FACTS ABOUT THE GAUSSIAN PROBABILITY DENSITY FUNCTION

B. Aldershof[1], **J. S. Marron**[2], **B. U. Park**[3] **and M. P. Wand**[4]

[1] *Department of Mathematics, Lafayette College, Easton, PA, 18402*

[2] *Department of Statistics, University of North Carolina, Chapel Hill, NC, 27514*

[3] *Department of Computer Science and Statistics, Seoul National University, SAN 56–1 Shinrim-Dong, Kwanak-ku, Seoul, South Korea*

[4] *Australian Graduate School of Management, University of New South Wales, Sydney, 2052, Australia*

ABSTRACT

The Gaussian probability density function plays a central role in probability, statistics and Fourier analysis. This paper presents formulas involving various combinations of moments, derivatives, integral, products and convolutions of this function. The results are useful in statistical curve estimation, but also have a beauty of their own.

AMS 1980 Subject Classification: Primary 60E05, secondary 42A99.

The research of the second and third authors was partially supported by NSF Grant DMS-8902973. Research of the second and fourth authors was partially supported by the U. S. Army Research Office through the Mathematical Sciences Institute of Cornell University. Research of the third author was partially supported by the Korean Science and Engineering Foundation 89–90, and the Basic Research Institute Program, Ministry of Education, 1995 Project No. BSRI-95-1418.

1. INTRODUCTION

The Gaussian, or Normal, probability density function is very special, both in probability theory and in statistics, because it is the limit law of the Central Limit Theorem, see Bhattacharya and Ranga Rao (1976) for access to the large literature on this topic. The main idea behind it is that sums of independent random variables, when properly normalized, all tend to have the same limiting distribution, the Gaussian.

The Normal probability density function also plays an important role in Fourier analysis, because it is the unique fixed point of the Fourier transform isometry between L_2 and itself, see (7.6) of Rudin (1973). For other interesting characterizations and properties of this function, see Patel and Read (1982).

In this paper some properties of this function are reviewed, and others are explored. In particular explicit formulas for quantities involving various combinations of moments, derivatives, products and convolutions are given.

The motivation for this investigation is recent work in statistical nonparametric curve estimation. However, we were pleased to find the results have a simple and elegant mathematical structure, which we feel is of independent interest. The proofs are mostly based on straightforward Fourier transform theory, and we feel they once again demonstrate the utility of that methodology.

Applications of these results have become important in two aspects of curve estimation. The first of these is the derivation of exact mean integrated squared errors for kernel density estimation of normal mixture densities, see Marron and Wand (1992). This promising new tool is proving to be far more powerful and efficient than the currently widely used Monte Carlo methods for assessing issues such as when, and how well, asymptotic analyses describe the actual situation.

The second is for the analysis, and also implementation, of data based smoothing parameter selectors. This area has recently seen a good deal of activity, and the proposal of many new methods. A useful tool in the comparison of these is the asymptotic limiting distributions, see Section 3 of Park and Marron (1990) for example (and for access to earlier references). Results of this paper are useful for this because the constants found in those distributions are usually of this form. In addition, some of the new smoothing parameter selectors require calibration in terms of a reference distribution. A popular means of doing this, see for example (2.10) of Park and Marron (1990), is to use the Normal scale family, with some scale estimate, which again requires results of the type in the present paper for implementation.

Notation and preliminary results are given in Section 2. Section 3 deals with moment results and Section 4 gives product formulas. In Section 5 convolution and integrals of products are presented while Section 6 contains miscellaneous extensions related to results from previous sections. For several of the results the application to curve estimation are pointed out through remarks. All proofs are given in the Appendix.

2. NOTATION AND PRELIMINARIES

The standard Normal probability density function is defined by

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

A subscript means a rescaling of the type

$$\phi_\sigma(x) = \phi(x/\sigma)/\sigma. \tag{2.1}$$

A random variable X has a Normal distribution with parameters μ and σ^2, written as $X \sim N(\mu, \sigma^2)$, if its probability density function in the variable x is $\phi_\sigma(x - \mu)$.

The convention concerning derivatives and rescalings is that rescalings are done first, so

$$\phi_\sigma^{(r)}(x) = (d^r/dx^r)\phi_\sigma(x) = \phi^{(r)}(x/\sigma)/\sigma^{r+1}. \tag{2.2}$$

Integrals with no explicit limits are understood to mean definite integrals from $-\infty$ to ∞. Given functions f and g, the convolution (when it exists) is

$$(f * g)(x) = \int f(x-u)g(u)\,du. \tag{2.3}$$

This version of the definition of convolution is preferred by probabilists because $f * g$ is the probability density function of $X + Y$, where X and Y are independent random variables having density f and g respectively.

For $r = 1, 2, \ldots,$ the "Odd Factorial" is defined by

$$\mathrm{OF}(2r) = (2r-1)(2r-3)\cdots 1 = \frac{(2r)!}{2^r r!}. \tag{2.4}$$

For $r = 0, 1, 2, \ldots$ the Hermite polynomials are defined by

$$H_r(x) = (-1)^r \phi^{(r)}(x)/\phi(x). \tag{2.5}$$

A handy representation of $\phi^{(r)}(x)$ is therefore

$$\phi^{(r)}(x) = (-1)^r H_r(x)\phi(x) \tag{2.6}$$

from which it follows that

$$\phi_\sigma^{(r)}(x-\mu) = (-1)^r H_r\left(\frac{x-\mu}{\sigma}\right)\phi\left(\frac{x-\mu}{\sigma}\right)\Big/\sigma^{r+1}. \tag{2.7}$$

