STATISTICA, anno LXVIII, nn. 3-4, 2008

EDGEWORTH AND CORNISH FISHER EXPANSIONS AND CONFIDENCE INTERVALS FOR THE DISTRIBUTION, DENSITY AND QUANTILES OF KERNEL DENSITY ESTIMATES

C.S. Withers, S. Nadarajah

1. INTRODUCTION

Our aim here is to present nonparametric confidence intervals (CIs) for densities on $\Re$ based on kernel estimates. Suppose we have a random sample $X_1,...,X_n$ of size n on $\Re$ with empirical distribution $\hat{F},(x)$ from a distribution X~$F(x)$ with density $f(x_0)$ and derivatives $f_{\cdot r}(x_0), r=1,2,...$ finite at some point x_0. Let $K(z)$ be a function on $\Re$ with the finite values $B_{ri}=B_{ri}(K)=\int z^r K(z)^i dz$ for r=0,1,..., i=1,2,... and $B_{01}=1$. The *kernel estimate* for $f(x_0)$ of *kernel* K and *bandwidth* (or *smoothing*) parameter h is

$$\hat{f}(x_0)=\int K_h(x_0-y)d\hat{F}(y)=n^{-1}\sum_{i=1}^{n} U_{1i}, \tag{1}$$

where $K_h(x)=h^{-1}K(x/h)$ and $U_{1i}=K_h(x_0-X_i)$. The bandwidth $h=h(n)$ is assumed to satisfy $h\to 0$ as $n\to\infty$. This estimate has mean

$$T_h(F,K)=\int K_h(x_0-y)dF(y)=\int K(z)f(x_0-hz)dz=\sum_{r=0}^{\infty} f_{\cdot r}(x_0)(-h)^r B_{r1}/r!. \tag{2}$$

Note that the existence of the full Taylor series in (2) is not needed for the results in this paper. The boundedness of the required number of derivatives of f is sufficient to validate the results presented.

Now suppose that the kernel K is of order $p\geq 1$, that is, $B_{r1}=0$ for $1\leq r<p$ and $B_{rp}\neq 0$. So, $E\hat{f}(x_0)=f(x_0)+O(h^p)$. Also $var\{\hat{f}(x_0)\}=\{f(x_0)B_{02}+O(h)\}/(nh)$ so its minimum asymptotic mean square error (AMSE) is $E|\hat{f}(x_0)-f(x_0)|^2=O(n^{-\gamma})$ for $y=2p/(2p+1)$ and is obtained by choosing

$$h = cn^{-\alpha} \tag{3}$$

for $c > 0$ and $\alpha = 1/(2p+1)$. This rate γ holds of course for any $c > 0$, not just the AMSE-optimal $c = c(f, x_0)$ say. Compare Parzen (1962). (Allowing c to depend on x_0 means that $\hat{f}(x_0)$ no longer integrates to one.) In fact, this choice of α achieves minimum asymptotic integrated MSE (AIMSE)

$$\int E|\hat{f}(x) - f(x)|^2 \, dx = O(n^{-\gamma}) \tag{4}$$

under suitable regularity conditions for $\gamma = 2p/(2p+1)$ and $c > 0$, not just the AIMSE-optimal $c = c(f)$ say. See Theorem 2.1.7 of Prakasa Rao (1983), and (3.4) of Terrell and Scott (1992). Using either AMSE or AIMSE, optimal c can be estimated using an initial estimate for f. See, for example, Scott *et al.* (1977). However, the use of such an estimate may destroy its optimality property.

Kernel estimates were introduced by Rosenblatt (1956). Since then there has been a huge literature on the subject. We mention only a few here. Silverman (1986) and Terrell and Scott (1992) show that the gain in optimizing c is generally not great compared with fixed c estimators. For example, the latter have asymptotic efficiency about 0.9 or 0.8 for f normal or Cauchy. (Terrell and Scott (1992) scale K so that $B_{p1} = \pm 1$.) Prakasa Rao (1983) has many related results. For example, some give convergence rates for $sup_x |\hat{f}(x) - f(x)|$. See Theorems 2.1.10, 12, 15 and 20 and Theorems 2.2.3 and 2.2.4 for similar results. Silverman (1986) and Terrell and Scott (1992) also consider estimates of the form $\tilde{f}(x_0) = \int K_{h(y)}(x_0 - y) d\hat{F}(y)$ for K_h of (1). When $p = 2$ one can choose $h(y)$ to achieve (4) as if $p = 4$, that is with $\gamma = 8/9$. Optimizing c is one way of deciding how to choose h or equivalently how to scale K. Another way is to replace h by $h(F_n) = cn^{-\alpha}\mu_2(F_n)^{1/2}$, where $\mu_2(F)$ is the variance and F_n the empirical distribution. One can show that this does not affect the optimal rate γ. Silverman (1986, page 45) suggests this with $\alpha = 1/(2p+1)$ for $p = 2$ and $c = 1.06$.

Turning now to CIs for $f(x_0)$, the AMSE and AIMSE criteria are no longer relevant. The obvious criterion is to minimise the coverage error - the difference between the actual coverage probability and the nominal level of the CI - or rather the *asymptotic coverage error* (ACE). Theorem 2.1.18 of Prakasa Rao (1983) gives a consistent CI for $f(x_0)$ but the ACE is not given. The *first order* CIs are based on Studentizing. Their optimal bandwidth h again has the form (3), giving $ACE = O(n^{-\beta})$. We shall refer to β as the *ACE rate*. Its possible values are $\beta > 0$.

The contents of this paper are organized as follows. Section 2 provides Edgeworth and Cornish Fisher expansions for Studentized versions of the kernel density estimate, $\hat{f}(x_0)$, and its extensions. These expansions are used to derive various CIs for $f(x_0)$. Section 3 derives first order one- and two-sided CIs using *asymptotic Studentization* $(\varepsilon = 1)$ and *empirical Studentization* $(\varepsilon = p)$, where ε is an indicator variable. Section 4 derives second order one- and two-sided CIs. Section 5 considers choosing the ACE-optimal constant *c* in (3) for first order CIs based on empirical Studentization. Finally, some conclusions are noted in Section 6.

2. CORNISH FISHER EXPANSIONS

Here, we lay the groundwork by showing that the kernel density estimate has Cornish-Fisher expansions with parameter *m*=*nh*. Moments and cumulants of the kernel estimate $\hat{f}(x_0)$ in (1) are derived in Section 2.1. These quantities are used in Sections 2.2–2.5 to provide Cornish Fisher expansions for Studentized versions of $\hat{f}(x_0)$. Section 2.2 derives Cornish Fisher expansions for

$$Y_n = m^{1/2}(\hat{w}_1 - w_1)k_{2h}^{-1/2}, \tag{5}$$

where $w_1 = f(x_0)$, $\hat{w}_1 = \hat{f}(x_0)$, $U_1 = K_h(x_0 - X)$, $\kappa_{rh} = \kappa_r(U_1)$ and $k_{rh} = h^{r-1}\kappa_{rh}$. Section 2.3 derives Cornish Fisher expansions for

$$Y_n = m^{1/2}(\hat{W} - W), \tag{6}$$

where $W = (w_1, ..., w_q)'$, $\hat{W} = (\hat{w}_1, \cdots, \hat{w}_q)'$, $q \geq 1$, $w_{ih} = T_h(F, K^i)$, $K^i(z) = K(z)^i$ and $\hat{w}_{ih} = T_h(\hat{F}, K^i)$. Note that we have suppressed the dependency of w_{ih} on h by simply writing w_i for w_{ih}. Section 2.4 derives Cornish Fisher expansions for

$$Y_{n0} = m^{1/2}(t(\hat{W}) - t(W))a_{21}^{-1/2}, \tag{7}$$

where $t(W)$ is a real function with finite derivatives and satisfies the asymptotic expansion (see (19) and Withers (1982)):

$$\kappa_r(t(\hat{W})) \approx \sum_{i=r-1}^{\infty} a_{ri} m^{-i} \tag{8}$$

for $r \geq 1$ and for certain bounded functions $a_{ri} = a_{rih}(w)$ given by Withers (1982). Finally, Section 2.5 derives Cornish Fisher expansions for

$$Y_{n1} = m^{1/2} t(\hat{W}) \tag{9}$$

for $t(W) = (w_1 - f(x_0))\hat{k}_{2h}^{-1/2}$ and $\hat{k}_{2h}$ an estimate of k_{2h}.

