



Accurate tests and intervals based on non-linear cusum statistics

Christopher S. Withers, Saralees Nadarajah

► To cite this version:

Christopher S. Withers, Saralees Nadarajah. Accurate tests and intervals based on non-linear cusum statistics. *Statistics and Probability Letters*, 2009, 79 (21), pp.2242. 10.1016/j.spl.2009.07.023 . hal-00582595

HAL Id: hal-00582595

<https://hal.science/hal-00582595>

Submitted on 2 Apr 2011

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 non-linear cusum statistics

Christopher S. Withers, Saralees Nadarajah

PII: S0167-7152(09)00285-5
DOI: [10.1016/j.spl.2009.07.023](https://doi.org/10.1016/j.spl.2009.07.023)
Reference: STAPRO 5483

To appear in: *Statistics and Probability Letters*

Received date: 20 October 2008

Revised date: 27 July 2009

Accepted date: 27 July 2009

Please cite this article as: Withers, C.S., Nadarajah, S., Accurate tests and intervals based on non-linear cusum statistics. *Statistics and Probability Letters* (2009), doi:10.1016/j.spl.2009.07.023

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 Non-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. Withers and Nadarajah [*Statistics and Probability Letters*, **79**, 2009, 689–697] showed how to improve the order of these approximations for linear statistics. Here, an extension is provided for non-linear statistics with non-zero asymptotic variances. This involves development of two calculi analogous to that of von Mises.

Keywords: Approximations; Cusums; Edgeworth-Cornish-Fisher; Non-linear statistics.

1 Introduction and Summary

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. Withers and Nadarajah (2009) showed how to improve on the order of these approximations for asymptotically normal linear statistics. For smooth *non-linear* statistics with non-zero asymptotic covariance, asymptotic normality still holds and *I*th order approximations to their distributions and quantiles can be obtained by expanding their cumulants as power series in n^{-1} then applying Edgeworth-Cornish-Fisher technology. To do this two calculi are developed analogous but simpler to that of von Mises. In the case, where the mean is not changing, the corresponding tests can be viewed as one- and two-sided confidence intervals for the mean.

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 $S_0 = 0$, $S_i = \sum_{j=1}^i X_j$ 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

$$m_n(t) = E M_n(t) = \mu[nt]/n \rightarrow m(t) = \mu t \quad (1.2)$$

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 . We shall assume that $\hat{\theta}$ satisfies the standard cumulant expansion given by

$$\kappa_r(\hat{\theta}) = \sum_{i=r-1}^{\infty} a_{ri} n^{-i} \quad (1.3)$$

for $r = 1, 2, \dots$.

For univariate data the most common prospective and retrospective two-sided cusum statistics and functionals are

$$A_n = A(M_n) = \max_{k=1}^n |S_k - k\mu|/n, \quad (1.4)$$

$$B_n = B(M_n) = \max_{k=1}^n |S_k - k\bar{X}_n|/n \quad (1.5)$$

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 $n \rightarrow \infty$

$$\begin{aligned} \hat{\sigma}^{-1} n^{1/2} \{M_n(t) - t\mu\} &\xrightarrow{\mathcal{L}} W(t), \\ \hat{\sigma}^{-1} n^{1/2} \{M_n(t) - tM_n(1)\} &\xrightarrow{\mathcal{L}} W_0(t) \end{aligned}$$

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).

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 $\int_0^1 g(t) dt$. This is an example of the statistics and functionals we consider here. They include $T(M_n)$ for

$$T(g) = \int_0^1 w(g(t), t) dt \text{ or } \int_0^1 w(g(t), t) dg(t) \quad (1.8)$$

for $w(x, t) : \mathcal{R}^p \times [0, 1] \rightarrow \mathcal{R}^q$, as well as smooth functions of such functionals. 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 (1.4)–(1.7), are not available.

The aim of this note is to extend the work of Withers and Nadarajah (2009) for *non-linear* functionals. In this case, the cumulant expansions in (1.3) and hence the Edgeworth-Cornish-Fisher expansions (Section 2, Withers and Nadarajah, 2009) are obtained in terms of *functional derivatives* of T at m using the Euler-McLaurin expansion. Section 2 defines the functional derivatives needed. There are two types depending on whether integration is with respect to dt or $dg(t)$. These are analogous to those of von Mises (1947) as extended in Withers (1983, 1988). We show that the chain rule applies to such derivatives. Examples include (1.8) and smooth functions of such functionals. Section 3 gives the cumulants of $T(M_n)$ needed to obtain its Edgeworth-Cornish-Fisher expansions (Section 2, Withers and Nadarajah, 2009). For example, as $n \rightarrow \infty$,

$$n \operatorname{var}(T(M_n)) \rightarrow \sigma^2 \int T_m(t)^2 dt$$

for $m(t)$ of (1.2) and $T_g(t)$ the first derivative of $T(g)$ in the calculus of Section 2.

