situations, several charts are run based on the same estimated parameters. In these situations the ARL conditional on the estimated in-control state is more relevant than the unconditional ARL. The middle and upper part of Fig. 1 shows that the conditional ARL can be much lower than 100, meaning that both the unadjusted threshold and the threshold adjusted for bias in the unconditional ARL lead, with a substantial probability, to charts that have a considerably decreased time until false alarms.

To overcome these problems we will look at the performance of the chart conditional on the estimated in-control distribution, averaging only over different runs of the chart. This will lead to charts that with high probability have an in-control distribution with desired properties conditional on the observed past data, thus reducing the situations in which there are many false alarms due to estimation error.

The bottom part of Fig. 1 shows the distribution of the in-control ARL when the threshold is adjusted to guarantee an in-control ARL of at least 100 with probability 90%. The adjustment uses a bootstrap procedure explained later in the article. The adjustment succeeds to avoid the too low ARLs with the prescribed probability, and we will see later that the cost in a higher out-of-control ARL is modest. Using hitting probabilities instead of ARL as criterion leads to similar results.

Our approach is similar in spirit to the exceedance probability concept developed by Albers and Kallenberg for various types of Shewhart (Albers & Kallenberg, 2004a, 2005; Albers *et al.*, 2005) and negative binomial charts (Albers & Kallenberg, 2009, 2010). They calculate approximate adjusted thresholds such that there is only a small prescribed probability that some performance measure, for instance an ARL, will be a certain amount below or above a specified target.

The main difference between their approach and what we present is that our approach applies far more widely, to many different types of charts and without having to derive specific approximation formulas in each setting. If we apply a non-parametric bootstrap, the proposed procedure will be robust against model misspecification. In addition to that, our approach allows not only to adjust the threshold but also to give a confidence interval for the in-control performance of a chart for a fixed threshold. Lastly, even though not strongly advocated in this study, the bootstrap procedure we propose can also be used to do a bias correction for the unconditional performance of the chart, as in the middle part of Fig. 1.

Next, we describe our approach more formally. We want to use a monitoring scheme and the in-control distribution P of the observations is unknown. We will assume throughout that we have a sample of past in-control observation, and based on these observations we calculate an estimate $\hat{P}$ of the in-control distribution. Let q denote the in-control property of the chart we want to compute, such as the ARL, the false alarm probability or the threshold needed for a certain ARL or false alarm probability. In the above example we were interested in finding a threshold such that the in-control ARL is 100. Generally, q may depend on both the true in-control distribution P and on estimated parameters of this distribution which for many charts are needed to run the chart. We denote these parameters by $\hat{\xi}=\xi(\hat{P})$. In the above CUSUM chart example $\hat{\xi}=\hat{\mu}$. We are interested in $q(P;\hat{\xi})$, that is the in-control performance of the chart conditional on the estimated parameter. In the above CUSUM example, $q(P;\hat{\xi})$ is the threshold needed to give an ARL of 100 if the observations are from the true in-control distribution P and the estimated parameter $\hat{\mu}$ is used. As P is not observed $q(P;\hat{\xi})$ is not observable. As mentioned above, many studies pretend that the estimated in-control distribution $\hat{P}$ equals the true in-control distribution P and thus use $q(\hat{P};\hat{\xi})$. Our suggestion is to use bootstrapping of past data to construct an approximate one-sided confidence interval for $q(P; \hat{\xi})$. From this we get a guaranteed conditional performance of the control scheme.

In section 2 we present the general idea in the setting with homogeneous observations, and discuss this for Shewhart and CUSUM charts. The main theoretical results are presented in section 3, with most of the proofs given in the Appendix. Section 4 contains simulations illustrating the performance of charts for homogeneous observations. In section 5 extensions to charts based on regression models are presented. Some concluding comments are given in section 6. The suggested methods are implemented in the R-package spcadjust, which is available on the Comprehensive R Archive Network (CRAN).

2. Monitoring homogeneous observations

2.1. General idea

Suppose that in control we have independent observations $X_1, X_2, \ldots$ following an unknown distribution P. We want to use some monitoring scheme/control chart that detects when X_i is no longer coming from P. The particular examples we consider are Shewhart and CUSUM charts, but the methodology we suggest applies more widely.

To run the charts, one often needs certain parameters ξ. For example, in the CUSUM control chart of the introduction, we needed $\xi = \mu$, the assumed in-control mean. These parameters will usually be estimated.

Let τ denote the time at which the chart signals a change. As τ may depend on ξ, we sometimes write $\tau(\xi)$. The charts we consider use a threshold c, which determines how quickly the chart signals (larger c lead to a later signal).

The performance of such a control chart with the in-control distribution P and the parameters ξ can, for example, be expressed as one of the following.

- $\mathrm{ARL}(P; \xi) = E(\tau(\xi))$, where E is the expectation with respect to P.
- $\mathrm{hit}(P; \xi) = P(\tau(\xi) \leq T)$ for some finite $T > 0$, where $X_1, X_2, \ldots \sim P$. This is the false alarm probability in T time units.
- $c_{\mathrm{ARL}}(P; \xi) = \inf\{c > 0 : \mathrm{ARL}(P; \xi) \geq \gamma\}$ for some $\gamma > 0$. Assuming appropriate continuity, this is the threshold needed to give an in-control ARL of γ.
- $c_{\mathrm{hit}}(P; \xi) = \inf\{c > 0 : \mathrm{hit}(P; \xi) \leq \beta\}$ for some $0 < \beta < 1$. This is the threshold needed to give a false alarm probability of β.

The latter two quantities are very important in practice, as they are needed to decide which threshold to use to run a chart. In the notation we have suppressed the dependence of the quantities on c, T, γ, β and Δ.

In the following, q will denote one of ARL, hit, c_{ARL} or c_{hit}, or simple transformations such as $\log(\mathrm{ARL})$, $\mathrm{logit}(\mathrm{hit})$, $\log(c_{\mathrm{ARL}})$ and $\log(c_{\mathrm{hit}})$, where $\mathrm{logit}(x) = \log(x/(1-x))$.

The true in-control distribution P and the parameters $\xi = \xi(P)$ needed to run the chart are usually estimated. We assume that we have past in-control observations $X_{-n}, \ldots, X_{-1}$ (independent of $X_1, X_2, \ldots$), which we use to estimate the in-control distribution P parametrically or non-parametrically. We denote this estimate by $\hat{P}$. The estimate of ξ will be denoted by $\hat{\xi} = \xi(\hat{P})$. For example, in the CUSUM control chart of the introduction, $\hat{\xi} = \hat{\mu}$ is the estimated in-control mean.

The observed performance of the chart will depend on the true in-control distribution P as well as on the estimated parameters $\hat{\xi}$ that are used to run the chart. Thus we are interested in $q(P; \hat{\xi})$, the performance of the control chart *conditional* on $\hat{\xi}$. This is an unknown quantity as P is not known. Based on the estimator $q(\hat{P}; \hat{\xi})$, we construct a one-sided confidence interval for this quantity to guarantee, with high probability, a certain performance for the chart. We choose to call the interval a confidence interval, even though the quantity $q(P;\hat{\xi})$ is random.

We suggest the following for guaranteeing an upper bound on q (which is relevant for $q=\text{hit}$, $q=c_{\text{ARL}}$ or $q=c_{\text{hit}}$). For $\alpha\in(0,1)$, let p_α be a constant such that

$$\mathrm{P}(q(\hat{P};\hat{\xi})-q(P;\hat{\xi})>p_\alpha)=1-\alpha,$$

assuming that such a p_α exists. Hence,

$$\mathrm{P}(q(P;\hat{\xi})<q(\hat{P};\hat{\xi})-p_\alpha)=1-\alpha.$$

Thus $(-\infty, q(\hat{P};\hat{\xi})-p_\alpha)$ could be considered an exact lower one-sided confidence interval of $q(P;\hat{\xi})$. Now p_α is unknown, and we suggest to obtain an approximation by bootstrapping. In the following, $\hat{P}^*$ denotes a parametric or non-parametric bootstrap replicate of the estimated in-control distribution $\hat{P}$. We can approximate p_α by p^*_α such that

$$\mathrm{P}(q(\hat{P}^*;\hat{\xi}^*)-q(\hat{P};\hat{\xi}^*)>p^*_\alpha\,|\,\hat{P})=1-\alpha.$$

Thus

$$(-\infty, q(\hat{P};\hat{\xi})-p^*_\alpha) \tag{1}$$

is a one-sided (approximate) confidence interval for $q(P;\hat{\xi})$. In this study, we will use the following generic algorithm to implement the bootstrap.

Algorithm 1 (Bootstrap).

1. From the past data $X_{-n},\dots,X_{-1}$, estimate $\hat{P}$ and $\hat{\xi}$.
2. Generate bootstrap samples $X^*_{-n},\dots,X^*_{-1}$from $\hat{P}$. Compute the corresponding estimate $\hat{P}^*$ and $\hat{\xi}^*$. Repeat B times to get $\hat{P}^*_1,\dots,\hat{P}^*_B$ and $\hat{\xi}^*_1,\dots,\hat{\xi}^*_B$.
3. Let p^*_α be the $1-\alpha$ empirical quantile of $q(\hat{P}^*_b;\hat{\xi}^*_b)-q(\hat{P};\hat{\xi}^*_b)$, $b=1,\dots,B$.

For guaranteeing a lower bound on q, which is, for example, relevant for $q=\text{ARL}$, a similar upper one-sided confidence interval can be constructed.

In a practical situation, the focus would be on deciding which threshold to use to obtain desired in-control properties. We suggest to use either $q=c_{\text{ARL}}$ or $q=c_{\text{hit}}$, or log transforms of these, and then run the chart with the adjusted threshold

$$q(\hat{P};\hat{\xi})-p^*_\alpha. \tag{2}$$

This will guarantee that in (approximately) $1-\alpha$ of the applications of this method, the control chart actually has the desired in-control properties.