The expansions given are "formal". We have given the expansions in the fullest possible forms - rather than say to the first or the second order - because they could lead to better approximations and have wider applicability. Often expansions give better approximations even if they diverge. We have not attempted to check the existence or the validity of the expansions. Some Cramer-type conditions and conditions on differentiability may be required to ensure existence and validity, see, for example, Hall (1992) and Garcia-Soidan (1998). This task is beyond the purpose of the present paper.

2.1 *Moments and cumulants*

The estimate $\hat{f}(x_0)$ of (1) can be viewed as the mean of a random sample of size n, where each observation of the sample is distributed as U_1. Its *i*th moment is

$$m_{ih} = EU_1^i = \int K_h(x_0 - y)^i \, dF(y) = h^{1-i} w_{ih}.$$

By (2),

$$w_{ih} = \sum_{r=0}^{\infty} f_{\cdot r}(x_0) B_{ri} (-h)^r / r! \tag{10}$$

$$= w_{i0} + O(h) \tag{11}$$

for $w_{i0} = f(x_0) \int K(z)^i \, dz$. So, w_{ih} is of magnitude 1 and $m_{ih} = O(h^{1-i})$. Typically *K* is symmetric so that (10) is a power series in h^2. Both parametric and nonparametric approaches, using ŵ and $\hat{F}$, have their merits. We can view $\hat{w}_i$ as the mean of a random sample of size *n*, where each observation of the sample is distributed as

$$U_i = K_h^i(x_0 - X) = h^{-1} K((x_0 - X)/h)^i. \tag{12}$$

The *r*th cumulant of $\hat{w}_1$ is given by

$$\kappa_r(\hat{w}_1) = n^{1-r} \kappa_{rh} = m^{1-r} k_{rh}. \tag{13}$$

The moments and cumulants of $\hat{w}$ are polynomials in *h* and *w*. An expression for the general $k_{rh} = k_{rh}(w)$ as a polynomial in *h* and *w* is given in Appendix B. One can now use (10) to expand k_{rh} as a formal power series in *h* with coefficients functions of the derivatives of *f* at x_0. Note that k_{rh} is bounded as $h \to 0$ and

$$k_{r0} = \lim_{h\to 0} h^{r-1} m_{rh} = f(x_0) B_{0r}. \tag{14}$$

2.2 *Cornish Fisher expansions for (5)*

By (13), $\hat{f}(x_0)$ behaves like the mean of a random sample of size $m = nh$ with rth cumulant k_{rh}. (Hall (1992, page 211) gave a heuristic justification for this.) By (13), the cumulants of $\hat{w}_1 = \hat{f}(x_0)$ satisfy the Cornish-Fisher assumption with respect to $m = nh$. That is $\kappa_r(\hat{w}_1) = O(m^{1-r})$ as $m \to \infty$. So, the distribution of Y_n in (5) satisfies the Cornish-Fisher expansions

$$P_n(x) = P(Y_n \le x) \approx \Phi(x) - \phi(x)\sum_{r=1}^{\infty} m^{-r/2} h_r(x, \lambda_{rh}), \tag{15}$$

$$dP_n(x)/dx \approx \phi(x)\sum_{r=1}^{\infty} m^{-r/2} \overline{h}_r(x, \lambda_{rh}), \tag{16}$$

$$\Phi^{-1}(P_n(x)) \approx x - \sum_{r=1}^{\infty} m^{-r/2} f_r(x, \lambda_{rh}), \tag{17}$$

$$P_n^{-1}(\Phi(x)) \approx x + \sum_{r=1}^{\infty} m^{-r/2} g_r(x, \lambda_{rh}), \tag{18}$$

where Φ and ϕ are the unit normal distribution and density. Note that h_r, $\overline{h}_r$, f_r, g_r are functions given in Cornish and Fisher (1960) for $r \le 4$ or in Withers (1984) and that λ_{rh} are the coefficients of the expansions. For example, $h_1(x, \lambda_{rh}) = f_1(x, \lambda_{rh}) = g_1(x, \lambda_{rh}) = \lambda_{3h}(x^2 - 1)/6$ and $\overline{h}_1(x, \lambda_{rh}) = \lambda_{3h}(x^3 - 3x)/6$. (An alternative derivation is to apply Example 1 of Withers (1983) with $T(F) = T_h(F) = \int K_h(x_0 - y)dF(y)$, $a_{ri} = 0$ if $i \ne r-1$, and $a_{ri} = \kappa_{rh}$ if $i = r-1$. So, $\kappa_r(\hat{f}(x_0)) = \kappa_{rh} n^{1-r} = k_{rh} m^{1-r}$, and $Y_n = n^{1/2}(f\hat{}(x_0) - \kappa_{1h})\kappa_{2h}^{-1/2} = m^{1/2}(\hat{w}_1 - k_{1h})k_{2h}^{-1/2}$ has an Edgeworth-Cornish-Fisher expansion with respect to m in terms of the coefficients $A_{r,r-1} = \kappa_{rh}\kappa_{2h}^{-r/2} = \lambda_{rh} h^{1-r/2}$ for λ_{rh} of (17). By (18), one can replace n and $\{A_{r,r-1}\}$ in the Cornish-Fisher expansions by m and $\{\lambda_{rh}\}$.) The expansion (15) was noted in (4.83) of Hall (1992). (There is a slip there: μ_{40} in $p_2(y)$ in second equation of page 212 should be $\mu_{40} - 3h\mu_{20}^2$. This slip is also on pages 269 and 271.)

2.3 *Cornish Fisher expansions for (6)*

The joint rth order cumulants of $\hat{w}$ have magnitude m^{1-r} for the same reason that k_{rh} is bounded:

$$\kappa(\hat{w}_{i_1},...,\hat{w}_{i_r}) = m^{r-1} k_h(i_1...i_r) \tag{19}$$

for $k_h(i_1...i_r) = h^{r-1}\kappa(U_{i_1},...,U_{i_r}) = O(1)$ and U_i of (12). This expresses $k_h(i_1...i_r)$ in the form $k_{i_1...i_r h}(F)$. We can also write $k_h(i_1...i_r) = k_{i_1...i_r h}(w)$ as a polynomial in h and w: see Appendix A. For example, $k_h(ij) = w_{i+j} - hw_i w_j$. Now fix q and set $V = (k_h(ij) : 1 \le i, j \le q)$. So, V is $q \times q$. For x in $\Re^q$, let $\Phi_V(x)$ and $\phi_V(x)$ be the distribution and density of a normal q-vector with mean 0 and covariance V. Then for Y_n in (6) the Edgeworth expansion (15) has an extension of the form

$$P(Y_n \le x) \approx \Phi_V(x) + \sum_{r=1}^{\infty} m^{-r/2} b_r(-\partial/\partial x, k_h)\Phi_V(x)/r!, \tag{20}$$

where, for s in $\Re^q$, $b_r(s, k_h) = B_r(c(s))$, $B_r(c) = \sum_{i=0}^{r} B_{ri}(c)$, $B_r(c)$ and $B_{ri}(c)$ are the complete and partial exponential Bell polynomials tabled on page 307 of Comtet (1974), $c_r(s) = k_{r+2,r+1}(s)/\{(r+2)(r+1)\}$ for $k_{r,r-1}(s) = \kappa(U_{i_1},...,U_{i_r})s_{i_1}...s_{i_r}$, and we use the tensor sum convention of implicitly summing repeated pairs of indices $i_1,...,i_r$ over their range $1,...,q$. (See Appendix B for a definition of these Bell polynomials.) Equation (5.67) of Hall (1992) gives a double expansion version of this for $q = 2$.