2 Functional Derivatives

Since $M_n(t)$ is not a distribution function, the calculus of von Mises (1947) does not apply to functionals of it. In this section we develop two differential calculi which can be applied to such functionals. The first calculus is for functionals like $\int_0^1 w(g(t), t) dt$ when $g(0) = 0$ and for $f(\int w dg)$ without the condition $g(0) = 0$. The second calculus covers functionals like $\int_0^1 w(g(t), t) dt$ without the condition $g(0) = 0$.

2.1 First Calculus

Consider functions $g, h : I \rightarrow \mathcal{R}$, where $I \subset \mathcal{R}^p$, and a functional $T(g)$ in \mathcal{R} . Suppose that for real ϵ ,

$$T(g + \epsilon h) = T(g) + \sum_{r=1}^{\infty} T_r(g, h) \epsilon^r / r!, \quad (2.1)$$

where

$$T_r(g, h) = \int_I \cdots \int_I T_g(t_1, \dots, t_r) dh(t_1) \dots dh(t_r)$$

is assumed to exist. We call $T_g(t_1, \dots, t_r)$ the *rth order (functional) derivative* of $T(g)$ with respect to the first calculus, or simply the first *rth order derivative*. It is made unique by the condition

$$T_g(t_1, \dots, t_r) \text{ is symmetric in } t_1, \dots, t_r. \quad (2.2)$$

Restrictions on h may be required in order that (2.1) hold, as the examples show. That is, for a given T the expansion for $T(g_n) - T(g)$ may place restrictions on eligible $g_n - g$.

Note 2.1 Condition (2.1) can be weakened in two ways. Firstly, it need only hold in a neighbourhood of zero, say $|\epsilon| < \delta$. This is satisfied by $\epsilon = n^{-1/2}$ if $n > \delta^{-2}$. One can then take, for example, $g(t) = m(t)$ and $h(t) = n^{1/2}(M_n(t) - m(t))$, so that the left hand side of

(2.1) is $T(M_n)$. Secondly, the infinite Taylor series (2.1) can be replaced by a partial series with a remainder term, analogous to the usual ordinary Taylor series for a real function of a real variable, given in Section 3.6.1 of Abramowitz and Stegun (1964).

Like von Mises' calculus, these calculi cannot be applied to sup functionals like that of (1.4) and (1.5).

Set

$$\Delta_t(s) = I(t \leq s) = \prod_{i=1}^p I(t_i \leq s_i). \quad (2.3)$$

Taking $\lambda_1, \dots, \lambda_r$ in \mathcal{R} and $h(s) = \sum_{i=1}^r \lambda_i \Delta_{t_i}(s)$, where $\Delta_{t_i}(\cdot)$ is defined by (2.3), we see that $T_g(t_1, \dots, t_r)$ is the coefficient of $\lambda_1 \cdots \lambda_r$ in $T_r(g, h)/r!$, so that $T_g(t_1, \dots, t_r)$ is the coefficient of $\lambda_1 \cdots \lambda_r \epsilon^r$ in $T(g + \epsilon h)/r!$. So, the first derivative is

$$S(g) = T_g(t) = \lim_{\epsilon \rightarrow 0} \{T(g + \epsilon \Delta_t) - T(g)\} / \epsilon, \quad (2.4)$$

where Δ_t is defined by (2.3). Similarly, by the method of proof of Theorem 2.1 of Withers (1983), it follows that

$$T_g(t_1, \dots, t_{r+1}) = S_g(t_{r+1}) \text{ for } S(g) = T_g(t_1, \dots, t_r). \quad (2.5)$$

That is, the derivative of the r th derivative is the $r + 1$ st derivative, just as for ordinary partial derivatives. The situation is analogous to, but simpler to the theory of derivatives of functionals of distribution functions developed by Von Mises (1947) and refined by Withers (1983). Note that in general $\int_0^1 T_g(t_1, \dots, t_r) dt_1 \neq 0$, whereas for the von Mises derivative $\int T_F(x_1, \dots, x_r) dF(x_1) = 0$: see (2.3) of Withers (1983).

Example 2.1 Suppose that $p = 1$ and $T(g) = \int_0^1 w(t) dg(t)$, where $w(t) : [0, 1] \rightarrow \mathcal{R}$ and $g : [0, 1] \rightarrow \mathcal{R}$ are chosen such that $T(g)$ exists. Set $w_r = (\int w^r)^{1/r}$ for $0 < r < \infty$. Then $T_g(t) = w(t)$ and higher derivatives are zero. So, $\int T_m dm = \mu w_1 \neq 0$ in general.

Example 2.2 Suppose that $p = 1$ and $T(g) = \int_0^1 W(t)g(t)dt$. Then one needs to impose the condition $h(0) = 0$ giving $T_g(t) = \int_t^1 W(s)ds$ or the condition $h(1) = 0$ giving $T_g(t) = \int_0^t W(s)ds$. The higher derivatives are zero. The condition $h(0) = 0$ restricts application to $T(g_n) - T(g)$ with $g_n(0) - g(0) = 0$, a condition satisfied by $(g_n, g) = (M_n, m)$ of (1.1), (1.2). The condition $h(1) = 0$ restricts application to $T(g_n) - T(g)$ with $g_n(1) - g(1) = 0$, a condition not satisfied by $(g_n, g) = (M_n, m)$.