Another form (from Gradshteyn and Ryzhik, 1980, but observe that our Hermites are equal to their's with an argument of $x/2^{1/2}$ and multiplied by $2^{-r/2}$) is

$$H_r(x) = x^r - \mathrm{OF}(2)\binom{r}{2}x^{r-2} + \mathrm{OF}(4)\binom{r}{4}x^{r-4} - \mathrm{OF}(6)\binom{r}{6}x^{r-6} + \dots. \tag{2.8}$$

where the sum is taken to $[r/2]+1$ terms, and $[\cdot]$ is the greatest integer function. Useful recursion formulas for these are (from 8.952 of Gradshteyn and Ryzhik, 1980):

$$H_r(x) = xH_{r-1}(x) - (r-1)H_{r-2}(x). \tag{2.9}$$

$$(d/dx)H_r(x) = rH_{r-1}(x). \tag{2.10}$$

The y intercept (from 8.956 of Gradshteyn and Ryzhik (1980)) is

$$H_r(0) = \begin{cases} (-1)^{r/2}\mathrm{OF}(r) & r \text{ even} \\ 0 & r \text{ odd.} \end{cases} \tag{2.11}$$

From (2.8) we have

$$(-i)^r H_r(ix) = x^r + \mathrm{OF}(2)\binom{r}{2}x^{r-2} + \mathrm{OF}(4)\binom{r}{4}x^{r-4} + \mathrm{OF}(6)\binom{r}{6}x^{r-6} + \dots \tag{2.12}$$

where the sum is taken to $[r/2]+1$ terms. Note that the coefficients are as for $H_r(x)$ except that all signs are positive.

From (2.6), (2.11) and (2.2) it follows that

$$\phi_\sigma^{(r)}(0) = \begin{cases} (-1)^{r/2}(2\pi)^{-1/2}\mathrm{OF}(r)\sigma^{-r-1} & r \text{ even} \\ 0 & r \text{ odd.} \end{cases} \tag{2.13}$$

3. MOMENTS AND RELATED QUANTITIES

Theorem 3. *For $\sigma > 0$, $X \sim N(\mu, \sigma^2)$ and $\nu > -1$,*

$$E(X^\nu) = \int x^\nu \phi_\sigma(x-\mu)\,dx = \sum_{k=0}^{\infty} \frac{\Gamma(\nu+1)\mu^{\nu-2k}\sigma^{2k}}{\Gamma(\nu+1-2k)k!2^k}$$

where the first equality defines $E(X^\nu)$, the νth central moment of X.

Corollary 3.1. *For $\sigma > 0$, $r = 0, 1, 2, \dots$ and $X \sim N(\mu, \sigma^2)$,*

$$E(X^r) = \int x^r \phi_\sigma(x-\mu)\,dx = (-i\sigma)^r H_r(i\mu/\sigma).$$

See (2.12) for a means of representing these polynomials.

Corollary 3.2. *For $X \sim N(0, \sigma^2)$, when r is even,*

$$E(X^r) = \int x^r \phi_\sigma(x)\,dx = \sigma^r \mathrm{OF}(r),$$

and when r is odd

$$E(X^r) = \int x^r \phi_\sigma(x)\,dx = 0.$$

Corollary 3.3. *When $r \geq r'$,*

$$\int x^r \phi_\sigma^{(r')}(x-\mu)\,dx = (-1)^r \frac{r!}{(r-r')!}(i\sigma)^{r-r'} H_{(r-r')}(i\mu/\sigma).$$

Otherwise

$$\int x^r \phi_\sigma^{(r')}(x-\mu) = 0.$$

Again, see (2.12) for convenient representation.

Corollary 3.4. *When $r + r'$ is even and $r \geq r'$,*

$$\int x^r \phi_\sigma^{(r')}(x)\,dx = (-1)^r \frac{r!}{(r-r')!}\sigma^{r-r'} \mathrm{OF}(r-r').$$

Otherwise

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

Remark: This result, together with Theorem 4 appears in the limiting distribution of Biased Cross-Validation, see for example Theorem 3.2 of Park and Marron (1990).

4. PRODUCTS

Theorem 4. *For $\sigma_i > 0$, $i = 1, \ldots, m$,*

$$\prod_{i=1}^m \phi_{\sigma_i}(x-\mu_i) = (2\pi)^{1-m/2}\left(\prod_{i=1}^m \sigma_i\right)^{-1} \phi_{\tilde{\sigma}}(\tilde{\mu})\phi_{\tilde{\tilde{\sigma}}}(x-\tilde{\tilde{\mu}})$$

where

$$\tilde{\sigma} = \left(\sum_{i=1}^m \sigma_i^{-2}\right)^{1/2}, \qquad \tilde{\mu} = \left[\sum\sum\nolimits_{i<j}\{(\mu_i-\mu_j)/(\sigma_i\sigma_j)\}^2\right]^{1/2},$$

$$\tilde{\tilde{\sigma}} = \tilde{\sigma}^{-1}, \qquad \tilde{\tilde{\mu}} = \sum_{i=1}^m \sigma_i^{-2}\mu_i \Big/ \sum_{i=1}^m \sigma_i^{-2}.$$

Corollary 4.1.

$$\int \prod_{i=1}^{m} \phi_{\sigma_i}(x-\mu_i)\,dx = (2\pi)^{1-m/2} \left(\prod_{i=1}^{m} \sigma_i\right)^{-1} \phi_{\tilde{\sigma}}(\tilde{\mu}).$$

Corollary 4.2. *For $\sigma, \sigma' > 0$,*

$$\phi_\sigma(x-\mu)\phi_{\sigma'}(x-\mu') = \phi_{\sigma^*}(\mu-\mu')\phi_{\sigma\sigma'/\sigma^*}(x-\mu^*)$$

where

$$\sigma^* = (\sigma^2 + {\sigma'}^2)^{1/2}, \qquad \mu^* = \frac{{\sigma'}^2\mu + \sigma^2\mu'}{\sigma^2 + {\sigma'}^2}.$$

