Journal of Statistical Planning and Inference 167 (2015) 174–181

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Expansions for multivariate densities

Ruby C. Weng

Department of Statistics, National Chengchi University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 23 April 2015
Received in revised form 4 May 2015
Accepted 5 May 2015
Available online 19 May 2015

Keywords:
Gram–Charlier series
Edgeworth series
Hermite polynomials
Woodroofe–Stein's identity

ABSTRACT

The Gram–Charlier and Edgeworth series are expansions of probability distribution in terms of its cumulants. The expansions for the multivariate case have not been fully explored. This paper aims to develop the multivariate Gram–Charlier series by Woodroofe–Stein's identity, and improve its approximation property by using the scaled normal density and Hermite polynomials. The series are useful to reconstruct the probability distribution from measurable higher moments.

1. Introduction

There exist several types of expansions of a univariate probability density function (pdf) in the orthogonal sets of Hermite polynomials. The expansions express the probability distribution in terms of its moments or cumulants. Three well-known expansions are the Gram–Charlier, Gauss–Hermite, and Edgeworth series. The first two series differ from each other in using different sets of Hermite polynomials, while the Edgeworth series differs from the first two series in that it collects terms of the same order. So, the Edgeworth series is an asymptotic expansion, whereas the other two are not.

It is known that the Gram–Charlier series may diverge in many cases of interest, and that even the asymptotic Edgeworth series may not converge. For the one-dimensional case, if the pdf $p(z)$ is of bounded variation in $(-\infty, \infty)$ and the integral $\int_{-\infty}^{\infty} p(z)\exp(z^2/4)dz$ exists, the Gram–Charlier series converges; on the other hand, if the conditions are not satisfied, the expansions may diverge (Cramér, 1946, Chapter 17). Blinnikov and Moessner (1998) gave a comparison of the Gram–Charlier, Gauss–Hermite, and Edgeworth series. They showed that the Edgeworth expansion is the best among them; however, for strongly non-Gaussian cases, like χ_ν^2 with degrees of freedom $\nu = 2$, the Edgeworth series also diverges like the Gram–Charlier series.

Despite divergence in some cases, in many practical applications what we really concern is whether a small number of terms suffice to give a good approximation (Cramér, 1946). In fact, these expansions are useful to measure the deviations of a pdf from the normal pdf, to provide correction terms for density approximation, and to reconstruct a pdf by measurable higher order moments. They have been used in a variety of areas. For instance, Sargan (1975, 1976) introduced it into econometrics; Van Der Marel and Franx (1993), Scherrer and Bertschinger (1991), and Blinnikov and Moessner (1998) applied the Gram–Charlier, Gauss–Hermite, and Edgeworth expansions in astrophysics; Comon (1994) and Amari et al. (1996) used the Edgeworth series in Independent Component Analysis to approximate the one-dimensional differential entropy; Hall (1992) showed how Edgeworth expansion and bootstrap methods can help explain each other; Rubinstein (1998) employed the Edgeworth expansion to value derivatives, among others.

Though there are many applications based on the expansion for the one-dimensional probability density, relatively few studies are available for the multivariate densities. To the best of our knowledge, the most comprehensive account of the

E-mail address: chweng@nccu.edu.tw.

http://dx.doi.org/10.1016/j.jspi.2015.05.001
0378-3758/

expansions for the multivariate case is by Barndorff-Nielsen and Cox (1989). Basically, they take the traditional approach which inverts expansions of the characteristic functions by the inverse Fourier transform. Other studies for the multivariate cases include Csörgő and Hegyi (2000), Van Hulle (2005), among others.

Instead of inverting the characteristic function, Weng (2010) derived the Gram–Charlier type and Edgeworth type expansions of a univariate pdf based on a version of Stein's identity. The identity is essentially repeated integration by parts, which naturally leads to a series expansion. This version of Stein's identity was developed by Woodroofe (1989, 1992a) for obtaining integrable expansions for posterior distributions. It is closely related to the well-known Stein's lemma (Stein, 1981) — the latter considers the expectation with respect to a normal distribution, while the former the expectation with respect to a "nearly normal distribution" Γ in the sense of (7); and both are proved by an application of Fubini theorem. Stein's lemma (Stein, 1981) is famous for its applications to James–Stein estimator (James and Stein, 1961) and empirical Bayes methods. As for Stein's identity, it has been applied to set frequentist corrected confidence sets following sequentially designed experiments (e.g. Woodroofe, 1992b; Coad and Woodroofe, 1996; Weng and Woodroofe, 2006, among others), and to Bayesian inference (e.g. Weng, 2003; Baghishania and Mohammadzadeh, 2012, among others). Recently Weng and Lin (2011) applied this identity to derive Bayesian online algorithms for ranking players; and to distinguish it from the well-known Stein's lemma, they coined it as *Woodroofe–Stein*'s identity. The identity may be further explored.

The main contribution of present paper is a theoretical advancement in generalizing Weng's (2010) approach to give a closed-form expression for the multivariate Gram–Charlier expansion. Furthermore, as the Gram–Charlier series suffers from poor convergence, a modified series is proposed for better convergence properties. The modification starts by suitably scaling the variable and applying the expansion to pdf of the scaled variable, and then converted back to the original variable.

