Extension of Stein’s lemma derived by using an integration by differentiation technique

Konstantinos Mami\textsuperscript{s}

\textsuperscript{*}Department of Mathematics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27695-8205, USA

Abstract

We extend Stein’s lemma for averages that explicitly contain the Gaussian random variable at a power. We present two proofs for this extension of Stein’s lemma, with the first being a rigorous proof by mathematical induction. The alternative, second proof is a constructive formal derivation in which we express the average not as an integral, but as the action of a pseudodifferential operator defined via the Gaussian moment-generating function. In extended Stein’s lemma, the absolute values of the coefficients of the probabilist’s Hermite polynomials appear, revealing yet another link between Hermite polynomials and normal distribution.

Keywords: normal distribution, Stein’s lemma, Hermite polynomials, generalized factorial coefficients, pseudodifferential operator

2000 MSC: primary 60E05, secondary 60E10, 47G30, 05A10

1. Introduction and main results

Stein’s lemma [12] is a celebrated result in probability theory with many applications in statistics, see e.g. [8]. For a scalar, zero-mean Gaussian random variable $X$ with variance $\sigma^2$, Stein’s lemma reads

$$\mathbb{E}[g(X)X] = \sigma^2 \mathbb{E}[g'(X)],$$

(1)

where $\mathbb{E}[\cdot]$ is the mean value operator, and prime denotes the first derivative of function $g$. By Theorem 1 of the present work, we extend Stein’s lemma for the average $\mathbb{E}[g(X)X^n]$, which is expressed as a finite series that contains averages of derivatives of $g(X)$ up to the $n$th order. We present two ways of proving Theorem 1; in Sec. 2, the extended Stein’s lemma is proven rigorously by mathematical induction on index $n$ of power $X^n$ inside the average. Furthermore, and in order to provide the reader with more insight, we present an alternative, constructive proof in Sec. 3. The constructive proof is based on a formal expression of the average as the action of a pseudodifferential operator, introduced by Definition 1 via the Gaussian moment-generating function. Series expansion of this pseudodifferential operator allows us to treat the averages not as integrals, but as Taylor-like infinite series. Thus, and under the formal assumption that all infinite series involved are summable, we are able to calculate $\mathbb{E}[g(X)X^n]$ without performing any integrations; instead, we perform only the differentiations that appear in the series terms, justifying thus the name integration by differentiation for this technique.

**Theorem 1** (Extension of Stein’s lemma). Let $X$ be a scalar Gaussian variable with zero mean value and variance $\sigma^2$. Given an $n \in \mathbb{N}$, let also $g$ be a $C^n(\mathbb{R} \to \mathbb{R})$ function for which averages $\mathbb{E}[g(X)X^n]$ and $\mathbb{E}[g^{(n-2k)}(X)]$ for $k = 0, \ldots, \lfloor n/2 \rfloor$ exist, with $g^{(\ell)}$ denoting the $\ell$th derivative of $g$, $g^{(0)} := g$, and $\lfloor \cdot \rfloor$ being the floor function. It holds true that:

$$\mathbb{E}[g(X)X^n] = \sum_{k=0}^{\lfloor n/2 \rfloor} H_{n,k} \sigma^{2(n-k)} \mathbb{E}[g^{(n-2k)}(X)],$$

(2)

where

$$H_{n,k} = \frac{n!}{2^k k! (n-2k)!}, \quad k = 0, \ldots, \lfloor n/2 \rfloor,$$

(3)

are the absolute values of the coefficients appearing in the $n$th-order probabilist’s Hermite polynomial $H_n(x)$ [1, expression 22.3.11]:

$$H_n(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k H_{n,k} x^{n-2k}.$$  

(4)

In the On-Line Encyclopedia of Integer Sequences (OEIS) [9], $H_{n,k}$ are referred to as the Bessel numbers. In this work, we shall call $H_{n,k}$ the signless Hermite coefficients, as a more suggestive term in our context.

**Proof.** Theorem 1 is proven by mathematical induction, performed in Sec. 2. In Sec. 3, we also present a constructive derivation of Eq. (2) that employs an integration by differentiation technique. $\Box$

**Remark 1** (Orders of derivatives in Eq. (2)). We easily observe that for an even (odd) power $n$ of $X$, only the averages of the even (odd)-order derivatives of $g(X)$, up to the $n$th order, appear in the right-hand side of Eq. (2). The average of $g(X)$ itself appears in Eq. (2) only for even powers.