For the next examples it is helpful to use \wedge and \vee to denote min and max:

$$t_{r\wedge} = \min_{i=1}^r t_i, \quad t_{r\vee} = \max_{i=1}^r t_i. \quad (2.6)$$

If $p > 1$ then min and max are interpreted element-wise.

Example 2.3 Take $p = 1$. Let s be any point in $[0, 1]$. Consider functions $g : [0, 1] \rightarrow \mathcal{R}$, and a smooth function $w : \mathcal{R} \rightarrow \mathcal{R}$. Denote its r th derivative by $w_{,r}$. If $h(0) = 0$ and $T(g) = w(g(s))$ then

$$T_g(t_1, \dots, t_r) = w_{,r}(g(s)) \Delta_{t_1}(s) \cdots \Delta_{t_r}(s) = w_{,r}(g(s)) \Delta_{t_{r\vee}}(s),$$

where $\Delta_{t_1}(\cdot)$ is defined by (2.3). If $h(s+) = h(s)$ and $T(g) = w(g(s) - g(1))$ then

$$\begin{aligned} T_g(t_1, \dots, t_r) &= w_{.r}(g(s) - g(1))(\Delta_{t_1}(s) - 1) \cdots (\Delta_{t_r}(s) - 1) \\ &= (-1)^r w_{.r}(g(s) - g(1))I(s < t_{r\wedge}). \end{aligned}$$

Example 2.4 Let ν be a measure on some space L . For s in L let $T_s(g)$ be a functional with finite r th derivative $T_{sg}(t_1, \dots, t_r)$. For $S(g) = \int T_s(g) d\nu(s)$,

$$S_g(t_1, \dots, t_r) = \int T_{sg}(t_1, \dots, t_r) d\nu(s).$$

Putting these two examples together, we have

Example 2.5 Take $p = 1$. Let $w(x, t) : \mathcal{R} \times [0, 1] \rightarrow \mathcal{R}$ have finite partial derivatives $w_{.r0}(x, t) = (\partial/\partial x)^r w(x, t)$. If $g(0) = 0$ and $S(g) = \int_0^1 w(g(s), s) ds$ then

$$S_g(t_1, \dots, t_r) = \int_{t_{r\vee}}^1 w_{.r0}(g(s), s) ds. \quad (2.7)$$

If g is right-continuous (for example, $g = M_n - m$) and

$$S(g) = \int_0^1 w(g(s) - g(1), s) ds \quad (2.8)$$

then

$$S_g(t_1, \dots, t_r) = (-1)^r \int_0^{t_{r\wedge}} w_{.r0}(g(s) - g(1), s) ds. \quad (2.9)$$

For example, alternatives to the statistics A_n and B_n of (1.4) and (1.5) are

$$\begin{aligned} C_n &= C_n(M_n) = n^{-3} \sum_{j=1}^n (S_j - j\mu)^2 w(j/n)/2, \\ D_n &= D_n(M_n) = n^{-3} \sum_{j=1}^n (S_j - jS_n/n)^2 w(j/n)/2 \end{aligned}$$

for

$$\begin{aligned} C_n(g) &= \int_0^1 (g(s) - \mu[ns]/n)^2 w([ns]/n) ds/2, \\ D_n(g) &= \int_0^1 (g(s) - g(1)[ns]/n)^2 w([ns]/n) ds/2. \end{aligned}$$

Their non-zero derivatives are

$$\begin{aligned} C_{ng}(t_1) &= \int_{t_1}^1 (g(s) - \mu[ns]/n) w([ns]/n) ds, \\ C_{ng}(t_1, t_2) &= \int_{t_{m2}}^1 w([ns]/n) ds, \\ D_{ng}(t_1) &= \int_0^1 (g(s) - g(1)[ns]/n) w([ns]/n) e_n(s, t_1) ds, \\ D_{ng}(t_1, t_2) &= \int_0^1 e_n(s, t_1) e_n(s, t_2) ds, \end{aligned}$$

where $e_n(s, t) = I(t \leq s) - [ns]/n$. Now replace $C_n(g), D_n(g)$ by their limits

$$C(g) = \int_0^1 (g(s) - \mu s)^2 w(s) ds / 2,$$

$$D(g) = \int_0^1 (g(s) - g(1)s)^2 w(s) ds / 2,$$

and set $C'_n = C(M_n), D'_n = D(M_n)$. If $w(s) = 1$ then one may show that $C_n - C'_n$ and $D_n - D'_n$ are $O_p(n^{-3/2})$ in probability.