The remaining of this paper is organized as the following. The next section provides some reviews. Section 3 presents the Gram–Charlier type expansion for multivariate densities and describes the proposed modified series. Section 4 concludes.

2. Reviews

2.1. Woodroofe–Stein's identity

We review the identity. Some results here will be generalized in Section 3.1. Some definitions and notations are needed. Let ϕ_p and Φ_p denote the density and distribution function of a standard p-variate normal distribution, and abbreviate Φ_1 and ϕ_1 as Φ and ϕ. For a function $h : R^p \to R$, we may write

$$\Phi_p h \equiv \int h d\Phi_p \tag{1}$$

for simplicity, provided the integral is finite. Let $h_0 = \Phi_p h$ be a constant, $h_p(\boldsymbol{z}) = h(\boldsymbol{z})$,

$$h_j(z_1, \dots, z_j) = \int_{R^{p-j}} h(z_1, \dots, z_j, \boldsymbol{w}) d\Phi_{p-j}(\boldsymbol{w}), \quad \text{and} \tag{2}$$

$$g_j(z_1, \dots, z_p) = e^{z_j^2/2} \int_{z_j}^{\infty} [h_j(z_1, \dots, z_{j-1}, w) - h_{j-1}(z_1, \dots, z_{j-1})] e^{-w^2/2} dw, \tag{3}$$

for $-\infty < z_1, \dots, z_p < \infty$ and $j = 1, \dots, p$.

Lemma 2.1. *For a function $h : R^p \to R$, let h_j and g_j be as in (2) and (3). If $h(\boldsymbol{z})$ depends on $\boldsymbol{z}$ only through $z_1, \dots, z_i$, then we have $h_j(z_1, \dots, z_j) = h(\boldsymbol{z})$ for all $j \ge i$; and consequently, $g_j(\boldsymbol{z}) \equiv 0$ for all $j > i$.*

The proof of the lemma is straightforward from (2) and (3), so we omit it. Now we shall define operators $U(h)$ and $U^2(h)$ associated with h. Let $U(h)$ denote the vector of the functions g_j in (3),

$$U(h) = [g_1, \dots, g_p]^T. \tag{4}$$

For example, for $\boldsymbol{z} \in \Re^p$, if $h(\boldsymbol{z}) = z_1$, then $Uh(\boldsymbol{z}) = (1, 0, \dots, 0)^T$ and if $h(\boldsymbol{z}) = \|\boldsymbol{z}\|^2$, then $Uh(\boldsymbol{z}) = \boldsymbol{z}$. Further, let U^2 denote the composition of U with itself:

$$U^2(h) = U(U(h)) = [U(g_1), \dots, U(g_p)]^T, \tag{5}$$

which is a $p \times p$ matrix whose jth row is $U(g_j)$ and g_j is as in (3). Since $g_j(\boldsymbol{z})$ defined in (3) depends only on $(z_1, \dots, z_j)$, by Lemma 2.1 we have that $U^2(h)$ is a lower triangular matrix. Next, define

$$V(h) = \frac{U^2(h) + (U^2(h))^T}{2} = \frac{1}{2}\{[U(g_1), \dots, U(g_p)]^T + [U(g_1), \dots, U(g_p)]\}. \tag{6}$$

Then, $V(h)$ is a symmetric matrix.

A function $f : R^p \to R$ is said to be almost differentiable if there exists a function $\nabla f : R^p \to R^p$ such that

$$f(\boldsymbol{z} + \boldsymbol{x}) - f(\boldsymbol{z}) = \int_0^1 \boldsymbol{x}^T \nabla f(\boldsymbol{z} + t\boldsymbol{x}) dt$$

for $\boldsymbol{z}, \boldsymbol{x} \in R^p$. We note that a continuously differentiable function f is almost differentiable with ∇f equal to its gradient. Let Γ be a measure of the form:

$$d\Gamma(\boldsymbol{z}) = f(\boldsymbol{z})\phi_p(\boldsymbol{z})d\boldsymbol{z}, \tag{7}$$

where f is a real-valued function (not necessarily non-negative) defined on R^p.

Proposition 2.1 (*Woodroofe–Stein's Identity*). *Suppose that $d\Gamma$ is defined as in (7), where f is almost differentiable. Let h be a real-valued function defined on R^p. Then, $\int h(\boldsymbol{z})d\Gamma(\boldsymbol{z}) = \int h(\boldsymbol{z})f(\boldsymbol{z})d\Phi_p(\boldsymbol{z})$ and*

$$\int h(\boldsymbol{z})f(\boldsymbol{z})d\Phi_p(\boldsymbol{z}) = \int h(\boldsymbol{z})d\Phi_p(\boldsymbol{z}) \cdot \int f(\boldsymbol{z})d\Phi_p(\boldsymbol{z}) + \int (Uh(\boldsymbol{z}))^T \nabla f(\boldsymbol{z})d\Phi_p(\boldsymbol{z}), \tag{8}$$

provided all the integrals are finite.

