



HAL
open science

Accurate tests and intervals based on linear cusum statistics

Christopher S. Withers, Saralees Nadarajah

► **To cite this version:**

Christopher S. Withers, Saralees Nadarajah. Accurate tests and intervals based on linear cusum statistics. *Statistics and Probability Letters*, 2009, 79 (5), pp.689. 10.1016/j.spl.2008.10.018 . hal-00508916

HAL Id: hal-00508916

<https://hal.science/hal-00508916>

Submitted on 7 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Accurate tests and intervals based on linear cusum statistics

Christopher S. Withers, Saralees Nadarajah

PII: S0167-7152(08)00499-9
DOI: 10.1016/j.spl.2008.10.018
Reference: STAPRO 5251

To appear in: *Statistics and Probability Letters*

Received date: 24 July 2008
Revised date: 15 October 2008
Accepted date: 17 October 2008

Please cite this article as: Withers, C.S., Nadarajah, S., Accurate tests and intervals based on linear cusum statistics. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.10.018

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Accurate Tests and Intervals Based on Linear Cusum Statistics

by

Christopher S. Withers
Applied Mathematics Group
Industrial Research Limited
Lower Hutt, NEW ZEALAND

Saralees Nadarajah
School of Mathematics
University of Manchester
Manchester M13 9PL, UK

Abstract: Suppose that we are monitoring incoming observations for a change in mean via a cusum test statistic. The usual nonparametric methods give first and second order approximations for the one- and two-sided cases. We show how to improve the order of these approximations for linear statistics.

Keywords: Approximations; Confidence intervals; Cusums; Edgeworth-Cornish-Fisher.

1 Introduction

Suppose that we are monitoring incoming observations for a change in mean via a cusum test statistic. The usual nonparametric asymptotic methods give a first order approximation in the one-sided case and a second order approximation in the two-sided case, where by *I*th order we mean an error of magnitude $n^{-I/2}$ for n the sample size. We show how to improve on the order of these approximations using simple linear statistics. These are most appropriate when one wishes to ensure against one-sided alternatives like a monotonic trend or jump. We give second order one-sided approximations; when the skewness is known (for example, for a symmetric population) we give third order one-sided and fourth order two-sided approximations.

Set $X_0 = 0$ and let $X = X_1, X_2, \dots$ be independent random variables in \mathcal{R}^p from some distribution $F(x)$ say, with mean μ , and finite moments. (If $p = 1$ we denote the r th cumulant by κ_r and set $\sigma^2 = \kappa_2$.) These observations can be considered as a random process which may at some point go “out of control”, changing their distribution. We define *the average process* of the observations as

$$M_n(t) = n^{-1}S_{[nt]} \quad (1.1)$$

for $S_0 = 0$, $S_i = \sum_{k=1}^i X_k$ and $0 \leq t \leq 1$, where $[x]$ is the integral part of x . A change in mean can be tracked by monitoring the average process via some functional of it, say $T(M_n)$, referred to as *a cusum statistic*.

We denote the mean of the average process (1.1) by $m_n(t) = E M_n(t) = \mu[nt]/n \rightarrow m(t) = \mu t$ as $n \rightarrow \infty$ for $0 \leq t \leq 1$. Given a functional T , $\hat{\theta} = T(M_n)$ can be thought of as an estimate of $\theta = T(m)$. In this way, we can, if desired, use $\hat{\theta}$ to provide an estimate of μ .

For univariate data the most common prospective (or offline) and retrospective (or online) two-sided cusum statistics and functionals are

$$A_n = A(M_n) = \max_{k=1}^n |S_k - k\mu|/n, \quad (1.2)$$

$$B_n = B(M_n) = \max_{k=1}^n |S_k - k\bar{X}_n|/n \quad (1.3)$$

for $A(g) = \sup_{[0,1]} |g(t) - t\mu|$ and $B(g) = \sup_{[0,1]} |g(t) - tg(1)|$, where $\bar{X}_n = M_n(1) = S_n/n$. If $\hat{\sigma}$ is a consistent estimate of the standard deviation σ , then as $n \rightarrow \infty$

$$\hat{\sigma}^{-1}n^{1/2}\{M_n(t) - t\mu\} \xrightarrow{\mathcal{L}} W(t), \quad (1.4)$$

$$\hat{\sigma}^{-1}n^{1/2}\{M_n(t) - tM_n(1)\} \xrightarrow{\mathcal{L}} W_0(t) \quad (1.5)$$

for $W(t)$ a Wiener process and $W_0(t) = W(t) - tW(1)$ a Brownian Bridge. So,

$$\hat{\sigma}^{-1}n^{1/2}A_n \xrightarrow{\mathcal{L}} \sup_{[0,1]} |W(t)|, \quad (1.6)$$

$$\hat{\sigma}^{-1}n^{1/2}B_n \xrightarrow{\mathcal{L}} \sup_{[0,1]} |W_0(t)|. \quad (1.7)$$

See Billingsley (1968) and Anderson and Darling (1952).

If μ is known, (1.6) is commonly used for monitoring the process *online* for a change in mean: at the 95% level one concludes that the mean of the observations has changed when, for some predetermined n , LHS(1.6) > the 95% level of RHS(1.6). If μ is not known, then (1.7) is commonly used for monitoring the process *retrospectively* - that is after the sample is taken. At the 95% level one concludes that the mean of the observations has changed when for some n LHS(1.7) > the 95% level of RHS(1.7).