2.2. *Specific charts*

Shewhart charts. The one-sided Shewhart chart (Shewhart, 1931) signals at

$$\tau=\inf\{t\in\{1,2,\dots\}:f(X_t,\xi)>c\} \tag{3}$$

for some threshold c, where f is some function, X_t is the observation at time t and ξ are some parameters. X_t can be a single measurement or for example, the average, range or standard deviation of several measurements, or some other statistic like a proportion. It is common to use a Shewhart chart with a threshold of the mean plus three times the standard deviation, resulting in $c=3$ and $f(x,\xi)=(x-\xi_1)/\xi_2$, where ξ_1 is the mean and ξ_2 is the standard deviation. For two-sided charts one can use $f(x,\xi)=|x-\xi_1|/\xi_2$.

Conditionally on fixed parameters ξ, the stopping time τ follows a geometric distribution with parameter $p=p(c;P,\xi)=\mathrm{P}(f(X_t,\xi)>c)$. Then the performance measures mentioned in the previous section simplify to

$$\mathrm{ARL}(P;\xi)=\frac{1}{p(c;P,\xi)},\quad \mathrm{hit}(P;\xi)=1-(1-p(c;P,\xi))^{\mathrm{T}},$$
$$c_{\mathrm{ARL}}(P;\xi)=p^{-1}\left(\frac{1}{\gamma};P,\xi\right)\quad\text{and}\quad c_{\mathrm{hit}}(P;\xi)=p^{-1}\left(1-(1-\beta)^{\frac{1}{\mathrm{T}}};P,\xi\right),$$

where $p^{-1}(\cdot;P,\xi)$ is the inverse of $p(\cdot;P,\xi)$.

Suppose that the in-control distribution is from a parametric family $P_\theta,\theta\in\Theta$ and that we have some way of computing an estimate $\hat\theta$ of θ based on the past observations $X_{-n},\dots,X_{-1}$. Then Algorithm 1 can be used with $\hat P=P_{\hat\theta}$.

Shewhart charts depend heavily on the tail behaviour of the distribution of the observations. This is particularly problematic when the sample size is small and we use non-parametric methods or a simple non-parametric bootstrap. We thus primarily suggest to use a parametric bootstrap for Shewhart charts.

Remark 1. In certain cases the parametric bootstrap will actually be exact when $B\to\infty$. This happens when the distribution of $q(P_{\hat\theta};\hat\xi)-q(P_\theta;\hat\xi)$ under P_θ does not depend on θ. In particular, this implies that $q(P_{\hat\theta^*};\hat\xi^*)-q(P_{\hat\theta};\hat\xi^*)$ has the same distribution and $p_\alpha^*\to p_\alpha$ as $B\to\infty$.

As an example, consider the case when $f(x,\xi)=(x-\xi_1)/\xi_2$ and X_t follows an $N(\xi_1,\xi_2^2)$ distribution and q is any of the performance measures described above. We use $\theta=\xi$ and as estimators we use the sample mean $\hat\xi_1$ and standard deviation $\hat\xi_2$. Then

$$p(c;P_\xi,\hat\xi)=\mathrm{P}_\xi\left(\frac{X_t-\hat\xi_1}{\hat\xi_2}>c\right)=1-\Phi\left(\frac{c\hat\xi_2+\hat\xi_1-\xi_1}{\xi_2}\right),$$

where Φ is the cdf of the standard normal distribution, and under P_ξ,

$$\frac{c\hat\xi_2+\hat\xi_1-\xi_1}{\xi_2}=c\frac{\hat\xi_2}{\xi_2}+\frac{\hat\xi_1-\xi_1}{\xi_2}\sim\frac{c}{\sqrt{n-1}}\sqrt{W}+\frac{1}{\sqrt{n}}Z,$$

where $W\sim\chi^2_{n-1}$ and $Z\sim N(0,1)$ are independent. Thus the distribution of $p(c;P_\xi,\hat\xi)$, and hence $q(P_\xi;\hat\xi)$, is completely known. As $p(c;P_{\hat\xi},\hat\xi)=\mathrm{P}_{\hat\xi}((X_t-\hat\xi_1)/\hat\xi_2>c)=1-\Phi(c)$, and thus $q(P_{\hat\xi};\hat\xi)$, is not random, the distribution of $q(P_{\hat\xi};\hat\xi)-q(P_\xi;\hat\xi)$ does not depend on any unknown parameters. Thus the parametric bootstrap is exact in this example.

CUSUM charts. We now consider the one-sided CUSUM chart (Page, 1954). The classical CUSUM chart was designed to detect a shift of size $\Delta>0$ in the mean of normally distributed observations. Let μ and σ denote, respectively, the in-control mean and standard deviation. A CUSUM chart can be defined by

$$S_t=\max(0,S_{t-1}+(X_t-\mu-\Delta/2)/\sigma),\quad S_0=0 \tag{4}$$

with hitting time $\tau=\inf\{t>0:S_t\ge c\}$ for some threshold $c>0$.

Alternatively, we could drop the scaling and not divide by the standard deviation σ in (4). See chapter 1.4 in Hawkins & Olwell (1998) for a discussion around this.

More generally, to accommodate observations with general in-control distribution with density f_0 and general out-of-control distribution with density f_1, it is optimal in a certain sense (Moustakides, 1986) to modify the CUSUM chart by replacing $(X_t - \mu - \Delta/2)/\sigma$ by the log likelihood ratio $\log(f_1(X_t,\theta)/f_0(X_t,\theta))$ such that the CUSUM chart is

$$S_t = \max(0, S_{t-1} + \log(f_1(X_t,\theta)/f_0(X_t,\theta))), \quad S_0 = 0. \tag{5}$$

Let ξ denote either (μ,σ) in (4) or θ in (5). Usually, ξ needs to be estimated, and we can then use Algorithm 1 to compute a confidence interval (1) for the performance measure $q(P;\hat{\xi})$. For (5) it is most natural to use a parametric bootstrap with $\hat{P} = P_{\hat{\theta}}$, while for (4) we can use either a parametric or a non-parametric bootstrap.

Remark 2. Similar as for Shewhart charts, this parametric bootstrap is exact when the distribution of $q(P_{\hat{\theta}};\hat{\xi}) - q(P_\theta;\hat{\xi})$ does not have any unknown parameters. This is, for instance, the case if we use (5) for an exponential distribution with the out-of-control distribution specified as an exponential distribution with mean $\Delta\lambda$, where λ is the in-control hazard rate. Another example of this is when we have normally distributed data and use a CUSUM with the increments $(X_t - \hat{\mu})/\hat{\sigma} - \Delta/2$.

3. General theory

In this section, we show that asymptotically, as the number of past observations n increases, our procedure works. An established way of showing asymptotic properties of bootstrap procedures is via a functional delta method (van der Vaart & Wellner, 1996; Kosorok, 2008). Whilst we will follow a similar route, our problem does not fit directly into the standard framework, because the quantity of interest, $q(P,\hat{\xi})$, contains the random variable $\hat{\xi}$. We present the setup and the main result in section 3.1, followed by examples in section 3.2.

3.1. Main theorem

Let D_q be the set in which P and its estimator $\hat{P}$ lie, that is, a set describing the potential probability distribution of our observations. This could be a subset of $\mathbb{R}^d$ for parametric distributions, the set of cumulative distribution functions for non-parametric situations, or the set of joint distributions of covariates and observations. We assume that D_q is a subset of a complete normed vector space D. Let Ξ be a non-empty topological space containing the potential parameters ξ used for running the chart. In our examples, we will let $\Xi \subset \mathrm{R}^d$ be an open set.

We assume that $\hat{P}^* = \hat{P}^*(\hat{P}, W_n)$ is a bootstrapped version of $\hat{P}$ based both on the observed data $\hat{P}$ and on an independent random vector W_n. For example, when resampling with replacement then W_n is a weight vector of length n, multinomially distributed, that determines how often a given observation is resampled. In a parametric bootstrap, W_n is the vector of random variables needed to generate observations from the estimated parametric distribution.

In the main theorem we will need that the mapping $q: D_q \times \Xi \to \mathbb{R}$, which returns the property of interest, satisfies the following extension of Hadamard differentiability. For the usual definition of Hadamard differentiability see, for example, van der Vaart (1998, section 20.2). The extension essentially consists in requiring Hadamard differentiability in the first component when the second component is converging.

Definition 1. *Let D, E be metric spaces, let $D_f \subset D$ and let Ξ be a non-empty topological space. The family of functions $\{f(\cdot;\xi): D_f \to E : \xi \in \Xi\}$ is called Hadamard differentiable at $\theta \in D_f$ around $\xi \in \Xi$ tangentially to $D_0 \subset D$ if there exists a continuous linear map $f'(\theta;\xi): D_0 \to E$ such that*

$$\frac{f(\theta+t_nh_n;\xi_n)-f(\theta;\xi_n)}{t_n}\to f'(\theta;\xi)(h) \quad (n\to\infty)$$

for all sequences $(\xi_n)\subset\Xi$, $(t_n)\subset\mathbb{R}$, $(h_n)\subset D$ *that satisfy* $\theta+t_nh_n\in D_f\forall n$ *and* $\xi_n\to\xi$, $t_n\to 0$, $h_n\to h\in D_0$ *as* $n\to\infty$.

In the following, we understand convergence in distribution, denoted by $\rightsquigarrow$, as defined in van der Vaart & Wellner (1996, def 1.3.3) or in Kosorok (2008, p. 108).