Proposition 2.1 was given by Woodroofe (1989). The last term on the right side of (8) can be further expanded. The derivation is in Proposition 2 of Woodroofe and Coad (1997) and Lemma 1 of Weng and Woodroofe (2000). We sketch the proof below as it is needed in Theorem 3.1. First assume that $\partial f/\partial z_j,\ j = 1, \ldots, p$ are almost differentiable. Next, by writing

$$(Uh(\boldsymbol{z}))^T \nabla f(\boldsymbol{z}) = \sum_{i=1}^{p} g_i(\boldsymbol{z}) \frac{\partial f(\boldsymbol{z})}{\partial z_i}$$

and applying (8) with h and f replaced by g_i and $\partial f/\partial z_i$, we obtain

$$\int g_i \frac{\partial f}{\partial z_i} d\Phi_p(\boldsymbol{z}) = \int g_i(\boldsymbol{z})d\Phi_p(\boldsymbol{z}) \cdot \int \frac{\partial f}{\partial z_i} d\Phi_p(\boldsymbol{z}) + \int (U(g_i))^T \nabla \left(\frac{\partial f}{\partial z_i}\right) d\Phi_p(\boldsymbol{z}), \tag{9}$$

provided all the integrals are finite. Then, summing up both sides of (9) over $i = 1, \ldots, p$ and making use of the notations (5) and (6), we can rewrite (8) as

$$\begin{aligned} \int h(\boldsymbol{z})f(\boldsymbol{z})d\Phi_p(\boldsymbol{z}) &= \int h(\boldsymbol{z})d\Phi_p(\boldsymbol{z}) \cdot \int f(\boldsymbol{z})d\Phi_p(\boldsymbol{z}) \\ &\quad + \left(\int (Uh(\boldsymbol{z}))d\Phi_p(\boldsymbol{z})\right)^T \int \nabla f(\boldsymbol{z})d\Phi_p(\boldsymbol{z}) + \int \mathrm{tr}\left[(Vh(\boldsymbol{z}))\nabla^2 f(\boldsymbol{z})\right] d\Phi_p(\boldsymbol{z}); \end{aligned} \tag{10}$$

where "tr" denotes the trace of a matrix, and $\nabla^2 f$ the Hessian matrix of f. Recall in (1) we denote by $\Phi_p h$ the integral of h with respect to Φ_p. Analogously, we denote by $\Phi_p(Uh)$ for integrals of a vector of functions Uh, $\Phi_p(Uh) = (\Phi_p(g_1), \ldots, \Phi_p(g_p))^T$; and similarly $\Phi_p(Vh)$ for a matrix function Vh.

Lemma 2.2. *The integrals $\Phi_p(Uh)$ and $\Phi_p(Vh)$ can be expressed as*

$$\Phi_p(Uh) = \int_{R^p} \boldsymbol{z}h(\boldsymbol{z})\Phi_p(d\boldsymbol{z}), \tag{11}$$

$$\Phi_p(Vh) = \int_{R^p} \frac{1}{2}(\boldsymbol{z}\boldsymbol{z}^T - I_p)h(\boldsymbol{z})\Phi_p(d\boldsymbol{z}). \tag{12}$$

Eq. (11) can be obtained by taking $f(z)$ in (8) as z_i, and Eq. (12) by taking $f(z)$ in (10) as $z_i z_j$ for $i, j = 1, 2, \ldots, p$.

Now suppose that the pdf of a random vector $\boldsymbol{Z}$ takes the form

$$p(\boldsymbol{z}) = c\phi_p(\boldsymbol{z})f(\boldsymbol{z}), \tag{13}$$

where c is the normalizing constant (not necessarily known). Clearly (13) is of the form (7). So, we can apply (8) and (10) to obtain the expectations:

$$E[h(\boldsymbol{Z})] = \Phi_p h + E\left\{[Uh(\boldsymbol{Z})]^T \frac{\nabla f(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right\}, \tag{14}$$

$$E[h(\boldsymbol{Z})] = \Phi_p h + (\Phi_p Uh)^T E\left[\frac{\nabla f(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right] + E\left\{\mathrm{tr}\left[Vh(\boldsymbol{Z})\frac{\nabla^2 f(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right]\right\}. \tag{15}$$

By (4), if $h(\boldsymbol{z}) = z_i$, then $Uh(\boldsymbol{z}) = \boldsymbol{e}_i$; and if $h(\boldsymbol{z}) = z_i z_j$, where $i \le j$, then $Uh(\boldsymbol{z}) = z_i \boldsymbol{e}_j$, where $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p\}$ denotes the standard basis for R^p. With these h in (14) and (15), we have the following result:

Lemma 2.3. *Let $\boldsymbol{Z}$ be a random vector whose pdf takes the form (13). Then,*

$$E(\boldsymbol{Z}) = E\left(\frac{\nabla f(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right) \quad \text{and} \quad E(Z_i Z_j) = \delta_{ij} + E\left[\frac{\nabla^2 f(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right]_{ij}, \tag{16}$$

where $[\cdot]_{ij}$ denotes the (i, j) component of a matrix and $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