These asymptotic results are easily extended to functionals like $\sup\{|g(t) - t\mu_0| - b(t)\}$, that is to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ by rejecting H_0 if $M_n(t) - t\mu_0$ crosses a given boundary $\pm b(t)$. An alternative is to use the L_1 norm. For example, the one-sided test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ one can use $\int_0^1 \{g(t) - t\mu_0 - b(t)\}dt$, or equivalently $\int_0^1 g(t)dt = \int g$ say. This is an example of the statistics and functionals we consider here: $T(M_n)$ for $T(g) = \int_0^1 g(t)dw(t)$ or $\int_0^1 w(t)dg(t) = \int wdg$ say for $w(t) : \mathcal{R} \rightarrow \mathcal{R}$ a given function. These functionals have the advantage of being asymptotically normal, unlike (1.6) and (1.7), and of having distribution, density and quantiles given by their Edgeworth-Cornish-Fisher expansions. In contrast, expansions for the distribution, density and quantiles of the cusum statistics (1.2)–(1.7), are not available.

The basic higher order approximations for a *general* estimate with standard cumulant expansions are derived from the Edgeworth-Cornish-Fisher expansions in Section 2 in terms of *the cumulant coefficients*. These coefficients - and the approximations - are given in Section 3 for $T(g)$ linear, including a fourth order confidence interval for μ for the case when the third central moment is known, e.g. for a symmetric population. We would like to point out that the derivations in Section 2 are “formal” in the sense that regularity conditions such as Cramer-type conditions and conditions on differentiability are not explicitly stated.

Lai (1974, 1995) has useful discussion and references on cusum statistics, in particular on linear statistics of the type considered here. In his context one repeatedly takes a small sample of size m : X_i is not the i th observation but a statistic based on this i th sample. For some references on cusum statistics and a first order test based on an alternative one-sided cusum statistic, see Sparks (2000).

2 Some Higher Order Approximations

Here we derive from the Edgeworth-Cornish-Fisher expansions the basic higher order approximations we shall be using. These are given in terms of the *cumulant coefficients*.

Let $\hat{\theta}$ be any real estimate whose cumulants have the standard expansion

$$\kappa_r(\hat{\theta}) = \sum_{i=r-1}^{\infty} a_{ri} n^{-i} \quad (2.8)$$

for $r = 1, 2, \dots$. It follows from Withers (1984) that for $a_{21} \neq 0$,

$$Y_n = (n/a_{21})^{1/2}(\hat{\theta} - \theta), \quad (2.9)$$

where $\theta = a_{10}$, has Edgeworth-Cornish-Fisher expansions of the form

$$\begin{aligned} P_n(x) &= P(Y_n \leq x) \approx \Phi(x) - \phi(x) \sum_{r=1}^{\infty} n^{-r/2} h_r(x), \\ p_n(x) &= dP_n(x)/dx \approx \phi(x) \left\{ 1 + \sum_{r=1}^{\infty} n^{-r/2} \tilde{h}_r(x) \right\}, \\ \Phi^{-1}(P_n(x)) &\approx x - \sum_{r=1}^{\infty} n^{-r/2} f_r(x), \\ P_n^{-1}(\Phi(x)) &\approx \sum_{r=0}^{\infty} n^{-r/2} g_r(x) \end{aligned}$$

for $g_0(x) = x$, where $\Phi(x)$ and $\phi(x)$ are the distribution and density of a unit normal random variable and $h_r(x)$, $\tilde{h}_r(x)$, $f_r(x)$, $g_r(x)$ are certain polynomials in x and the standardised cumulant coefficients

$$A_{ri} = a_{21}^{-r/2} a_{ri}. \quad (2.10)$$

Explicit forms for these polynomials are given in Withers (1984). They involve the Hermite polynomials

$$H_{rx} = \phi(x)^{-1} (-d/dx)^r \phi(x) = E(x + jN)^r$$

for $j = \sqrt{-1}$ and $N \sim \mathcal{N}(0, 1)$: $H_{1x} = x$, $H_{2x} = x^2 - 1$, $H_{3x} = x^3 - 3x$, $H_{4x} = x^4 - 6x^2 + 3$, \dots (See Withers (2000) for a proof of the second expression for H_{rx} .) For $r = 1$ and 2 they are as follows.

$$\begin{aligned} h_1(x) &= f_1(x) = g_1(x) = A_{11} + A_{32} H_{2x}/6, \\ \tilde{h}_1(x) &= A_{11} x + A_{32} H_{3x}/6, \end{aligned}$$

$$\begin{aligned}
h_2(x) &= (A_{11}^2 + A_{22})x/2 + (A_{11}A_{32} + A_{43}/4)H_{3x}/6 + A_{32}^2H_{5x}/72, \\
\tilde{h}_2(x) &= (A_{11}^2 + A_{22})H_{2x}/2 + (A_{11}A_{32} + A_{43}/4)H_{4x}/6 + A_{32}^2H_{6x}/72, \\
f_2(x) &= (A_{22}/2 - A_{11}A_{32}/3)x + A_{43}H_{3x}/24 - A_{32}^2(4x^3 - 7x)/72, \\
g_2(x) &= A_{22}x/2 + A_{43}H_{3x}/24 - A_{32}^2(2x^3 - 5x)/36.
\end{aligned}$$

If $\hat{\theta}$ is lattice there may be correction terms to add.