Theorem 1. *Let* $q: D_q\times\Xi\to\mathbb{R}$ *be a mapping, let* $P\in D_q$ *and let* $\xi: D_q\to\Xi$ *be a continuous function. Suppose that the following conditions are satisfied.*

(a) *q is Hadamard differentiable at P around ξ tangentially to D_0 for some $D_0\subset D$.*
(b) *$\hat{P}$ is a sequence of random elements in D_q such that $\sqrt{n}(\hat{P}-P)\rightsquigarrow Z$ as $n\to\infty$ where Z is some tight random element in D_0.*
(c) *$\sqrt{n}(\hat{P}^*-\hat{P})\underset{W}{\overset{P}{\rightsquigarrow}} Z$ as $n\to\infty$ where $\underset{W}{\overset{P}{\rightsquigarrow}}$ denotes weak convergence conditionally on $\hat{P}$ in probability as defined in Kosorok (2008, p. 19).*
(d) *The cumulative distribution function of $q'(P;\xi)Z$ is continuous.*
(e) *Outer-almost surely, the map $W_n\mapsto h(\hat{P}^*(\hat{P}, W_n))$ is measurable for each n and for every continuous bounded function $h: D_q\to\mathbb{R}$.*
(f) *$q(\hat{P};\hat{\xi})-q(P;\hat{\xi})$ and p_α^* are random variables, that is, measurable, where $\hat{\xi}=\xi(\hat{P})$ and $p_\alpha^*=\inf\{t\in\mathbb{R}:\hat{\mathrm{P}}(q(\hat{P}^*;\hat{\xi}^*)-q(\hat{P};\hat{\xi}^*)\le t)\ge\alpha\}$.*

Then

$$\mathrm{P}(q(P;\hat{\xi})\in(-\infty, q(\hat{P};\hat{\xi})-p_\alpha^*))\to 1-\alpha \quad (n\to\infty).$$

A similar result holds for upper confidence intervals.

The proof is in appendix A. The theorem essentially is an extension of the delta-method. Condition (a) ensures the necessary differentiability. Conditions (b) and (c) are standard assumptions for the functional delta method; (b) for the ordinary delta method and (c) for the bootstrap version of it. Condition (d) ensures that, after using an extension of the delta-method, the resulting confidence interval will have the correct asymptotic coverage probability. Condition (e) is a technical measurability condition, which will be satisfied in our examples. Condition (f) is a measurability condition, which should usually be satisfied.

3.2. Examples

The following sections give examples in which theorem 1 applies. We consider hitting probabilities ($q=\mathrm{hit}$) and thresholds to obtain certain hitting probabilities ($q=c_{\mathrm{hit}}$).

These examples are meant to be illustrative rather than exhaustive. For example, other parametric setups could be considered along similar lines to the section 'CUSUM charts with normally distributed observations'. Furthermore, other performance measures such as $\log(c_{\mathrm{hit}})$ or logit(hit) would essentially require application of chain rules to show differentiability.

Simple non-parametric setup for CUSUM charts. We show how the above theorem applies to the CUSUM chart described in (4) when using a non-parametric bootstrap version of algorithm 1.

Let $D=l_\infty(\mathbb{R})$ be the set of bounded functions $\mathbb{R}\to\mathbb{R}$ equipped with the sup-norm $\|x\|=\sup_{t\in\mathbb{R}}|x_t|$. Let $D_q\subset D$ be the set of cumulative distribution functions on $\mathbb{R}$ with finite second moment. The parameters needed to run the chart are the mean and the standard deviation of the in-control observations, thus we may choose $\Xi = \mathbb{R} \times (0, \infty)$ and $\xi : D_q \to \Xi, P \mapsto (\int xP(\mathrm{d}x), \int x^2 P(\mathrm{d}x) - (\int xP(\mathrm{d}x))^2)$.

As quantities q of interest we are considering hitting probabilities ($q = \text{hit}$) and thresholds ($q = c_{\text{hit}}$) needed to achieve a certain hitting probability. The probability $\text{hit} : D_q \times \Xi \to \mathbb{R}$ of hitting a threshold $c > 0$ up to step $T > 0$ can be written as $\text{hit}(P; \xi) = \mathrm{P}(m(Y) \geq c)$, where $m(Y) = \max_{i=1,\dots,T} R_i(Y)$ is the maximum value of the chart up to time T, $R_i(Y) = \sum_{j=1}^{i} Y_j - \min_{0 \leq k \leq i} \sum_{j=1}^{k} Y_j$ is the value of the CUSUM chart at time i, $Y = (Y_1, \dots, Y_T)$, $Y_t = (X_t - \xi_1 - \Delta/2)/\xi_2$ and $X_1, \dots, X_T \sim P$ are the independent observations. The threshold needed to achieve a certain hitting probability $\beta \in (0, 1)$ is $c_{\text{hit}} : D_q \times \Xi \to \mathbb{R}$, $c_{\text{hit}}(P; \xi) = \inf\{c > 0 : \text{hit}(P; \xi) \leq \beta\}$.

The setup for the non-parametric bootstrap is as follows. W_n is an n-variate multinomially distributed random vector with probabilities $1/n$ and n trials. The resampled distribution is $\hat{P}^* = \frac{1}{n} \sum_{j=1}^{n} W_{nj} \delta_{X_{-j}}$, where δ_x is the Dirac measure at x.

The following lemma shows condition (a) of theorem 1, the Hadamard differentiability of hit and c_{hit}.

Lemma 1. *For every $P \in D_q$, and every $\xi \in \mathbb{R} \times (0, \infty)$, the function* hit *is Hadamard differentiable at P around ξ tangentially to $D_0 = \{H : \mathbb{R} \to \mathbb{R} : H$ continuous, $\lim_{t \to \infty} H(t) = \lim_{t \to -\infty} H(t) = 0\}$. If, in addition, P has a continuous bounded positive derivative f with $f(x) \to 0$ as $x \to \pm\infty$, then c_{hit} is also Hadamard differentiable at P around ξ tangentially to D_0.*

The proof is in appendix B.4, with preparatory results in appendix B.1–B.3.

Conditions (b) and (c) of theorem 1 follow directly from empirical process theory (see e.g. Kosorok, 2008, p. 17 theorems 2.6 and 2.7). Condition (e) is satisfied as well, see bottom of p. 189 and after theorem 10.4 (p. 184) of Kosorok (2008). Verifying condition (d) in full is outside the scope of the present study. A starting point could be the fact that by the Donsker theorem, $Z \sim G \circ P$, where G is a Brownian bridge.

CUSUM charts with normally distributed observations. In this section, we consider a setup similar to the one considered in the previous subsection with the difference that we now use parametric assumptions. More specifically, the observations X_i follow a normal distribution with unknown mean μ and variance σ^2. We will use this both for computing the properties of the chart as well as in the bootstrap, which will be a parametric bootstrap version of algorithm 1.

The distribution of the observations can be identified with its parameters which we estimate by $\hat{P} = (\hat{\mu}, \hat{\sigma}^2)$, where $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_{-i}$ and $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_{-i} - \hat{\mu})^2$. The set of potential parameters is $D_q = \mathbb{R} \times (0, \infty)$ which is a subset of the Euclidean space $D = \mathbb{R}^2$. The parameters needed to run the chart (4) are just the same, thus $\Xi = D_q$ and $\xi : D_q \to \Xi, (\mu, \sigma) \mapsto (\mu, \sigma)$ is just the identity.

We are interested in hitting probabilities within the first T steps. Using the function hit defined in the previous subsection, we can write the hitting probability in this parametric setup as $\text{hit}^N : D_q \times \Xi \to \mathbb{R}$, $(\mu, \sigma; \xi) \mapsto \text{hit}(\Phi_{\mu,\sigma^2}; \xi)$, where the superscript N stands for normal distribution and Φ_{μ,σ^2} is the cdf of the $N(\mu, \sigma^2)$ distribution. Using c_{hit} from the previous subsection, the threshold needed to achieve a given hitting probability is $c_{\text{hit}}^N : D_q \times \Xi \to \mathbb{R}$, $(\mu, \sigma; \xi) \mapsto c_{\text{hit}}(\Phi_{\mu,\sigma^2}; \xi)$.

The resampling is a parametric resampling. To put this in the framework of the main theorem, we let $W_n = (W_{n1}, \dots, W_{nn})$, where $W_{n1}, \dots, W_{nn} \sim N(0, 1)$ are independent. The resampled parameters are then $\hat{\mu}_n^* = \frac{1}{n} \sum_{i=1}^{n} X_{ni}^*$ and $\hat{\sigma}_n^{*2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ni}^* - \hat{\mu}_n^*)^2$ where $X_{ni}^* = \hat{P}_2 W_{ni} + \hat{P}_1$.

The following lemma shows that condition (a) of theorem 1 is satisfied.

Lemma 2. *For every $\theta \in \mathbb{R} \times (0, \infty)$ and every $\xi \in \mathbb{R} \times (0, \infty)$, the functions* hit^N *and* c_{hit}^N *are Hadamard differentiable at θ around ξ.*

The proof can be found in appendix B.4, using again the preparatory results of appendix B.1–B.3.

Concerning the other conditions of theorem 1: condition (b) can be shown using standard asymptotic theory, for example, maximum likelihood theory, which will yield that Z is normally distributed. Condition (c) is essentially the requirement that the parametric bootstrap of normally distributed data is working. As Z is a normally distributed vector, condition (d) holds unless q' equals 0. Condition (e) is satisfied, as the mapping $W_n \mapsto \hat{P}^*(\hat{P}, W_n) = (\hat{\mu}_n^*, \hat{\sigma}_n^{*2})$ is continuous and hence measurable.

Setup for Shewhart charts. For Shewhart charts, the same setup as in the previous two sections can be used, the only difference is the choice of q. Conditions (b), (c) and (e) are as in the previous two sections. We conjecture that it is possible to show the Hadamard differentiability more directly, as the properties are available in closed form, see 'Shewhart charts' in section 2.2.

4. Simulations for homogeneous observations

We now illustrate our approach by some simulations. The simulations were done in R (R Development Core Team, 2012).