2.2. Bayesian connections

Woodroofe–Stein's identity fits to Bayesian framework. Let $\ell(\boldsymbol{\theta})$ be the log-likelihood based on data D, where $\boldsymbol{\theta} \in R^p$. Assume that $\ell(\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$. Assume also that the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ satisfies $\nabla \ell(\hat{\boldsymbol{\theta}}) = 0$ and $-\nabla^2 \ell(\hat{\boldsymbol{\theta}})$ being positive definite, where ∇ indicates differentiation with respect to $\boldsymbol{\theta}$. Define a $p \times p$ matrix Σ and a p-dimensional random vector $\boldsymbol{Z}$ as

$$\Sigma^T \Sigma = -\nabla^2 \ell(\hat{\boldsymbol{\theta}}) \quad \text{and} \quad \boldsymbol{Z} = \Sigma(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}). \tag{17}$$

Consider a Bayesian model in which $\boldsymbol{\theta}$ has a prior density ξ. Then the posterior density of $\boldsymbol{\theta}$ given data D is $p(\boldsymbol{\theta}|D) \propto \exp(\ell(\boldsymbol{\theta}))\xi(\boldsymbol{\theta})$, and the posterior density of $\boldsymbol{Z}$ is

$$p(\boldsymbol{z}|D) \propto \exp[\ell(\boldsymbol{\theta}) - \ell(\hat{\boldsymbol{\theta}})]\xi(\boldsymbol{\theta}), \tag{18}$$

where the relation of $\boldsymbol{\theta}$ and $\boldsymbol{z}$ is given in (17). Now define

$$u(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \ell(\hat{\boldsymbol{\theta}}) + \frac{1}{2}\|\boldsymbol{z}\|^2 \quad \text{and} \quad f(\boldsymbol{z}) = \xi(\boldsymbol{\theta}(\boldsymbol{z}))\exp[u(\boldsymbol{\theta})].$$

So, (18) can be rewritten as

$$p(\boldsymbol{z}|D) \propto \phi_p(\boldsymbol{z})f(\boldsymbol{z}). \tag{19}$$

Since (19) is of the form (7), Proposition 2.1 can be applied.

In the context of Bayesian inference, Weng (2010) specialized Woodroofe–Stein's identity to the 1-dimensional case to obtain asymptotic expansions for marginal posterior densities. Let q_k denote Hermite polynomials, given by

$$q_k(x)\phi(x) = (-d/dx)^k \phi(x). \tag{20}$$

For instance, for $k = 0, 1, \dots, 4$ the Hermite polynomials are $q_0(x) = 1$, $q_1(x) = x$, $q_2(x) = x^2 - 1$, $q_3(x) = x^3 - 3x$, and $q_4(x) = x^4 - 6x^2 + 3$. The Hermite polynomials are orthogonal with respect to the standard normal pdf:

$$\int q_k(x)q_j(x)d\Phi(x) = k! \quad \text{if } k = j, \quad \text{and} \quad 0 \quad \text{if } k \neq j. \tag{21}$$

Let $\boldsymbol{Z} = (Z_1, \dots, Z_p)^T$ as in (17) and denote the posterior expectation given data D as $E(\cdot|D)$. Weng (2010) obtained the marginal posterior distribution of Z_i:

$$p(z_i) = \phi(z_i) + \sum_{\substack{j\in\{1,\dots,3s\} \\ j\neq 3s-1}} \frac{1}{j!} q_j(z_i)\phi(z_i)E[q_j(Z_i)|D] + O(t^{-\frac{3s+1}{2}+s}), \tag{22}$$

provided some regularity conditions hold. If we disregard the error order and arrange terms in (22) according to j, it becomes

$$p(x) \sim \sum_{j=0}^{\infty} c_j q_j(x)\phi(x), \quad \text{with } c_j = \frac{1}{j!}\int_{-\infty}^{\infty} p(x)q_j(x)dx. \tag{23}$$

The series (23) is commonly known as the Gram–Charlier series of type A (see Kendall and Stuart, 1977). As pointed out in Blinnikov and Moessner (1998), (23) is in fact a Fourier expansion of $p(x)/\phi(x)$ in Hermite polynomials:

$$\frac{p(x)}{\phi(x)} = \sum_j \left\langle \frac{p}{\phi}, \frac{q_j}{\sqrt{j!}} \right\rangle \frac{q_j(x)}{\sqrt{j!}},$$

where the inner product $\langle \cdot, \cdot \rangle$ of two functions f_1 and f_2 is defined as

$$\langle f_1, f_2 \rangle \equiv \int_{R^1} f_1(x)f_2(x)\phi(x)dx. \tag{24}$$

3. Main results

The method in Weng (2010) may readily be generalized to the Gram–Charlier series for multivariate densities, but not restricted to a posterior density as in Section 2.2.

3.1. Expansions for multivariate case

We shall expand the identities (8) and (10) to (28) in Theorem 3.1, extend Lemmas 2.2 and 2.3 to Propositions 3.2 and 3.3, respectively, and then establish the multivariate Gram–Charlier expansion in Theorem 3.2.