We shall call a probability statement (such as a confidence interval) based on a normal percentile x , R th order, if it holds with probability $p + O(n^{-R/2})$, where $p = \Phi(x)$ in the one-sided case and $p = 2\Phi(x) - 1$ in the two-sided case. We shall generally write such statements in square brackets. For example, since $Y_n \leq P_n^{-1}(\Phi(x))$ with probability $\Phi(x)$, for $R = 1, 2, \dots$

$$[Y_n \leq y_{nR}(x)] \quad (2.11)$$

is R th order, where

$$y_{nR}(x) = \sum_{r=0}^{R-1} n^{-r/2} g_r(x).$$

For $R = 1$, (2.11) gives inference on θ when a_{21} is known. For $R = 2$, (2.11) gives inference on θ when, for example, a_{11} is a function of θ and a_{21}, a_{32} are both known. For $R = 3$, (2.11) gives inference on θ when a_{11} is a function of θ and $a_{21}, a_{32}, a_{22}, a_{43}$ are known. And so on.

Replacing x by $-x$, and assuming for convenience that the distribution of $\hat{\theta}$ is continuous, we have for $R = 1, 2, \dots$

$$[Y_n \geq y_{nR}(-x)] \text{ is } R\text{th order :} \quad (2.12)$$

$$[Y_n \geq -x] \text{ is first order,} \quad (2.12)$$

$$[Y_n \geq -x + n^{-1/2}g_1(x)] \text{ is second order,} \quad (2.13)$$

$$[Y_n \geq -x + n^{-1/2}g_1(x) - n^{-1}g_2(x)] \text{ is third order,}$$

and so on, since $g_r(-x) = (-1)^{r-1}g_r(x)$. So, for $S = R$, for $R = 1, 2, \dots$,

$$[y_{nR}(-x) \leq Y_n \leq y_{nR}(x)] \quad (2.14)$$

is S th order. In fact, by (5.11) of Withers (1983) - see also Withers (1982, 1988) - (2.14) holds for $S = R + 1$ if R is odd:

$$[|Y_n| \leq x] \text{ is second order,} \quad (2.15)$$

$$[|Y_n - n^{-1/2}g_1(x)| \leq x] \text{ is second order,} \quad (2.16)$$

$$[|Y_n - n^{-1/2}g_1(x)| \leq x + n^{-1}g_2(x)] \text{ is fourth order.}$$

For $\hat{\theta}$ having a continuous distribution, (2.14) can be written

$$|Y_n - \sum_{r=1}^{[R/2]} e_{2r-1}| \leq \sum_{r=0}^{[(R-1)/2]} e_{2r} \quad (2.17)$$

for $e_r = n^{-r/2}g_r(x)$, where $[x]$ is the integral part of x . Now suppose that $\hat{g}_r(x) = g_r(x) + O_p(n^{-1/2})$. Set

$$\hat{y}_{nR}(x) = \sum_{r=0}^{R-2} n^{-r/2}g_r(x) + n^{-(R-1)/2}\hat{g}_{R-1}(x).$$

Then under regularity conditions one can show that for $R = 1, 2, \dots$

$$[Y_n \leq \hat{y}_{nR}(x)] \text{ is } R\text{th order,} \quad (2.18)$$

$$[Y_n \geq \hat{y}_{nR}(-x)] \text{ is } R\text{th order,} \quad (2.19)$$

$$[\hat{y}_{nR}(-x) \leq Y_n \leq \hat{y}_{nR}(x)] \text{ is } S\text{th order,} \quad (2.20)$$

where $S = R$ for R even and $S = R + 1$ for R odd. Note that (2.20) can be written as (2.17) with $g_{R-1}(x)$ replaced by $\hat{g}_{R-1}(x)$.

Now consider the Studentized version of (2.9) $Y_{n0} = (n/\hat{a}_{21})^{1/2}(\hat{\theta} - \theta) = n^{1/2}\hat{\theta}_0$ say. For a wide class of estimates $\hat{\theta}$ with $\hat{a}_{21} = a_{21} + O_p(n^{-1/2})$, the basic cumulant expansion also holds for $\hat{\theta}_0$, say

$$\kappa_r(\hat{\theta}_0) = \sum_{i=r-1}^{\infty} a_{ri0}n^{-i}.$$

Let us denote $g_r(x)$ for $\hat{\theta}_0$ by $g_{r0}(x)$, and the Studentized versions of the approximations (2.11)–(2.20), that is with (Y_n, g_r) replaced by (Y_{n0}, g_{r0}) , by (2.11)₀–(2.20)₀.