We use two past sample sizes, $n = 50$ and $n = 500$. The in-control distribution of X_t is $N(0, 1)$ and we use 1000 replications and $B = 1000$ bootstrap replications.

We first consider a simple Shewhart chart situation to verify that our bootstrap approach is correct in a situation were an exact solution exist. We use the chart (3) with $f(x, \xi) = (x - \xi_1)/\xi_2$ and ξ_1 and ξ_2 estimated by the sample mean and sample standard deviation of $X_{-n}, \ldots, X_{-1}$.

We would like to adjust the threshold such that we can guarantee with 90% probability an in-control ARL of at least 100. As explained in Albers & Kallenberg (2004a) exact thresholds can be calculated in this situation, and they also derive simple approximations for these thresholds. As explained in section 2.2 'Shewhart charts' the parametric bootstrap is also exact in this situation. In Fig. 2 we show the in-control distribution of the conditional ARL with these three adjustments and with no adjustment for estimation error. We see that the bootstrap is exact as claimed, the approximation by Albers & Kallenberg (2004a) is slightly non-conservative for $n = 50$ and the unadjusted threshold may lead to substantially too low ARLs.

In the remainder of this section, we consider CUSUM charts (4) with $\Delta = 1$. As far as we know there exist no other approaches in the literature for giving a guaranteed performance for such charts. We employ both the parametric and the non-parametric bootstrap.

For the performance measures ARL, log(ARL), hit and logit(hit) we use a threshold of $c = 3$. For c_{ARL} we calibrate to an ARL of 100 in control and for c_{hit} we calibrate to a false alarm probability of 5% in 100 steps.

To compute ARLs and hitting probabilities, we use a Markov chain approximation (with 75 grid points), similar to the one suggested in Brook & Evans (1972).

4.1. Coverage probabilities

Table 1 contains coverage probabilities of nominal 90% confidence intervals. These are the

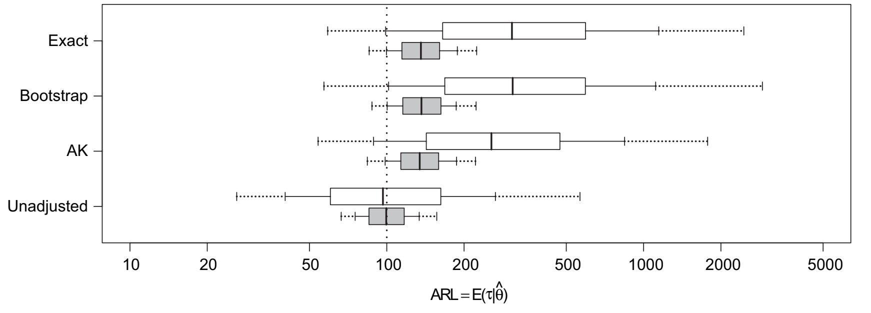

Fig. 2. Distribution of the conditional ARL for Shewhart charts in a normal distribution setup. Thresholds are calibrated to an in-control ARL of 100 with a guarantee of 90% using respectively the exact threshold (Exact) our bootstrap approach (Bootstrap) and the approximation formula by Albers & Kallenberg (2004a) (AK). In the two boxplot at bottom the thresholds are unadjusted. The boxplots show the 2.5%, 10%, 25%, 50%, 75%, 90% and 97.5% quantiles of the ARL distributions. The white boxplots are for $n=50$, the gray for $n=500$. Log-scale on the horizontal axis.

Table 1. *Coverage probabilities of nominal* 90% *confidence intervals for CUSUM charts*

	Parametric		Non-parametric	
q	$n=50$	$n=500$	$n=50$	$n=500$
ARL	1.000	0.929	0.999	0.944
log(ARL)	0.928	0.899	0.902	0.915
hit	0.923	0.896	0.878	0.910
logit(hit)	0.892	0.893	0.870	0.904
c_{ARL}	0.881	0.892	0.846	0.893
$\log(c_{\mathrm{ARL}})$	0.896	0.895	0.868	0.904
c_{hit}	0.878	0.890	0.843	0.891
$\log(c_{\mathrm{hit}})$	0.897	0.893	0.856	0.901

one-sided lower confidence intervals given by (1), except for $q=\mathrm{ARL}$ and log(ARL) where the corresponding upper interval is used.

In the parametric case, for $n=50$, the coverage probabilities are somewhat off for untransformed versions, in particular for $q=\mathrm{ARL}$. Using log or logit transformations seems to improve the coverage probabilities considerably. In the parametric case, for $n=500$, all coverage probabilities seem to be fine, except for $q=\mathrm{ARL}$, which nevertheless shows considerable improvement compared to $n=50$. A similar picture emerges in the non-parametric case, with slightly worse coverage probabilities.

Remark 3. For $q=\log(c_{\mathrm{ARL}})$ and $q=\log(c_{\mathrm{hit}})$ the division by $\hat{\sigma}$ in (4) could be skipped without making a difference to the coverage probabilities. Indeed, the division by $\hat{\sigma}$ just scales the chart (and the resulting threshold) by a multiplicative factor, which is turned into an additive factor by log and which then cancels out in our adjustment.

4.2. *The benefit of an adjusted threshold*

In this section, we consider both the in- and out-of-control performance of CUSUM charts when adjusting the threshold c to give an in-control ARL of 100. Setting the threshold is, in our opinion, the most important practical application of our method.

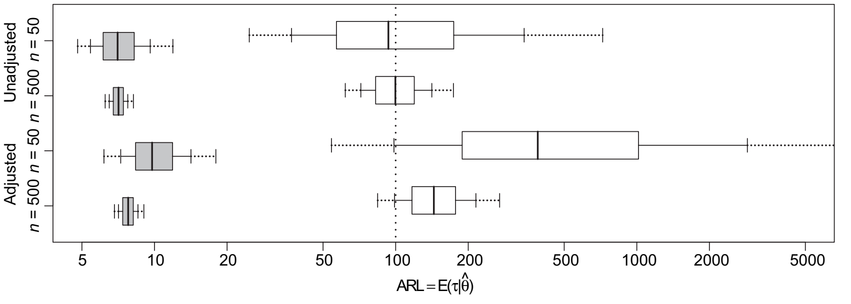

Fig. 3. Distribution of the conditional ARL for CUSUMs in a normal distribution setup. Thresholds are calibrated to an in-control ARL of 100. The adjusted thresholds have a guarantee of 90%. A log transform is used in the calibration. The boxplots show the 2.5%, 10%, 25%, 50%, 75%, 90% and 97.5% quantiles. The white boxplots are in-control, the gray boxplots out-of-control. Log-scale on the horizontal axis.

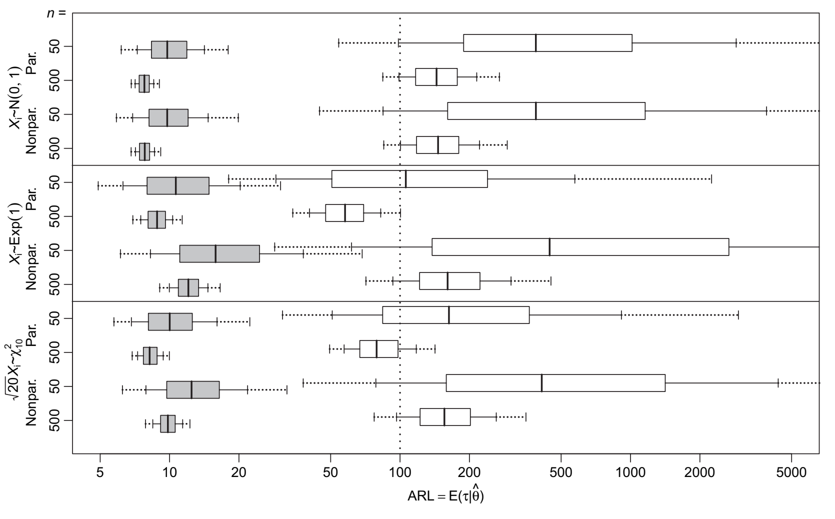

Fig. 4. Effects of misspecification. Thresholds are calibrated to an in-control ARL of 100 and adjusted to the estimation error with a guarantee of 90%. A log transform is used. The white boxplots are in-control, the gray boxplots are out-of-control. The boxplots show the 2.5%, 10%, 25%, 50%, 75%, 90% and 97.5% quantiles. Log-scale on the horizontal axis.

Figure 3 shows average run lengths for both the unadjusted threshold $c(\hat{P};\hat{\mu},\hat{\sigma})$ and the adjusted threshold $\exp(\log(c(\hat{P};\hat{\mu},\hat{\sigma}))-p^*_{0.1})$, where $p^*_{0.1}$ is computed via the parametric bootstrap using $q=\log(c_{\mathrm{ARL}})$. Thus, with 90% probability, the adjusted threshold should lead to an ARL that is above 100. In this and in all following simulations, the out-of-control ARL refers to the situation where the chart is out-of-control from the beginning, that is, from time 0 onwards.

For the unadjusted threshold, the desired in-control ARL is only reached in roughly half the cases. More importantly, for $n=50$, the probability of having an in-control ARL of below 50 is $>20\%$.

With the adjusted threshold we should get an ARL of at least 100 in 90% of the cases. This is achieved. The out-of-control ARL using the adjusted thresholds increases only slightly compared to the unadjusted version.

Removing the scaling by $\hat{\sigma}$ in (4) would not change these results, see remark 3.

4.3. Non-parametric bootstrap – advantages and disadvantages

In this section, we compare the parametric and the non-parametric bootstrap. We consider CUSUM charts calibrated to an in-control average run length of 100 assuming a normal distribution. We use the adjusted threshold $\exp(\log(c_{\mathrm{ARL}}(\hat{P};\hat{\mu},\hat{\sigma}))-p^{*}_{0.1})$.

