A generalized Isserlis theorem for location mixtures of Gaussian random vectors

C.Vignat*

E.P.F.L., L.T.H.I., Lausanne, Switzerland

Abstract

In a recent paper, Michalowicz et al. provide an extension of Isserlis theorem to the case of a Bernoulli location mixture of a Gaussian vector. We extend here this result to the case of any location mixture of Gaussian vector; we also provide an example of the Isserlis theorem for a "scale location" mixture of Gaussian, namely the $d-$dimensional generalized hyperbolic distribution.

Keywords: Isserlis theorem, normal-variance mixture, generalized hyperbolic distribution

1. Introduction

Isserlis theorem, as discovered by Isserlis [1] in 1918, allows to express the expectation of a monomial in an arbitrary number of components of a zero mean Gaussian vector $X \in \mathbb{R}^d$ in terms of the entries of its covariance matrix only. Before providing in Thm 1 the slightly generalized version of Isserlis theorem due to Withers [3], we introduce the following notations: for any set $A = \{\alpha_1, \ldots, \alpha_N\}$ of integers such that $1 \le \alpha_i \le d$ and any vector $X \in \mathbb{R}^d$, we use the multi-index notation and denote

$$X_A = \prod_{\alpha_i \in A} X_{\alpha_i}$$

with the convention that for the empty set

$$X_\emptyset = 1.$$

A pairing in a set A is a partition of A into disjoint pairs. We denote by $\Pi(A)$ the set of all pairings σ in A: note that $\Pi(A)$ is empty if A has an odd number of elements. For a given $\sigma \in \Pi(A)$, we denote by A/σ the set $\{i;\, \sigma = (i, \sigma(i))\}$; finally, $\sum_A \prod \mathbb{E}(X_i X_j)$ denotes the sum

$$\sum_{\sigma \in \Pi(A)} \prod_{i \in A/\sigma} \mathbb{E}\left(X_{\alpha_i} X_{\alpha_{\sigma(i)}}\right).$$

In other words, for a given pairing σ in the set A, we compute the product of all possible moments $\mathbb{E}(X_i X_j)$ where i and j are paired by σ; then, $\sum_A \prod \mathbb{E}(X_i X_j)$ denotes the sum of these products

*on leave from L.S.S., Supélec, Orsay, France

Email address: `christophe.vignat@epfl.ch` (C.Vignat)

over all possible pairings in A. As an example

$$\sum_{\{1,1,2,4\}}\prod\mathbb{E}\left(X_iX_j\right)=\mathbb{E}\left(X_1^2\right)\mathbb{E}\left(X_2X_4\right)+2\mathbb{E}\left(X_1X_2\right)\mathbb{E}\left(X_1X_4\right).$$

A general form of Isserlis theorem, due to Withers, is as follows.

Theorem 1. *If $A=\{\alpha_1,\ldots,\alpha_{2N}\}$ is a set of integers such that $1\le\alpha_i\le d$, $\forall i\in[1,2N]$ and $X\in\mathbb{R}^d$ is a Gaussian vector with zero mean then*

$$\mathbb{E}X_A=\sum_A\prod\mathbb{E}\left(X_iX_j\right) \tag{1}$$

Moreover, if $A=\{\alpha_1,\ldots,\alpha_{2N+1}\}$ then, under the same assumptions,

$$\mathbb{E}X_A=0.$$

For example, choosing $\alpha_i=i,\ 1\le i\le 4$ yields the well-known identity

$$\mathbb{E}\left(X_1X_2X_3X_4\right)=\mathbb{E}\left(X_1X_2\right)\mathbb{E}\left(X_3X_4\right)+\mathbb{E}\left(X_1X_3\right)\mathbb{E}\left(X_2X_4\right)+\mathbb{E}\left(X_1X_4\right)\mathbb{E}\left(X_2X_3\right).$$