2.4 *Cornish Fisher expansions for (7)*

By Withers (1982), the Cornish-Fisher expansions (15)–(18) also hold with Y_n replaced by Y_{n0} in (7) and λ_{rh} replaced by

$$A_h = \{A_{ri} = a_{ri} a_{21}^{-r/2}\} \tag{21}$$

provided that a_{21} is bounded away from 0 and that some regularity conditions ensuring

$$sup_{|t_1|+...+|t_q|>\varsigma} \left| \int_{-\infty}^{\infty} exp\left\{ \sum_{j=1}^{q} \sqrt{-1} t_j K^j(u) \right\} dF(x-hu) \right| < 1 - C(\varsigma)h \tag{22}$$

hold, where $C(\cdot)$ is some bounded real valued function. Also by Withers (1982), the first few coefficients in (8) are:

$$a_{21} = t_{\cdot i}t_{\cdot j}k_h(ij),$$
$$a_{11} = t_{\cdot ij}k_h(ij)/2,$$
$$a_{32} = c_{21} + 3c_{23} \tag{23}$$

for $t_{\cdot i_1\cdots i_r} = \partial^r t(W)/\partial w_{i_1}\ldots\partial w_{i_r}$, $c_{21} = t_{\cdot i}t_{\cdot j}t_{\cdot l}k_h(ijl)$, and $c_{23} = t_{\cdot i}k_h(ij)t_{\cdot jl}k_h(lq)t_{\cdot q}$, where we use the tensor summation convention. For example,

$$P(Y_{n0} \le x) \approx \Phi(x) - \phi(x)\sum_{r=1}^{\infty} m^{-r/2} h_r(x) \tag{24}$$

for $h_r(x, A_h) = h_r(x) = \sum_{i=0}^{3r-1} P_{rih} He_r(x)$, where $He_r(x) = \phi(x)^{-1}(-d/dx)^i \phi(x)$ is the rth Hermite polynomial, P_{rih} is a polynomial in A_h, and the last summation is restricted to r–i odd. For example, $h_1(x, A_h) = \sum_{i=0,2} P_{1ih} He_i(x)$ and $h_2(x, A_h) = \sum_{i=1,3,5} P_{2ih} He_i(x)$ for $P_{10h} = A_{11h}$, $P_{12h} = A_{32h}/6$, $P_{21h} = (A_{11h}^2 + A_{22h})/2$, $P_{23h} = (4A_{11h}A_{32h} + A_{43h})/24$, and $P_{25h} = A_{32h}^2/72$. By Withers (2000), $He_i(x) = E(x + \sqrt{-1}N)^r = \psi_i$ say, where $N \sim N(0,1)$, $\psi_0 = 1$, $\psi_1 = x$, $\psi_2 = x^2 - 1$, $\psi_3 = x^3 - 3x$, $\psi_4 = x^4 - 6x^2 + 3$, $\psi_5 = x^5 - 10x^3 + 15x$, ...

2.5 *Cornish Fisher expansions for (9)*

Now consider the Studentized estimate of $f(x_0)$ given by (9). We shall consider two types of estimates with $a_{21} = 1 + O(h^\varepsilon)$ for ε=1 or p. We can apply (24) by noting that $Y_{n1} \le z$ if and only if $Y_{n0} \le x = a_{21}^{-1/2}(z - m^{1/2}t(W))$. So, $x = z + \delta$ for $\delta = \delta_0 + \delta_1 z$, $\delta_0 = -a_{21}^{-1/2} m^{1/2} t(W) = O_p(m^{1/2}h^p)$, and

$$\delta_1 = a_{21}^{-1/2} - 1 = O(h^\varepsilon). \tag{25}$$

Also at $x = z + \delta$,

$$\phi(x)He_i(x) = \phi(z)He_i(z) - \delta\phi(z)He_{i+1}(z) + O(\delta^2),$$

$$\phi(x)h_1(x) = \phi(z)h_1(z) - \delta\phi(z)\sum_{i=0,2} P_{1ih}He_{i+1}(z) + O(\delta^2),$$

$$P(Y_{n1} \leq z) = P(Y_{n0} \leq x) = \Phi(x) - \phi(x)\sum_{r=1}^{3} m^{-r/2} h_r(x) + O(m^{-2})$$

$$= \Phi(z) - \phi(z)\{-\delta + z\delta^2/2 + \sum_{r=1}^{3} m^{-r/2} h_r(z)$$

$$= -\delta m^{-1/2} \sum_{i=0,2} P_{1ih} He_{i+1}(z) + O(\delta^3 + m^{-1/2}\delta + m^{-1}\delta + m^{-2}), \qquad (26)$$

giving

$$P(|Y_{n1}| \leq z) = 2\Phi(z) - 1 - 2\phi(z)\{z(-\delta_1 + \delta_0^2/2) + m^{-1}h_2(z)$$

$$= -\delta_0 m^{-1/2} \sum_{i=0,2} P_{1ih} He_{i+1}(z) + O(\delta_0{}^3 + m^{-1/2}\delta_0{}^2 + m^{-1}\delta_0 + m^{-2}), \quad (27)$$

3. FIRST ORDER CIs

The first order CIs for $f(x_0)$ are obtained as usual by Studentizing. This can be done using the empirical estimate of $mvar\hat{f}(x_0) = k_{2h}(w)$ or by estimating its asymptotic approximation of (14), $k_{2h}(w) \approx f(x_0)\int K(z)^2\, dz$.

3.1 *First order one-sided CIs based on asymptotic studentization*

Choose

$$t(w_1) = (w_1 - f(x_0))(w_1 B_{02})^{-1/2}. \qquad (28)$$

Then $a_{21} = t_{\cdot 1}^2 k_{2h} = 1 + O(h)$. So, ε=1 in (25) and, by (26),

$$P(Y_{n1} \leq z) = \Phi(z) + O(e_{n1}) \qquad (29)$$

for $e_{n1} = m^{-1/2} + h + m^{1/2}h^p$. Taking h as in (3), $e_{n1} = O(n^{-\beta})$ with $\beta = \min((1-\alpha)/2,\ \alpha,\ (p+1/2)\alpha - 1/2)$.

The case $p \geq 2$: β has its maximum $1/3$ at α=$1/3$. So, a one-sided lower CI for $f(x_0)$ of level $\Phi(z) + O(n^{-1/3})$ and "half-width" $O(m^{-1/2}) = O(n^{-1/3})$ is given by $Y_{n1} \leq z$, that is

$$f(x_0) \geq L = \hat{w}_1 - (B_{02}\hat{w}_1/m)^{1/2} z. \qquad (30)$$

Replacing z by $-z$, a one-sided upper CI for $f(x_0)$ of level $\Phi(z)+O(n^{-1/3})$ and "half-width" $O(n^{-1/3})$ is given by $Y_{n1} \geq z$, that is

$$f(x_0) \leq U = \hat{w}_1 + (B_{02}\hat{w}_1 / m)^{1/2} z. \tag{31}$$

The case $p=1$: β has its maximum $1/4$ at $\alpha=1/4$. So, both the error and width of these one-sided CIs of nominal level $\Phi(z)$ are $O(n^{-1/4})$.

