

# NORMAL REFERENCE BANDWIDTHS FOR THE GENERAL ORDER, MULTIVARIATE KERNEL DENSITY DERIVATIVE ESTIMATOR

DANIEL J. HENDERSON AND CHRISTOPHER F. PARMETER

**ABSTRACT.** This note derives the general form of the approximate mean integrated squared error for the  $q$ -variate,  $\nu^{\text{th}}$ -order kernel density  $r^{\text{th}}$  derivative estimator. This formula allows for normal reference rule-of-thumb bandwidths to be derived. We give tables for some of the most common cases in the literature.

JEL Classification: C1 (General), C13 (Estimation), C14 (Semiparametric and nonparametric methods).

## 1. INTRODUCTION

The general order kernel density derivative estimator for multivariate data is of general interest. Chaçon, Duong & Wand (2010) derive the basic asymptotic properties of this estimator using a multivariate kernel with bandwidth matrix  $\mathbf{H}$ . They explicitly consider exact mean integrated square error (MISE) properties for this estimator both assuming the matrix form of  $\mathbf{H}$  and the common applied setting of a product kernel with a vector of bandwidths  $\mathbf{h}$  (in this case  $\mathbf{H}$  is diagonal). They are able to obtain a closed form expression for a normal scale bandwidth matrix for density derivative estimators. Here we discuss this general setting as well but also consider the kernel order. Our focus resides with approximate mean integrated squared error (AMISE) as this quantity is useful for deriving parametric reference rule-of-thumb bandwidths that are useful when beginning analysis prior to more computationally expensive data driven methods such as least-squares cross-validation.

These rule-of-thumb constants are important since in economics it appears that a standard rule-of-thumb mechanism is deployed even in settings where these constants are not directly linked to the estimator being used. For example in the `np` package (Hayfield & Racine 2008) in R (R

---

*Date:* September 30, 2011.

*Key words and phrases.* Derivative Estimation, Smoothing, AMISE.

Daniel J. Henderson, Department of Economics, State University of New York, Binghamton, NY 13902-6000, (607) 777-4480, Fax: (607) 777-2681, e-mail: [djhender@binghamton.edu](mailto:djhender@binghamton.edu).

Christopher F. Parmeter, Department of Economics, University of Miami, Coral Gables, FL 33124, (305) 284-4397, Fax: (305) 284-2985, e-mail: [cparmeter@bus.miami.edu](mailto:cparmeter@bus.miami.edu).

Development Core Team 2008) when one selects the `normal-reference` option in the `npudensbw` command (to generate bandwidths for the multivariate kernel density estimator), the  $q$  bandwidths are constructed as  $1.06 * \hat{\sigma}_j n^{-1/(4+q)}$  where  $\hat{\sigma}_j$  is the standard deviation of the  $j^{\text{th}}$  variable and  $q$  is the dimensionality of the density. The 1.06 scale is the appropriate normal reference rule-of-thumb bandwidth if one is using a second order Gaussian kernel in the univariate setting. This is troublesome in three aspects. First, if one uses any kernel other than the second order Gaussian kernel in the univariate setting this scale factor is inappropriate. Second, if one uses any kernel order other than the second order Gaussian kernel in the univariate setting this scale factor is inappropriate. And third, if one uses any kernel (even the second order Gaussian kernel) in the multivariate setting this scale factor is inappropriate. See Silverman (1986, Table 4.1) and Epanechnikov (1969) for further details.

Li & Racine (2006, pg. 26) suggest generalizing the 2nd order Gaussian kernel bandwidth normal reference rule-of-thumb bandwidth of  $1.06\hat{\sigma}_x n^{-1/5}$  to the multivariate setting using  $1.06\hat{\sigma}_{x_j} n^{-1/(4+q)}$  for  $j = 1, \dots, q$ . Similarly, Li, Perrigne & Vuong (2002) use a rule of thumb bandwidth designed for the univariate case with a product kernel (but for the Triweight kernel). Silverman (1986, pg. 87) notes that “... it will often be appropriate to use a slightly smaller value” rather than using the one dimensional normal reference rule-of-thumb bandwidth in the multidimensional case. Indeed, our results show that the normal reference rule-of-thumb bandwidths in higher dimensional settings for 2nd order kernels produces smaller scaling factors than those that are commonplace in univariate work.