Corollary 4.3.

$$\phi_\sigma(x)\phi_{\sigma'}(x) = (2\pi)^{-1/2}\phi_{\sigma\sigma'/\sigma^*}(x)/\sigma^*.$$

Corollary 4.4. *For $\sigma_1, \ldots, \sigma_r > 0$,*

$$\prod_{i=1}^{r} \phi_{\sigma_i}(x) = (2\pi)^{(1-r)/2}\phi_{(\prod_i \sigma_i)/(\sigma^+)}(x)/\sigma^+,$$

where

$$\sigma^+ = \left(\sum_{i=1}^{r}\prod_{j\neq i}\sigma_j\right)^{1/2} = \left[\left(\prod_{i=1}^{r}\sigma_j^2\right)\left(\sum_{i=1}^{r}\sigma_j^{-2}\right)\right]^{1/2}.$$

Corollary 4.5.

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

Corollary 4.6. *For $\sigma_1, \sigma_2, \sigma_3 > 0$,*

$$\phi_{\sigma_1}(x-\mu_1)\phi_{\sigma_2}(x-\mu_2)\phi_{\sigma_3}(x-\mu_3) = (2\pi)^{-1/2}\phi_{\sigma^{**}}(\mu^{**})\phi_{\sigma^{***}}(x-\mu^{***}),$$

where

$$\sigma^{**} = (\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2)^{1/2}.$$

$$\mu^{**} = \{\sigma_1^2(\mu_2-\mu_3)^2 + \sigma_2^2(\mu_1-\mu_3)^2 + \sigma_3^2(\mu_1-\mu_2)^2\}^{1/2},$$

$$\sigma^{***} = \sigma_1\sigma_2\sigma_3/\sigma^{**},$$

$$\mu^{***} = (\sigma_2^2\sigma_3^2\mu_1 + \sigma_1^2\sigma_3^2\mu_2 + \sigma_1^2\sigma_2^2\mu_3)/(\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2).$$

Theorem 5. *For $\sigma, \sigma' > 0$, and $r, r' = 0, 1, 2, \ldots$*

$$\phi_\sigma^{(r)}(\cdot - \mu) * \phi_{\sigma'}^{(r')}(\cdot - \mu')(x) = \phi_{\sigma^*}^{(r+r')}(x - \mu - \mu')$$

where $\sigma^* = (\sigma^2 + {\sigma'}^2)^{1/2}$ as in Corollary 4.2.

Corollary 5.1. *For $r_1, r_2, \ldots, r_m = 0, 1, 2, \ldots$*

$$\phi^{(r_1)} * \phi^{(r_2)} * \cdots * \phi^{(r_m)}(x) = \phi_{m^{1/2}}^{(r_1 + r_2 + \cdots + r_m)}(x).$$

Corollary 5.2.

$$\int \phi_\sigma^{(r)}(x - \mu) \phi_{\sigma'}^{(r')}(x - \mu')\, dx = (-1)^r \phi_{\sigma^*}^{(r+r')}(\mu - \mu').$$

Remark: This result is used in the calculation of both asymptotic and exact mean integrated squared error formulas, as in Marron and Wand (1992). It also is used in the calibration of many different "plug-in" smoothing parameter selectors, for example that treated in Hall, Sheather, Jones and Marron (1991).

Corollary 5.3. *For $r + r'$ even*

$$\int \phi_\sigma^{(r)}(x) \phi_{\sigma'}^{(r')}(x)\, dx = (-1)^{(r-r')/2} (2\pi)^{-1/2} \mathrm{OF}(r + r') / (\sigma^*)^{r+r'+1}$$

where σ^* is as in Corollary 4.2 and for $r + r'$ odd

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

Corollary 5.4.

$$\int \phi^{(r)}(x)^2\, dx = \frac{\mathrm{OF}(2r)}{\pi^{1/2} 2^{r+1}}.$$

Remark: This form appears in both the limiting distribution of the smoothing parameters, and also in the data-based calibration, for both Smoothed Cross-Validation, as discussed in Theorems 3.1 and 4.2 of Hall, Marron and Park (1992), and also for the main method discussed in Park and Marron (1990), see Theorem 3.3 there.

Corollary 5.5.

$$\int \{\phi^{(2)} * \phi^{(2)}(x)\}^2\, dx = \frac{105}{2^9 (2\pi)^{1/2}}.$$

Remark: As with Corollary 5.4, this form also appears in both the calibration and the analysis of many different types of recently proposed smoothing parameter selectors, see Park and Marron (1990).

Corollary 5.6.

$$\int \{\phi^{(r)} * \phi(x)\}^2\, dx = \frac{\mathrm{OF}(2r)}{(2\pi)^{1/2} 2^{2r+1}}.$$

Remark: This also appears in both places in the method of Park and Marron (1990).

6. MISCELLANEOUS EXTENSIONS

Theorem 6.1.

$$\int \prod_{i=1}^{m} \phi_{\sigma_i}^{(r_i)}(x-\mu_i)\, dx$$

$$= (2\pi)^{1-m/2}\phi_{\tilde\sigma}(\tilde\mu) \sum_{j_1=0}^{r_1} \cdots \sum_{j_m=0}^{r_m} \prod_{k=1}^{m} \left\{ \binom{r_k}{j_k} H_{r_k-j_k}(\mu_k^\dagger/\sigma_k)/\sigma_k^{r_k+j_k+1} \right\} \tilde{\tilde\sigma}^{\sum_{\ell=1}^m j_\ell} \mathrm{OF}\left(\sum_{\ell=1}^{m} j_\ell\right)$$

where

$$\mu_k^\dagger = \mu_k - \tilde{\tilde\mu} \quad \textit{and} \quad \mathrm{OF}(r) = 0 \quad \textit{if } r \textit{ is odd}$$

and $\tilde\sigma$, $\tilde\mu$, $\tilde{\tilde\sigma}$ *and* $\tilde{\tilde\mu}$ *are as in Theorem 4.*

Remark: This result is useful for exact mean squared error calculations in the estimation of integrated squared derivatives of probability density functions.