Some new notations are needed. For a function $h : R^p \to R$, let $h^{(0)} = h$, and write $U(h)$ and $U^2(h)$ in (4) and (5) as $h^{(1)}$ and $h^{(2)}$:

$$h^{(1)} = U(h) = [h_1^{(1)}, \ldots, h_p^{(1)}]^T, \qquad h^{(2)} = U^2(h) = \{h_{i_1 i_2}^{(2)} : i_1, i_2 = 1, \ldots, p\}.$$

So, $h^{(0)}$ is a function, $h^{(1)}$ is a $p \times 1$ vector (1-dimensional array) of functions, and $h^{(2)}$ is a $p \times p$ matrix (2-dimensional array) of functions whose jth row is $U(h_j^{(1)})$. Analogously, denote a $p \times p \times \cdots \times p$ array (k-dimensional array) of functions

$$h^{(k)} = U^k(h) = U(U^{k-1}h) = \{h_{i_1 i_2 \cdots i_k}^{(k)} : i_1, i_2, \ldots, i_k = 1, \ldots, p\},$$

where $U(h_{i_1 \cdots i_{k-1}}^{(k-1)}) = [h_{i_1 \cdots i_{k-1} 1}^{(k)}, h_{i_1 \cdots i_{k-1} 2}^{(k)}, \ldots, h_{i_1 \cdots i_{k-1} p}^{(k)}]^T$. For any function $f : R^p \to R$, denote by $f_{i_1,\ldots,i_s}^{(s)}(\boldsymbol{z})$ the sth derivative of f with respect to $z_{i_1}, \ldots, z_{i_s}$:

$$f_{i_1,\ldots,i_s}^{(s)}(\boldsymbol{z}) = \frac{\partial^s f(\boldsymbol{z})}{\partial z_{i_1} \cdots \partial z_{i_s}}. \tag{25}$$

With these definitions and notation (1), we can rewrite (8) and (10) as

$$\Phi_p(hf) = \Phi_p(h) \cdot \Phi_p(f) + \sum_{i_1} \Phi_p(h_{i_1}^{(1)} f_{i_1}^{(1)}) \tag{26}$$

$$\Phi_p(hf) = \Phi_p(h) \cdot \Phi_p(f) + \sum_{i_1} \Phi_p(h_{i_1}^{(1)}) \Phi_p(f_{i_1}^{(1)}) + \sum_{i_1,i_2} \Phi_p(h_{i_1 i_2}^{(2)} f_{i_1 i_2}^{(2)}), \tag{27}$$

where the summation is over $i_1, i_2 = 1, \ldots, p$.

Theorem 3.1. *Let s be any positive integer. Then,*

$$\begin{aligned} \Phi_p(hf) &= \Phi_p(h) \cdot \Phi_p(f) + \sum_{i_1} \Phi_p(h_{i_1}^{(1)}) \Phi_p(f_{i_1}^{(1)}) + \sum_{i_1,i_2} \Phi_p(h_{i_1 i_2}^{(2)}) \Phi_p(f_{i_1 i_2}^{(2)}) + \cdots \\ &\quad + \sum_{i_1,i_2,\ldots,i_{k-1}} \Phi_p(h_{i_1 i_2 \cdots i_{k-1}}^{(k-1)}) \Phi_p(f_{i_1 i_2 \cdots i_{k-1}}^{(k-1)}) + \sum_{i_1,i_2,\ldots,i_k} \Phi_p(h_{i_1 i_2 \cdots i_k}^{(k)} f_{i_1 i_2 \cdots i_k}^{(k)}), \end{aligned} \tag{28}$$

provided all the integrals exist.

Proof. It will be proved by induction. First, if $k = 1$ and 2, (28) reduces to (26) and (27). Next, suppose that (28) holds for $k = j - 1$. So, the last term in the equation will be $\sum_{i_1,i_2,\ldots,i_{j-1}} \Phi_p(h_{i_1 i_2 \cdots i_{j-1}}^{(j-1)} f_{i_1 i_2 \cdots i_{j-1}}^{(j-1)})$. It now suffices to show that (28) holds for $k = j$. To verify this, similar to (9) we apply (8) but with h and f replaced by $h_{i_1 i_2 \cdots i_{j-1}}^{(j-1)}$ and $f_{i_1 i_2 \cdots i_{j-1}}^{(j-1)}$, which gives

$$\Phi_p(h_{i_1 i_2 \cdots i_{j-1}}^{(j-1)} f_{i_1 i_2 \cdots i_{j-1}}^{(j-1)}) = \Phi_p(h_{i_1 i_2 \cdots i_{j-1}}^{(j-1)}) \Phi_p(f_{i_1 i_2 \cdots i_{j-1}}^{(j-1)}) + \sum_{i_j} \Phi_p(h_{i_1 i_2 \cdots i_j}^{(j)} f_{i_1 i_2 \cdots i_j}^{(j)}). \tag{29}$$

Then, summing up both sides of (29) over $i_j = 1, \ldots, p$ gives the desired result. $\square$

We remark here that the derivation of Woodroofe–Stein's identity involves interchanging the order of integration, and (28) is essentially repeated applications of interchange of the order of integration.