Furthermore, for the complete treatment of $\mathbb{E}[g(X)X^n]$, we also have to consider the case where the mean value of $X$ is non-zero. Thus, we state the following Theorem:
Theorem 2 (Extension of Stein’s lemma for non-zero mean). For a Gaussian variable $X$ with non-zero mean value $\mu$, the extended Stein’s lemma reads

$$
\mathbb{E}[g(X)X^{n}] = \sum_{k=0}^{n} \binom{n}{k} \mu^{n-k} \sum_{\ell=0}^{\lfloor k/2 \rfloor} H_{\ell,k} \sigma^{2(n-k)} \mathbb{E}\left[ g^{(\ell)}(X) \right],
$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient.

Proof. Eq. (5) can be proven by mathematical induction, similar to the one of Sec. 2 for Theorem 1. Alternatively, one can perform a constructive proof for Eq. (5) similar to the one of Sec. 3 for Theorem 1, by expressing the averages via Eq. (B.3).

The formulas of extended Stein’s lemma, stated in Theorems 1, 2, are useful in engineering applications where averages $\mathbb{E}[g(X)X^n]$ arise, see e.g. the recent work [11] in tribology. Note that the respective extensions of Stein’s lemma for multivariate Gaussian variables are also feasible and will be the topic of a forthcoming work.

Remark 2 (The need for extended Stein’s lemma). One could argue that the evaluation of averages $\mathbb{E}[g(X)X^n]$ can always be performed by repetitive applications of the classical Stein’s lemma. However, the advantage of closed-form formulas (2), (5) over the recursive ones is manifested for increasing $n$, see e.g. [11, Sec. 2.3], where keeping track of the repetitive applications of Eq. (1) becomes progressively more difficult. On the other hand, the averages and coefficients appearing in the right-hand sides of Eqs. (2), (5) are a priori known for any $n$, offering thus an easy and tractable calculation.

Corollary 1 (Central moments of a zero-mean Gaussian variable). The higher order moments of a zero-mean Gaussian variable $X$ are given by the formula

$$
\mathbb{E}[X^n] = \begin{cases} 
0 & \text{for } n \text{ odd}, \\
\sigma^n (n-1)!! & \text{for } n \text{ even}.
\end{cases}
$$

Proof. Eq. (6) is usually derived by repetitive integrations by parts, see e.g. [10, p. 148]. Here, we easily derive it from Eq. (2), by considering $g(x) = 1$. In this case, all derivatives of $g$ appearing in the right-hand side of Eq. (2) are zero except for the zeroth-order one, which equals to 1. Since Eq. (2) does not contain $\mathbb{E}[g(X)]$ for odd $n$ (see Remark 1), all odd moments of $X$ are equal to zero. For even $n = 2\ell$, zeroth-order derivative appears for $k = \ell$, and Eq. (2) for $g(x) = 1$ reads

$$
\mathbb{E}[X^{2\ell}] = H_{\ell,\ell} \sigma^{2\ell} = \frac{(2\ell)!}{2^\ell} \sigma^{2\ell}.
$$

Then, by using factorial relations $n! = n!(n-1)!!$, and $n!! = 2^\ell \ell!$ for $n = 2\ell$, we obtain the branch of Eq. (6) for even $n$.

2. Proof of Theorem 1 by mathematical induction

Theorem 1 can be proven by mathematical induction. For $n = 1$, it is easy to see that Eq. (2) results in Stein’s lemma (1), or

$$
\mathbb{E}[g(X)X^{n}] = \sum_{k=0}^{\lfloor n/2 \rfloor} H_{n,k} \sigma^{2(n-k)} \mathbb{E}\left[ g^{(n-k)}(X) \right],
$$

since, from Eq. (3), $H_{1,0} = 1$. We then assume that Eq. (2) holds true for a particular $n$ (inductive hypothesis). Now, for $n + 1$, and by using the inductive hypothesis, we obtain:

$$
\mathbb{E}[g(X)X^{n+1}] = \mathbb{E}[g(X)X^n] = \sum_{k=0}^{\lfloor n/2 \rfloor} H_{n,k} \sigma^{2(n-k)} \mathbb{E}\left[ g^{(n-k)}(X) \right].
$$