Now consider the case $\hat{\theta} = T(M_n)$, $\theta = T(m)$. So, for inference on θ above we can read inference on μ . In Sections 3 we derive the basic expansion (2.8) for linear statistics. It may be shown that for the Studentized version of $\hat{\theta} = T(M_n)$, $a_{210} = 1$, $a_{110} = A_{11} - \gamma_{10}\lambda_3/2$ and $a_{320} = A_{32} - 3\gamma_{10}\lambda_3$, so that

$$g_{10}(x) = g_1(x) - \gamma_{10}\lambda_3x^2/2, \quad (2.21)$$

where

$$\lambda_r = \kappa_r/\sigma^r, \quad \gamma_{10} = (T_1)_1/(T_1^2)_1^{1/2} \quad (2.22)$$

for $(T_1^i)_1 = \int_0^1 T_m(t)^i dt$ and $T_g(t)$ the (suitably defined) functional derivative of $T(g)$. For $T(g) = \int w dg$, γ_{10} reduces to $\gamma_{10} = \int w/(\int w^2)^{1/2}$ as in (3.23) below.

3 Expansions for Linear Functionals of M_n

Here we obtain the basic cumulant expansion (2.8) for univariate data and the linear statistics $\hat{\theta} = T(M_n)$, where $T(g) = \int w dg$ or $\int g dw$ for a given scalar weight function $w : [0, 1] \rightarrow \mathcal{R}$. First consider

$$\hat{\theta} = \int w dM_n = n^{-1} \sum_{i=1}^n w(i/n)X_i.$$

Set $w_r = (\int w^r)^{1/r}$ for $0 < r < \infty$. So, $T(m) = \mu w_1$ and for $r = 1, 2, \dots$ the r th cumulant is

$$\kappa_r(\hat{\theta}) = n^{1-r} \kappa_r s_n(w^r),$$

where $w^r(t) = w(t)^r$ and

$$s_n(g) = n^{-1} \sum_{i=1}^n g(i/n) = \sum_{k=0}^{\infty} n^{-k} \alpha_{1k}(g)$$

by the Euler-McLaurin expansion for $\alpha_{1k}(g)$ of (3.30) in the appendix. So, the basic cumulant expansion (2.8) holds with coefficients $a_{ri} = \kappa_r \beta_{ri}$, $a_{r,r-1} = \kappa_r \beta_{r,r-1}$, $a_{rr} = \kappa_r \beta_{r,r}$, $a_{r,r+1} = \kappa_r \beta_{r,r+1}$ and $a_{r,r+2} = 0, \dots$, where $\beta_{ri} = \alpha_{1,i-r+1}(w^r)$, $\beta_{r,r-1} = \int w^r$, $\beta_{r,r} = \{w(1)^r - w(0)^r\}/2$, $\beta_{r,r+1} = r\{w(1)^{r-1}w_1(1) - w(0)^{r-1}w_1(0)\}/12$ and $w_{.r}(t)$ is the r th derivative of $w(t)$. So, Y_n of (2.9) can be written

$$Y_n = n^{1/2} \left(\int w dM_n - \mu w_1 \right) \sigma^{-1} w_2^{-1},$$

and the standardised coefficients (2.10) are given by $A_{ri} = \lambda_r \gamma_{ri}$, where

$$\lambda_r = \kappa_r / \sigma^r, \quad \gamma_{ri} = \beta_{ri} w_2^{-r}. \quad (3.23)$$

Although γ_{ri} and β_{ri} are known, λ_r and κ_r are generally unknown. Let $\hat{\lambda}_r$ be a suitably regular estimate satisfying $\hat{\lambda}_r = \lambda_r + O_p(n^{-1/2})$, such as the empirical estimate. Let us consider how the various probability statements of Section 2 can be written as *confidence intervals* for μ . We first suppose that the variance κ_2 is known, and then the contrary.

3.1 Case 1: $w_1 = \int w \neq 0$

Without loss of generality let us assume that $w_1 > 0$ and that $d_{nw} = w_1 + n^{-1}\{w(1) - w(0)\}/2 > 0$, where w_1, d_{nw} appear as divisors. Set

$$L_{1,x} = \{\hat{\theta} - n^{-1/2} w_2 \sigma x\} / w_1, \quad (3.24)$$

$$L_{2,x}(\kappa_3) = \{\hat{\theta} - n^{-1/2} w_2 \sigma x - \kappa_3 \left(\int w^3 \right) (n \kappa_2 \int w^2)^{-1} H_{2x} / 6\} / d_{nw}, \quad (3.25)$$

$$L_{3,x}(\kappa_4) = L_{2,x}(\kappa_3) - n^{-3/2} w_2 \sigma g_2(x) / d_{nw}. \quad (3.26)$$

3.1.1 Case 1.1: κ_2 Known, κ_3 Unknown

Note that (2.11) at $R = 1$, (2.12), (2.15), (2.18)–(2.20) at $R = 2$ give

$$\begin{aligned} & [L_{1,x} \leq \mu] \text{ and } [L_{1,-x} \geq \mu] \text{ are first order,} \\ & [L_{1,x} \leq \mu \leq L_{1,-x}], [L_{2,x}(\hat{\kappa}_3) \leq \mu], [L_{2,-x}(\hat{\kappa}_3) \geq \mu], \\ & \text{and } [L_{2,x}(\hat{\kappa}_3) \leq \mu \leq L_{2,-x}(\hat{\kappa}_3)] \text{ are second order.} \end{aligned} \quad (3.27)$$