Since $D(g)$ does not have the form (2.8), (2.9) does not apply. This can be fixed by applying the method not to $M_n(t)$ but to

$$N_n(t) = [nt]^{-1} \sum_{i=1}^{[nt]} X_i,$$

interpreted as 0 for $t \leq 1/n$. However, $\kappa(\hat{w}_1, \dots, \hat{w}_r)$ for $\hat{w}_i = N_n(t_i)$ is more complicated than for $\hat{w}_i = M_n(t_i)$, so we do not pursue this idea beyond the next example.

Example 2.6 Consider

$$D'_n = D'_n(N_n) = n^{-1} \sum_{j=1}^{n-1} (S_j/j - S_n/n)^2 w(j/n) / 2,$$

where

$$D'_n(g) = \int_{n^{-1}}^1 (g(s) - g(1))^2 w([ns]/n) ds / 2.$$

The non-zero derivatives are

$$D'_{ng}(t_1) = - \int_{n^{-1}}^1 (g(s) - g(1)) w([ns]/n) I(s \leq t) ds,$$

$$D'_{ng}(t_1, t_2) = \int_{n^{-1}}^1 w([ns]/n) I(s < t_{m2}) ds.$$

Note that $D_n[w] = D_n$ and $D'_n[w] = D'_n$ are equivalent in that $D'_n[w'] = D_n[w]$ for $w'(t) = t^2 w(t)$. Note that

$$2n^3 D_n[1] = 2n^3 D'_n[t^2] = \sum_{j=1}^n (S_j - j\bar{X}_n)^2$$

provides a simple alternative to

$$nB_n = \max_{j=1}^n |S_j - j\bar{X}_n|,$$

since they provide L_2 and L_∞ norms of $S_j - j\bar{X}_n$.

2.2 Chain Rule

Since $T_g(t_1, \dots, t_r)$ is an exact analogue of the ordinary r th order partial derivative, it satisfies the chain rule. We now spell this out in detail. By Comtet (1974, page 137), Faà di Bruno's chain rule for the derivatives of a function of a function is most simply stated in terms of the exponential Bell polynomials tabled on page 307 of Comtet (1974): for $r \geq 1$, the r th derivative of a real function of a real function, say $S(g) = f(T(g))$, for g in \mathcal{R} , $T : \mathcal{R} \rightarrow \mathcal{R}$, $f : \mathcal{R} \rightarrow \mathcal{R}$ is given by

$$S^{(r)}(g) = \sum_{k=1}^r f_k B_{rk}(\mathbf{T}) \quad (2.10)$$

at $f_k = f^{(k)}(T(g))$ and $\mathbf{T} = (T_1, T_2, \dots)$, where $T_r = T^{(r)}(g)$, the r th derivative of $T(g)$. This extends to g and $T(g)$ multi-dimensional by a simple reinterpretation. For example, for $r = 4$, (2.10) gives

$$S^{(4)}(g) = f_1 T_4 + f_2 (4T_1 T_3 + 3T_2^2) + f_3 (6T_1^2 T_2) + f_4 T_1^4.$$

For g in p dimensions and $T(g)$ in q dimensions this becomes

$$\begin{aligned} S_{.i_1 i_2 i_3 i_4}(g) &= f_{.j_1} T_{j_1 .i_1 i_2 i_3 i_4} + f_{.j_1 j_2} \left(\sum_i^4 T_{j_1 .i_1} T_{j_2 .i_2 i_3 i_4} + \sum_i^3 T_{j_1 .i_1 i_2} T_{j_2 .i_3 i_4} \right) \\ &\quad + f_{.j_1 j_2 j_3} \sum_i^6 T_{j_1 .i_1} T_{j_2 .i_2} T_{j_3 .i_3 i_4} + f_{.j_1 j_2 j_3 j_4} T_{j_1 .i_1} T_{j_2 .i_2} T_{j_3 .i_3} T_{j_4 .i_4}, \end{aligned} \quad (2.11)$$

where, for i_1, i_2, \dots in $1, 2, \dots, p$ $S_{.i_1 i_2 \dots}(g) = \partial_{i_1} \partial_{i_2} \dots S(g)$ for $\partial_i = \partial / \partial g_i$. Similarly, for the partial derivatives of $T_j(g)$ for $j = 1, \dots, q$; and for j_1, j_2, \dots in $1, 2, \dots, q$ $f_{.j_1 j_2 \dots}(T(g)) = \partial_{j_1} \partial_{j_2} \dots f(y)$ at $y = T(g)$ for $\partial_j = \partial / \partial y_j$. The convention is that j_1, j_2, \dots , the repeated pairs of suffixes in (2.11), are implicitly summed over their range $1, 2, \dots, q$. The expressions \sum_i^N in (2.11) mean summation over all N permutations of the i 's giving different terms. For example,

$$\sum_i^4 T_{j_1 .i_1} T_{j_2 .i_2 i_3 i_4} = T_{j_1 .i_1} T_{j_2 .i_2 i_3 i_4} + T_{j_1 .i_2} T_{j_2 .i_1 i_3 i_4} + T_{j_1 .i_3} T_{j_2 .i_1 i_2 i_4} + T_{j_1 .i_4} T_{j_2 .i_1 i_2 i_3}.$$