Figure 4 shows the distribution of ARL for $n=50$ and $n=500$ for both the parametric bootstrap that assumes a normal distribution of the updates and the non-parametric bootstrap. We consider both a correctly specified model ($X_t \sim N(0,1)$) as well as two misspecified models ($X_t \sim$ Exponential(1), $\sqrt{20}X_t \sim \chi^2_{10}$). All of the X_t have variance 1. We display both the in- as well as the out-of-control performance.

In the correctly specified model ($X_t \sim N(0,1)$), the performance of the parametric and the non-parametric chart seems to be almost identical. The only difference is a slightly worse in-control performance for the non-parametric chart for $n=50$.

In the misspecified models the parametric charts do not have the desired in-control probabilities. The non-parametric chart are doing well, in particular for $n=500$.

5. Regression models

In many monitoring situations, the units being monitored are heterogeneous, for instance when monitoring patients at hospitals or bank customers. To make sensible monitoring systems in such situations, the explainable part of the heterogeneity should be accounted for by relevant regression models. The resulting charts are often called risk adjusted. Some charts of this type can be found in Grigg & Farewell (2004).

To run risk-adjusted charts, the regression model needs to be estimated based on past data, and this estimation needs to be accounted for. Our approach for setting up charts with a guaranteed performance applies also to risk-adjusted charts. We will only look at linear models in this study, however the approach also works for other regression models such as logistic models and survival analysis models.

Suppose we have independent observations (Y_1, X_1), $(Y_2, X_2), \ldots,$ where Y_i is a response of interest and X_i is a corresponding vector of covariates, with the first component usually equal to 1. Let P denote the joint distribution of (Y_i, X_i) and suppose that in control $\mathrm{E}(Y_i \mid X_i)=X_i\xi$. From some observation κ there is a shift in the mean response to $E(Y_i \mid X_i)=\Delta+X_i\xi$ for $i=\kappa,\kappa+1,\ldots.$

Monitoring schemes for detecting changes in regression models can naturally be based on residuals of the model, see for instance Brown *et al.* (1975) and Horváth *et al.* (2004). We can, for instance, define a CUSUM to monitor changes in the conditional mean of Y by

$$S_t=\max(0, S_{t-1}+Y_t-X_t\xi-\Delta/2), \quad S_0=0,$$

with hitting time $\tau=\inf\{t>0: S_t \geq c\}$ for some threshold $c>0$. In a similar manner we could also set up charts for monitoring changes in other components of ξ.

The parameter vector ξ is estimated from past in control data, for example, by the standard least squares estimator. We suggest to use a non-parametric version of the general Algorithm 1 with $\hat{P}$ being the empirical distribution putting weight $1/n$ on each of the past observations $(Y_{-n}, X_{-n}), \ldots, (Y_{-1}, X_{-1})$.

Note that the non-parametric bootstrap should give good coverage probabilities even if the linear model is misspecified, that is, $\mathrm{E}(Y_i \mid X_i)=X_i\xi$ does not necessarily hold.

An analogous approach can be used for Shewhart charts. In settings where it is reasonable to consider the covariate vector to be non-random one could alternatively use bootstrapping of residuals (see e.g. Freedman, 1981).

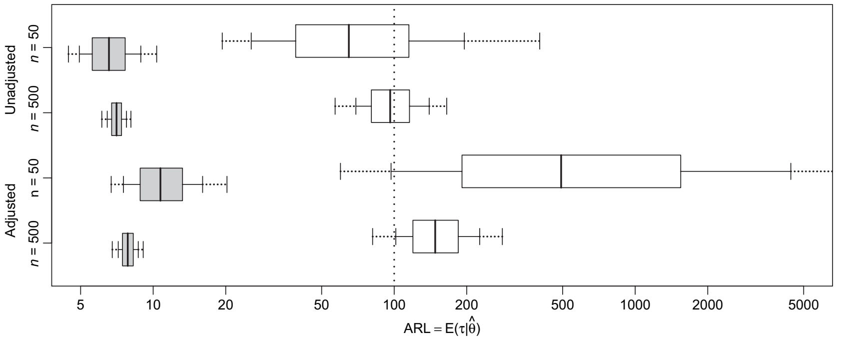

Fig. 5. Distribution of the conditional ARL for CUSUMs in a linear regression setup. Thresholds are calibrated to an in-control ARL of 100 using a log transform. Adjusted thresholds have a 90% guarantee. White boxplots are in-control, gray boxplots are out-of-control. The boxplots show the 2.5%, 10%, 25%, 50%, 75%, 90% and 97.5% quantiles. Log-scale on the horizontal axis.

5.1. Theoretical considerations

Obtaining precise results is more demanding than in the examples without covariates in section 3.2. We only give an idea of the setup that might be used.

The set of distributions of the observations D_q can be chosen as the set of cdfs on $\mathbb{R}^{d+1}$ with finite second moments, where d is the dimension of the covariate. The first cdf corresponds to the responses, the others to the covariates. D_q is contained in the vector space $D = l_\infty(\mathbb{R}^{d+1})$, the set of bounded functions $\mathbb{R}^{d+1} \to \mathbb{R}$. The parameters needed to run the chart are the regression coefficients contained in the set $\Xi = \mathbb{R}^d$. These parameters are obtained from the distribution of the observations via $\xi : D_q \to \Xi$, $F \mapsto (E(X^{\mathrm{T}}X))^{-1}E(XY)$ where $(Y, X) \sim F$ and X is a row vector.

We conjecture that the conditions of theorem 1 are broadly satisfied if the cdf of $Y - X\xi$ is differentiable and if for the property q we use hitting probabilities or thresholds to achieve a given hitting probability. In particular, it should be possible to show Hadamard differentiability similarly to lemma 1: write q as concatenation of two functions and use the chain rule in lemma 4. The first mapping returns the distribution of the updates of the chart depending on $F \in D_q$ and $\xi \in \Xi$ via $(F; \xi) \mapsto \mathcal{L}(Y - X\xi - \Delta)$, where $\mathcal{L}$ denotes the law of a random variable. The second takes the distribution of the updates and returns the property of interests. The differentiability of the second map has been shown in lemmas 5 and 6.

5.2. Simulations

We illustrate the performance of the bootstrapping scheme using a CUSUM chart and the linear in-control model $Y = X_1 + X_2 + X_3 + \epsilon$. Let $\epsilon \sim N(0, 1)$, $X_1 \sim$ Bernoulli(0.4), $X_2 \sim U(0, 1)$ and $X_3 \sim N(0, 1)$, where X_1, X_2, X_3 and ϵ are all independent. The out-of-control model is $Y = 1 + X_1 + X_2 + X_3 + \epsilon$, that is, $\Delta = 1$. Figure 5 shows the distribution of the attained ARL for CUSUMs with thresholds calibrated to give an in control ARL of 100. The behaviour of the adjusted versus unadjusted thresholds are very similar to what we observed for the simpler model in Fig. 3. The coverage probabilities for this regression model, not reported here, are very similar to the coverage probabilities in Table 1, though with a tendency to be slightly worse.

6. Discussion and conclusions

We have presented a general approach for handling estimation error in control charts with estimated parameters and unknown in-control distributions. Our suggestion is, by bootstrap methods, to tune the monitoring scheme to guarantee, with high probability, a certain conditional in-control performance (conditional on the estimated in-control distribution). If we apply a non-parametric bootstrap, the approach is robust against model specification error.

In our opinion, focusing on a guaranteed conditional in-control performance is generally more relevant than focusing on some average performance, as an estimated chart usually is run for some time without independent reestimation. Our approach can also easily be adapted to make for instance bias adjustments. Bias adjustments, in contrast to guaranteed performance, tend to be substantially influenced by tail behaviour for heavy tailed distributions which for instance the average run length has. This implies that the bias adjustments need not be useful in the majority of cases as the main effect of the adjustment is to adjust the tail behaviour.

We have demonstrated our approach for various variants of Shewhart and CUSUM charts, but the approach will apply to other charts as well. The method is generally relevant when the in-control distribution is unknown and the conditions of theorem 1 hold. We conjecture that this will be the case for many of the most commonly used control charts. Numerous extensions of control charts to other settings exist, for example to other regression models, to autocorrelated date and to multivariate data. We do conjecture that our approach will also apply in many of these settings. The method is implemented for various situations in the R-package spcadjust.

References

Albers, W. & Kallenberg, W. C. (2004a). Are estimated control charts in control? *Statistics* **38**, 67–79.

Albers, W. & Kallenberg, W. C. (2004b). Estimation in Shewart control charts: effects and corrections. *Metrika* **59**, 207–234.

Albers, W. & Kallenberg, W. C. (2005). New corrections for old control charts. *Quality Eng.* **17**, 467–473.

Albers, W. & Kallenberg, W. C. (2006). Self-adapting control charts. *Statist. Neerl.* **60**, 292–308.

Albers, W. & Kallenberg, W. C. (2009). Cumin charts. *Metrika* **70**, 111–113.

Albers, W. & Kallenberg, W. C. (2010). The optimal choice of negative binomial charts for monitoring high-quality processes. *J. Statist. Plann. Inference* **140**, 214–225.

Albers, W., Kallenberg, W. C. & Nurdiati, S. (2005). Exceedance probabilities for parametric control charts. *Statistics* **39**, 429–443.

Aue, A., Horváth, L., Hušková, M. & Kokoszka, P. (2006). Change-point monitoring in linear models. *Econom. J.* **9**, 373–403.

Biswas, P. & Kalbfleisch, J. D. (2008). A risk-adjusted CUSUM in continuous time based on the Cox model. *Statist. Med.* **27**, 3382–3406.

Bottle, A. & Aylin, P. (2008). Intelligent information: a national system for monitoring clinical performance. *Health Serv. Res.* **43**, 10–31.

Brook, D. & Evans, D. A. (1972). An approach to the probability distribution of CUSUM run length. *Biometrika* **59**, 539–549.