Note 1. Note $t(w_1)$ *is strictly increasing. So,* $Y_{n1} \leq z$ *if and only if* $\hat{w}_1 \leq t^{-1}(m^{-1/2}z)$ *if and only if* $Y_n \leq z_0 \;=\; m^{1/2}\{t^{-1}(m^{-1/2}z)-k_{1h}\}k_{2h}^{-1/2}$. *So, an asymptotic expansion for the left hand side of (29) is given by the right hand side of (15) at* $x=z_0$. *Alternatively, we can "stabilize the variance" by choosing* $s(\hat{w}_1)$ *with variance* $\approx s_{.1}(w_1)^2 w_1 B_{02}/n = 1/n$, *that is* $s(w_1)=2(w_1/B_{02})^{1/2}$. *That is, choosing* $t(w_1)=s(w_1)-s(f(x_0))$, *a CI with the same properties as (30) is given by* $Y_{n1}=m^{1/2}t(\hat{w}_1)\leq z$, *that is*

$$f(x_0) \geq \tilde{L} = \{\hat{w}_1^{1/2} - (B_{02}/m)^{1/2} z/2\}^2. \tag{32}$$

Replacing z *by* $-z$, *a one-sided upper CI for* $f(x_0)$ *of level* $\Phi(z)+O(n^{-1/3})$ *and "half-width"* $O(n^{-1/3})$ *is given by*

$$f(x_0) \leq \tilde{U} = \{\hat{w}_1^{1/2} + (B_{02}/m)^{1/2} z/2\}^2. \tag{33}$$

A similar idea was used in S_3, *see page 210 of Hall (1992).*

Note (11) suggests estimating $\theta = w_1$ by $\hat{\theta}=\sum_{i=1}^{\infty}\tau_i\hat{\theta}_i$ for $\hat{\theta}_i = \hat{w}_i / B_{0i}$ and $\{\tau_i\}$ constants adding to 1. Its bias is $O(h)$ or $O(h^2)$ if K is symmetric (about 0). Also $mvar(\hat{\theta}) = \sum_{i,j=1}^{\infty}\tau_i\tau_j k_h(ij)$ so that $E(\hat{\theta}-\theta)^2 = m^{-1}\theta\tau' B\tau + O(h^2+h/m)$ for $B=(B_{0,i+j})$. To this degree of accuracy this MSE is minimized by $\tau = B^{-1}1/1'B^{-1}1$, where 1 is an infinite vector of 1's, giving $E(\hat{\theta}-\theta)^2 = m^{-1}\theta/1'B^{-1}1 + O(h^2+h/m)$. Apart from the question of if and how these infinite sums need to be truncated, clearly this estimate can be used to provide an alternative CI to that of (30). However, we shall see that its asymptotic efficiency relative to the following method is 0 for $p\geq 3$.

3.2 *First order one-sided CIs based on empirical studentization*

Choose

$$t(W) = (w_1 - f(x_0))k_{2h}(w)^{-1/2}, \tag{34}$$

where $k_{2h}(w) = k_{2h} = w_2 - hw_1^2$. Then
$a_{21} = t_{\cdot 1}^2 k_h(11) + 2t_{\cdot 1}t_{\cdot 2}k_h(12) + t_{\cdot 2}^2 k_h(22) = 1 + O(h^p)$ since $t_{\cdot 1} = k_{2h}^{-1/2} + O(h^{p+1})$ and $t_{\cdot 2} = O(h^p)$. So, $\varepsilon = p$ in (25). So, by (26),

$$P(Y_{n1} \leq z) = \Phi(z) + O(e_{n1}{}') = O(n^{-\beta})$$

for $e_{n1}{}' = m^{-1/2} + m^{1/2}h^p$ and $= \min((1\quad)/2, (p+1/2)\quad 1/2)$ with the maximum

$$\beta = p/(2p+2) \tag{35}$$

at $\alpha = 1/(p+1)$. So, a one-sided lower CI for $f(x_0)$ of level $\Phi(z) + O(n^{-p/(2p+2)})$ and "half-width" $O(m^{-1/2}) = O(n^{-p/(2p+2)})$ is given by $Y_{n1} \leq z$, that is

$$f(x_0) \geq L' = \hat{w}_1 - (k_{2h}(\hat{w})/m)^{1/2} z. \tag{36}$$

Replacing z by $-z$ gives the corresponding one-sided upper CI

$$f(x_0) \leq U' = \hat{w}_1 + (k_{2h}(\hat{w})/m)^{1/2} z. \tag{37}$$

So, for $p = 1$ or 2, empirical and asymptotic Studentization achieve the same β but for $p \geq 3$ empirical Studentization is superior.

3.3 *First order two-sided CIs*

By (27) with h of (3), $P(|Y_{n1}| \leq z) = 2\Phi(z) - 1 + O(e_{n2})$ for $e_{n2} = \delta_1 + \delta_0^2 + m^{-1} \sim h^\varepsilon + mh^{2p} + m^{-1} \sim n^{-\beta}$ and $\beta = \min(\varepsilon\alpha, 1-\alpha, (2p+1)\alpha - 1)$. If $\varepsilon = 1$ the corresponding two-sided CI of level $2\Phi(z) - 1 + O(n^{-\beta})$ is given by $L \leq f(x_0) \leq U$ for L and U of (30) and (31). An alternative with the same asymptotic properties is $\tilde{L} \leq f(x_0) \leq \tilde{U}$, for $\tilde{L}$ and $\tilde{U}$ of (32) and (33). If $\varepsilon = p$ (i.e. using empirical Studentization), the corresponding two-sided CI of level $2\Phi(z) - 1 + O(n^{-\beta})$ is given by $L' \leq f(x_0) \leq U'$ for L' and U' of (36) and (37).

4. SECOND ORDER CIs

Here, we give second order one- and two-sided CIs for $f(x_0)$.

4.1 *Second order one-sided CIs using asymptotic studentization*

Taking t of (28), by (26), $P(Y_{n1} \le z) = \Phi(z) + m^{-1/2} h_1(z) + O(\delta_0 + m^{-1})$ so $P(Y_{n1} - m^{-1/2} h_1(z) \le z) = \Phi(z) + O(\delta_0 + m^{-1})$. So, one would expect that

$$P(Y_{n1} - m^{-1/2} \hat{h}_1(z) \le z) = \Phi(z) + O(\delta_0 + m^{-1}) \tag{38}$$

if $\hat{h}_1(z) = h_1(z) + O_p(m^{-1/2})$. Let us take $\hat{h}_1(z) = h_{1h}(z, \hat{w})$. We now confirm (38). Given z, set $t_m(W) = t(w_1) - m^{-1} h_{1h}(z, w)$ for $t(w_1)$ of (28). By Lemma 5.1 of Withers (1983), $\kappa_r(t_m(\hat{W})) \approx \sum_{i=r-1}^{\infty} a_{ri}'' m^{-i}$ for certain functions $a_{ri}'' = a_{ri}''(w)$.

Also $a_{10}'' = a_{10}$, $a_{21}'' = a_{21}$, $a_{11}'' = a_{11} - h_1(z)$, and $a_{32}'' = a_{32}$. Set $Y_{n0}'' = m^{1/2}(t_m(\hat{W}) - t(W)) a_{21}^{-1/2}$ and $Y_{n1}'' = m^{1/2} t_m(\hat{W})$. Then $Y_{n0}'' \le x$ if and only if $m^{1/2}(t_m(\hat{w}) - t(w)) \le y = a_{21}^{1/2} x$ if and only if $Y_{n1}'' = m^{1/2} t_m(\hat{w}) \le z = y + m^{1/2} t(w)$ and this occurs with probability $\Phi(x) - m^{-1/2} h_1''(x)\phi(x) + O(m^{-1}) = \Phi(z) - m^{-1/2} h_1''(z)\phi(z) + O(e_{n2})$ since $x = z + O(h + m^{1/2} h^p)$. Here, h_1'' is h_1 for the coefficients a_{ri}''. So, $h_1''(z)$ = $A_{11}'' + A_{32}(z^2 - 1)/6$ = $h_1(z) - a_{21}^{-1/2} h_1(z) = O(h)$ since $A_{11}'' = A_{11} - a_{21}^{-1/2} h_1(z)$. So, $P(Y_{n1}'' \le z) = \Phi(z) + O(e_{n2})$. This confirms (38).