3.1.2 Case 1.2: κ_2, κ_3 Known, κ_4 Unknown

Let $\hat{g}_2(x)$ denote $g_2(x)$ with κ_4 replaced by $\hat{\kappa}_4$. Note that (2.11) at $R = 2$, (2.13), (2.16) and (2.18)–(2.20) at $R = 3$ give

$$\begin{aligned} & [L_{2,x}(\kappa_3) \leq \mu], [L_{2,-x}(\kappa_3) \geq \mu], [L_{2,x}(\kappa_3) \leq \mu \leq L_{2,x}(\kappa_3)] \text{ are second order,} \\ & [L_{3,x}(\hat{\kappa}_4) \leq \mu], [L_{3,-x}(\hat{\kappa}_4) \geq \mu], \text{ are third order,} \\ & [L_{3,x}(\hat{\kappa}_4) \leq \mu \leq L_{3,-x}(\hat{\kappa}_4)] \text{ is fourth order.} \end{aligned} \quad (3.28)$$

Now let us consider briefly some of the *tests* generated by these confidence intervals. If κ_2 is known but not κ_3 , a second order one-sided test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ is to reject H_0 if $[L_{2,-x}(\hat{\kappa}_3) \geq \mu_0]$, that is (2.19) at $R = 2$.

For κ_2 known and $\kappa_3 = 0$ a fourth order one-sided test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ is to accept H_0 if $[L_{3,x}(\hat{\kappa}_4) \leq \mu_0 \leq L_{3,-x}(\hat{\kappa}_4)]$, where

$$L_{3x}(\kappa_4) = \{\hat{\theta} - n^{-1/2}w_2\sigma(x + n^{-1}g_2(x))\}/d_{nw}$$

and

$$g_2(x) = A_{22}x/2 + A_{43}H_{3x}/24 = \{w(1)^2 - w(0)^2\}w_2^{-2}x/4 + \kappa_4\kappa_2^{-2}w_4^4w_2^{-4}H_{3x}/24.$$

3.2 Case 2: $\int w = 0$

In this case the behaviour of Y_n can be used either for a test statistic (see below) or to make inference on σ . The equations for Case 1.1 now give for $x > 0$

$$\begin{aligned} [n^{1/2}\hat{\theta}w_2^{-1}x^{-1} \leq \sigma] & \text{ is first order,} \\ [-n^{1/2}\hat{\theta}w_2^{-1}x^{-1} \leq \sigma] & \text{ is first order,} \\ [n^{1/2}|\hat{\theta}|w_2^{-1}x^{-1} \leq \sigma] & \text{ is second order,} \\ [n^{1/2}\hat{\theta}w_2^{-1}/\{x + n^{-1/2}\hat{g}_1(x)\} \leq \sigma] & \text{ is second order,} \\ [-n^{1/2}\hat{\theta}w_2^{-1}/\{x - n^{-1/2}\hat{g}_1(x)\} \leq \sigma] & \text{ is second order,} \\ [-x + n^{-1/2}\hat{g}_1(x) \leq n^{1/2}\hat{\theta}w_2^{-1}\sigma^{-1} \leq x + n^{-1/2}\hat{g}_1(x)] & \text{ is second order.} \end{aligned}$$

The last equation can be written as a two-sided confidence interval for σ .

The other equations, such as those for Case 1.2, can be similarly restated.

3.3 Case 3: κ_2 Unknown

Let us denote the Studentized forms of $L_{r,x}$ of (3.24)–(3.26) above by $L_{r,x,0}$, that is with $\sigma^2 = \kappa_2$ replaced by its empirical or its unbiased estimate. Then one can use the Studentized forms of (3.27) and (3.28). The Studentised form of $L_{1,x}(\sigma) = L_{1,x}$ is simply $L_{1,x,0} = L_{1,x}(\hat{\sigma})$. By (2.21) that for $L_{2,x}(\kappa_3)$ is

$$L_{2,x,0}(\kappa_3) = \{\hat{\theta} - n^{-1/2}w_2\hat{\sigma}x - \kappa_3(n\hat{\kappa}_2)^{-1}[(\int w^3)(\int w^2)^{-1}H_{2x}/6 - w_1/2]\}/d_{nw},$$

as γ_{10} of (2.22) reduces to γ_{10} of (3.23). So, we have

3.3.1 Case 3.1: κ_2, κ_3 Unknown

We have

$$\begin{aligned} [L_{1,x,0} \leq \mu] \text{ and } [L_{1,-x,0} \geq \mu] & \text{ are first order,} \\ [L_{1,x,0} \leq \mu \leq L_{1,-x,0}], [L_{2,x,0}(\hat{\kappa}_3) \leq \mu], [L_{2,-x,0}(\hat{\kappa}_3) \geq \mu], \\ \text{and } [L_{2,x,0}(\hat{\kappa}_3) \leq \mu \leq L_{2,-x,0}(\hat{\kappa}_3)] & \text{ are second order.} \end{aligned}$$

3.3.2 Case 3.2: κ_3 Known, κ_2 and κ_4 Unknown

Let $\hat{g}_{20}(x)$ denote $g_{20}(x)$ with κ_4 replaced by $\hat{\kappa}_4$. The Studentized forms of (2.11) at $R = 2$, (2.13), (2.16) and (2.18)–(2.20) at $R = 3$ give

$$\begin{aligned} [L_{2,x,0}(\kappa_3) \leq \mu], [L_{2,-x,0}(\kappa_3) \geq \mu], [L_{2,x,0}(\kappa_3) \leq \mu \leq L_{2,x,0}(\kappa_3)] &\text{ are second order,} \\ [L_{3,x,0}(\hat{\kappa}_4) \leq \mu], [L_{3,-x,0}(\hat{\kappa}_4) \geq \mu], &\text{ are third order,} \\ [L_{3,x,0}(\hat{\kappa}_4) \leq \mu \leq L_{3,-x,0}(\hat{\kappa}_4)] &\text{ is fourth order.} \end{aligned}$$