Brown, R., Durbin, J. & Evans, J. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **37**, 149–192.

Capizzi, G. & Masarotto, G. (2009). Bootstrap-based design of residual control charts. *IIE Trans.* **41**, 275–286.

Carey, R. G. (2003). *Improving healthcare with control charts: basic and advanced SPC methods and case studies*. ASQ Quality Press, Milwaukee.

Castagliola, P. & Maravelakis, P. E. (2011). A CUSUM control chart for monitoring the variance when parameters are estimated. *J. Statist. Plann. Inference* **141**, 1463–1478.

Chakraborti, S. & Graham, M. (2007). Nonparametric control charts. In *Encyclopedia of statistics in quality and reliability* (eds F. Ruggeri, R. Kenettt & F. Faltin), 415–429. Wiley, Chichester.

Champ, C. W. & Jones-Farmer, L. A. (2007). Properties of multivariate control charts with estimated parameters. *Sequential Anal.* **26**, 153–169.

Chatterjee, S. & Qiu, P. (2009). Distribution-free cumulative sum control charts using bootstrap-based control limits. *Ann. Appl. Stat.* **3**, 349–369.

Fouladirad, M., Grall, A. & Dieulle, L. (2008). On the use of on-line detection for maintenance of gradually deteriorating systems. *Reliability Eng. Syst. Safety* **93**, 1814–1820.

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9**, 1218–1228.

Frisén, M. (ed.) (2008). *Financial surveillance*. Wiley, Chichester.

Frisén, M. & Andersson, E. (2009). Semiparametric surveillance of monotonic changes. *Sequential Anal.* **28**, 434–454.

Gandy, A., Kvaløy, J., Bottle, A. & Zhou, F. (2010). Risk-adjusted monitoring of time to event. *Biometrika* **97**, 375–388.

Grigg, O. & Farewell, V. (2004). An overview of risk-adjusted charts. *J. Roy. Statist. Soc. Ser. A* **167**, 523–539.

Hawkins, D. M. (1987). Self-starting cusum charts for location and scale. *J. Roy. Statist. Soc. Ser. D* **36**, 299–316.

Hawkins, D. M. & Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*. Springer, New York.

Horváth, L., Hušková, M., Kokoszka, P. & Steinebach, J. (2004). Monitoring changes in linear models. *J. Statist. Plann. Inference* **126**, 225–251.

Hušková, M. & Kirch, C. (2010). Bootstrapping sequential change-point tests for linear regression. *Metrika* **75**, 673–708.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W. & Woodall, W. H. (2006). Effects of parameter estimation on control chart properties; a literature review. *J. Quality Technol.* **38**, 349–364.

Jones, L. A. (2002). The statistical design of EWMA control charts with estimated parameters. *J. Quality Technol.* **34**, 277–288.

Jones, L. A., Champ, C. W. & Rigdon, S. E. (2004). The run length distribution of the cusum with estimated parameters. *J. Quality Technol.* **36**, 95–108.

Kirch, C. (2008). Bootstrapping sequential change-point tests. *Sequential Anal.* **27**, 330–349.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer, New York.

Lawson, A. B. & Ken, K. (eds) (2005). *Spatial and syndromic surveillance for public health*. Wiley, Chichester.

Maravelakis, P. E. & Castagliola, P. (2009). An EWMA chart for monitoring the process standard deviation when parameters are estimated. *Comput. Statist. Data Anal.* **53**, 2653–2664.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14**, 1379–1387.

O'Connor, P. D. (2002). *Practical reliability engineering*. Wiley, Chichester.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–115.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Van Nostrand, New York.

Steiner, S. H., Cook, R. J., Farewell, V. T. & Treasure, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* **1**, 441–452.

Stoumbos, Z. G., Marion, R., Reynolds, J., Ryan, T. P. & Woodall, W. H. (2000). The state of statistical process control as we proceed into the 21st century. *J. Amer. Statist. Assoc.* **95**, 992–998.

Sullivan, J. H. & Jones, L. A. (2002). A self-starting control chart for multivariate individual observations. *Technometrics* **44**, 24–33.

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.

van der Vaart, A. & Wellner, J. (1996). *Weak convergence and empirical processes*. Springer, New York.

Woodall, W. H. (2006). The use of control charts in health-care and public-health surveillance. *J. Quality Technol.* **38**, 89–134. With discussion.

Xie, M., Goh, T. & Ranjan, P. (2002). Some effectice control chart procedures for reliability monitoring. *Reliability Eng. Syst. Safety* **77**, 143–150.

Zhang, Y., Castagliola, P., Wu, Z. & Khoo, M. B. C. (2011). The synthetic $\bar{X}$ chart with estimated parameters. *IIE Trans.* **43**, 676–687.

Received May 2012, in final form November 2012

Axel Gandy, Department of Mathematics, Imperial College London, London SW7 2AZ, UK.
E-mail: a.gandy@imperial.ac.uk

Appendix A. Proof of the main theorem

The following extension of the functional delta method will help in the proof of theorem 1.

Lemma 3. *Suppose that $q: D_q \times \Xi \to E$ is Hadamard differentiable at $P \in D_q$ around $\xi \in \Xi$ tangentially to $D_0 \subset D$ and that $\xi: D_q \to \Xi$ is continuous. Let $\hat{P}$ be a sequence of random elements in D_q such that $\sqrt{n}(\hat{P} - P) \rightsquigarrow Z$ as $n \to \infty$,* where *Z is some tight random element in D_0. Then*

$$\sqrt{n}(q(\hat{P}; \xi(\hat{P})) - q(P; \xi(\hat{P}))) \rightsquigarrow q'(P; \xi(P))Z.$$

Proof. Note that $\sqrt{n}(q(\hat{P}; \hat{\xi}) - q(P; \hat{\xi})) = g_n(\sqrt{n}(\hat{P} - P))$, where $g_n: \tilde{D}_n \to F$, $g_n(h) = \sqrt{n}[q(P + n^{-\frac{1}{2}}h; \xi(P + n^{-\frac{1}{2}}h)) - q(P; \xi(P + n^{-\frac{1}{2}}h))]$ and $\tilde{D}_n = \{h \in D: P + n^{-\frac{1}{2}}h \in D_q\}$.

Let h_n be a sequence such that $h_n \in \tilde{D}_n$ and $h_n \to h$ for some $h \in D_0$. Let $\xi_n = \xi(P + n^{-\frac{1}{2}}h_n)$. The continuity of ξ implies $\xi_n \to \xi(P)$. Thus by the Hadamard differentiability of q we get $g_n(h_n) \to q'(P; \xi(P))(h)$. Using the extended continuous mapping theorem (van der Vaart & Wellner, 1996, theorem 1.11.1) finishes the proof.

Proof of theorem 1. Let $\tilde{Z}_1$ and $\tilde{Z}_2$ be independent copies of Z. Arguing as in the first part of the proof of (Kosorok, 2008, theorem 12.1) one can see that unconditionally

$$\sqrt{n}\left(\begin{pmatrix}\hat{P}^*\\ \hat{P}\end{pmatrix} - \begin{pmatrix}P\\ P\end{pmatrix}\right) \rightsquigarrow \begin{pmatrix}\tilde{Z}_1 + \tilde{Z}_2\\ \tilde{Z}_2\end{pmatrix}.$$

Applying lemma 3 to this with the mappings $(x, y; \xi) \mapsto (q(x; \xi), q(y; \xi), x, y)$ and $(x, y) \mapsto \xi(x)$ gives

$$\sqrt{n}\begin{pmatrix}q(\hat{P}^*; \hat{\xi}^*) - q(P; \hat{\xi}^*)\\ q(\hat{P}; \hat{\xi}^*) - q(P; \hat{\xi}^*)\\ \hat{P}^* - P\\ \hat{P} - P\end{pmatrix} \rightsquigarrow \begin{pmatrix}q'(P; \xi)(\tilde{Z}_1 + \tilde{Z}_2)\\ q'(P; \xi)(\tilde{Z}_2)\\ \tilde{Z}_1 + \tilde{Z}_2\\ \tilde{Z}_2\end{pmatrix}.$$

After that one can argue exactly as in the remainder of the proof of (Kosorok, 2008, theorem 12.1, p. 237), to show that

$$A_n := \sqrt{n}(q(\hat{P}^*; \xi(\hat{P}^*)) - q(\hat{P}; \xi(\hat{P}^*)) \underset{W}{\overset{P}{\rightsquigarrow}} G \quad (n \to \infty) \tag{6}$$

for $G = q'(P; \xi)Z$. Furthermore, lemma 3 shows that

$$B_n := \sqrt{n}(q(\hat{P}; \hat{\xi}) - q(P; \hat{\xi})) \rightsquigarrow G. \tag{7}$$

Similarly to the ideas in van der Vaart (1998, Lemma 23.3), one can show that (6) and (7) imply the correct coverage probabilities. Indeed, for any subsequence there is a further subsequence such that $A_n \rightsquigarrow G$ a.s. conditionally on $\hat{P}$. Using van der Vaart (1998, lemma 21.2), we get $F_{A_n}^{-1} \rightsquigarrow F_G^{-1}$ along this subsequence, where F^{-1} denotes the quantile function of the random variable in the subscript, that is, $F_G^{-1}(x) = \inf\{t \in \mathbb{R}: P(G \leq t) \geq x\}$. Thus for any continuity point β of F_G^{-1}, we get $F_{A_n}^{-1}(\beta)\to F_G^{-1}(\beta)$ a.s. along the subsequence. Thus overall, we have $F_{A_n}^{-1}(\beta)\xrightarrow{\mathrm{P}} F_G^{-1}(\beta)$. By Slutsky's lemma and (7), $B_n - F_{A_n}^{-1}(\beta)\rightsquigarrow G - F_G^{-1}(\beta)$. Thus, as G is continuous,

$$\mathrm{P}(B_n\le F_{A_n}^{-1}(\beta))\to \mathrm{P}(G\le F_G^{-1}(\beta))=\beta. \tag{8}$$

This holds for all β, because there are at most countably many points β at which F_G^{-1} is not continuous, because both the left- and the right-hand side of (8) are monotone in β, and because the right-hand side is continuous. As $p_\alpha^* = F_{A_n}^{-1}(\alpha)/\sqrt{n}$, (8) implies

$$\mathrm{P}(q(P;\hat\xi)<q(\hat P;\hat\xi)-p_\alpha^*)=1-\mathrm{P}(q(\hat P;\hat\xi)-q(P;\hat\xi)\le p_\alpha^*)=1-\mathrm{P}(B_n\le F_{A_n}^{-1}(\alpha))\to 1-\alpha.$$

Appendix B. Proofs for Hadamard differentiability

The main goal of this section is to prove lemmas 1 and 2. Before we do this in appendix B.4, we first show a chain rule in appendix B.1, then a lemma about the uniform Hadamard differentiability of inverse maps in appendix B.2. After that we show general differentiability of hitting probability in CUSUM charts with respect to the updating distribution in appendix B.3. The results in B.1–B.3 may also be useful in other situations.