So, a second order lower one-sided CI is $f(x_0) \ge L_2 = \hat{w}_1 - (B_{02}\hat{w}_1 / m)^{1/2}(z + m^{-1/2}\hat{h}_1(z))$. For h of (3), its error behaves as $\sim \delta + m^{-1} \sim h^{\varepsilon} + m^{1/2} h^p + m^{-1} \sim n^{-\beta}$ for $\beta = \min((1-\alpha), \varepsilon\alpha, (p + 1/2)\alpha - 1/2)$. (For $p \ge 2$ this rate improves on β using the method of Section 3.2.) Replacing z by $-z$, a second order upper one-sided CI of level $\Phi(z) + O(n^{-\beta})$ is $f(x_0) \le U_2 = \hat{w}_1 + (B_{02}\hat{w}_1 / m)^{1/2}(z - m^{-1/2}\hat{h}_1(z))$.

Recall that $h_1(z) = A_{11} + A_{32}(z^2 - 1)/6$ is given by (21)-(23) and (20).

Now suppose we take instead $\hat{h}_{10}(z) = \lim_{h \to 0} h_{1h}(z, \hat{w})$. This introduces another error of magnitude $O_p(h)$ since $h_{1h}(z, w)$ has a power series expansion in h. The effect is to add a term of magnitude $m^{-1/2} h$ into e_{n2} in (38). But $m^{-1/2} h = O(e_{n2})$ so there is no change to the above α and β. Using $\lambda_{30} = \lim_{h \to 0} \lambda_{3h} = w_{20}^{-3/2} w_{30}$ = $f(x_0)^{-1/2} B_{02}^{-3/2} B_{03}$, we obtain

$$a_{210}=1, a_{110}=-\lambda_{30}/2, a_{320}=-2\lambda_{30}, h_{10}(z)=-\lambda_{30}(2z^2+1)/6. \quad (39)$$

4.2 *Second order one-sided CIs using empirical studentization*

Now consider the second type of Studentizing. By (26), $P(Y_{n1}-m^{-1/2}h_1(z)\le z)=\Phi(z)+O(e_{n2}{}')$ for $e_{n2}{}'=m^{-1}+m^{1/2}h^p\sim n^{-\beta}$ and β=min$(1-\alpha, p\alpha, (p+1/2)\alpha-1/2)$. As before one can show that $P(Y_{n1}-m^{-1/2}\hat{h}_1(z)\le z)=\Phi(z)+O(e_{n2}{}')$. So, a second order lower one-sided CI of level $\Phi(z)+O(n^{-\beta})$ for $\beta=2p/(2p+3)$ is given by α=3/(2p+3) and $f(x_0)\ge L_2{}'=\hat{w}_1-k_{2h}(\hat{w})^{1/2}\{m^{-1/2}z-m^{-1}h_{1h}(z,\hat{w})\}$. The corresponding upper one-sided CI is $f(x_0)\le U_2{}' = \hat{w}_1+k_{2h}(\hat{w})^{1/2}\{m^{-1/2}z-m^{-1}h_{1h}(z,\hat{w})\}$. So, a two-sided CI of level $2\Phi(z)-1+O(n^{-\beta})$ is given by $L_2{}'\le f(x_0)\le U_2{}'$.

4.3 *An alternative approach*

For $j\ge 1$, Example 1 of Withers (1983) gives a one-sided CI of level $\Phi(x)+O(n^{-j/2})$ of the form $V_{jm}(\hat{F},x)\le\mu(F)$ for a mean $\mu(F)=E\,Y$ for $Y=h(X)$ and $h(x)$ a given function, where, for example, $V_{jn}(\hat{F},x)=\mu(\hat{F})+\sum_{r=1}^{j}n^{-r/2}q_r(\hat{F},x)$, $q_1(F,x)=-\mu_2^{1/2}x$, $q_2(F,x)=\mu_2^{-1}\mu_3(1+2x^2)/6$, and $\mu_r=\mu_r(Y)$.

Taking $Y=K_h(x_0-X)$, we have $n^{-r/2}q_r(F,z)=m^{-r/2}q_r(w,z)$, where $q_1(w,z)=-v_{2h}^{1/2}z$, $q_2(w,z)=v_{2h}^{-1}v_{3h}(1+2z^2)/6$ and $v_{rh}=h^{r-1}\mu_r(Y)$. See Appendix B for these as polynomials in h and w. Replacing x by z, this suggests evaluating the coverage probability of the following jth order CI for $f(x_0)$:

$$L_j''=V_{jm}(\hat{w},z)\le f(x_0), \quad (40)$$

where $V_{jn}(\hat{w},z)=\hat{w}_1+\sum_{r=1}^{j}m^{-r/2}q_r(\hat{w},z)$. We now evaluate the probability of (40) for the case j=2. Note (40) holds if and only if $m^{1/2}t_m(\hat{W})\le z$ for $t_m(W)=t(W)+m^{-1}t_1(W)$, $t_1(W)=w_2^{-3/2}w_3(1+2z^2)/6$, and t of (34). By Lemma 5.1 of Withers (1983),

$$\kappa_r(t_m(\hat{W}))\approx\sum_{i=r-1}^{\infty}a'_{ri}m^{-i}, \quad (41)$$

where the leading a'_{ri} are given in terms of a_{ri} for $t = t_0$ by $a_{r,r-1}{}' = a_{r,r-1}$ and $a_{11}{}' = a_{11} + t_1(W)$. So, the Cornish-Fisher expansions hold for $Y_{n0} = m^{1/2}(t_m(\hat{W}) - t(W))a_{21}^{-1/2}$ with polynomials $h'_i(x)$ say. By an argument similar to that proving (38) one obtains the probability of (40) as $\Phi(z) - m^{-1/2}h_1{}'(z)\phi(z) + O(e'_{n2})$ for $h_1{}'(z) = h_1(z) + a_{21}^{-1/2}t_1(W) = O(h)$ so that the ACE has magnitude $m^{-1/2}h + e'_{n2}$. If $p \geq 3$ this gives ACE rate $\beta = 2/3$ for $\alpha = 1/3$. If $p \leq 3$ this gives ACE rate $\beta = 2p/(2p+3)$ for $\alpha = 3/(2p+3)$, the same as achieved by a one-sided CI by empirical Studentization.

4.4 *An improvement*

One can show that we can improve the rate given above for $p \geq 3$ to that given for $p \geq 3$ if we choose $t_1(w)$ so that $h_1{}'(z) = O(h^\varepsilon)$ with $3\varepsilon \geq p$. (This follows since the ACE then has magnitude $m^{-1/2}h^\varepsilon + e'_{n2}$.) We now show how to achieve this with $\varepsilon = p$.