For the evaluation of the derivative inside the average in the right-hand side of Eq. (7), we use the general Leibniz rule [1, expression 3.3.8]:

$$
(g(x)x)^{(n-2k)} = \sum_{\ell=0}^{n-2k} \binom{n-2k}{\ell} g^{(n-2k-\ell)}(x) x^{\ell}.
$$

Since $x^{(0)} = x$, $x^{(1)} = 1$, and $x^{(\ell)} = 0$ for $\ell \geq 2$, Eq. (8) is simplified into

$$
(g(x)x)^{(n-2k)} = g^{(n-2k)}(x) x + (n-2k)g^{(n-2k-1)}(x),
$$

under the convention that $g^{(-1)}(x) = 0$. By using Leibniz rule (9) in Eq. (7), we have

$$
\mathbb{E}[g(X)X^{n+1}] = \sum_{k=0}^{n/2} H_{n,k} \sigma^{2(n-k)} \mathbb{E}\left[ g^{(n-k)}(X) \right],
$$

and

$$
\sum_{k=0}^{\lfloor n/2 \rfloor} (n-2k)H_{n,k} \sigma^{2(n-k)} \mathbb{E}\left[ g^{(n-k-1)}(X) \right].
$$

Note that the upper limit of $k$-sum in the right-hand side of Eq. (12) has been changed from $n/2$ to $(n-1)/2$, in order to exclude any term containing $g^{(-1)}$. The only such term is for even $n$ and $k = \lfloor n/2 \rfloor$. Thus, the upper limit of $k$-sum is $n/2$ for $n$ odd and $n/2 - 1$ for $n$ even. These two values can be expressed in a unified way as $(n-1)/2$. By also performing a change in index, and using the fact that $\lfloor (n+1)/2 \rfloor = \lfloor (n-1)/2 \rfloor + 1$, $\Sigma_2$ reads

$$
\sum_{k=1}^{\lfloor (n+1)/2 \rfloor} (n-2k+2)H_{n,k-1} \sigma^{2(n-k+1-k)} \mathbb{E}\left[ g^{(n-k+1-2k)}(X) \right].
$$

For $\Sigma_1$, we apply Stein’s lemma (1) again, at the average appearing in the right-hand side of Eq. (11), resulting into

$$
\sum_{k=0}^{n/2} H_{n,k} \sigma^{2(n-k+1-k)} \mathbb{E}\left[ g^{(n-k+1-2k)}(X) \right].
$$

Under Eqs. (13), (14), the right-hand side of Eq. (10) is rearranged into

$$
\mathbb{E}[g(X)X^{n+1}] = \sum_{k=0}^{\lfloor (n+1)/2 \rfloor} \left[ (n-2k+2)H_{n,k-1} + H_{n,k} \right] \times \sigma^{2(n-k+1-2k)} \mathbb{E}\left[ g^{(n-k+1-2k)}(X) \right],
$$

under the convention that $H_{n,k} = 0$ for $k < 0$ or $k > n/2$.
Lemma 1 (Recurrence relation for $H_{n,k}$). For $n \in \mathbb{N}$, $k = 0, \ldots, [(n + 1)/2]$, and with $H_{n,k} = 0$ for $k < 0$ or $k > n/2$, it holds true that

$$H_{n+1,k} = (n - 2k + 2)H_{n,k-1} + H_{n,k}.$$  

Proof. See Appendix A. Eq. (16) is also stated in the relevant OEIS entry [9].

Substitution of Eq. (16) into Eq. (15) results in

$$E\left[g(X)X^{n+1}\right] = \sum_{k=0}^{[(n+1)/2]} H_{n+1,k} \sigma^{2(n+1-k)} E\left[g^{n+1-2k}(X)\right] ,$$  

which is Eq. (2) for $n + 1$, completing thus the proof of Theorem 1 by induction.

3. Formal derivation of Theorem 1 using integration by differentiation

Stein’s lemma (1) is usually proven by integration by parts (see e.g. [12, lemma 1]), by employing the definition of mean value operator $E\lfloor \cdot \rceil$ as an integral over $\mathbb{R}$. That is, since random variable $X$ follows the univariate normal distribution $f(x) = (\sqrt{2\pi})^{-1} \exp(-x^2/2\sigma^2)$, we calculate:

$$E[g(X)X] = \int_{\mathbb{R}} g(x)xf(x)dx = -\sigma^2 \int_{\mathbb{R}} g'(x)f(x)dx = \sigma^2 \int_{\mathbb{R}} g'(x)f(x)dx = \sigma^2 E[g'(X)].$$

While one could also go on to try to find the adequate repetitive integrations by parts for the calculation of $E[g(X)X^n]$, here we shall follow a different, more tractable path, based on the interpretation of average not as an integral, but as the action of a pseudodifferential operator, defined as follows.