Here $L_{3,x,0}(\hat{\kappa}_4) \leq \mu$ is a re-arrangement of $Y_{n0} \leq x + n^{-1/2}g_{10}(x) + n^{-1}\hat{g}_{20}(x)$, where $\hat{g}_{20}(x) = \hat{a}_{220}x/2 + \hat{a}_{430}H_{3x}/24 - A_{32}^2(2x^3 - 5x)/36$. However, we shall not give a_{220} and a_{430} here.

The obvious application of Case 3.2 is to symmetrically distributed observations.

Note 3.1 *If we replace $w(t)$ by $w'(t) = w(1 - t)$, then in L_{rx} only $\hat{\theta}$ and d_{nw} change: $d_{nw'} = w_1 + n^{-1}\{w(0) - w(1)\}/2$.*

The usual strategy will be to weight recent observations more heavily, as in the following

Example 3.1 *Take $w(t) = t$. Set $I(A) = 1$ or 0 for A true or false. Then*

$$\begin{aligned} \hat{\theta} &= n^{-2} \sum_{i=1}^n iX_i, \\ n\hat{\theta} - (n+1)\mu/2 &= n^{-1} \sum_{i=1}^n i(X_i - \mu), \\ \theta &= a_{10} = \mu/2, a_{21} = \kappa_2/3, \beta_{r,r-1} = (r+1)^{-1}, \beta_{r,r} = -2^{-1}, \\ A_{r,r-1} &= \lambda_r 3^{r/2}/(r+1), A_{r,r} = -\lambda_r 3^{r/2}/2, \\ A_{r,r+1} &= -r\lambda_r I(r > 1) 3^{r/2}/12, A_{r,r+2} = 0, \\ Y_n &= \sigma^{-1}(3n)^{1/2}(\hat{\theta} - \mu/2), \\ h_1(x) &= f_1(x) = g_1(x) = 3^{1/2}\{-\lambda_1/2 + \lambda_3 H_{2x}/8\}, \\ g_2(x) &= 3\{-x + \lambda_4 H_{3x}/10 - \lambda_3^2(2x^3 - 5x)/16\}/4, \\ L_{1,x} &= 2\hat{\theta} - 2(3n)^{-1/2}\sigma x, \\ L_{2,x}(\kappa_3) &= \{\hat{\theta} - (3n)^{-1/2}\sigma x - n^{-1}\sigma^{-2}\kappa_3 H_{2x}/8\}/d_{nw}, \\ L_{3,x}(\kappa_4) &= L_{2,x}(\kappa_3) - n^{-3/2}3^{-1/2}\sigma g_2(x)/d_{nw}, \\ L_{2,x,0}(\kappa_3) &= \{\hat{\theta} - (3n)^{-1/2}\hat{\sigma}x - (n\hat{\kappa}_2)^{-1}\kappa_3(x^2 - 3)/8\}/d_{nw}, \\ d_{nw} &= (1 + n^{-1})/2. \end{aligned}$$

Example 3.2 *Take $w(t) = 1 - t$. Then*

$$\hat{\theta} = n^{-2} \sum_{k=1}^{n-1} S_k = n^{-1} \sum_{i=1}^{n-1} (1 - i/n)X_i.$$

So,

$$n\hat{\theta} - (n-1)\mu/2 = n^{-1} \sum_{k=1}^{n-1} (S_k - k\mu)$$

and

$$\hat{\theta} - \mu/2 = \int_0^1 (M_n(t) - \mu t) dt$$

are one-sided L_1 versions of the two-sided L_∞ statistic A_n of (1.2) for a sample of size $n - 1$. (This shows up a weakness in these statistics: one would generally prefer to give recent observations more weight rather than less weight.) Note that $\beta_{r,r-1}$, $a_{r,r-1}$, $A_{r,r-1}$, $A_{r,r+2}$, Y_n and $L_{r,x}$ are all as given in the previous example but with $d_{nw} = (1 - n^{-1})/2$ while $\beta_{r,r}$, $A_{r,r}$, $A_{r,r+1}$, λ_1 in $h_1(x) = f_1(x) = g_1(x)$ and $-x$ in $g_2(x)$ all change sign.