Write *t* of (34) as $ND^{-1/2}$. So, $N = O(h^p)$ is not a function of w, but *D* and the derivatives of *N* and *D* are. For example, $D_{\cdot 1} = -2hw_1$. Also

$$t_{\cdot 1} = \tilde{t}_1 - ND^{-1/2}D_{\cdot 1}/2 = \tilde{t}_1 + O(h^{p+1}),$$

$$t_{\cdot 2} = -ND^{-3/2}/2 = O(h^p),$$

$$t_{\cdot 11} = \tilde{t}_{11} + 3ND^{-5/2}D_{\cdot 1}/4 - ND^{-3/2}D_{\cdot 11}/2 = \tilde{t}_{11} + O(h^{p+2}),$$

$$t_{\cdot 12} = \tilde{t}_{12} + 3ND^{-5/2}D_{\cdot 1}^2/4 = \tilde{t}_{12} + O(h^{p+1}),$$

$$t_{\cdot 22} = 3ND^{-5/2}/4 = O(h^p)$$

for $\tilde{t}_1 = D^{-1/2}$, $\tilde{t}_{11} = -D^{-3/2}D_1$ and $\tilde{t}_{12} = -D^{-3/2}/2$. In this way we can approximate the derivatives of *t* to $O(h^p)$, and so using (21)-(23), approximate the coefficients a_{ri} and $h_r(z)$ to $O(h^p)$. Call these approximations $\tilde{a}_{ri} = \tilde{a}_{ri}(w)$ and $\tilde{h}_r(z) = \tilde{h}_{rh}(z, w)$. For example,

$$\tilde{a}_{11} = -\lambda_{3h}/2,$$

$$\tilde{a}_{32} = \tilde{c}_{21} + 3\tilde{c}_{23},$$

$$\tilde{c}_{21} = D^{-3/2}k_{3h},$$

$$\tilde{c}_{23} = -D^{-3/2} D_{.1} k_{2b}^2 - D^{-5/2} k_{2b} k_b(12) = 2hw_1 k_{2b}^{1/2} - (w_3 - hw_1 w_2) k_{2b}^{-3/2},$$

$$\tilde{h}_1(z) = \tilde{h}_{1b}(z, w) = \tilde{a}_{11} + \tilde{a}_{32}(z^2 - 1)/6.$$

Now choose $t_m(w) = t(w) - \tilde{h}_{1b}(z, w)$. By (41), $h_1'(z) = h_1(z) + a_{21}^{-1/2} t_1(w) = O(h^p)$ so that the ACE of the CI $m^{1/2} t_m(\hat{w}) \leq z$ has magnitude e_{n2}' giving ACE rate $\beta = 2p/(2p+3)$ for $\alpha = 3/(2p+3)$, the same as achieved for a one-sided CI by empirical Studentization. This gives the second order lower one-sided CI $f(x_0) \geq \tilde{L}_2 = \hat{w}_1 - k_{2b}(\hat{w})^{1/2}(m^{-1/2} z + m^{-1} \tilde{h}_{1b}(z, \hat{w}))$ of level $\Phi(z) + O(n^{-\beta})$. The corresponding upper CI is $f(x_0) \leq \tilde{U}_2 = \hat{w}_1 + k_{2b}(\hat{w})^{1/2}(m^{-1/2} z - m^{-1} \tilde{h}_{1b}(z, \hat{w}))$.

4.5 *Second order two-sided CIs using empirical studentization*

By (27),

$$P(|Y_{n1}| \leq z) = 2\Phi(z) - 1 - 2\phi(z) m^{-1} h_2(z) + O(\delta_0 + m^{-2}) =$$
$$2\Phi(z) - 1 - 2\phi(z) m^{-1} \tilde{h}_{2b}(z, w) + O(e_{n3})$$

for $e_{n3} = h^p + mh^{2p} + m^{-2}$ so that

$$P(|Y_{n1}| \leq z + m^{-1} \tilde{h}_2(z, w)) = 2\Phi(z) - 1 + O(e_{n3})$$

and this also holds with w replaced by $\hat{w}$. Also $e_{n3} \sim n^{-\beta}$ for

$$\beta = \min(p\alpha, -1 + (2p+1)\alpha, 2 - 2\alpha) = 2p/(p+2) \text{ at } \alpha = 2/(p+2).$$

This gives the two-sided CI $|f(x_0) - \hat{w}_1| \leq (k_{2b}(\hat{w})/m)^{1/2}(z + m^{-1} \tilde{h}_2(z, \hat{w}))$ of asymptotic level $2\Phi(z) - 1$ and ACE rate $\beta = 2p/(p+2)$.

5. OPTIMIZING THE CONSTANT C IN THE SMOOTHING PARAMETER (3)

5.1 *One-sided first order CI based on empirical studentization*

Consider the one-sided first order CI (36) with $\alpha = 1/(p+1)$. We noted that, for any $c > 0$, its coverage error is $O(n^{-\beta})$, where $\beta = p/(2p+2)$. We now show how to choose $c = c(f, x_0)$ to minimize the asymptotic coverage error (ACE). Set $\rho_r = f_{.r}(x_0) B_{r1}(-1)^r / r!$. Since K is of order p, $\rho_p \neq 0$. We now prove the following in terms of $h_{10}(z)$ of (39). For $\rho_p > 0$, $c(f, x_0) = [h_{10}(z)/\{(2p+1)\rho_p\}]^{1/(p+1)}$. For $\rho_p < 0$,

$$c(f,x_0)=[h_{10}(z)/\rho_p]^{1/(p+1)} \tag{42}$$

and the ACE is improved to $O(n^{-\beta_1})$, where $\beta_1=(p+2)/(2p+2)>\beta$. This surprising result is analogous to the result of Silverman (1986) quoted in the introduction. (As for that result, in practice one needs to study the effect of replacing ρ_p by an estimate. We shall not consider that here.)

We now prove these results. By (25),

$$x=z-m^{1/2}h^p\rho_p k_{2h}^{-1/2}+O(\delta_2+h^p)$$

for $\delta_2=m^{1/2}h^{p+1}$, and

$$P(Y_{n1}\le z)=\Phi(z)-\phi(z)e_{n3}+O(\delta_2+m^{-1}+h^p)$$

for

$$e_{n3}=m^{1/2}h^p\rho_p k_{2h}^{-1/2}+m^{-1/2}h_1(z)=e_{n30}+O(m^{-1/2}h),$$
$$e_{n30}=m^{1/2}h^p\rho_p+m^{-1/2}h_{10}(z)=n^{-\beta}a(c),$$
$$a(c)=\rho_p k_{2h}^{-1/2}c^{p+1/2}+c^{-1/2}h_{10}(z),$$

and α, β as in (35). We want to minimize the asymptotic value of the error $|e_{n30}|$, that is, minimize $|a(c)|$. We assume that $\int K^3>0$ so that $w_{30}>0$. The minimizing c is as given. For $\rho_p<0$, $|a(c)|=0$ at c of (42) giving ACE rate behaves as $\varepsilon+m^{-1/2}h\sim n^{-\beta_1}$ since $m^{-1}+h^p\sim n^{-p/(p+1)}$ and $p/(p+1)>\beta_1$ for $p>2$. For $\rho_p<0$, one can show that we can improve this ACE rate β_1 to $\beta_2=(p/2+1)/(p+1)>\beta_1$ by replacing $h=cn^{-\alpha}$ by $h=cn^{-\alpha}(1+c_3n^{-\alpha})$ for c of (42) and $c_3=N/D$ for $N=c^{p+3/2}\rho_{p+1}+ch_{11}(z)$, and $D=-(p+1/2)c^{p+1/2}\rho_p+c^{-1/2}h_{10}(z)/2$, where $h_{1r}(z)$ is defined by the expansion $h_1(z)=h_{1h}(z)=\sum_{r=0}^{\infty}h_{1r}(z)h^r$. In fact, this process may be repeated using a third term in h to further increase the ACE rate above β_2.

5.2 *Two-sided first order CI based on empirical studentization*

First note that $a_{21}=1-2w_{20}^{-2}\rho_p h^p+O(h^{p+1})$ and $\delta_1=w_{20}^{-2}\rho_p h^p+O(h^{p+1})$. So, by (27), $P(|Y_{n1}|\le z)=2\Phi(z)-1-2\phi(z)n^{-\beta_1}a(c)+O(n^{-\beta_2})$ for $\alpha_1=1/(p+1)$, $\beta_1=p/(p+1)$, $\beta_2=\beta_1+\alpha_1=1$, $a(c)=\gamma_0c^{-1}-\gamma_1c^p+\gamma_2c^{2p+1}$, $\gamma_0=h_{20}(z)$,

$\gamma_1 / \rho_p = z w_{20}^{-2} + w_{20}^{-1/2} \sum_{i=0,2} P_{1i0} He_{i+1}(z)$, and $\gamma_2 = z w_{20}^{-1} \rho_p^2$. The optimal c minimizes$|a(c)|$. Again there are two cases. If the minimum is positive the CI has ACE rate $\beta_1 = p/(p+1)$. But if there exists c such that $a(c) = 0$ then the CI has ACE rate $\beta_2 = 1$. Since $\gamma_2 > 0$ there one such c exists if $\gamma_0 < 0$ while if $\gamma_1 > 0$ there will be two such c if γ_1 is sufficiently large but otherwise no such c; that is, if $\gamma_1 \geq \gamma_{10}$, where γ_{10} is given by eliminating c_0 from $a(c_0) = a^{\cdot}(c_0) = 0$ for $a^{\cdot}(c)$ the derivative of $a(c)$.