This chain rule in turn has a simple reinterpretation for functions of functionals. Consider the real functional $S(g) = f(T(g))$, where $T(g)$ is a q -dimensional functional on functions $g : I \rightarrow \mathcal{R}$, where $I \subset \mathcal{R}^p$, and $f(T)$ is an ordinary function from \mathcal{R}^q to \mathcal{R} . Then, for example, the fourth order functional derivatives of $S(g)$ are given in terms of the functional derivatives of $T(g)$ by (2.11) with $S_{.i_1 i_2 \dots}(g)$ replaced by $S_g(t_1 t_2 \dots)$ and $T_{j .i_1 i_2 \dots}$ replaced by $T_{jg}(t_1 t_2 \dots)$, the functional derivatives of $T_j(g)$:

$$\begin{aligned} S_g(t_1 t_2 t_3 t_4) &= f_{.j_1} T_{j_1 g}(t_1 t_2 t_3 t_4) + f_{.j_1 j_2} \sum_t^4 T_{j_1 g}(t_1) T_{j_2 g}(t_2 t_3 t_4) \\ &\quad + \sum_t^3 T_{j_1 g}(t_1 t_2) T_{j_2 g}(t_3 t_4) + f_{.j_1 j_2 j_3} \sum_t^6 T_{j_1 g}(t_1) T_{j_2 g}(t_2) T_{j_3 g}(t_3 t_4) \\ &\quad + f_{.j_1 j_2 j_3 j_4} T_{j_1 g}(t_1) T_{j_2 g}(t_2) T_{j_3 g}(t_3) T_{j_4 g}(t_4). \end{aligned}$$

2.3 Extension to Vector g

We now extend the results of this section to g, h, ϵ s -vectors. Let ϵh denote the s -vector with i th component $\epsilon_i h_i$ for $i = 1, \dots, s$. For r in \mathcal{N}^s , where $\mathcal{N} = \{0, 1, 2, \dots\}$, set

$$|r| = \sum_{i=1}^s r_i, \quad \epsilon^r / r! = \prod_{i=1}^s \epsilon^i / r_i!.$$

Now consider functionals $T(g)$ such that for all s -vector functions g, h on $I \subset \mathcal{R}^p$,

$$T(g + \epsilon h) = \sum_{r \in \mathcal{N}^s} T_r(g, h) \epsilon^r / r!$$

for

$$T_r(g, h) = \int_I^{|r|} T_g(\{u_{ij}, i = 1, \dots, s; j = 1, \dots, r_i\}) \prod_{i=1}^s \prod_{j=1}^{r_i} dh_i(u_{ij}),$$

where $\int_I^{|r|}$ denotes $|r|$ integrals over I . The r th derivative, the above integrand, is again made unique by the condition that it is symmetric in each of its s rows. Taking $h_i(u) = \sum_{j=1}^{r_i} \lambda_{ij} \Delta_{t_i}(u)$, where $\Delta_{t_i}(\cdot)$ is defined by (2.3), we see that this derivative is the coefficient of $\prod_{i=1}^s \prod_{j=1}^{r_i} \lambda_{ij}$ in $T_r(g, h)/r!$.

Let e_{is} be the i th unit vector in \mathcal{R}^s . Then as in (2.4), for t in \mathcal{R} the s first derivatives are

$$S_i(g) = \lim_{\epsilon \rightarrow 0} \{T(g + \epsilon e_{is} \Delta_t) - T(g)\} / \epsilon,$$

where Δ_t is defined by (2.3). Similarly, (2.5) holds if t_1, \dots, t_{r+1} and t_1, \dots, t_r are interpreted as i th rows with the other rows implicit and the same on both sides, that is

Example 2.7 Take

$$I \subset \mathcal{R}^p, \quad g : I \rightarrow \mathcal{R}, \quad w : I \rightarrow \mathcal{R}^q, \quad T(g) = \int w dg, \quad S(g) = f(\int w dg).$$

Then

$$S_g(t_1, \dots, t_r) = f_{j_1 \dots j_r}(\int w dg) w_{j_1}(t_1) \dots w_{j_r}(t_r).$$

2.4 Second Calculus

Consider functions $g, h : I \rightarrow \mathcal{R}$, where $I \subset \mathcal{R}^p$, and a functional $T(g)$ in \mathcal{R} . Suppose that for real ϵ , (2.1) holds, where now

$$T_r(g, h) = \int_I \dots \int_I T_g(t_1, \dots, t_r) h(t_1) dt_1 \dots h(t_r) dt_r$$

is assumed to exist. We call $T_g(t_1, \dots, t_r)$ the r th order (functional) derivative of $T(g)$ with respect to the second calculus, or simply the second r th order derivative. It is made unique by the condition (2.2).