It is important to note that normal reference rule-of-thumb scale factors are irrelevant when performing data-driven bandwidth selection (which we advocate wholeheartedly). They are also not optimal (in the AMISE sense) when the assumed density differs from the true density. However, rule-of-thumb bandwidths are commonly used when no data driven methods exist for a particular estimator (as in Li et al. 2002), are used in applied work as a benchmark, or are used in simulation studies to avoid computationally expensive Monte Carlo analysis. As noted by Silverman (1986, pg. 87), these rule-of-thumb bandwidths “... give a quick and easy choice of at least an initial value of the window width.” Thus, developing the appropriate scaling factors for a variety of kernels and kernel orders for  $q$ -dimensions has practical merit. Our general formula will allow easy derivation

of a normal reference rule-of-thumb bandwidth for general derivative settings as well given that closed form solutions for the derivatives of a standard multivariate normal are easily obtained.

## 2. ESTIMATOR

The general  $\nu^{\text{th}}$ -order kernel estimator for a  $q$ -variate density derivative estimator will require precise notation prior to the discussion of the estimator. For this we will follow Masry (1996). Let the vector  $r = (r_1, r_2, \dots, r_q)$  and define

$$r! = r_1! \times \dots \times r_q!, \quad |r| = \sum_{s=1}^q r_s, \quad x^r = x_1^{r_1} \times \dots \times x_q^{r_q}.$$

For taking Taylor expansions and multivariate partial derivatives we also use the notation

$$\sum_{0 \leq |r| \leq \nu} = \sum_{|r|=0}^{\nu} \sum_{r_1=0}^{|r|} \dots \sum_{r_q=0}^{|r|}, \quad g^{(r)}(x) = \frac{\partial^r g(x)}{\partial^{r_1} x_1 \dots \partial^{r_q} x_q},$$

for a given function  $g(\cdot)$ . We define the Taylor expansion of total order  $p$  of  $g(z)$  around the point  $u$  to be

$$g(z) \approx \sum_{0 \leq |r| \leq p} \frac{1}{r!} g^{(r)}(x) \Big|_{x=u} (z - u)^r.$$

For a univariate kernel  $k_\nu(u)$  we say that  $k_\nu(\cdot)$  is a  $\nu^{\text{th}}$ -order kernel if  $\int k_\nu(u) du = 1$ ,  $\int u^j k_\nu(u) du = 0$  for  $j = 1, \dots, \nu - 1$  and  $\int u^\nu k_\nu(u) du = \kappa_\nu(k_\nu) < \infty$ . We define the roughness of a function  $g(x)$  as  $R(g) = \int g(x)^2 dx$ .

The  $\nu^{\text{th}}$ -order, multivariate kernel density estimator of the  $r^{\text{th}}$  derivative of the density of  $x$  ( $f(x)$ ) is defined as

$$(1) \quad \hat{f}^{(r)}(x) = n^{-1} \sum_{i=1}^n K_{\nu,h}^{(r)}(x_i - x),$$

where the product kernel is defined as

$$(2) \quad K_{\nu,h}^{(r)}(x_i - x) = \prod_{s=1}^q h_s^{-1} k_\nu^{(r_s)} \left( \frac{x_{is} - x_s}{h_s} \right).$$

If the same univariate kernel is used for each of the  $q$  variables then we have

$$R\left(K_{\nu,h}^{(r)}\right) = \int K_{\nu,h}^{(r)}(u)^2 du = \int \cdots \int \left( \prod_{s=1}^q h_s^{-(2+2r_s)} k_{\nu}^{(2r_s)}(u/h_s) \right) du_1 \cdots du_q = h^{-(2+2r)} \prod_{s=1}^q R\left(k_{\nu}^{(r_s)}\right).$$

For  $r = (0, \dots, 0)$ ,  $R\left(K_{\nu,h}^{(r)}\right) = h^{-2}R(k_{\nu})^q$ .

To derive a normal reference rule-of-thumb bandwidth we need to derive the asymptotic mean integrated squared error (AMISE) which requires derivation of the bias and variance of  $\hat{f}^{(r)}(x)$ . For the derivations that follow we will assume that  $x$  is distributed *iid*.