Corollary 6.1.1.

$$\int \prod_{i=1}^{m} \phi_{\sigma_i}^{(r_i)}(x)\, dx$$

$$= (2\pi)^{(1-m)/2} \sum_{j_1=0}^{r_1} \cdots \sum_{j_m=0}^{r_m} \prod_{k=1}^{m} \left\{ \binom{r_k}{j_k} \sigma_k^{-r_k-j_k-1} \mathrm{OF}(r_k-j_k) \right\} \tilde{\tilde\sigma}^{1+\sum_{\ell=1}^m j_\ell} \mathrm{OF}\left(\sum_{\ell=1}^{m} j_\ell\right).$$

Corollary 6.1.2.

$$\int \prod_{i=1}^{m} \phi^{(r_i)}(x)\, dx$$

$$= (2\pi)^{(1-m)/2} \sum_{j_1=0}^{r_1} \cdots \sum_{j_m=0}^{r_m} \prod_{k=1}^{m} \left\{ \binom{r_k}{j_k} \mathrm{OF}(r_k-j_k) \right\} m^{-\frac{1}{2}(1+\sum_{\ell=1}^m j_\ell)} \mathrm{OF}\left(\sum_{\ell=1}^{m} j_\ell\right).$$

Corollary 6.1.3.

$$\int \{\phi^{(2)}\}^2 \phi = \frac{1}{3^{3/2}\pi}.$$

Corollary 6.1.4.

$$\int \{\phi^{(4)}\}^2 \phi = \frac{76}{3^{7/2}\pi}.$$

Remarks: This appears in both the limiting distribution and the calibration of Smoothed Cross-Validation, see Theorems 3.1 and 3.2 and (4.2) of Hall, Marron and Park (1992).

Theorem 6.2. *For $k = 1, 2, \ldots$ and $r + r' \equiv k$ (mod 2),*

$$\int x^k \phi^{(r)}(x) \phi^{(r')}(x)\, dx$$

$$=(-1)^{(r'-r+2k)/2} 2^{-(r'+r+2k+2)/2} \pi^{-1/2} k!$$
$$\times \sum \sum\nolimits_{(p,p') \in \mathcal{A}} \binom{r}{p} \binom{r'}{p'} (-1)^{(p-p')/2} 2^{(p+p')/2} \mathrm{OF}(r + r' - p - p')/\{(k - p - p')/2\}!$$

where

$$\mathcal{A} = \{(a_1, a_2) \in \mathcal{Z}^+ \times \mathcal{Z}^+ : a_1 + a_2 \equiv k \ (\mathrm{mod}\ 2),\ a_1 + a_2 \leq k, a_1 \leq r, a_2 \leq r'\}$$

and $\mathcal{Z}^+$ is the set of positive integers. For $r + r' \not\equiv k$ (mod 2),

$$\int x^k \phi^{(r)}(x) \phi^{(r')}(x)\, dx = 0.$$

Corollary 6.2.1. *For $k = 1, 2, \ldots$ and $r + r' \equiv k$ (mod 2),*

$$\int x^k \phi_\sigma^{(r)}(x) \phi_\sigma^{(r')}(x)\, dx$$

$$= \sigma^{-(r'+r-k+1)} (-1)^{(r'-r+2k)/2} 2^{-(r'+r+2k+2)/2} \pi^{-1/2} k!$$
$$\times \sum \sum\nolimits_{(p,p') \in \mathcal{A}} \binom{r}{p} \binom{r'}{p'} (-1)^{(p-p')/2} 2^{(p+p')/2} \mathrm{OF}(r + r' - p - p')/\{(k - p - p')/2\}!$$

while for $r + r' \not\equiv k$ (mod 2),

$$\int x^k \phi_\sigma^{(r)}(x) \phi_\sigma^{(r')}(x)\, dx = 0.$$

Corollary 6.2.2. *For $r + r'$ odd,*

$$\int x \phi_\sigma^{(r)}(x) \phi_\sigma^{(r')}(x)\, dx = \sigma^{-(r+r')} \frac{(-1)^{(r-r'+1)/2} (r - r') \mathrm{OF}(r + r' - 1)}{\pi^{1/2} 2^{(r+r'+3)/2}},$$

while for $r + r'$ even,

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

Corollary 6.2.3. *For $r + r'$ even and $r + r' > 0$,*

$$\int x^2 \phi_\sigma^{(r)}(x) \phi_\sigma^{(r')}(x)\, dx = \sigma^{-(r+r'-1)} \frac{(-1)^{(r-r')/2} \{4rr' - (r + r' - 1)^2\} \mathrm{OF}(r + r' - 2)}{\pi^{1/2} 2^{(r+r'+4)/2}},$$

while for $r = r' = 0$,

$$\int x^2 \phi_\sigma^{(r)}(x)\phi_\sigma^{(r')}(x)\,dx = \sigma/(4\pi^{1/2}),$$

and for $r + r'$ *odd,*

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

Theorem 6.3. *For* $\sigma, \sigma' > 0$,

$$\int x^2 \phi_\sigma^{(1)}(x)\phi_{\sigma'}^{(1)}(x)\,dx = 3(2\pi)^{-1/2}(\sigma\sigma')^2/(\sigma^2 + {\sigma'}^2)^{5/2}.$$

Remark: This form comes up in the analysis and calibration of the smoothing parameter factorization version of Smoothed Cross-Validation, see Theorem 1 of Jones, Marron and Park (1991).