Definition 1 (Mean value as the action of an averaged shift operator). Let $X$ be a zero-mean Gaussian random variable with variance $\sigma^2$ and $g$ be a $C^\infty(\mathbb{R} \rightarrow \mathbb{R})$ function. Then, average $E[g(X)]$ is expressed as

$$E[g(X)] = \exp\left(\frac{\sigma^2}{2} \frac{d^2}{dx^2}\right) g(x)\Bigg|_{x=0}.$$  

In Eq. (19), $\exp\left(\frac{\sigma^2}{2} \frac{d^2}{dx^2}\right)$ is a pseudodifferential operator, called the averaged shift operator (see Appendix B), whose action is to be understood by its series form

$$E[g(X)] = \sum_{m=0}^{\infty} \frac{\sigma^{2m}}{2^m m!} \frac{d^m g(x)}{dx^m}\Bigg|_{x=0} ,$$

with $\downarrow_{x=0}$ denoting that all derivatives appearing in the right-hand side of Eq. (20) are calculated at $x = 0$.

Proof. Definition 1 is formally derived in Appendix B by employing the moment-generating function of $X$. Its infinite dimensional version is presented in [2], and this concept is also found in [7, Ch. 4].

Remark 3 (Integration by differentiation). Choosing Eq. (20) for the evaluation of $E[g(X)X^n]$ is more convenient than integration by parts, since, by using Eq. (20), only the calculation of derivatives is needed. Other integration by differentiation techniques have also been recently proposed, see e.g. [6], exploiting the fact that differentiation is generally easier than integration.

Thus, under the additional assumption that $g$ is $C^\infty(\mathbb{R} \rightarrow \mathbb{R})$, we express $E[g(X)X^n]$ via Eq. (20) as

$$E[g(X)X^n] = \sum_{m=0}^{\infty} \frac{\sigma^{2m}}{2^m m!} \frac{d^m [g(x)x^n]}{dx^{2m}}\bigg|_{x=0} .$$

We evaluate the derivatives appearing in the right-hand side of Eq. (21) by using the general Leibniz rule

$$\frac{d^m [g(x)x^n]}{dx^{2m}} = \sum_{\ell=0}^{m} \binom{m}{\ell} g^{2m-\ell}(x)(x^n)^{(\ell)} ,$$

and since $(x^n)^{(\ell)} = (n!(n-\ell))x^{n-\ell}$ for $n \geq \ell$ and zero for $n < \ell$:

$$\frac{d^m [g(x)x^n]}{dx^{2m}} = \sum_{\ell=0}^{\min(m,n)} \binom{m}{\ell} \frac{n!}{(n-\ell)!} g^{2m-\ell}(x)x^{n-\ell} .$$

(23)

Note that, in Eq. (21), derivative (23) has to be calculated for $x = 0$. Since, for $x = 0$, all terms of the sum in the right-hand side of (23) are zero except for $\ell = n$, we obtain:

$$\frac{d^m [g(x)x^n]}{dx^{2m}}\bigg|_{x=0} = \left\{ \begin{array}{ll} 0 & \text{for } m < n/2, \\ \frac{(2m)!}{2^{m-n} m!} g^{2m-n}(0) & \text{for } m \geq n/2. \end{array} \right.$$  

(24)

By substitution of expression (24) into Eq. (21), and after some algebraic manipulations, we have

$$E[g(X)X^n] = \sum_{m=\max(0,\lfloor n/2 \rfloor)}^{\infty} \frac{\sigma^{2m}}{2^m m!} \frac{(2m)!}{m!} g^{2m-n}(0)\bigg|_{x=0} ,$$