**2.1. Approximate Mean Intergrated Square Error.** The bias of our  $q$ -variate,  $\nu^{\text{th}}$ -order kernel density  $r^{\text{th}}$  derivative estimator presented in (1) is

$$(3) \quad \text{Bias} \left\{ \hat{f}^{(r)}(x) \right\} = E \left[ \hat{f}^{(r)}(x) \right] - f^{(r)}(x) \approx \frac{\kappa_{\nu}(k_{\nu})}{\nu!} \sum_{s=1}^q h_s^{\nu} f^{(r_s+\nu)}(x),$$

while the variance is

$$(4) \quad \text{Var} \left[ \hat{f}^{(r)}(x) \right] \approx \frac{f^{(r)}(x) R\left(K_{\nu}^{(r)}(u)\right)}{n^{-1}h^{(1+2r)}} = \frac{f^{(r)}(x) \prod_{s=1}^q R\left(k_{\nu}^{(r_s)}(u)\right)}{n^{-1}h^{(1+2r)}}.$$

Combing our results in (3) and (4) we can produce the approximate mean integrated squared error (AMISE) for our  $q$ -variate,  $\nu^{\text{th}}$ -order kernel density  $r^{\text{th}}$  derivative estimator:

$$(5) \quad \begin{aligned} \text{AMISE} \left\{ \hat{f}^{(r)}(x) \right\} &= \int \text{Bias} \left\{ \hat{f}^{(r)}(x) \right\}^2 dx + \int \text{Var} \left[ \hat{f}^{(r)}(x) \right] dx \\ &= \int \left( \frac{\kappa_{\nu}(k_{\nu})}{\nu!} \sum_{s=1}^q h_s^{\nu} f^{(r_s+\nu)}(x) \right)^2 dx + \frac{\prod_{s=1}^q R\left(k_{\nu}^{(r_s)}(u)\right)}{n^{-1}h^{(1+2r)}} \\ &= \frac{\kappa_{\nu}^2(k_{\nu})}{(\nu!)^2} \int \left( \sum_{s=1}^q h_s^{\nu} f^{(r_s+\nu)}(x) \right)^2 dx + \frac{\prod_{s=1}^q R\left(k_{\nu}^{(r_s)}(u)\right)}{n^{-1}h^{(1+2r)}}. \end{aligned}$$

It should be clear from (5) that a closed form solution for our vector of optimal bandwidths is not obtainable in closed form solution for  $q > 1$ . Equation (5) is the general form for the  $q$ -variate,  $\nu^{\text{th}}$ -order kernel density  $r^{\text{th}}$  derivative estimator, i.e. it holds for any  $\nu^{\text{th}}$ -order kernel and any derivative order.

**2.2. Derivation of Optimal Bandwidth.** To derive a general form for the optimal bandwidths we assume that  $h_1 = \dots = h_q = h$ . In this case we have

$$AMISE \left\{ \hat{f}^{(r)}(x) \right\} = \frac{\kappa_\nu^2(k_\nu) h^{2\nu} R(\nabla^{(\nu)} f^{(r)})}{(\nu!)^2} + \frac{\prod_{s=1}^q R(k_\nu^{(r_s)}(u))}{n^{-1} h^{q+2|r|}},$$

where we have used the notation  $\nabla^{(\nu)} f^{(r)} = \sum_{s=1}^q f^{(r_s+\nu)}(x)$ . Differentiating *AMISE* with respect to  $h$  to find the optimal bandwidth produces

$$(6) \quad h_{opt} = \left[ \frac{(\nu!)^2 (q + 2|r|) \prod_{s=1}^q R(k_\nu^{(r_s)}(u))}{(2\nu) \kappa_\nu^2(k_\nu) R(\nabla^{(\nu)} f^{(r)})} \right]^{\frac{1}{2\nu+2|r|+q}} n^{-\frac{-1}{2\nu+2|r|+q}}.$$