Corollary 6.4.1.

$$\int \{x\phi * \phi^{(1)}(x) - 2x\phi^{(1)}(x)\}^2\,dx = \pi^{-1/2}\{3(2)^{1/2}/16 - 4(6)^{1/2}/9 + 3/2\}.$$

Remark: This form appears in the limiting distribution of Least Squares Cross-Validation, see Theorem 3.1 in Park and Marron (1990).

APPENDIX: PROOFS

Notation and Preliminaries:

A rescaling of our convolution defined at (1.3), that is used by Rudin (1973), is

$$(f *_R g)(x) = (f * g)(x)/(2\pi)^{1/2}. \tag{A.1.1}$$

This version of the convolution is cleaner and more convenient when dealing with Fourier transforms, as done in these proofs, because fewer distracting factors of $(2\pi)^{1/2}$ are needed in the formulas. Rudin's version of the Fourier transform of f is

$$\mathrm{FT}_f^R(t) = (2\pi)^{-1/2}\int f(x)e^{-itx}\,dx. \tag{A.1.2}$$

Note that

$$\mathrm{FT}_\phi^R(t) = \phi(t). \tag{A.1.3}$$

By 7.1d and 7.4c of Rudin (1973) (using Rudin's notation of $D^r = i^{-r}(d^r/dx^r)$),

$$\mathrm{FT}_{\phi^{(r)}}^R(t) = \mathrm{FT}_{i^r D^r\phi}^R(t) = i^r t^r \mathrm{FT}_\phi^R(t) = i^r t^r \phi(t), \tag{A.1.4}$$

and similarly, using 7.2a and 7.2d of Rudin (1973),

$$\mathrm{FT}^R_{\phi_\sigma^{(r)}(\cdot\,-\mu)}(t) = i^r t^r e^{-it\mu}\mathrm{FT}^R_\phi(\sigma t) = i^r t^r e^{-it\mu}\phi(\sigma t) = i^r t^r e^{-it\mu}\phi_{1/\sigma}(t)/\sigma. \tag{A.1.5}$$

where we have used the identity

$$\phi(\sigma t) = \phi\{t/(1/\sigma)\}\{1/(1/\sigma)\}(1/\sigma) = \phi_{1/\sigma}(t)/\sigma. \tag{A.1.6}$$

A related useful result, which follows from 7.4c of Rudin (1973), from (A.1.5) (with r taken there as 0) and by completing the square, is

$$\mathrm{FT}^R_{\cdot^r\phi_\sigma(\cdot\,-\mu)}(t) = (-i^{-1})^r(d^r/dt^r)e^{-it\mu}\phi(\sigma t) = i^r(d^r/dt^r)\phi(\sigma t + i\mu/\sigma)e^{-\mu^2/(2\sigma^2)}.$$

Hence by (A.1.6) and (2.7)

$$\begin{aligned}\mathrm{FT}^R_{\cdot^r\phi_\sigma(\cdot\,-\mu)}(t) =& i^r e^{-\mu^2/(2\sigma^2)}\phi^{(r)}_{1/\sigma}(t + i\mu/\sigma^2)/\sigma \\ =&(-i\sigma)^r e^{-\mu^2/(2\sigma^2)}H_r(\sigma t + i\mu/\sigma)\phi(\sigma t + i\mu/\sigma).\end{aligned} \tag{A.1.7}$$

A closely related fact, which uses (A.1.5), is

$$\mathrm{FT}^R_{\cdot^r\phi_\sigma^{(r')}(\cdot\,-\mu)}(t) = (-i^{-1})^r(d^r/dt^r)\{i^{r'}t^{r'}e^{-it\mu}\phi(\sigma t)\} = i^{r+r'}(d^r/dt^r)\{t^{r'}e^{-it\mu}\phi_{1/\sigma}(t)\}/\sigma. \tag{A.1.8}$$

Proof of Theorem 3.

Using Nishimoto's (1984) definition of a fractional derivative (Definition 1),

$$\frac{d^\nu}{dx^\nu}\phi_\sigma(x-\mu) = (2\pi i)^{-1}\Gamma(\nu+1)\int_{-\infty}^{0+}\eta^{-\nu-1}\phi_\sigma(t+\eta-\mu)\,d\eta$$

$$= (2\pi)^{-3/2}(i\sigma)^{-1}\sum_{k=0}^{\infty}\frac{(-1)^k\Gamma(\nu+1)}{k!(2\sigma^2)^k}\int_{-\infty}^{0+}\eta^{2k-\nu-1}\exp[\{2(t-\mu)\eta + (t-\mu)^2\}/(2\sigma^2)]\,d\eta$$

$$= \phi_\sigma(x-\mu)\sum_{k=0}^{\infty}\frac{(-1)^k\Gamma(\nu+1)}{\Gamma(\nu+1-2k)k!(2\sigma^2)^k}\left(\frac{t-\mu}{\sigma^2}\right)^{\nu-2k}. \tag{A.1.9}$$

Using 3.1.1 of Nishimoto,

$$\frac{d^\nu}{dt^\nu}\mathrm{FT}^R_f(t) = (2\pi)^{-1/2}(-i)^\nu\int x^\nu f(x)e^{-itx}\,dx.$$