B.1. Chain rule

In this section we present a chain rule for Hadamard differentiable functions. For this we need the following stronger version of Hadamard differentiability.

Definition 2. *Let D, E be metric spaces. A function $\phi: D_\phi\subset D\to E$ is called uniformly Hadamard differentiable at $\theta\in D_0$ along $d: D\times D\to\mathbb{R}$ tangentially to $D_0\subset D$ if there exists a linear map $\phi'_\theta: D_0\to E$ such that*

$$\frac{\phi(\theta_n+t_nh_n)-\phi(\theta_n)}{t_n}\to\phi'_\theta(h),$$

for all $\theta_n\to\theta$ with $d(\theta_n,\theta)\to 0$, $t_n\to 0$ and all converging sequences (h_n) with $h_n\to h\in D_0$ and $\theta_n+t_nh_n\in D_\phi$.

Lemma 4 (Chain rule). *Let D, E, F be metric spaces and let H be a non-empty set. Let $\{f_\xi: D_f\to E:\xi\in\Xi\}$ be a family of functions that is Hadamard differentiable at $\theta\in D_f$ around $\xi\in\Xi$ tangentially to $D_0\subset D$.* Let *$\phi: E_\phi\to F$ be uniformly Hadamard differentiable at $f_\xi(\theta)$ along $d: E_\phi\times E_\phi\to\mathbb{R}$ tangentially to $f'_\xi(\theta;\xi)(D_0)$. Furthermore, suppose that $\xi_n\to\xi$ implies $d(f(\theta;\xi_n), f(\theta;\xi))\to 0$. Then $\{\phi\circ f_\xi: D_f\to F:\xi\in\Xi\}$ is Hadamard differentiable at θ around $\xi\in\Xi$ tangentially to D_0.*

Proof. Let $(\xi_n)\subset\Xi$, $(t_n)\subset\mathbb{R}$, $(h_n)\subset D$ satisfying $\theta+t_nh_n\in D_f\forall n$ and $\xi_n\to\xi$, $t_n\to 0$, $h_n\to h\in D_0$ as $n\to\infty$. Let $k_n=(f_{\xi_n}(\theta+t_nh_n)-f_{\xi_n}(\theta))/t_n$. Hadamard differentiability of f implies $k_n\to q'_\xi(h)$. Then by uniform Hadamard differentiability of ϕ,

$$\frac{\phi(f_{\xi_n}(\theta+t_nh_n))-\phi(f_{\xi_n}(\theta))}{t_n}=\frac{\phi(f_{\xi_n}(\theta)+t_nk_n)-\phi(f_{\xi_n}(\theta))}{t_n}\to\phi'_{f_\xi(\theta)}(q'_\xi(\theta)(h)).$$

B.2. Uniform Hadamard differentiability of the inverse map

Let D_ϕ be the set of non-decreasing functions in $D[u,v]$, for some $-\infty<u<v<\infty$, that cross $\beta\in\mathbb{R}$, that is,

$$D_\phi=\{F\in D[u,v]:F \text{ non-decreasing}, \exists x\in(u,v]:F(x-)\leq\beta\leq F(x)\}.$$

Suppose that $F\in D_\phi$ and $G\in D_\phi$ are differentiable on $[u,v]$ with derivatives f and g. Let $d(F,G)=\sup_{x\in[u,v]}|f(u)-g(u)|$. If either F or G are not differentiable on $[u,v]$ then we set $d(F,G)=\infty$. Let $\phi:D_\phi\to\mathbb{R}, \phi(F)=\inf\{x:F(x)\geq\beta\}$, the first point at which the function crosses the threshold.

Lemma 5. *Let $\theta\in D_\phi$ such that θ is differentiable on $[u,v]$ with continuous bounded positive derivative. Then ϕ is uniformly Hadamard differentiable at θ along d tangentially to $C[u,v]$.*

Proof. Let $\xi=\phi(\theta)$. Let $(h_n)\subset D[u,v]$ such that $h_n\to h\in C[u,v]$. Let $(t_n)\subset[0,\infty)$ such that $t_n\to 0$. Let $(\theta_n)\subset D_\phi$ such that $\theta_n\to\theta$ and $d(\theta_n,\theta)\to 0$. Let $\xi_n=\phi(\theta_n+t_nh_n)$. By the definition of ϕ, we have

$$(\theta_n+t_nh_n)(\xi_n-\epsilon_n)\leq\beta\leq(\theta_n+t_nh_n)(\xi_n), \tag{9}$$

for every $\epsilon_n>0$. Let (ϵ_n) be positive and such that $\epsilon_n=o(t_n)$.

First, we show $\xi_n\to\xi$. The sequence (h_n) is uniformly bounded because $h_n\to h$ and h is bounded. Thus, $\theta_n(\xi_n-\epsilon_n)+O(t_n)\leq\beta\leq\theta_n(\xi_n)+O(t_n)$. As $t_n\to 0$ and $\theta_n\to\theta$,

$$\theta(\xi_n-\epsilon_n)+o(1)\leq\beta\leq\theta(\xi_n)+o(1).$$

For every $\delta>0$, the function θ is bounded away from β outside $(\xi-\delta,\xi+\delta)$. Furthermore, θ is strictly increasing. Thus, to satisfy the previous display we must have eventually $\xi_n\geq\xi-\delta$ and $\xi_n-\epsilon_n\leq\xi+\delta$, which implies $\xi_n\to\xi$.

Let $\tilde{\xi}_n=\phi(\theta_n)$. Using the mean value theorem in (9) gives

$$\theta_n(\tilde{\xi}_n)+(\xi_n-\epsilon_n-\tilde{\xi}_n)\theta_n'(\rho_{1n})+t_nh_n(\xi_n-\epsilon_n)\leq\beta\leq\theta_n(\tilde{\xi}_n)+(\xi_n-\tilde{\xi}_n)\theta_n'(\rho_{2n})+t_nh_n(\xi_n)$$

for some ρ_{1n} between $\xi_n-\epsilon_n$ and $\tilde{\xi}_n$ and for some ρ_{2n} between ξ_n and $\tilde{\xi}_n$. Rewriting this using $\theta_n(\tilde{\xi}_n)=\beta$ gives

$$(\xi_n-\tilde{\xi}_n)\theta_n'(\rho_{1n})+t_nh_n(\xi_n-\epsilon_n)-\epsilon_n\theta_n'(\rho_{1n})\leq 0\leq(\xi_n-\tilde{\xi}_n)\theta_n'(\rho_{2n})+t_nh_n(\xi_n).$$

By the uniform convergence of h_n and the continuity of h, we have $h_n(\xi_n-\epsilon_n)\to h(\xi)$ and $h_n(\xi_n)\to h(\xi)$. Using this, the fact that we have chosen ϵ_n such that $\epsilon_n=o(t_n)$ and that θ_n' is uniformly bounded, we get

$$(\xi_n-\tilde{\xi}_n)\theta_n'(\rho_{1n})-o(t_n)\leq -t_nh(\xi)\leq(\xi_n-\tilde{\xi}_n)\theta_n'(\rho_{2n})+o(t_n).$$

Hence,

$$-\frac{h(\xi)}{\theta_n'(\rho_{2n})}-o(1)\leq\frac{\xi_n-\tilde{\xi}_n}{t_n}\leq-\frac{h(\xi)}{\theta_n'(\rho_{1n})}+o(1).$$

We have already shown $\xi_n\to\xi$ which implies $\rho_{1n}\to\xi$ and $\rho_{2n}\to\xi$. Using the assumptions that $\theta_n'\to\theta$ uniformly and that θ' is continuous shows $\theta_n'(\rho_{1n})\to\theta'(\xi)$ and $\theta_n'(\rho_{2n})\to\theta'(\xi)$, which finishes the proof.

B.3. Differentiability of hitting probabilities

We are interested in hitting probabilities for CUSUM charts within the first $T \in \mathbb{N}$ steps. We show that the mapping from the distribution of the updates Y_i to the hitting probabilities is uniformly Hadamard differentiable. The Y_i are the adjusted observations, for example, in the notation of section 3.2 they are $Y_i = (X_i - \xi_1 - \Delta/2)/\xi_2$, where X_i is the observed value.

Let D_ϕ be the set of cumulative distribution functions on $\mathbb{R}$, considered as a subset of $D = l_\infty(\mathbb{R})$ equipped with the uniform norm. Consider the mapping

$$\phi : D_\phi \to l_\infty(\mathbb{R}), q(F)(c) = P(\text{hit threshold c within T}) = \int g(y, c)\, \mathrm{d}F(y_1) \ldots \mathrm{d}F(y_T),$$

where $g(y, c) = 1(m(y) \geq c)$ with $1(\cdot)$ being the indicator function and $m(y)$ is as defined in section 3.2.