(25)

where $\lfloor x \rfloor$ denotes the ceiling function and $(2m)! = (2m)(2m-1) \cdots (2m-n + 1)$ is the falling factorial. In order to evaluate further the right-hand side of Eq. (25), the falling factorial of $2m$ has to be expressed in terms of the falling factorial of $m$. Following Charalambides [4, Sec. 8.4], this is performed by using the generalized factorial coefficients with parameter $2$ $C(n, \ell; 2)$, that have the property

$$(2m)! = \sum_{\ell=0}^{m} C(n, \ell; 2) m^{2\ell}. \tag{26}$$

Generalized factorial coefficients $C(n, \ell; 2)$ are defined in terms of the Stirling numbers of first $s(n, k)$ and second $S(n, k)$ kind [4, theorem 8.13]:

$$C(n, \ell; 2) = \sum_{k=\ell}^{n} 2^k s(n, k) S(k, \ell), \tag{27}$$

and obey the following recurrence relation for $n \in \mathbb{N}$, $\ell = 1, 2, \ldots, n + 1$ [4, theorem 8.19]:

$$C(n + 1, \ell; 2) = (2\ell - n)C(n, \ell; 2) + 2C(n, \ell - 1; 2). \tag{28}$$
with initial conditions \( C(0, 0; 2) = 1, C(n, 0; 2) = 0 \) for \( n > 0 \), and \( C(n, \ell; 2) = 0 \) for \( \ell > n \). Also, since \( s(n, n) = S(n, n) = 1 \), see [3, proposition 5.3.2], we deduce from Eq. (27) that
\[
C(n, n; 2) = 2^n, \quad \text{for } n \in \mathbb{N}. \tag{29}
\]
Furthermore, in [4, remark 8.8], it is proved that \( C(n, \ell; 2) = 0 \) for \( \ell < n/2 \). This property results in Eq. (26) to be updated into
\[
(2m)! = \sum_{\ell = [n/2]}^{n} C(n, \ell; 2)m^\ell \tag{30}
\]
Last, from the definition of falling factorial, \( m^\ell \) is zero for \( n > m \), and thus Eq. (30) is finally expressed as
\[
(2m)! = \sum_{\ell = [n/2]}^{\min(n,m)} C(n, \ell; 2)m^\ell \tag{31}
\]
Substituting Eq. (31) into Eq. (25), and use of \( m^\ell /m! = 1/(m - \ell)! \) results in
\[
\mathbb{E}[g(X)^m] = \sum_{m=0}^{\infty} \sum_{\ell=[m/2]}^{m} C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0} = \frac{n}{\ell} \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0} + \frac{1}{\ell} \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} C(n, \ell: 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0}. \tag{32}
\]
Double summations in the rightmost side of Eq. (32) are easily rearranged into
\[
\mathbb{E}[g(X)^m] = \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0} + \frac{1}{\ell} \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0} = \frac{1}{\ell} \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} \frac{C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0}}{\ell} = \frac{1}{\ell} \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} \frac{C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0}}{\ell}. \tag{33}
\]
An index change in the second sum, as well as the use of formula (20), results in
\[
\mathbb{E}[g(X)^m] = \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} \frac{C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0}}{\ell} = \frac{1}{\ell} \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} \frac{C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0}}{\ell} = \frac{1}{\ell} \sum_{\ell=2}^{n} \sum_{m=\ell}^{n} \frac{C(n, \ell; 2) \frac{\sigma_2^{2m}}{(m - \ell)!} \frac{g^{(2m-n)}(x)}{2^m} |_{x=0}}{\ell}. \tag{34}
\]
By also performing the change of summation index \( k = n - \ell \) in Eq. (34), we obtain
\[
\mathbb{E}[g(X)^m] = \sum_{k=0}^{[n/2]} \frac{C(n, n - k; 2) \frac{\sigma_2^{2(n-k)}}{2^{n-k}} \frac{g^{(2n-2k)}(x)}{2^m} |_{x=0}}{2^{n-k}}. \tag{35}
\]
Eq. (35) is of the same form as the extended Stein’s lemma, Eq. (2). What remains to be proven is the identification of \( C(n, n-k; 2)/2^{n-k} \) as the signless Hermitian coefficients \( H_{n,k} \). This is performed in the following Lemma, which concludes the constructive formal derivation of Theorem 1.

**Lemma 2** (Signless Hermite coefficients as rearranged, rescaled, generalized factorial coefficients with parameter 2). For \( H_{n,k} \) being the signless Hermite coefficients defined by Eq. (3) and \( C(n, \ell; 2) \) being the generalized factorial coefficients with the property (26), it holds true that
\[
H_{n,k} = \frac{C(n, n-k; 2)}{2^{n-k}}, \quad \text{for } n \in \mathbb{N}, \quad k = 0, \ldots, \lfloor n/2 \rfloor. \tag{36}
\]

**Proof.** See Appendix C. To the best of our knowledge, relation (36) has not been pointed out before. \( \square \)

**Declaration of Competing Interest**

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgement**

Part of this work was conducted while the author was a PhD candidate in the National Technical University of Athens, under the supervision of professor G.A. Athanassoulis.