Hence,

$$E(X^\nu) = (2\pi)^{1/2}(-i)^{-\nu}\left[\frac{d^\nu}{dt^\nu}\mathrm{FT}^R_{\phi_\sigma(\cdot-\mu)}(t)\right]_{t=0}.$$

From the first part of (A.1.7) and (A.1.9) we obtain

$$\frac{d^\nu}{dt^\nu}\mathrm{FT}^R_{\phi_\sigma(\cdot-\mu)}(t) = \sigma^{-1}e^{-\mu/(2\sigma^2)}\phi_{1/\sigma}(t+i\mu/\sigma^2)\sum_{k=0}^{\infty}\frac{(-1)^k\Gamma(\nu+1)}{\Gamma(\nu+1-2k)k!2^k}\left(\frac{t+i\mu}{\sigma^2}\right)^{\nu-2k}(1/\sigma^2)^{k-\nu}.$$

Therefore,

$$\left[\frac{d^\nu}{dt^\nu}\mathrm{FT}^R_{\phi_\sigma(\cdot-\mu)}(t)\right]_{t=0} = i^\nu(2\pi)^{-1/2}\sum_{k=0}^{\infty}\frac{\Gamma(\nu+1)\mu^{\nu-2k}\sigma^{2k}}{\Gamma(\nu+1-2k)k!2^k}$$

and the result follows.

Proof of Corollary 3.1.

This follows from Theorem 3, $\Gamma(-r)=\infty$ for r a non-negative integer and (2.12).

Proof of Corollary 3.3.

Using integration by parts

$$\int x^r\phi_\sigma^{(r')}(x-\mu) = (-1)^{r'}\frac{r!}{(r-r')!}\int x^{r-r'}\phi_\sigma(x-\mu)\,dx.$$

Corollary 3.3 now follows from Corollary 3.1.

Proof of Theorem 4.

$$\begin{aligned}
\prod_{i=1}^m\phi_{\sigma_i}(x-\mu_i) &= (2\pi)^{-m/2}\left(\prod_{i=1}^m\sigma_i\right)^{-1}\exp\left\{-\tfrac12\sum_{i=1}^m\sigma_i^{-2}(x-\mu_i)^2\right\}\\
&= (2\pi)^{-m/2}\left(\prod_{i=1}^m\sigma_i\right)^{-1}\exp\left\{-\tfrac12\sum_{i=1}^m\sigma_i^{-2}\mu_i^2\right\}\exp\left\{-\tfrac12\sum_{i=1}^m\sigma_i^{-2}(x^2-2x\mu_i)\right\}\\
&= (2\pi)^{-m/2}\left(\prod_{i=1}^m\sigma_i\right)^{-1}\exp\left\{-\tfrac12\left(\sum_{i=1}^m\sigma_i^{-2}\mu_i^2+\tilde{\tilde{\mu}}\sum_{i=1}^m\sigma_i^{-2}\right)\right\}\\
&\qquad\times\exp\left\{-\tfrac12\sum_{i=1}^m\sigma_i^{-2}(x^2-2x\tilde{\tilde{\mu}}+\tilde{\tilde{\mu}}^2)\right\}\\
&= (2\pi)^{1-m/2}\left(\prod_{i=1}^m\sigma_i\right)^{-1}\phi_{\tilde{\sigma}}(\tilde{\tilde{\mu}})\phi_{\tilde{\tilde{\sigma}}}(x-\tilde{\tilde{\mu}}).
\end{aligned}$$

Proof of Theorem 5.

First note that by 7.2c of Rudin (1973), by (A.1.5), and by Corollary 4.2,

$$\begin{aligned}
&\mathrm{FT}^R_{\phi_\sigma^{(r)}(\cdot-\mu)*_R\phi_{\sigma'}^{(r')}(\cdot-\mu')}(t)\\
&\quad= \mathrm{FT}^R_{\phi_\sigma^{(r)}(\cdot-\mu)}(t)\mathrm{FT}^R_{\phi_{\sigma'}^{(r')}(\cdot-\mu')}(t)\\
&\quad= i^{r+r'}t^{r+r'}e^{-it(\mu+\mu')}\phi_{1/\sigma}(t)\phi_{1/\sigma'}(t)/(\sigma\sigma')\\
&\quad= (2\pi)^{-1/2}i^{r+r'}t^{r+r'}e^{-it(\mu+\mu')}\phi_{(\sigma^2+\sigma'^2)^{-1/2}}(t)/(\sigma^2+\sigma'^2)^{1/2}.
\end{aligned}$$

But now using (A.1.5) the other way around,

$$\mathrm{FT}^{R}_{\phi_\sigma^{(r)}(\cdot-\mu)*_R\phi_{\sigma'}^{(r')}(\cdot-\mu')}(t) = (2\pi)^{-1/2}\mathrm{FT}^{R}_{\phi^{(r+r')}_{(\sigma^2+\sigma'^2)^{1/2}}(\cdot-\mu-\mu')}(t).$$

Theorem 5 now follows from the Inversion Theorem 7.7c in Rudin (1973).

Proof of Corollary 5.2.