However, indices α_i need not be distinct: for example, choosing $\alpha_i=1,\ 1\le i\le 4$ yields

$$\mathbb{E}X_1^4=3\mathbb{E}X_1^2.$$

Several extensions of this result have been provided recently: in [3], Withers extends Isserlis theorem to the case of noncentral Gaussian vectors and relates the result with multivariate Hermite polynomials; in [4], a general formula for Gaussian scale mixtures, and more generally for elliptically distributed vectors is derived; it is applied to the computation of moments of the uniform distribution on the sphere. In [5], Isserlis theorem is extended to the computation of the moments of linear combinations of independent Student-t vectors. In [6], Isserlis theorem is extended to Gaussian matrix mixtures, i.e. random vectors of the form

$$X=AN$$

where N is a standard Gaussian vector in $\mathbb{R}^d$ and A is a $(d\times d)$ random matrix. Let us also mention the reference [8] where the author tackles the computational complexity of formula (1), using Magnus lemma to replace a product of n variables by sums of polynomials of degree n in these variables.

Recently, Michalowicz et al. [2] addressed the case of Gaussian location mixtures: they provided an extension of Isserlis theorem to the case of a random vector $X\in\mathbb{R}^d$ with probability density

$$f_X\left(x\right)=\frac{1}{2}\phi_R\left(x+\mu\right)+\frac{1}{2}\phi_R\left(x-\mu\right) \tag{2}$$

where

$$\phi_R\left(x\right)=\frac{1}{\left|2\pi R\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}x^tR^{-1}x\right)$$

is the $d-$variate Gaussian density with zero mean and covariance matrix R.

In the following, we give a new and simple proof of the result by Michalowicz et al., adopting a formalism that allows us to extend their results to the general case of an arbitrary Gaussian location mixture. We also provide an extension of these results to the case of a scale-location mixture of Gaussian.

2. Extensions of the result by Michalowicz et al.

A key observation is that the random vector X with density (2) reads

$$X = \epsilon\mu + \zeta \tag{3}$$

where ϵ is a Bernoulli random variable ($\Pr\{\epsilon = -1\} = \Pr\{\epsilon = 1\} = \frac{1}{2}$), $\zeta \in \mathbb{R}^d$ is a zero mean Gaussian vector and equality is in the sense of distributions. This stochastic representation allows to prove easily a generalized version of the main result of [2], namely

Theorem 2. *If $X \in \mathbb{R}^d$ is distributed according to (2) and $A = \{\alpha_1, \dots, \alpha_{2N}\}$ with $1 \le \alpha_i \le d$ then*

$$\mathbb{E}X_A = \sum_{k=0}^{N} \sum_{\substack{S\subset A\\|S|=2k}} \mu_S \sum_{A\backslash S} \prod \mathbb{E}\left(\zeta_i\zeta_j\right).$$

If $A = \{\alpha_1, \dots, \alpha_{2N+1}\}$ then $\mathbb{E}X_A = 0$.

The simplified proof we propose is as follows: by (3),

$$\mathbb{E}X_A = \mathbb{E}\left(\epsilon\mu + N\right)_A$$

and since the product $(a+b)_A$ can be expanded as

$$(a+b)_A = \sum_{k=0}^{2N} \sum_{\substack{S\subset A\\|S|=k}} a_S b_{A\backslash S},$$

we deduce that

$$\mathbb{E}X_A = \sum_{k=0}^{2N} \sum_{\substack{S\subset A\\|S|=k}} \mu_S \mathbb{E}\left(\epsilon^k\right) \mathbb{E}\left(\zeta_{A\backslash S}\right).$$

By Isserlis theorem, the expectation of the product of an odd number of centered Gaussian random variables ζ_i is equal to zero so that this expression simplifies to

$$\sum_{k=0}^{N} \sum_{\substack{S\subset A\\|S|=2k}} \mu_S \mathbb{E}\left(\epsilon^{2k}\right) \mathbb{E}\left(\zeta_{A\backslash S}\right).$$

Since ϵ is Bernoulli distributed, all its even moments are equal to 1; moreover, since ζ has zero mean, by Isserlis theorem, $E\left(\zeta_{A\backslash S}\right) = \sum_{A\backslash S}\prod E\left(\zeta_i\zeta_j\right)$ and we obtain

$$\mathbb{E}X_A = \sum_{k=0}^{N} \sum_{\substack{S\subset A\\|S|=2k}} \mu_S \sum_{A\backslash S} \prod \mathbb{E}\left(\zeta_i\zeta_j\right)$$

which is the desired result. The case where A has an odd number of elements is equally simple.