**Appendix A. Proof of Lemma 1**

By using definition relation (3), it is easily calculated that
\[
H_{n,0} = H_{n+1,0} = 1. \tag{A.1}
\]
Since, by convention, \( H_{n,-1} = 0 \), Eq. (A.1) coincides with recurrence relation (16) for \( k = 0 \). For \( k = 1, \ldots, \lfloor n/2 \rfloor \), we have
\[
(n - 2k + 2)H_{n,k-1} + H_{n,k} = \frac{2k! + n - 2k + 1}{2^k}! = \frac{(n + 1)!}{2^k!} \frac{n - 2k + 1}{2^k!}! = H_{n+1,k}. \tag{A.2}
\]
Eqs. (A.1) and (A.2) constitute the proof of recurrence relation (16) for even \( n \) and \( k = 0, \ldots, \lfloor n/2 \rfloor \), since \( \lfloor (n + 1)/2 \rfloor = \lfloor n/2 \rfloor \) for even \( n \). For odd \( n = 2\ell + 1 \), Eq. (16) has to be also proven for \( k = \lfloor (n + 1)/2 \rfloor = \ell + 1 \):
\[
H_{2\ell+1,\ell} = \frac{(2\ell + 1)!}{2\ell!} = \frac{(2\ell + 2)!}{2\ell!}, \tag{A.3}
\]
Since, by convention, \( H_{2\ell+1,\ell+1} = 0 \), we can easily see that Eq. (A.3) coincides with recurrence relation (16) for \( n = 2\ell + 1, k = \ell + 1 \). Thus, the proof of recurrence relation (16) for both odd and even \( n \), and for \( k = 0, \ldots, \lfloor (n + 1)/2 \rfloor \) is completed.

**Remark 4.** Note that \( H_{n+1,0} = 1 \) is the initial condition supplementing recurrence relation (16). Linear recurrence relation (16) and initial condition (A.1), under the convention that \( H_{n,k} = 0 \) for \( k < 0 \) or \( k > n/2 \), constitute a complete definition for signless Hermite coefficients \( H_{n,k} \), whose unique solution is formula (3) (see [4, Sec. 7.2]).
Appendix B. Formal derivation of Definition 1

First, we recall that, for a deterministic $C^\infty(\mathbb{R} \to \mathbb{R})$ function $g(x)$, its Taylor series expansion around $x_0$ is alternatively expressed via the translation (shift) pseudodifferential operator (see e.g. [5, Sec. 1.1]), first introduced by Lagrange:

$$g(x) = \sum_{m=0}^{\infty} \frac{\hat{x}^m}{m!} \frac{d^m g(x)}{dx^m} \bigg|_{x=x_0} \exp \left( \frac{d}{dx} \right) g(x) \bigg|_{x=x_0},$$

where $\hat{x} = x - x_0$ is the shift argument. Now, we substitute $X$ as the argument of function $g$, with $X$ being a random variable with non-zero, in general, mean value $\mu$. By choosing $x_0 = \mu$, and taking the average in both sides, Eq. (B.1) results into

$$E[g(X)] = E \left[ \exp \left( \frac{\hat{X} d}{dx} \right) g(x) \right]_{\mu} = M_{\hat{X}} \left( \frac{d}{dx} \right) g(x) \bigg|_{\mu}.$$  

In Eq. (B.2), $M_{\hat{X}}(u)$ is identified as the moment-generating function of centered variable $\hat{X} := X - \mu$ [10, Sec. 5.5]; $M_{\hat{X}}(u) = E \left[ \exp \left( \hat{X} u \right) \right]$. Due to its resemblance to the shift operator of Eq. (B.1), we shall call pseudodifferential operator $M_{\hat{X}}(d/dx)$ the averaged shift operator. For a Gaussian variable $X$ with variance $\sigma^2$, the moment-generating function of $\hat{X}$ takes the form $M_{\hat{X}}(u) = \exp \left( \sigma^2 u^2 / 2 \right)$ [10, example 5.28], and Eq. (B.2) is specified into

$$E[g(X)] = \exp \left( \frac{\sigma^2}{2} \frac{d^2}{dx^2} \right) g(x) \bigg|_{\mu}.$$  

For the zero-mean value case, $\mu = 0$, Eq. (B.3) results in Eq. (19).