Using integration by substitution and the (skew) symmetry of $\phi^{(r')}$,

$$\begin{aligned}\int \phi_\sigma^{(r)}(x-\mu)\phi_{\sigma'}^{(r')}(x-\mu')\,dx &= \int \phi_\sigma^{(r)}(t)\phi_{\sigma'}^{(r')}(t+\mu-\mu')\,dt\\ &=(-1)^r\int \phi_\sigma^{(r)}(t)\phi_{\sigma'}^{(r')}(\mu-\mu'-t)\,dt\\ &=(-1)^r\phi_\sigma^{(r)} * \phi_{\sigma'}^{(r')}(\mu-\mu').\end{aligned}$$

So using Theorem 5,

$$\int \phi_\sigma^{(r)}(x-\mu)\phi_{\sigma'}^{(r')}(x-\mu')\,dx = (-1)^r\phi^{(r+r')}_{(\sigma^2+\sigma'^2)^{1/2}}(\mu-\mu').$$

Proof of Theorem 6.1.

$$\begin{aligned}&\int\prod_{i=1}^m \phi_{\sigma_i}^{(r_i)} = \int\prod_{i=1}^m \phi_{\sigma_i}^{(r_i)}(x-\mu_i^\dagger)\,dx\\
&= (-1)^{\sum_{i=1}^m r_i}\left(\prod_{i=1}^m \sigma_i^{-r_i-1}\right)\int\prod_{i=1}^m H_{r_i}\left(\frac{x-\mu_i^\dagger}{\sigma_i}\right)(2\pi)^{1-m/2}\phi_{\tilde\sigma}(\tilde\mu)\phi_{\tilde{\tilde\sigma}}(x)\,dx\\
&= (-1)^{\sum_{i=1}^m r_i}(2\pi)^{1-m/2}\phi_{\tilde\sigma}(\tilde\mu)\left(\prod_{i=1}^m \sigma_i^{-r_i-1}\right)\sum_{i_1=0}^{[r_1/2]}\cdots\sum_{i_m=0}^{[r_m/2]}(-1)^{\sum_{k=1}^m i_k}\\
&\times\left\{\prod_{j=1}^m \mathrm{OF}(2i_j)\binom{r_j}{2i_j}\right\}\int\prod_{j=1}^m\left(\frac{x-\mu_j^\dagger}{\sigma_j}\right)^{r_j-2i_j}\phi_{\tilde{\tilde\sigma}}(x)\,dx\end{aligned}$$

(where $[\cdot]$ is the greatest integer function)

$$\begin{aligned}&= (2\pi)^{1-m/2}\phi_{\tilde\sigma}(\tilde\mu)\left(\prod_{i=1}^m \sigma_i^{-r_i-1}\right)\sum_{i_1=0}^{[r_1/2]}\cdots\sum_{i_m=0}^{[r_m/2]}\sum_{j_1=0}^{r_1-2i_1}\cdots\sum_{j_m=0}^{r_m-2i_m}(-1)^{\sum_{k=1}^m i_k}\\
&\times\left\{\prod_{k=1}^m \mathrm{OF}(2i_k)\binom{r_k}{2i_k}\binom{r_k-2i_k}{j_k}(\mu_k^\dagger/\sigma_k)^{r_k-2i_k-j_k}\sigma_k^{-j_k}\right\}\tilde{\tilde\sigma}^{\sum_{\ell=1}^m j_\ell}\mathrm{OF}\left(\sum_{\ell=1}^m j_\ell\right)\\
&= (2\pi)^{1-m/2}\phi_{\tilde\sigma}(\tilde\mu)\sum_{j_1=0}^{r_1}\cdots\sum_{j_m=0}^{r_m}\prod_{k=1}^m\binom{r_k}{j_k}\sigma_k^{-r_k-j_k-1}\end{aligned}$$

$$\times \sum_{i_k=0}^{[(r_k-j_k)/2]} \left\{(-1)^{i_k}\mathrm{OF}(2i_k)\binom{r_k-j_k}{2i_k}(\mu_k^\dagger/\sigma_k)^{r_k-j_k-2i_k}\right\}\tilde{\tilde{\sigma}}^{\sum_{\ell=1}^m j_\ell}\mathrm{OF}\left(\sum_{\ell=1}^m j_\ell\right)$$

$$= (2\pi)^{1-m/2}\phi_{\tilde{\sigma}}(\tilde{\mu})\sum_{j_1=0}^{r_1}\cdots\sum_{j_m=0}^{r_m}\prod_{k=1}^m\left\{\binom{r_k}{j_k}\sigma_k^{-r_k-j_k-1}H_{r_k-j_k}(\mu_k^\dagger/\sigma_k)\right\}\tilde{\tilde{\sigma}}^{\sum_{\ell=1}^m j_\ell}\mathrm{OF}\left(\sum_{\ell=1}^m j_\ell\right).$$

Proof of Theorem 6.2.

Set $\tilde{\sigma} = 2^{1/2}$, $\tilde{\mu} = \mu_1 - \mu_2$, $\tilde{\tilde{\sigma}} = 2^{-1/2}$ and $\tilde{\tilde{\mu}} = (\mu_1+\mu_2)/2$. Then by Corollaries 4.2 and 3.1,