The statistic $\int w dM_n$ arises naturally in change point problems. Consider the one-parameter exponential family $f_\theta(x) = \exp\{a(\theta)T(x) + b(x) - c(\theta)\}$. Suppose we observe Y_1, \dots, Y_n independent with $Y_i \sim f_\theta$ for $1 \leq i \leq k$ and $Y_i \sim f_{\theta+\delta}$ for $k < i \leq n$, where k , θ , δ are unknown. Suppose we assume that $k = i$ with probability $p_{in} \propto p(i/n)$, $i = 2, \dots, n$, where $\int p = 1$. Then the likelihood ratio of $H_0 : \delta = 0$ versus $H_1 : \delta > 0$ is $\{1 + A_n\delta + o(\delta)\} \exp\{na(\theta)S - nc(\theta)\}$ as $\delta \rightarrow 0$, where

$$S = \sum_{i=1}^n X_i/n, \quad X_i = T(Y_i),$$

$$A_n = \sum_{k=1}^n p_{kn} \sum_{i=k+1}^n \{\dot{a}(\theta)X_i - \dot{c}(\theta)\} = \sum_{i=2}^n \{\dot{a}(\theta)X_i - \dot{c}(\theta)\} q_{in}$$

for

$$q_{in} = \sum_{k=1}^{i-1} p_{kn} \approx n w_p(i/n),$$

$$w_p(t) = \int_0^t p \approx n^2 \dot{a}(\theta) \left(\int w_p dM_n - \mu \int w_p \right)$$

for $\mu = EX_1$. For uniform prior $p(t) = 1$, this gives $w_p(t) = t$, a result due to Kander and Zacks (1966) for the case $a(\theta) = \theta$, and to Chernoff and Zacks (1964) for the normal case. Kander and Zacks (1966) gave the Edgeworth expansion to $O(n^{-3/2})$. For related references and the full likelihood ratio test see Sections 1.8 and 1.5 of Csorgo and Horvath (1997). They also consider the ‘‘epidemic alternative’’ $H_2 : Y_i \sim f_{\theta+\delta}$ for $k_1 < i \leq k_2$ and $Y_i \sim f_\theta$ otherwise, where $k_1 < k_2$ are unknown change points. If one takes a uniform prior on (k_1, k_2) one obtains in the same way the statistic $\int w_u dM_n$, where $w_u(t) = t - t^2$.

If one only wishes to construct a test of $\delta = 0$ (rather than a confidence interval for μ) one can replace μ by S above in the approximation to A_n , giving the statistic $\int w dM_n$, where $w(t) = w_p(t) - \int w_p$, or in the case of the epidemic alternative, $w(t) = w_u(t) - \int w_u$. In either case one has $\int w = 0$. This approach was advocated by Ramanayake (2004) for the case of a gamma distribution with known scale parameter, uniform prior and alternative H_1 .

Note 3.2 *Lai (1974) considers statistics which are moving averages of the last k observations: $Y_n = \sum_{i=n-k+1}^n c_{n-i} X_i$. If one takes $k = n(1 - t_0)$, where $0 < t_0 < 1$ and $c_{n-i}/n = w(i/n)$ then $Y_n/n^2 = \int w_0 dM_n$, where $w_0(t) = w(t)I(t_0 < t)$. His condition that c_k are non-increasing is achieved if $w(t)$ is non-decreasing. He also considers exponentially weighted moving average schemes. These correspond to choosing $w(t) = p^{1-t}$.*

We now consider the functional $\int gdw$ firstly for w continuous, and then for w consisting of atoms.

Example 3.3 *If w is continuous it is easy to check that*

$$\int M_n dw = n^{-1} \sum_{i=1}^n X_i \tilde{w}(i/n) = \int \tilde{w} dM_n,$$

where $\tilde{w}(t) = w(1) - w(t)$. So, we can apply our previous results with w replaced by \tilde{w} . This gives

$$\begin{aligned} \beta_{r,r-1} &= \tilde{w}_r^r = \int \{w(1) - w(t)\}^r dt, \quad \beta_{r,r} = -\{w(1) - w(0)\}^r / 2, \\ \beta_{r,r+1} &= r[\{w(1) - w(0)\}^{r-1} w_{\cdot 1}(0) - \delta_{r1} w_{\cdot 1}(1)] / 12, \\ d_{n\tilde{w}} &= w(1) - w_1 - n^{-1} \{w(1) - w(0)\} / 2. \end{aligned}$$

Finally we show how to deal with $\int gdw$ for discrete w .

Example 3.4 *Fix m points in $[0, 1]$, say $0 \leq t_1 \leq \dots \leq t_m \leq 1$. Suppose that $T(g) = m^{-1} \sum_{i=1}^m g(t_i)$. That is, $T(g) = \int gdw$, where w puts weight m^{-1} at t_i for $i = 1, \dots, m$. Set $t_0 = 0, U_i = [nt_i], u_i = U_i/n$. Then*

$$\hat{\theta} = n^{-1} \sum_{i=0}^{m-1} (S_{U_{i+1}} - S_{U_i})$$

and $\kappa_r(\hat{\theta}) = \kappa_r \beta_{rn} n^{1-r}$ for

$$\beta_{rn} = \sum_{i=0}^{m-1} (1 - i/m)^r (u_{i+1} - u_i).$$

So, (2.8) holds with $a_{r,r-1} = \kappa_r \beta_{rn}$ and the other $a_{ri} = 0$. In terms of the standardised cumulants of (3.23), this gives $A_{r,r-1} = \lambda_r \gamma_{rn}$, where $\gamma_{rn} = \beta_{rn} / \beta_{2n}^{r/2}$, and the other $A_{ri} = 0$. Alternatively for

$$\ell_n(t) = [nt] - nt, \tag{3.29}$$

$u_i = t_i + n^{-1} \ell_n(t_i)$ so that $\kappa_r(\hat{\theta}) = a_{r,r-1} n^{1-r} + a_{r,r-1} n^{-r}$ for

$$a_{r,r-1} = \kappa_r \sum_{i=0}^{m-1} (1 - i/m)^r (t_{i+1} - t_i)$$

and

$$a_{r,r} = \kappa_r \sum_{i=0}^{m-1} (1 - i/m)^r \{\ell_n(t_{i+1}) - \ell_n(t_i)\}.$$

Acknowledgments

The authors would like to thank the Editor and the referee for carefully reading the paper and for their comments which greatly improved the paper.