Let $\delta(t)$ denote the Dirac function on \mathcal{R} . For $t, s \in \mathcal{R}^q$, set $\delta(t) = \prod_{i=1}^q \delta(t_i)$ and $\delta_t(s) = \delta(t - s)$. Taking $\lambda_1, \dots, \lambda_r$ in \mathcal{R} and $h(s) = \sum_{i=1}^r \lambda_i \delta_{t_i}(s)$, we see that $T_g(t_1, \dots, t_r)$ is the coefficient of $\lambda_1 \dots \lambda_r$ in $T_r(g, h)/r!$. So, the first derivative is (2.4) with Δ_t replaced by δ_t . Similarly, (2.5) and the chain rule hold.

Example 2.8 Take $p = 1$. Suppose that $T(g) = \int_0^1 v(t)g(t)dt = \int vg$ say. Then $T_g(t) = v(t)$ and higher derivatives are zero.

Example 2.9 Take $p = 1$. Let s be any point in $[0, 1]$. Consider functions $g : [0, 1] \rightarrow \mathcal{R}$, and $w : \mathcal{R} \rightarrow \mathcal{R}$. If $T(g) = w(g(s))$ then

$$T_g(t_1, \dots, t_r) = w_{.r}(g(s))\delta_{t_1}(s) \cdots \delta_{t_r}(s).$$

If $T(g) = w(g(s) - g(1))$ then

$$T_g(t_1, \dots, t_r) = w_{.r}(g(s) - g(1))(\delta_{t_1}(s) - \delta_{t_1}(1)) \cdots (\delta_{t_r}(s) - \delta_{t_r}(1)).$$

Since the analog of Example 3.4 in Withers and Nadarajah (2009) holds for the second calculus, we have in the notation of Example 2.5,

Example 2.10 Take

$$I \subset \mathcal{R}^p, g : I \rightarrow \mathcal{R}, v : I \rightarrow \mathcal{R}^q, T(g) = \int_I vg, S(g) = f(\int_I vg).$$

Then

$$S_g(t_1, \dots, t_r) = f_{j_1 \dots j_r}(\int_I vg)v_{j_1}(t_1) \cdots v_{j_r}(t_r).$$

We note in passing that our method avoids the complexities of Gateaux, Fréchet and Hadamard derivatives. For more on these see, for example, Dudley (1992), Bednarski et al. (1991), Ren and Sen (1991, 1995) and Pons and de Turckheim (1991).

3 Cumulant Expansions

Here we apply the first calculus of the previous section with $p = 1$, $I = [0, 1]$ in order to obtain the leading coefficients $\{a_{r_i}\}$ in the cumulant expansion (1.3) for $\hat{\theta} = T(M_n)$. The higher order tests and confidence intervals can then be obtained as in Withers and Nadarajah (Section 2, 2009). By (2.1)

$$\begin{aligned} T(M_n) &= T(m) + \sum_{r=1}^{\infty} \int_0^1 \cdots \int_0^1 T_m(t_1, \dots, t_r) \\ &\quad d(M_n(t_1) - m(t_1)) \cdots d(M_n(t_r) - m(t_r))/r!. \end{aligned}$$

Cumulant expansions for $T(M_n)$ could be obtained from this, following the method in Withers (1983), where cumulant expansions for $T(F_n)$ were obtained for F_n the empirical distribution of a random sample from F . However, in Withers (1988) we showed that these could be obtained much more easily by applying the cumulant expansions of Withers (1982) for $t(\hat{w})$ by identifying \hat{w}_i with $F_n(x_i)$, $i = 1, \dots, p$ and letting $p \rightarrow \infty$. Here we take the same short cut with $\hat{w}_i = M_n(t_i)$. We begin with some notation.

For $f : [0, 1]^r \rightarrow \mathcal{R}$, set

$$(f)_r = \int_0^1 \cdots \int_0^1 f(t_1, \dots, t_r) dt_1 \cdots dt_r.$$

For π_1, π_2, \dots sets with elements in $\{1, 2, \dots, r\}$ set

$$[\pi_1, \pi_2, \dots] = [\pi_1, \pi_2, \dots]_T = (h)_r \text{ for } h(t_1, \dots, t_r) = T_{\pi_1} T_{\pi_2} \dots,$$

where, for $\pi = \{i, j, \dots\}$, $T_\pi = T_m(t_i, t_j, \dots)$. For example,

$$\begin{aligned} [1^i] &= \int T_m(t_1)^i dt_1 = (T_1^i)_1, \\ [1^i, 11^j] &= (T_1^i T_{11}^j)_1, \\ [1^i, 12^j, 2^k] &= (T_1^i T_{12}^j T_2^k)_2 \end{aligned}$$