6. CONCLUSIONS

In this paper, we have given first order and second order one- and two-sided CIs for $f(x_0)$. These CIs are based on Edgeworth and Cornish Fisher expansions for Studentized versions of the kernel density estimate, $\hat{f}(x_0)$. Tables 1 and 2 summarize the main CIs we give in terms of their ACE rate β achieved at their optimal choice of α and their subsection.

TABLE 1

ACE rate β and α for (3) for one-sided CIs ($\varepsilon = 1$ for asymptotic Studentization, $\varepsilon = p$ for empirical Studentization, p^* = best ACE rate achievable using the c in (3), p = order of the kernel)

ε	first order CIs	§	second order CIs	§
1	1/3 at 1/3 if p≥2	3.1	1/2 at 1/2 if p≥2	4.1.1, Hall
1	1/4 at 1/4 if p=1	3.1	2/5 at 3/5 if p=1	4.1.1, Hall
P	p/(2p+2) at 1/(p+1)	3.2	2p/(2p+3) at 3/(2p+3)	4.1.2, 4.1.4, Hall, page 222
p*	(p+2)/(2p+2) at 1/(p+1)	5.1		

TABLE 2

ACE rate β and α for (3) for twosided CIs ($\varepsilon = 1$ for asymptotic Studentization, $\varepsilon = p$ for empirical Studentization, p^* = best ACE rate achievable using the c in (3), p = order of the kernel)

ε	first order CIs	§	second order CIs	§
1	1/2 at 1/2	3.3	2/3 at 2/3	4.5
P	p/(p+1) at 1/(p+1)	3.3	2p/(p+2) at 2/(p+2)	4.5
p*	1 at 1/(p+1)	5.2		

Hall (1992, Section 4.4.3, page 222) gives one- and two-sided CIs based on bootstrapping with $\beta = 2p/(2p+3)$ and $\beta = p/(p+1)$ using $\alpha = 3/(2p+3)$ and $\alpha = 1/(p+1)$, respectively. (This is for the case where bias is not estimated. Otherwise the formulas for β become too complicated.) We have used two types of Studentizations for the first order CIs given in Section 3. Using asymptotic Studentization, we obtain $\beta = 1/3$ and $1/2$ for one- and two-sided CIs using $\alpha = 1/3$ and $1/2$ for $p \geq 2$. Using the usual empirical Studentization, these values of β are improved to $\beta = p/(2p+2)$ for one-sided CIs, and

$\beta = p/(p+1)$ for two-sided CIs both using $\alpha = 1/(p+1)$. The second order one- and two-sided CIs given (Section 4) have $\beta = 2p/(2p+3)$ for one-sided CIs, and $\beta = 2p/(p+2)$) for two-sided CIs using $\alpha = 3/(2p+3)$ and $2/(p+2)$.

We have also considered two cases for choosing the ACE-optimal constant c in (3) for first order CIs based on empirical Studentization. In one case, using this c increases the ACE rate to $(p+2)/(2p+2)$ for one-sided CIs and to 1 for two-sided CIs. This result is analogous to that of Silverman (1986).

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.

Applied Mathematics Group CHRISTOPHER S. WITHERS
Industrial Research Limited

School of Mathematics SARALEES NADARAJAH
University of Manchester

REFERENCES

L. BIGGERI, (1999), *Diritto alla 'privacy' e diritto all'informazione statistica*, in Sistan-Istat, "Atti della Quarta Conferenza Nazionale di Statistica", Roma, 11-13 novembre 1998, Roma, Istat, Tomo 1, pp. 259-279.

L. COMTET, (1974), *Advanced Combinatorics*, Reidel, Dordrecht, Holland.

R. A. FISHER, E. A. CORNISH, (1960), *The percentile points of distributions having known cumulants,* "Technometrics", 2, pp. 209-225.

P. GARCIA-SOIDAN, (1998), *Edgeworth expansions for triangular arrays,* "Communications in Statistics-Theory and Methods", 27, pp. 705-722.

P. HALL, (1992), *The Bootstrap and Edgeworth Expansion,* Springer-Verlag, New York.

E. PARZEN, (1962), *On the estimation of a probability density and mode,* "Annals of Mathematical Statistics", 33, pp. 1065-1076.

B. L. S. PRAKASA RAO, (1983), *Nonparametric Functional Estimation,* Academic Press, Orlando.

M. ROSENBLATT, (1956), *Remarks on some nonparametric estimates of a density function,* "Annals of Mathematical Statistics", 27, pp. 883-835.

D. W. SCOTT, R. A. TAPIA, J. R. THOMPSON, (1977), *Kernel density estimation by discrete maximum penalized-likelihood criteria,* "Annals of Statistics", 8, pp. 820-832.

B. W. SILVERMAN, (1986), *Density Estimation for Statistics and Data Analysis,* Chapman and Hall, New York.

G. R. TERRELL, D. W. SCOTT, (1992), *Variable kernel density estimation,* "Annals of Statistics", 20, pp. 1236-1265.

C. S. WITHERS, (1982), *The distribution and quantiles of a function of parameter estimates,* "Annals of the Institute of Statistical Mathematics, Series A", 34, pp. 55-68.

C. S. WITHERS, (1983), *Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric CIs,* "Annals of Statistics", 11, pp. 577-587.

C. S. WITHERS, (1984), *Asymptotic expansions for distributions and quantiles with power series cumulants,* "Journal of the Royal Statistical Society, Series B", 46, pp. 389-396.

C. S. WITHERS, (2000), *A simple expression for the multivariate Hermite polynomial,* "Statistics and Probability Letters", 47, pp. 165-169.

APPENDIX A: WAYS TO CONSTRUCT KERNELS OF HIGHER ORDER

We saw in Section 1 that $\hat{f}(x_0)$, the kernel estimate with kernel K of order p has bias $O(h^p)$. One can reduce the bias to $O(h^{p+1})$ by using $\tilde{f}(x_0)=\hat{f}(x_0)-\hat{f}_{\cdot p}(x_0)B_{p1}(-h)^p/p!$, where $\hat{f}_{\cdot p}(x)=(d/dx)^p\hat{f}(x)$. But this amounts to replacing the kernel K by a kernel of order p+1, giving $K_0(z)=K(z)-K_{\cdot p}(z)B_{p1}(-1)^p/p!=(1-D^pB_{p1}(-1)^p/p!)K(z)$ for $D=\partial/\partial z$. If K is symmetric then p is even so K_0 is of order $p+2$.

Example 1. Take $K(z)=\phi(z)$ so $p=2$. Set $D=\partial/\partial z$. Then

$$K_0(z)=(1-D^2/2)\phi(z)=(1-He_2(z)/2)\phi(z)=(3-z^2)\phi(z)/2$$

has order 4,

$$K_1(z)=(1+D^4/8)(1-D^2/2)\phi(z)=(1-He_2(z)/2+He_4(z)/8+He_6(z)/16)\phi(z)=$$
$$=(15-45z^2+15z^4-z^6)/16\phi(z)/16$$

has order 8,

$$K_2(z)=(1-D^6/96)(1+D^4/8)(1-D^2/2)\phi(z)=$$
$$=(1-He_2(z)/2+He_4(z)/8-He_6(z)(1/16+1/96)+He_8(z)/(96\times 2)=$$
$$=-He_{10}(z)/(96\times 8)+He_{12}(z)/(96\times 8\times 2)$$

has order 8, and so on.