References

- [1] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*. U.S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series **55**.
- [2] Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain ‘goodness of fit’ criteria based on stochastic processes. *Annals of Mathematical Statistics*, **23**, 193–212.
- [3] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- [4] Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, **35**, 999–1018.
- [5] Csorgo, M. and Horvath, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester, England.
- [6] Kander, Z. and Zacks, S. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Annals of Mathematical Statistics*, **37**, 1196–1210.
- [7] Lai, T. L. (1974). Control charts based on weighted sums. *Annals of Statistics*, **2**, 134–147.
- [8] Lai, T. L. (1995). Sequential change point detection in quality control and dynamical systems. *Journal of the Royal Statistical Society, B*, **57**, 613–658.
- [9] Ramanayake, A. (2004). Tests for a change point in the shape parameter of gamma random variables. *Communications in Statistics—Theory and Methods*, **33**, 821–833.
- [10] Sparks, R. S. (2000). CUSUM charts for AR1 data: are they worth the effort? *Australian and New Zealand Journal of Statistics*, **42**, 25–42.
- [11] Withers, C. S. (1982). Second order inference for asymptotically normal random variables. *Sankhyā, B*, **44**, 1–9.
- [12] Withers, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Annals of Statistics*, **11**, 577–587.
- [13] Withers, C. S. (1984). Asymptotic expansions for distributions and quantiles with power series cumulants. *Journal of the Royal Statistical Society, B*, **46**, 389–396.
- [14] Withers, C. S. (1988). Nonparametric confidence intervals for functions of several distributions. *Annals of the Institute of Statistical Mathematics*, **40**, 727–746.
- [15] Withers, C. S. (2000). A simple expression for the multivariate Hermite polynomial. *Statistics and Probability Letters*, **47**, 165–169.

Appendix

Here we give the Euler-McLaurin expansion (Abramowitz and Stegun, 1964, equation (23.1.30), page 806), and related results. For $g : [0, 1]^r \rightarrow R$, set

$$(g)_{rn} = n^{-r} \sum_{i_1=1}^n \cdots \sum_{i_r=1}^n g(i_1/n, \dots, i_r/n).$$

Suppose that g has finite derivatives. Then for $r = 1$ we have the expansion

$$(g)_{1n} = \sum_{k=0}^{\infty} \alpha_{1k}(g) n^{-k}, \quad (3.30)$$

where $\alpha_{10}(g) = \int_0^1 g(t) dt$, $\alpha_{1k}(g) = \{g^{(k-1)}(1) - g^{(k-1)}(0)\} e_k B_k / k!$ for $k = 1, 2, 3, \dots$, $e_1 = -1$, $e_k = 1$ for $k = 2, 3, \dots$ and B_k is the k th Bernoulli number, given by the left column on page 809 of Abramowitz and Stegun (1964): $B_1 = -1/2$, $B_2 = 1/6$, $B_3 = 0$, $B_4 = -1/30$, \dots and $B_k = 0$ for $k = 3, 5, 7, \dots$. So, $\alpha_{11}(g) = \{g(1) - g(0)\} / 2$ and $\alpha_{1k}(g) = 0$ for $k = 3, 5, 7, \dots$. Note that (3.30) implies that for $\ell_n(t)$ of (3.29),

$$\int_0^1 g(t) d\ell_n(t) = \sum_{k=0}^{\infty} \alpha_{1,k+1}(g) n^{-k}.$$

Also from (3.30) it follows that

$$(g)_{rn} = \sum_{k=0}^{\infty} \alpha_{rk}(g) n^{-k},$$

where

$$\begin{aligned} \alpha_{r0}(g) &= \int_0^1 \cdots \int_0^1 g(t_1, \dots, t_r) dt_1 \cdots dt_r, \\ \alpha_{r1}(g) &= \sum_{i=1}^r \{g_i(1) - g_i(0)\} / 2, \\ g_i(t_i) &= \left(\int_0^1 \right)^{r-1} g(t_1, \dots, t_r) dt_1 \cdots dt_{i-1} dt_{i+1} \cdots dt_r, \\ (g)_{rn} &= \prod_{i=1}^r \left(\sum_{k=0}^{\infty} n^{-k} \beta_{ik} \right) g(t_1, \dots, t_r), \end{aligned}$$

where the operator β_{ik} is defined by $\beta_{ik} g(t_1, \dots, t_r) = \alpha_{1k}(h)$ for $h(t_i) = g(t_1, \dots, t_r)$. For example,

$$\beta_{i0} g(t_1, \dots, t_r) = \int_0^1 g(t_1, \dots, t_r) dt_i.$$

So,

$$\alpha_{rk}(g) = \sum \{ \beta_{1k_1} \cdots \beta_{rk_r} g(t_1, \dots, t_r) : k_1 + \cdots + k_r = k, k_i \geq 0 \}.$$