To determine a reference rule-of-thumb bandwidth we act as though we have *a priori* knowledge of  $f(x)$ . For follow the standard convention and use the normal family for  $f(x)$ . Using the properties of normal probability density functions we have that for  $r = (0, \dots, 0)$

$$\begin{aligned} R(\nabla^{(\nu)} \phi_I(x)) &= q \phi_{2I}^{(2\nu)}(0) + q(q-1) \phi_{2I}^{(\nu+\nu')}(0) \\ &= q \left( (2\pi)^{-q/2} (2\nu)!! 2^{-\frac{\nu+q}{2}} \right) + q(q-1) \left( (2\pi)^{-q/2} (\nu!!)^2 2^{-\frac{\nu+q}{2}} \right) \\ &= \frac{q}{\pi^{q/2} 2^{\nu+q}} \left[ (2\nu)!! + (q-1)(\nu!!)^2 \right], \end{aligned}$$

where  $\phi_I(\cdot)$  represents the multivariate normal variance-covariance matrix. The double factorial notation is defined as  $(2\nu)!! = (2\nu-1) \dots 5 \cdot 3 \cdot 1$  (commonly known as the odd factorial). When  $\nu = 2$  we have the well known result that  $R(\nabla^{(2)} \phi_I(x)) = (2\sqrt{\pi})^{-q} ((1/2)q + (1/4)q^2)$  (see Silverman 1986, pg. 86). Using this definition for  $R(\nabla^{(\nu)} f^{(r)})$  inside of (6) yields

$$(7) \quad h_{opt} = \left[ \frac{\pi^{q/2} 2^{\nu+q-1} (\nu!)^2 R(k_\nu)^q}{\nu \kappa_\nu^2(k_\nu) [(2\nu)!! + (q-1)(\nu!!)^2]} \right]^{\frac{1}{2\nu+q}} n^{-\frac{1}{2\nu+q}} = C(k_\nu) n^{-\frac{1}{2\nu+q}}.$$

For  $\nu = 2$  and using the Gaussian kernel we know that  $R(k_\nu) = (2\sqrt{\pi})^{-1}$  and  $\kappa_\nu^2(k_\nu) = 1$ . In this setting we have

$$C(k_2(u)) = \left[ \frac{4}{2+q} \right]^{\frac{1}{4+q}}.$$

This is the same as equation 6.41 (pp. 152) in Scott (1992). There appears to be a typo in Table 4.1 of Silverman (1986) where  $C(k_2(u))$  is given as  $[4/(2d+1)]^{1/(d+4)}$  and this typo is repeated in Pagan & Ullah (1999).

Given that we assumed our parametric reference family was the  $q$ -variate standard normal, we rescale each of our bandwidths by the variables standard deviation to obtain the normal reference rule-of-thumb bandwidths for the  $q$ -dimensional setting. These bandwidths are of the form

$$(8) \quad h_j^{ROT} = C(k_\nu) \hat{\sigma}_j n^{-\frac{1}{2\nu+q}}.$$

### 3. ILLUSTRATION

We derive normal reference rule-of-thumb bandwidths for the class of kernels defined as

$$(9) \quad k_s(u) = \frac{(2s+1)!!}{2^{s+1}s!} (1-u^2)^s \mathbf{1}\{|u| \leq 1\}.$$

This class of kernels contains many of the most common kernels deployed in empirical work, including the Epanechnikov ( $s = 1$ ) and the Gaussian kernel ( $s = \infty$ ). For the class of polynomial kernels of order  $s$ , a  $\nu^{\text{th}}$ -order  $s$ -kernel can be constructed as (Hansen 2005, Theorem1)

$$k_{\nu,s}(u) = B_{\nu/2,s}(u) k_s(u),$$

where

$$B_{\nu/2,s}(u) = \frac{\left(\frac{3}{2}\right)_{\nu/2-1} \left(\frac{3}{2} + s\right)_{\nu/2-1}}{(s+1)_{\nu/2-1}} \sum_{j=0}^{\nu/2-1} \frac{(-1)^j \left(\frac{1}{2} + s + \nu/2\right)_j x^{2j}}{j!(\nu/2-1-j)! \left(\frac{3}{2}\right)_j}.$$

The notation  $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)}$  is Pochhammer's symbol, where  $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ . See Wand & Schucany (1990, Theorem 2.1) for a similar expression for higher order Gaussian kernels. In what follows we will dispense with the  $s$  notation of the kernel as our key results will be independent of the choice of kernel.