We note that Theorem 2 can be also easily deduced using generating functions as done in [3, Theorem 1.1] who proves a version of Wick's theorem for a Gaussian vector with mean $\mu \neq 0$: choosing a Bernoulli randomized version of this mean as in (3) yields the result.

3. The general case of Gaussian location mixture

With the useful representation (3), we can generalize the preceding result to any kind of location mixture of Gaussian: namely, we consider a random vector $X \in \mathbb{R}^d$ that reads

$$X = \mu + \zeta \tag{4}$$

where ζ is a zero-mean Gaussian vector in $\mathbb{R}^d$, independent of the random vector $\mu \in \mathbb{R}^d$ with probability distribution F_μ; note that the vector μ may be discrete - taking values μ_i with probabilities p_i - or not, but we don't need to assume the existence of a density f_μ. In the discrete case, the density of X reads

$$f_X(x) = \sum_{i=0}^{+\infty} p_i \phi_R(x - \mu_i);$$

and in the most general case,

$$f_X(x) = \int_{\mathbb{R}^d} \phi_R(x - \mu) \, dF_\mu(\mu).$$

We now state our main theorem.

Theorem 3. *Assume that $X \in \mathbb{R}^d$ follows model (4) and that all the first-order moments $m_k = \mathbb{E}\mu_k$ of μ exist. Then if $A = \{\alpha_1, \ldots, \alpha_{2N+\epsilon}\}$, with $\epsilon \in \{0, 1\}$,*

$$\mathbb{E}X_A = \sum_{k=0}^{N} \sum_{\substack{S \subset A \\ |S|=2k+\epsilon}} \mathbb{E}(\mu_S) \sum_{A \setminus S} \prod \mathbb{E}(\zeta_i \zeta_j) \tag{5}$$

We remark that if all elements α_i of A are different and if the vector μ has independent components, this expression can be further simplified to

$$\mathbb{E}X_A = \sum_{k=0}^{N} \sum_{\substack{S \subset A \\ |S|=2k+\epsilon}} (\mathbb{E}\mu)_S \sum_{A \setminus S} \prod \mathbb{E}(\zeta_i \zeta_j), \tag{6}$$

noting the difference between $\mathbb{E}(\mu_S) = \mathbb{E}\prod_{\alpha_i \in S} \mu_{\alpha_i}$ in (5) and $(\mathbb{E}\mu)_S = \prod_{\alpha_i \in S} \mathbb{E}\mu_{\alpha_i}$ in (6).

The proof is as follows.

Proof 1. *From (4), we deduce*

$$\mathbb{E}X_A = \sum_{k=0}^{2N+\epsilon} \sum_{\substack{S \subset A \\ |S|=k}} \mathbb{E}(\mu_S) \, \mathbb{E}(\zeta_{A \setminus S}).$$

Since the cardinality of $A \setminus S$ is $2N + \epsilon - k$, $\mathbb{E}(\zeta_{A \setminus S}) = 0$ unless $|S| = k$ has the same parity as ϵ, in which case it is equal to $\sum_{A \setminus S} \prod \mathbb{E}(\zeta_i \zeta_j)$, hence formula (5). Formula (6) is easily deduced from formula (5) assuming that the components of μ are independent and that all elements of A are distinct.

We now provide a further generalization of Isserlis theorem by considering a Gaussian vector with both random scale and location parameters.

4. A Normal variance-mean mixture application

The generalized $d-$dimensional hyperbolic distribution was introduced by Barndorff-Nielsen in 1978 [7]. It is the distribution of a random vector that reads

$$X = \mu + \sigma^2 \Delta \beta + \sigma \Delta^{\frac{1}{2}} \zeta \tag{7}$$

where μ and β are two deterministic vectors in $\mathbb{R}^d$, Δ is a deterministic $(d \times d)$ matrix with $|\Delta| = 1$, ζ is a standard Gaussian vector in $\mathbb{R}^d$ and σ^2 is a scalar random variable that follows the Generalized Inverse Gaussian $GIG\left(\psi, \chi, \lambda\right)$ distribution

$$f_{\psi,\chi,\lambda}\left(x\right) = \frac{\left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}}}{2K_\lambda\left(\sqrt{\psi\chi}\right)} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right),\ x > 0 \tag{8}$$

with parameters $\psi > 0,\ \chi > 0$ and $\lambda \in \mathbb{R}$. We note that in (7), the GIG random variable σ^2 appears both as a scale and location parameter of the Gaussian vector, hence the "normal variance-mean mixture" name. From the stochastic representation (7), we derive a version of the Isserlis theorem as follows.