for $(T_1^i)(t_1) = T_m(t_1)^i$, $(T_1^i T_{11}^j)(t_1) = T_m(t_1)^i T_m(t_1 t_1)^j$ and $(T_1^i T_{12}^j T_2^k)(t_1 t_2) = T_m(t_1)^i T_m(t_1 t_2)^j T_m(t_2)^k$. For $\hat{w}_i = M_n(t_i)$,

$$\kappa(\hat{w}_{i_1}, \dots, \hat{w}_{i_r}) = \kappa_r n^{1-r} k_n(t_{r\wedge})$$

for $t_{r\wedge}$ of (2.6), where $k_n(t) = [nt]/n = t + n^{-1} \ell_n(t)$ for $\ell_n(t) = [nt] - nt$. So, (1.3) holds with a_{ri} given by Withers (1982) with $k_{r-1}^{i_1 \dots i_r}$ replaced by $\kappa_r dt_{r\wedge}$, $k_r^{i_1 \dots i_r}$ replaced by $\kappa_r d\ell_n(t_{r\wedge})$, other $k_j^{i_1 \dots i_r}$ replaced by 0, \sum replaced by \int and $t_{ij\dots}$ replaced by $T_m(t_1 t_2 \dots)$. For example,

$$a_{21} = \sum_{ij} t_i k_1^{ij} t_j = \kappa_2 \int \int T_m(t_1) T_m(t_2) dt_{2\wedge} = \kappa_2 \int T_m(t)^2 dt = \kappa_2 [1^2],$$

and the second term in a_{11} is

$$\sum_{ij} t_{ij} k_1^{ij} / 2 = \kappa_2 \int \int T_m(t_1, t_2) dt_{2\wedge} / 2 = \kappa_2 \int \int T_m(t, t) dt / 2 = \kappa_2 [11] / 2.$$

The first term in a_{11} is

$$\begin{aligned} t_i k_1^i &= \kappa_1 \int T_m(t) d\ell_n(t) = \kappa_1 \alpha_1(T_1) + n^{-1} \kappa_1 \alpha_2(T_1) + O(n^{-2}) \\ &= \kappa_1 (T_m(1) - T_m(0)) / 2 + n^{-1} \kappa_1 (T_m(1) - T_m(0)) / 12 + O(n^{-2}), \end{aligned}$$

for $\alpha_i(g)$ of the Euler-McLaurin expansion (A.1) of Withers and Nadarajah (2009, Appendix A), since this implies that

$$\int_0^1 g(t) d\ell_n(t) = \sum_{k=0}^{2K-2} \alpha_{k+1}(g) n^{-k} + n r_{n,2K},$$

where $r_{nk} = O(n^{-k})$ if $g^{(k)}$ is bounded on $[0, 1]$. The first term in a_{22} is

$$\begin{aligned} t_i k_2^{ij} t_j &= \kappa_2 \int \int T_1 T_2 d\ell_n(t_{2\wedge}) = \kappa_2 \int T_1^2 d\ell_n(t_1) = \kappa_2 \alpha_1(T_1^2) + O(n^{-1}) \\ &= \kappa_2 (T_m(1)^2 - T_m(0)^2) / 2 + O(n^{-1}). \end{aligned}$$

By the corrigendum to Withers (1982), the fifth term in a_{22} is

$$\begin{aligned} 2t_i k_1^{ij} t_j k_1^k &= 2\kappa_1 \kappa_2 \int \int \int T_1 T_{23} dt_{2\wedge} d\ell_n(t_3) = 2\kappa_1 \kappa_2 \alpha_1(g_3) + O(n^{-1}) \\ &= \kappa_1 \kappa_2 (g_3(1) - g_3(0)) + O(n^{-1}) \end{aligned}$$

for $g_3(t) = \int_0^1 T_m(s)T_m(s,t)ds$. In this way, discarding the terms in a_{rr} of magnitude $O(n^{-1})$ or less, one obtains

$$\begin{aligned} a_{10} &= T(m), \\ a_{21} &= \kappa_2 [1^2], \end{aligned} \tag{3.1}$$

$$a_{11} = \kappa_1 \alpha_1(T_m) + \kappa_2 [11]/2, \tag{3.2}$$

$$a_{32} = \kappa_3 [1^3] + 3\kappa_2^2 [1, 12, 2], \tag{3.3}$$

$$\begin{aligned} a_{22} &= \kappa_2 \alpha_1(T_m^2) + \kappa_3 [1, 11] + \kappa_2^2 [12^2]/2 + \kappa_2^2 [1, 122] \\ &\quad + 2\kappa_1 \kappa_2 \alpha_1(g_3), \end{aligned}$$

$$\begin{aligned} a_{43} &= \kappa_4 [1^4] + 12\kappa_2 \kappa_3 [1^2, 12, 2] + 12\kappa_2^3 [1, 12, 23, 3] \\ &\quad + 4\kappa_2^3 [1, 2, 3, 123]. \end{aligned}$$