Some notations are needed for Proposition 3.2. Given a positive integer s and nonnegative integers r_i, $i = 1, \ldots, p$ for which $\sum_{j=1}^{p} r_j = s$, we define

$$I_{r_1,\ldots,r_p}^s = \left\{ (i_1, \ldots, i_s) : \sum_{j=1}^{s} 1_{\{i_j = k\}} = r_k, \text{ for } k = 1, \ldots, p \right\}, \tag{30}$$

which represents the set of all possible ways to split n objects into p distinct groups of sizes $r_1, \ldots, r_p$, respectively. For instance, if $p = 2$, then $I_{2,1}^3 = \{(1, 1, 2), (1, 2, 1), (2, 1, 1)\}$. Clearly, the elements in $I_{r_1,\ldots,r_p}^s$ are permutations of one another, and the number of elements in this set is the multinomial coefficient defined by

$$\binom{s}{r_1, r_2, \ldots, r_p} = \frac{s!}{r_1! r_2! \cdots r_p!}. \tag{31}$$

For $p = 2$, (31) becomes the binomial coefficient $C_{r_1}^s$.

Proposition 3.2. *Let s be any positive integer and $r_j, j = 1, \ldots, p$ be nonnegative integers for which $\sum_{j=1}^p r_j = s$. Then,*

$$\sum_{(i_1,\ldots,i_s)\in I^s_{r_1,r_2,\ldots,r_p}} \Phi h^{(s)}_{i_1\cdots i_s} = \frac{\binom{s}{r_1,r_2,\ldots,r_p}}{s!} \int_{R^p} \prod_{j=1}^p q_{r_j}(x_j) h(\boldsymbol{x}) d\Phi_p(\boldsymbol{x}), \tag{32}$$

where q_i are the Hermite polynomials defined in (20).

The proof is by induction and we omit it. This proposition extends Lemma 2.2, which contains the cases of $s = 1$ and 2. More explicitly, for $s = 1$, if we take $f(\boldsymbol{z}) = z_i$ in (26) for $1 \le i \le p$, it gives

$$\Phi_p h_i^{(1)} = \int x_i h(\boldsymbol{x}) \phi_p(\boldsymbol{x}) d\boldsymbol{x}.$$

This is the component-wise expression of (11), and is exactly a special case of (32) with $s = r_i = 1$ and $r_j = 0$ for $j \ne i$. For $s = 2$, we can take $f(\boldsymbol{z}) = z_i z_j$ in (27) for $1 \le i, j \le p$ to obtain expressions for $\Phi_p h^{(2)}_{ij}$. For instance, if $f(z) = z_i^2$, we obtain

$$\Phi_p h^{(2)}_{ii} = \int \frac{1}{2}(x_i^2 - 1)\phi_p(\boldsymbol{x}) d\boldsymbol{x}, \tag{33}$$

which is a special case of (32) when $s = r_i = 2$ and $r_j = 0$ for $j \ne i$; and if $f(z) = z_i z_j$ for $i \ne j$, we obtain

$$\Phi_p h^{(2)}_{ij} + \Phi_p h^{(2)}_{ji} = \int x_i x_j \phi_p(\boldsymbol{x}) d\boldsymbol{x}, \tag{34}$$

which is a special case of (32) when $s = 2, r_i = r_j = 1$, and $r_k = 0$ for $k \ne i, j$. Moreover, combining (33) and (34) gives (12).

Now suppose that the pdf of a random vector $\boldsymbol{Z}$ takes the form $p(\boldsymbol{z}) = c\phi_p(\boldsymbol{z}) f(\boldsymbol{z})$ as in (13), where the normalizing constant c is not necessarily tractable. Then, $E[h(\boldsymbol{Z})] = \Phi_p(chf)$ and by (28) we obtain

$$\begin{aligned} E[h(\boldsymbol{Z})] &= \Phi_p h + \sum_{i_1} \Phi_p h^{(1)}_{i_1} E\left[\frac{f^{(1)}_{i_1}(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right] + \sum_{i_1,i_2} \Phi_p h^{(2)}_{i_1 i_2} E\left[\frac{f^{(2)}_{i_1 i_2}(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right] + \cdots \\ &\quad + \sum_{i_1,\ldots,i_{s-1}} (\Phi_p h^{(s-1)}_{i_1\ldots i_{s-1}}) E\left[\frac{f^{(s-1)}_{i_1\ldots i_{s-1}}(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right] + \sum_{i_1,\ldots,i_s} E\left[h^{(s)}_{i_1\ldots i_s}(\boldsymbol{Z}) \frac{f^{(s)}_{i_1\ldots i_s}(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right], \end{aligned} \tag{35}$$

provided all the expectations exist. When $s = 1$ and 2, the above equations become (14) and (15); furthermore, (16) in Lemma 2.3 can be rewritten as

$$E\left(\frac{f_i^{(1)}(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right) = E(Z_i), \qquad E\left(\frac{f_{ii}^{(2)}(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right) = E(Z_i^2 - 1), \qquad E\left(\frac{f_{ij}^{(2)}(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right) = E(Z_i Z_j), \quad \text{if } i \ne j.$$

The proposition below extends Lemma 2.3 to higher moments. The proof is by induction and we omit it.