Appendix C. Proof of Lemma 2

By taking into consideration that $C(n, \ell; 2) = 0$ for $\ell < n/2$ [4, remark 8.8], recurrence relation (28) for generalized factorial coefficients with parameter $\ell$ is rewritten as

$$C(n + 1, \ell; 2) = (2\ell - n)C(n, \ell; 2) + 2C(n, \ell - 1; 2), \quad n \in \mathbb{N}, \quad \ell = [(n + 1)/2], \ldots, n + 1,$$

supplemented by the final condition deduced from Eq. (29):

$$C(n + 1, n + 1; 2) = 2^{n+1}, \quad n \in \mathbb{N}.$$  

By performing the index change $k = n + 1 - \ell$, Eq. (C.1) is expressed as

$$C(n + 1, n + 1 - k; 2) = (n - 2k + 2)C(n, n - (k - 1); 2) + 2C(n, n - k; 2), \quad k = 0, \ldots, [(n + 1)/2].$$  

Dividing both sides of Eq. (C.3) by $2^{n-1+k}$ results in

$$\frac{C(n + 1, n + 1 - k; 2)}{2^{n-1+k}} = \frac{(n - 2k + 2)C(n, n - (k - 1); 2)}{2^{n-(k-1)}} + \frac{C(n, n - k; 2)}{2^{n-k}}, \quad k = 0, \ldots, [(n + 1)/2].$$  

and dividing Eq. (C.2) by $2^n$ results in

$$\frac{C(n + 1, n + 1; 2)}{2^{n+1}} = 1, \quad n \in \mathbb{N}.$$  

Thus, recurrence relation (C.4) and initial condition (C.5) for $C(n, n - k; 2)/2^{n-k}$ coincides with the recurrence relation (B.1) and initial condition (A.1) for $H_{n,k}$. This constitutes the proof of Eq. (36).

References

[1] M. Abramowitz and I. A. Stegun. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications (1983 reprint), 10th edition, 1964.

[2] G.A. Alhamassouls and K.I. Mamis. Extensions of the Novikov-Furutsu theorem, obtained by using Volterra functional calculus. Physica Scripta, 94(11):151217, 2019. https://doi.org/10.1088/1402-4896/ab10b5

[3] P. J. Cameron. Combinatorics: Topics, Techniques, Algorithms. Cambridge University Press, Cambridge UK, 1994.

[4] Ch. A. Charalambides. Enumerative combinatorics. Chapman & Hall/CRC, Boca Raton, 2002.

[5] H.-J. Glaeske, A.P. Prudnikov, and K.A. Skôrnik. Operational Calculus and Related Topics. Chapman & Hall/CRC, Boca Raton, 2006.

[6] D. Jia, E. Tang, and A. Kempf. Integration by differentiation: new proofs, methods and examples. Journal of Physics A: Mathematical and Theoretical, 50:235201, 2017. https://doi.org/10.1088/1751-8121/aa6f32

[7] V. I. Klyatskin. Stochastic Equations through the eye of the Physicist. Elsevier, Amsterdam, 2005.

[8] N. Mukhopadhyay. On Rereading Stein’s Lemma: Its Intrinsic Connection with Cramér-Rao Identity and Some New Identities. Methodology and Computing in Applied Probability, 23:355–367, 2021. https://doi.org/10.1007/s11009-020-09830-w

[9] OEIS Foundation Inc. Triangle of Bessel numbers read by rows, Entry A100861 in The On-Line Encyclopedia of Integer Sequences. url: https://oeis.org/A100861

[10] A. Papoulis and S. U. Pillai. Probability, Random variables, and Stochastic processes. McGraw-Hill, New York, 4th edition, 2002.

[11] D. Skalskas, G. N. Rossopoulos and Ch. I. Papadopoulos. A Comparative Study of the Reynolds Equation Solution for Slider and Journal Bearings with Stochastic Roughness on the Stator and the Rotor. Tribology International, 167:107410, 2022. https://doi.org/10.1016/j.triboint.2021.107410

[12] Ch. M. Stein. Estimation of the Mean of a Multivariate Normal Distribution. The Annals of Statistics, 9(6):1135–1151, 1981. https://doi.org/10.1214/aos/1176345632