Table 1 provides the normal reference rule-of-thumb constants ( $C(k_\nu)$  in (8)) for the  $\nu^{\text{th}}$ -order  $q$ -variate kernel density estimator ( $r = (0, \dots, 0)$ ). The Biweight kernel corresponds to  $s = 2$  and the Triweight kernel to  $s = 3$  in our  $s$  polynomial kernel family in (9). We point out several striking features. First, in the common setting of a second order kernel ( $\nu = 2$ ) the rule-of-thumb constants are decreasing as  $q$  increases. Scott (1992) notes that these reach a minimum when  $q = 11$ . The

$\nu = 2$  case is the only one he considers. When  $\nu > 2$ , we see that the rule-of-thumb constants are increasing in the dimensionality of the problem. The basic idea behind this is given that higher-order kernels reduce bias, larger bandwidths are needed to minimize AMISE. However, note that the increase is not uniform over  $\nu$ . For example, the optimal bandwidth for  $q = 2$  and  $\nu = 6$  for the Gaussian kernel is 1.1318 and the rule-of-thumb bandwidth for that same kernel with  $\nu = 8$  is 1.1235. There appears to be an interplay between the roughness and variance which goes into calculating the rule-of-thumb bandwidth.

TABLE 1. Normal reference rule-of-thumb constants ( $C(k_\nu)$ ) for the multivariate  $\nu^{\text{th}}$  order kernel density estimator.

Kernel	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$	$q = 6$	$q = 7$	$q = 8$	$q = 9$	$q = 10$
$\nu = 2$										
Epanechnikov	2.3449	2.1991	2.1200	2.0730	2.0437	2.0246	2.0121	2.0037	1.9982	1.9947
Biweight	2.7779	2.6073	2.5150	2.4604	2.4263	2.4044	2.3900	2.3805	2.3744	2.3705
Triweight	3.1545	2.9636	2.8606	2.8000	2.7624	2.7383	2.7226	2.7124	2.7059	2.7020
Gaussian	1.0592	1.0000	0.9686	0.9506	0.9397	0.9330	0.9289	0.9265	0.9251	0.9245
$\nu = 4$										
Epanechnikov	3.0286	3.1203	3.1993	3.2680	3.3285	3.3823	3.4304	3.4737	3.5130	3.5488
Biweight	3.3930	3.4975	3.5874	3.6657	3.7347	3.7959	3.8507	3.9001	3.9449	3.9857
Triweight	3.7230	3.8403	3.9412	4.0291	4.1066	4.1753	4.2369	4.2923	4.3426	4.3884
Gaussian	1.0834	1.1233	1.1576	1.1875	1.2139	1.2373	1.2583	1.2771	1.2943	1.3099
$\nu = 6$										
Epanechnikov	3.5284	3.6886	3.8334	3.9648	4.0846	4.1942	4.2949	4.3877	4.4734	4.5528
Biweight	3.8434	4.0191	4.1780	4.3223	4.4539	4.5743	4.6849	4.7868	4.8809	4.9683
Triweight	4.1350	4.3261	4.4991	4.6561	4.7994	4.9306	5.0511	5.1621	5.2648	5.3600
Gaussian	1.0767	1.1318	1.1819	1.2275	1.2693	1.3076	1.3429	1.3756	1.4058	1.4338
$\nu = 8$										
Epanechnikov	3.9504	4.1336	4.3047	4.4647	4.6146	4.7552	4.8874	5.0118	5.1290	5.2397
Biweight	4.2308	4.4279	4.6121	4.7844	4.9458	5.0973	5.2396	5.3736	5.4999	5.6191
Triweight	4.4940	4.7050	4.9022	5.0868	5.2598	5.4222	5.5748	5.7185	5.8539	5.9818
Gaussian	1.0685	1.1235	1.1751	1.2236	1.2691	1.3120	1.3525	1.3906	1.4267	1.4608
$\nu = 10$										
Epanechnikov	4.3250	4.5167	4.6992	4.8729	5.0384	5.1961	5.3466	5.4902	5.6274	5.7585
Biweight	4.5798	4.7835	4.9775	5.1622	5.3381	5.5059	5.6659	5.8187	5.9646	6.1041
Triweight	4.8212	5.0371	5.2426	5.4383	5.6249	5.8027	5.9724	6.1344	6.2892	6.4372
Gaussian	1.0615	1.1134	1.1629	1.2103	1.2556	1.2989	1.3403	1.3800	1.4179	1.4543