Proposition 3.3. *Suppose that the pdf of $\boldsymbol{Z}$ takes the form (13). Let s be any positive integer and $r_j, j = 1, \ldots, p$ be nonnegative integers for which $\sum_{j=1}^p r_j = s$. Suppose that $E(\prod_{j=1}^p Z_j^{r_j}) < \infty$. Then,*

$$E\left[\frac{f^{(s)}_{i_1\ldots i_s}(\boldsymbol{Z})}{f(\boldsymbol{Z})}\right] = E\left(\prod_{j=1}^p q_{r_j}(Z_j)\right). \tag{36}$$

The next theorem gives an expansion for the multivariate density. We need the following notations. For $s = 1, 2, \ldots,$ define

$$Q^{(s)}(\boldsymbol{z}) = \sum_{\substack{r_1,\ldots,r_p \ge 0 \\ \sum_{j=1}^p r_j = s}} \binom{s}{r_1, r_2, \ldots, r_p} \prod_{j=1}^p q_{r_j}(z_j) \cdot E\left(\prod_{j=1}^p q_{r_j}(Z_j)\right); \tag{37}$$

for instance, for $s = 1, 2$, we have

$$Q^{(1)}(\boldsymbol{z}) = \sum_{j=1}^p z_j E(Z_j) \quad \text{and} \quad Q^{(2)}(\boldsymbol{z}) = \sum_{j=1}^p (z_j^2 - 1)E(Z_j^2 - 1) + 2\sum_{i\ne j} z_i z_j E(Z_i Z_j).$$

Theorem 3.2. *Suppose that the pdf of $\boldsymbol{Z}$ takes the form (13). Let $h(\boldsymbol{z}) = 1_{\{Z_i \le z_i,\ i=1,\dots,p\}}$. Then, the pdf of $\boldsymbol{Z}$ can be expressed as*

$$p(\boldsymbol{z}) = \phi_p(\boldsymbol{z}) \left(1 + Q^{(1)}(\boldsymbol{z}) + \frac{1}{2!} Q^{(2)}(\boldsymbol{z}) + \frac{1}{3!} Q^{(3)}(\boldsymbol{z}) + \cdots + \frac{1}{(k-1)!} Q^{(k-1)}(\boldsymbol{z}) \right) + \sum_{i_1,\dots,i_k} \frac{\partial^p}{\partial z_1 \cdots \partial z_p} E \left\{ h^{(k)}_{i_1\dots i_k}(\boldsymbol{Z}) \frac{f^{(k)}_{i_1\dots i_k}(\boldsymbol{Z})}{f(\boldsymbol{Z})} \right\}, \tag{38}$$

provided all the integrals involved are finite.

Proof. First, plug this indicator function h into (35) and replace the terms on the right hand side of (35) by (32) and (36). This will give the cumulative distribution function (cdf) of $\boldsymbol{Z}$ in terms of its moments. Then, taking the pth partial derivative of the cdf with respect to $z_1, \dots, z_p$ yields the desired pdf. □

Corollary 3.1. *Let $p = 1$. Then, (38) reduces to the Gram–Charlier series in (23):*

$$p(z) \sim \phi(z) \sum_{j=0}^{\infty} \frac{1}{j!} q_j(z) E(q_j(Z)). \tag{39}$$

Note that an arbitrary pdf $p(\boldsymbol{x})$ fits into the form (13) by writing

$$p(\boldsymbol{x}) = \phi_p(\boldsymbol{x}) \frac{p(\boldsymbol{x})}{\phi_p(\boldsymbol{x})}.$$

3.2. Modified series

Cramér (1946) gave the convergence criteria (see Section 1 for details) for the Gram–Charlier series. Though the series has poor convergence properties, Cramér commented that in many practical applications we were concerned whether the truncated series gives a good approximation, instead of the convergence properties of the series.

To give a better approximation, we proposed to first consider the Gram–Charlier series of the pdf of a scaled variable Y defined as

$$Y = \alpha Z, \quad 0 \le \alpha \le 1, \tag{40}$$

and then transform the series to that of $p(z)$. The idea is that since $p(y)$ will fall to zero faster than $p(z)$, the Gram–Charlier series of $p(y)$ may have better convergence property. Once we obtain the Gram–Charlier series of $p(y)$,

$$p(y) \sim \phi(y) \sum_{j=0}^{\infty} \frac{1}{j!} q_j(y) E(q_j(Y)), \tag{41}$$

we can easily transform it to obtain a Gram–Charlier type expansion for $p(z)$:

$$p(z) \sim \alpha\phi(\alpha z) \sum_{j=0}^{\infty} \frac{1}{j!} q_j(\alpha z) E(q_j(\alpha Z)), \tag{42}$$

where $E(q_j(\alpha Z)) = E(q_j(Y))$. The above equation is actually a Fourier series of $p(z)/\phi(\alpha z)$ in the *scaled* orthonormal polynomials $\{q_j(\alpha z)\sqrt{\alpha j!}\}$, with the inner product defined in (24).