$$\begin{aligned}
\int x^k\phi^{(r)}(x)\phi^{(r')}(x)\,dx =&(-1)^{r+r'}\left[\frac{d^{r+r'}}{d\mu_1^r d\mu_2^{r'}}\int x^k\phi(x-\mu_1)\phi(x-\mu_2)\,dx\right]_{\mu_1=\mu_2=0}\\
=&(-1)^{r+r'}\left[\frac{d^{r+r'}}{d\mu_1^r d\mu_2^{r'}}\left\{\phi_{\tilde{\sigma}}(\tilde{\mu})\int x^k\phi_{\tilde{\tilde{\sigma}}}(x-\tilde{\tilde{\mu}})\,dx\right\}\right]_{\mu_1=\mu_2=0}\\
=&(-1)^{r+r'}\left[\frac{d^{r+r'}}{d\mu_1^r d\mu_2^{r'}}\left\{\phi_{\tilde{\sigma}}(\tilde{\mu})(-i\tilde{\tilde{\sigma}})^kH_k(i\tilde{\tilde{\mu}}/\tilde{\tilde{\sigma}})\right\}\right]_{\mu_1=\mu_2=0}\\
=&(-1)^{r+r'}\left[\frac{d^{r+r'}}{d\mu_1^r d\mu_2^{r'}}\left\{\phi_{\tilde{\sigma}}(\tilde{\mu})\sum_{j=0}^{[k/2]}2^{j-k}\binom{k}{2j}\mathrm{OF}(2j)(\mu_1+\mu_2)^{k-2j}\right\}\right]_{\mu_1=\mu_2=0}\\
=&(-1)^{r+r'}\sum_{p=0}^{r}\sum_{p'=0}^{r'}\binom{r}{p}\binom{r'}{p'}\left[\frac{d^{r+r'-p-p'}}{d\mu_1^{r-p}d\mu_2^{r'-p'}}\phi_{\tilde{\sigma}}(\tilde{\mu})\right]_{\mu_1=\mu_2=0}\\
&\times\left[\frac{d^{p+p'}}{d\mu_1^p d\mu_2^{p'}}\sum_{j=0}^{[k/2]}2^{j-k}\binom{k}{2j}\mathrm{OF}(2j)(\mu_1+\mu_2)^{k-2j}\right]_{\mu_1=\mu_2=0}
\end{aligned}\tag{A.1.10}$$

Repeated application of (2.6) shows that for $r+r'\equiv p+p' \pmod 2$,

$$\left[\frac{d^{r+r'-p-p'}}{d\mu_1^{r-p}d\mu_2^{r'-p'}}\phi_{\tilde{\sigma}}(\tilde{\mu})\right]_{\mu_1=\mu_2=0} = (-1)^{(r'-r+p-p')/2}2^{-(r'+r-p-p'+2)/2}\pi^{-1/2}\mathrm{OF}(r+r'-p-p').$$

By expansion we obtain

$$\left[\frac{d^{p+p'}}{d\mu_1^p d\mu_2^{p'}}(\mu_1+\mu_2)^{k-2j}\right]_{\mu_1=\mu_2=0} = \begin{cases}(p+p')! & p+p'=k-2j\\ 0 & \text{otherwise.}\end{cases}$$

Hence, for $p+p'\equiv k \pmod 2$,

$$\left[\frac{d^{p+p'}}{d\mu_1^p d\mu_2^{p'}}\sum_{j=0}^{[k/2]}2^{j-k}\binom{k}{2j}\mathrm{OF}(2j)(\mu_1+\mu_2)^{k-2j}\right]_{\mu_1=\mu_2=0}$$

$$= 2^{-(p+p'-k)/2}\binom{k}{k-p-p'}\mathrm{OF}(k-p-p')(p+p')!.$$

Combining these, (A.1.10) becomes

$$\begin{aligned}
&(-1)^{(r'-r+2k)/2}2^{-(r'+r+k+2)/2}\pi^{-1/2}\sum\sum\nolimits_{(p,p')\in\mathcal{A}}\binom{r}{p}\binom{r'}{p'}(-1)^{(p-p')/2}\mathrm{OF}(r+r'-p-p')\\
&\qquad\times\binom{k}{k-p-p'}\mathrm{OF}(k-p-p')(p+p')!\\
=&(-1)^{(r'-r+2k)/2}2^{-(r'+r+k+2)/2}\pi^{-1/2}k!\\
&\quad\times\sum\sum\nolimits_{(p,p')\in\mathcal{A}}\binom{r}{p}\binom{r'}{p'}(-1)^{(p-p')/2}2^{(p+p')/2}\frac{\mathrm{OF}(r+r'-p-p')}{\{(k-p-p')/2\}!}
\end{aligned}$$

as required.

Proof of Theorem 6.3.

By (2.7),

$$\phi_\sigma^{(1)}(x) = -x\phi_\sigma(x)/\sigma^2, \qquad \phi_{\sigma'}^{(1)}(x) = -x\phi_{\sigma'}(x)/{\sigma'}^2.$$

Hence by Corollary 4.3,

$$\phi_\sigma^{(1)}(x)\phi_{\sigma'}^{(1)}(x) = (\sigma\sigma')^{-2}x^2\phi_{\sigma^*}(0)\phi_{\sigma\sigma'/\sigma^*}(x).$$

By Corollary 3.2,

$$\begin{aligned}
\int x^2\phi_\sigma^{(1)}(x)\phi_{\sigma'}^{(1)}(x)\,dx =&(\sigma\sigma')^{-2}\phi_{\sigma^*}(0)(\sigma\sigma'/\sigma^*)^4\mathrm{OF}(2)\\
=&3(2\pi)^{-1/2}(\sigma\sigma')^2/(\sigma^2+{\sigma'}^2)^{5/2}.
\end{aligned}$$

REFERENCES

Bhattacharya, R. N. and Ranga Rao, R. (1976). *Normal Approximations and Asymptotic Expansions.* Wiley, New York.

Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series and Products.* Academic Press, San Diego.

Hall, P., Marron, J. S. and Park, B. U. (1992). Smoothed cross-validation. *Prob. Theor. Rel. Fields*, **92**, 1–20.

Hall, P., Sheather, S. J., Jones, M. C. and Marron, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, **78**, 263–269.

Jones, M. C., Marron, J. S. and Park, B. U. (1991). A simple root-n bandwidth selector. *Annals of Statistics*, **19**, 1919–1932.

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *Annals of Statistics*, **20**, 712–736.

Nishimoto, K. (1984). *Fractional Calculus.* Descartes Press, Koriyama, Japan.

Park, B. U. and Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.*, **85**, 66–72.

Patel, J. K. and Read, C. B. (1982). *Handbook of the Normal Distribution.* Dekker, New York.

Rudin, W. (1973). *Functional Analysis.* MacGraw Hill, New York.