Table 2 provides the rule-of-thumb constants for the setting  $|r| = 1$ , which represents estimating a derivative of the multivariate kernel density estimator in a specific dimension. Here the normal

reference rule-of-thumb bandwidth can be shown to be

$$(10) \quad h_{ROT}^{|r|=1} = \left[ \frac{\pi^{q/2} 2^{\nu+q-1} (\nu!)^2 (q+2) R(k_\nu)^{q-1} R(k_\nu^{(1)})}{\nu \kappa_\nu^2(k_\nu) [(2\nu)!!(q+\nu-1/2) + (q-1)^2(\nu!)^2]} \right]^{\frac{1}{2\nu+2+q}} n^{-\frac{1}{2\nu+2+q}}.$$

We see exactly the same patterns for  $|r| = 1$  as we did when  $|r| = 0$ . For 2nd order kernels the rule-of-thumb constants decrease for all kernels as the dimensionality increases while for higher order kernels the rule-of-thumb constants increase as the dimensionality increases. Note that Table 2 does not list rule-of-thumb constants for the class of Epanechnikov kernels as in this setting this kernel is not continuously differentiable over its support.

TABLE 2. Normal reference rule-of-thumb constants for the multivariate  $\nu^{\text{th}}$  order kernel density derivative estimator for  $|r| = 1$  provided in equation (10).

Kernel	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$	$q = 6$	$q = 7$	$q = 8$	$q = 9$	$q = 10$
$\nu = 2$										
Biweight	2.4912	2.4531	2.4085	2.374	2.3499	2.3337	2.3232	2.3166	2.3128	2.311
Triweight	2.8322	2.7905	2.741	2.7028	2.6761	2.6583	2.6469	2.6398	2.6359	2.6342
Gaussian	0.9686	0.9557	0.9397	0.9274	0.9189	0.9133	0.9098	0.9078	0.9068	0.9065
$\nu = 4$										
Biweight	3.1805	3.3219	3.4327	3.5238	3.601	3.6679	3.7268	3.7795	3.8269	3.87
Triweight	3.4918	3.6489	3.7724	3.8739	3.9601	4.0348	4.1007	4.1596	4.2127	4.2609
Gaussian	1.0308	1.0805	1.1199	1.1526	1.1805	1.2048	1.2263	1.2455	1.2628	1.2787
$\nu = 6$										
Biweight	3.6793	3.8759	4.0458	4.1962	4.3312	4.4534	4.5651	4.6675	4.7621	4.8497
Triweight	3.9598	4.173	4.3575	4.5209	4.6676	4.8006	4.9221	5.0336	5.1366	5.232
Gaussian	1.0435	1.1033	1.1556	1.2021	1.244	1.2822	1.3172	1.3494	1.3792	1.4069
$\nu = 8$										
Biweight	4.0964	4.3136	4.5083	4.6861	4.8501	5.0024	5.1445	5.2777	5.4029	5.5209
Triweight	4.3521	4.5843	4.7926	4.9829	5.1584	5.3215	5.4738	5.6165	5.7507	5.8772
Gaussian	1.0454	1.1049	1.1586	1.2079	1.2536	1.2962	1.3362	1.3737	1.4091	1.4425
$\nu = 10$										
Biweight	4.465	4.6902	4.8963	5.0877	5.2671	5.4363	5.5965	5.7486	5.8933	6.0312
Triweight	4.7012	4.9395	5.1576	5.3603	5.5504	5.7297	5.8995	6.0607	6.2141	6.3604
Gaussian	1.0444	1.1009	1.1529	1.2014	1.2472	1.2905	1.3316	1.3708	1.4082	1.4439



## REFERENCES

- Aldershof, B., Marron, J. S., Park, B. U. & Wand, M. P. (1995), ‘Facts about the Gaussian probability density function’, *Applicable Analysis* **59**(1), 289–306.
- Chaçon, J. E., Duong, T. & Wand, M. P. (2010), Asymptotics for general multivariate kernel density derivative estimators. Working Paper.
- Epanechnikov, V. A. (1969), ‘Nonparametric estimation of a multidimensional probability density’, *Theory of Probability and its Application* **14**, 153–158.
- Hansen, B. E. (2005), ‘Exact mean integrated squared error of higher order kernel estimators’, *Econometric Theory* **21**, 1031–1057.
- Hayfield, T. & Racine, J. S. (2008), ‘Nonparametric econometrics: The np package’, *Journal of Statistical Software* **27**(5).  
**URL:** <http://www.jstatsoft.org/v27/i05/>
- Li, Q. & Racine, J. S. (2006), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton.
- Li, T., Perrigne, I. & Vuong, Q. (2002), ‘Structural estimation of the affiliated private value auction model’, *RAND Journal of Economics* **33**, 171–193.
- Masry, E. (1996), ‘Multivariate regression estimation: local polynomial fitting for time series’, *Stochastic Processes and Their Applications* **65**, 81–101.
- Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
**URL:** <http://www.R-project.org>
- Scott, D. W. (1992), *Multivariate density estimation: Theory, practice, and visualization*, John Wiley and Sons, New York.
- Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, Chapman and Hall, New York.
- Wand, M. P. & Jones, M. C. (1993), ‘Comparison of smoothing parameterizations in bivariate kernel density estimation’, *Journal of the American Statistical Association* **88**(422), 520–528.
- Wand, M. P. & Schucany, W. R. (1990), ‘Gaussian-based kernels’, *Canadian Journal of Statistics* **18**, 197–204.