For the multivariate case, we define variable variables Y_i, $i = 1, \dots, p$ as

$$Y_i = \alpha_i Z_i, \quad 0 \le \alpha_i \le 1,$$

and proceed as (42) to obtain a multivariate Gram–Charlier series for $p(\boldsymbol{z})$. The choice of α_i can be flexibly determined by users. Typically, the more the non-Gaussian requires the smaller the α_i.

4. Concluding remarks

We have derived the Gram–Charlier type expansion for multivariate cases by Woodroofe–Stein's identity, and proposed a modified series in terms of scaled Hermite polynomials. Some questions deserve further study. First, it is desirable to effectively derive the coefficients in the multivariate Edgeworth series. Secondly, since the expansion is useful to recover a distribution by its moments, it is of interest to explore whether the analytical expansions and sampling methods such as MCMC can benefit each other.

Acknowledgments

The author would like to thank the reviewers for helpful comments. The author is partially supported by the Ministry of Science and Technology of Taiwan and the National Science Council of Taiwan.

References

Amari, S.-i., Cichocki, A., Yang, H.H., et al., 1996. A new learning algorithm for blind signal separation. Adv. Neural Inf. Process. Syst. 757–763.
Baghishania, H., Mohammadzadeh, M., 2012. Asymptotic normality of posterior distributions for generalized linear mixed models. J. Multivariate Anal. 111, 66–77.
Barndorff-Nielsen, O.E., Cox, D.R., 1989. Asymptotic Techniques for Use in Statistics. Chapman & Hall.
Blinnikov, S., Moessner, R., 1998. Expansions for nearly Gaussian distributions. Astron. Astrophys. Suppl. Ser. 130, 193–205.
Coad, D.S., Woodroofe, M., 1996. Corrected confidence intervals after sequential testing with applications to survival analysis. Biometrika 83, 763–777.
Comon, P., 1994. Independent component analysis, a new concept? Signal Process. 36 (3), 287–314.
Cramér, H., 1946. Mathematical Methods of Statistics. Princeton University Press, Princeton.
Csörgő, T., Hegyi, S., 2000. Model independent shape analysis of correlations in 1, 2 or 3 dimensions. Phys. Lett. B 489 (1), 15–23.
Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer, New York.
James, W., Stein, C., 1961. Estimation with quadratic loss. In: Proc. Fourth Berkeley Symp. Math. Statist. Prob., vol. 1. pp. 361–379.
Kendall, M., Stuart, A., 1977. The advanced theory of statistics, fourth ed. In: Distribution Theory, vol. 1. Griffin, London, p. 1. 1977.
Rubinstein, M., 1998. Edgeworth binomial trees. J. Derivatives 5 (3), 20–27.
Sargan, J.D., 1975. Gram-Charlier approximations applied to t ratios of k-class estimators. Econometrica 327–346.
Sargan, J.D., 1976. Econometric estimators and the Edgeworth approximation. Econometrica 421–448.
Scherrer, R.J., Bertschinger, E., 1991. Statistics of primordial density perturbations from discrete seed masses. Astrophys. J. 381, 349–360.
Stein, C., 1981. Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135–1151.
Van Der Marel, R.P., Franx, M., 1993. A new method for the identification of non-gaussian line profiles in elliptical galaxies. Astrophys. J. 407, 525–539.
Van Hulle, M.M., 2005. Multivariate edgeworth-based entropy estimation. Mach. Learn. Signal Process. 2005, 311–316.
Weng, R.C., 2010. A Bayesian Edgeworth expansion by Stein's Identity. Bayesian Anal. 5 (4), 741–764.
Weng, R.C., 2003. On Stein's identity for posterior normality. Statist. Sinica 13, 495–506.
Weng, R.C., Lin, C.-J., 2011. A Bayesian approximation method for online ranking. J. Mach. Learn. Res. 12, 267–300.
Weng, R.C., Woodroofe, M., 2000. Integrable expansions for posterior distributions for multiparameter exponential families with applications to sequential confidence levels. Statistica Sinica 10, 693–713.
Weng, R.C., Woodroofe, M., 2006. Approximate confidence sets for a stationary AR(p) process. J. Statist. Plann. Inference 136, 2719–2745.
Woodroofe, M., 1989. Very weak expansions for sequentially designed experiments: linear models. Ann. Statist. 17, 1087–1102.
Woodroofe, M., 1992a. Integrable expansions for posterior distributions for one-parameter exponential families. Statistica Sinica 2, 91–111.
Woodroofe, M., 1992b. Estimation after sequential testing: A simple approach for a truncated sequential probability ratio test. Biometrika 79, 347–353.
Woodroofe, M., Coad, D.S., 1997. Corrected confidence sets for sequentially designed experiments. Statistica Sinica 7, 53–74.