APPENDIX B: CUMULANTS IN TERMS OF MOMENTS

Here, we show how to express $k_{rh} = m^{r-1}\kappa_r(\hat{w}_1)$ and $k_h(i_1...i_r) = m^{r-1}\kappa(\hat{w}_{i_1},...,\hat{w}_{i_r})$ of (13) and (19) as polynomials in h and $w = (w_1, w_2, ...)$ of (10). Let U be a real random variable with finite moments $m_r = EU^r$ and cumulants κ_r defined as usual by $\sum_{r=1}^{\infty}\kappa_r t^r / r! \equiv \ln(1+S(t))$ for all t in C for which the moment generating function $1+S(t) = Ee^{Ut}$ exists. By equation (2), page 160 of Comtet (1974),

$$\kappa_r = \sum_{i=1}^{r}(-1)^{i-1}(i-1)!B_{ri}(m), \tag{43}$$

where for $m = (m_1, m_2, ...)$, $B_{ri}(m)$ is the *partial exponential Bell polynomial* defined by

$$\left(\sum_{r=1}^{\infty} m_r t^r / r!\right)^i / i! = \sum_{r=i}^{\infty} B_{ri}(m) t^r / r!$$

for i=0,1,... . These polynomials $B_{ri}(m)$ are tabled by Comtet on page 307. For example, $B_{r1}(m) = m_r$, $B_{rr}(m) = m_1^r$, $B_{32}(m) = 3m_1 m_2$, $B_{42}(m) = 4m_1 m_3 + 3m_2^2$, and $B_{43}(m) = 6m_1^2 m_2$. An explicit formula for $B_{ri}(m)$ is given by

$$B_{ri}(m) = \sum_{n \in N^r}\{\pi(n) m_1^{n_1}...m_r^{n_r} : n_1 + ... + n_r = i, n_1 + 2n_2 + ... + rn_r = r\}$$

for $\pi(n)$ the *partition function* defined by $\pi(n) = r!/\prod_{i=1}^{r}(i!^{n_i}\, n_i!)$ for $r = n_1 + 2n_2 + ... + rn_r$. Now take $U = U_1 = K_h(x_0 - X)$, where X has distribution F on $\Re$. In the notation of Section 2, $m_r = m_{rh} = h^{1-r} w_r$, and $\kappa_r = \kappa_{rh} = h^{1-r} k_{rh}$. It follows that

$$k_{rh} = \sum_{i=1}^{r}(-1)^{i-1}(i-1)!h^{i-1}B_{ri}(w), \tag{44}$$

a polynomial in h of degree $r-1$. For example,

$$k_{1h} = w_1,$$

$$k_{2h} = w_2 - hw_1^2,$$

$$k_{3h} = w_3 - 3hw_2 w_1 + 2!h^2 w_1^3,$$

$$k_{4h} = w_4 - h(4w_1w_3 + 3w_2^2) + 2!h^2(6w_1^2w_2) - 3!h^3w_1^4,$$

$$k_{5h} = w_5 - h(5w_1w_4 + 10w_2w_3) + 2!h^2(10w_1^2w_3 + 15w_1w_2^2) - 3!h^3(10w_1^3w_2) + 4!h^4w_1^5.$$

Analogous to $w_r = w_{rh} = h^{r-1}EY^r$ and $k_{rh} = h^{r-1}\kappa_r(Y)$, define $\nu_{rh} = h^{r-1}\mu_r(Y)$. Expanding in terms of non-central moments gives

$$\nu_{rh} = \sum_{j=0}^{r}\binom{r}{j}(-hw_1)^j w_{r-j} = \sum_{j=0}^{r-2}\binom{r}{j}(-hw_1)^j w_{r-j} + (r-1)(-h)^{r-1}w_1^r$$

for $w_0 = h^{-1}$. For example, $\nu_{2h} = k_{2h} = w_2 - hw_1^2$ and $\nu_{4h} = k_{4h} + 3hk_{2h}^2$. The multivariate forms of (43) and (44) can now be written as

$$\kappa_{i_1 \ldots i_r} = \sum_{i=1}^{r}(-1)^{i-1}(i-1)!B_i^{i_1\cdots i_r}(w)$$

and

$$k_h(i_1 \ldots i_r) = \sum_{i=1}^{r}(-1)^{i-1}(i-1)!h^{i-1}B_i^{i_1\cdots i_r}(w),$$

where $B_i^{i_1\ldots i_r}(m)$ is the multivariate form of $B_{ri}(m)$. These may be written down on sight. For example, $B_{42}(m) = 4m_1m_3 + 3m_2^2$ is replaced by $B_2^{i_1\ldots i_4}(m) = \sum^4 m_{i_1}m_{i_2\cdots i_4} + \sum^3 m_{i_1i_2}m_{i_3i_4}$, where $\sum^N f(i_1 \ldots i_r)$ is $f(j_1 \ldots j_r)$ summed over all N permutations $j_1 \ldots j_r$ of $i_1 \ldots i_r$ giving different terms. So, $B_2^{i_1\ldots i_4}(m)$ is the sum of $\sum^4 m_{i_1}m_{i_2\cdots i_4} = m_{i_1}m_{i_2\ldots i_4} + m_{i_2}m_{i_3i_4i_1} + m_{i_3}m_{i_4i_1i_2} + m_{i_4}m_{i_1i_2i_4}$ and $\sum^3 m_{i_1i_2}m_{i_3i_4} = m_{i_1i_2}m_{i_3i_4} + m_{i_1i_3}m_{i_2i_4} + m_{i_1i_4}m_{i_2i_3}$. So,

$$k_h(i_1) = w_{i_1},$$

$$k_h(i_1i_2) = w_{i_1+i_2} - hw_{i_1}w_{i_2},$$

$$k_h(i_1i_2i_3) = w_{i_1+i_2+i_3} - h\sum^3 w_{i_1+i_2}w_{i_3} + 2!h^2w_{i_1}w_{i_2}w_{i_3},$$

$$k_h(i_1i_2i_3i_4) =$$
$$= w_{i_1+i_2+i_3+i_4} - h\{\sum^4 w_{i_1+i_2+i_3}w_{i_4} + \sum^3 w_{i_1+i_2}w_{i_3+i_4}\} + 2!h^2\sum^6 w_{i_1+i_2}w_{i_3}w_{i_4} - 3!h^3w_{i_1}w_{i_2}w_{i_3}w_{i_4}.$$

For example, $k_h(112) = w_4 - h(2w_1w_3 + w_2^2) + 2h^2w_1^2w_2$.

SUMMARY

Edgeworth and Cornish Fisher expansions and confidence intervals for the distribution, density and quantiles of Kernel density estimates

We show that kernel density estimates $\hat{f}(x_0)$ of bandwidth $h = h(n) \to 0$ satisfy the Cornish-Fisher assumption with parameter *m*=*nh*. This allows Cornish-Fisher expansions about the normal for standardized and Studentized kernel density estimates in powers of $m^{-1/2}$ for smooth functions *t*. The expansions given are formal and the conditions for existence/validity are not explored. The expansions lead to first order confidence intervals (CIs) for $f(x_0)$ of level $1-\omega+O(n^{-\beta})$, where $\beta = p/(2p+2)$ for one-sided CIs and $\beta = p/(p+1)$ for two-sided CIs, where *p* is the order of the kernel used. The second order one- and two-sided CIs are given with $\beta = 2p/(2p+3)$ and $\beta = 2p/(p+2)$. We show how to choose the bandwidth for asymptotic optimality.