Lemma 6. *ϕ is uniformly Hadamard differentiable tangentially to $D_0 = \{H \in C(\mathbb{R}) : \lim_{t\to\infty} H(t) = \lim_{t\to-\infty} H(t) = 0\}$ with derivative*

$$\phi'(F)(H)(c) = \sum_{i=1}^{T} \int g(y, c) \left(\prod_{j \neq i} \mathrm{d}F(y_j) \right) \mathrm{d}H(y_i).$$

Since H may be of infinite variation, the above integral is defined by partial integration, that is,

$$\phi'(F)(H)(c) = -\sum_{i=1}^{T} \int H(y_i)\, \mathrm{d}\left(\int g(y, c) \left(\prod_{j \neq i} \mathrm{d}F(y_j) \right) \right).$$

Proof. Suppose $F_n \to F$, $t_n \to 0$, $H_n \to H \in D_0$ such that $F_n + t_n H_n \in D_\phi$ for all n. The difference quotient can be written as

$$\frac{\phi(F_n + t_n H_n)(c) - \phi(F_n)(c)}{t_n} = \sum_{i=1}^{T} \int g(y, c) \left(\prod_{j \neq i} \mathrm{d}F_n(y_j) \right) \mathrm{d}H_n(y_i) + \sum_{\substack{I \subset \{1,\ldots,T\} \\ |I| \geq 2}} t_n^{|I|-1} A_I, \quad (10)$$

where $A_I = \int g(y, c) \left(\prod_{i \in I} \mathrm{d}F_n(y_i) \right) \left(\prod_{i \in I} \mathrm{d}H_n(y_i) \right)$. We first show that the second terms converge uniformly in c to 0. Partial integration (applied several times) gives that for $I \subset \{1, \ldots, T\}, |I| \geq 2$,

$$A_I = (-1)^{|I|} \int \left(\prod_{i \in I} H_n(y_i) \right) \mathrm{d}B_I(y_I), \quad (11)$$

where $B_I = \int g(y, c) \prod_{i \in I} \mathrm{d}F_n(y_i)$. $g(y, c)$ is monotonically increasing in y thus B_I is increasing in $y_I = (y_i)_{i \in I}$. Thus the total variation of B_I is bounded by 1. Hence, using (11),

$$t_n^{|I|-1} A_I \leq t_n^{|I|-1} \left(\sup_{z \in \mathbb{R}} |H_n(z)| \right)^{|I|},$$

which converges to 0 uniformly in c. Thus the second term of (10) converges to 0 uniformly in c. Next, we show that first term on the right-hand side of (10), henceforth denoted by ζ, converges uniformly in c to $\phi'(F)(H)$. Consider the decomposition

$$\zeta - \phi'(F)(H) = C_n + \sum_{i=1}^{T} \int g(y, c) \left(\left(\prod_{j \neq i} \mathrm{d}F_n(y_j) \right) - \prod_{j \neq i} \mathrm{d}F(y_j) \right) \mathrm{d}H(y_i), \quad (12)$$

where $C_n = \sum_{i=1}^{T} \int g(y,c)\left(\prod_{j\neq i} \mathrm{d}F_n(y_j)\right)\left(\mathrm{d}H_n(y_i) - \mathrm{d}H(y_i)\right)$. As mentioned above, H might be of infinite variation, thus C_n is defined via partial integration, that is

$$C_n = -\sum_{i=1}^{T} \int (H_n(y_i) - H(y_i))\,\mathrm{d}\left(\int g(y,c) \prod_{j\neq i} \mathrm{d}F_n(y_j)\right).$$

As above, the total variation of the integrator is bounded by 1, thus $|C_n| \leq T\|H_n - H\|$, which converges to 0 as $n \to \infty$.

We can rewrite the second term of (12) as

$$\sum_{i=1}^{T} \sum_{k=1,k\neq i}^{T} \int D_{ik}\left(\left(\prod_{j\neq i, j<k} \mathrm{d}F_n(y_j)\right) \prod_{j\neq i, j>k} \mathrm{d}F(y_j)\right),$$

where $D_{ik} = \int g(y,c)\,\mathrm{d}H(y_i)(\mathrm{d}F_n(y_k) - \mathrm{d}F(y_k))$. Using partial integration,

$$\begin{aligned} D_{ik} &= -\int H(y_i)\,\mathrm{d}g(y_{-i}, \mathrm{d}y_i, c)(\mathrm{d}F_n(y_k) - \mathrm{d}F(y_k)) \\ &= \int H(y_i)(F_n(y_k) - F(y_k))\,\mathrm{d}g(y_{-i,-k}, \mathrm{d}y_i, \mathrm{d}y_k, c), \end{aligned}$$

where negative subscripts denote removal of the corresponding component of the vector (e.g. y_{-i} is the vector y with the ith component removed). Since g is of bounded variation with respect to y_i and y_k independent of c and $y_{-i,-k}$, we can bound this uniformly above by $K \sup_z |H(z)| \sup_z |F_n(z) - F(z)|$ for some fixed $K > 0$. Thus, since the variation of $F_n - F$ is bounded by 2, the second term of (12) converges to 0 uniformly in c.

The following lemma is needed to use the result about the inverse mapping, see lemma 5.

Lemma 7. *Let F be a cumulative distribution function with continuous bounded positive derivative f. Let $Y = (Y_1, \ldots, Y_T)$ where $Y_1, \ldots, Y_T \sim F$ independently. Then:*

(a) *$c \mapsto \mathrm{P}(m(Y) \leq c)$ is continuously differentiable for $c > 0$ (call this derivative g).*
(b) *g is bounded away from 0 (at least on some compact sets).*
(c) *Let f_n be densities that converge uniformly to f. Let $Y^{(n)} = (Y_1^{(n)}, \ldots, Y_T^{(n)}) \sim f_n$ and denote the derivative of $c \mapsto \mathrm{P}(m(Y^{(n)}) \leq c)$ for $c > 0$ by $g^{(n)}$. Then $g^{(n)}$ converges uniformly to g on any compact set $K \subset (0, \infty)$.*

Proof. For $0 \leq i \leq T$, let $A_i = \{y \in \mathbb{R}^{\mathrm{T}} : R_i(y) > R_v(y) \forall v \neq i\}$. Then A_i are disjoint sets with $\mathrm{P}(Y \in \bigcup_i A_i) = 1$. Thus, $\mathrm{P}(m(Y) \leq c) = \sum_i \mathrm{P}(m(Y) \leq c, Y \in A_i)$ and $g(c) = \sum_{i=1}^{T} g_i(c)$, where

$$\begin{aligned} g_i(c) &= \frac{\partial}{\partial c}\mathrm{P}(m(Y) \leq c, Y \in A_i) = \frac{\partial}{\partial c}\mathrm{P}(R_i(Y) \leq c, Y \in A_i) \\ &= \frac{\partial}{\partial c}\mathrm{P}(Y_i \leq c - R_{i-1}(Y), Y \in A_i) = \int \frac{\partial}{\partial c} \int^{c - R_{i-1}(y)} 1(y \in A_i) f(y_i)\,\mathrm{d}y_i \prod_{v\neq i} f(y_v)\,\mathrm{d}y_{-i} \\ &= \int 1((y_1, \ldots, y_{i-1}, c - R_{i-1}(y), y_{i+1}, \ldots, y_T) \in A_i) f(c - R_{i-1}(y)) \prod_{v\neq i} f(y_v)\,\mathrm{d}y_{-i}. \end{aligned}$$

The continuity of g follows because of the dominated convergence theorem. This shows (a). Statement (b) follows from g being positive and its continuity. For (c), use a telescoping sum to go from the product of fs to the product of f_ns. Then use the dominated convergence theorem.

B.4. Hadamard differentiability of hitting probability in simple examples

Proof of lemma 1. hit can be written as $\phi \circ g$, where ϕ is as in appendix B.3 and where $g: D_q \to D_q, g(P;\xi) = (x \mapsto P(x\xi_2 + \xi_1 + \Delta/2))$. As lemma 6 allows the use of the chain rule (lemma 4), it suffices to show that g is Hadamard differentiable at P around ξ tangentially to D_0. Clearly, g is linear in P. Thus for $t_n \to 0$, $h_n \to h \in D_0$, $\xi_n \to \xi$,

$$\begin{aligned}\frac{g(P + t_n h_n; \xi_n) - g(P; \xi_n)}{t_n} - g(h; \xi) &= g(h_n; \xi_n) - g(h; \xi) \\ &= (g(h_n; \xi_n) - g(h; \xi_n)) + (g(h; \xi_n) - g(h; \xi)).\end{aligned}$$

The first term converges uniformly to 0 as $h_n \to h$. The second term converges to 0 as $h \in D_0$ implies that h is uniformly continuous.

To see the differentiability of c_{hit}: $\xi_n \to \xi$ implies that the derivative of $g(P; \xi_n)$ converges uniformly to the derivative of $g(P; \xi)$. As $g(P; \xi_n)'(x) = f(x\xi_{2n} + \xi_{1n} + \Delta/2)\xi_{2n}$, this is implied by the uniform continuity of f. Thus, by lemma 7, the derivative of hit$(P; \xi_n)$ converges uniformly to the derivative of hit$(P; \xi)$. It remains to apply the chain rule (lemma 4), the differentiability of hit, and the differentiability of the inverse (lemma 5).

Proof of lemma 2. Let $g: (\mathbb{R} \times (0, \infty))^2 \to D_q$, $g(\mu, \sigma, \xi) = (x \mapsto \Phi((\xi_1 + \Delta/2 + \xi_2 x - \mu)/\sigma))$. Then, as in the proof of lemma 1, $\text{hit}^N = \phi \circ g$. The proof can be completed with similar steps.