## APPENDIX A. FACTS PERTAINING TO THE GAUSSIAN PROBABILITY DENSITY FUNCTION

To obtain specific values for the normal reference rule-of-thumb bandwidths we need to invoke several properties of the  $q$ -variate normal probability density. The results presented here can either be found directly in Aldershof, Marron, Park & Wand (1995) and Wand & Jones (1993) or are straightforward extensions of their results.

The standard univariate normal probability density is defined as

$$\phi(x) = (2\pi)^{-1/2} e^{-(1/2)x^2},$$

while the  $r^{\text{th}}$  derivative of  $\phi(x)$  is defined as

$$\phi^{(r)}(x) = (-1)^r H_r(x) \phi(x),$$

where  $H_r(x)$  is the  $r^{\text{th}}$  Hermite polynomial. The Hermite polynomials are defined as

$$H_r(x) = xH_{r-1}(x) - (r-1)H_{r-2}(x), \quad H_0(x) = 1, \quad H_1(x) = x.$$

For the case where the random variable does not have unit variance the rescaling

$$\phi_\sigma(x) = \phi(x/\sigma)/\sigma$$

will be used. In this case the  $r^{\text{th}}$  derivative of  $\phi_\sigma(x)$  is defined as

$$\phi_\sigma^{(r)}(x) = \phi^{(r)}(x/\sigma)/\sigma^{r+1}.$$

We will be interested in the  $r^{\text{th}}$  derivative of  $\phi_\sigma(x)$  evaluated at 0. For this we have

$$\phi_\sigma^{(r)}(0) = (-1)^{r/2} \sigma^{-(r+1)} (2\pi)^{-1/2} r!! \quad \text{for } r \text{ even}$$

and 0 otherwise.

In the multivariate setting we define the standard  $q$ -variate normal probability density as

$$\phi(x) = \prod_{s=1}^q \phi(x_s) = (2\pi)^{-q/2} e^{-(1/2)x'x}.$$

For  $\Sigma$  a  $q \times q$  symmetric, positive definite matrix, we have

$$\phi_{\Sigma}(x) = |\Sigma|^{-1/2} \phi(\Sigma^{-1/2}x).$$

When  $\Sigma$  is diagonal with elements  $\sigma_s^2$  we have

$$\phi_{\Sigma}(x) = \prod_{s=1}^q \phi_{\sigma_s}(x_s).$$

Wand & Jones (1993, Theorem 2) show that for vectors  $r = (r_1, \dots, r_q)$  and  $r' = (r'_1, \dots, r'_q)$

$$\int \phi_{\Sigma}^{(r)}(x) \phi_{\Sigma'}^{(r')}(x) dx = \phi_{\Sigma+\Sigma'}^{(r+r')}(0).$$

For the development of our normal reference rule we assume that  $\Sigma = \Sigma' = I_q$  which yields

$$\int \phi_{I_q}^{(r)}(x) \phi_{I_q}^{(r')}(x) dx = \phi_{2I_q}^{(r+r')}(0).$$

Moreover, we have

$$\phi_{2I_q}^{(r+r')}(0) = \prod_{s=1}^q \phi_{\sqrt{2}}^{(r_s+r'_s)}(0).$$

It is simple to check that when  $r = r' = (0, \dots, 0)$  we have

$$\phi_{2I_q}(0) = \prod_{s=1}^q \phi_{\sqrt{2}}(0) = (2\sqrt{\pi})^{-q}.$$