Theorem 4. *If $X \in \mathbb{R}^d$ is a generalized hyperbolic vector as in (7) and $A = \{\alpha_1, \ldots, \alpha_{2N+\epsilon}\}$ with $\epsilon \in \{0, 1\}$ then*

$$\mathbb{E}X_A = \sum_{\substack{0 \le l \le N \\ 0 \le p \le 2l+\epsilon}} \sum_{\substack{T \subset S \subset A \\ |T|=p, |S|=2l+\epsilon}} \mu_T \gamma_{S \setminus T} m_{N+l-p+\epsilon} \sum_{A\setminus S} \prod \mathbb{E}Z_i Z_j$$

where $\gamma = \Delta\beta$, where Z is a centered Gaussian vector with covariance matrix Δ and

$$m_l = \mathbb{E}\sigma^{2l} = \left(\frac{\psi}{\chi}\right)^{-\frac{l}{2}} \frac{K_{\lambda+l}\left(\sqrt{\psi\chi}\right)}{K_\lambda\left(\sqrt{\psi\chi}\right)}.$$

Proof 2. *Assuming first $\epsilon = 0$, we have*

$$\mathbb{E}X_A = \mathbb{E}\sum_{l=0}^{N} \sum_{\substack{S \subset A \\ |S|=2l}} \left(\mu + \sigma^2\gamma\right)_S \left(\sigma Z\right)_{A\setminus S} = \sum_{l=0}^{N} \sum_{\substack{S \subset A \\ |S|=2l}} \mathbb{E}\left(\sigma^{2N-2l}\left(\mu + \sigma^2\gamma\right)_S\right) \mathbb{E}Z_{A\setminus S}$$

with

$$\mathbb{E}Z_{A\setminus S} = \sum_{A\setminus S} \prod \mathbb{E}Z_i Z_j$$

and

$$\mathbb{E}\sigma^{2N-2l}\left(\mu + \sigma^2\gamma\right)_S = \sum_{p=0}^{2l} \sum_{\substack{T \subset S \\ |T|=p}} \mu_T \gamma_{S\setminus T} \mathbb{E}\left(\sigma^2\right)^{N+l-p}.$$

The moment of order l of the GIG random variable σ^2 can be easily computed from (8) as

$$m_l = \left(\frac{\psi}{\chi}\right)^{-\frac{l}{2}} \frac{K_{\lambda+l}\left(\sqrt{\psi\chi}\right)}{K_\lambda\left(\sqrt{\psi\chi}\right)},$$

hence the result. The case $\epsilon = 1$ follows the same steps.

References

[1] Isserlis L., On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables, Biometrika, 1918, 12, 134–139

[2] Michalowicz J.V., Nichols J.M., Bucholtz F., Olson C.C., A general Isserlis theorem for mixed-Gaussian random variables, Statistics & Probability Letters, August 2011, 81-8,1233-1240

[3] Withers C.S., The moments of the multivariate normal, Bulletin of the Australian Mathematical Society, 1985, 32, 103-107

[4] Vignat C. and Bhatnagar S., An extension of Wick's theorem, Statistics & Probability Letters, 2008, 78-15, 2404-2407

[5] Repetowicz P. and Richmond P., The Wick theorem for non-Gaussian distributions and its application for noise filtering of correlated q-Exponentially distributed random variables, unpublished, arXiv:math-ph/0411020 v1, Nov 2004

[6] Grigelionis B., On the Wick theorem for mixtures of centered Gaussian distributions, Lithuanian Mathematical Journal, 2009, 49-4, 372-380

[7] Barndorff-Nielsen O., Hyperbolic Distributions and Distributions on Hyperbolae, Scandinavian Journal of Statistics, 1978, 5-3, 151-157

[8] Kan R., From Moments of Sum to Moments of Product, Journal of Multivariate Analysis, 99, 542-554, 2008.