Note the formal equivalence, apart from the bias terms α_k , with the expressions for the cumulant coefficients of $T(F_n)$ given by Withers (1983, page 580). (The negative term in a_{43} does not appear here due to the different form there for the r th order cumulants of \hat{w} .) So, we obtain from Withers (1983) the coefficients needed to extend inference to $O(n^{-2})$:

$$\begin{aligned} a_{12} &= \kappa_1 \alpha_2(T_m) + \kappa_3 [111]/6 + \kappa_2^2 [1122]/8, \\ a_{33} &= 3\kappa_4 [1^2, 11]/2 + 3\kappa_2 \kappa_3 [1, 12, 22] + 3\kappa_2 \kappa_3 [1, 12^2] + 3\kappa_2 \kappa_3 [1, 2, 122] \\ &\quad + 3\kappa_2 \kappa_3 [1^2, 122]/2 + 3\kappa_2^3 [1, 2, 1233]/2 + \kappa_2^3 [12, 23, 31] \\ &\quad + 3\kappa_2^3 [1, 12, 233] + 3\kappa_2^3 [1, 23, 123] + d_3, \\ a_{54} &= \kappa_5 [1^5] + 20\kappa_2 \kappa_4 [1, 2^3, 12] + 60\kappa_2^2 \kappa_3 [1, 2, 3, 12, 23] \\ &\quad + 60\kappa_2^2 \kappa_3 [1, 2^2, 13, 23] + 30\kappa_2^2 \kappa_3 [1, 2, 3^2, 123] \\ &\quad + 5\kappa_2^4 [1, 2, 3, 4, 1234] + 60\kappa_2^4 [1, 2, 34, 13, 24] \\ &\quad + 60\kappa_2^4 [1, 2, 3, 14, 234], \end{aligned}$$

where d_3 of a_{33} is the bias term, obtainable from ${}_3\Delta^{abc}$ of Withers (1983, page 67) as $d_3 = \alpha_1(g) = (g(1) - g(0))/2$ for

$$\begin{aligned} g(s) &= 3\kappa_1 \kappa_3 \int_0^1 T_m(s,t)T_m(t)^2 dt + 3\kappa_1 \kappa_2^2 \int_0^1 \int_0^1 \{T_m(s,t,u)T_m(t)T_m(u) \\ &\quad + 2T_m(s,t)T_m(t,u)T_m(u)\} dt du. \end{aligned}$$

Example 3.1 Suppose that $T(g) = \int_0^1 w(g(s), s)ds$. By (2.7) the r th derivative at $g = m$ is $\int_{t_{rv}}^1 f_r(s)ds$, where $f_r(s) = w_{,r0}(m(s), s)$. So, a_{21}, a_{11}, a_{32} are given by (3.1)–(3.3) in terms of

$$\begin{aligned} [1^r] &= \int \cdots \int s_r \wedge \prod_{i=1}^r \{f_1(s_i)ds_i\}, \\ [11] &= \int s f_2(s)ds, \\ \alpha_1(T_m) &= - \int f_1(s)ds/2, \\ [1, 12, 2] &= \int \int \int f_2(s_0)ds_0 \prod_{i=1}^2 \{\min(s_0, s_i)f_1(s_i)ds_i\}. \end{aligned}$$

In particular,

$$\begin{aligned} a_{21} &= \kappa_2 \int ds_1 \int ds_2 f_1(s_1) f_1(s_2) s_{2\wedge}, \\ 2a_{11} &= -\mu \int f_1(s) ds + \kappa_2 \int s f_2(s) ds. \end{aligned}$$

Acknowledgments

The authors would like to thank the Editor-in-Chief 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, Volume 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] Bednarski, T., Clarke, B. R. and Kolkiewicz, W. (1991). Statistical expansions and locally uniform Fréchet differentiability. *Journal of the Australian Mathematical Society, A*, **50**, 88–97.
- [4] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- [5] Comtet, L. (1974). *Advanced Combinatorics*. Reidel, Dordrecht.
- [6] Dudley, R. M. (1992). Fréchet differentiability, p -variation and uniform Donsker classes. *Annals of Probability*, **20**, 1968–1982.
- [7] Pons, O. and de Turckheim, E. (1991). Von Mises method, bootstrap and Hadamard differentiability for nonparametric general models. *Statistics*, **22**, 205–214.
- [8] Ren, J. J. and Sen, P. K. (1991). On Hadamard differentiability of extended statistical functional. *Journal of Multivariate Analysis*, **39**, 30–43.
- [9] Ren, J. J. and Sen, P. K. (1995). Hadamard differentiability on $D[0, 1]^p$. *Journal of Multivariate Analysis*, **55**, 14–28.
- [10] Von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Annals of Mathematical Statistics*, **18**, 309–342.
- [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. (1988). Nonparametric confidence intervals for functions of several distributions. *Annals of the Institute of Statistical Mathematics*, **40**, 727–746.
- [14] Withers, C. S. and Nadarajah, S. (2009). Accurate tests and intervals based on linear cusum statistics. *Statistics and Probability Letters*, **79**, 689–697.